Inferability of Unions of Certain Closed Set Systems

Yuichi Kameda

2011

Abstract

In this thesis we study inferability in the limit from positive data for the classes of bounded and unbounded unions of certain class of languages. In order to show inferability, we put an emphasis on a characteristic set of a given language.

This thesis consists of two parts: one for bounded unions, and the other for unbounded unions. In both cases, the notion of characteristic sets plays an important role to show inferability and to construct learning algorithms concretely.

We consider a class of languages called a closed set system $\mathcal{C}$. For bounded unions, we consider the bounded union $\cup^{\leq k}\mathcal{C}$ of closed set systems $\mathcal{C}$ and we assume the following three conditions: (1) $\mathcal{C}$ is Noetherian, (2) $\mathcal{C}$ is compact, and (3) a characteristic set of a given closed set in $\cup^{\leq k}\mathcal{C}$ can be constructed from its characteristic set in $\mathcal{C}$. Then we have a learning algorithm of $\cup^{\leq k}\mathcal{C}$ concretely under these conditions, by using the notion of hypergraphs. We give two examples of closed set systems that satisfy the above three conditions, and apply our algorithm to them.

For unbounded unions, we consider the unbounded union $\mathcal{C}^\star$ of closed set systems such that there exists an algorithm for generating a characteristic set consisting of one element. We construct a learning algorithm of $\mathcal{C}^\star$ concretely, and give two examples. Furthermore, we investigate relation between those examples and transaction databases, and attempt to apply our algorithm to the transaction databases.

0.1 Introduction

In this thesis we study inferability of classes of bounded or unbounded unions of certain closed set systems, and we consider concrete learning algorithms of them.

Machine learning is originally to study algorithms to simulate the mechanism that human beings learn from their experiences. Today, after development of information technology, machine learning is more used as theoretical basics of getting patterns or tendencies behind given large quantity of data, and expected to be progressed further.

There are a few models in machine learning. In this thesis our approach is called identification in the limit from positive data. Identification in the limit from positive data is defined as follows:
Given an enumerable set U (elements of U are called words) and a class $\mathcal{L}$ of subsets of U (elements of this class are called languages). Every language is labeled by the elements of some enumerable set $\mathcal{H}$, called a hypothesis space. For a language L, an enumeration of L is called a positive data of L. Let P is an algorithm that runs as follows. Let L is an unknown given language of $\mathcal{L}$ and suppose that a positive data σ of L is given. When the elements of σ are presented one by one, then each time P outputs a hypothesis of language that seems to be indicated by the positive data. If the outputs of P converge to a hypothesis that indicates a given language, then we say that P identifies L from σ in the limit. If P identifies L from σ in the limit for every $L \in \mathcal{L}$ and every positive data σ of L, then we say that P identifies $\mathcal{L}$ from positive data in the limit.

The idea of identification in the limit is introduced by Gold [7]. In 1980, Angluin [1] gave a necessary and sufficient condition for identifying a given language from positive data ((**EC1**) in Theorem 1.2). Angluin also presented an instance of a class of languages that is inferable from positive data for the first time. After that, some sufficient conditions more convenient to deal with than Angluin's have been presented, such as existence of characteristic sets ((**C2**) in Definition 1.3) and finite elasticity ((**C3**)). However, those conditions are not always appropriate for constructing a learning algorithm concretely. Therefore, concrete learning algorithms are needed for the languages that decided identifiable by such conditions. A class of unions of languages, that is we considered in this thesis, is a typical instance in such circumstances. Our goal is to construct learning algorithms for classes of unions of languages called closed set systems concretely under certain conditions.

A class of unions of languages is the class of subsets of U that can be expressed as the set unions of finite number of languages. If the number of languages is bounded uniformly, the class is called bounded unions of languages. Otherwise, then it is called unbounded unions. A class of unions of languages can be regarded as a class that deals with the combinations of languages. It is, however, not easy to handle a class of unions of languages. In particular, in case of unbounded unions, the condition for identifying a class of unbounded unions of languages seems to be fairly strong by the results of de Brecht et al. [2].

We deal with closed set systems as classes of unions of languages. A closed set system is a class constructed by using a mapping called a closure operator. One can say that closed set systems are appropriate to deal with some algebraic objects, such as vector spaces, as the target of learning.

In this thesis we consider both bounded and unbounded unions. In both cases, we suggest new conditions for constructing learning algorithms and construct learning algorithms concretely by using the conditions. We also present a few instances that satisfy the conditions and apply the algorithms to them.

This thesis goes as follows: In Chapter 1 we review definitions and theorems about the inferability from positive data and closed set systems. We summarize preliminaries from algebra, such as the definition of ideals of polynomial ring, in Chapter 2. Then in Chapter 3, we investigate that how the algebraic preliminaries are connected to the theory of inductive inference. In Chapter 4 we consider the case of classes of bounded unions of languages. In §4.1 we introduce the condition $(*)$ for constructing an algorithm learning a class of bounded unions of languages, and construct a learning algorithm by making use of $(*)$ concretely. We present two instances of applications of the learning algorithm in §§4.2 and 4.3. Chapter 5 proceeds similarly for the case of unbounded unions. We introduce the condition $(\star)$ and give a learning algorithm for unbounded unions in §5.1, and then we present instances in §§5.2 – 5.4. We conclude our argument in Chapter 6.

Chapter 1

Preliminaries from Inductive Inference

1.1 Inferability from Positive Data

In this article, a *language* L is a subset of some countable set U such that L is expressed $L(G)$ by some finite expression G. We call this finite expression a *hypothesis*. A set of all hypotheses $\mathcal{H}$ is called a *hypothesis space*. Let $\mathcal{L}$ be the set of all languages $\{L(G) \mid G \in \mathcal{H}\}$. We assume that $\mathcal{L}$ is *uniformly recursive*, that is, there is a recursive function $f(w, G)$ such that $f(w, G) = 1$ if and only if $w \in L(G)$ for every $w \in U$ and $G \in \mathcal{H}$.

A *positive data* (or *positive presentation*) of $L \in \mathcal{L}$ is an infinite sequence $\sigma : s_1, s_2, \ldots$ of elements of L such that $L = \{s_1, s_2, \ldots\}$. An *inference algorithm* M is that:

- M receives incrementally elements of a positive data σ of a language,
- M outputs a hypothesis $G_n \in \mathcal{H}$ when M receives n-th element of σ.

$\mathcal{L}$ is *inferable in the limit from positive data* if there exists an inference algorithm M satisfies that for all $L \in \mathcal{L}$ and an arbitrary positive data of L, the output sequence of M converges to a hypothesis G such that $L(G) = L$.

It is known that inferability of a class of languages $\mathcal{L}$ can be characterized as follows:

Definition 1.1 Let L be a language of $\mathcal{L}$. A finite subset S of L is called a *finite tell-tale* of L in $\mathcal{L}$ if $L' \in \mathcal{L}$ includes S, then $L' \not\subset L$. In other words, L is a minimal language in the class $\{L' \in \mathcal{L} \mid S \subseteq L'\}$ with respect to set inclusion.

Theorem 1.2 ([1]) *$\mathcal{L}$ is inferable in the limit from positive data if and only if:* (**EC1**) *there exists a procedure to enumerate elements of a finite tell-tale of every $L \in \mathcal{L}$.*

Moreover, there are some sufficient conditions for inferability of $\mathcal{L}$.

Definition 1.3 1. Let L be a language of $\mathcal{L}$. A finite subset $S \subseteq L$ is called a *characteristic set* of L in $\mathcal{L}$ if $L' \in \mathcal{L}$ includes S, then $L \subseteq L'$, that is, L is the minimum language in the class $\{L' \in \mathcal{L} \mid S \subseteq L'\}$.
2. $\mathcal{L}$ has *infinite elasticity* if there exists an infinite sequence $w_0, w_1, \ldots$ of elements of U and infinite sequence $L_1, L_2, \ldots$ of languages of $\mathcal{L}$ such that, for every n, L_n contains $w_0, \ldots, w_{n-1}$ but w_n. $\mathcal{L}$ has *finite elasticity* if it does not have infinite elasticity.
3. $\mathcal{L}$ has *finite thickness* if the set $\{L \in \mathcal{L} \mid w \in L\}$ is finite for any $w \in U$.

Theorem 1.4 ([10],[13]) *Consider the following conditions:*
(**C2**) *For each L in $\mathcal{L}$, there exists a characteristic set of L in $\mathcal{L}$,*
(**C3**) *$\mathcal{L}$ has finite elasticity,*
(**C4**) *$\mathcal{L}$ has finite thickness.*
Then it holds that;

$$(\mathbf{C4}) \Rightarrow (\mathbf{C3}) \Rightarrow (\mathbf{C2}) \Rightarrow (\mathbf{EC1}).$$

In particular, (**C2**), (**C3**) *and* (**C4**) *are sufficient conditions for inferability of $\mathcal{L}$.*

Let $\mathcal{L}'$ be a subclass of $\mathcal{L}$. Then, it clearly holds that:

Proposition 1.5 *1. If $\mathcal{L}$ is inferable from positive data, then $\mathcal{L}'$ is.*
2. If $L \in \mathcal{L}$ has a characteristic set in L, then the characteristic set is also a characteristic set of L in L'.

1.2 Closed Set System

First we define a closure operator. Let 2^U be the power set of U.

Definition 1.6 A mapping $C : 2^U \to 2^U$ is called a *closure operator* if C satisfies:
(CO1) $X \subseteq C(X)$,
(CO2) $X \subseteq Y \Rightarrow C(X) \subseteq C(Y)$, and
(CO3) $C(C(X)) = C(X)$
for each $X, Y \in 2^U$.

A set $X \subseteq U$ is called *closed* if $X = C(X)$. A *closed set system* $\mathcal{C}$ is the class of all closed sets of a closure operator. A closed set system can be characterized by the property intersection closed as the following way.

Proposition 1.7 *1. Let $\mathcal{C}$ be a closed set system. $\mathcal{C}$ is intersection closed, that is, the intersection of arbitrary number of closed sets of $\mathcal{C}$ is an element of $\mathcal{C}$.*
2. Let $\mathcal{F}$ be a class of subsets of U. Suppose that $\mathcal{F}$ is intersection closed and, for each $S \subseteq U$, there is at least one $X \in \mathcal{F}$ such that $S \subseteq X$. Then there is a closure operator C such that the closed set system associated with C is $\mathcal{F}$.

Proof. 1. Let $\{X_i\} \subseteq \mathcal{C}$. Since $\cap X_i \subseteq X_i$, $C(\cap X_i) \subseteq C(X_i) = X_i$ for every i. This implies $C(\cap X_i) \subseteq \cap X_i$. On the other hand, $C(\cap X_i) \supseteq \cap X_i$ by (CO1). Therefore $C(\cap X_i) = \cap X_i$, thus $\cap X_i$ is closed.
2. For $A \subseteq U$, we define $C(A) = \cap\{X \in \mathcal{F} \mid A \subseteq X\}$. Since $\mathcal{F}$ is intersection closed, $C(A)$ is in $\mathcal{F}$. It is easy to verify that C becomes a closure operator.

Remark 1.8 In a closed set system, the union of closed sets is not necessarily closed.

Remark 1.9 According to the proposition above, a subclass of a closed set system is closed set system if it is intersection closed and there exists an element of the subclass for each subset of U.

In the following, we regard a closed set system $\mathcal{C}$ as a class of languages and assume that it is recursive.

Definition 1.10 Let X is a closed set of $\mathcal{C}$. If there is a finite set $Y \subseteq U$ such that $X = C(Y)$, then X is called a *finitely generated closed set.*

Lemma 1.11 ([2]) *Let $X = C(Y)$ be a closed set. The followings are equivalent:*
1. Y is finite,
2. Y is a finite tell-tale of X, and
3. Y is a characteristic set of X.

An immediate consequence of Lemma 1.11 and Theorem 1.2 is as follows:

Corollary 1.12 *$\mathcal{C}$ is inferable from positive data if and only if every closed set is finitely generated.*

Proof. ($\Rightarrow$) If $\mathcal{C}$ is inferable, then $\mathcal{C}$ satisfies (**EC1**) by Theorem 1.2, so each closed set X of $\mathcal{C}$ has a finite tell-tale Y. By Lemma 1.11, it holds $C(Y) = X$ and X is finitely generated.
($\Leftarrow$) Let X be an arbitrary closed set of $\mathcal{C}$ and Y be a finite generating set of X. By Lemma 1.11, Y is a characteristic set of X. Hence $\mathcal{C}$ satisfies the condition (**C2**). By applying Theorem 1.4, it is shown that $\mathcal{C}$ is inferable from positive data.

Next we define a Noetherian closed set system.

Definition 1.13 A closed set system $\mathcal{C}$ is *Noetherian* if there is no closed sets $C_1, C_2, \ldots$ of $\mathcal{C}$ such that $C_1 \subsetneq C_2 \subsetneq \ldots$, that is, $\mathcal{C}$ contains no infinite strictly ascending chain of closed sets.

It is known that

Theorem 1.14 ([2, Theorem 7]) *A closed set system $\mathcal{C}$ is Noetherian if and only if $\mathcal{C}$ has finite elasticity.*

Hence it follows that:

Corollary 1.15 *Let $\mathcal{C}$ be a Noetherian closed set system.*
1. Every closed set C of $\mathcal{C}$ is finitely generated.
2. Every closed set C of $\mathcal{C}$ has a characteristic set.
3. $\mathcal{C}$ is inferable from positive data.

From Remark 1.9, a intersection closed subclass of a closed set system inherits the properties such as inferability. Henceforth, we regard a intersection closed subclass of a closed set system as a closed set system.

1.3 Bounded and Unbounded Unions of Closed Set Systems

We start this section by defining the bounded union of languages.

Definition 1.16 Let $\mathcal{L}$ be a class of languages and k be a fixed positive integer. The bounded union $\cup^{\leq k}\mathcal{L}$ of $\mathcal{L}$ is the class defined by

$$\cup^{\leq k}\mathcal{L} = \{L_1 \cup \ldots \cup L_m \mid m \leq k, L_i \in \mathcal{L}\ (i = 1, \ldots, m)\}.$$

It is known that

Theorem 1.17 ([19]) *If $\mathcal{L}$ has finite elasticity, then $\cup^{\leq k}\mathcal{L}$ also has finite elasticity. In particular, $\cup^{\leq k}\mathcal{L}$ is inferable from positive data.*

If L is a language of $\mathcal{L}$, then L can be regarded as an element of $\cup^{\leq k}\mathcal{L}$. We use the next lemma in the proof of Lemma 4.7.

Lemma 1.18 *Let L be a language of $\mathcal{L}$ and S be a characteristic set of L in $\cup^{\leq k}\mathcal{L}$. Then S becomes a characteristic set of L in $\mathcal{L}$.*

Proof. Obviously $S \subseteq L$ and S is finite. By the definition of characteristic set, every element $L_1 \cup \ldots \cup L_m$ of $\cup^{\leq k}\mathcal{L}$ that includes S satisfies $L \subseteq L_1 \cup \ldots \cup L_m$. Since m and $L_1 \cup \ldots \cup L_m$ are arbitrary, this implies that every element L' of $\mathcal{L}$ that includes S satisfies $L \subseteq L'$. Therefore S is a characteristic set of L in $\mathcal{L}$.

We need the following definition later.

Definition 1.19 $\cup^{\leq k}\mathcal{L}$ is said to be *compact* if it satisfies the following condition:
For each $m \leq k$ and $L, L_i \in \mathcal{L}$ $(i = 1, \ldots, m)$, if $L \subseteq L_1 \cup \ldots \cup L_m$, then there exists i_0 such that $L \subseteq L_{i_0}$.

Next, the unbounded union of languages is defined as follows:

Definition 1.20 ([15]) Let $\mathcal{L}$ be a language. The unbounded union $\mathcal{L}^*$ of $\mathcal{L}$ is the class

$$\mathcal{L}^* = \{L_1 \cup \ldots \cup L_m \mid \forall m \in \mathbb{N}, L_i \in \mathcal{L} \ (i = 1, \ldots, m)\}.$$

where $\mathbb{N}$ denotes the set of all positive integers $\{1, 2, \ldots\}$.

Remark 1.21 For an element of $\cup^{\leq k}\mathcal{L}$ or $\mathcal{L}^*$, we always assume that the expression $L_1 \cup \ldots \cup L_m$ is not redundant, that is, $L_i \not\subseteq L_j$ for any $i, j (i \neq j)$ in the following.

In [2], de Brecht et al. gave a necessary and sufficient condition for unbounded unions of closed set systems to be inferable.

Theorem 1.22 ([2]) *Let $\mathcal{L}$ be a closed set system. $\mathcal{L}^*$ is inferable from positive data if and only if every closed set $L \in \mathcal{L}$ is equal to a union of finitely many closed sets generated by a single element.*

1.4 Transaction Databases

Let I be a countable set $\{p_1, p_2, \ldots\}$ and we regard I as the set of items.

Definition 1.23 A finite subset of I is called itemset. A transaction database $\mathcal{D}$ over I is a sequence of itemsets $X_1, X_2, \ldots$. Elements of $\mathcal{D}$ are called transactions of $\mathcal{D}$.

For a subset $X \subseteq I$, the support of X in $\mathcal{D}$ is defined by $\{i \in \mathbb{N} \mid X \subseteq X_i\}$ and denoted by $t(X)$. Note that $t(\emptyset) = \mathbb{N}$. t is a mapping that maps 2^I to $2^{\mathbb{N}}$. By definition clearly holds that

Lemma 1.24 *Let $X, Y \subseteq I$. If $X \subseteq Y$ then $t(X) \supseteq t(Y)$.*

Definition 1.25 The number of elements of $t(X)$ is called the frequency of X and denoted by $|X|$. An itemset $X \subseteq I$ is called closed if $|Y| < |X|$ for every $Y \supsetneq X$. To avoid confusions, we call this DB-closed here.

Remark 1.26 By Lemma 1.24, one can express the definition of DB-closed as follows: X is DB-closed if $t(Y) \subsetneq t(X)$ for every $Y \supsetneq X$.

Note that every DB-closed itemset X is finite. In fact, if X is infinite, then $t(X)$ is empty. So X can not be closed. Next we define another mapping $\iota : 2^{\mathbb{N}} \to 2^I$. For a set of indexes $A \subseteq \mathbb{N}$, $\iota(A) = \{p_i \mid p_i \in X_a$ for every $a \in A\}$. If $A = \emptyset$, we define $\iota(\emptyset) = I$. Similarly to t, it holds:

Lemma 1.27 *Let $A, B \subseteq \mathbb{N}$. If $A \subseteq B$ then $\iota(A) \supseteq \iota(B)$.*

It is known that:

Proposition 1.28 *$\iota \circ t : 2^I \to 2^I$ is a closure operator on I.*

Proof. Let $X, Y \subseteq I$. (CO1) Let p_i in X. Since every element of $t(X)$ includes p_i, $p_i \in \iota(t(X))$.
(CO2) Assume that $X \subseteq Y$. Let p_i in $\iota \circ t(X)$. $p_i \in \iota(t(X))$ implies that, for every $a \in t(X)$, $p_i \in X_a$. Now $t(Y) \subseteq t(X)$ by Lemma 1.24, so we have that $p_i \in X_a$ for every $a \in t(Y)$. Therefore $p_i \in \iota(t(Y))$.
(CO3) Since (CO1) and (CO2), $\iota \circ t(X) \subseteq \iota \circ t(\iota \circ t(X))$. Let $p_i \in \iota \circ t(\iota \circ t(X))$. Then $p_i \in X_a$ for every $a \in t(\iota(t(X)))$. Now, one can show that $A \subseteq t \circ \iota(A)$ in a similar way of (CO1). Thus $t(\iota(t(X))) \supseteq t(X)$, so we have $p_i \in X_a$ for every $a \in t(X)$. This means $p_i \in \iota(t(X))$.

Remark 1.29 Similarly, one can show that $t \circ \iota : 2^{\mathbb{N}} \to 2^{\mathbb{N}}$ becomes a closure operator on $\mathbb{N}$.

Then the closure operator $\iota \circ t$ makes DB-closed sets closed. More precisely,

Proposition 1.30 *Let X is an itemset. Then, X is DB-closed if and only if X is closed with respect to $\iota \circ t$.*

Proof. $\Rightarrow$) Suppose that X is DB-closed but closed. Let $Y = \iota(t(X))$. Since X is not closed, $X \subsetneq Y$. Then $t(X) \supsetneq t(Y)$ since X is DB-closed. By Remark 1.29, we have $t(X) \supsetneq t(Y) = t(\iota(t(X))) \supseteq t(X)$. This is contradiction.
$\Leftarrow$) Suppose that $X = \iota(t(X))$ and X is not DB-closed. Since X is not DB-closed, there exists $Y \supsetneq X$ such that $t(X) = t(Y)$. $t(X) = t(Y)$ implies that $X = \iota(t(X)) = \iota(t(Y)) \supseteq Y$. This contradicts to $X \subsetneq Y$.

Chapter 2

Preliminaries from Algebra

2.1 Ideals of Polynomial ring

We refer to [6] for details in this section. We denote the set of all polynomials of n variables with $\mathbb{Q}$-coefficients by $\mathbb{Q}[x_1, \ldots, x_n]$.

Definition 2.1 A nonempty subset I of $\mathbb{Q}[x_1, \ldots, x_n]$ is called an *ideal* if it satisfies the following two conditions:
1. For each $f, g \in I$, $f \pm g \in I$, and
2. For each $f \in I$ and $h \in \mathbb{Q}[x_1, \ldots, x_n]$, $hf \in I$.

We denote the set of all ideals by $\mathcal{I}$.

Definition 2.2 For a finite subset $F = \{f_1, \ldots, f_r\} \subset \mathbb{Q}[x_1, \ldots, x_n]$, we define the ideal generated by $f_1, \ldots, f_r$, which is denoted by $\langle f_1, \ldots, f_r \rangle$ or $\langle F \rangle$, as follows:

$$\langle F \rangle := \Big\{ \sum_{i=1}^{r} h_i f_i \mid \ h_i \in \mathbb{Q}[x_1, \ldots, x_n] \Big\}.$$

Similarly, for a subset S of $\mathbb{Q}[x_1, \ldots, x_n]$ that is not necessarily finite,

$$\langle S \rangle := \Big\{ \sum_{\text{finite}} h_f f \mid f \in S, h_f \in \mathbb{Q}[x_1, \ldots, x_n] \Big\}.$$

S is called a generating set of I if $I = \langle S \rangle$.

An ideal I is called *finitely generated* if there exists a finite subset $F \subset I$ such that $I = \langle F \rangle$. The following theorem is well known as the consequence of Hilbert's basis theorem in algebra.

Theorem 2.3 ([6]) *1. Every ideal of* $\mathbb{Q}[x_1, \ldots, x_n]$ *is finitely generated.*
2. There is no infinite ascending chain of ideals of $\mathbb{Q}[x_1, \ldots, x_n]$*. That is, If* I_j*'s are ideals and* $I_1 \subseteq I_2 \subseteq \ldots$*, then there exists* N *such that* $I_N = I_{N+1} = I_{N+2} = \ldots$*.*

Let $\mathcal{M}$ be the set of all monomials.

Definition 2.4 An ideal I is called a *monomial ideal* if there exists a set of monomials $F \subseteq \mathcal{M}$ such that $I = \langle F \rangle$.

Monomial ideals are characterized as follows:

Proposition 2.5 *Let* I *be an ideal of* $\mathbb{Q}[x_1, \ldots, x_n]$*. The following four conditions are equivalent:*
(a) I *is a monomial ideal,*
(b) I *is generated by the set of all monomials in* I*,*
(c) for each $f \in I$*, every monomial occurring in* f *is also in* I*, and*
(d) I *is generated by a set of finitely many monomials.*

We denote the class of all monomial ideals by $\mathcal{MI}$. By Proposition 2.5, it clearly holds that:

Lemma 2.6 *Let* I *be a monomial ideal. 1. Let* m *be a monomial. Suppose that* $I = \langle m_1, \ldots, m_s \rangle$*. Then* $m \in I$ *if and only if there exists* i *such that* $m_i | m$*.*
2. Let f *be a polynomial. Then* $f \in I$ *if and only if all monomials that appear in* f *are in* I*.*

Next we review the theory of Groebner basis. For the details of Groebner basis, see [6]. For that purpose we first consider a monomial ordering.

Definition 2.7 Let $<$ be an order of $\mathcal{M}$. $<$ is called a *monomial ordering* if it satisfies:
1. for each $m, u, v \in \mathcal{M}$, $u < v \Rightarrow mu < mv$, and
2. for all $m \in \mathcal{M}$, $1 < m$.

Example 2.8 Let u and v are monomials and suppose that $u = x_1^{u_1} \dots x_n^{u_n}$ and $v = x_1^{v_1} \dots x_n^{v_n}$. The lexicographic order $<_{\text{lex}}$ on $\mathcal{M}$ is defined as follows: $u <_{\text{lex}} v$ if, $u_{i_0} < v_{i_0}$ where $i_0 = \min\{i \mid u_i \neq v_i\}$. Then it is easy to verify that $<_{\text{lex}}$ is a monomial ordering.

In the following, we consider a fixed monomial ordering $<$.

Definition 2.9 Let f be a polynomial. The *leading term* (or *initial term*) of f is the maximum monomial appears in f with respect to $<$, denoted by $LT(f)$.

Definition 2.10 Let $I \in \mathcal{I}$. The *initial ideal* of I is defined by $\langle \{LT(f) \mid f \in I\} \rangle$, and it is denoted by $LT(I)$.

We define a Groebner basis of I as follows.

Definition 2.11 Let I be an ideal. A finite generating set $\{g_1, \dots, g_s\}$ of I is called a *Groebner basis* of I if $LT(I) = \langle LT(g_1), \dots, LT(g_s) \rangle$.

Then it is known that:

Theorem 2.12 ([6]) *For every $I \in \mathcal{I}$, there exists a Groebner basis of I.*

There is a special Groebner basis called reduced.

Definition 2.13 A Groebner basis $\{g_1, \dots, g_s\}$ of I is called *reduced* if it satisfies (a) the coefficients of all leading terms of g_i's are 1, and (b) for each $i \neq j$, every term of g_i is not divisible by $LT(g_j)$.

Theorem 2.14 ([6]) *For every ideal $I \in \mathcal{I}$, there uniquely exists the reduced Groebner basis. Moreover, there is an algorithm that computes the reduced Groebner basis for given I.*

2.2 Infinite Dimensional Vector Spaces

Let V be a vector space over the set of rational numbers $\mathbb{Q}$. First we state the definition of infinite dimensional vector space.

Definition 2.15 A subset $S \subseteq V$ is called *linearly dependent* if there exists a finite subset $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\} \subseteq S$ and $c_1, \dots, c_n \in \mathbb{Q}$ that at least one of c_i's is not zero, such that $c_1\boldsymbol{v}_1 + \dots + c_n\boldsymbol{v}_n = 0$. If S is not linearly dependent, then S is called *linearly independent.*

We denote the cardinality of S by $\sharp(S)$.

Definition 2.16 A vector space V is called *finite dimensional* if there exists a positive integer n such that all subsets consist of more than n elements are linearly dependent. V is called *infinite dimensional* if there exists linearly independent subsets $S_n \subseteq V$ such that $\sharp(S_n) = n$ for every n.

In case of infinite dimensional, a basis of V is defined as follows.

Definition 2.17 A subset $\mathcal{B} \subseteq V$ is called a *basis* of V if it satisfies:
1. each $\boldsymbol{v} \in V$ can be uniquely written by a linear combination of finite number of elements of $\mathcal{B}$, and
2. each $\boldsymbol{g} \in \mathcal{B}$ can not be written by any linear combination of finite number of elements of $\mathcal{B} \setminus \{\boldsymbol{g}\}$.

It is known that

Proposition 2.18 *Every basis of V has the same cardinality. That is to say, if $\mathcal{B}$ and $\mathcal{B}'$ are bases, then $\sharp(\mathcal{B}) = \sharp(\mathcal{B}')$. The dimension of V is the cardinality of bases of V.*

The following statement can be shown by using Zorn's lemma.

Proposition 2.19 *Every vector space has a basis.*

In the following we assume that V has countable basis. We fix one basis $\mathcal{B} = \{\boldsymbol{g_1}, \boldsymbol{g_2}, \ldots\}$ of V.

Remark 2.20 Note that V is enumerable. For instance, let i be positive integer and its prime factorization be $i = \prod_j p_j^{c_j}$, where p_j denotes the j-th prime number. Since $\mathbb{Q}$ is enumerable, one can index all rational numbers, so $\mathbb{Q}$ can be expressed as $\{q_1, q_2, \ldots\}$. We define $\boldsymbol{v}_i = \sum_j q_{c_j} \boldsymbol{g}_j$. Then $\boldsymbol{v}_i \in V$ and $V = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots\}$.

The followings are defined as the same as the case of finite dimensional.

Definition 2.21 A *subspace* of V is a subset of V that becomes a vector space with respect to the same addition and scalar multiplication of V.

Definition 2.22 Let S be a subset of V. The subspace generated by S, denoted by $\langle S \rangle$, is the minimum subspace of V that includes S with respect to set inclusion. $\langle S \rangle$ can be written as follows:

$$\langle S \rangle = \left\{ \sum_{\text{finite}} c_i \boldsymbol{v}_i \mid c_i \in \mathbb{Q}, \boldsymbol{v}_i \in S \right\}.$$

Definition 2.23 Let V and W be vector spaces. A mapping $T : V \to W$ is called a *linear mapping* if it satisfies:
1. for every $\boldsymbol{v}, \boldsymbol{v'} \in V$, $T(\boldsymbol{v} + \boldsymbol{v'}) = T(\boldsymbol{v}) + T(\boldsymbol{v'})$, and
2. for every $\boldsymbol{v} \in V$ and $c \in \mathbb{Q}$, $T(c\boldsymbol{v}) = cT(\boldsymbol{v})$.
If $V = W$ then T is called a *linear transformation*.

Definition 2.24 Let T be a linear transformation on V and W be a subspace of V. W is called T-*invariant* if $T(W) \subseteq W$.

Here we give a few examples of infinite dimensional vector space with countable basis.

Example 2.25 Sequence space. Let V be the set of sequences $\{(x_n)_{n=1,2,\ldots} \mid x_n \in \mathbb{Q}\}$. Addition and scalar multiplication are defined as follows:

$$(x_n) + (y_n) = (x_n + y_n),\ \ c(x_n) = (cx_n)\ \ ((x_n), (y_n) \in V, c \in \mathbb{Q}).$$

Then V becomes an infinite dimensional vector space. A countable basis of V is given by $\{(e_n^{(k)}) \mid k = 1, 2, \ldots\}$, where if $k = n$ then $e_n^{(k)} = 1$, else $e_n^{(k)} = 0$.

Example 2.26 Polynomial ring over $\mathbb{Q}$. Clearly $\mathbb{Q}[x_1, \ldots, x_n]$ together with usual addition and scalar multiplication is an infinite dimensional vector space, and the set of all monomials $\mathcal{M}$ can be regarded as a countable basis. From this viewpoint, ideals of $\mathbb{Q}[x_1, \ldots, x_n]$ become subspaces. In particular, a monomial ideal of $\mathbb{Q}[x_1, \ldots, x_n]$ is a subspace that generated by some subset M of the set of monomials $\mathcal{M}$, by Proposition 2.5. On $\mathbb{Q}[x_1, \ldots, x_n]$, one can define various linear transformations. For instance, multiplication by some $f \in \mathbb{Q}[x_1, \ldots, x_n]$, substitution of $a \in \mathbb{Q}$ for some variable x_i, or derivation by some x_i are linear transformations.

Example 2.27 The class of periodic function. Let $\mathbb{R}$ be the set of all real numbers. A function $f : \mathbb{R} \to \mathbb{R}$ is called periodic if there exists a $p \in \mathbb{R}$ such that $f(x + p) = f(x)$ for all $x \in \mathbb{R}$. For simplicity we fix $p = 2\pi$ here. Let V is the set of all periodic function with $f(x+2\pi) = f(x)$. Then V becomes an infinite dimensional vector space over $\mathbb{R}$. The theory of Fourier series implies that the set $\{1, \sin(nx), \cos(nx) \mid n = 1, 2, \ldots\}$ is a basis of V. We will see this example in Example 5.13 again.

Chapter 3

Closed Set Systems and Algebra

3.1 Closed Set Systems and Ideals of Polynomial Ring

In this section we show that the class $\mathcal{I}$ of ideals of polynomial ring can be regarded as a closed set system, and investigate what properties $\mathcal{I}$ has. We also consider the class of monomial ideals $\mathcal{MI}$. At first we show that the operation $\langle\cdot\rangle$ satisfies the condition of closure operator.

Lemma 3.1 *The mapping $F \mapsto \langle F\rangle$ can be regarded as a closure operator on $\mathbb{Q}[x_1, \ldots, x_n]$.*

Proof. (CO1) and (CO2) are obvious. (CO3) $\langle F\rangle \subseteq \langle\langle F\rangle\rangle$ since (CO1) and (CO2). Let $f \in \langle\langle F\rangle\rangle$. Suppose that f is expressed by the form

$$f = \sum_{i=1}^{m} h_i f_i, \text{ where } f_i \in \langle F\rangle, h_i \in \mathbb{Q}[x_1, \ldots, x_n].$$

Since each f_i is an element of $\langle F\rangle$, f_i can be written in the form

$$f_i = \sum_{j=1}^{m_i} k_{i,j} f_{i,j}, \text{ where } f_{i,j} \in F, k_{i,j} \in \mathbb{Q}[x_1, \ldots, x_n].$$

Thus $f = \sum_{i,j} h_i k_{i,j} f_{i,j}$ $(h_i k_{i,j} \in \mathbb{Q}[x_1, \ldots, x_n], f_{i,j} \in F)$, so $f \in \langle F\rangle$.

Remark 3.2 One can easily show that the intersection of arbitrary number of ideals is also an ideal. Thus one can define a closure operator according to Proposition 1.7(2). Then this closure operator is identical with $\langle\cdot\rangle$.

According to Lemma 3.1, we have:

Proposition 3.3 *$\mathcal{I}$ and $\mathcal{MI}$ are closed set systems.*

Proof. It is clear for $\mathcal{I}$. For $\mathcal{MI}$, it suffices to show that (1) $\mathcal{MI}$ is intersection closed, and (2) for every $S \in \mathbb{Q}[x_1, \ldots, x_n]$, there exists $I \in \mathcal{MI}$ such that $S \subseteq I$, since Remark 1.9. (2) is easy: in fact, every S is included in $\langle 1\rangle = \mathbb{Q}[x_1, \ldots, x_n]$. We show (1). Let $\{I_a \mid a \in A\}$ be monomial ideals. Here we denote the set of all monomials in I_a, that is $I_a \cap \mathcal{M}$, by M_a. Then we claim that

$$\bigcap_{a\in A} I_a = \Big\langle \bigcap_{a\in A} M_a \Big\rangle.$$

To show that, suppose $f \in \cap I_a$. Let M_f be the set of all monomials occur in f. Since Proposition 2.5(c), $M_f \subseteq M_a$ for every $a \in A$, so $M_f \subseteq \cap M_a$. Thus $f \in \langle \cap M_a\rangle$, and then we have $\cap I_a \subseteq \langle \cap M_a\rangle$. On the other hand, if we suppose $f \in \langle \cap M_a\rangle$, then we can express f by $f = \sum f_i m_i$, where $f_i \in \mathbb{Q}[x_1, \ldots, x_n]$ and $m_i \in \cap M_a$. This means that $f \in I_a$ for each $a \in A$. Hence $f \in \cap I_a$, and thus $\cap I_a \supseteq \langle \cap M_a\rangle$. Therefore the claim holds, and then we have that $\mathcal{MI}$ is intersection closed.

The closure operator associated with $\mathcal{MI}$ is not clear from the proof of the above proposition. The following lemma says how the closure operator can be written.

Lemma 3.4 *Let C be the closure operator associated with $\mathcal{MI}$. Then,*
1. For $M \subseteq \mathcal{M}$, $C(M) = \langle M\rangle$.
2. For $S \subseteq \mathbb{Q}[x_1, \ldots, x_n]$, $C(S) = \langle M_S\rangle$, where $M_S = \{m \in \mathcal{M} \mid m$ occurs in some $f \in S\}$.

Proof. 1. According to Proposition 1.7(2), $C(M) = \cap\{I \in \mathcal{MI} \mid M \subseteq I\}$. Here we put $\mathcal{A}_M = \{I \in \mathcal{MI} \mid M \subseteq I\}$. $C(M) \subseteq \langle M\rangle$ since $\langle M\rangle \in \mathcal{A}_M$. By the definition of closure operator, $M \subseteq I$ implies $\langle M\rangle \subseteq I$. Hence every $I \in \mathcal{A}_M$ includes $\langle M\rangle$. Thus $C(M) = \cap\mathcal{A}_M \supseteq \langle M\rangle$.
2. By Proposition 2.5, if a monomial ideal I includes S, then I must include M_S. Since $S \subseteq C(S)$, $M_S \subseteq C(S)$, and thus $\langle M_S\rangle \subseteq \langle C(S)\rangle$. Now $C(S)$ is an ideal and $\langle\cdot\rangle$ is a closure operator, $\langle C(S)\rangle = C(S)$ holds. Therefore

$C(S) \supseteq \langle M_S \rangle$. On the other hand, $\langle M_S \rangle \in \mathcal{A}_S$ since $S \subseteq \langle M_S \rangle$. Thus $C(S) = \cap \mathcal{A}_S \subseteq \langle M_S \rangle$.

By applying Proposition 2.3 to $\mathcal{I}$ and $\mathcal{MI}$, we have:

Lemma 3.5 *1. All elements of $\mathcal{I}$ and $\mathcal{MI}$ have characteristic sets.*
2. $\mathcal{I}$ and $\mathcal{MI}$ have finite elasticity. In particular, $\mathcal{I}$ and $\mathcal{MI}$ are Noetherian closed set systems.

Proof. 1. For $\mathcal{I}$, it is obvious from Theorem 1.14(1). For $\mathcal{MI}$, it holds since Proposition 2.5(d).
2. It immediately follows from Theorem 1.14 and Proposition 2.3(2).

Remark 3.6 Neither $\mathcal{I}$ nor $\mathcal{MI}$ satisfies (**C4**). In fact, every ideal includes 0.

Therefore it holds that:

Corollary 3.7 *Both $\mathcal{I}$ and $\mathcal{MI}$ are inferable from positive data.*

For the classes of bounded unions, by Theorem 1.17, next holds:

Corollary 3.8 *$\cup^{\leq k}\mathcal{I}$ and $\cup^{\leq k}\mathcal{MI}$ are inferable from positive data.*

We consider a learning algorithm of $\cup^{\leq k}\mathcal{I}$ in §4.2.

Furthermore, in case of monomial ideals, one can show that $\mathcal{MI}^{\star}$ is inferable from positive data. We will show it and give a learning algorithm in §5.3.

3.2 Closed Set Systems and Vector Spaces

At first we consider a finite dimensional vector space. Let V_n be a n-dimensional vector space over $\mathbb{Q}$. Then it holds:

Lemma 3.9 ([17, Lemma 1]) *Let $M \leq n$. There are vectors $\{\boldsymbol{v}_i \in V_n \mid i = 1, \ldots, M\}$ such that any n vectors of $\boldsymbol{v}_i$'s are linearly independent.*

We give an instance of Lemma 3.9.

Example 3.10 We consider a fixed basis $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$ of V_n. Let $c_1, \ldots, c_M$ be mutually distinct rational numbers and let $\boldsymbol{v}_i = \boldsymbol{e}_1 + c_i\boldsymbol{e}_2 + \ldots + c_i^{n-1}\boldsymbol{e}_n$. Then any n of $\boldsymbol{v}_i$'s are linearly independent. In fact, for each combination of n indexes $i_1 < i_2 < \ldots < i_n$, one can write by using matrix as follows:

$$\begin{pmatrix} \boldsymbol{v}_{i_1} \\ \boldsymbol{v}_{i_2} \\ \vdots \\ \boldsymbol{v}_{i_n} \end{pmatrix} = \begin{pmatrix} 1 & c_{i_1} & \ldots & c_{i_1}^{n-1} \\ 1 & c_{i_2} & \ldots & c_{i_2}^{n-1} \\ \vdots & \vdots & & \vdots \\ 1 & c_{i_n} & \ldots & c_{i_n}^{n-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{e}_1 \\ \boldsymbol{e}_2 \\ \vdots \\ \boldsymbol{e}_n \end{pmatrix}.$$

Let $C_{i_1,\ldots,i_n}$ be the matrix at the right of above equation. This matrix is known as a Vandermonde's matrix, and it holds that

$$\det(C_{i_1,\ldots,i_n}) = \prod_{j<k}(c_{i_j} - c_{i_k}).$$

Since we assume that c_i's are mutually distinct, we have $\det(C_{i_1,\ldots,i_n}) \neq 0$, and thus $\{\boldsymbol{v}_{i_1}, \ldots, \boldsymbol{v}_{i_n}\}$ is linearly independent.

After this we consider infinite dimensional vector spaces. Let V be an infinite dimensional vector space with countable basis over $\mathbb{Q}$.

Lemma 3.11 *A mapping $\langle\cdot\rangle$ is a closure operator on V.*

Proof. (CO1) and (CO2) are obvious. (CO3) Let $S \subseteq V$. Clearly $\langle S\rangle \subseteq \langle\langle S\rangle\rangle$. Let $\boldsymbol{v} \in \langle\langle S\rangle\rangle$. By definition, there is some $c_i \in \mathbb{Q}$ and $\boldsymbol{v}_i \in \langle S\rangle$ such that $\boldsymbol{v} = \sum_{\text{finite}} c_i\boldsymbol{v}_i$. Similarly, there exists some $c_j^{(i)} \in \mathbb{Q}$ and $\boldsymbol{v}_{i,j}^{(i)} \in S$ such that $\boldsymbol{v}_i = \sum_{\text{finite}} c_j^{(i)}\boldsymbol{v}_{i,j}^{(i)}$. Thus $\boldsymbol{v}$ can be expressed as $\boldsymbol{v} = \sum_i\sum_j c_ic_j^{(i)}\boldsymbol{v}_{i,j}^{(i)}$. Since the summation on the right of above formula is finite, we have $\boldsymbol{v} \in \langle S\rangle$, and therefore $\langle\langle S\rangle\rangle \subseteq \langle S\rangle$.

We denote the class of all finite dimensional subspaces of V by $\mathcal{SV}$. Although $\langle\cdot\rangle$ is a closure operator, $\langle\cdot\rangle$ is not the closure operator associated with $\mathcal{SV}$. In fact, $\langle\cdot\rangle$ makes infinite dimensional subspaces closed. In addition, $\mathcal{SV}$ does not become a closed set system from the method of Proposition 1.7, because if $S \subseteq V$ satisfies that $\langle S\rangle$ is infinite dimensional, then there is no $W \in \mathcal{SV}$ includes S. Then we consider $\mathcal{SV} \cup \{V\}$.

Lemma 3.12 *$\mathcal{SV} \cup \{V\}$ is intersection closed.*

Proof. In general the intersection of arbitrary number of subspaces is also a subspace, and the intersection of finite dimensional subspaces must be finite dimensional. Moreover, since $W \cap V = W$ for every $W \in \mathcal{SV}$, the existence of V does not affect intersections. Hence $\mathcal{SV} \cup \{V\}$ is intersection closed.

Therefore we have:

Proposition 3.13 *$\mathcal{SV} \cup \{V\}$ is a closed set system.*

Proof. As we saw in Lemma 3.11, $\langle\cdot\rangle$ defines a closed set system. This closed set system includes $\mathcal{SV} \cup \{V\}$. According to Proposition 1.7, it suffices to show that for each $S \subseteq V$, there exists $W \in \mathcal{SV} \cup \{V\}$ such that $S \subseteq W$. Every S is included in V, therefore the statement holds.

The closure operator associated with $\mathcal{SV} \cup \{V\}$ can be expressed as follows.

Lemma 3.14 *Let $S \subseteq V$ and let C is the closure operator associated with $\mathcal{SV} \cup \{V\}$. Then:*
1. If $\langle S\rangle$ is finite dimensional, then $C(S) = \langle S\rangle$.
2. If $\langle S\rangle$ is infinite dimensional, then $C(S) = V$.

Proof. 1. According to Proposition 1.7, $C(S) = \cap\{W \in \mathcal{SV} \cup \{V\} \mid S \subseteq W\}$. $\langle S\rangle$ is in the set of the right of this formula. Hence $C(S) \subseteq \langle S\rangle$. On the other hand, $S \subseteq C(S)$ implies $\langle S\rangle \subseteq \langle C(S)\rangle$. Now $C(S)$ is a vector space, that is, a closed set with respect to $\langle\cdot\rangle$. Thus $\langle C(S)\rangle = C(S)$, so we have $C(S) \supseteq \langle S\rangle$.
2. Clearly no $W \in \mathcal{SV}$ includes S. Thus V is only element of $\mathcal{SV} \cup \{V\}$ that includes S, so $C(S) = \cap\{W \in \mathcal{SV} \cup \{V\} \mid S \subseteq W\} = V$.

$\mathcal{SV} \cup \{V\}$ can not be inferable from positive data. In fact, $V \in \mathcal{SV} \cup \{V\}$ has no finite tell-tale: for every finite set F of V, $\langle F\rangle \in \mathcal{SV} \cup \{V\}$ and $F \subseteq \langle F\rangle \subseteq V$. Then we modify the class further. We introduce a new element $\boldsymbol{g}_0$ that is not included in V and let $V' = \langle V \cup \{\boldsymbol{g}_0\}\rangle$. Elements of $\mathcal{SV}$ can be regarded as finite dimensional subspaces of V'. Let $\mathcal{SV}' = \mathcal{SV} \cup \{V'\}$. Then,

Proposition 3.15 *$\mathcal{SV}'$ is a closed set system.*

Proof. We define a mapping $C : 2^{V'} \to 2^{V'}$ by:

$$C(S) = \begin{cases} \langle S\rangle & \text{if } \boldsymbol{g}_0 \notin \langle S\rangle \text{ and } \langle S\rangle \text{ is finite dimensional, that is, } \langle S\rangle \in \mathcal{SV} \\ V' & \text{else.} \end{cases}$$

Then it is easy to show that C becomes a closure operator and the closed set system defined by C is $\mathcal{SV}'$.

Moreover, the following holds:

Proposition 3.16 *Every element of $\mathcal{SV}'$ has a characteristic set.*

Proof. Let $W \in \mathcal{SV}'$. If $W \in \mathcal{SV}$, then there exists a finite basis G of W. Clearly $C(G) = W$. Since Lemma 1.11, G is a characteristic set of W. If $W = V'$, then the set $\{\boldsymbol{g}_0\}$ is a characteristic set of V'.

Remark 3.17 $\mathcal{SV}'$ is not Noetherian. In fact, if we fix a countable basis $\{\boldsymbol{g}_1, \boldsymbol{g}_2, \ldots\}$ of V, then there exists an infinite ascending chain

$$\langle \boldsymbol{g}_1 \rangle \subsetneq \langle \boldsymbol{g}_1, \boldsymbol{g}_2 \rangle \subsetneq \langle \boldsymbol{g}_1, \boldsymbol{g}_2, \boldsymbol{g}_3 \rangle \subsetneq \cdots$$

of $\mathcal{SV}'$.

Chapter 4

Bounded Unions of Closed Set Systems

4.1 Bounded Unions of Closed Set Systems

Let $\mathcal{L}$ be a Noetherian closed set systems over some set U and C denotes its closure operator. As we have seen in Theorem 1.14, $\mathcal{L}$ has finite elasticity and Theorem 1.17 implies that the class $\cup^{\leq k}\mathcal{L}$ also has finite elasticity, thus $\cup^{\leq k}\mathcal{L}$ is inferable from positive data. In this section, we consider a learning procedure for $\cup^{\leq k}\mathcal{L}$.

Let F be a finite subset of U. By Lemma 1.11, F is a characteristic set of $C(F)$ in $\mathcal{L}$. Since all elements of $\cup^{\leq k}\mathcal{L}$ have characteristic sets and $C(F)$ can be regarded as a member of $\cup^{\leq k}\mathcal{L}$, $C(F)$ has a characteristic set in $\cup^{\leq k}\mathcal{L}$. Throughout this section, we assume the following:

($*$) There exists an algorithm to compute a characteristic set of $C(F)$ in $\cup^{\leq k}\mathcal{L}$ from F. Moreover, If we denote the yielding characteristic set by $\chi(C(F), \cup^{\leq k}\mathcal{L})$, then we assume that

$$\chi(C(\chi(C(F), \cup^{\leq k}\mathcal{L})), \cup^{\leq k}\mathcal{L}) = \chi(C(F), \cup^{\leq k}\mathcal{L}).$$

In §§4.2 and 4.3, we give examples of Noetherian closed set systems satisfying ($*$).

In the algorithm we present later we will use the idea of hypergraph. First we review the definition of hypergraph briefly.

Definition 4.1 A *hypergraph* is a pair (V, HE) where V is a finite set and HE be a subset of 2^V that does not contain the empty set $\emptyset$. An element of

V is called a *vertex*, and an element of HE is called a *hyperedge*. We denote the set of vertices and hyperedges of a hypergraph $\mathcal{G}$ by $V(\mathcal{G})$ and $HE(\mathcal{G})$ respectively.

Now we construct our learning algorithm. Let $L_1 \cup \ldots \cup L_m \in \cup^{\leq k}\mathcal{L}$ be a target language of the algorithm and let $\sigma : a_1, a_2, \ldots, a_n, \ldots$ be a positive data of $L_1 \cup \ldots \cup L_m$. We inductively define a hypergraph denoted by $\mathcal{G}_n$ having the set of vertices $V(\mathcal{G}_n) = \{a_1, \ldots, a_n\}$ as follows:

Inductive definition of $\mathcal{G}_n$:
For $n = 1$, we put

$$V(\mathcal{G}_1) = \{a_1\}, \quad HE(\mathcal{G}_1) = \{\{a_1\}\}.$$

Suppose that $\mathcal{G}_n$ is already given and a_{n+1} is presented. We construct $\mathcal{G}_{n+1}$ in the following way:

Procedure 1: Construction of $\mathcal{G}_{n+1}$ from $\mathcal{G}_n$;
Input: a_{n+1} and $\mathcal{G}_n$;
Output: a hypergraph $\mathcal{G}_{n+1}$;
begin
1. Put $V = V(\mathcal{G}_n) \cup \{a_{n+1}\}$ and $HE = HE(\mathcal{G}_n)$;
2. **for** each subset $F \subseteq V$ such that $a_{n+1} \in F$ **do begin**
3. Let $E = \chi(C(F), \cup^{\leq k}\mathcal{L})$;
4. **if** $E \subseteq V$ and there is no $\mathcal{E} \in HE$ such that $E \subseteq \mathcal{E}$ **then begin**
5. **for** each element $\mathcal{E}$ of HE **do**
6. **if** $\mathcal{E} \subsetneq E$ **then** remove $\mathcal{E}$ from HE;
7. Add E to HE;
8. **end**;
9. **end**;
10. **return** $\mathcal{G}_{n+1} = (V, HE)$;
end.

We make use of the assumption $(*)$ at 3. As a result of the algorithm above, the following lemma clearly holds:

Lemma 4.2 *Let F be a finite set of U and N be a fixed positive integer. Suppose that there is $\mathcal{E} \in HE(\mathcal{G}_N)$ such that $F \subseteq \mathcal{E}$. Then for each $n \geq N$, there exists $\mathcal{E}_n \in HE(\mathcal{G}_n)$ such that $F \subseteq \mathcal{E}_n$.*

In additional, the next holds:

Proposition 4.3 *Let $\mathcal{G}_n$ be the yielding hypergraph of Procedure 1 and $F \subseteq \{a_1, \ldots, a_n\}$. Let $E = \chi(C(F), \cup^{\leq k}\mathcal{L})$. Then it holds:*
1. If $E \subseteq \{a_1, \ldots, a_n\}$, then there exists $\mathcal{E} \in HE(\mathcal{G}_n)$ such that $E \subseteq \mathcal{E}$.
2. If else, there exists $\mathcal{E}' \in HE(\mathcal{G}_M)$ such that $E \subseteq \mathcal{E}'$, where M is the largest index of elements of E.

Proof. (1) is obvious. (2) At the point of construction of $\mathcal{G}_M$, the algorithm checks whether $\chi(C(E), \cup^{\leq k}\mathcal{L})$ is included in $\{a_1, \ldots, a_M\}$. Since the assumption $(*)$, $\chi(C(E), \cup^{\leq k}\mathcal{L}) = E$, and by the choice of M, $E \subseteq \{a_1, \ldots, a_M\}$. Hence either E is added to $HE(\mathcal{G}_M)$ or there is $\mathcal{E}' \in HE(\mathcal{G}_M)$ such that $E \subseteq \mathcal{E}'$.

Remark 4.4 In general, $\{a_i\}$ is a characteristic set of $C(\{a_i\})$ in $\cup^{\leq k}\mathcal{L}$. Otherwise, there is $L'_1 \cup \ldots \cup L'_s \in \cup^{\leq k}\mathcal{L}$ such that $a_i \in L'_1 \cup \ldots \cup L'_s \subsetneq C(\{a_i\})$. Suppose that $a_i \in L'_j$. This implies $C(\{a_i\}) \subseteq L'_j$ and it contradicts $L'_1 \cup \ldots \cup L'_s \subsetneq C(\{a_i\})$. Note that although $\chi(C(\{a_i\}), \cup^{\leq k}\mathcal{L})$ is not necessarily equal to $\{a_i\}$, the proposition above holds.

We are now in a position to give our learning procedure:

Procedure 2: Learning $\cup^{\leq k}\mathcal{L}$;
Input: a positive presentation $\sigma : a_1, a_2, \ldots, a_n, \ldots$ for $L_1 \cup \ldots \cup L_m$;
Output: a sequence of at most k-tuples of characteristic sets
$(\chi_1^{(1)}, \ldots, \chi_{m_1}^{(1)})$, $(\chi_1^{(2)}, \ldots, \chi_{m_2}^{(2)}), \ldots$;
begin
1. $S = \emptyset$; /*Possible candidates for characteristic sets*/
2. Put $n = 1$;
3. **repeat**
4. Construct the hypergraph $\mathcal{G}_n$ for $a_1, a_2, \ldots, a_n$;
5. Put $S = HE(\mathcal{G}_n)$;
6. Choose at most k maximal elements from S with respect to the order as below;
7. Output (at most) k-tuple in 6;
8. Add 1 to n;
9. **forever**;

end.

We define an ordering on S as follows:

$$\chi_1 < \chi_2 \quad \Leftrightarrow \quad \begin{array}{l} C(\chi_1) \subsetneq C(\chi_2) \\ \texttt{ELSE} \quad C(\chi_1) = C(\chi_2) \text{ and } \chi_1 \prec \chi_2 \text{ under a certain suitable} \\ \text{ordering } \prec. \end{array}$$

The ordering $\prec$ does not affect the validity of Procedure 2, so we can adopt a convenient ordering (for example, the order of appearance in S).

Remark 4.5 Note that $C(\chi_i^{(n)}) \not\subseteq C(\chi_j^{(n)})$ for any i, j $(i \neq j)$. Otherwise, if the case $C(\chi_i^{(n)}) \subsetneq C(\chi_j^{(n)})$ then $\chi_i^{(n)}$ could not be maximal, or if $C(\chi_i^{(n)}) = C(\chi_j^{(n)})$ then either $\chi_i^{(n)}$ or $\chi_j^{(n)}$ could not be maximal.

Now our theorem is the following:

Theorem 4.6 *Suppose that $\cup^{\leq k}\mathcal{L}$ is compact. $\cup^{\leq k}\mathcal{L}$ is identifiable in the limit from positive data via Procedure 2.*

We need some lemmas to prove Theorem 4.6.

Lemma 4.7 *Let $\mathcal{E}$ be an arbitrary hyperedge of $\mathcal{G}_n$. Then $C(\mathcal{E}) \subset L_1 \cup \ldots \cup L_m$. Moreover, if $\mathcal{L}$ is compact, then there exists L_i such that $C(\mathcal{E}) \subseteq L_i$.*

Proof. By our construction of $\mathcal{G}_n$, there exists $F \subseteq V(\mathcal{G}_n) = \{a_1, \ldots, a_n\}$ such that $\mathcal{E} = \chi(C(F), \cup^{\leq k}\mathcal{L})$. By Lemma 1.18, $\mathcal{E}$ is also a characteristic set of $C(F)$ in $\mathcal{L}$. By combining this with Lemma 1.11, we obtain $C(\mathcal{E}) = C(F)$. Moreover, $\mathcal{E} \subseteq \{a_1, \ldots, a_n\} \subseteq L_1 \cup \ldots \cup L_m$ by construction of $\mathcal{G}_n$. Since $\mathcal{E}$ is a characteristic set of $C(F)$ in $\cup^{\leq k}\mathcal{L}$, the definition of characteristic set implies that $C(F) \subseteq L_1 \cup \ldots \cup L_m$. Therefore $C(\mathcal{E}) \subseteq L_1 \cup \ldots \cup L_m$. The second statement is immediate from the definition of compactness.

Lemma 4.8 *Suppose that $\cup^{\leq k}\mathcal{L}$ is compact.*
1. Let $L_1, \ldots, L_m$ be distinct members of $\mathcal{L}$ that satisfy $L_i \not\subseteq L_j$ for all i, j $(i \neq j)$. Then $L_i \not\subseteq \cup_{j \neq i} L_j$.
2. Let $M \in \mathcal{L}$ and let $L_1, \ldots, L_m$ be as above. If $M \subseteq L_1 \cup \ldots \cup L_m$ and $L_i \subseteq M$ for some i, then $L_i = M$.

Proof. (1) If $L_i \subseteq \cup_{j \neq i} L_j$, then $L_i \subseteq L_{j_0}$ for some j_0 by compactness. This contradicts to our assumption. (2) By compactness, there exists L_{j_0} such that $M \subseteq L_{j_0}$. Hence $L_i \subseteq M \subseteq L_{j_0}$. By our assumption, $L_i = L_{j_0}$.

Proof of Theorem 4.6. Let $L_1 \cup \ldots \cup L_m$ be an arbitrary element in $\cup^{\leq k}\mathcal{L}$. Suppose that $L_i = C(F_i)$, where F_i is a finite subset of L_i. Since F_i is finite, all elements of F_i's are presented at a certain finite step N_0. Therefore, at a certain step N after the step N_0, one can assume that all elements of all $\chi(C(F_i), \cup^{\leq k}\mathcal{L})$'s are presented. Let $\mathcal{E}_{1,N}, \ldots, \mathcal{E}_{m_N,N}$ be maximal hyperedges of $\mathcal{G}_N$. By the proof of Lemma 4.7, each $\mathcal{E}_{j,N}$ is a characteristic set of closed set $C(\mathcal{E}_{j,N})$ in $\cup^{\leq k}\mathcal{L}$ and $C(\mathcal{E}_{j,N})$ is contained in $L_1 \cup \ldots \cup L_m$. Note that $C(\mathcal{E}_{i,N}) \not\subseteq C(\mathcal{E}_{j,N})$ for any i, j $(i \neq j)$ by Remark 4.5. Now we claim that:

Claim. For each $1 \leq i \leq m$, there exists a unique $\mathcal{E}_{j_i,N}$ of $HE(\mathcal{G}_N)$ with $L_i = C(\mathcal{E}_{j_i,N})$.

By our construction of $\mathcal{G}_N$, $\chi(C(F_i), \cup^{\leq k}\mathcal{L})$ is either added as a hyperedge or contained in a hyperedge added at a certain step. Hence there exists a hyperedge $\mathcal{E}_i$ of $\mathcal{G}_N$ such that $\chi(C(F_i), \cup^{\leq k}\mathcal{L}) \subseteq \mathcal{E}_i$ for each i (See Lemma 4.2 and Proposition 4.3). By the definition of closure operator, $C(\chi(C(F_i), \cup^{\leq k}\mathcal{L})) \subseteq C(\mathcal{E}_i)$. The proof of Lemma 4.7 implies that $C(F_i) = C(\chi(C(F_i), \cup^{\leq k}\mathcal{L}))$, so $L_i = C(F_i) \subseteq C(\mathcal{E}_i)$. Applying Lemma 4.8(2), we obtain $L_i = C(\mathcal{E}_i)$. Now, since each $\mathcal{E}_i$ is in $HE(\mathcal{G}_N)$ and each $\mathcal{E}_{j,N}$ is maximal element of $HE(\mathcal{G}_N)$, there is a j_i such that $\mathcal{E}_i < \mathcal{E}_{j_i,N}$, and this means $L_i = C(\mathcal{E}_i) \subseteq C(\mathcal{E}_{j_i,N})$. Applying Lemma 4.8(2) again, we find that $L_i = C(\mathcal{E}_{j_i,N})$. The uniqueness part of the claim follows from Remark 4.5 immediately.

We finally show that $m_N = m$. By Claim, $m_N \geq m$. If $m_N > m$, then there exists $\mathcal{E}_{j_0,N}$ such that (i) $C(\mathcal{E}_{j_0,N}) \neq L_i$ for $i = 1, \ldots, m$ and (ii) $C(\mathcal{E}_{j_0,N}) \subset L_1 \cup \ldots \cup L_m$. These conditions imply that $C(\mathcal{E}_{j_0,N}) \subset C(\mathcal{E}_{j_i,N})$ for some j_i. This contradicts to Remark 4.5.

Remark 4.9 Note that the hypotheses in our algorithm are not necessarily consistent. However, one can modify them into consistent ones without difficulty.

4.2 Learning Bounded Set Unions of Polynomial Ideals

In this section we present the class of bounded unions of ideals of a polynomial ring as an application of the procedure we gave in the previous section.

As we have seen in §3.1, $\mathcal{I}$ is a Noetherian closed set system associated with the closure operator $F \mapsto \langle F \rangle$. At first we check the condition $(*)$.

Lemma 4.10 *There exists an algorithm that constructs a characteristic set $\chi(I, \cup^{\leq k}\mathcal{I})$ if a finite generating set F of I is given. Moreover, it holds that*

$$\chi(\langle \chi(I, \cup^{\leq k}\mathcal{I}) \rangle, \cup^{\leq k}\mathcal{I}) = \chi(I, \cup^{\leq k}\mathcal{I}).$$

The key to proof Lemma 4.10 is the following proposition.

Proposition 4.11 ([17, Theorem 9]) *For each element $I_1 \cup \ldots \cup I_m$ of $\cup^{\leq k}\mathcal{I}$, a characteristic set of $I_1 \cup \ldots \cup I_m$ can be constructed concretely if generating sets of I_i's are given.*

Proof. For simplification we assume $m = 1$, since we need only the case. (For general m, see [17]). Suppose that $I_1 = \langle f_1, \ldots, f_n \rangle$. Put $M = k(n-1)+1$. As we have seen in Lemma 3.9 and Example 3.10, one can construct a (n, M)-matrix $C = (c_{ij})_{1 \leq j \leq M}^{1 \leq i \leq n}$ over $\mathbb{Q}$ such that any n column vectors of C are linearly independent. Now let $h_j = \sum_{i=1}^{n} c_{ij} f_i \quad (j = 1, \ldots, M)$. Then $\{h_1, \ldots, h_M\}$ becomes a characteristic set of I_1 in $\cup^{\leq k}\mathcal{I}$. To prove this, suppose that $\{h_1, \ldots, h_M\} \subseteq J_1 \cup \ldots \cup J_m \in \cup^{\leq k}\mathcal{I}$. Since $m < k$, the pigeon-hole principle implies that there exists an a such that J_a includes at least n of h_i's. Suppose that $h_{i_1}, \ldots, h_{i_n} \in J_a$. Since the matrix $C_{i_1, \ldots, i_n} = (c_{ij})_{j=i_1,\ldots,i_n}^{1 \leq i \leq n}$ is nonsingular, so $f_1, \ldots, f_n$ can be written by linear combinations of $h_{i_1}, \ldots, h_{i_n}$. This means $f_1, \ldots, f_n \in J_a$, thus we have $I_1 = \langle f_1, \ldots, f_n \rangle \subseteq J_a \subseteq J_1 \cup \ldots \cup J_m$. Therefore it is proved that $\{h_1, \ldots, h_M\}$ is a characteristic set of I_1 in $\cup^{\leq k}\mathcal{I}$.

Proof of Lemma 4.10. We fix some monomial order $<$. One can compute the reduced Groebner basis G of I from F. Applying the procedure in the proof of Proposition 4.11 to G, we obtain a characteristic set $\{h_1, \ldots, h_M\}$ of I in $\cup^{\leq k}\mathcal{I}$. We define $\chi(I, \cup^{\leq k}\mathcal{I}) = \{h_1, \ldots, h_M\}$. The reduced Groebner basis of $\langle \chi(I, \cup^{\leq k}\mathcal{I}) \rangle$ is equal to G since the uniqueness of the reduced Groebner basis. This implies $\chi(\langle \chi(I, \cup^{\leq k}\mathcal{I}) \rangle, \cup^{\leq k}\mathcal{I}) = \chi(I, \cup^{\leq k}\mathcal{I})$.

Remark 4.12 For instance we can give an example of $\chi(I, \cup^{\leq k}\mathcal{I})$ by using Example 3.10. Let $\{g_1, \ldots, g_r\}$ be the reduced Groebner basis of I. Let

$$h_i = g_1 + c_i g_2 + \ldots + c_i^{r-1} g_r \ (i = 1, \ldots, M)$$

where $M = k(r-1)+1$ and c_i's are distinct elements of $\mathbb{Q}$. Note that no h_i will vanish since $\{g_1, \ldots, g_r\}$ is the reduced Groebner basis.

Lemma 4.10 also implies that $\cup^{\leq k}\mathcal{I}$ is compact:

Lemma 4.13 *$\cup^{\leq k}\mathcal{I}$ is compact.*

Proof. Suppose that $I = \langle g_1, \ldots, g_r \rangle$ is contained in $I_1 \cup \ldots \cup I_m$. Let $M = m(r-1)+1$ and take $h_1, \ldots, h_M$ as above. By the pigeon-hole principle, there exists some j such that I_j includes at least r of h_i's. Suppose that $h_{i_1}, \ldots, h_{i_r} \in I_j$. Since the construction of h_i's, each g_i can be represented by linear combination of $h_{i_1}, \ldots, h_{i_r}$. This means $I \subseteq I_j$.

According to above arguments, we have:

Theorem 4.14 *$\cup^{\leq k}\mathcal{I}$ is identifiable in the limit from positive data via Procedure 2.*

Example 4.15 Let us consider learning $\langle x^2, y^3 \rangle \cup \langle x^3, y^2 \rangle \in \cup^{\leq 2}\mathcal{I}$. Let a positive presentation σ be $x^2, y^3, y^2, x^2+y^3, x^3, x^3+y^2, \ldots$. By the argument of §3 of [17], we can take a characteristic set $\chi(\langle f, g \rangle, \cup^{\leq 2}\mathcal{I}) = \{f, g, f+g\}$ for distinct polynomials f and g. The hyperedges of hypergraphs constructed by Procedure 1 are as follows:

$$
\begin{aligned}
HE_1 &= \{\{x^2\}\}, \\
HE_2 &= \{\{x^2\}, \{y^3\}\}, \\
HE_3 &= \{\{x^2\}, \{y^3\}, \{y^2\}\}, \\
HE_4 &= \{\{x^2, y^3, x^2+y^3\}, \{y^2\}\}, \\
HE_5 &= \{\{x^2, y^3, x^2+y^3\}, \{y^2\}, \{x^3\}\}, \\
HE_6 &= \{\{x^2, y^3, x^2+y^3\}, \{y^2, x^3, x^3+y^2\}\}.
\end{aligned}
$$

Hence Procedure 2 learns $\langle x^2, y^3 \rangle \cup \langle x^3, y^2 \rangle$ when $n = 6$.

We finish this section by giving an improved learning algorithm concretely. The algorithm outputs reduced Groebner bases instead of characteristic sets, and we make use of uniqueness of the reduced Groebner basis to simplify the algorithm. Consequently it does not construct a hypergraph. Here we employ the characteristic sets in Remark 4.12 and suppose that $c_1, c_2, \ldots$ are fixed nonzero rational numbers such that $i \neq j \Rightarrow c_i \neq c_j$ (for example, $c_i = i$).

Procedure 3: Learning $\cup^{\leq k}\mathcal{I}$;

Input: a positive presentation $\sigma : f_1, f_2, \ldots, f_n, \ldots$ for $I_1 \cup \ldots \cup I_m$;
Output: a sequence of at most k-tuples of reduced Groebner bases
$(G_1^{(1)}, \ldots, G_{m_1}^{(1)}), (G_1^{(2)}, \ldots, G_{m_2}^{(2)}), \ldots$;
begin
1. Put $V_1 = \{f_1\}$, $S_1 = \{\{f_1\}\}$;
2. Put $n = 1$;
3. Output $(\{f_1\})$;
4. **repeat**
5. Add 1 to n;
6. Let $V_n = V_{n-1} \cup \{f_n\}$ and $S = S_{n-1}$;
7. **for** each subset $F \subseteq V_n$ such that $a_n \in F$ **do begin**
8. **if** there is no $\mathcal{E} \in S$ such that $\langle F \rangle \subseteq \langle \mathcal{E} \rangle$ **then begin**
9. Compute the reduced Groebner basis $G = \{g_1, \ldots, g_r\}$ of $\langle F \rangle$ from F;
10. Put $M = k(r-1)+1$;
11. Put $E = \{g_1 + c_i g_2 + \ldots + c_i^{r-1} g_r \mid i = 1, \ldots, M\}$;
12. **if** $E \subseteq V_n$ **then begin**
13. **for** each element $\mathcal{E}$ of S **do**
14. **if** $\langle \mathcal{E} \rangle \subsetneq \langle F \rangle$ **then** remove $\mathcal{E}$ from S;
15. Add G to S;
16. **end**;
17. **end**;
18. **end**;
19. Put $S_n = S$;
20. Choose at most k maximal elements from S_n with respect to the order $<$ that satisfies:
$G_1 < G_2 \Leftrightarrow G_1 \prec G_2$ under a certain suitable ordering $\prec$;
21. Output (at most) k-tuple in 20;
22. **forever**;

end.

Theorem 4.16 $\cup^{\leq k}\mathcal{I}$ *is identifiable in the limit from positive data via Procedure 3.*

Proof of Theorem 4.16 is an analogy to the proof of Theorem 4.6. We need the following lemmas.

Lemma 4.17 *Let S_n be the same as in Procedure 3 and $G \in S_n$. Then $\langle G\rangle \subseteq I_1 \cup \ldots \cup I_m$.*

Proof. Let E and V_n be as in Procedure 3. E can be regarded as a characteristic set of $\langle G\rangle$ in $\cup^{\leq k}\mathcal{I}$. On the other hand, $G \in S_n$ implies $E \subseteq V_n \subseteq I_1 \cup \ldots \cup I_m$. By the definition of characteristic set, $\langle G\rangle \subseteq I_1 \cup \ldots \cup I_m$.

Lemma 4.18 *Let V_n and S_n be the same as in Procedure 3 and let $F \subseteq \{a_1, \ldots, a_n\}$. Suppose that $a_n \in F$. Let G be the reduced Groebner basis of $\langle F\rangle$ and E be the set constructed at 11 of Procedure 3 from G. Then it holds:*
1. If $E \subseteq V_n$, then there exists $\mathcal{E} \in S_n$ such that $\langle G\rangle \subseteq \langle\mathcal{E}\rangle$.
2. If else, there exists $\mathcal{E}' \in S_M$ such that $\langle G\rangle \subseteq \langle\mathcal{E}'\rangle$, where M is the largest index of elements of E.

Proof. (1) is obvious. (2) Since $\langle E\rangle = \langle F\rangle$, the uniqueness of the reduced Groebner basis assures that G is the reduced Groebner basis of $\langle E\rangle$. Hence at the step M, the same E appears again, and this time $E \subseteq V_M$. Thus the statement holds.

Lemma 4.19 *Let G_1, G_2 be elements of S_n. Then $\langle G_1\rangle \not\subseteq \langle G_2\rangle$.*

Proof. Obvious by the construction of S_n.

Proof of Theorem 4.16. Let $G_i = \{g_{1,i}, \ldots, g_{r_i,i}\}$ be the reduced Groebner basis of I_i. Let $M_i = k(r_i - 1) + 1$. Suppose that $E_i = \{g_{1,i} + c_i g_{2,i} + \ldots + c_i^{r_i-1} g_{r_i,i} \mid i = 1, \ldots, M_i\}$. At a certain step N, one can assume that all elements of all E_i's are presented. By applying Lemma 4.18(2), there exists $G'_i \in S_N$ such that $I_i = \langle G_i\rangle \subseteq \langle G'_i\rangle$. On the other hand, Lemma 4.17 says that $\langle G'_i\rangle \subseteq I_1 \cup \ldots \cup I_m$. Thus Lemma 4.8 implies that $\langle G'_i\rangle = \langle G_i\rangle = I_i$. From the uniqueness of the reduced Groebner basis, we have $G_i = G'_i$. If there is $G' \in S_n \setminus \{G_1, \ldots, G_m\}$, there exists j such that $\langle G'\rangle \subseteq I_j = \langle G_j\rangle$ since Lemma 4.17 and compactness. This contradicts Lemma 4.19.

4.3 Learning Bounded Unions of Tree Pattern Languages

4.3.1 Tree Pattern Languages

Let Σ be a finite set and V be a countable set disjoint from Σ. The elements of Σ and V are called *symbols* and *variables*, respectively. We assume that there is a mapping *rank* that maps an element of Σ to a non-negative integer. We define the rank of elements of V to be zero.

Definition 4.20 1. A *tree pattern* p over Σ is a tree satisfying following properties:
• p has the root.
• p is directed.
• p is ordered.
• Each node of p is labeled by an element of $\Sigma \cup V$.
• The number of children of each node is equal to the rank of the label of the node.
2. A *tree* over Σ is a tree pattern over Σ that has no nodes labeled by an element of V.

$\mathcal{TP}_\Sigma$ and $\mathcal{T}_\Sigma$ denote the set of all tree patterns and all trees over Σ, respectively.

Definition 4.21 A *substitution* is a mapping θ from V to $\mathcal{TP}_\Sigma$. $p\theta$ denotes the tree pattern obtained from applying a substitution θ to p. We define a relation on $\mathcal{TP}_\Sigma$ as follows: $p \preceq q \Leftrightarrow$ there exists a substitution θ such that $p = q\theta$. We denote $p \equiv q$ if $p \preceq q$ and $q \preceq p$, and call that p and q are equivalent.

Note that $p \equiv q$ if and only if $p = q\theta$ for some renaming θ of variables. On substitutions, the next lemma holds. The fact that $\equiv$ is an equivalence relation follows from (1). Throughout this section, we regard two equivalent tree patterns as the same.

Lemma 4.22 *1. If $p \preceq q$ and $q \preceq r$, then $p \preceq r$.*
2. Let $|p|$ be the number of nodes of p. If $p \preceq q$, then $|p| \geq |q|$.
3. For any $p \in \mathcal{TP}_\Sigma$, there are finitely many $q \in \mathcal{TP}_\Sigma$ such that $p \preceq q$ (up to renaming of variables).

Proof. 1. Suppose that $p = q\theta_1$ and $q = r\theta_2$. Clearly $p = r(\theta_2 \circ \theta_1)$.
2. By definition, every substitution replaces a node by a tree pattern that has at least one node. Therefore the statement holds.
3. According to (2), $p \preceq q \Rightarrow |p| \geq |q|$. Since Σ is finite, there are finitely many tree patterns that has at most $|p|$ nodes up to renaming of variables.

Definition 4.23 Let S be a nonempty subset of $\mathcal{TP}_\Sigma$. A tree pattern p is called the *least common anti-instance* of S if

(i) $q \preceq p$ for any $q \in S$, and
(ii) if $q \preceq r$ for any $q \in S$, then $p \preceq r$.

Lemma 4.24 ([13]) *For any subset $S \neq \emptyset$ of $\mathcal{TP}_\Sigma$, there uniquely exists the least common anti-instance of S up to equivalence.*

We denote the least common anti-instance of S by $lca(S)$. As a duality one can define the greatest common instance.

Definition 4.25 ([13]) Let S be a nonempty subset of $\mathcal{TP}_\Sigma$. There uniquely exists a tree pattern called the *greatest common instance* of S, denoted by $gsi(S)$, that satisfies:
(i) $gsi(S) \preceq p$ for any $p \in S$, and
(ii) if $q \preceq p$ for any $p \in S$, then $q \preceq gsi(S)$.

If S is finite, then $lca(S)$ and $gsi(S)$ can be computed in polynomial time [13]. Now we define a tree pattern language.

Definition 4.26 Let p be a tree pattern. A *tree pattern language* defined by p is the set of trees (not tree patterns) $L(p) = \{t \in \mathcal{T}_\Sigma \mid t \preceq p\}$. We denote the set of all tree pattern languages $\{L(p) \mid p \in \mathcal{TP}_\Sigma\}$ by $\mathcal{TPL}(\Sigma, V)$. We may omit (Σ, V) if it is clear from the context.

Lemma 4.27 ([13]) *$p \preceq q \Rightarrow L(p) \subseteq L(q)$ for any $p, q \in \mathcal{TP}_\Sigma$. If $\sharp(\Sigma) \geq 2$, then $p \preceq q \Leftrightarrow L(p) \subseteq L(q)$.*

4.3.2 Learning Bounded Unions of Tree Pattern Languages

In [4], Arimura et al. studied learnability of bounded union of tree pattern languages. However, they did not seem to use characteristic sets explicitly. We here give a procedure learning bounded unions of tree pattern languages by using our result in §4.1.

Lemma 4.28 ([4]) *$\mathcal{TPL}$ is a Noetherian closed set system.*

Proof. First we show that $\mathcal{TPL}$ is intersection closed. Let $\{L(p_i)\} \subseteq \mathcal{TPL}$. Then clearly $\cap L(p_i) = L(gsi(\{p_i\}))$ by definition. Hence $\mathcal{TPL}$ is intersection closed, thus $\mathcal{TPL}$ is a closed set system since Proposition 1.7. Then it is enough to show that $\mathcal{TPL}$ has finite thickness. For any fixed $k \in \mathbb{N}$, the set

$\{p \in \mathcal{TP}_\Sigma \mid |p| \leq k\}$ is finite up to equivalence. This fact and Lemma 4.22(2) imply that, for any $t \in \mathcal{T}_\Sigma$, there are finitely many $q \in \mathcal{TP}_\Sigma$ such that $t \preceq q$, that is, $t \in L(q)$. This means that $\mathcal{TPL}$ has finite thickness. Therefore, $\mathcal{TPL}$ has finite elasticity by Theorem 1.4.

Lemma 4.29 ([3]) *If $\sharp(\Sigma) > k$, $\cup^{\leq k}\mathcal{TPL}$ is compact.*

Now we introduce a closed set system $\mathcal{C}$. For $S \subseteq \mathcal{TP}_\Sigma$, we define $C(S) = \{p \in \mathcal{TP}_\Sigma \mid p \preceq lca(S)\}$. Note that $C(S) = C(lca(S))$.

Lemma 4.30 *C is a closure operator on $\mathcal{TP}_\Sigma$.*

Proof. (CO1) is obvious by the definition of lca. (CO2) Suppose $S_1 \subseteq S_2 \subseteq \mathcal{TP}_\Sigma$. Clearly, $lca(S_1) \preceq lca(S_2)$. Lemma 4.22(1) implies $C(S_1) = C(lca(S_1)) \subseteq C(lca(S_2)) = C(S_2)$. (CO3) In general, $lca(C(S)) = lca(S)$ holds. Thus $C(C(S)) = C(lca(C(S))) = C(lca(S)) = C(S)$.

Remark 4.31 Let S be a subset of $\mathcal{T}_\Sigma$. As an analogy of C, one can define $L(S) = \{t \in \mathcal{T}_\Sigma \mid t \preceq lca(S)\}$. Then $L : 2^{\mathcal{T}_\Sigma} \to 2^{\mathcal{T}_\Sigma}$ becomes a closure operator.

$\mathcal{C}$ denotes the closed set system defined by C. The following lemma clearly holds by definition.

Lemma 4.32 *For every $p \in \mathcal{TP}_\Sigma$, $L(p) = C(p) \cap \mathcal{T}_\Sigma$.*

Lemma 4.33 *1. $\mathcal{C}$ has finite elasticity. 2. $\cup^{\leq k}\mathcal{C}$ is compact.*

Proof. 1. Similar to the proof of Lemma 4.28. 2. Suppose that $C(p) \subseteq C(p_1) \cup \ldots \cup C(p_m)$ $(m \leq k)$. Since $p \in C(p)$, there exists i_0 such that $p \in C(p_{i_0})$. Hence $C(p) \subset C(p_{i_0})$.

Now we consider characteristic sets. Let $\Sigma_0 = \{a \in \Sigma \mid rank(a) = 0\}$ and $\Sigma_+ = \{f \in \Sigma \mid rank(f) > 0\}$. In the following, we assume that neither Σ_0 nor Σ_+ is empty.

Lemma 4.34 *1. For every $p \in \mathcal{TP}_\Sigma$, there exists a characteristic set of $L(p)$ in $\mathcal{TPL}$ consisting of at most two elements.*
2. For every $S \subseteq \mathcal{TP}_\Sigma$, there exists a characteristic set of $C(S)$ in $\mathcal{C}$ consisting of one element.

Proof. 1. Let $a \in \Sigma_0$ and $f \in \Sigma_+$. For simplicity, we assume $rank(f) = 1$. Suppose that all variables that appear in p are $\{x_1, \ldots, x_n\}$. Let a tree $f^{(i)}$ be $f(f(\ldots(f(a))\ldots))$, where f occurs i times. Now we define substitutions θ and θ' by

$$\theta : x_i \mapsto a, \theta' : x_i \mapsto f^{(i)},$$

then $\{p\theta, p\theta'\}$ becomes a characteristic set of $L(p)$ in $\mathcal{TPL}$. In fact, one can show that $lca(\{p\theta, p\theta'\}) = p$. If $\{p\theta, p\theta'\} \subseteq L(q)$, then $p \preceq q$ by the definition of lca. By Lemma 4.27, we have $L(p) \subseteq L(q)$.
2. Since C is a closure operator and $C(S) = C(lca(S))$, clearly $\{lca(S)\}$ is a characteristic set of $C(S)$ in $\mathcal{C}$.

Corollary 4.35 *$lca(L(p)) \equiv p$.*

Proof. By the argument of the proof of Lemma 4.34(1), there are $t_1, t_2 \in L(p)$ such that $lca(\{t_1, t_2\}) = p$. In general $S_1 \subseteq S_2 \Rightarrow lca(S_1) \preceq lca(S_2)$ holds, hence $p = lca(\{t_1, t_2\}) \preceq lca(L(p))$. On the other hand, since $t \preceq p$ for all $t \in L(p)$, the definition of lca implies that $L(p) \preceq p$. Therefore $lca(L(p)) \equiv p$.

Lemma 4.34 indicates that characteristic sets of $\mathcal{TPL}$ and $\mathcal{C}$ are bounded uniformly. From this property, one can show that characteristic sets of the unions are also bounded uniformly.

Lemma 4.36 *1.* ([18]) *Suppose $\sharp(\Sigma_+) \geq k$. For every $p \in \mathcal{TP}_\Sigma$, there exists a characteristic set of $L(p)$ in $\cup^{\leq k}\mathcal{TPL}$ consisting of at most $k+1$ elements.*
2. For every $S \subseteq \mathcal{TP}_\Sigma$, there exists a characteristic set of $C(S)$ in $\cup^{\leq k}\mathcal{C}$ consisting of one element.

Proof. 1. The proof is a general case of the proof of Lemma 4.34(1). Let $a \in \Sigma_0$ and $f_1, \ldots, f_k \in \Sigma_+$. For simplicity, we assume that $rank(f_i) = 1$ for each i. Suppose that all variables that appear in p are $\{x_1, \ldots, x_n\}$. Let a tree $f_j^{(i)}$ be $f_j(f_j(\ldots(f_j(a))\ldots))$, where f_j occurs i times. Let

$$\theta_0 : x_i \mapsto a, \theta_1 : x_i \mapsto f_1^{(i)}, \ldots, \theta_k : x_i \mapsto f_k^{(i)}.$$

We show that $\{p\theta_0, p\theta_1, \ldots, p\theta_k\}$ is a characteristic set of $L(p)$ in $\cup^{\leq k}\mathcal{TPL}$. Suppose that $\{p\theta_0, p\theta_1, \ldots, p\theta_k\} \subseteq L(p_1) \cup \ldots \cup L(p_m)$. Then there exists a $L(p_r)$ such that at least two $p\theta_i$'s are in $L(p_r)$. Now it holds that, for each $i \neq j$, $lca(\{p\theta_i, p\theta_j\}) = p$. Thus, $p \preceq p_r$ by the definition of lca. This means

$L(p) \subseteq L(p_r)$, so we have $L(p) \subseteq L(p_1) \cup \ldots \cup L(p_r)$.
2. We show $\{lca(S)\}$ is a characteristic set of $C(S)$ in $\mathcal{C}$. If $\{lca(S)\} \subseteq C(p_1) \cup \ldots \cup C(p_m)$, then there exists r such that $lca(S) \in C(p_r)$. This implies that $C(S) \subseteq C(p_r)$, so $C(S) \subseteq C(p_1) \cup \ldots \cup C(p_m)$.

In the proof of above lemma, a characteristic set of a closed set in $\mathcal{TPL}$ or $\mathcal{C}$ is constructed concretely. Therefore,

Lemma 4.37 *For both $\mathcal{TPL}$ and $\mathcal{C}$, there is an algorithm satisfies $(*)$.*

Proof. In case of $\mathcal{C}$, it immediately follows if we define $\chi(C(S), \cup^{\leq k}\mathcal{C}) = \{lca(S)\}$. In case of $\mathcal{TPL}$, we consider the following algorithm. Suppose that a finite subset $S \subseteq \mathcal{T}_\Sigma$ is given. (1) Compute $lca(S)$. (2) Apply the method in the proof of Lemma 4.36(1) to $lca(S)$ and define $\chi(L(S), \cup^{\leq k}\mathcal{TPL})$ by the obtained characteristic set. By Corollary 4.35, $lca(L(S)) = lca(S)$. Thus we have $\chi(L(\chi(L(S), \cup^{\leq k}\mathcal{TPL})), \cup^{\leq k}\mathcal{TPL}) = \chi(L(S), \cup^{\leq k}\mathcal{TPL})$.

According to above arguments, it holds:

Theorem 4.38 *$\cup^{\leq k}\mathcal{TPL}$ and $\cup^{\leq k}\mathcal{C}$ are identifiable in the limit from positive data via Procedure 2.*

Finally we consider concrete learning algorithm. Lemma 4.36(2) makes the algorithm learning $\cup^{\leq k}\mathcal{C}$ much simpler, because the condition $E \subseteq V$ at the step 4 of Procedure 1 is always true. (In fact, $\mathcal{C}^*$ can be identified by almost the same procedure. Note that $\mathcal{C}$ satisfies the condition in Theorem 1.22).

Procedure 4: Learning $\cup^{\leq k}\mathcal{C}$;
Input: a positive presentation $\sigma : q_1, q_2, \ldots, q_n, \ldots$ of tree patterns for $C(p_1) \cup \ldots \cup C(p_m)$;
Output: a sequence of at most k-tuples of tree patterns $(r_1^{(1)}, \ldots, r_{m_1}^{(1)}), (r_1^{(2)}, \ldots, r_{m_2}^{(2)}), \ldots$;
begin
1. $S = \emptyset$; /*The set to memorize a given sequence of $q_1, \ldots, q_n$*/
2. Put $n = 1$;
3. **repeat**
4. Add q_n to S;
5. Choose at most k maximal elements from S with respect to $\preceq$ up to equivalence;
6. Output (at most) k-tuple in 5;

7. Add 1 to n;
8. **forever**
end.

We assume $\sharp(\Sigma_+) \geq k$ in order to make Lemma 4.36(1) holds. For $\cup^{\leq k}\mathcal{TPL}$, if we apply Procedure 1 and 2 directly, then we obtain:

Procedure 5: Learning $\cup^{\leq k}\mathcal{TPL}$;
Input: a positive presentation $\sigma : t_1, t_2, \ldots, t_n, \ldots$ of trees for $L(p_1) \cup \ldots \cup L(p_m)$;
Output: a sequence of at most k-tuples of tree patterns $(q_1^{(1)}, \ldots, q_{m_1}^{(1)}), (q_1^{(2)}, \ldots, q_{m_2}^{(2)}), \ldots$;
begin
1. Put $n = 1$;
2. Put $\mathcal{G}_1 = (\{t_1\}, \{\{t_1\}\})$;
3. Output $\{t_1\}$;
4. **repeat**
5. Add 1 to n;
6. Put $V_n = \mathcal{G}_{n-1} \cup \{t_n\}$ and $HE_n = HE(\mathcal{G}_{n-1})$;
7. **for** each subset $F \subseteq V_n$ such that $F \ni t_n$ **do begin**
8. Let $E = \chi(L(F), \cup^{\leq k}\mathcal{TPL})$;
9. **if** $E \subseteq V_n$ and there is no $\mathcal{E} \in HE_n$ such that $E \subseteq \mathcal{E}$ **then begin**
10. **for** each element $\mathcal{E}$ of HE_n **do**
11. **if** $\mathcal{E} \subsetneq E$ **then** remove $\mathcal{E}$ from HE_n;
12. Add E to HE_n;
13. end;
14. end;
15. Put $\mathcal{G}_n = (V_n, HE_n)$;
16. Choose at most k maximal elements from HE_n with respect to $\preceq$ up to equivalence;
17. Output at most k-tuple in 16;
18. **forever**
end.

If we make use of Procedure 4, $\cup^{\leq k}\mathcal{TPL}$ is inferred alternatively as follows:

Procedure 6: Learning $\cup^{\leq k}\mathcal{TPL}$;
Input: a positive presentation $\sigma : t_1, t_2, \ldots, t_n, \ldots$ of trees for
$L(p_1) \cup \ldots \cup L(p_m)$;
Output: a sequence of at most k-tuples of tree patterns
$(q_1^{(1)}, \ldots, q_{m_1}^{(1)}), (q_1^{(2)}, \ldots, q_{m_2}^{(2)}), \ldots$;
begin
1. Generate "positive data" of $C(p_1) \cup \ldots \cup C(p_m)$ from σ;
2. Run Procedure 4 by "positive data" generated in 1;
3. Output the output of 2;
end.
Generation of "positive data" ($\mathcal{GPD}$);
4. $S = \emptyset$;
5. Put $n = 1$;
6. **repeat**
7. Add t_n to S;
8. **output** t_n;
9. **for** each subset F of S with $t_n \in F$ and $\sharp(F) = k + 1$ **do**
10. **if** $lca(t_i, t_j) = lca(F)$ for all $t_i, t_j \in F$ $(i \neq j)$ **then**
11. **output** $lca(F)$;
12. Add 1 to n;
13. **forever**;
end.

Theorem 4.39 $\cup^{\leq k}\mathcal{TPL}$ *is identifiable in the limit from positive data via Procedure 6.*

Proof. It suffices to show that $\mathcal{GPD}$ generates a positive data for $A = C(p_1) \cup \ldots \cup C(p_m)$. Let p be an arbitrary element of A. If $p \in \mathcal{T}_\Sigma$, then $p \in A \cap \mathcal{T}_\Sigma = L(p_1) \cup \ldots \cup L(p_m)$, so there exists a number j such that $t_j = p$. Thus, p is enumerated by step 8 of Procedure 6. If not, then there exists a set F that satisfies the condition of step 10 by Lemma 4.36(1). Let n_0 be the least n satisfying $\{t_1, \ldots, t_n\} \supseteq F$. It is clear that p is enumerated at step 11 when $n = n_0$.

We end this section by giving an example.

Example 4.40 Suppose $\Sigma = \{a, b, f, g\}, rank(a) = rank(b) = 0, rank(f) = 2, rank(g) = 1$, and $x, y \in V$. Let us consider learning $L(f(a, x)) \cup L(f(x, b)) \in \cup^{\leq 2}\mathcal{TPL}$. Let a positive presentation σ be as follows:

$$\begin{aligned} t_1 &= f(a, a),\ t_2 = f(a, f(a, b)),\ t_3 = f(b, b), \\ t_4 &= f(a, g(a)),\ t_5 = f(a, b),\ t_6 = f(g(a), b), \ldots \end{aligned}$$

This time Procedure 6 learns $L(f(a, x)) \cup L(f(x, b))$ as follows:

•$n = 1$: $\mathcal{GPD}$ outputs t_1 and Procedure 6 outputs (t_1).

•$n = 2$: $\mathcal{GPD}$ outputs t_2 and Procedure 6 outputs (t_1, t_2).

•$n = 3$: $lca(t_1, t_2) = f(a, x)$, $lca(t_1, t_3) = f(x, x)$, $lca(t_2, t_3) = f(x, y)$. Hence $\mathcal{GPD}$ outputs only t_3. Procedure 6 chooses two larger elements from $\{t_1, t_2, t_3\}$ and output them.

•$n = 4$: Since $lca(t_1, t_2) = lca(t_1, t_4) = lca(t_2, t_4) = f(a, x)$, $\mathcal{GPD}$ outputs t_4 and $f(a, x)$. Procedure 6 outputs two maximal elements of $\{t_1, \ldots, t_4, f(a, x)\}$, that is, $f(a, x)$ and t_3.

•$n = 5$: Since $lca(t_1, t_5) = lca(t_2, t_5) = lca(t_4, t_5) = f(a, x)$, $\mathcal{GPD}$ outputs t_5 and $f(a, x)$. Procedure 6 outputs $f(a, x)$ and the larger element of $\{t_3, t_5\}$.

•$n = 6$: Since $lca(t_3, t_5) = lca(t_3, t_6) = lca(t_5, t_6) = f(x, b)$, $\mathcal{GPD}$ outputs t_6 and $f(x, b)$. Procedure 6 outputs $(f(a, x), f(x, b))$.

Chapter 5

Unbounded Unions of Closed Set Systems

5.1 Learning Unbounded Unions of Closed Set Systems

In this chapter we consider unbounded unions of languages. To study the inferability of $\mathcal{L}^*$, let us start with the following proposition.

Proposition 5.1 *Let $\mathcal{L}$ be a closed set system such that every $L \in \mathcal{L}$ has a characteristic set and let $\mathcal{U}$ be the family of all finite nonempty subset of U. If there exists a mapping $\delta : \mathcal{U} \to U$ satisfying the condition*

$$(\star) \qquad \delta(S) \in L \Leftrightarrow S \subseteq L$$

for all $S \in \mathcal{U}$ and $L \in \mathcal{L}$, then every nonempty $L \in \mathcal{L}$ has a characteristic set consisting of one element.

Proof. Let S be a characteristic set of an arbitrary $L \in \mathcal{L}$. We show that $\{\delta(S)\}$ is a characteristic set of L. From the condition $(\star)$, $\{\delta(S)\} \subseteq L$. Assume that there exists a $L' \in \mathcal{L}$ such that $\delta(S) \in L'$. By applying $(\star)$ for L', we have $S \subseteq L'$. Since S is a characteristic set of L, L must be a subset of L'.

By combining Proposition 5.1 and Theorem 1.22, we have:

Corollary 5.2 *Let $\mathcal{L}$ be the same as in the previous proposition. If there is a mapping δ satisfies the condition $(\star)$, then $\mathcal{L}^*$ is inferable from positive data.*

Remark 5.3 In general, the converse of Corollary 5.2 is not always true. To show that we present an instance of it. Let $U = \mathbb{Z}_{\geq 0}$ and $\mathcal{L} = \{A \subset \mathbb{Z}_{\geq 0} \mid 0 \notin A, \sharp(A) \leq 2\} \cup \{\mathbb{Z}_{\geq 0}\}$. We define a closure operator C by

$$C(S) = \begin{cases} S & \text{if } 0 \notin S \text{ and } \sharp(S) \leq 2, \\ \mathbb{Z}_{\geq 0} & \text{else.} \end{cases}$$

This closure operator makes $\mathcal{L}$ a closed set system. Then every element of $\mathcal{L}$ is equal to a union of at most two closed sets generated from a single element. In fact, $\{a\} = C(a)$, $\{a, b\} = C(a) \cup C(b)$ and $\mathbb{Z}_{\geq 0} = C(0)$. Thus, by Theorem 1.22, $\mathcal{L}^*$ is inferable from positive data. Now, assume that there is a mapping δ that satisfies $(\star)$. According to the proof of Proposition 5.1, $\delta(\{a, b\})$ must be a characteristic set of $\{a, b\}$ in $\mathcal{L}$, nevertheless $\{a, b\}$ has no characteristic set consisted by one element. This is contradiction.

Example 5.4 As we have mentioned in §4.3, the class $\mathcal{C}^*$ is inferable from positive data. This time the mapping lca satisfies the condition $(\star)$. In fact, if $lca(S) \in C(p)$ then, for every $q \in S$, $q \preceq lca(S) \preceq p$ holds, and hence $q \in C(p)$. On the other hand, if $S \subseteq C(p)$, then $lca(S) \preceq p$ by the definition of lca. Thus $lca(S) \in C(p)$.

Example 5.5 Another simple example is the class of ideals of polynomial ring in one variable. It is known that every ideal of polynomial ring $\mathbb{Q}[x]$ in one variable over $\mathbb{Q}$ is generated by one polynomial. As we have seen in Lemma 3.1, the operation generating ideals is a closure operator. By Lemma 1.11, a generating set is a characteristic set. Therefore all ideals of $\mathbb{Q}[x]$ has a characteristic set consists of one element, thus the condition of Theorem 1.22 is satisfied.

Let I be an ideal of $\mathbb{Q}[x]$. Suppose that $I = \langle f \rangle$. Clearly $g \in I \Leftrightarrow g$ is divisible by f. Now we consider the mapping gcd that maps a finite set of polynomials F to the greatest common divisor $gcd(F)$ of F. Let $G = \{g_1, \ldots, g_s\} \subset \mathbb{Q}[x]$. If $G \subseteq I$, then for each i there exists $a_i \in \mathbb{Q}[x]$ such that $g_i = a_i f$. Hence $gcd(G)$ is divisible by f, so $gcd(G) \in I$. On the other hand if $gcd(G) \in I$, then f divides $gcd(G)$ and $gcd(G)$ divides every g_i, and thus $G \subseteq I$. Therefore the mapping gcd satisfies $(\star)$.

In the following sections, we present two examples of $\mathcal{L}$ and δ.

5.2 Learning Invariant Subspaces of a Linear Transformation of a Vector Space

Let V be an infinite dimensional vector space with countable basis over $\mathbb{Q}$. We fix a countable basis $\{\boldsymbol{g}_1, \boldsymbol{g}_2, \ldots\}$ of V. The class of all finite dimensional subspaces of V itself is not a closed set system as we have seen in §3.2. So we introduce a new element $\boldsymbol{g}_0$ that is not included in V and let $V' = \langle\{\boldsymbol{g}_0, \boldsymbol{g}_1, \boldsymbol{g}_2, \ldots\}\rangle$. We define

$$\mathcal{V} = \{W \subseteq V \mid W \text{ is a finite dimensional subspace of } V\} \cup \{V'\}.$$

Then we showed that $\mathcal{V}$ is a closed set system in Proposition 3.15, and that every $W \in \mathcal{V}$ has a characteristic set in Proposition 3.16.

We saw that $\mathcal{V}$ is not Noetherian in Remark 3.17. Nevertheless one can show that the bounded union $\cup^{\leq k}\mathcal{V}$ is also inferable. Note that we can not apply Theorem 1.17 to show the inferability of $\cup^{\leq k}\mathcal{V}$. Instead, we make use of the argument in Proposition 4.11. Choose $W = \langle \boldsymbol{w}_1, \ldots, \boldsymbol{w}_r\rangle \in \mathcal{V}$ and let M be a $(r \times (rk-1))$-matrix in $\mathbb{Q}$-entries such that any k column vectors are linearly independent. Put

$$(\boldsymbol{w}_{1'}, \ldots, \boldsymbol{w}_{rk-1'}) = (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_r)\, M.$$

Then by the same argument in the proof of [17, Theorem 9], we can show that $\{\boldsymbol{w}_{1'}, \ldots, \boldsymbol{w}_{rk-1'}\}$ is a characteristic set of W in $\cup^{\leq k}\mathcal{V}$.

On the other hand, $\mathcal{V}^*$ is not inferable. For instance, put $W = \langle \boldsymbol{g}_1, \boldsymbol{g}_2\rangle$. Choose any finite subset $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\}$ of W. Then, the union $\langle \boldsymbol{v}_1\rangle \cup \ldots \cup \langle \boldsymbol{v}_r\rangle$ of 1-dimensional vector spaces $\langle \boldsymbol{v}_i\rangle$ $(i = 1, \ldots, r)$ is proper subset of W. This means that any finite subset of W can not be a finite tell-tale. Thus $\mathcal{V}^*$ is not inferable.

Now we consider invariant subspaces of a certain linear transformation. Let $T : V \to V$ be a fixed linear transformation such that $T(\boldsymbol{g}_i) = a_i\boldsymbol{g}_i$ $(a_i \in \mathbb{Q})$ for each i. Here we assume that a_i's are distinct. Then the property T-invariant is characterized by the following:

Lemma 5.6 *Let $W \neq \{\boldsymbol{0}\}$ be a finite dimensional subspace of V. W is T-invariant if and only if there exists $\boldsymbol{g}_{i_1}, \ldots, \boldsymbol{g}_{i_n}$ such that $W = \langle \boldsymbol{g}_{i_1}, \ldots, \boldsymbol{g}_{i_n}\rangle$.*

Proof. If there exists $\boldsymbol{g}_{i_1}, \ldots, \boldsymbol{g}_{i_n}$ such that $W = \langle \boldsymbol{g}_{i_1}, \ldots, \boldsymbol{g}_{i_n}\rangle$, then clearly W is T-invariant. Suppose that W is T-invariant. Let $\boldsymbol{w}$ be a nonzero vector of W. We can write $\boldsymbol{w} = c_1\boldsymbol{g}_{k_1} + \ldots + c_m\boldsymbol{g}_{k_m}$, where every c_i is not zero. Since W is T-invariant, $T(\boldsymbol{w}) = c_1 a_{k_1}\boldsymbol{g}_{k_1} + \ldots + c_m a_{k_m}\boldsymbol{g}_{k_m} \in W$. Similarly, $T^2(\boldsymbol{w}), \ldots, T^{m-1}(\boldsymbol{w}) \in W$. Now, let A be the $(m \times m)$-matrix on the right in the equation

$$\big(\boldsymbol{w}, T(\boldsymbol{w}), \ldots, T^{m-1}(\boldsymbol{w})\big) = \big(\boldsymbol{g}_{k_1}, \ldots, \boldsymbol{g}_{k_m}\big) \begin{pmatrix} c_1 & c_1 a_{k_1} & \ldots & c_1 a_{k_1}^{m-1} \\ c_2 & c_2 a_{k_2} & \ldots & c_2 a_{k_2}^{m-1} \\ \vdots & \vdots & & \vdots \\ c_m & c_m a_{k_m} & \ldots & c_m a_{k_m}^{m-1} \end{pmatrix}.$$

A is a matrix obtained by multiplying each column of a Vandermonde's matrix by nonzero scalars. Since a_i's are distinct,

$$\det A = \left(\prod_{j=1}^{m} c_j\right) \cdot \left(\prod_{p<q} (a_{k_p} - a_{k_q})\right) \neq 0,$$

so there is the inverse matrix A^{-1}. This means that $\boldsymbol{g}_{k_j}$'s can be written by linear combinations of $\boldsymbol{w}, T(\boldsymbol{w}), \ldots, T^{m-1}(\boldsymbol{w})$. Thus, $G = \{\boldsymbol{g}_{k_1} \cdots \boldsymbol{g}_{k_m}\} \subseteq W$. If $\langle G \rangle \neq W$, repeat the method above for $\boldsymbol{w}^{(1)} \in W \setminus \langle G \rangle$, and add the yielding $\{\boldsymbol{g}_{k_1}^{(1)}, \ldots, \boldsymbol{g}_{k_{m_1}}^{(1)}\}$ to G. This procedure ends in finite number of steps since W is finite dimensional.

Let $\mathcal{V}_T$ be the class as follows:

$$\begin{aligned} \mathcal{V}'_T &= \{W \subseteq V \mid W \text{ is a finite dimensional } T\text{-invariant subspaces of } V\}, \text{ and} \\ \mathcal{V}_T &= \mathcal{V}'_T \cup \{V'\}. \end{aligned}$$

Then, we have

Lemma 5.7 *$\mathcal{V}_T$ is a closed set system.*

Proof. By Proposition 1.7, it suffices to show that (1) $\mathcal{V}_T$ is intersection closed, and (2) for every $S \in V'$, there is $W \in \mathcal{V}_T$ such that $S \subseteq W$.
(1) Let $\{W_i\}$ be elements of $\mathcal{V}_T$. If there is i_0 such that $W_{i_0} = \emptyset$, then clearly $\cap W_i = \emptyset \in \mathcal{V}_T$. Thus we suppose that all W_i's are nonempty. By Lemma 5.6, if $W_i \neq V'$ then it can be written by $W_i = \langle \boldsymbol{g}_{k,1}^{(i)}, \ldots, \boldsymbol{g}_{k,m_i}^{(i)} \rangle$. In addition, we defined $V' = \langle \boldsymbol{g}_0, \boldsymbol{g}_1, \ldots \rangle$. Thus it holds that either $\cap W_i$ is generated some finite subset of $\{\boldsymbol{g}_1, \boldsymbol{g}_2, \ldots\}$ or $\cap W_i = V'$. In both cases $\cap W_i$ is in $\mathcal{V}_T$.
(2) Obviously $S \subseteq V' \in \mathcal{V}_T$.

Remark 5.8 $\mathcal{V}_T'$ does not satisfies (2). Thus $\mathcal{V}_T'$ is not a closed set system. Nevertheless, we show that $\mathcal{V}_T'^{\star}$ is inferable later.

Lemma 5.9 *Every $W \in \mathcal{V}_T$ has a characteristic set of the form $\{\boldsymbol{g}_{i_1}, \ldots, \boldsymbol{g}_{i_n}\}$ in $\mathcal{V}_T$.*

Proof. It is clearly that V' has a characteristic set $\{\boldsymbol{g}_0\}$. Let $W \neq V'$ be the element of $\mathcal{V}_T$. By Lemma 5.6, there exists $\{\boldsymbol{g}_{k_1}, \ldots, \boldsymbol{g}_{k_m}\}$ such that $W = \langle \boldsymbol{g}_{k_1}, \ldots, \boldsymbol{g}_{k_m} \rangle$. Then $\{\boldsymbol{g}_{k_1}, \ldots, \boldsymbol{g}_{k_m}\}$ is a characteristic set of W: if $\{\boldsymbol{g}_{k_1}, \ldots, \boldsymbol{g}_{k_m}\} \subseteq W' \in \mathcal{V}_T$, then $W = \langle \boldsymbol{g}_{k_1}, \ldots, \boldsymbol{g}_{k_m} \rangle \subseteq W'$ since the property of $\langle \cdot \rangle$.

Now we show that $\mathcal{V}_T^*$ is inferable.

Theorem 5.10 $\mathcal{V}_T^*$ *is inferable from positive data.*

Proof. First we define the mapping δ as follows. Let $S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\} \subseteq V'$. For each i, we can write $\boldsymbol{v}_i = \sum_{\text{finite}} c_{i,j} \boldsymbol{g}_j \quad (c_{i,j} \in \mathbb{Q})$. Let $A_i = \{\boldsymbol{g}_j \mid c_{i,j} \neq 0\}$. $\delta(S)$ is defined as follows:

$$A = \bigcup_{i=1}^{n} A_i, \ \delta(S) = \sum_{\boldsymbol{g} \in A} \boldsymbol{g}.$$

According to Corollary 5.2, it suffices to show that δ satisfies

$$(\star) \qquad \delta(S) \in W \Leftrightarrow S \subseteq W$$

for each finite subset $\emptyset \neq S \subseteq V$ and for each $W \in \mathcal{V}_T$. Let W be an arbitrary element of $\mathcal{V}_T$. If $W = \langle \mathbf{0} \rangle$ then $S = \{\mathbf{0}\} \Leftrightarrow \delta(S) = \mathbf{0}$, hence $(\star)$ holds. If $W = V'$ then $(\star)$ is obvious. So we assume that $W \neq \langle \mathbf{0} \rangle, V'$. By Lemma 5.6, we can write $W = \langle \boldsymbol{g}_{i_1}, \ldots, \boldsymbol{g}_{i_n} \rangle$, where $i_j > 0$.
$(\Rightarrow)$ By applying the argument in the proof of Lemma 5.6 to $\delta(S) = \sum_{\boldsymbol{g} \in A} \boldsymbol{g} \in W$, we get that A is a subset of $\{\boldsymbol{g}_{i_1}, \ldots, \boldsymbol{g}_{i_n}\}$, hence A_i is. Since $\boldsymbol{v}_i$ is a linear combination of the elements of A_i, each $\boldsymbol{v}_i$ is in W.
$(\Leftarrow)$ $S \subseteq W$ implies that $A_i \subseteq \{\boldsymbol{g}_{i_1}, \ldots, \boldsymbol{g}_{i_n}\}$ for every i, so A is. Thus $\delta(S) \in W$.

A concrete inference algorithm of $\mathcal{V}_T^*$ is shown as follows:

Procedure 7: Learning $\mathcal{V}_T^*$;

Input: a positive presentation $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n, \ldots$ for $V_1 \cup \ldots \cup V_m \in \mathcal{V}_T^*$;
Output: a sequence $V_1^{(1)} \cup \ldots \cup V_{m_1}^{(1)}, V_1^{(2)} \cup \ldots \cup V_{m_2}^{(2)}, \ldots$ of elements of $\mathcal{V}_T^*$;
begin
$S = \emptyset$;
Put $n = 1$;
repeat
  if there is no $W \in S$ such that $\boldsymbol{v}_n \in W$ **then begin**
    Set $A_n = \{\boldsymbol{g}_j \mid \boldsymbol{g}_j$ occurs in $\boldsymbol{v}_n$ with nonzero coefficient $\}$;
    if $\boldsymbol{g}_0 \in A_n$ **then** Set $A_n = \{\boldsymbol{g}_0, \boldsymbol{g}_1, \boldsymbol{g}_2, \ldots\}$;
    Remove all $W' \in S$ such that $W' \subsetneq \langle A_n \rangle$ from S;
    Add $\langle A_n \rangle$ to S;
  end;
  Output $\cup_{W \in S} W$;
  Add 1 to n;
forever;
end.

Theorem 5.11 *$\mathcal{V}_T^*$ is identifiable in the limit from positive data via Procedure 7.*

Proof. If there is i such that $V_i = V'$, then there is n that it holds $\boldsymbol{g}_0 \in A_n$ for the first time. At this step V' is added to S and afterwards the algorithm keeps on outputting V'. (Note that this time the target language is V' since Remark 1.21).

Then suppose that there is no V_i such that $V_i = V'$. By Lemma 5.6, V_i can be expressed as $V_i = \langle \boldsymbol{g}_{k,1}^{(i)}, \ldots, \boldsymbol{g}_{k,m_i}^{(i)} \rangle$. For each i, there is n_i such that $\boldsymbol{v}_i$ is an element of the form $c_1 \boldsymbol{g}_{k,1}^{(i)} + \ldots + c_{m_i} \boldsymbol{g}_{k,m_i}^{(i)}$, where all c_j's are nonzero, for the first time. At this step V_i is added to S. Thus after the step $n = \max\{n_i\}$, the algorithm outputs $V_1 \cup \ldots \cup V_m$.

Moreover,

Corollary 5.12 *$\mathcal{V}_T'^*$ is identifiable via Procedure 7.*

Proof. This statement clearly follows from the fact:

$$\text{the target language is not } V' \Leftrightarrow \boldsymbol{g}_0 \text{ is never appeared in } A_n.$$

We end this section with presenting an example of $\mathcal{V}$ and T.

Example 5.13 Let V be the subspace of the vector space consisting of Fourier series as follows:

$$V = \left\{ \sum_{n=0}^{r} a_n \cos nx \;\middle|\; a_n \in \mathbb{Q},\ r \in \mathbb{Z}_{\geq 0} \right\}.$$

The set $\{1, \cos x, \cos 2x, \dots\}$ forms a basis of V. Let $T : V \ni f \mapsto \frac{d^2 f}{dx^2} \in V$. Then, $T(1) = 0, T(\cos x) = -\cos x, T(\cos 2x) = -4\cos 2x, \dots$, so the class of unbounded unions of finite dimensional T-invariant subspaces is inferable from positive data since Corollary 5.12.

This situation can be generalized to that of Hilbert spaces. A vector space H over the field of real or complex numbers is called Hilbert space if an inner product is defined over H and H is complete with respect to the metric induced by the inner product. It is known that H has an orthonormal basis under a certain condition. This means that every element of H can be approximated by a linear combination of finite number of elements of the orthonormal basis within an arbitrary error. This example can be considered to be the situation treating an approximation cut off after the r-th term. A Hilbert space is one of the most typical and important examples of infinite dimensional vector spaces which appear in mathematics, and it is closely related to functional analysis and approximation theory. This might shed new light on the connection between learning theory and analysis.

5.3 Learning Monomial Ideals of Polynomial Ring

As we have seen in Lemma 3.5, the class of all monomial ideals $\mathcal{MI}$ can be regarded as a Noetherian closed set system. According to Lemma 3.4, the closure operator associated with $\mathcal{MI}$ is the same as the mapping $\langle\cdot\rangle$ if it is restricted on $2^{\mathcal{M}}$. So we treat $\langle\cdot\rangle$ as the closure operator of $\mathcal{MI}$, provided that it deal with only monomials.

First we show $\mathcal{MI}^*$ is inferable by defining δ.

Theorem 5.14 *$\mathcal{MI}^*$ is inferable from positive data.*

Proof. Let $S = \{s_1, \dots, s_n\} \subseteq \mathbb{Q}[x_1, \dots, x_n]$. For each i, we can write $s_i = \sum_{m \in \mathcal{M}} c_{i,m} m$, where all but finite $c_{i,m}$'s are equal to 0. Let $M_i = \{m \mid c_{i,m} \neq 0\}$. $\delta(S)$ is defined as follows:

$$M = \bigcup_{i=1}^{n} M_i, \ \delta(S) = \sum_{m \in M} m.$$

According to Corollary 5.2, it suffices to show that δ satisfies

$$\delta(S) \in I \Leftrightarrow S \subseteq I$$

for each finite subset $S \subseteq \mathbb{Q}[x_1, \dots, x_n]$ and for each $I \in \mathcal{MI}$.
($\Rightarrow$) Suppose $\delta(S) = \sum_{m \in M} m \in I$. Applying Proposition 2.5(c) to $\delta(S)$, the set of all monomials occurring in $\delta(S)$, that is M, is a subset of I. Hence $M_i \subseteq I$ for each i. Since s_i is a linear combination of the elements of M_i, each s_i is in I.
($\Leftarrow$) $S \subseteq I$ implies that $M_i \subseteq I$ for every i, so M is. Thus $\delta(S) \in I$.

A learning algorithm of $\mathcal{MI}^*$ resembles Procedure 7.

Procedure 8: Learning $\mathcal{MI}^*$;
Input: a positive presentation $f_1, f_2, \dots, f_n, \dots$ for $I_1 \cup \dots \cup I_m \in \mathcal{MI}^*$;
Output: a sequence $V_1^{(1)} \cup \dots \cup V_{m_1}^{(1)}, V_1^{(2)} \cup \dots \cup V_{m_2}^{(2)}, \dots$ of elements of $\mathcal{MI}^*$;
begin
  $S = \emptyset$;
  Put $n = 1$;
  repeat
    if there is no $I \in S$ such that $f_n \in I$ **then begin**
      Set $A_n = \{m_i \mid m_i$ occurs in f_n with nonzero coefficient$\}$;
      Remove all $J \in S$ such that $J \subsetneq \langle A_n \rangle$ from S;
      Add $\langle A_n \rangle$ to S;
    end;
    Output $\cup_{I \in S} I$;
    Add 1 to n;
  forever;
end.

Example 5.15 Consider a monomial ideal $I = \langle x^3, xy, y^2 \rangle$ of the polynomial ring in two variables $\mathbb{Q}[x, y]$. I equals to the set

$$\left\{ \sum_{\text{finite}} c_m m \;\middle|\; c_m \in \mathbb{Q}, m \in \mathcal{M} \text{ is divisible by } x^3, xy \text{ or } y^2 \right\}.$$

Then, the set

$$\{\delta(\{x^3, xy, y^2\})\} = \{x^3 + xy + y^2\}$$

forms a characteristic set of I in $\mathcal{MI}$.

5.4 Closed Set Systems and Transaction Databases

In this section, we apply our arguments of closed set systems considered in §§5.2 and 5.3 to the study of transaction databases.

5.4.1 Vector Spaces and Transaction Databases

Let $I = \{p_1, p_2, \ldots\}$ be the set of all items and let V be the set of all formal linear combinations of elements of I:

$$V = \left\{ \sum_{\text{finite}} c_k p_k \;\middle|\; c_k \in \mathbb{Q}, p_k \in I \right\}.$$

I forms a countable basis of V. We take a fixed linear transformation $T : V \to V$ that is the form $T(p_i) = a_i p_i$, where a_i's are distinct rational numbers (for example, a linear transformation T defined by $T(p_i) = ip_i$ is suitable). By Lemma 5.6, we have

$$\mathcal{V}'_T = \{\langle S \rangle \mid S \subset I, S \text{ is finite}\}.$$

Similar to the argument of §5.2, we introduce p_0 and let $V' = \langle p_0, p_1, p_2, \ldots \rangle$, $\mathcal{V}_T = \mathcal{V}'_T \cup \{V'\}$. Lemma 5.7 says that $\mathcal{V}_T$ is a closed set system. Then,

Lemma 5.16 *A finite subset of I, that is an itemset, can be regarded as a closed set of $\mathcal{V}_T$.*

Proof. Let $\mathcal{S}$ be the class of all itemsets. By Lemma 5.6, for each $W \in \mathcal{V}_T$, if $W \neq V'$ then there exists unique $S_W \in \mathcal{S}$ such that $W = \langle S_W \rangle$. Then the mapping

$$\Phi : \mathcal{V}_T \to \mathcal{S}, \quad W \mapsto S_W$$

is the inverse of

$$\Psi : \mathcal{S} \to \mathcal{V}_T, \quad S \mapsto \langle S \rangle.$$

Thus an itemset corresponds to an element of $\mathcal{V}'_T$, that is $\mathcal{V}_T \setminus \{V'\}$.

Now we apply our result in §5.2 to consider the inference of $\mathcal{V}'^{\star}_T$. Let $W_1 \cup \ldots \cup W_m \in \mathcal{V}'^{\star}_T$ be a target and $\sigma : \boldsymbol{v}_1, \boldsymbol{v}_2, \ldots$ be a positive data of $W_1 \cup \ldots \cup W_m$. As we have seen in Corollary 5.12, $W_1 \cup \ldots \cup W_m$ is inferable from $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots$.

Here we define an itemset $X_i \subset I$ by

$$X_i = \{p_i \mid p_i \text{ appears in } \boldsymbol{v_i} \text{ with nonzero coefficient}\}.$$

Since every X_i is finite, the sequence $\{X_1, X_2, \ldots\}$ forms a transaction database, which we denote by $\mathcal{D}$. Put $C_i = \{p_i \mid p_i \in W_i\}$ $(= \Phi(W_i))$. From the correspondence in Lemma 5.16, C_i is the subset of I that corresponds to W_i. Then it holds that:

Proposition 5.17 *C_i is DB-closed. Moreover, C_i is a maximal DB-closed set with respect to set inclusion.*

Proof. Let $C_i \subsetneq X \subsetneq I$ be an arbitrary finite set. Since $W_1 \cup \ldots \cup W_m$ is not redundant, there is no W_k such that $\Psi(X) = \langle X \rangle \subseteq W_k$. This means that there is no j such that $X_j \supseteq X$. Hence $t(X) = \emptyset$. On the other hand, $t(C_i) \neq \emptyset$, since $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots$ is positive data and so there is a $\boldsymbol{v}_k$ that is the form $\sum_{p \in C_i} c_p p$ $(c_p \in \mathbb{Q})$. Thus $t(X) \subsetneq t(C_i)$, and then C_i is DB-closed. In addition, $t(X) = \emptyset$ implies that $X \supsetneq C_i$ can not be DB-closed. So C_i is maximal.

Furthermore, we have:

Proposition 5.18 *If X is a maximal DB-closed set of $\mathcal{D}$, then there exists unique i such that $C_i = X$.*

Proof. If $t(X) = \emptyset$ then X is not DB-closed, so $t(X) \neq \emptyset$. Let $j \in t(X)$. There is a W_k such that $\boldsymbol{v}_j \in W_k$. By definition, it holds that $X \subseteq X_j \subseteq C_k$. The assumption and Proposition 5.17 imply $X = C_k$. If there is a $k' \neq k$

such that $X = C_k = C_{k'}$, then it means that $W_1 \cup \ldots \cup W_m$ is redundant, and this is a contradiction.

Thus "learning $\mathcal{V}_T'^{\star}$" corresponds to "mining maximal closed sets of $\mathcal{D}$".

Remark 5.19 This example would be uninteresting, since a maximal DB-closed itemset is only a maximal itemset. But, if we choose the linear transformation T appropriately or make coefficients of $\boldsymbol{g_i}$'s have a meaning, it might be interesting.

Example 5.20 In practice, both the set of items I and the database $\mathcal{D}$ are finite. Here we consider such a case as an example. Let $I = \{p_1, p_2, p_3, p_4\}$ and $\mathcal{D}$ as follows:

	p_1	p_2	p_3	p_4
X_1	○	○	○	
X_2		○	○	
X_3				○
X_4		○	○	
X_5	○			○

Then DB-closed sets of $\mathcal{D}$ are

$$\{p_1\}, \{p_4\}, \{p_1, p_4\}, \{p_2, p_3\}, \{p_1, p_2, p_3\}$$

and maximal DB-closed sets are $\{p_1, p_4\}$ and $\{p_1, p_2, p_3\}$. Now let $\boldsymbol{v}_i \in V$ $(i = 1, \ldots, 5)$ be the element corresponding to X_i:

$$\boldsymbol{v}_1 = p_1 + p_2 + p_3, \boldsymbol{v}_2 = p_2 + p_3, \boldsymbol{v}_3 = p_4, \boldsymbol{v}_4 = p_2 + p_3, \boldsymbol{v}_5 = p_1 + p_4.$$

Then Procedure 7 outputs $\langle p_1, p_2, p_3 \rangle \cup \langle p_1, p_4 \rangle$ when $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_5$ is taken as input. Since

$$\Phi(\langle p_1, p_2, p_3 \rangle) = \{p_1, p_2, p_3\},\ \Phi(\langle p_1, p_4 \rangle) = \{p_1, p_4\},$$

the maximal DB-closed sets of $\mathcal{D}$ can be given by Procedure 7.

5.4.2 Monomial Ideals and Transaction Databases

We have consider an application of Theorem 5.14 to the study of transaction databases. Our advantage is that, by using Theorem 5.14, we can deal with transaction databases that contains data of quantities of items. Let $I = \{x_1, x_2, \ldots x_n\}$ be the set of items. An itemset considered in this section is a set of pairs of an item and its quantity $\{(x_{a_1}, p_1), (x_{a_2}, p_2), \ldots, (x_{a_k}, p_k)\}$, $p_i \in \mathbb{N}$. For convenience, we always assume that $a_1 < a_2 < \ldots < a_k$. A transaction database $\mathcal{D}$ is a sequence of itemsets $\{T_1, T_2, \ldots\}$.

Definition 5.21 Let $T = \{(x_{a_1}, p_1), (x_{a_2}, p_2), \ldots, (x_{a_k}, p_k)\}$ and $T' = \{(x_{b_1}, q_1), (x_{b_2}, q_2), \ldots, (x_{b_l}, q_l)\}$ be itemsets. If there is an j_i such that $b_{j_i} = a_i$ and $p_i \leq q_{j_i}$ for each $i = 1, 2, \ldots, k$, then T is said to be included in T', and denotes by $T \preceq T'$.

Let $\mathcal{T}$ be the set of all itemsets. Formally we define $T_\infty = \{(x_1, \infty), (x_2, \infty), \ldots, (x_n, \infty)\}$ and $T \preceq T_\infty$ for all $T \in \mathcal{T}$. We denote $\mathcal{T}' = \mathcal{T} \cup \{T_\infty\}$. The next lemma obviously follows from definition.

Lemma 5.22 *$\preceq$ is a partial order on $\mathcal{T}$ or $\mathcal{T}'$. That is, $\preceq$ satisfies the following two conditions: (1) $T \preceq T'$ and $T' \preceq T$ if and only if $T = T'$, and (2) if $T \preceq T'$ and $T' \preceq T''$ then $T \preceq T''$.*

There is a natural correspondence between $\mathcal{T}$ and the set of all monomials $\mathcal{M}$ as follows.

Definition 5.23 1. Let $T = \{(x_{a_1}, p_1), (x_{a_2}, p_2), \ldots, (x_{a_k}, p_k)\}$ be an transaction. We define a monomial $\mu(T) = x_{a_1}^{p_1} x_{a_2}^{p_2} \ldots x_{a_k}^{p_k}$.
2. Let $m = x_{a_1}^{p_1} x_{a_2}^{p_2} \ldots x_{a_k}^{p_k} \in \mathcal{M}$ be an monomial. We define an itemset $\tau(m) = \{(x_{a_1}, p_1), (x_{a_2}, p_2), \ldots, (x_{a_k}, p_k)\}$.

These mappings define bijections between $\mathcal{T}$ and $\mathcal{M}$. Moreover, if we define $m_\infty = x_1^\infty x_2^\infty \ldots x_n^\infty$ formally and define $\tau(m_\infty) = T_\infty$ and $\mu(T_\infty) = m_\infty$, then τ and μ define a correspondence between $\mathcal{T}'$ and $\mathcal{M} \cup \{m_\infty\}$. We denote $\mathcal{M}' = \mathcal{M} \cup \{m_\infty\}$.

Example 5.24 Here we give an example of the correspondence.

	x_1	x_2	x_3	x_4
T_1	2			1
T_2		1	3	2
T_3	1			
T_4	1	2	1	

$$\overset{\mu}{\longrightarrow} \qquad \underset{\tau}{\longleftarrow}$$

m_1	$x_1^2 x_4$
m_2	$x_2 x_3^3 x_4^2$
m_3	x_1
m_4	$x_1 x_2^2 x_3$

Then $\mu(T_i) = m_i$ and $\tau(m_i) = T_i$ in the tables above.

Clearly it holds that:

Lemma 5.25 *Let T and T' be itemsets. Then $T \preceq T'$ if and only if $\mu(T)|\mu(T')$.*

Next we define the mappings that correspond to ι and t in §1.4.

Definition 5.26 1. Let T be an itemset and $\mathcal{D} = \{T_1, T_2, \ldots\}$ be a transaction database. We define $t_{\mathcal{D}} : \mathcal{T}' \to 2^{\mathbb{N}}$ by

$$t_{\mathcal{D}}(T) := \{a \in \mathbb{N} \mid T \preceq T_a\} = \{a \in \mathbb{N} \mid \mu(T)|\mu(T_a)\}.$$

2. Let $A \subseteq \mathbb{N}$. We define $\iota_{\mathcal{D}} : 2^{\mathbb{N}} \to \mathcal{T}'$ by

$$\iota_{\mathcal{D}}(A) := \max_{\preceq}\{T \in \mathcal{T} \mid \forall a \in A, T \preceq T_a\} = \tau(\gcd\{\mu(T_a) \mid a \in A\}).$$

If $A = \emptyset$, we define $\iota_{\mathcal{D}}(\emptyset) = T_\infty$.

For simplicity, we will write ι and t instead of $\iota_{\mathcal{D}}$ and $t_{\mathcal{D}}$, respectively. Similar to Lemma 1.24 and Lemma 1.27, it holds that:

Lemma 5.27 *1. Let $T, T' \in \mathcal{T}'$ and suppose that $T \preceq T'$. Then $t(T) \supseteq t(T')$.*
2. Let $A, B \subseteq \mathbb{N}$ and suppose that $A \subseteq B$. Then $\iota(A) \succeq \iota(B)$.

Proof. 1. If $t(T) = \emptyset$ then the statement is obvious. Suppose that $t(T) \neq \emptyset$. Let $a \in t(T')$. By definition $T' \preceq T_a$. By applying Lemma 5.22, we have $T \preceq T_a$. This means $a \in t(T)$.
2. Since $A \subseteq B$, $\gcd\{\mu(T_a) \mid a \in B\} \mid \gcd\{\mu(T_a) \mid a \in A\}$. Thus $\iota(B) \preceq \iota(A)$.

Proposition 5.28 *Let $\mathcal{D} = \{T_1, T_2, \ldots\}$ be a fixed transaction database and ι, t be the mappings above. The composition $C_{\mathcal{D}} = \iota \circ t : \mathcal{T}' \to \mathcal{T}'$ satisfies the following conditions: for each $T, T' \in \mathcal{T}'$,*
(CO1') *$T \preceq C_{\mathcal{D}}(T)$,*
(CO2') *$T \preceq T' \Rightarrow C_{\mathcal{D}}(T) \preceq C_{\mathcal{D}}(T')$, and*
(CO3') *$C_{\mathcal{D}}(C_{\mathcal{D}}(T)) = C_{\mathcal{D}}(T)$.*

Proof. (CO1') Let $T \in \mathcal{T}'$. Since $C_{\mathcal{D}}(T_\infty) = T_\infty$ (note that $t(T_\infty) = \emptyset$ because all T_i's are itemsets, that is, elements of $\mathcal{T}$), we can assume that $T \neq T_\infty$. Clearly it holds that $\mu(T)|\gcd\{\mu(T_a) \mid \mu(T)|\mu(T_a)\}$, hence we have $T \preceq \tau(\gcd\{\mu(T_a) \mid \mu(T)|\mu(T_a)\}) = \tau(\gcd\{\mu(T_a) \mid a \in t(T)\}) = \iota(t(T))$.

(CO2') $\mu(T)|\mu(T')$ since $T \preceq T'$, so we have $\{\mu(T_a) \mid \mu(T)|\mu(T_a)\} \supseteq \{\mu(T_a) \mid \mu(T')|\mu(T_a)\}$, and thus $\gcd\{\mu(T_a) \mid \mu(T')|\mu(T_a)\}$ is divisible by $\gcd\{\mu(T_a) \mid \mu(T)|\mu(T_a)\}$. Therefore it holds that $\iota(t(T)) = \tau(\gcd\{\mu(T_a) \mid \mu(T)|\mu(T_a)\}) \preceq \tau(\gcd\{\mu(T_a) \mid \mu(T')|\mu(T_a)\}) = \iota(t(T'))$.
(CO3')By definition, $\mu(C_{\mathcal{D}}(C_{\mathcal{D}}(T))) = \gcd\{\mu(T_a) \mid \mu(C_{\mathcal{D}}(T))|\mu(T_a)\}$ and $\mu(C_{\mathcal{D}}(T)) = \gcd\{\mu(T_a) \mid \mu(T)|\mu(T_a)\}$. Now it holds that $\mu(T_k)$ is divisible by $\mu(T)$ if and only if $\mu(T_k)$ is divisible by $\gcd\{\mu(T_a) \mid \mu(T)|\mu(T_a)\}$. Hence $\{\mu(T_a) \mid \mu(T)|\mu(T_a)\} = \{\mu(T_a) \mid \mu(C_{\mathcal{D}}(T))|\mu(T_a)\}$, and therefore $C_{\mathcal{D}}(C_{\mathcal{D}}(T)) = C_{\mathcal{D}}(T)$.

Therefore $C_{\mathcal{D}}$ can be regarded as a variety of closure operator. Here we define that $T \in \mathcal{T}$ is $C_{\mathcal{D}}$-closed if $C_{\mathcal{D}}(T) = T$. Then one can show the following proposition as an analogy to Proposition 1.30.

Proposition 5.29 *Let $T \in \mathcal{T}$. T is $C_{\mathcal{D}}$-closed if and only if T is DB-closed.*

Now we fix a transaction database $\mathcal{D}$. As $\mathcal{D}$ is regarded to a sequence of monomials, the argument in §5.3 can be applied to $\mathcal{D}$. The algorithm becomes much simpler since the sequence presented as a positive data is a sequence of only monomials, instead of a sequence of polynomials.

Procedure 9-1:
Input: a sequence of monomials $m_1, m_2, \ldots$;
Output: a sequence of a set of monomials $S_1, S_2, \ldots$;
begin
    $S = \emptyset$;
    Put $n = 1$;
    repeat
        if there is no $m' \in S$ such that $m'|m_n$ **then begin**
            Remove all $m' \in S$ such that $m_n|m'$ from S;
            Add m_n to S;
        end;
        Set $S_n = S$ and output S;
        Add 1 to n;
    forever;
end.

It is clear that

Proposition 5.30 *Let S_n be the output of Procedure 9-1 at the n-th step. Then S_n is equal to the set $\{m_i \mid m_j \nmid m_i \quad (i, j = 1, 2, \ldots, n)\}$.*

By translating Proposition 5.30 into the language of transaction databases, it follows that:

Proposition 5.31 *Let S_n be the same as Proposition 5.30. S_n is the set of all maximal DB-closed itemset of $\{m_1, m_2, \ldots, m_n\}$.*

Now we improve Procedure 9-1. We set a threshold k as follows.

Procedure 9-2:
Input: a sequence of monomials $m_1, m_2, \ldots$ and a threshold k;
Output: a sequence of a set of monomials $S_1, S_2, \ldots$;
begin
$S = \emptyset$, $M = \emptyset$;
Put $n = 1$;
repeat
  Set $m = m_n$.
  if there is an element of the form (c_m, m) in M, **then** $c_m = c_m + 1$;
  else set $c_m = 1$ and add (c_m, m) to M;
  if $c_m \geq k$ and there is no $m' \in S$ such that $m'|m$ **then begin**
    Remove all $m' \in S$ such that $m|m'$ from S;
    Add m to S;
  end;
  Set $S_n = S$ and output S;
  Add 1 to n;
forever;
end.

Note that we can take quantities as well as frequency into account in Procedure 9-2. This is an advantage of Procedure 9-2.

Chapter 6

Conclusion

In this thesis we have constructed learning algorithms for the classes of bounded and unbounded unions of closed set systems concretely under certain conditions. As we have observed through several examples, closed set systems are closely related to some algebraic objects, such as vector spaces. As a result, we showed that the scheme of learning from positive data can be applied to some objects in both algebra and analysis, such as polynomial rings or Hilbert spaces. Moreover, we have seen that our algorithm can be applied to the study of transaction databases in the last section. In this way, the notion of characteristic set and its computability plays important role to give a learning algorithm of bounded or unbounded unions of closed set systems.

Bibliography

[1] D. Angluin: Inductive Inference of Formal Languages from Positive Data, Information and Control **45**, 117–135, 1980.

[2] M. de Brecht, M. Kobayashi, H. Tokunaga and A. Yamamoto: Inferability of Closed Set Systems From Positive Data, New Frontiers in Artificial Intelligence, Lecture Notes in Artificial Intelligence **4384**, 265–275, 2007.

[3] H. Arimura, T. Shinohara, S. Otsuki: Polynomial Time Inference of Unions of Tree Pattern Languages, Proc. the 2nd Workshop on Algorithmic Learning Theory (ALT'91), 105–114, 1991.

[4] H. Arimura, T. Shinohara, S. Otsuki: A Polynomial Time Algorithm for Finding Finite Unions of Tree Pattern Languages, Proc. the 2nd International Workshop on Nonmonotonic and Inductive Inference (NIL'91), Lecture Notes in Artificial Intelligence **659**, 118–131, 1991.

[5] M. de Brecht, M. Kobayashi, H. Tokunaga and A. Yamamoto: Inferability of Closed Set Systems From Positive Data, New Frontiers in Artificial Intelligence (Lecture Notes in Artificial Intelligence **4384**), 265–275, 2007.

[6] D. Cox, J. Little and D. O'Shea: *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*, Springer, 1992.

[7] E. M. Gold: Language Identification in the Limit, Information and Control **10**, 447–474, 1967.

[8] S. Hayashi: Mathematics Based on Learning, Proc. of 13th International Conference on Algorithmic Learning Theory, Lecture Notes in Artificial Intelligence **2533**, 7–21, 2002.

[9] S. Kapur and G. Bilardi: On uniform learnability of language families, Information Processing Letters **44**, 35–38, 1992.

[10] S. Kobayashi: Approximate Identification, Finite Elasticity and Lattice Structure of Hypothesis Space, Technical Report, CSIM 96-04, Dept. of Compt. Sci. and Inform. Math., Univ. of Electro-Communications, 1996.

[11] Y. Kameda and H. Tokunaga: Inferability of Unbounded Unions of Certain Closed Set Systems, New Frontiers in Artificial Intelligence, Lecture Notes in Artificial Intelligence **6284**, 317–330, 2010.

[12] Y. Kameda, H. Tokunaga and A. Yamamoto: Learning bounded unions of Noetherian closed set systems via characteristic sets, Grammatical Inference: Algorithms and Applications, Lecture Notes in Artificial Intelligence **5278**, 98–110, 2008.

[13] J. L. Lassez, M. J. Maher, and K. Marriott: Unification Revisited, in Minker J. (ed): *Foundations of Deductive Databases and Logic Programming*, 587–626, Morgan-Kaufman, 1988.

[14] T. Motoki, T. Shinohara and K. Wright.: The Correct Definition of Finite Elasticity: Corrigendum to Identification of Unions, Proceedings of COLT'91, 1991.

[15] T. Shinohara and H. Arimura: Inductive Inference of Unbounded Unions of Pattern Languages From Positive Data, Theoretical Computer Science **241**, 191–209, 2000.

[16] F. Stephan and Y. Ventsov: Learning Algebraic Structures from Text, Theoretical Computer Science, **268**, 221–273, 2001.

[17] I. Takamatsu, M. Kobayashi, H. Tokunaga, and A. Yamamoto: Computing Characteristic Sets of Bounded Unions of Polynomial Ideals, Lecture Notes in Artificial Intelligence **4914**, 318–329, 2008.

[18] H. Tokunaga and A. Yamamoto: Inductive Inference of Bounded Unions of Languages with Complete Characteristic Sets (in Japanese), SIG-FPAI-A703-01, 01–08, 2008.

[19] K. Wright: Identification of Unions of Languages Drawn from an Identifiable Class, in Proc. of COLT'89, 328–388, Morgan-Kaufman, 1989.