Adding a Path Connectedness Operator to $FO + poly$ ($linear$)

Chris Giannella*and Dirk Van Gucht

December 8, 2000

Abstract

In the constraint database community, $FO+poly$ and $FO+linear$ have been proposed as foundations for spatial database query languages. One of the strengths of this approach is that these languages are a clean and natural generalization of Codd's relational model to a spatial setting. As a result, rigorous mathematical study of their expressiveness and complexity can be carried out.

Along this line, important geometric queries involving connectivity have been shown to be inexpressible in $FO+poly$ and $FO+linear$. To address this problem, we extend both languages with a parameterized path-connectivity predicate, $Pconn$. We show that: $FO + linear + Pconn$ and $FO+poly+Pconn-3D$ are closed and have PTIME data complexity. We also examine the expressiveness of $FO+poly+Pconn$ and $FO+linear+Pconn$ and show that parity and transitive closure are expressible in each.

1 Introduction

Due to the rapidly increasing speeds of CPUs and sizes of memories, large amounts of spatial data (e.g. maps, pictures, etc.) can readily be accumulated and stored. For this data to be of use, it must be manageable and accessible in a fast and flexible manner. The development of software to provide such access is an important goal.

*Partially supported by US Department of Education grant P200A502367. Contact: cgiannel@cs.indiana.edu

Guting et al. have approached this goal by integrating well-chosen 2D abstract data types (ADTs), into relational databases ([16], [17], [15]). Their fundamental data structure is a *realm* and is described in detail in [16]. Based on this, a formal specification of spatial data types is described in detail in [17] and is called the *ROSE algebra.* This specification has been implemented and the details are described in [15]. To combat problems caused by the numerical representation of points and line segments in a realm, a new underlying data structure,*a dual grid*, has recently been developed [21]. An implementation of the ROSE system over this data structure is currently under way.

Although these ADTs perform well in their specific domain of application, they are difficult to generalize and implement in higher dimensions. Also, their mathematical properties remain fairly unexplored. Polynomial (and linear) spatial constraint database models were introduced (1) to generalize ADT approaches and (2) to provide richer mathematical foundations for ADTs. This approach was first proposed in [18] and has been investigated substantially since. It is a clean and natural generalization of Codd's relational model ([11]) to a spatial data setting. As a result, rigorous mathematical study of its fundamental properties can be carried out.

Spatial databases are represented as a collection of semi-algebraic (or semi-linear) relations (FO-definable sets over $\mathbb{R}$ with $<, +, *, 0, 1$ - just $<, +, 0, 1$ in the semi-linear case) with an associated relation name. Example: the relation with name INDIANA is the collection of points defining the state of Indiana and is finitely represented as a semi-linear set. Notice that the constituents of each tuple are real numbers and that the relation is finitely represented but not finite. Semi-algebraic sets are sufficiently rich to model all geometric objects that appear in spatial databases. Semi-linear sets are argued in [27] to be sufficiently rich to model most geometric objects that appear as well.

Strictly speaking a spatial database is represented as a first order structure with universe $\mathbb{R}$ and relation names interpreted by their associated finitely represented relations. Query languages for spatial data are modeled using first order logic over $\mathbb{R}$. The most prominent ones are $FO{+}poly$ and $FO{+}linear$. A query is specified in $FO{+}poly$ by FO formulae over a vocabulary consisting of a collection of relation names plus $<, *, +, 0, 1$. Each formula defines a mapping from spatial databases to finitely represented relations in the standard way. $FO + linear$ is $FO + poly$ without "+".

$FO + poly$ can express many natural spatial queries, e.g. computing the convex hull of a semi-algebraic set ([12]). However, two important types of queries cannot be expressed adequately: aggregation and connectivity. For an aggregation example, consider the query Q_{ct} which, given a point, p, on a map and a distance, ϵ, returns the count of the number of cities lying within ϵ from p. Q_{ct} is not expressible in $FO + poly$ (follows from the result in [14] that parity is not expressible). For a connectivity example, consider the query Q_{cn} which determines if two given points on a map can be connected by a series of roads (also not expressible in $FO + poly$ [14]). The underlying data structures, semi-algebraic or semi-linear sets, are rich enough to model geometric objects, but their associated query languages ($FO + poly$, $FO + linear$) are incommensurate to the task of querying the objects. Richer query languages are necessary.

$FO+poly$ has been extended to deal with aggregate queries. Benedikt and Libkin [7] augment $FO + poly$ with a finite, bag sum operator. Grumbach et al. [13] augment $FO + poly$ with finite bag operators, sum, product, and average obtaining $FO + poly + Agg$ (the authors denote the language $FO^{agg}_{\mathbb{R}}$). In $FO + poly + Agg$, it is possible to express query Q_{ct}.

1.1 Purpose and Summary of Results

The purpose of this paper is to address the problem that $FO + poly$ and $FO + linear$ cannot express important geometric queries like connectivity (e.g. Q_{cn}). To do so, we extend $FO + poly$ and $FO + linear$ with a parameterized path-connectivity predicate, $Pconn$. We denote the resulting languages $FO + poly + Pconn$ and $FO + linear + Pconn$.

Our main results are the following. (1) $FO + linear + Pconn$ and $FO + poly + Pconn - 3D$ are closed and have PTIME data complexity. $FO + poly + Pconn - mD$ is the resulting sub-language from restricting formulae inside $Pconn$ to have at most m free variables. (2) $FO + poly + Pconn$ and $FO + linear + Pconn$ can capture queries seemingly unrelated to connectivity, such as determining if a point is inside a polygon and determining the parity of a finite semi-algebraic set.

1.2 Related Work

Benedikt et al. [6], independently of us, prove $FO + poly + Pconn$ and $FO + linear + Pconn$ are closed. Their proof of closure of $FO +$

$linear + Pconn$ (Proposition 6 of [6]) is constructive, very similar to ours and yields a PTIME data complexity proof (both ours and their proofs do not require a change of coordinates). Their proof of closure of $FO + poly + Pconn$ does not yield a PTIME data complexity proof. Instead they define an extension of $FO + poly$, $L_{Path}(poly)$, which they call a "path logic". $L_{Path}(poly)$ subsumes $FO + poly + Pconn$ and they prove that it has PTIME data complexity. These results are described in greater detail in section 6.

Our closure results are not as general as those of Benedikt et al.. However, our proof techniques are simpler and interesting in their own right as a contrast to those of Benedikt et al.. In particular we establish closure of $FO + poly + Pconn - 3D$ using only a variant of the Collins cylindrical algebraic decomposition algorithm which does not require a change of coordinates ([4]). Benedikt et al. use semi-algebraic triviality (section 9.3 of [8]) in addition to the standard Collins cylindrical algebraic decomposition algorithm ([10], [3]) to obtain closure (semi-algebraic triviality requires a change of coordinates). Therefore, we show how closure (and PTIME data complexity) of $FO + poly + Pconn - 3D$ can be proven *without requiring a change of coordinates.*

Also, our proof techniques highlight more clearly the essential properties and limitations of the decomposition produced by the Collins algorithm in proving closure and PTIME data complexity. Specifically, there is insufficient knowledge of the geometry of the boundary of cells in the decomposition to allow our technique to go through in the general case (only in the 3D case). In section 5.4 we point out an open problem summarizing this lack of knowledge. A solution to this problem would not only allow our technique to go through, but would also provide a better understanding of the geometric properties of decompositions produced by the Collins algorithm. To the best of our knowledge, literature on the Collins algorithm does not throughly address the geometry of the boundary of cells. We feel that this represents a significant gap in the understanding of the Collins algorithm.

2 Preliminaries

Given a database schema, $\mathcal{S}$ (a finite set of relation symbols), a semi-algebraic (semi-linear) instance, $\mathcal{I}$, of $\mathcal{S}$ is a function with domain $\mathcal{S}$ such that $\forall P \in \mathcal{S}, \mathcal{I}(P)$ is a semi-algebraic (semi-linear) subset of $\mathbb{R}^{arity(P)}$. We now define the query language $FO + poly$ ($FO + linear$). The syntax of $FO + poly$ ($FO + linear$) is the same as that of FO+$(0, 1, +, *, <)$ (FO+$(0, 1, +, <)$) augmented with $\mathcal{S}$. The semantics is described as follows. Let ϕ be an $FO{+}poly$ ($FO{+}linear$) formula (with n free variables). The query defined by ϕ takes semi-algebraic (semi-linear) instances $\mathcal{I}$ of $\mathcal{S}$ to a set $\phi(\mathcal{I})$ which is defined in the standard way ([1]).

Example 1 Let Q_{dis} be a query over schema $\{P\}$, $arity(P) = 1$, where $Q_{dis}(\mathcal{I})$ is true if and only if $\mathcal{I}(P)$ is discrete. $Q_{dis} \in FO{+}linear$ as the following $FO + linear$ sentence defines it (taken from [27]): $\exists d \forall x[(d > 0) \wedge P(x) \rightarrow \neg\exists y(\neg(x = y) \wedge (x - d < y < x + d))]$.

An essential property of $FO + poly$ and $FO + linear$ (and most traditional query languages) is that the result of a query can be used as input to another query. This property relies on the fact that $FO{+}poly$ and $FO + linear$ are closed. Namely, queries in $FO + poly$ ($FO + linear$) return semi-algebraic (semi-linear) sets. The proof of closure is immediate from the definition of these languages. However, the proofs of closure of $FO + poly + Pconn$ and $FO + linear + Pconn$ are non-trivial.

Before going on to the next section and defining $FO{+}poly{+}Pconn$, some basic definitions are needed. Let $1 \leq i \leq n$ and $X \subseteq \mathbb{R}^i$; $Z_n(X) = X \times \mathbb{R}^{n-i}$ if $n > i$, otherwise, X. If $i = n - 1$, $Z_n(X)$ is called the *cylinder* above X (and is denoted $Z(X)$ in [3]). For $\alpha \in \mathbb{R}^i$, $Z_n(\{\alpha\})$ is shortened to $Z_n(\alpha)$. $Z_n(\alpha)$ is said to be a *fiber*. The next definition will be crucial in the what follows.

Definition 1 *Given* $1 \leq i \leq n$ *and* $X \subseteq \mathbb{R}^n$, $PC_i(X) = \{(\alpha, b_1, \dots, b_n, c_1, \dots, c_n) \in \mathbb{R}^{i+2n} \mid \overrightarrow{b}$ *and* $\overrightarrow{c}$ *are path connected in* $X \cap Z_n(\alpha)\}$. $PC_0(X) = \{(\overrightarrow{b}, \overrightarrow{c}) \in \mathbb{R}^{2n} \mid \overrightarrow{b}$ *and* $\overrightarrow{c}$ *are path connected in* $X\}$.

3 FO+poly+Pconn

The syntax of $FO{+}poly{+}Pconn$ is defined exactly as that of $FO{+}poly$ ($FO + linear$) with one additional type of complex formula. Let ϕ be an $FO + poly + Pconn$ ($FO + linear + Pconn$) formula with free variables $x_1, \dots, x_n$. Let $\overrightarrow{w}, \overrightarrow{z}$ be n-tuples of distinct variables not appearing in ϕ, and $0 \leq i \leq n$,

$$Pconn_{(x_1,\dots,x_i)}(\phi(x_1, \dots, x_n))(\overrightarrow{w}, \overrightarrow{z}) \tag{1}$$

is a formula of $FO + poly + Pconn$. Moreover, it has free variables $x_1, \ldots, x_i$ and the variables in $\overrightarrow{w}$ and $\overrightarrow{z}$. An example is $Pconn_{(x_1)}(x_1 \leq 1 \wedge x_2 \leq 2)(w, z)$ which has free variables x_1, w, z. Given $m \geq 1$, $FO + poly + Pconn - mD$ denotes the sub-language of $FO + poly + Pconn$ where complex formula (1) is restricted so that ϕ has no more than m free variables.

The semantics is defined just as for $FO+poly$ ($FO+linear$) with a single addition. Given ϕ an $FO+poly+Pconn$ ($FO+linear+Pconn$) formula with free variables $x_1, \ldots, x_n$, and $\mathcal{I}$ a semi-algebraic (semi-linear) instance over $\mathcal{S}$, define $Pconn_{(x_1,\ldots,x_i)}(\phi(x_1, \ldots, x_n))(\overrightarrow{w}, \overrightarrow{z})(\mathcal{I})$ as $PC_i(\phi(\mathcal{I}))$.

Example 2 Let Q_{pconn} be a query over schema $\{P\}$, $arity(P) = 2$, where $Q_{pconn}(\mathcal{I})$ is true if and only if $\mathcal{I}(P)$ is path connected. Consider the $FO + poly + Pconn$ formula $\psi := \forall \overrightarrow{y} \forall \overrightarrow{z} [(P(\overrightarrow{y}) \wedge P(\overrightarrow{z})) \rightarrow Pconn_{()}(P(x_1, x_2))(\overrightarrow{y}, \overrightarrow{z})]$. ψ has no free variables and $\psi(\mathcal{I})$ is true if and only if $\forall \overrightarrow{a}, \overrightarrow{b} \in \mathcal{I}(P)$, $\overrightarrow{a}$, $\overrightarrow{b}$ are path connected in $\mathcal{I}(P)$.

4 Definitions and Properties for the Closure Theorem

In order for $FO+poly+Pconn$ ($FO+linear+Pconn$) to be closed, it must return semi-algebraic (semi-linear) sets on semi-algebraic (semi-linear) instances. By an induction on formulae, it suffices to show:

Claim 1 *Given semi-algebraic (semi-linear) $S \subseteq \mathbb{R}^n$ and $0 \leq i \leq n$, $PC_i(S)$ is semi-algebraic (semi-linear).*

We are not able to show this claim for any semi-algebraic S. However, we are able to show it in two special cases: S is semi-linear and $n \leq 3$. As a result, we prove $FO + linear + Pconn, FO + poly + Pconn - 3D$ are closed. The proofs depend heavily on the concept of a *cylindrical algebraic decomposition*. Many of the definitions are taken from [3]. Henceforth, $\mathcal{P}$ denotes a finite defining set of polynomials for S.

4.1 Definition of a Cylindrical Decomposition Sequence

Given $1 \le i \le n, x \in \mathbb{R}^i$, and $X \subseteq \mathbb{R}^i$ let $\Pi(x)$ denote the natural projection which removes the last coordinate of x and let $\Pi(X)$ denote $\{\Pi(x) | x \in X\}$. Let $\overline{X}$ denote the topological closure of X. Given $Y \subseteq \mathbb{R}^i$, X and Y are said to be *adjacent* if $X \cap \overline{Y} \neq \emptyset$ or $Y \cap \overline{X} \neq \emptyset$. X is said to be Y-*invariant* if $X \subseteq Y$ or $X \cap Y = \emptyset$. A collection of sets is Y-invariant if each set in the collection is. Given $x, y \in \mathbb{R}^i$ and $\epsilon > 0$, let $Cube(x, \epsilon)$ denote the closed, filled-in cube in $\mathbb{R}^i$ centered at x with side length ϵ. Let $D(x, y)$ denote the standard Euclidean distance between x and y.

A non-empty, connected subset of $\mathbb{R}^i$ is called a *region*. Given region, $R \subseteq \mathbb{R}^{n-1}$, and continuous $f : R \rightarrow \mathbb{R}$, an *f-section* of $Z_n(R)$ is $graph(f)$ (i.e. $\{(a, f(a)) \mid a \in R\}$). Given continuous $f, g : R \rightarrow \mathbb{R}$ with $f < g$ (i.e. for all $a \in R, f(a) < g(a)$), an (f, g)-*sector* of $Z_n(R)$ is a set of points (x, b) where $x \in R$ and $f(x) < b < g(x)$. Note that "constant" functions $f = -\infty$ and $g = \infty$ are permitted.

A *decomposition* of X is a *finite* collection of disjoint regions whose union is X. Given continuous $f_1 < \ldots < f_k : R \rightarrow \mathbb{R}$ a natural decomposition of the cylinder over R consists of regions (1) (f_i, f_{i+1})-sectors for $0 \le i \le k$ where $f_0 = -\infty$ and $f_{k+1} = \infty$; (2) f_i-sections for $1 \le i \le k$. Such a decomposition is called a *stack* over R (determined by $f_1, \ldots, f_k$).

A decomposition $\mathcal{C}^n$ of $\mathbb{R}^n$ is said to be *cylindrical* if (1) $n = 1$ and $\mathcal{C}^1$ contains a finite number of singletons or (2) $n > 1$ and there exists a cylindrical decomposition, $\mathcal{C}^{n-1}$ of $\mathbb{R}^{n-1}$ where for each $C_{n-1} \in \mathcal{C}^{n-1}$, some subset of $\mathcal{C}^n$ forms a stack over C_{n-1}. $\mathcal{C}^{n-1}$ is unique for $\mathcal{C}^n$. Therefore, $\mathcal{C}^n$ induces unique cylindrical decompositions $\mathcal{C}^{n-1}, \ldots, \mathcal{C}^2, \mathcal{C}^1$ of $\mathbb{R}^{n-1}, \mathbb{R}^{n-2}, \ldots, \mathbb{R}$, respectively. $\mathcal{C}^n, \ldots, \mathcal{C}^1$ is said to be a *cylindrical decomposition sequence*.

$\mathcal{C}^n$ is said to be a *stratification* if for any $C \in \mathcal{C}^n$, $\overline{C}$ is the union of regions in $\mathcal{C}^n$. We shall see that this is a very desirable but not always obtainable property.

4.2 Properties of Cylindrical Decomposition Sequences

A fundamental property which provides a nice framework for induction proofs on n is: for all $1 \le i < n, C \in \mathcal{C}^n$, $\Pi^{n-i}(C) \in \mathcal{C}^i$ where Π^{n-i}

denotes Π applied $n-i$ times.

The next three theorems describe some useful geometric properties of cylindrical decomposition sequences. The first, Theorem 1, describes some properties of the interaction between a fiber originating in some region in a lower decomposition and a region in the top decomposition. The second, Theorem 2, describes how sections are smooth inside their stacks. The third, Theorem 3, describes how regions are smooth with respect to fibers inside their stacks.

Theorem 1 *For all* $1 \leq i < n, C \in \mathcal{C}^n$, *and* $\alpha \in \Pi^{n-i}(C)$, $Z_n(\alpha) \cap C$ *is non-empty and path connected.*

Proof: Assume $i = n-1$. If C is an f_e-section, then, by definition, it follows that $Z_n(\alpha) \cap C = \{(\alpha, f_e(\alpha))\}$ which is non-empty and path connected. If C is an (f_e, f_{e+1})-sector, then $Z_n(\alpha) \cap C = \{(\alpha, a) | f_e(\alpha) < a < f_{e+1}(\alpha)\}$ (clearly path connected). Since $f_e < f_{e+1}$, then $f_e(\alpha) < f_{e+1}(\alpha)$, so, $Z_n(\alpha) \cap C$ is also non-empty. Assume henceforth that $i < n-1$. The rest of the proof proceeds by induction on n. The base case, $n = 2$, is vacuous, so, assume $n > 2$. By induction, $Z_{n-1}(\alpha) \cap \Pi(C)$ is non-empty and path connected. It can easily be seen that $Z_n(\alpha) \cap C$ must, therefore, be non-empty. Showing that $Z_n(\alpha) \cap C$ is path connected, however, is more difficult.

There exists continuous functions $f_e < f_{e+1} : \Pi(C) \to \mathbb{R}$ such that C is an f_e-section or an (f_e, f_{e+1})-sector. Recall that $f_e = -\infty$ or $f_{e+1} = \infty$ are allowed. If $f_e \neq -\infty$, define $F_e : \Pi(C) \to \mathbb{R}^n$ as $\hat{x} \in \Pi(C) \mapsto (\hat{x}, f_e(\hat{x}))$. If $f_{e+1} \neq \infty$, define F_{e+1} in an analogous fashion. Clearly F_e, F_{e+1} are continuous.

Case C is an f_e-section ($f_e \neq -\infty$). $F_e(Z_{n-1}(\alpha) \cap \Pi(C))$ is path connected. Moreover, it can be checked that $F_e(Z_{n-1}(\alpha) \cap \Pi(C)) = Z_n(\alpha) \cap C$.

Case C is an (f_e, f_{e+1})-sector. Then $C = \{(\hat{x}, a) \in \mathbb{R}^n \mid \hat{x} \in \Pi(C)$ and $f_e(\hat{x}) < a < f_{e+1}(\hat{x})\}$. Let $x, y \in Z_n(\alpha) \cap C$. There exists, P, a path in $Z_{n-1}(\alpha) \cap \Pi(C)$ between $\Pi(x)$ and $\Pi(y)$. Assume, for the moment, that $f_e \neq -\infty$ and $f_{e+1} \neq \infty$. $F_e(P)$ and $F_{e+1}(P)$ are both compact, so, they can be separated by a distance $\delta > 0$ (i.e. for any $z_1 \in F_e(P), z_2 \in F_{e+1}(P)$, $D(z_1, z_2) > \delta$).

Let $Q = \{(\hat{x}, \delta/2 + f_e(\hat{x})) | \hat{x} \in P\}$. Clearly, $Q \subseteq Z_n(P) \cap C \subseteq Z_n(\alpha) \cap C$ and is path connected. Moreover, $\forall \hat{x} \in P$, $Z_n(\hat{x}) \cap C \subseteq Z_n(\alpha) \cap C$ is path connected and intersects Q. Hence, $\hat{Q} = (Q \cup (Z_n(\Pi(x)) \cap C) \cup (Z_n(\Pi(y)) \cap C)) \subseteq Z_n(\alpha) \cap C$ is path connected, and contains x, y. There exists a path in $\hat{Q}$ between x and y.

If $f_e = -\infty$ or $f_{e+1} = \infty$ the argument above can be modified slightly to get the desired result.

$\square$

Theorem 2 *For any section $C \in \mathcal{C}^n, C = Z_n(\Pi(C)) \cap \overline{C}$.*

Proof: Clearly $C \subseteq Z_n(\Pi(C)) \cap \overline{C}$. Let $x \in Z_n(\Pi(C)) \cap \overline{C}$. Since C is a section, then there exists $f_e : \Pi(C) \to \mathbb{R}$ continuous such that C is an f_e-section ($f_e \neq -\infty$). Consider F_e from the "C is an $f_e - section$" case in the proof of Theorem 1.

Let $y = F_e(\Pi(x))$ (y is in C) and (x_m) be a sequence in C which converges on x. $(\Pi(x_m))$ is a sequence in $\Pi(C)$ which converges on $\Pi(x)(= \Pi(y))$. Since F_e is continuous at $\Pi(y)$, then $(F_e(\Pi(x_m)))$ converges to y. Since C is an f_e-section, then for all $m, F_e(\Pi(x_m)) = x_m$. Thus, (x_m) converges to y, so, $x = y$. We conclude that $x \in C$, so, $Z_n(\Pi(C)) \cap \overline{C} \subseteq C$, as desired.

$\square$

Theorem 3 *For all $C \in \mathcal{C}^n$ and all $\alpha \in \Pi(C)$, $Z_n(\alpha) \cap \overline{C} \subseteq \overline{Z_n(\alpha) \cap C}$.*

Proof: Let $C \in \mathcal{C}^n$ and $\alpha \in \Pi(C)$.

Assume C is a section. $Z_n(\alpha) \cap \overline{C} = Z_n(\alpha) \cap Z_n(\Pi(C)) \cap \overline{C} = Z_n(\alpha) \cap C \subseteq \overline{Z_n(\alpha) \cap C}$ (the last equality is due to Theorem 2).

Assume C is a sector (an (f_e, f_{e+1})-sector). Let $x \in Z_n(\alpha) \cap \overline{C}$. Then $x = (\alpha, a_x)$ where $a_x \in \mathbb{R}$. Suppose $a_x > f_{e+1}(\alpha)$. Then, by Theorem 2, there exists $\epsilon > 0$ such that $Cube(x, \epsilon) \cap Z_n(\Pi(C)) \cap graph(f_{e+1}) = \emptyset$. It follows that $Cube(x, \epsilon)$ is entirely above $graph(f_{e+1})$ (i.e. for all $\hat{y} \in Cube(\Pi(x), \epsilon) \cap \Pi(C), f_{e+1}(\hat{y}) < a_x - \epsilon$). Hence $Cube(x, \epsilon)$ is entirely above C, therefore, $x \notin \overline{C}$. This is a contradiction, so, we conclude that $a_x \leq f_{e+1}(\alpha)$. A similar argument shows that $a_x \geq f_e(\alpha)$. But $f_e < f_{e+1}$, so, both are not equality. Without loss of generality, assume, $f_e(\alpha) \leq a_x < f_{e+1}(\alpha)$.

Let $\epsilon > 0$. There exists $f_e(\alpha) < b < f_{e+1}(\alpha)$ such that $(\alpha, b) \in Cube(x, \epsilon)$. Therefore, $Cube(x, \epsilon) \cap Z_n(\alpha) \cap C \neq \emptyset$, so, $x \in \overline{Z_n(\alpha) \cap C}$. We conclude that $Z_n(\alpha) \cap \overline{C} \subseteq \overline{Z_n(\alpha) \cap C}$ as desired.

$\square$

Theorem 3 shows for $n = 2$ an additional "smoothness" property which will be important later: $\mathcal{C}^2$ has 1-well-behaved fibers.

4.3 Definition of $\mathcal{P}$-invariant CAD Sequences

Given $p \in \mathcal{P}$, let $V(p) = \{\overrightarrow{a} \in \mathbb{R}^n | p(\overrightarrow{a}) = 0\}$ (i.e. the *variety* of p). Let $V_<(p) = \{\overrightarrow{a} \in \mathbb{R}^n | p(\overrightarrow{a}) < 0\}$ and $V_>(p) = \{\overrightarrow{a} \in \mathbb{R}^n | p(\overrightarrow{a}) > 0\}$. Given $X \subseteq \mathbb{R}^n$, X is $\mathcal{P}$-*invariant* if $\forall p \in \mathcal{P}, X \subseteq V(p)$ or $X \subseteq V_<(p)$ or $X \subseteq V_>(p)$. A collection of subsets is $\mathcal{P}$-invariant if each set in the collection is. A given decomposition $\mathcal{D}$ is said to be *algebraic* if each set in $\mathcal{D}$ is semi-algebraic. A *CAD sequence* of $\mathbb{R}^n, \ldots, \mathbb{R}^1$ is a cylindrical, algebraic, decomposition sequence $\mathcal{C}^n, \ldots, \mathcal{C}^1$ such that each $\mathcal{C}^i$ is algebraic. If $\mathcal{C}^n$ is also $\mathcal{P}$-invariant, then $\mathcal{C}^n, \ldots, \mathcal{C}^1$ is a $\mathcal{P}$-invariant CAD sequence.

5 Main Proofs

This section gives proofs of Claim 1 for the two special cases of S semi-linear and $n \leq 3$. In addition, data complexity proofs are given.

Section 5.1 describes the framework for the proofs of the special cases. This framework relies on the existence of a $\mathcal{P}$-invariant CAD sequence which has two additional nice properties. It is shown that if such a nice CAD sequence can be found, then Claim 1 holds.

Section 5.2 describes an algorithm using a nice CAD sequence for constructing $PC_i(S)$. The conclusion is that if a nice CAD sequence can be constructed from $\mathcal{P}$ in PTIME, then $PC_i(S)$ can be constructed from $\mathcal{P}$ in PTIME.

Sections 5.3, 5.4 describe two PTIME algorithms for constructing $\mathcal{P}$-invariant CAD sequences from $\mathcal{P}$. The first is the standard Collins cylindrical algebraic decomposition algorithm and the second is a slight variant for 3D.

Section 5.5 gives a proof that, for linear $\mathcal{P}$, the $\mathcal{P}$-invariant CAD sequence produced by the Collins algorithm has the two additional nice properties. Section 5.6 gives a proof that for $n \leq 3$, the $\mathcal{P}$-invariant CAD sequence produced by the variant of the Collins algorithm for 3D has the two additional nice properties.

Finally section 5.7 puts together all of the special cases into the main closure and data complexity theorem (Theorem 12).

5.1 Framework

Claim 1 for $i = 0, n$ is straightforward to prove and is treated separately. We shall deal with the Claim first for $1 \leq i < n$, then in

Section 5.7, deal with the Claim for $i = 0, n$.

The main idea for proving Claim 1 for $1 \leq i < n$ is to construct, for a given $1 \leq i < n$, a $\mathcal{P}$-invariant CAD sequence such that $\mathcal{C}^n$ has two additional properties which make it well-behaved.

The first additional property is the following. $\mathcal{C}^n$ is said to have *$i-$nice adjacencies* ($1 \leq i < n$) if for any $C, \hat{C} \in \mathcal{C}^n$ such that $\Pi^{n-i}(C) = \Pi^{n-i}(\hat{C})$, C and $\hat{C}$ are adjacent implies that for all $\alpha \in \Pi^{n-i}(C), Z_n(\alpha) \cap \overline{C} \cap \hat{C} \neq \emptyset$.

Take note that this property does not necessarily imply that $\mathcal{C}^n$ is a stratification. But if $\mathcal{C}^n$ is a stratification, then $Z_n(\alpha) \cap \overline{C} \cap \hat{C} = Z_n(\alpha) \cap \hat{C} \neq \emptyset$ (the inequality is due to Theorem 1). So, $\mathcal{C}^n$ has i-nice adjacencies. Moreover, $\mathcal{C}^n$ has $(n-1)$-nice adjacencies (but is not necessarily a stratification). $\Pi^{n-(n-1)}(C) = \Pi^{n-(n-1)}(\hat{C})$ implies that $C, \hat{C}$ are in the same stack. If $\overline{C} \cap \hat{C} \neq \emptyset$ then by definition of stacks, it can be seen that $\hat{C} \subseteq \overline{C}$, so, $Z_n(\alpha) \cap \overline{C} \cap \hat{C} = Z_n(\alpha) \cap \hat{C} \neq \emptyset$.

The second additional property is the following. $\mathcal{C}^n$ is said to have *i-well-behaved fibers* ($1 \leq i < n$) if for any $C \in \mathcal{C}^n$, and any $\alpha \in \Pi^{n-i}(C), Z_n(\alpha) \cap \overline{C} \subseteq \overline{Z_n(\alpha) \cap C}$. By Theorem 3 $\mathcal{C}^n$ has $(n-1)$-well-behaved fibers.

Using these two additional properties, a crucial theorem can be proved allowing the computation of connectivity along fibers to be made "globally" along regions.

Theorem 4 *If $\mathcal{C}^n$ has i-nice adjacencies and i-well-behaved fibers, then for any $C, \hat{C} \in \mathcal{C}^n$ with $\Pi^{n-i}(C) = \Pi^{n-i}(\hat{C})$, it is the case that: C, $\hat{C}$ are adjacent if and only if for any $\alpha \in \Pi^{n-i}(C), (C \cup \hat{C}) \cap Z_n(\alpha)$ is path connected.*

Proof: Assume $\mathcal{C}^n$ has i-nice adjacencies and i-well-behaved fibers.

($\Rightarrow$:) Assume $C, \hat{C}$ are adjacent (say, $\overline{C} \cap \hat{C} \neq \emptyset$) and let $\alpha \in \Pi^{n-i}(C)$. Since $\mathcal{C}^n$ has i-nice adjacencies, then $Z_n(\alpha) \cap \overline{C} \cap \hat{C} \neq \emptyset$. Moreover, since it also has i-well-behaved fibers, then $\overline{Z_n(\alpha) \cap C} \cap (Z_n(\alpha) \cap \hat{C}) \neq \emptyset$. By Theorem 1, $Z_n(\alpha) \cap C$ and $Z_n(\alpha) \cap \hat{C}$ are both non-empty and path-connected. Therefore, $(C \cup \hat{C}) \cap Z_n(\alpha)$ is connected. Since $C, \hat{C}$ are semi-algebraic, then by [8] Theorem 2.4.5, $(C \cup \hat{C}) \cap Z_n(\alpha)$ is path connected.

($\Leftarrow$:) Follows easily without using the two additional properties.

□

Based on Theorem 4, $PC_i(S)$ can be constructed as follows. Given $\alpha \in \mathbb{R}^i$, let $PC_{i,\alpha} = \{(\alpha, \overrightarrow{b}, \overrightarrow{c}) \in \mathbb{R}^{i+2n} \mid \overrightarrow{b}$ and $\overrightarrow{c}$ are path connected in $S \cap Z_n(\alpha)\}$. Since $\mathcal{C}^i$ is a decomposition of $\mathbb{R}^i$, then, by definition 1, $PC_i(S) = \bigcup_{C_j \in \mathcal{C}^i} \bigcup_{\alpha \in C_j} PC_{i,\alpha}(S)$. $\mathcal{C}^i$ is finite and each of its regions are semi-algebraic (semi-linear if S is semi-linear). If for each $C_j \in \mathcal{C}^i, \bigcup_{\alpha \in C_j} PC_{i,\alpha}(S)$ is semi-algebraic (semi-linear), then $PC_i(S)$ is semi-algebraic (semi-linear). Hence Claim 1 can be proven by showing that: for each $C_j \in \mathcal{C}^i, \bigcup_{\alpha \in C_j} PC_{i,\alpha}(S)$ is semi-algebraic (semi-linear if S is semi-linear).

Given $C_j \in \mathcal{C}^i$, let $\mathcal{D}_{C_j} = \{C \in \mathcal{C}^n | C \subseteq S, \Pi^{n-i}(C) = C_j\}$. $\mathcal{D}_{C_j}$ is a decomposition of $S \cap Z_n(C_j)$ (because $\mathcal{C}^n$ is $\mathcal{P}$-invariant and for all $C \in \mathcal{C}^n$, $\Pi^{n-i}(C) = C_j$ or $\Pi^{n-i}(C) \cap C_j = \emptyset$). Define an equivalence relation, $\equiv_{C_j}$, on $\mathcal{D}_{c_j}$ as $D \equiv_{C_j} \hat{D}$ if D is related to $\hat{D}$ in the transitive closure of the "adjacent" relation on $\mathcal{D}_{C_j}$. Let $\mathcal{E}_{C_j,1}, \ldots, \mathcal{E}_{C_j,m_j}$ denote the equivalence classes of $\equiv_{C_j}$. Let $\mathbb{E}_{C_j,\ell} = \bigcup \mathcal{E}_{C_j,\ell}$ for all $1 \leq \ell \leq m_j$.

Theorem 5 *If $\mathcal{C}^n$ has i-nice adjacencies and i-well-behaved fibers, then for all $C_j \in \mathcal{C}^i$ and $\alpha \in C_j$, it is the case that $PC_{i,\alpha}(S) = \{\alpha\} \times \bigcup_{\ell=1}^{m_j}((Z_n(\alpha) \cap \mathbb{E}_{C_j,\ell}) \times (Z_n(\alpha) \cap \mathbb{E}_{C_j,\ell}))$.*

Proof: Assume $\mathcal{C}^n$ has i-nice adjacencies and i-well-behaved fibers. Let $C_j \in \mathcal{C}^i, \alpha \in C_j$.

($\subseteq$:) Let $(\alpha, x, y) \in PC_{i,\alpha}(S)$. Then x and y are path connected in $Z_n(\alpha) \cap S$. There exists continuous $f : [0,1] \to Z_n(\alpha) \cap S$ such that $f(0) = x$ and $f(1) = y$. Since $\mathcal{D}_{C_j}$ is a decomposition of $S \cap Z_n(\alpha)$, then $S \cap Z_n(\alpha) \subseteq \bigcup \mathcal{D}_{C_j}$, so, $f([0,1]) \subseteq \bigcup \mathcal{D}_{C_j}$.

Let D_x be the region in $\mathcal{D}_{C_j}$ containing x. Let $\mathcal{D}_0 = \{D_x\}$ and for all $0 \leq k, \mathcal{D}_{k+1} = \mathcal{D}_k \cup \{D \in \mathcal{D}_{C_j} | D, (\bigcup \mathcal{D}_k)$ are adjacent $\}$. Since $\mathcal{D}_{C_j}$ is finite and $\mathcal{D}_0, \mathcal{D}_1, \mathcal{D}_2, \ldots$ is a monotonically increasing sequence, then there exists $q \geq 0$ such that for all $\hat{q} \geq q, \mathcal{D}_q = \mathcal{D}_{\hat{q}}$. Two facts follow immediately:

1. there exists $1 \leq \ell \leq m_j$ such that $\mathcal{D}_q \subseteq \mathcal{E}_{C_j,\ell}$;
2. if $\mathcal{D}_{C_j} - \mathcal{D}_q \neq \emptyset$, then $(\bigcup \mathcal{D}_q), (\bigcup(\mathcal{D}_{C_j} - \mathcal{D}_q))$ are not adjacent.

We assert that $f([0,1]) \subseteq (\bigcup \mathcal{D}_q)$. By (1) it will follow that $(\alpha, x, y) \in \{\alpha\} \times (Z_n(\alpha) \cap \mathbb{E}_{C_j,\ell}) \times (Z_n(\alpha) \cap \mathbb{E}_{C_j,\ell})$ as needed. Suppose the assertion is false, then there exists $t \in [0,1]$ such that $f(t) \notin \bigcup \mathcal{D}_q$. Hence $\mathcal{D}_{C_j} - \mathcal{D}_q \neq \emptyset$, so, by (2) $(\bigcup \mathcal{D}_q), (\bigcup(\mathcal{D}_{C_j} - \mathcal{D}_q))$ are not adjacent. Since f is continuous and $[0,1]$ connected, then $f([0,1]) \subseteq \bigcup \mathcal{D}_q$ or $f([0,1]) \subseteq \bigcup(\mathcal{D}_{C_j} - \mathcal{D}_q)$. Since $f(0) = x \in D_x \subseteq \bigcup \mathcal{D}_q$, then $f([0,1]) \subseteq \bigcup \mathcal{D}_q$. Therefore, $f(t) \in \bigcup \mathcal{D}_q$, which is a contradiction. We conclude that the assertion is true.

($\supseteq$:) Let $(\alpha, x, y) \in \{\alpha\} \times \bigcup_{\ell=1}^{m_j}((Z_n(\alpha) \cap \mathbb{E}_{C_j,\ell}) \times (Z_n(\alpha) \cap \mathbb{E}_{C_j,\ell}))$. Then there exists $1 \leq \ell \leq m_j$ and $D_x, D_y \in \mathcal{E}_{C_j,\ell}$ such that $x \in$

$Z_n(\alpha) \cap D_x$ and $y \in Z_n(\alpha) \cap D_y$. If D_x, D_y are adjacent then by Theorem 4, it follows that x, y are path connected in $Z_n(\alpha) \cap (D_x \cup D_y)$. Since $D_x, D_y \subseteq S$, then $(\alpha, x, y) \in PC_{i,\alpha}(S)$. If D_x, D_y are not adjacent, then there exists $D_1, \dots, D_k \in \mathcal{E}_{C_j,\ell}$ for $k \geq 1$ such that: D_x, D_1 are adjacent, D_1, D_2 are adjacent, $\cdots$, D_{k-1}, D_k are adjacent, and D_k, D_y are adjacent. By Theorem 4 it follows that x and y are path connected in $Z_n(\alpha) \cap (D_x \cup D_1 \cup D_2 \cup \cdots \cup D_k \cup D_y)$. Since $D_x, D_1, \dots, D_k, D_y \subseteq S$, then x, y are path connected in $Z_n(\alpha) \cap S$. Thus $(\alpha, x, y) \in PC_{i,\alpha}(S)$.

□

Under appropriate assumptions, Claim 1 can now be proven (for $1 \leq i < n$).

Theorem 6 *If $\mathcal{C}^n$ has i-nice adjacencies and i-well-behaved fibers, then $PC_i(S)$ is semi-algebraic (semi-linear if S is semi-linear).*

Proof: Assume $\mathcal{C}^n$ has i-nice adjacencies and i-well-behaved fibers. By Theorem 5 it suffices to show that: for each $C_j \in \mathcal{C}^i, \bigcup_{\alpha \in C_j}(\{\alpha\} \times \bigcup_{\ell=1}^{m_j}((Z_n(\alpha) \cap \mathbb{E}_{C_j,\ell}) \times (Z_n(\alpha) \cap \mathbb{E}_{C_j,\ell})))$ is semi-algebraic (semi-linear if S is semi-linear). Let $C_j \in \mathcal{C}^i$. $\mathcal{C}^n$ is finite and each of its regions are semi-algebraic (semi-linear if S is semi-linear). Thus, for all $1 \leq \ell \leq m_j, \mathbb{E}_{C_j,\ell}$ is semi-algebraic (semi-linear). Let ϕ_ℓ be an FO+poly (FO+linear) formula which defines $\mathbb{E}_{C_j,\ell}$ and ϕ_{C_j} which defines C_j.

Let ϕ be the following FO+poly (FO+linear) formula:

$$\phi_{C_j}(\overrightarrow{\alpha}) \wedge (x_1, \dots, x_i) = \overrightarrow{\alpha} = (y_1, \dots, y_i) \wedge (\bigvee_{\ell=1}^{m_j} (\phi_\ell(\overrightarrow{x}) \wedge \phi_\ell(\overrightarrow{y}))).$$

Given $\alpha \in \mathbb{R}^i, x, y \in \mathbb{R}^n$, (α, x, y) satisfies ϕ if and only if $\alpha \in C_j$, $x, y \in Z_n(\alpha)$, and $(x, y) \in \bigcup_{\ell=1}^{m_j}(\mathbb{E}_{C_j,\ell} \times \mathbb{E}_{C_j,\ell})$ if and only if $\alpha \in C_j$ and $(x, y) \in \bigcup_{\ell=1}^{m_j}((\mathbb{E}_{C_j,\ell} \cap Z_n(\alpha)) \times (\mathbb{E}_{C_j,\ell} \cap Z_n(\alpha)))$. Therefore ϕ defines $\bigcup_{\alpha \in C_j}(\{\alpha\} \times \bigcup_{\ell=1}^{m_j}((Z_n(\alpha) \cap \mathbb{E}_{C_j,\ell}) \times (Z_n(\alpha) \cap \mathbb{E}_{C_j,\ell})))$

□

5.2 Data Complexity

The problem of constructing $PC_i(S)$ from $\mathcal{P}$ in PTIME (for $1 \leq i < n$) reduces to showing that a $\mathcal{P}$-invariant CAD sequence with i-nice adjacencies and i-well-behaved fibers can be constructed from $\mathcal{P}$ in PTIME.

Theorem 7 *If an $\mathcal{P}$-invariant CAD sequence, $\mathcal{C}^n$, with i-nice adjacencies and i-well-behaved fibers can be constructed from $\mathcal{P}$ in PTIME, then so can $PC_i(S)$.*

Proof: Assume the antecedent. Since $PC_i(S) = \bigcup_{C_j \in \mathcal{C}^i} \bigcup_{\alpha \in C_j} PC_{i,\alpha}(S)$, then it suffices to show that for any $C_j \in \mathcal{C}^i$, $\bigcup_{\alpha \in C_j} PC_{i,\alpha}(S)$ can be constructed in PTIME. The equivalence classes of $\equiv_{C_j}$ over $\mathcal{D}_{C_j}$ ($\mathcal{E}_{C_j,1}, \ldots, \mathcal{E}_{C_j,m_j}$) can be constructed in PTIME since $\mathcal{C}^n$ can. Hence, the formula ϕ from the proof of Theorem 6 can be constructed.

□

5.3 Construction of $\mathcal{P}$-invariant CAD Sequences

The first PTIME (n fixed) algorithm for constructing a $\mathcal{P}$-invariant CAD sequence was obtained by Collins in 1975 ([10]). We briefly outline the exposition of the algorithm given in [3] stating a few of its properties (without proof) which we will use explicitly. The algorithm proceeds in three stages: projection, base, and extension (these terms are from [3]).

In the projection phase, a series of finite sets of polynomials is produced from $\mathcal{P}$: $\mathcal{P}^n, \mathcal{P}^{n-1}, \ldots, \mathcal{P}^1$ where $\mathcal{P}^n = \mathcal{P}$. For all $1 \leq j < n$, $\mathcal{P}^j \subseteq \mathbb{Q}[x_1, \ldots, x_j]$ and all polynomials in $\mathcal{P}^j$ are linear if all in $\mathcal{P}^{j+1}$ are linear. In the base phase, a $\mathcal{P}^1$-invariant, algebraic decomposition of $\mathbb{R}$ is produced, $\mathcal{C}^1$. $\mathcal{C}^1$ consists of a finite set of points and open intervals (all of which are all semi-linear if all polynomials in $\mathcal{P}^1$ are linear). In the extension phase, a series of algebraic decompositions of $\mathbb{R}^2, \ldots, \mathbb{R}^n$, respectively, are produced from $\mathcal{C}^1, \mathcal{C}^2, \ldots, \mathcal{C}^n$. For all $1 < j \leq n$, $\mathcal{C}^j$ is $\mathcal{P}^j$-invariant and each of its regions are semi-linear if all polynomials in $\mathcal{P}^j$ are linear. Take note that $\mathcal{C}^n$ is S-invariant since $\mathcal{P}^n$ is a defining set of polynomials for S.

Given a polynomial $q \in \mathbb{Q}[x_1, \ldots, x_m]$ ($m > 1$) and $X \subseteq \mathbb{R}^{m-1}$, X is said to *nullify* q if for all $x \in X, q(x,.)$ is the zero polynomial. For each $1 \leq j < n$ and each $C_j \in \mathcal{C}^j$, let $\mathcal{P}^{j+1}_{C_j}$ denote $\{p \in \mathcal{P}_{j+1} | C_j$ does not nullify $p\}$. If $\mathcal{P}^{j+1}_{C_j} = \emptyset$, then the stack above C_j consists of one sector, $Z_{j+1}(C_j)$.

If $\mathcal{P}^{j+1}_{C_j}$ is non-empty then let $P^{j+1}_{C_j}$ be the product of all polynomials in $\mathcal{P}^{j+1}_{C_j}$. There exists $k_{C_j} \geq 0$ such that for all $x \in C_j, P^{j+1}_{C_j}(x,.)$ has k_{C_j} roots (denoted $r_{1,x} < r_{2,x} < \cdots < r_{k_{C_j},x}$). Moreover, for each $1 \leq \ell \leq k_{C_j}$, the map $f_\ell : x \in C_j \mapsto r_{\ell,x}$ is continuous and

$f_1 < f_2 < \cdots < f_{k_{C_j}}$ (see [10]). If $k_{C_j} = 0$, then the stack above C_j consists of one sector, $Z_{j+1}(C_j)$. Otherwise, the graphs of $f_1, \ldots, f_{k_{C_j}}$ form the sections in the stack above C_j (the spaces in-between form the sectors). The sections in the stack above $\Pi(C)$ are the connected components of $V(P^{j+1}_{C_j}) \cap Z_n(\Pi(C))$.

Fact 1 *Let $C \in \mathcal{C}^n$ be a section. There exists $p \in \mathcal{P}$ such that $\Pi(C)$ does not nullify p. Moreover, the following also holds.*

1. *C is a connected component of $V(p) \cap Z_n(\Pi(C))$;*
2. *if $\mathcal{P}$ consists of linear polynomials, then $V(p) \cap Z_n(\Pi(C)) = C$.*

Proof: Since C is a section, then by the discussion immediately preceeding the statement of Fact 1, there exists $k_{\Pi(C)} > 0$ and $1 \leq \ell \leq k_{\Pi(C)}$ such that C is the graph of the map $\hat{x} \in \Pi(C) \mapsto r_{\ell,\hat{x}}$. Therefore, there exists $p \in \mathcal{P}^n_{\Pi(C)}$ and $\hat{x} \in \Pi(C)$ where $(\hat{x}, r_{\ell,\hat{x}}) \in V(p) \cap C$. Since $p \in \mathcal{P}^n_{\Pi(C)}$, then $\Pi(C)$ does not nullify p.

1. Since $V(p) \cap C \neq \emptyset$, C is $\mathcal{P}$-invariant and p divides $P^n_{\Pi(C)}$, then $C \subseteq V(p) \cap Z_n(\Pi(C)) \subseteq V(P^{j+1}_{C_j}) \cap Z_n(\Pi(C))$. Also, by the discussion immediately preceeding statement of Fact 1, C is a connected component of $V(P^{j+1}_{C_j}) \cap Z_n(\Pi(C))$. We conclude that C is a connected component of $V(p) \cap Z_n(\Pi(C))$.

2. Assume p is a linear polynomial. Let $x \in V(p) \cap Z_n(\Pi(C))$. By Theorem 1, $Z_n(\Pi(x)) \cap C \neq \emptyset$. Moreover since p is linear, then $p(\Pi(x), .)$ has at most one root, so, $|Z_n(\Pi(x)) \cap V(p)| = 1$. By 1. $C \subseteq V(p) \cap Z_n(\Pi(C))$, therefore, $x \in C$. We conclude that $C = V(p) \cap Z_n(\Pi(C))$.

□

5.4 A Variant of the Collins Algorithm

As seen in Theorems 4, 5, 6, 7, i-well-behaved fibers and i-nice adjacencies are crucial properties of $\mathcal{C}^n$ in proving closure and PTIME data complexity. In the linear case (e.g. the polynomials in $\mathcal{P}$ are linear), we shall see in Section 5.5 that these properties hold. Moreover, they can also be shown to hold in the non-linear case where $i = n - 1$. However, to the best of our knowledge, the question of whether these properties hold for the Collins algorithm with $n \geq 3$ and $1 \leq i < n - 1$ is open.

Open Problem 1 *Let $\mathcal{P} \subseteq \mathbb{Z}[x_1, \ldots, x_n]$ with $n \geq 3$, and $\mathcal{C}^n, \ldots, \mathcal{C}^1$ be the $\mathcal{P}$-invariant CAD sequence produced by the Collins algorithm and $1 \leq i < n-1$. Does $\mathcal{C}^n$ have i-nice adjacencies or i-well-behaved fibers?*

However, a small variant of the Collins algorithm is described in [4] for $\mathcal{P} \subseteq \mathbb{Z}[x_1, x_2, x_3]$ (not necessarily linear) on which these properties hold (shown in Section 5.6). This algorithm produces a $\mathcal{P}$-invariant CAD sequence $\mathcal{C}^3, \mathcal{C}^2, \mathcal{C}^1$, in PTIME, with some additional nice properties: (1) $\mathcal{C}^3$ is a stratification; (2) for all $C_2 \in \mathcal{C}^2$, if there exists $p \in \mathcal{P}$ nullified by C_2 then C_2 is a singleton. Henceforth we refer to this variant as *the 3D Collins algorithm.*

The 3D Collins algorithm allows us to get around this open problem in proving Claim 1 for $n = 3$. But, for $n > 3$ we cannot get around the open problem.

5.5 Claim 1 for S Semi-linear

Assume S is semi-linear ($\mathcal{P}$ consists of linear polynomials). Let $\mathcal{C}^n, \ldots,$ $\mathcal{C}^1$ be a $\mathcal{P}$-invariant CAD sequence produced by the Collins algorithm ($n \geq 1$). The regions of $\mathcal{C}^n$ have a very nice geometric structure. We develop this structure here.

The next lemma (Lemma 1) provided a syntactic characterization of the sections in the stack above a region in $\mathcal{C}^{n-1}$. Each polynomial in $\mathcal{P}$ is of the form $a_n x_n + \cdots + a_1 x_1 + a_0$ where $a_0, \ldots, a_n$ are integers. Let $\mathcal{P}_{a_n \neq 0}$ be the polynomials in $\mathcal{P}$ such that $a_n \neq 0$.

Lemma 1 *For all $C_{n-1} \in \mathcal{C}^{n-1}$, if $p \in \mathcal{P}_{a_n \neq 0}$, then $V(p) \cap Z_n(C_{n-1})$ is a section in the stack above C_{n-1}. Moreover, if $C \in \mathcal{C}^n$ is a section in the stack above C_{n-1}, then there exists $p \in \mathcal{P}_{a_n \neq 0}$ such that $V(p) \cap$ $Z_n(C_{n-1})$.*

Proof: Let $C_{n-1} \in \mathcal{C}^{n-1}$. Let $p \in \mathcal{P}_{a_n \neq 0}$. Since $a_n \neq 0$, then C_{n-1} does not nullify p. Moreover, given $\hat{x} \in C_{n-1}$, there exists $r_{\hat{x}}$ a unique root of $p(\hat{x}, .)$. Recall from the discussion in Section 5.3 that $P^n_{C_{n-1}}$ is the product of the set of polynomials from $\mathcal{P}$ which are not nullified by C_{n-1}. p is included in this product. So, $r_{\hat{x}}$ is a root of $P^n_{C_{n-1}}(\hat{x}, .)$. It follows that $(\hat{x}, r_{\hat{x}})$ is in a section, C, in the stack above C_{n-1}, so, $V(p) \cap C \neq \emptyset$. Since $\mathcal{C}^n$ is $\mathcal{P}^n$-invariant, then $C \subseteq V(p) \cap Z_n(C_{n-1})$.

Let $x \in V(p) \cap Z_n(C_{n-1})$. From the discussion in the previous paragraph, it follows that $\{x\} = Z_n(\Pi(x)) \cap V(p)$. Moreover by Theorem 1 $Z_n(\Pi(x)) \cap C \neq \emptyset$, thus, $x \in C$. We conclude that $C = V(p) \cap Z_n(C_{n-1})$.

On the other hand, let $C \in \mathcal{C}^n$ be a section in the stack above C_{n-1}. By Fact 1, there exists $p \in \mathcal{P}$ such that C_{n-1} does not nullify p and $V(p) \cap Z_n(C_{n-1}) = C$. $a_n \neq 0$ or else C_{n-1} nullifies p, therefore, $p \in \mathcal{P}_{a_n \neq 0}$. $\square$

Lemma 2 *For any $C \in \mathcal{C}^n$, C is convex and $C = ri(C)$ where $ri(C)$ denotes the relative interior of C ([24]).*

Proof: By induction on n. The base case of $n = 1$ is immediate since C is an open interval in $\mathbb{R}$. Assume $n > 1$ and $\Pi(C)$ is convex and equals $ri(\Pi(C))$. Since $\Pi(C)$ is convex, then $Z_n(\Pi(C))$ is too.

Let $C \in \mathcal{C}^n$. First we show that C is convex. Assume C is a section, then by Lemma 1, $C = V(p) \cap Z_n(\Pi(C))$ for some $p \in \mathcal{P}_{a_n \neq 0}$. Since $V(p)$ and $Z_n(\Pi(C))$ are both convex, then their intersection is too.

Assume C is a sector. If C has no sections immediately above or below in its stack then $C = Z_n(\Pi(C))$. If C has a section immediately above but none below, then $C = V_<(p) \cap Z_n(\Pi(C))$ for some $p \in \mathcal{P}_{a_n \neq 0}$. If C has a section immediately below but not above then $C = V_>(p) \cap Z_n(\Pi(C))$ for some $p \in \mathcal{P}_{a_n \neq 0}$. If C has a section immediately above and a section immediately below then $C = V_>(p) \cap V_<(p') \cap Z_n(\Pi(C))$ for some $p, p' \in \mathcal{P}_{a_n \neq 0}$. Since $V_>(p)$, $V_<(p')$, $V_<(p)$, and $Z_n(\Pi(C))$ are all convex, then in all cases C is convex.

Now we show that $ri(C) = C$. By definition $ri(C) \subseteq C$. $\Pi(C) = \Pi(ri(C))$ because $\Pi(.)$ is a linear transformation, C is convex, and $\Pi(C) = ri(\Pi(C))$. Let $w \in C$. $\Pi(w) \in \Pi(ri(C))$, so, there exists $x \in ri(C)$ such that $\Pi(x) = \Pi(w)$. If C is a section then $w = x$, so, $w \in ri(C)$. Assume C is a sector. There exists $y \in C$ such that $y \neq x$ and the vertical line segment between x and y contains w. Theorem 6.1 pg. 45 of [24] implies that $w \in ri(C)$. We conclude that $C \subseteq ri(C)$ as desired.

$\square$

The previous two lemmas show that the regions in a stack have a nice geometric structure. Now we show that the regions from one stack "interface" nicely with the regions from an adjacent stack. Specifically, we show that $\mathcal{C}^n$ is a stratification.

Lemma 3 *$\mathcal{C}^n$ is a stratification.*

Proof: By induction on n. In the base case of $n = 1$, $\mathcal{C}^n$ consists of a finite number of points and open intervals - clearly a stratification. Consider the induction case of $n > 1$ and let $C \in \mathcal{C}^n$. If suffices to show that for all $\hat{C} \in \mathcal{C}^n, \overline{C} \cap \hat{C} = \emptyset$ or $\hat{C} \subseteq \overline{C}$.

Let $\hat{C} \in \mathcal{C}^n$ such that $\overline{C} \cap \hat{C} \neq \emptyset$. By induction $\Pi(\hat{C}) \subseteq \overline{\Pi(C)}$.

Case C is a section. By Lemma 1, there exists $p \in \mathcal{P}_{a_n \neq 0}$ such that $C = V(p) \cap Z_n(\Pi(C))$. Since $\overline{C} \cap \hat{C} \neq \emptyset$, then $\hat{C} \cap V(p) \neq \emptyset$ because $V(p)$ is closed. Further since $\mathcal{C}^n$ is $\mathcal{P}$-invariant, then $\hat{C} \subseteq V(p)$.

Let $x \in \hat{C}$, then $\Pi(x) \in \Pi(\hat{C}) \subseteq \overline{\Pi(C)}$. So, there exists $(\hat{x}_m)$ a sequence in $\Pi(C)$ which converges on $\Pi(x)$. p is of the form $a_n x_n + \hat{p}$ where $\hat{p} = a_{n-1}x_{n-1} + \cdots + a_1 x_1 a_0$ and $a_n \neq 0$. Let $a_m = \frac{-\hat{p}(\hat{x}_m)}{a_n}$ and $x_m = (\hat{x}_m, a_m)$. Since $(\hat{x}_m)$ converges, then (a_m) does too; let a be the real number to which (a_m) converges. Clearly (x_m) is a sequence in $V(p) \cap Z_n(\Pi(C))$ which converges to $(\Pi(x), a)$, hence, $p(\Pi(x), a) = 0$. Since p is a linear polynomial and $a_n \neq 0$, then a is the unique root of $p(\Pi(x), .)$. It follows that $x = (\Pi(x), a)$, so, (x_m) converges to x. We conclude that $x \in \overline{V(p) \cap Z_n(\Pi(C))} = \overline{C}$, so, $\hat{C} \subseteq \overline{C}$ as desired.

Case C is a sector. Assume C has sections, C_b, C_a in its stack immediately below and above, respectively. By Lemma 1, there exists $p_b, p_a \in \mathcal{P}_{a_n \neq 0}$ such that $C_b = V(p_b) \cap Z_n(\Pi(C))$ and $C_a = V(p_a) \cap Z_n(\Pi(C))$. It follows that $C = V_<(p_a) \cap V_>(p_b) \cap Z_n(\Pi(C))$.

Assume $\hat{C}$ is a section. Since $\overline{C} \cap \hat{C} \neq \emptyset$, then by Lemma 1, it follows that $\hat{C} = V(p_a) \cap Z_n(\Pi(\hat{C}))$ or $V(p_b) \cap Z_n(\Pi(\hat{C}))$. Hence, $\hat{C} \subseteq \overline{C_a}$ or $\overline{C_b}$, respectively. Clearly, $C_a, C_b \subseteq \overline{C}$, therefore, $\overline{C_a}, \overline{C_b} \subseteq \overline{C}$. We conclude that $\hat{C} \subseteq \overline{C}$ as desired.

Assume $\hat{C}$ is a sector, then $\hat{C} \cap V(p_a), \hat{C} \cap V(p_b) = \emptyset$. There exists $y \in \hat{C} \cap \overline{C}$, so, $y \in V_<(p_a) \cap V_>(p_b)$. Because $\hat{C}$ is $\mathcal{P}$-invariant, then $\hat{C} \subseteq V_<(p_a) \cap V_>(p_b)$. Let $x \in \hat{C}$. There exists $\epsilon > 0$ such that $Cube(x, \epsilon) \subseteq V_<(p_a) \cap V_>(p_b)$. $\Pi(x) \in \Pi(\hat{C}) \subseteq \overline{\Pi(C)}$, so, there exists $y \in Cube(x, \epsilon)$ such that $\Pi(y) \in \Pi(C)$. It follows that $y \in Cube(x, \epsilon) \cap V_<(p_a) \cap V_>(p_b) \cap Z_n(\Pi(C)) = Cube(x, \epsilon) \cap C$. We conclude that $x \in \overline{C}$, so, $\hat{C} \subseteq \overline{C}$ as desired.

The three remaining sub-cases are: C only has a section immediately above, C only has a section immediately below, C does not have a section immediately above or below. Analogous arguments can be given in these sub-cases.

$\square$

Now we prove the main result.

Theorem 8 *For $1 \leq i < n$, $PC_i(S)$ is semi-linear and can be constructed from $\mathcal{P}$ in PTIME.*

Proof By Theorems 6 and 7 it suffices to show that $\mathcal{C}^n$ has i-nice adjacencies and i-well-behaved fibers. Since $\mathcal{C}^n$ is a stratification by Lemma 3, then $\mathcal{C}^n$ has i-nice adjacencies (see comments immediately after the definition of i-nice adjacencies in Section 5.1). Let $C \in \mathcal{C}^n$. By Lemma 2 and Theorem 1, $Z_n(\alpha) \cap ri(C) \neq \emptyset$. Corollary 6.5.1 pg. 48 in [24] (with $M = Z_n(\alpha)$) implies $(Z_n(\alpha) \cap \overline{C}) \subseteq \overline{Z_n(\alpha) \cap C}$. Hence $\mathcal{C}^n$ has i-well-behaved fibers.

□

5.6 Claim 1 for $n \leq 3$

Assume $S \subseteq \mathbb{R}^3$ and is semi-algebraic (with a finite set of defining polynomials $\mathcal{P}$). Let $\mathcal{C}^3, \mathcal{C}^2, \mathcal{C}^1$ be the $\mathcal{P}$-invariant CAD sequence produced by the 3D Collins algorithm. We shall show that $\mathcal{C}^3$ has i-nice adjacencies and i-well-behaved fibers for $i = 1, 2$. First we prove two technical lemmas. These lemmas lead to Theorem 9 which state that "ill-behaved fibers" cause some polynomial in $\mathcal{P}$ to be nullified. As a result, property (2) of the 3D Collins algorithm described in Section 5.4 is applied to rule out the existence of any such fibers.

The next lemma shows that regions in $\mathcal{C}^2$ have arbitrarily small neighborhoods about any point in $\mathbb{R}^2$ whose intersection with the region is connected.

Lemma 4 *For all $\hat{C} \in \mathcal{C}^2$, all $\delta > 0$, and all $\hat{x} \in \mathbb{R}^2$ there exists $\delta/2 > \epsilon_1 > 0$ such that for all $\epsilon_1 > \epsilon > 0$, $Cube(\hat{x}, \epsilon) \cap \hat{C}$ is connected.*

Proof: If $\hat{C}$ is an f_e-section, then the result follows from the Monotonicity Theorem pg 43 of [25]. Assume $\hat{C}$ is an (f_e, f_{e+1})-sector. Let $\sigma_e = inf\{D(\hat{x}, \hat{z}) | \hat{z} \in graph(f_e)\}$, $\sigma_{e+1} = inf\{D(\hat{x}, \hat{z}) | \hat{z} \in graph(f_{e+1})\}$. There are several cases to consider: $\sigma_e = 0, \sigma_{e+1} = 0$; $\sigma_e = 0, \sigma_{e+1} > 0$; $\sigma_e > 0, \sigma_{e+1} = 0$; $\sigma_e > 0, \sigma_{e+1} > 0$. In all of these cases the Monotonicity Theorem is used to show the desired result. Details are omitted.

□

Let $C \in \mathcal{C}^3$ be an f_e-section.

Lemma 5 *For all $y = (\hat{x}, b) \in \mathbb{R}^3 - \overline{C}$ where $\hat{x} \in \mathbb{R}^2$ and all $\delta > 0$, there exists $\delta/2 > \epsilon > 0$ such that: $\forall \hat{z} \in \Pi(C) \cap Cube(\hat{x}, \epsilon), f_e(\hat{z}) < b - \epsilon/2$ or $\forall \hat{z} \in \Pi(C) \cap Cube(\hat{x}, \epsilon), f_e(\hat{z}) > b + \epsilon/2$.*

Proof: Let $\delta > 0$. Let ϵ_1 be as in Lemma 4 (with $\hat{C} = \Pi(C)$). Since $x \notin \overline{C}$, then there exists $\epsilon_1 > \epsilon > 0$ such that $Cube(y,\epsilon) \cap C = \emptyset$. So, for any $\hat{z} \in \Pi(C) \cap Cube(\hat{x},\epsilon)$, $f_e(\hat{z}) < b - \epsilon/2$ or $b + \epsilon/2 < f_e(\hat{z})$ (or else $(\hat{z}, f_e(\hat{z})) \in Cube(y,\epsilon) \cap C$ which is impossible).

By Lemma 4, it follows that, $f_e(\Pi(C) \cap Cube(\hat{x},\epsilon))$ is connected. The result follows.

□

Theorem 9 *For any $x \in \overline{C} - \overline{Z_3(\Pi^2(x)) \cap C}$, there exists $p \in \mathcal{P}$ where $p(\Pi(x), .)$ is the zero polynomial.*

Proof: Let $x = (\hat{x}, a) \in \overline{C} - \overline{Z_3(\Pi^2(x)) \cap C}$ where $\hat{x} \in \mathbb{R}^2$. Since $x \notin \overline{Z_3(\Pi^2(x)) \cap C}$, then, there exists $\delta > 0$ such that $Cube(x,\delta) \cap Z_3(\Pi^2(x)) \cap C = \emptyset$. Hence for all $\hat{z} \in Cube(\hat{x},\delta) \cap Z_2(\Pi^2(x)) \cap \Pi(C), f_e(\hat{z}) < a - \delta/2$ or $f_e(\hat{z}) > a + \delta/2$ (or else $(\hat{z}, f_e(\hat{z})) \in Cube(x,\delta) \cap Z_3(\Pi^2(x)) \cap C$ which is impossible). $\Pi(C) \cap Z_2(\Pi^2(x))$ is an interval thus $f_e(Cube(\hat{x},\delta) \cap Z_2(\Pi^2(x)) \cap \Pi(C)$ is connected. Hence for all $\hat{z} \in Cube(\hat{x},\delta) \cap Z_2(\Pi^2(x)) \cap \Pi(C), f_e(\hat{z}) < a - \delta/2$ or for all $\hat{z} \in Cube(\hat{x},\delta) \cap Z_2(\Pi^2(x)) \cap \Pi(C), f_e(\hat{z}) > a + \delta/2$. Without loss of generality assume that:

$$\text{for all } \hat{z} \in Cube(\hat{x},\delta) \cap Z_2(\Pi^2(x)) \cap \Pi(C), f_e(\hat{z}) < a - \delta/2. \qquad (2)$$

Let $a - \delta/2 < b < a$. Suppose $y = (\hat{x}, b) \notin \overline{C}$. By Lemma 5, there exists $\delta/2 > \epsilon > 0$ such that for all $\hat{z} \in \Pi(C) \cap Cube(\hat{x},\epsilon), f_e(\hat{z}) < b - \epsilon/2$ or for all $\hat{z} \in \Pi(C) \cap Cube(\hat{x},\epsilon), f_e(\hat{z}) > b + \epsilon/2$.

$x \in \overline{C}$ implies that there exists a sequence $((\hat{x_m}, a_m))$ in C which converges to x where $\hat{x_m} \in \Pi(C)$. Then $(\hat{x_m})$ is a sequence in $\Pi(C)$ which converges to $\hat{x}$ and (a_m) is a sequence in $\mathbb{R}$ which converges to a. Since C is a section, then for all m, $f_e(\hat{x_m}) = a_m$; $(f_e(\hat{a_m}))$ converges to a. Therefore, there exists $\hat{x_m} \in \Pi(C) \cap Cube(\hat{x},\delta)$ such that $f_e(\hat{x_m}) \geq b - \epsilon/2$ (since $a > b$). We have that:

$$\text{for all } \hat{z} \in \Pi(C) \cap Cube(\hat{x},\epsilon), f_e(\hat{z}) > b + \epsilon/2. \qquad (3)$$

$x = (\hat{x}, b) \in Z_3(\Pi^2(x)) \cap \overline{C}$, so, $\hat{x} \in Z_2(\Pi^2(x)) \cap \overline{\Pi(C)} \subseteq \overline{Z_2(\Pi^2(x)) \cap \Pi(C)}$ by Theorem 3. There exists $\hat{z} \in Cube(\hat{x},\epsilon) \cap Z_2(\Pi^2(x)) \cap \Pi(C)$. So, by inequality (3), $f_e(\hat{z}) > b + \epsilon/2$. However, by inequality (2), $f_e(\hat{z}) < a - \delta/2$. Therefore $b + \epsilon/2 < f_e(\hat{z}) < a - \delta/2$, so, $b + \delta/2 < a$ which is a contradiction.

We conclude that for all $a - \delta < b < a, (\hat{x}, b) \in \overline{C}$. Since C is a section then by Fact 1 part 1., there exists $p \in \mathcal{P}$ such that

$C \subseteq V(p)$. Hence $\overline{C} \subseteq V(p)$ because $V(p)$ is closed. So, for all $a - \delta < b < a, p(\hat{x}, b) = 0$. $p(\hat{x}, .)$ is a univariate polynomial with an infinite number of roots. Thus, $p(\hat{x}, .) = p(\Pi(x), .)$ is the zero polynomial.

□

Now we prove the main result.

Theorem 10 *For $1 \leq i < 3, PC_i(S)$ is semi-algebraic and can be produced from $\mathcal{P}$ in PTIME.*

Proof: By Theorems 6 and 7, it suffices to show that $\mathcal{C}^3$ has i-nice adjacencies and i-well-behaved fibers. But $\mathcal{C}^3$ is a stratification by property (1) of the 3D Collins algorithm described in Section 5.4. So, it has i-nice adjacencies (see the comment immediately following the definition of i-nice-adjacencies in Section 5.1). Moreover, by Theorem 3, $\mathcal{C}^3$ has 2-well-behaved fibers (see the comment immediately following the definition of i-well-behaved fibers). All that remains to be proven is that $\mathcal{C}^3$ has 1-well-behaved fibers.

Let $C \in \mathcal{C}^3$ and $\alpha \in \Pi^2(C)$.

Case C is a section. Suppose there exists $x \in (Z_3(\alpha) \cap \overline{C} - \overline{Z_3(\alpha) \cap C})$. By Theorem 9, there exists $p \in \mathcal{P}$ such that $p(\Pi(x), .)$ is the zero polynomial. Let C_2 be the region in $\mathcal{C}^2$ which contains $\Pi(x)$. Hence C_2 nullifies p, so, C_2 must be a singleton by property (2) of the 3D Collins algorithm discussed in section 5.4. Since $x \in Z_3(\alpha)$ and $\alpha \in \Pi^2(C)$ then $\Pi(C_2) = \Pi^2(C)$. Therefore $\Pi^2(C)$ is a singleton, so, $C \subseteq Z_3(\alpha)$. $Z_3(\alpha)$ is closed, so, $\overline{C} \subseteq Z_3(\alpha)$. Thus $Z_3(\alpha) \cap \overline{C} = \overline{C}$. Hence, $x \in \overline{Z_3(\alpha) \cap C}$ which is a contradiction. We conclude that $Z_3(\alpha) \cap \overline{C} \subseteq \overline{Z_3(\alpha \cap C)}$.

Case C is an (f_e, f_{e+1})-sector. Recall that $f_e = -\infty$ or $f_{e+1} = \infty$ are allowed. Let $x \in Z_3(\alpha) \cap \overline{C}$. If $x \in \overline{graph(f_e)}$, then the argument from the "C is a section case" applied to $graph(f_e)$ shows that $x \in \overline{Z_3(\alpha) \cap graph(f_e)}$. It can be seen, then that $x \in \overline{Z_3(\alpha) \cap C}$ as desired. If $x \in \overline{graph(f_{e+1})}$, then an analogous argument shows that $x \in \overline{Z_3(\alpha) \cap C}$.

Assume henceforth that $x \notin \overline{graph(f_e)}, \overline{graph(f_{e+1})}$. Then there exists $\epsilon > 0$ such that $Cube(x, \epsilon) \cap graph(f_e) = \emptyset$ and $Cube(x, \epsilon) \cap graph(f_{e+1}) = \emptyset$.

Let $x = (\hat{x}, a)$ where $\hat{x} \in \mathbb{R}^2$. We say that $Cube(x, \epsilon) > f_{e+1}$ if for all $\hat{z} \in Cube(\hat{x}, \epsilon) \cap \Pi(C), f_{e+1}(\hat{z}) < a - \epsilon/2$ (similarly for f_e). Likewise, $Cube(x, \epsilon) < f_{e+1}$ if for all $\hat{z} \in Cube(\hat{x}, \epsilon) \cap \Pi(C), f_{e+1}(\hat{z}) > a + \epsilon/2$

(similarly for f_e). Clearly, $Cube(x,\epsilon) > f_{e+1}$ or $Cube(x,\epsilon) < f_{e+1}$ (similarly for f_e).

Since $x \in \overline{C}$, then $\hat{x} \in \overline{\Pi(C)}$, so, there exists $\hat{z} \in Cube(\hat{x},\epsilon) \cap \Pi(C)$. Hence there are four cases to consider: $Cube(x,\epsilon) < f_e$, $Cube(x,\epsilon) > f_{e+1}$; $Cube(x,\epsilon) > f_e$, $Cube(x,\epsilon) > f_{e+1}$; $Cube(x,\epsilon) < f_e$, $Cube(x,\epsilon) < f_{e+1}$; $Cube(x,\epsilon) > f_e$, $Cube(x,\epsilon) < f_{e+1}$.

Sub-case $Cube(x,\epsilon) < f_e$**,** $Cube(x,\epsilon) > f_{e+1}$**.** This sub-case is not possible since $f_e < f_{e+1}$.

Sub-case $Cube(x,\epsilon) > f_e$**,** $Cube(x,\epsilon) > f_{e+1}$**.** Since $Cube(x,\epsilon) > f_{e+1}$, then it can be seen that $Cube(x,\epsilon) \cap C = \emptyset$. But $x \in \overline{C}$, so, this sub-case is also impossible.

Sub-case $Cube(x,\epsilon) < f_e$**,** $Cube(x,\epsilon) < f_{e+1}$**.** Analogous to the previous sub-case.

Sub-case $Cube(x,\epsilon) > f_e$**,** $Cube(x,\epsilon) < f_{e+1}$**.** Since $x \in Z_3(\alpha) \cap \overline{C}$, then $\hat{x} \in Z_2(\alpha) \cap \overline{\Pi(C)} \subseteq \overline{Z_2(\alpha) \cap \Pi(C)}$ ($\subseteq$ is due to Theorem 3). Hence there exists $(\hat{z}_m)$ a sequence in $Cube(\hat{x},\epsilon) \cap Z_2(\alpha) \cap \Pi(C)$ which converges to $\hat{x}$. Since $f_e < Cube(x,\epsilon) < f_{e+1}$, then for each m, $f_e(\hat{z}_m) < a < f_{e+1}(\hat{z}_m)$. Hence $((\hat{z}_m, a))$ is a sequence in $Z_3(\alpha) \cap C$ which converges to $(\hat{x}, a) = x$. We conclude that $x \in \overline{Z_3(\alpha) \cap C}$, as desired.

□

5.7 The Main Closure and Data Complexity Theorem

First we deal with the cases of $i = 0, n$.

Theorem 11 *For any semi-algebraic (semi-linear) $S \subseteq \mathbb{R}^n$ with finite defining set of polynomials $\mathcal{P}$, $PC_0(S)$ and $PC_n(S)$ are semi-algebraic (semi-linear) and can be produced from $\mathcal{P}$ in PTIME.*

Proof: First consider $PC_0(S)$. Let $\mathcal{C}^n, \ldots, \mathcal{C}^1$ be the $\mathcal{P}$-invariant CAD sequence produced by the Collins algorithm on $\mathcal{P}$. It can be seen that there exists $m \geq 1$ such that S has m path connected components, $PCC_1, \ldots, PCC_m$ each of which is a union of regions from $\mathcal{C}^n$. By definition, $PC_0(S) = \bigcup_{j=1}^{m} PCC_j^2$, so, the desired result holds for $PC_0(S)$.

Now consider $PC_n(S)$. By definition $PC_n(S) = \{(\overrightarrow{a}, \overrightarrow{a}, \overrightarrow{a}) \in S^{3n}\}$ which is semi-algebraic (semi-linear), since S is. Moreover, $\{(\overrightarrow{a}, \overrightarrow{a}, \overrightarrow{a}) \in S^{3n}\}$ can clearly be computed in PTIME from $\mathcal{P}$.

$\square$

Putting together all of the special cases we arrive at the following theorem.

Theorem 12 *$FO + poly + Pconn - 3D$ and $FO + linear + Pconn$ are closed and have PTIME data complexity (n fixed).*

Proof: $FO + poly$, by itself, has data complexity NC [18]. The Collins algorithm and the 3D Collins algorithm both work in PTIME. Thus, a straightforward induction on formulae using Theorems 11, 8, and Theorem 10 will prove the desired results. $\square$

6 Related Work Revisited

As stated in section 1.2, Benedikt, et al. [6], independently of us, prove $FO+poly+Pconn$ and $FO+linear+Pconn$ are closed. In fact, they prove that $FO+poly+\mathcal{T}$ and $FO+linear+\mathcal{T}$ are closed where $\mathcal{T}$ is the set of definable topological maps (see section 3.3 of [6]) which includes $Pconn$. Their proof of closure of $FO + linear + Pconn$ (Proposition 6 of [6]) is constructive, very similar to ours and yields a PTIME data complexity proof (both ours and their proofs do not require a change of coordinates). They use the Collins algorithm and prove that query evaluation can be carried out "region-wise". This proof involves showing that regions have nice enough geometric properties to allow topological maps to be preserved across adjacencies ([22]). This technique is very similar to ours as illustrated in Section 5.5.

Their proof of closure of $FO+poly+Pconn$ does not yield a PTIME data complexity proof. They use semi-algebraic triviality (section 9.3 of [8]) along with the Collins algorithm. Semi-algebraic triviality requires a change of coordinates (the Collins algorithm and 3D Collins algorithm do not). This differs from our proof for $FO+poly+Pconn-3D$ which only uses the 3D Collins algorithm. Therefore our proof of closure (and PTIME data complexity) of $FO + poly + Pconn - 3D$ does not require a change of coordinates. To obtain PTIME data complexity, Benedikt et al. define a an extension of $FO + poly$, $L_{Path}(poly)$, which they call a "path logic". $L_{Path}(poly)$ subsumes $FO + poly + Pconn$ they prove that it has PTIME data complexity. This proof makes use of semi-algebraic triviality as well as the Collins algorithm. They also define $L_{Path}(linear)$ extending $FO+linear$ and providing an alternate PTIME data complexity proof of $FO+linear$.

7 Expressiveness

Since $FO + poly + Pconn$ and $FO + linear_P conn$ are closed, studying their expressiveness is a sensible endeavor. Some natural questions arise. (1) Is $FO+poly+Pconn$ ($FO+linear+Pconn$) more expressive than $FO + poly$ ($FO + linear$)? (2) Results are known which relate $FO + linear$ and $FO + poly$ [26]. What kinds of similar results hold for $FO+linear+Pconn, FO+poly+Pconn^{lin}$? $FO+poly+Pconn^{lin}$ are the linear queries of $FO+poly+Pconn$ (map semi-linear instances to semi-linear sets). Likewise $FO + poly^{lin}$ are the linear queries of $FO + poly$. (3) To what extent is the extension of $FO + poly$ with $Pconn$ orthogonal to that with aggregate operations?

7.1 Transitive Closure of an Undirected Graph

To answer question (1), consider the query Q_{trans} which on a finite, semi-algebraic $S \subseteq \mathbb{R}^2_{>0}$ ($\mathbb{R}_{>0}$ denotes the positive real numbers), returns its reflexive, symmetric transitive closure, $TC(S)$, defined in the following sense. Let $Adom(S)$ be the active domain of S (i.e. the set of all $a \in \mathbb{R}_{>0}$ such that $(a,b) \in S$ or $(b,a) \in S$ for some $b \in \mathbb{R}_{>0}$). Let $Ref(S) = S \cup \{(a,a) | a \in Adom(S)\}$. Let $\mathcal{G}_S$ denote the undirected graph with vertices $Adom(S)$ and edges $Ref(S)$. Define $TC(S)$ to be the set of all $(a,b) \in Adom(S)$ such that there exists an edge between a and b in the transitive closure of $\mathcal{G}_S$. Q^{lin}_{trans} is Q_{trans} taking only semi-linear input S (thus returning only semi-linear output).

Grumbach and Su [14] show that $Q_{trans} \notin FO+poly$ and $Q^{lin}_{trans} \notin FO + linear$. Using a similar reduction technique, we prove a result (Theorem 13) resolving question (1) in the positive.

Theorem 13 *$Q_{trans} \in FO + poly + Pconn$ and $Q^{lin}_{trans} \in FO + linear + Pconn$.*

Proof: First we show $Q_{trans} \in FO + poly + Pconn$. Let S be a finite semi-algebraic subset of $\mathbb{R}^2_{>0}$. We shall show how a set, $T_S \subseteq \mathbb{R}^3$, can be defined using S in $FO + poly$ such that for all $a, b \in Adom(S), (a,b) \in TC(S)$ if and only if $(a,0,0)$ and $(b,0,0)$ are path connected in T_S. It will follow that $Q_{trans} \in FO+poly+Pconn$. T_S is defined as follows. For each $(a,b) \in S$ construct the "connector line", $C_{(a,b)}$, as the set of all $(c_1, c_2, c_3) \in \mathbb{R}^3$, such that at least one of the following holds:

- $c_1 = a$, $0 \leq c_2 \leq a$, and $c_3 = 0$;

- $c_1 = a$, $c_2 = a$, and $0 \leq c_3 \leq 1$;
- $a \leq c_1 \leq b$, $c_2 = a$, and $c_3 = 1$;
- $c_1 = b$, $c_2 = a$, and $0 \leq c_3 \leq 1$;
- $c_1 = b$, $0 \leq c_2 \leq a$, and $c_3 = 0$.

T_S is the union of all connector lines. Consider a simple example where $S = \{(1,3),(3,5),(2,4)\}$. T_S depicted in figure 1. Note that $(1,5) \in TC(S)$ and $(1,0,0),(5,0,0)$ are path connected in T_S.

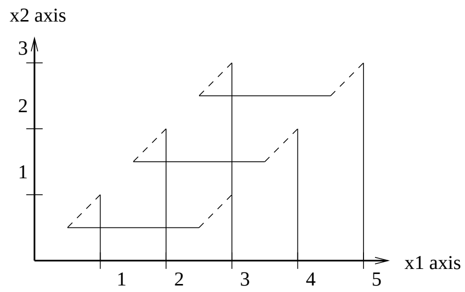

Figure 1: The dashed lines indicate the x3 axis direction.

Clearly T_S can be defined from S in $FO+poly$, so, all that remains to be shown: for all $a, b \in Adom(S)$, $(a,b) \in TC(S)$ if and only if $(a,0,0)$ and $(b,0,0)$ are path connected in T_S.

($\Rightarrow$:) Assume $(a,b) \in TC(S)$. Then $(a,b) \in Ref(S)$ or there exists (a,e_1), $(e_1,e_2),\ldots,(e_{k-1},e_k),(e_k,b) \in Ref(S)$ for $k \geq 1$. We assert that for any $c,d \in Adom(S)$ where $(c,d) \in Ref(S)$ it follows that: $(c,0,0)$ and $(d,0,0)$ are path connected in T_S. It will follow that $(a,0,0)$ and $(b,0,0)$ are path connected in T_S.

To see this assertion observe that $(c,d) \in S$ or $c = d$. In the first case there exists $C_{(c,d)}$ between $(c,0,0)$ and $(d,0,0)$. So, $(c,0,0)$ and $(d,0,0)$ are path connected in T_S. In the second case, since $c \in Adom(S)$, then there exists $e \in Adom(S)$ with $(c,e) \in S$ or $(e,c) \in S$. Hence $(c,0,0)$ is in $C_{(c,e)}$ or in $C_{(e,c)}$ and thus in T_S. Since $c = d$, then $(c,0,0)$ and $(d,0,0)$ are path connected in T_S.

($\Leftarrow$:) Assume $(a,0,0)$ and $(b,0,0)$ are path connected in T_S. Then there exists $(d_1,e_1),\ldots,(d_m,e_m) \in S$ for $m \geq 1$ such that $C_{(d_k,e_k)} \cap C_{(d_{k+1},e_{k+1})} \neq \emptyset$ for all $1 \leq k < m$, $(a,0,0) \in C_{(d_1,e_1)}$, and $(b,0,0) \in C_{(d_m,e_m)}$. By construction of T_S it can be seen that $C_{(d_k,e_k)} \cap C_{(d_{k+1},e_{k+1})} \neq \emptyset$ implies $\{d_k,e_k\} \cap \{d_{k+1},e_{k+1}\} \neq \emptyset$. It follows that $(a,b) \in TC(S)$ We conclude that $Q_{trans} \in FO + poly + Pconn$.

Now we show $Q_{trans}^{lin} \in FO + linear + Pconn$. Let S be a finite semi-linear subset of $\mathbb{R}^2$. Let T_S be defined as before. Since S is semi-linear and the definition of T_S does not involve multiplication, then T_S can be constructed from S in $FO + linear$. The rest of the argument goes through exactly as before.

□

Transitive closure allows for important graph-based spatial operations to be expressed in $FO + poly + Pconn$ (e.g. accessibility checking). Two interesting additional results are that the following queries can be expressed in $FO + poly + Pconn$: (i) the query which determines the parity of a finite semi-algebraic subset of $\mathbb{R}_{>0}$ and (ii) the query which computes the inside of a simple, closed, semi-algebraic curve in the plane (i.e. a set homeomorphic to the 1-sphere). To prove (i), let $\{a_1, ..., a_n\}$ be a semi-algebraic subset of $\mathbb{R}_{>0}$. The following set can be defined from $\{a_1, ..., a_n\}$ in $FO + poly$: $S = \{(a_1, a_3), (a_2, a_4), (a_3, a_5), (a_4, a_6), \ldots, (a_{n-3}, a_{n-1}), (a_{n-2}, a_n), (a_1, a_n)\}$. Its transitive closure, $TC(S)$, contains (a_1, a_2) if and only if n is even. By Theorem 13, parity is expressible in $FO + poly + Pconn$.

To prove (ii), let $C \subseteq \mathbb{R}^2$ be a simple, closed, semi-algebraic curve. By the Jordan Curve Theorem ([23] pg. 383), $\mathbb{R}^2 - C$ has two path connected components, W_1 and W_2, exactly one of which is bounded (assume W_1). W_1 is the inside of C, so, computing the inside query is reduced to computing W_1. Consider the $FO+poly+Pconn$ "formula" $Pconn_{()}(\mathbb{R}^2 - C)$. This formula defines the set $PC = W_1 \times W_1 \cup W_2 \times W_2$. Let $\psi := \overrightarrow{x} \notin C \wedge \exists \delta > 0 \forall \overrightarrow{y} [(\overrightarrow{x}, \overrightarrow{y}) \in PC \rightarrow \| \overrightarrow{x} - \overrightarrow{y} \| < \delta]$. It can be checked that ψ defines W_1.

(i) further answers question (1) in the positive (since parity is shown in [14] to not be expressible in $FO + poly$). Moreover it shows that a very limited form of counting is possible in $FO + poly + Pconn$. (ii) allows for important geometric operations to be expressed (e.g. the ROSE algebra of Guting, Schneider [17] has an "inside of" operation).

7.2 Linearity and the Extension

Recall that $FO + poly + Pconn^{lin}$ are the queries in $FO + poly + Pconn$ which map semi-linear instances to semi-linear instances (likewise for $FO + poly^{lin}$). The relationship between the classes of queries $FO + poly, FO + linear, FO + poly^{lin}$ and their $Pconn$ extensions are depicted below (arrows indicate orientation, e.g. $FO + poly^{lin} \subsetneq FO + poly$).

$$\begin{array}{ccc}
FO + poly & \subsetneq & FO + poly + Pconn \\
\subsetneq\uparrow & & \subsetneq\uparrow \\
FO + poly^{lin} & \subsetneq & FO + poly + Pconn^{lin} \\
\subsetneq\uparrow & & \subsetneq\uparrow \\
FO + linear & \subsetneq & FO + linear + Pconn
\end{array}$$

All of the horizontal $\subsetneq$ follow from Theorem 13, and the fact that $Q_{trans} \notin FO + poly$.

Consider the left side of the figure. $FO + poly^{lin} \subsetneq FO + poly$ because there exists queries in $FO + poly$ which map semi-linear instances to semi-algebraic but not semi-linear instances (e.g. the query, Q_{S^1}, which maps any instance to the 1-sphere which is not a semi-linear set by [2]). $FO + linear \subsetneq FO + poly^{lin}$ follows from results in [2] and [26].

Consider the right side of the figure. $FO+poly+Pconn^{lin} \subsetneq FO+poly+Pconn$ because $Q_{S^1} \in FO+poly+Pconn-FO+poly+Pconn^{lin}$. Showing $FO+linear+Pconn \subsetneq FO+poly+Pconn^{lin}$ is a bit trickier. By definition, $FO+linear+Pconn \subseteq FO+poly+Pconn^{lin}$; suppose they were equal. Let Q_{colin} be the boolean query which on an input semi-linear set returns true if all points in the input are colinear. It can be seen that $Q_{colin} \in FO+poly^{lin}$ hence $Q_{colin} \in FO+linear+Pconn$. It is shown in [26], however, that $Q_{colin} \in FO + linear$ implies that multiplication can be defined, hence, $FO + linear = FO + poly$. We have that $FO+poly \subseteq FO+linear+Pconn$ which is impossible since $FO + linear + Pconn$ is closed (returns only semi-linear sets).

7.3 Extension Orthogonality

There appear to be two *orthogonal* ways of extending $FO+poly$ ($FO+linear$): (1) add aggregate operators (e.g. count, sum, average); and (2) add geometric operators (e.g. path connectivity). Toward (1) Benedikt and Libkin [7] developed an extension of $FO + poly$ with a finite, bag sum operator. Additionally, Grumbach et al. [13] develop an extension of $FO + poly$ with finite bag operators, sum, product, and average. We denote their resulting language $FO + poly + Agg$. Toward (2) we and Benedikt et al [6] independently introduced $FO + poly + Pconn$. The question raised here is to what extent are (1) and (2) orthogonal?

In a very limited form, $FO + poly + Pconn$ and $FO + linear + Pconn$ have counting aggregates since parity is expressible in each.

Nonetheless, full counting does not seem possible in $FO+poly+Pconn$ (but is open). Consider the query, Q_{ct}, which returns the count of a finite semi-algebraic set. Clearly, $Q_{ct} \in FO + poly + Agg$, see [13] pg 177 and $Q_{conn} \in FO + poly + Pconn$ (see Example 2). However, whether $Q_{ct} \in FO + poly + Pconn$ or $Q_{conn} \in FO + poly + Agg$ is open.

Finally consider $FO+poly+Pconn+Agg$. To what extent do the two extensions combine and what queries still cannot be done? We could begin to address this question by answering such a question as: "Is the query which computes the minimum spanning tree of a graph in $FO + poly + Pconn + Agg$?". All of this is left as future work.

8 Conclusions

Spatial constraint database models were introduced to provide a rigorous foundation for spatial database systems. This approach is a clean and natural generalization of the relational model to spatial data. Query languages for spatial data are modeled using first order logic over $\mathbb{R}$. The most prominent ones are $FO + poly$ and $FO + linear$. These languages can express many natural spatial queries (e.g. computing the convex hull). However, many important queries involving connectivity cannot be expressed.

In an effort to address this weakness we extend $FO + poly$ and $FO+linear$ with a path connectivity operator, $Pconn$. We show that $FO + linear + Pconn$ and $FO + poly + Pconn - 3D$ are closed and have PTIME data complexity. Benedikt et al. [6], independently of us, prove $FO + poly + Pconn$ and $FO + linear + Pconn$ are closed. While our closure results are not as general as theirs, our proof techniques are simpler and interesting in their own right as a contrast to their. In particular our closure proof of $FO + poly + Pconn - 3D$ does not require a change of coordinates in the computation of the decomposition. Their proof of closure for $FO+poly+Pconn$ requires a coordinate change. Moreover our proof technique highlights more clearly the essential properties and limitations of the decomposition produced by the Collins algorithm in proving closure. In doing so we raise an open problem concerning the geometric properties of decompositions produced by the Collins algorithm.

Finally we examine the expressiveness of $FO + poly + Pconn$ and $FO+linear+Pconn$ and determine that parity and transitive closure are expressible in each. Parity demonstrates that a very limited form of counting is possible in these extension languages. We conjecture that general counting is not possible (but is open). The development of techniques for proving inexpressiveness results of this kind for $FO + poly + Pconn$ and $FO + linear + Pconn$ is an interesting direction for future work.

Acknowledgments: The authors wish to thank Marc Gyssens, Floris Geerts, Jan Paredaens, and Leonid Libkin for several useful discussions. The authors also wish to thank Catherine Wyss, Bart Kuijpers, Jan Van den Bussche, Ed Robertson, Dennis Groth, and Mehmet Dalkilic.

References

[1] Abiteboul S., Hull R., Vianu V.. *Foundations of Databases*. Published by Addison-Wesley, 1995.

[2] Afrati F., Cosmadakis S., Grumbach S., Kuper G., *Linear Versus Polynomial Constraints in Database Query Languages*. Proceedings of the 2nd International Conference on Principals and Practice of Constraint Programming, A. Borning, ed.. Lecture Notes in Computer Science 874, 1994, pg. 181-192.

[3] Arnon D., Collins G., and McCallum S., Cylindrical Algebraic Decomposition I: The Basic Algorithm Quantifier Elimination and Cylindrical Algebraic Decomposition, Text and Monographs in Symbolic Computation, in: B. Caviness, J. Johnson eds., (Springer-Verlag, New York, 1998), 136-148.

[4] Arnon D., Collins G., McCallum S., An Adjacency Algorithm for Cylindrical Algebraic Decompositions of Three-dimensional Space, Journal of Symbolic Computation 5 (1988) 163-187.

[5] Arnon D., Geometric Reasoning with Logic and Algebra, Artificial Intelligence 37 (1988) 37-60.

[6] Benedikt M., Grohe M., Libkin L., Segoufin L., Reachability and Connectivity Queries in Constraint Databases, in: ACM SIGMOD-SIGACT-SIGART Symposium on Principals of Database Systems (PODS 2000) 104-115

[7] Benedikt M., Libkin L. On the Decidability of Exact and Approximate Aggregation in Constraint Query Languages, in:

ACM SIGMOD-SIGACT-SIGART Symposium on Principals of Database Systems (PODS 1999) 102-113.

[8] Bochnak, J., Coste, M., Roy, M., Real Algebraic Geometry (Springer-Verlag, New York, 1998).

[9] Bovet, D., Crescenzi P., Introduction to the Theory of Complexity (Prentice Hall International, New Jersey, 1994).

[10] Collins, G. Quantifier Elimination for Real Closed Fields by Cylindrical Algebraic Decomposition, in: Second GI Conference on Automata and Formal Languages 1975, Lecture Notes in Computer Science, Vol. 33, 134-183.

[11] E. Codd, A Relational Model of Data for Large Shared Data Banks, Communications of the ACM 13(6) (1970) 377-387.

[12] Dumortier F., Gyssens M., Vandeurzen L., Van Gucht D., On the Decidability of Semi-Linearity for Semi-Algebraic Sets and its Implications for Spatial Databases, in: ACM SIGMOD-SIGACT-SIGART Symposium on Principals of Database Systems (PODS 1997) 68-77.

[13] Grumbach S., Rafanelli M., Tininini L., Querying Aggregate Data, ACM SIGMOD-SIGACT-SIGART Symposium on Principals of Database Systems (PODS 1999), 174-184.

[14] Grumbach S., Su J., Finitely Representable Databases, Journal of Computer and System Sciences 55 (1997) 273-298.

[15] Guting, R., de Ridder T., Schneider M., Implementation of the ROSE Algebra: Efficient Algorithms for Realm-Based Spatial Data Types, in: Proceedings of the 4th International Symposium on Large Spatial Database Systems (Portland, August 1995) 216-239.

[16] Guting, R., Schneider M., Realms: A Foundation for Spatial Data Types in Database Systems, in: Proceedings of the 3rd International Symposium on Large Spatial Database Systems (Singapore, 1993) 14-35.

[17] Guting, R., Schneider M., Realm-Based Spatial Data Types: The ROSE Algebra, Journal on Very Large Databases 4 (1995) 243-286.

[18] Kanellakis P., Kuper G., Revesz P., Constraint Query Languages, Journal of Computer and System Sciences 51 (1995) 26-52.

[19] Kuijpers B., Paredaens J., Van den Bussche J., On Topological Elementary Equivalence of Spatial Databases, in: 6th International Conference on Database Theory (ICDT 1997), Lecture Notes in Computer Science, Vol. 1186 432-446.

[20] Lang, S., Linear Algebra, Third Edition, (Springer-Verlag, New York, 1987)

[21] Lema J. A., Guting R. H., Dual Grid: A New Approach for Robust Spatial Algebra Implementation, FernUniversitat Hagen, Informatik-Report 268, May 2000.

[22] Libkin, L., *Personal communication*, November 1999.

[23] Munkres J., Topology a First Course, (Prentice-Hall, New Jersey 1975).

[24] Rockafellar R. T., Convex Analysis, (Princeton University press, 1970).

[25] van den Dries L., Tame Topology and O-minimal Structures, in: London Mathematical Society Lecture Note Series 248 (Cambridge University press 1998).

[26] Vandeurzen L., Gyssens M., Van Gucht D., On Query Languages for Linear Queries Definable with Polynomial Constraints, in: Proceedings of the 2nd International Conference on Principals and Practice of Constraint Programming (1996), Lecture Notes in Computer Science, Vol. 1118 468-481.

[27] Vandeurzen L., Gyssens M., Van Gucht D., On the Desirability and Limitations of Linear Spatial Query Languages, in: 4th International Symposium on Advances in Spatial Databases (1995), Lecture Notes in Computer Science, Vol. 951, (1995) 14-28.