Computing Roadmaps of General Semi-Algebraic Sets

JOHN CANNY*

Computer Science Division, University of California, Berkeley, CA 94720, USA

In this paper we study the problem of determining whether two points lie in the same connected component of a semi-algebraic set S. Although we are mostly concerned with sets $S \subseteq \mathbb{R}^k$, our algorithm can also decide if points in an arbitrary set $S \subseteq R^k$ can be joined by a semi-algebraic path, for any real closed field R. Our algorithm computes a one-dimensional semi-algebraic subset $\Re(S)$ of S (actually of an embedding of S in a space $\hat{R}^k$ for a certain real extension field $\hat{R}$ of the given field R). $\Re(S)$ is called the roadmap of S. The basis of this work is the roadmap algorithm described in [3, 4] which worked only for compact, regularly stratified sets.

We measure the complexity of the formula describing the set S by the number of polynomials n, their maximum degree d, the maximum length of their coefficients in bits c, and the number of variables k. With respect to the above measures, the bit complexity of our new algorithm is $(n^k \log n)d^{O(k^2)}(c \log n)^{(1+\epsilon)}$ randomized, or $(n^k \log n)d^{O(k^4)}c^{(1+\epsilon)}$ deterministic, for any $\epsilon > 0$. Note that the combinatorial complexity (complexity in terms of n) in both cases is within a polylog factor of the worst-case lower bound for the number of connected components $\Omega(n^k)$.

Received November 1992, revised May 1993

1. INTRODUCTION

Good sequential and parallel algorithms have been available for some time for deciding the theory of the reals [10, 11, 3], and for real quantifier elimination [2, 14, 21]. Geometrically, these problems amount to deciding the emptiness or non-emptiness of semi-algebraic sets and their projections. Recently, attention has turned to computing geometric properties of semi-algebraic sets, such as counting the number of connected components. The first algorithm for connectivity was described by Schwartz and Sharir [22] who observed that a cylindrical algebraic cell decomposition provides a convenient substrate from which to extract connectivity information. This idea was pursued by Kozen and Yap [18], who gave a simple formula for deciding adjacency between cells, and made use of a parallel algorithm for constructing the cell decomposition [2] (this paper contains errors that were fixed in [8], see also [6]). Their algorithm was fully parallel, and had single-exponential parallel running time. From Kozen and Yap's results it follows that all geometric properties of semi-algebraic sets are at hand, because a semi-algebraic set can be represented as a regular cell complex.

However, all known methods for constructing algebraic cell decompositions require double exponential sequential time and improving this bound remains a major open problem. So these two cell decomposition algorithms run in sequential time double exponential in the dimension, with the latter taking single exponential time in parallel. Neither looks practical in the near future even in low dimensions.

A different approach was taken in [3] and [4], based on the construction of a one-dimensional skeleton of the set. This construction, called a roadmap, gave much lower (single exponential) complexity. The original roadmap papers [3] and [4] (see also [24]) made use of regular stratification [13], rather than a cell decomposition, which allowed a very coarse (and efficient) partition of the set. The complexity of finding paths in [3] was $(n^k \log n)d^{O(k^2)}c^2$, and since the lower bound on the number of components is $\Omega((nd)^k)$, this algorithm is nearly optimal in terms of n. This is important for applications in geometric modeling and robotics where d and k are small and fixed, but n, representing the number of surfaces, may be large. The disadvantage of the algorithm in [3], which we will henceforth refer to as the roadmap algorithm, is that it required the semi-algebraic set to be compact and in general position.

Later Grigor'ev and Vorobjov [12] gave a $(dn)^{O(k^{19})}$ algorithm for finding paths in arbitrary semi-algebraic sets, and the Franco-Argentine school in [15] and [16] gave a solution for the general case with a running time of $(dn)^{k^{O(1)}}$. However, the double exponents of both algorithms appears to make them impractical for robotics

*Supported by a David and Lucile Packard Foundation Fellowship and by NSF Presidential Young Investigator Grant IRI-8958577

or geometric modeling applications. The problem of actually finding the connected components in exponential time, i.e. giving quantifier-free formulae for them, was solved in [7] (see also [17]). In [9] roadmaps are defined using arbitrary, not necessarily generic, projection maps.

In this paper we describe a method for path-finding in the general case which has a running time of $(n^k \log^2 n)d^{O(k^2)}c^2$ randomized, which is within $\log n$ of the original roadmap algorithm. Making the algorithm deterministic increases the complexity to $(n^k \log n)d^{O(k^4)}c^2$. Note that the combinatorial complexity in terms of n in both cases is again close to the worst-case lower bound of $\Omega(n^k)$ on the number of connected components. The new method performs a direct reduction from the general case to the compact, regularly stratified case, so that the algorithm of [3] can be applied. This reduction increases the length of the formula by a constant factor, and its coefficient size by a factor of $k \log(dn)$. Each polynomial in the input formula is replaced by polynomials which differ in their constant coefficients by various infinitesimals. The new collection of polynomials define a semi-algebraic set which is compact and regularly stratified, and has the same connected components as the original.

2. PRELIMINARIES

The input to the algorithm is a semi-algebraic set S defined by polynomials with rational coefficients. For a set defined in k-dimensional space, the polynomials lie in the ring $\mathbb{Q}[x] = \mathbb{Q}[x_1, \ldots, x_k]$. Formally, a semi-algebraic set is defined as:

Definition 2.1. *Let $f_1, \ldots, f_n \in \mathbb{Q}[x]$ be a collection of polynomials with rational coefficients. A* semi-algebraic set *$S \subseteq \mathbb{R}^k$ is a set derived from the sets*

$$S_i = \{x \in \mathbb{R}^k \mid f_i(x) > 0\} \tag{1}$$

by finite union, intersection and complement.

A semi-algebraic set can be defined as the set of points in $\mathbb{R}^k$ satisfying a certain predicate of the form $B(A_1, \ldots, A_n)$ where $B : \{0,1\}^n \to \{0,1\}$ is a Boolean function and each A_i is an atomic formula of one of the following types:

$$(f_i = 0),\ (f_i \neq 0),\ (f_i > 0),\ (f_i < 0),\ (f_i \geq 0),\ (f_i \leq 0) \tag{2}$$

with $f_i \in \mathbb{Q}[x]$. In the analysis that follows it will be helpful to assume a certain form for the defining predicate.

Definition 2.2. *A formula $B(A_1, \ldots, A_n)$ is said to be in* monotone standard form *if the Boolean function B is monotone, and all atomic formulae A_i are either $(f_i = 0)$ or $(f_i > 0)$.*

An arbitrary formula can be converted to monotone standard form with a constant factor increase in size. Assume we are given a Boolean circuit C to represent the function B. This circuit can be converted to a negation-free, and therefore monotone circuit C_M as follows. For each node $v \in C$ there are two in C_M, one of which represents v and the other $\neg v$. Now all primitive logical operations between nodes in C_M can be implemented with $\wedge$ and $\vee$, e.g. if $C_7 = C_3 \vee \neg C_5$ in the circuit C, in C_M, the node representing C_7 is the $\vee$ of the node representing C_3 and the node representing $\neg C_5$. We also need to compute $\neg C_7$ for later use, and this node is the $\wedge$ of the node representing $\neg C_3$ and the node representing C_5.

The circuit C_M defines a monotone Boolean function of the original atomic formulae and their negations. The negations can be pushed into the atomic formulae by replacing $\neg(f_i > 0)$ with $(f_i \leq 0)$ etc. This formula can be converted to standard form by substituting for the inequalities $\leq, \neq, \geq$ with a union of a pair of inequalities using $>$ and $=$. Overall the number of atomic formulae increases by a factor of at most four compared to the original predicate.

Definition 2.3. *We measure the complexity of a predicate with four quantities, the number of polynomials n, the number of variables k, the maximum degree of the polynomials d, and the maximum coefficient length c of the coefficients of the polynomials.*

There remains one sticky point with regard to the Boolean formula B. There is a certain cost associated with evaluating B, given the signs of the f_i's. This time is clearly linear for the first evaluation. The algorithm of [3] requires frequent re-evaluation of B when a single f_i changes sign. Our complexity bounds will be valid if the time to re-evaluate B when a single f_i changes sign is $O(\log n)$. In section 9, we show that this will be true if the function B is defined by a formula. There we show there that an arbitrary Boolean formula can be converted to an equivalent log-depth *trinary* formula in polynomial time. The predicate B has been assumed to be a formula in most previous work on semi-algebraic sets. Our algorithm will still work if the function B is represented by a general circuit, but if the time to re-evaluate B when a single input changes is greater than log, then we must substitute this larger time for $\log n$ in the complexity bounds above.

2.1. Stratifications

Definition 2.4. *A* stratification *$\underline{\underline{S}}$ of a set $S \subset \mathbb{R}^k$ is a partition of S into a finite number of disjoint subsets S_i called* strata *such that each S_i is a manifold.*

A regular stratification satisfies some additional conditions which are well described in [13]. There are several basic ways to construct regular stratifications. We will only need two:

- **Taking products.** If $\underline{\underline{C}}$ and $\underline{\underline{D}}$ are regular stratifications of the spaces C and D respectively, then the product $\underline{\underline{C}} \times \underline{\underline{D}}$ is a regular stratification of $C \times D$.

- **Preimage of a transversal map.** If $F : M \to N$ is transversal to $\underline{\underline{D}}$ for a regular stratification $\underline{\underline{D}}$ of a subset $D \subseteq N$, then $F^{-1}(\underline{\underline{D}})$ is a regular stratification of $F^{-1}(D)$.

$F \pitchfork \underline{\underline{D}}$ means F is transversal to all the strata of $\underline{\underline{D}}$, and $F^{-1}(\underline{\underline{D}})$ is the set $\{F^{-1}(\sigma) \mid \sigma \in \underline{\underline{D}}\}$. If we now define

DEFINITION 2.5. *Let $f_i \in \mathbb{Q}[x]$, $i = 1, \ldots, n$ be a collection of polynomials that define a map $F : \mathbb{R}^k \to \mathbb{R}^n$. If $\underline{\underline{\mathbb{R}}} = \{\mathbb{R}_-, \{0\}, \mathbb{R}_+\}$, a* sign sequence *$\sigma$ is an element of $\underline{\underline{\mathbb{R}}}^n$. The sets $F^{-1}(\sigma)$ are called* sign-invariant sets *of F.*

Then we can view a semi-algebraic set S as a finite union of sign-invariant sets of some polynomial map F. The sign partition $(\underline{\underline{\mathbb{R}}})^n$ of $\mathbb{R}^n$ is a regular stratification of $\mathbb{R}^n$. So if a map $F : \mathbb{R}^k \to \mathbb{R}^n$ is transversal to $(\underline{\underline{\mathbb{R}}})^n$, then the preimage $F^{-1}((\underline{\underline{\mathbb{R}}})^n)$, which is the collection of sign-invariant sets of F, is a regular stratification.

2.2. Infinitesimals

We will make extensive use of extensions of real fields by infinitesimals. This process is simple to implement computationally, and has been well formalized in [1] using the real spectrum. An elementary description of the use of infinitesimal elements is given in [5] in an algorithm for the existential theory of the reals.

One disadvantage of working over an infinitesimal extension field is that basic field operations become very expensive. Usually, an element of $\mathbb{R}(\epsilon, \delta)$ is represented as a polynomial in ϵ and δ. The degree of such elements will typically be $O(d^{O(k)})$, and clearly with 3 or 4 infinitesimals, each field operation is enormously expensive.

But in section 7.1 a method is described for computing with infinitesimals which costs only slightly more than integer arithmetic in typical cases. The idea is to do arithemetic using straight line programs, and recover only the lowest degree rational coefficient of the field element by differentiation. Thus the use of infinitesimals in quantifier elimination can be a practical proposition.

DEFINITION 2.6. *For a given real field R, we say that an element ϵ is infinitesimal with respect to R if the extension $R(\epsilon)$ is ordered such that ϵ is positive, but smaller than any positive element of R.*

We will have cause to make use of towers of such field extensions. We will use the suggestive notation $\delta \gg \epsilon$ for two infinitesimals to mean that ϵ is infinitesimal with respect to the real closure of the field $R(\delta)$.

Strictly speaking, in what follows we do not need true infinitesimals. Any result that we prove for an infinitesimal will hold for all sufficiently small real values. This follows because the "bad sets" or critical sets for the various calculations we perform are semi-algebraic, and there is a smallest bad real value. Even when we use towers of infinitesimals, as long as we choose a good real value for each variable, there will be a good real value for the next variable that is "small enough" with respect to it, and which avoids the bad set. Some of our proofs will be phrased as though the infinitesimals were real numbers. This saves us having to use awkward definitions of compactness, connectivity and regularity for arbitrary real closed fields. This idea may be seen as a special case of the "Transfer Principle" [23].

3. REDUCTION FROM AN UNBOUNDED TO A BOUNDED SET

This is a rather standard reduction which is used in a number of places, see for example [5]. First we show that the unbounded set $S \subset \mathbb{R}^k$ is homotopy equivalent to a bounded set. For this purpose, let $\rho(x_1, \ldots, x_k) = x_1^2 + \cdots + x_k^2$ be a polynomial radius function. Consider the set $S \cap D_r$ where $D_r = \rho^{-1}([0, r])$.

LEMMA 3.1. *There exists a positive r_0 such that for all $r > r_0$, $S \cap D_r$ is a deformation retract of S.*

Proof. Let $F = (f_1, \ldots, f_n)$ be a set of polynomials which define S. Let $\underline{\underline{\mathbb{R}}}^k$ be a Whitney regular stratification of $\mathbb{R}^k$ which is compatible with the sign-invariant sets of F. All sign-invariant sets are unions of strata. Note that this is a different stratification than $\underline{\underline{\mathbb{R}}}^k$ which is the stratification by sign of $\mathbb{R}^k$. By the semi-algebraic Sard theorem [1], the map ρ has only finitely many critical values when restricted to any of these strata. Let r_0 be the largest critical value, then for all $r > r_0$, $S \cap D_r$ has the same homotopy type (hence number of connected components) as S. To see this, use ρ to lift a vector field on $[r, \infty)$ to one on $\mathbb{R}^k - \mathrm{Int}(D_r)$ which is compatible with the stratification of $\mathbb{R}^k$. The flow along this vector field defines a retraction of S onto $S \cap D_r$. ∎

So to find connected components of S it suffices to find components of $S \cap D_r$ for sufficiently large r. In practice this is done by treating r as an indeterminate element of the base field. When it comes time to determine the sign of a base field element, which will be a polynomial in r, we use the sign of the highest degree term in r. This correctly gives the sign for sufficiently large r.

4. REDUCTION TO A REGULAR STRATIFICATION

As was shown in [3] one can obtain a regular stratification by taking the sign-invariant sets of a system of polynomials in sufficiently general position. In the present case, the given f_i will not be in general position. In [3] a fixed perturbation was applied to their constant coefficients to achieve this. Now consider the following symbolic perturbation of the $F = (f_1, \ldots, f_n) : \mathbb{R}^k \rightarrow \mathbb{R}^n$. Define

$$F_{\epsilon a} = F + \epsilon a \qquad (3)$$

where $a \in \mathbb{R}^n_+$ is supposed constant for the time being, and ϵ is a single infinitesimal. As shown in [3] for almost all choices of the constant coefficients ϵa, the map $F_{\epsilon a}$ is transversal to the natural stratification by sign $(\underline{\underline{\mathbb{R}}})^n$ of $\mathbb{R}^n$, where $\underline{\underline{\mathbb{R}}} = \{\mathbb{R}_-, \{0\}, \mathbb{R}_+\}$. In particular, if we fix $a = (a_1, \ldots, a_n)$, then for almost all such choices, the map $F_{\epsilon a}$ is transversal to $(\underline{\underline{\mathbb{R}}})^n$ for almost all ϵ. This implies that the sign-invariant sets of $F_{\epsilon a}$ form a Whitney regular stratification, a very important property for us later. We assume for now that such a_i's have been found. Later we will show how they can be determined either deterministically or probabilistically.

Assume the a_i's were chosen to be positive, the sign-invariant sets of (3) are the same as for

$$\begin{array}{cc} f_1/a_1 & + \ \epsilon \\ f_2/a_2 & + \ \epsilon \\ \vdots & \vdots \\ f_n/a_n & + \ \epsilon \end{array} \qquad (4)$$

Define $g_i = f_i/a_i$. Then another way to describe the sign-invariant stratification of the last paragraph is as the preimage under $G = (g_1, \ldots, g_n)$ of the stratification $(\underline{\underline{\mathbb{R}}})^n$.

Now suppose instead of a single polynomial, we replace each f_i by four polynomials as follows:

$$f_i/a_i + \delta, \qquad f_i/a_i + \epsilon, \qquad f_i/a_i - \epsilon, \qquad f_i/a_i - \delta \qquad (5)$$

this gives us a system of $4n$ polynomials H_j, and we claim that for almost all a and for almost all pairs (ϵ, δ), the map H is transversal to $(\underline{\underline{\mathbb{R}}})^{4n}$.

We can say this another way as follows. Let $\underline{\underline{\mathbb{R}_{\delta\epsilon}}}$ be the stratification of the real line into the points $-\delta$, $-\epsilon$, ϵ, and δ and the open intervals in between and to $\pm\infty$. Then we have the following general position lemma:

LEMMA 4.1. *For almost all $a \in \mathbb{R}^n$, the map $G = (f_1/a_1, \ldots, f_n/a_n)$ is transversal to the stratification $(\underline{\underline{\mathbb{R}_{\delta\epsilon}}})^n$ of $\mathbb{R}^n$, for almost all (δ, ϵ).*

Proof. We say that a value of a is *regular* if the set of (δ, ϵ)-values such that G is *not* transversal to $(\underline{\underline{\mathbb{R}_{\delta\epsilon}}})^n$ has measure zero in $\mathbb{R}^2$. The map G is transversal to $(\underline{\underline{\mathbb{R}_{\delta\epsilon}}})^n$ if it is transversal to all the strata of $(\underline{\underline{\mathbb{R}_{\delta\epsilon}}})^n$. If we can find a value of a which is regular with respect to every stratum, we are done because the set of values of (δ, ϵ) for which G is not transversal to $(\underline{\underline{\mathbb{R}_{\delta\epsilon}}})^n$ is the union of the sets of values for which it is not transversal to the individual strata. This is a finite union of measure zero sets, and so has measure zero.

We show below that the set of a values which are not regular with respect to a given stratum has measure zero in $\mathbb{R}^n$. From this it follows that the set of a values which are not regular with respect to the stratification $(\underline{\underline{\mathbb{R}_{\delta\epsilon}}})^n$ is measure zero in $\mathbb{R}^n$. Again this is because it is the union of measure zero sets corresponding to non-regular values for the individual strata. This will complete the proof.

So it remains to show that the set of non-regular values a of G with respect to a given stratum $\sigma_{\delta\epsilon}$ in $\mathbb{R}_{\delta\epsilon}$ has measure zero in $\mathbb{R}^n$.

Now $\sigma_{\delta\epsilon}$ is a product of points and open intervals in $\mathbb{R}$. From the definition of transversality given earlier, it should be clear that G will be transversal to $\sigma_{\delta\epsilon}$ if dG is surjective in the subspace V of $\mathbb{R}^n$ comprising those coordinates where $\sigma_{\delta\epsilon}$ is a point. Write $G|_V$ for the map G restricted to those coordinates, and $\sigma_{\delta\epsilon}|_V$ for the restriction of $\sigma_{\delta\epsilon}$. By definition, $p = \sigma_{\delta\epsilon}|_V$ is a single point in V. The condition that G be transversal to $\sigma_{\delta\epsilon}$ is equivalent to the condition that p be a regular value of $G|_V$.

Now define $q \in V$ by $q_i = a_i p_i$. The condition that p is a regular value of $G|_V$ is the same as the condition that q is a regular value of $F|_V$. By Sard's theorem, we know that the set of non-regular q values has measure zero in V. In fact the semi-algebraic version of Sard's theorem [1] tells us that the set of bad values is semi-algebraic. Since it is a measure zero set, it must be contained in a algebraic proper subset $Z \subset V$.

The i^{th} coordinate of q is either $\pm a_i\epsilon$ or $\pm a_i\delta$. Choose a q that avoids Z and temporarily suppose $\epsilon = \delta = 1$. This fixes the corresponding values a_i to $\pm q_i$. Let L denote the map taking q to a when $\epsilon = \delta = 1$, and suppose henceforth that a is fixed at some value $L(q)$ where $q \notin Z$. If ϵ and δ are supposed variable again, the plane in V parametrized by ϵ and δ intersects Z in an algebraic set. Since this set does not contain the image of the point $(1, 1)$, it must be a codimension one subset. Hence for almost all pairs (δ, ϵ), the point q is a regular value of $F|_V$.

Now notice that any value of a which is not in the codimension one set $L(Z)$ is a regular value in the sense defined above, with respect to a given stratum. Taking the union of the $L(Z)$ for all strata, we obtain a codimension one set of a values for which G is not transversal to $\sigma_{\delta\epsilon}$ for almost all δ and ϵ. ■

Once we have this regular stratification, we can use a certain subset of the strata to approximate an arbitrary semi-algebraic set. Define $\mathbb{R}_{\text{sep}} = (-\infty, -\delta] \cup [-\epsilon, \epsilon] \cup [\delta, \infty)$, then $\mathbb{R}_{\text{sep}}$ is a closed set. Intuitively, this is a partition of the real line separating values that are nearly zero from those that are definitely non-zero. We will show later that the connected components of the preimage $G^{-1}((\mathbb{R}_{\text{sep}})^n)$ are in one-to-one correspondence with the connected components of sign-invariant sets of F (if $\delta \gg \epsilon > 0$ are both sufficiently small).

For each sign-sequence $\sigma \in (\underline{\underline{\mathbb{R}}})^n$, there is also "separated" sign-sequence $\sigma_{\rm sep}$ defined as

$$(\sigma_{\rm sep})_i = \begin{cases} [\delta,\infty) & \text{if } \sigma_i = \mathbb{R}_+ \\ [-\epsilon,\epsilon] & \text{if } \sigma_i = \{0\} \\ (-\infty,-\delta] & \text{if } \sigma_i = \mathbb{R}_- \end{cases} \tag{6}$$

and the $\sigma_{\rm sep}$ are exactly the connected components of $(\mathbb{R}_{\rm sep})^n$.

The set $\mathbb{R}_{\rm sep}$ has a regular stratification, denoted $\underline{\underline{\mathbb{R}_{\rm sep}}}$:

$$\underline{\underline{\mathbb{R}_{\rm sep}}} = \{(-\infty,-\delta), \{-\delta\}, \{-\epsilon\}, (-\epsilon,\epsilon), \{\epsilon\}, \{\delta\}, (\delta,\infty)\} \tag{7}$$

but because we chose a carefully, the preimage $G^{-1}((\underline{\underline{\mathbb{R}_{\rm sep}}})^n)$ is a Whitney regular stratification of $G^{-1}((\mathbb{R}_{\rm sep})^n)$. Since the preimage is also a closed, bounded set, and therefore compact, the algorithm of [3] can be applied directly to compute its connected components. This takes us a long way toward computing the connected components of a given semi-algebraic set, and leaves us only with the task of determining adjacencies between connected components of sign-invariant sets. For now we must show

THEOREM 4.2. *Let $G = (f_1/a_1, \ldots, f_n/a_n)$ and $\mathbb{R}_{\rm sep}$ be defined as above. Then the connected components of sign-invariant sets of F are in one-to-one correspondence with the connected components of $G^{-1}((\mathbb{R}_{\rm sep})^n)$ for almost all $a \in \mathbb{R}^n$ and for all sufficiently small $\delta \gg \epsilon > 0$*

The result follows from the next two lemmas. There is a natural correspondence between each non-empty sign-invariant set $F_\sigma = F^{-1}(\sigma)$ and the set $G^{-1}(\sigma_{\rm sep})$. The lemmas show that these two sets have the same homotopy type, by showing that they can be retracted onto a common subset.

Consider a particular sign-invariant set F_σ of F, and number the f_i's such that $f_1, \ldots, f_m$ are all zero on F_σ, and the remaining polynomials are non-zero. Assume also for simplicity that all non-zero f_i's are positive on F_σ. We replace each inequality $f_i > 0$ for $i > m$ with a new inequality $f_i \geq \delta$. Together with the inequality $\rho \leq r$ (ρ is the radius function defined earlier), this defines a closed set $F_\sigma^-(\delta)$ which is a subset of F_σ for $\delta > 0$. Since it is also a subset of the compact ball of radius r, it is compact.

We first show that the homotopy type of $F_\sigma^-(\delta)$ is the same as that of F_σ for small enough δ. In fact we have

LEMMA 4.3. *There exists a positive δ_0 such that for all positive $\delta < \delta_0$, the set $F_\sigma^-(\delta)$ is a deformation retract of F_σ.*

Proof. Consider the set D in $\mathbb{R}^{k+1}$ defined as $D = \{(x_1, \ldots, x_k, \delta) \mid x \in F_\sigma^-(\delta)\}$. Think of D as the "graph" of $F_\sigma^-(\delta)$. D is certainly semi-algebraic, and so has a Whitney regular stratification compatible with the signs of the polynomials f_i and $(f_i - \delta)$. The projection $\pi_\delta : (x, \delta) \mapsto \delta$ has a finite number of critical values when restricted to these strata. Choose $\delta_0 > 0$ to be the smallest positive critical value, and let δ be any positive number less than δ_0. Then π_δ is regular on all strata for values in the range $(0, \delta]$. We can use π_δ to lift a vector field on $(0, \delta]$ to a vector field on D which is compatible with its strata. Since D is compact when restricted to $[0, \delta]$, this gives us a deformation retraction of the set $D|_{(0,\delta]} = \pi_\delta^{-1}(0, \delta]$ onto the compact set $D_\delta = \pi_\delta^{-1}(\delta)$.

But if we define $\pi_x : (x, \delta) \mapsto x$, then the projection $\pi_x(G|_{(0,\delta]})$ is just F_σ, and $\pi_x(D|_\delta)$ is just $F_\sigma^-(\delta)$. Furthermore, composing π_x with the deformation retraction of the last paragraph gives us a deformation retraction of F_σ onto $F_\sigma^-(\delta)$. ■

To guarantee that δ is small enough, we leave δ as an indeterminate element of the base field (like r), and when it comes time to evaluate the sign of a base field element, which is a polynomial in r and δ, we first find the term of lowest degree in δ, then among all terms with this degree in δ we take the sign of the highest degree term in r. This is equivalent to preceding all evaluations with the quantification $\exists r_0 \; \forall r > r_0 \; \exists \delta_0 \; \forall \delta < \delta_0 \ldots$.

Operationally, this is also equivalent to working in a real field extension $\mathbb{R}(r, \delta)$ where r is larger than any element of $\mathbb{R}$ and ϵ is smaller than any element of the real closure of $\mathbb{R}(r)$. However, in the algorithms that follow, all the numerical calculations we make will involve polynomials from $\mathbb{R}[r, \delta]$. We do not need r and δ to be values that do not lie in $\mathbb{R}$, but only "sufficiently large" or "sufficiently small" real values that the signs of all polynomials are correctly computed. If we work only over the reals, life is much easier, since the usual notions of compactness and connectivity apply.

In the last lemma we defined compact sets $F_\sigma^-(\delta)$ with the useful properties that they are compact and are deformation retracts of sign-invariant sets. This means that each *connected component* of a sign-invariant set F_σ contains a single component of $F_\sigma^-(\delta)$. Next we define a set $F_\sigma^+(\delta, \epsilon)$ which is a "neighborhood" of $F_\sigma^-(\delta)$ and can be retracted onto it. To define $F_\sigma^+(\delta, \epsilon)$, we take each inequality in an f_i and replace it with one or two inequalities:

$$\begin{array}{ll} (f_i + \epsilon a_i \geq 0) \wedge (f_i - \epsilon a_i \leq 0) & \text{if } \sigma_i = \{0\} \\ (f_i - \delta a_i \geq 0) & \text{if } \sigma_i = \mathbb{R}_+ \end{array} \tag{8}$$

and notice that $F_\sigma^+(\delta, \epsilon)$ is now compact. By the previous general position lemma, it is also regularly stratified by the signs of the polynomials that define it. Notice also that $F_\sigma^-(\delta)$ is a subset of $F_\sigma^+(\delta, \epsilon)$.

LEMMA 4.4. *Assume δ is chosen to satisfy lemma 4.3. There exists a positive ϵ_0 such that for all $0 < \epsilon < \epsilon_0$, the set $F_\sigma^-(\delta)$ is a deformation retract of $F_\sigma^+(\delta, \epsilon)$.*

Proof. We only sketch this proof since it is almost identical to the proof of lemma 4.3. Let D in $\mathbb{R}^{k+1}$ be defined as $D = \{(x_1, \ldots, x_k, \epsilon) \mid x \in F_\sigma^+(\delta, \epsilon)\}$, treating

δ as a constant. For all sufficiently small $\epsilon > 0$, the set D has a Whitney regular stratification $\underline{\underline{D}}$ into sign-invariant sets of the polynomials that define it. There is some ϵ_0 which is the smallest positive critical value of the projection $\pi : (x, \epsilon) \mapsto \epsilon$ restricted to $\underline{\underline{D}}$. For $\epsilon < \epsilon_0$, We can use π to lift a vector field on $(0, \overline{\epsilon}]$ and thereby define the desired retraction. ■

Notice that $F_\sigma^+(\delta, \epsilon)$ is exactly the set $G^{-1}(\sigma_{\text{sep}})$, the non-singular approximation of $F^{-1}(\sigma)$. Since F_σ and $F_\sigma^+(\delta, \epsilon)$ have a common retract $F_\sigma^-(\delta)$, they have the same homotopy type and hence the same number of connected components. This completes the proof of theorem 4.2.

Finally, we observe that the sets $F_\sigma^+(\delta, \epsilon)$ and $F_{\sigma'}^+(\delta, \epsilon)$ are disjoint for σ and σ' distinct. This follows because if the union of a component of $F_\sigma^+(\delta, \epsilon)$ and a component of $F_{\sigma'}^+(\delta, \epsilon)$ were connected, then its image under G would have to be connected also. But that image must lie in the union of the disjoint sets σ_{sep} and σ'_{sep}, and it must intersect both, which is impossible.

So the connected components of $G^{-1}((\mathbb{R}_{\text{sep}})^n)$, which are the union of connected components of $F_\sigma^+(\delta, \epsilon)$'s, correspond exactly to the connected components of sign-invariant sets of F. Since $G^{-1}((\mathbb{R}_{\text{sep}})^n)$ is a compact set which is regularly stratified as $G^{-1}((\mathbb{R}_{\text{sep}})^n)$, we can apply the algorithm of [3], modified to work over arbitrary real coefficient fields as described in [5].

All that remains to determine connectivity of a given semi-algebraic set is to determine the adjacencies between connected components of sign-invariant sets. This we deal with in the next section.

5. DETERMINING ADJACENCIES BETWEEN SIGN COMPONENTS

In the last section, we modeled connected components of sign-invariant sets F_σ with regularly stratified compact "neighborhood" sets $F_\sigma^+(\delta, \epsilon)$. In this section, we show that these neighborhood sets can also be used to determine adjacencies between components. We will need to make use of a "big" neighborhood of one set and a "small" neighborhood of the other.

First we remark that if disjoint sets A and B have a connected union, then either $A \cap \overline{B} \neq \phi$, or $B \cap \overline{A} \neq \phi$. For if this were not true, both A and B would be open in the union $A \cup B$, and therefore both closed in $A \cup B$. So to check whether connected A and B have a connected union, it suffices to check whether either $A \cap \overline{B} \neq \phi$, or $B \cap \overline{A} \neq \phi$.

Let A be a connected component of F_σ, and let $A^+(\delta_1, \epsilon_1)$ be the corresponding connected component of $F_\sigma^+(\delta_1, \epsilon_1)$. Let B be a connected component of a second sign-invariant set $F_{\sigma'}$, and $B^+(\delta_0, \epsilon_0)$ be the corresponding connected component of $F_\sigma^+(\delta_0, \epsilon_0)$. Note the use of two different sets of infinitesimals. The next lemma shows that for suitable choices of the infinitesimals, adjacency of A and B can be checked by testing overlap of $A^+(\delta_1, \epsilon_1)$ and $B^+(\delta_0, \epsilon_0)$:

LEMMA 5.1. *With $A^+(\delta_1, \epsilon_1)$ and $B^+(\delta_0, \epsilon_0)$ as defined above, and for all sufficiently small $\delta_1 \gg \epsilon_1 \gg \delta_0 \gg \epsilon_0 > 0$,*

$$(A \cap \overline{B} \neq \phi) \quad \Longleftrightarrow \quad A^+(\delta_1, \epsilon_1) \cap B^+(\delta_0, \epsilon_0) \neq \phi$$

Proof. First we clarify what we mean by "sufficiently small". Let $\overset{\circ}{\forall}\alpha$ mean that there is some positive α' such that the formula that follows the quantifier is true for all positive α less than α'. The lemma states that $A \cap \overline{B} \neq \phi$ is equivalent to

$$\overset{\circ}{\forall}\delta_1 \ \overset{\circ}{\forall}\epsilon_1 \ \overset{\circ}{\forall}\delta_0 \ \overset{\circ}{\forall}\epsilon_0 \ (A^+(\delta_1, \epsilon_1) \cap B^+(\delta_0, \epsilon_0) \neq \phi) \tag{9}$$

We can simplify this formula by doing quantifier elimination from the inside out. Let $B^-(\delta_0)$ be the component of $F_{\sigma'}^-(\delta_0)$ that is a retract of $B^+(\delta_0, \epsilon_0)$. We claim that

$$\begin{aligned} \overset{\circ}{\forall}\epsilon_0 \ (A^+(\delta_1, \epsilon_1) \cap B^+(\delta_0, \epsilon_0) \neq \phi) \quad &\Longleftrightarrow \\ (A^+(\delta_1, \epsilon_1) \cap B^-(\delta_0) \neq \phi) \quad & \end{aligned} \tag{10}$$

To see this, notice that the set $(A^+(\delta_1, \epsilon_1) \cap B^+(\delta_0, \epsilon_0))$ in (x, ϵ_0)-space is compact when restricted to $\epsilon_0 \in [0, \epsilon_0']$, and so achieves a minimum non-negative value of ϵ_0. If this value is greater than zero, both formulae are false, if it is zero, both formulae are true. Next we need to show that

$$\overset{\circ}{\forall}\delta_0 (A^+(\delta_1, \epsilon_1) \cap B^-(\delta_0) \neq \phi) \iff (A^+(\delta_1, \epsilon_1) \cap B \neq \phi) \tag{11}$$

Let p be a point in $(A^+(\delta_1, \epsilon_1) \cap B)$, and let δ_0' be the minimum of the values of the f_i's at p, excluding those f_i's which are zero at p (we assume wlog that non-zero f_i's are positive). The first formula will be true for all δ_0 less than δ_0'. Conversely, if the first formula is true for *any* δ_0, then there is a point p in $(A^+(\delta_1, \epsilon_1) \cap B^-(\delta_0))$. This point is also in $(A^+(\delta_1, \epsilon_1) \cap B)$ since $B^-(\delta_0)$ is a subset of B, so the second formula is true also. It remains to show that

$$\overset{\circ}{\forall}\delta_1 \overset{\circ}{\forall}\epsilon_1 (A^+(\delta_1, \epsilon_1) \cap B \neq \phi) \quad \Longleftrightarrow \quad (A \cap \overline{B} \neq \phi) \tag{12}$$

Suppose first that $A \cap \overline{B} \neq \phi$, and let p be a point in $A \cap \overline{B}$. Choose δ_1 small enough so that p is in the (relative) interior of $A^+(\delta_1, \epsilon_1)$, i.e. none of the inequalities $f_i(p) \geq a_i \delta_1$ has equal arguments. We know that all neighborhoods of p intersect B, and the rest of the formula checks this. Specifically, for every $\epsilon_1 > 0$, $A^+(\delta_1, \epsilon_1)$ contains an absolute neighborhood U of p. This follows because all the polynomials that define $A^+(\delta_1, \epsilon_1)$ are non-zero at p (some f_i's may be zero of course, but $A^+(\delta_1, \epsilon_1)$ is defined by these polynomials $\pm a_i \epsilon_1$). Since U intersects B, so does $A^+(\delta_1, \epsilon_1)$.

Conversely, suppose $A \cap \overline{B} = \phi$. We show that $\overset{\circ}{\forall}\epsilon_1 \ A^+(\delta_1, \epsilon_1) \cap B \neq \phi$ is false for any $\delta_1 > 0$. Pick a $\delta_1 > 0$. Now choose any monotonically decreasing sequence $v_i \to 0$ of positive ϵ_1 values. Then for at least one of these values $A^+(\delta_1, \epsilon_1) \cap B = \phi$.

Why? Suppose the intersection were non-empty for all v_i. Let p_i be a point in $A^+(\delta_1, v_i) \cap B$. The sequence (p_i) lies in the compact set $A^+(\delta_1, v_1) \cap \overline{B}$, and so has a convergent subsequence. But let p be a limit point, we must have $p \in A^-(\delta_1) \subset A$ by continuity of the polynomials defining A^+. So we have a subsequence of $(p_i) \in B$ which converges to $p \in A$. This shows $A \cap \overline{B} \neq \phi$, contrary to our assumption of the last paragraph. So the above assumption was false, and $A^+(\delta_1, \epsilon_1) \cap B$ must be empty for some (in fact almost all) $v_i > 0$. ■

To compute with this quantification, we once again treat δ_1, ϵ_1, δ_0 and ϵ_0 as real elements of the ground field. This is equivalent to a real field extension by infinitesimals $\mathbb{R}(\delta_1, \epsilon_1, \delta_0, \epsilon_0)$ where δ_1, ϵ_1, δ_0 and ϵ_0 are adjoined in that order, and each is taken to be smaller than any positive element of the previous extension field. However, the correct semantics is that they are real numbers that are small enough (relative to previously quantified values) that they give the same sign for the ground field elements computed by our algorithm.

An important corollary of the above theorem is that we can compute the connected components of an arbitrary semi-algebraic set by computing the connected components of a certain compact, regularly stratified set. First let us define a fine stratification of the real line $\underline{\underline{\mathbb{R}_{\text{fin}}}}$ which consists of the eight points $\pm\delta_1$, $\pm\epsilon_1$, $\pm\delta_0$, $\pm\epsilon_0$ and the open intervals in between and to $\pm\infty$.

Let S be the original semi-algebraic set defined by polynomials $f_1, \ldots, f_n$. Assume without loss of generality or efficiency, that the formula defining S is a monotone Boolean function of inequalities of the form $f_i = 0$ or $f_i > 0$. Suppose a suitable $a \in \mathbb{R}^n$ has been chosen so that $G = (f_1/a_1, \ldots, f_n/a_n)$ is transversal to $(\underline{\underline{\mathbb{R}_{\text{fin}}}})^n$. Replace each $(f_i = 0)$ in the formula for S with the conjunction

$$(f_i \leq \epsilon_0 a_i) \wedge (f_i \geq -\epsilon_0 a_i) \qquad (13)$$

and each $(f_i > 0)$ with $(f_i \geq \delta_0 a_i)$. Call this new set $S_{\text{sep}0}$. We need the defining Boolean formula to be monotone so that the new set $S_{\text{sep}0}$ is a union of separated sign-invariant sets $G^{-1}(\sigma_{\delta_0\epsilon_0})$.

Similarly, we can define a set $S_{\text{sep}1}$ as above using ϵ_1 and δ_1. We can now state the result relating the connected components of S and the regularly stratified sets $S_{\text{sep}0}$ and $S_{\text{sep}1}$:

COROLLARY 5.2. *There is a one-to-one correspondence between connected components of S and connected components of $S_{\text{sep}0} \cup S_{\text{sep}1}$.*

Proof. Let $A_1, \ldots, A_q$ be the connected components of sign-invariant sets of the map F that are contained in S. Note that this means there may be more than one A_i within the same sign-invariant set (in fact if one component of a given sign-invariant is in S, all the others must be).

We saw in the last section that each A_i corresponds in a simple way with a certain connected component $A_i^+(\delta_0, \epsilon_0)$ of $S_{\text{sep}0}$, and to a certain component $A_i^+(\delta_1, \epsilon_1)$ of $S_{\text{sep}1}$. Now consider the union $A_i^{++} = A_i^+(\delta_0, \epsilon_0) \cup A_i^+(\delta_1, \epsilon_1)$. This is a connected set, and it intersects A_i. We complete the proof by showing that $A_i \cup A_j$ is connected for distinct A_i and A_j if and only if $A_i^{++} \cup A_j^{++}$ is connected.

If $A_i \cup A_j$ is connected, then one set intersects the closure of the other. Assume for instance that $A_i \cap \overline{A_j} \neq \phi$. We know from the theorem of this section that this implies $A_i^+(\delta_1, \epsilon_1) \cap A_j^+(\delta_0, \epsilon_0) \neq \phi$, so that $A_i^{++} \cup A_j^{++}$ must be connected.

Conversely, suppose $A_i^{++} \cup A_j^{++}$ is connected. We know from the last section that $A_i^+(\delta_0, \epsilon_0)$ and $A_j^+(\delta_0, \epsilon_0)$ are disjoint, and similarly for $A_i^+(\delta_1, \epsilon_1)$ and $A_j^+(\delta_1, \epsilon_1)$. So the two remaining possibilities are $A_i^+(\delta_1, \epsilon_1) \cap A_j^+(\delta_0, \epsilon_0) \neq \phi$ which implies $A_i \cap \overline{A_j} \neq \phi$, or $A_i^+(\delta_0, \epsilon_0) \cap A_j^+(\delta_1, \epsilon_1) \neq \phi$ which implies $\overline{A_i} \cap A_j \neq \phi$. In either case $A_i \cup A_j$ is connected. ■

5.1. Transformation Algorithm

To summarize, the following algorithm reduces calculation of connected components of a general semi-algebraic set to calculation of connected components of a compact, regularly stratified set.

- Convert the input formula to monotone standard form, and if necessary, collapse the defining formula to a trinary formula of logarithmic depth as described in section 9.
- Add to the formula a conjunction with the polynomial inequality $\sum_{j=1}^{k} x_j^2 \leq r^2$ (converted to standard form), where r is an infinite positive value (larger than any real). Let the resulting formula be $B(A_1, \ldots, A_n)$, it defines a bounded set in the extension field $\mathbb{R}(r)$.
- Choose an $a \in (\mathbb{R}_+)^n$ at random, or let $a_1 \gg a_2 \gg \cdots \gg a_n > 0$ be a series of infinitesimals.
- Construct a formula B_0 from the input formula $B(A_1, \ldots, A_n)$ as follows. For each atomic predicate A_i, replace A_i with

$$\begin{array}{lll} \text{if } A_i \text{ is } f_i = 0 & \text{then } (f_i + \epsilon_0 a_i \geq 0) \wedge & \\ & \qquad (f_i - \epsilon_0 a_i \leq 0) & (14) \\ \text{if } A_i \text{ is } f_i > 0 & \text{then } (f_i - \delta_0 a_i \geq 0) & \end{array}$$

  where $\delta_0 \gg \epsilon_0$ are infinitesimals. Then the set $S_{\text{sep}0}$ defined by this formula is closed and bounded, therefore compact (or semi-algebraic compact, if we think in terms of true infinitesimals rather than sufficiently small reals). By the results of section 4 the connected components of the sign-invariant sets of B_0 are in one-to-one correspondence with those of B.
- Now define a new formula B_1 from $B(A_1, \ldots, A_n)$ as in the previous step but with ϵ_1 and δ_1 replacing ϵ_0 and δ_0. Set $\delta_1 \gg \epsilon_1 \gg \delta_0$. The formula B_1 defines a set $S_{\text{sep}1}$, and by the results of section 5, the

connected components of $S_{\text{sep}0} \cup S_{\text{sep}1}$ are in one-to-one correspondence with the connected components of S.

- Return $B_0 \vee B_1$, which defines a regularly stratified, compact semi-algebraic set whose connected components correspond one-to-one with those of S.

6. COMPLEXITY

A full analysis of the algorithm is given in [3], but the main ideas are simple enough to describe here. The main facts we need are the following:

1. The algorithm of [3], modified to use the BKR lemma as described in [5] makes all its branching decisions based on the signs of query polynomials whose degree in the coefficients of the input polynomials is $d^{O(k^2)}$. This bound is obtained by inspecting the resultant matrices used in calculating projections, and the 2-d point ordering algorithm.
2. The number of such polynomials that might ever occur in the calculation is $(nd)^{O(k^2)}$, obtained by considering all the possible slices that might be taken recursively.
3. Each query polynomial contains at most $O(k^2)$ infinitesimals, even if all of $a_1, \ldots, a_n$ are infinitesimal. This seems surprising at first, but one must remember that by using infinitesimal a_i's, we have guaranteed that all the algebraic surfaces defined by the input polynomials meet transversally. In particular, any collection of more than k surfaces wıll not have a common intersection point. If the query polynomials were only generated by intersection points, there could be at most $O(k)$ infinitesimals in each one. But the roadmap algorithm also generates hyperplanes which are defined by k input surfaces. These also meet the input surfaces transversally (except at one point each), and so an intersection point can actually depend on $O(k^2)$ input surfaces.

6.1. Deterministic Version

Making these observations allows us to determine the running time for the deterministic version of the algorithm, which uses infinitesimal a_i's. The algorithm of [3] must also use generic linear projection maps $\pi : \mathbb{R}^k \to \mathbb{R}^2$. Because the algorithm recurses on dimension, there need to be k choices of such maps. So all the π's can be specified with $2k^2$ real values. For the deterministic version, we define them as infinitesimals. Overall, we have the sequence

$$\rho \gg a_1 \gg \cdots a_n \gg \delta_1 \gg \epsilon_1 \gg$$
$$\delta_0 \gg \epsilon_0 \gg \pi_1 \gg \cdots \pi_{2k^2} \gg \mu > 0$$

The bounds in [3] and [5] show that the roadmap algorithm for a particular input requires $O(n^k \log n) d^{O(k^2)}$ evaluations of query polynomials. The basic query is determining the signs of the polynomials f_i at certain points along an algebraic curve. As explained in step 3 of the complexity summary above, each query may depend on k^2 of the a_i's and possibly all of the maps π, involving a total of $O(k^2)$ infinitesimals. From [3], we know the degree of the queries is $d^{O(k^2)}$. So each such query, as a polynomial in those infinitesimals, may have $d^{O(k^4)}$ coefficients. Therefore they require $d^{O(k^4)}$ time to evaluate. The running time is the product of the number of query polynomial evaluations and the time for each which is $(n^k \log n) d^{O(k^4)}$. Adding the cost of integer arithmetic, we get an overall bound for the deterministic algorithm of

$$(n^k \log n) d^{O(k^4)} c^{(1+\epsilon)} \qquad (15)$$

for any $\epsilon > 0$. Here c is the bound on the bit length of the input coefficients and we assume the cost of arithmetic on b-bit integers is $O(b^{(1+\epsilon)})$.

6.2. Randomized Version

For the randomized algorithm, we need to figure out the number of random bits required in the choice of a_i's and the maps π. Let Π denote all k of the projection maps $\pi : \mathbb{R}^k \to \mathbb{R}^2$. We could try to figure out explicitly the conditions for a particular (a, Π) to be a good choice, but there is a simpler argument we can use, which takes advantage of the fact that our calculation can be expressed as an algebraic decision tree. A particular (a, Π) must be a good choice if *all the query polynomials in the decision tree are non-zero* at that (a, Π) (excepting query polynomials which are identically zero, which can be ignored). This follows because for such an (a, Π), there is an open, connected neighborhood $N(a, \Pi)$ such that all the query polynomials have the same sign over all of $N(a, \Pi)$ as they do at (a, Π). Thus the algorithms output is that same for all these choices. But almost all of the points in $N(a, \Pi)$ must be good choices, since good points are dense. The algorithm must produce the correct output at these points, hence it produces the correct output at (a, Π).

So it suffices to choose (a, Π) to avoid the zero sets of all the query polynomials. The query polynomials have degree $d^{O(k^2)}$ and there are potentially $(nd)^{O(k^2)}$. The union of the zero sets gives us a bad set which is an algebraic set in the space of possible a and Π values which has degree $(nd)^{O(k^2)}$. By Schwartz's lemma, we will have probability p of hitting the bad set if we choose the a_i's and π_j's randomly with $\log(p^{-1}(nd)^{O(k^2)})$ bits. Fixing p, we see that $O(k^2 \log(nd))$ bits suffice. This contributes the extra factor of roughly $\log n$ to the running time of the original roadmap algorithm, and gives a randomized running time of

$$(n^k \log n) d^{O(k^2)} (c \log n)^{(1+\epsilon)} \qquad (16)$$

for any $\epsilon > 0$.

7. COMPUTING WITH INFINITESIMALS

This section addresses the practical problems of computing over an infinitesimal extension of the real numbers. In the algorithms of [14, 21, 12, 6] and this paper, various singularities are dealt with by perturbing the input polynomials with infinitesimals. This moves the problem away from the singularity, and when done carefully preserves the important properties (like connectivity or non-emptiness) of the input. Computations with an infinitesimal ϵ are done in the rational field $\mathbb{Q}(\epsilon)$. That is, each number a or b in this extension field is a rational function (a quotient of polynomials) in ϵ. To perform arithmetic, we use the usual rules for arithmetic on rational functions. To determine the sign of such an element, we exclusive-or the signs of its numerator and denominator, which are polynomials in ϵ. To determine the sign of a polynomial in ϵ, we use the sign of the lowest degree non-zero coefficient.

But it is very expensive to compute with explicit rational functions. For example, in the extension $\mathbb{R}(\mu, \epsilon, \delta, \rho)$ that we have been using, an element of degree 10 would have several hundred coefficients. But the sign of the element, which is all we need for the sign-determination algorithm, is determined by just one of these coefficients. This element is the lowest degree element under the lexicographic ordering $\mu \prec \epsilon \prec \delta \prec \rho$

If we knew that this element was say $\mu^4\epsilon\delta^2\rho$, we could find it by computing modulo the ideal $(\mu^5, \epsilon^2, \delta^3, \rho^2)$, which effectively discards higher-degree terms. Since we dont know the degree, we would have to do some search, gradually increasing degree until we obtain a non-zero term. Rather than doing this repeatedly, we can obtain the lowest degree term by differentiating a straight-line program.

7.1. Straight-Line Programs

Suppose we have computed an element $a \in \mathbb{R}$ from some other real values $b_1, \ldots, b_m$ via a series of arithmetic operations. For example, such a could be a coefficient of one of the Sturm query polynomials. We can represent a as a straight-line program (DAG whose vertices represent the arithmetic operations) rooted at the values $b_1, \ldots, b_m$. Now suppose that b_1 is specialized to the infinitesimal value ϵ, and that the other b_i take on integer values. We would like to know the sign of a. For simplicity we assume that $a = a(\epsilon)$ is a polynomial in ϵ. This is all we need in our applications.

We could substitute $\epsilon = 0$ and evaluate the straight-line program over the rationals. If we are lucky, $a(0)$ will have a non-zero value, and this gives the sign of $a(\epsilon)$. If not, we can construct a straight-line program for the derivative $\frac{da(\epsilon)}{d\epsilon}$. This has roughly double the size of the original straight-line program. Now evaluating this program at $\epsilon = 0$ gives us a_1, the coefficient of ϵ in $a(\epsilon)$. If this is non-zero, it gives us the sign of $a(\epsilon)$, otherwise we compute the second derivative, and continue. The extra program for the k^{th} derivative is about $k+1$ times the size of the original program, and it uses nodes from the first $k-1$ derivatives. The ***total*** program size to compute the k^{th} derivative is $\binom{k+2}{2}$ times the original.

This process generalizes easily to multivariate elements, using randomization. For example, to find the sign of $a(\mu, \epsilon, \delta)$ with $\mu \ll \epsilon \ll \delta$, we first substitute random integer values for ϵ and δ. With high probability, this doesnt change the degree of the lowest degree term in μ. Then we apply the procedure above to obtain a straight-line program for the first non-zero derivative at $\mu = 0$. Let $aa(\mu, \epsilon, \delta)$ denote this derivative. Then $aa(0, \epsilon, \delta)$ is the lowest-degree coefficient of a in μ, times the constant $k_1!$, where k_1 is the order of the derivative.

We iterate the process, and set δ to a random integer, μ to zero, and run the univariate procedure on the straight-line program for aa as a polynomial in ϵ. This gives us a straight-line program for the first non-zero derivative at $\epsilon = 0$, which we denote $aaa(\mu, \epsilon, \delta)$.

Finally, we run the univariate routine on the straight-line program aaa with μ and ϵ both set to zero. Evaluating the resulting program at $\mu = \epsilon = \delta = 0$ gives the sign of the lexicographically first term, which is what we need.

Some simple analysis shows that the straight-line program for computing the sign when the lowest degree term is $\mu^{k_1}\epsilon^{k_2}\delta^{k_3}$ is $k_1^2k_2^2k_3^2$ times the original. More generally we have

PROPOSITION 7.1. *Let $P(\epsilon_1, \ldots, \epsilon_m)$ be a polynomial, represented as a straight-line program of size L. If $\epsilon_1 \ll \cdots \ll \epsilon_m$ are infinitesimals, and if the lexicographically first term in P is $c\epsilon^{k_1}\cdots\epsilon^{k_m}$, then a straight-line program for this term can be constructed having size $\leq Lk_1^2\cdots k_m^2$, in the same number of steps.*

We claim this method is useful in practice because the k_i's are typically small constants, independent of the degree of a in μ, ϵ, δ. Each infinitesimal is used to perturb away from a possibly singular input, and the degree in that infinitesimal is a measure of the multiplicity of the singularity. Where the input is not singular at all, the degree in that infinitesimal will be zero. Most of the time, we expect small multiplicities, and the cost of working over the infinitesimal extension should be only a small constant factor more than integer arithmetic, this factor being the increase in the straight-line program size.

8. CONCLUSIONS

We presented a perturbation method that, given a formula describing a general semi-algebraic set S, produces a formula defining a set S_0 which is compact and regularly stratified by the signs of its defining polynomials. The connected components of S_0 are in one-to-one correspondence with those of S, and "contain" them using the natural embedding of $\mathbb{R}$ in the real closure of the extension field defined by some infinitesimals. The perturbation method uses only a constant number of infinitesimals in its randomized version (5), and can be performed with a number of arithmetic sets which is linear in the size of the original formula.

The perturbation method reduces the calculation of roadmaps, hence connected components, of general semi-algebraic sets to the case of compact, regularly stratified sets. This case was treated earlier in [3]. The resulting algorithm for general sets has a running time of $(n^k \log n)d^{O(k^2)}(c\log n)^{(1+\epsilon)}$ randomized, or $(n^k \log n)d^{O(k^4)}c^{(1+\epsilon)}$ deterministic. Similar algorithms for deciding non-emptiness only of semi-algebraic sets are presented in [6].

9. APPENDIX: DEPTH COMPRESSION OF BOOLEAN FORMULAE

Let $B(b_1,\ldots,b_n)$ be a Boolean function that is defined by an input formula B^F. We think of B^F as a rooted binary tree whose leaves are the b_i's. By pushing all negations through to the inputs, we may assume that B^F contains only $\vee$ and $\wedge$. In the worst case, such a tree with $2n-1$ vertices may have depth very close to n, which means it may take a long time to determine the value of the output when one of the input variables changes.

We show that for any B^F, there is a formula C^F using a three-valued (trinary) logic which computes the same function as B^F, such that the size of C^F is $O(n)$ and its depth is only $O(\log n)$. It should be clear that the time required to compute correct values at all nodes of an $O(\log n)$-depth formula when a single input changes is $O(\log n)$. We use the trinary formula to determine whether the curve segments along a silhouette curve are inside or outside of S in $O(\log n)$ time.

The trinary formulae uses three-state logic. Every vertex may have a value of 0, 1, or X. Certain vertices, like the leaves and the root, will only ever take on a value of 0 or 1. The trinary circuit will have $\vee$ and $\wedge$ vertices, which work as follows:

$$a \vee b = \begin{cases} 1 & \text{if either } a \text{ or } b \text{ is } 1 \\ 0 & \text{if both inputs are } 0 \\ X & \text{otherwise.} \end{cases} \tag{17}$$

$$a \wedge b = \begin{cases} 0 & \text{if either } a \text{ or } b \text{ is } 0 \\ 1 & \text{if both inputs are } 1 \\ X & \text{otherwise.} \end{cases} \tag{18}$$

There is also an asymmetric function $\succ$ defined as follows

$$a \succ b = \begin{cases} a & \text{if } a = 0 \text{ or } a = 1 \\ b & \text{otherwise} \end{cases} \tag{19}$$

For a vertex u in B^F, we let $D(u)$ denote the subtree rooted at u. For two vertices u, v in B^F, we use the notation $B^F(u,v)$ to denote the subtree $D(u)-D(v)+\{v\}$. Thus v should be a descendent of u.

Our construction is based on recursively constructing trinary formulae $C(u,v)$. The formula $C(u,v)$ takes as input all the leaves (b_i's) in the subtree $B^F(u,v)$, and outputs the following:

$$C(u,v) = \begin{cases} 0 & \text{if } u = 0 \text{ irrespective of the value of } v \\ 1 & \text{if } u = 1 \text{ irrespective of the value of } v \\ X & \text{otherwise} \end{cases} \tag{20}$$

Assuming we can construct such a formula, we can then build a trinary formula equivalent to B^F as follows: Let r be the root of B^F, and b_1 be some leaf of B^F. Then the formula we return is $C_1(r) = C(r,b_1) \succ b_1$ which is 0-1 valued, and correctly computes B.

Now we construct recursively a low depth $C(u,v)$. For the base case, if $u = v$ we return X. Otherwise the tree $B^F(u,v)$ has some number $m > 1$ of vertices. In any tree of m vertices there is an edge whose removal splits the tree into two subtrees with between $m/3$ and $2m/3$ vertices. Let p be the vertex below such a partitioning edge in $B^F(u,v)$, and let z be the lowest common ancestor of p and v. Now there are two cases:

If $p = z$, we recursively compute formulae for $C(p,v)$ and $C(u,p)$ and we return the formula $C(u,p) \succ C(p,v)$.

If $p \neq z$, then let z_1 be the child of z above p, and let z_2 be the child of z above v. We recursively compute the formulae $C(u,z)$, $C(z_1,p)$, and $C(z_2,v)$, and $C_1(p)$. The formula $C_1(p)$ is defined like $C_1(r)$ above. Assuming z to be an $\vee$-vertex, we return the formula:

$$C(u,z) \succ ((C(z_1,p) \succ C_1(p)) \vee C(z_2,v)) \tag{21}$$

and a similar formula is computed with $\wedge$ if z is a $\wedge$-vertex. Notice that all subformulae are computed from subtrees of size at most 2/3 of the size of $B^F(u,v)$. So the depth of the recursive calculation is at most $O(\log n)$. The depth of the formula increases by a constant amount with each recursion level, so the depth of $C_1(r)$ is at most $O(\log n)$. If we remove the constant X's from the formula, all its leaves will be b_i's, and since it is a tree it will have only $O(n)$ vertices.

REFERENCES

[1] J. Bochnak, M. Coste, and M.-F. Roy. *Géométrie algébrique réelle*. Number 12 in Ergebnisse der Mathematik 3. Springer-Verlag, Berlin, 1987.

[2] M. Ben-Or, D. Kozen, and J. Reif. The complexity of elementary algebra and geometry. *J. Comp. and Sys. Sci.*, 32:251–264, 1986.

[3] J.F. Canny. *The Complexity of Robot Motion Planning.* M.I.T. Press, Cambridge, 1988.

[4] J.F. Canny. Constructing roadmaps of semi-algebraic sets I: Completeness. *Artificial Intelligence*, 37:203–222, 1988.

[5] J.F. Canny. Some algebraic and geometric computations in PSPACE. In *ACM Symposium on Theory of Computing*, pages 460–467, 1988.

[6] J.F. Canny. Improved algorithms for sign determination and existential quantifier elimination. *The Computer Journal* 36(5):409–418, 1993.

[7] J.F. Canny, D.Y. Grigor'ev and N.N. Vorobjov. Finding Connected Components of a Semialgebraic Set in Subexponential Time. *Applicable Algebra in Engineering, Communication and Computing*, 2:217–239, 1992.

[8] N. Fitchas, A. Galligo, and J. Morgenstern. Algorithmes rapides en séquential et en parallèle pour l'élimination des quantificateur en géométrie élémentaire. *Sém. Structures Algébriques Ordonnées*, 1987.

[9] L. Gournay and J.-J. Risler. Construction of roadmaps in semialgebraic sets. Manuscript, 1991.

[10] D.Y. Grigor'ev. Complexity of deciding Tarski algebra. *Journal of Symbolic Computation*, 5:65–108, 1988.

[11] D.Y. Grigor'ev and N.N. Vorobjov. Solving systems of polynomial equations in subexponential time. *Journal of Symbolic Computation*, 5:37–64, 1988.

[12] D.Y. Grigor'ev and N.N. Vorobjov. Counting connected components of a semialgebraic set in subexponential time. *Computational Complexity*, 2:133–186, 1992.

[13] C.G. Gibson, K. Wirthmüller, A.A. Du Plessis, and E.J.N. Looijenga. *Topological Stability of Smooth Mappings.* Number 552 in Lecture Notes in Mathematics. Springer–Verlag, New York, 1976.

[14] J. Heintz, M.F. Roy, and P. Solerno. Complexité du principe de Tarski-Seidenberg. *Bull. Soc. Math. France*, 118:101–126, 1990.

[15] J. Heintz, M.F. Roy, and P. Solerno. Single-exponential path finding in semialgebraic sets I: The case of a smooth bounded hypersurface. In *AAECC-90, Springer LNCS 508*, 1990.

[16] J. Heintz, M.F. Roy, and P. Solerno. Single-exponential path finding in semialgebraic sets II: The general case. In *Proc. 60th Birthday Conf. for S. Abhyankar*, 1990.

[17] J. Heintz, M.F. Roy, and P. Solerno. Description des composantes connexes d'un ensemble semi-algebrique en temps simplement exponentiel. *Compte-Rendus Acad. Sci. Paris*, 313 Serié I, 1991.

[18] D. Kozen and C. Yap. Algebraic cell decomposition in NC. In *IEEE Conference on Foundations of Computer Science*, pages 515–521, 1985.

[19] P. Pedersen. Multivariate sturm theory. In *Proc. AAECC-9, New Orleans*, Springer LNCS **539**. 1991.

[20] P. Pedersen, M.-F. Roy, and A. Szpirglas. Counting real zeros in the multivariate case. In *Proc. MEGA-92*, 1992. To appear in *Computational Algebraic Geometry*, edited by F. Eyssette, and A. Galligo, Birkhauser.

[21] J. Renegar. On the computational complexity and geometry of the first-order theory of the reals, parts I, II and III. In *Journal of Symbolic Computation*, 13(3):255–352, 1992.

[22] J.T. Schwartz and M. Sharir. On the piano movers' problem, II: General techniques for computing topological properties of real algebraic manifolds. *Advances in Applied Mathematics*, 4:298–351, 1983.

[23] A. Tarski. *A Decision Method for Elementary Algebra and Geometry.* University of California Press, Berkeley, 1948.

[24] D. Trotman. On Canny's roadmap algorithm: Orienteering in semialgebraic sets. Technical report, Univ. Aix-Marseille, 1989.