Journal of Symbolic Computation 47 (2012) 131–152

Contents lists available at SciVerse ScienceDirect

Journal of Symbolic Computation

journal homepage: www.elsevier.com/locate/jsc

Complete subdivision algorithms, II: Isotopic meshing of singular algebraic curves

Michael Burr [a,1], Sung Woo Choi [b], Ben Galehouse [a,2], Chee K. Yap [a,c]

[a] Courant Institute, NYU, 251 Mercer Street, New York, NY 10012, United States
[b] Department of Mathematics, Duksung Women's University, Seoul 132-714, Republic of Korea
[c] Korea Institute of Advanced Study, 85 Hoegiro (Cheongnyangni-dong 207-43), Dongdaemun-gu, Seoul 130-722, Republic of Korea

ARTICLE INFO

Article history:
Received 28 January 2009
Accepted 15 November 2010
Available online 1 September 2011

Keywords:
Meshing
Singularity
Root bound
Evaluation bound
Implicit algebraic curve
Complete numerical algorithm
Subdivision algorithm

ABSTRACT

Given a real valued function $f(X, Y)$, a box region $B_0 \subseteq \mathbb{R}^2$ and $\varepsilon > 0$, we want to compute an ε-isotopic polygonal approximation to the restriction of the curve $S = f^{-1}(0) = \{p \in \mathbb{R}^2 : f(p) = 0\}$ to B_0. We focus on subdivision algorithms because of their adaptive complexity and ease of implementation. Plantinga & Vegter gave a numerical subdivision algorithm that is exact when the curve S is bounded and non-singular. They used a computational model that relied only on function evaluation and interval arithmetic. We generalize their algorithm to any bounded (but possibly non-simply connected) region that does not contain singularities of S. With this generalization as a subroutine, we provide a method to detect isolated algebraic singularities and their branching degree. This appears to be the first complete *purely numerical* method to compute isotopic approximations of algebraic curves with isolated singularities.

1. Introduction

Given $\varepsilon > 0$, a box region $B_0 \subseteq \mathbb{R}^2$ and a real valued function $f : \mathbb{R}^2 \to \mathbb{R}$, we want to compute a polygonal approximation P to the restriction of the implicit curve $S = f^{-1}(0)$ to B_0 (where $f^{-1}(0) =$

E-mail addresses: burr@cims.nyu.edu, mburr1@fordham.edu (M. Burr), swchoi@duksung.ac.kr (S.W. Choi), bgalehou@mpi-inf.mpg.de (B. Galehouse), yap@cs.nyu.edu (C.K. Yap).

[1] Present address: Mathematics Department, Fordham University, Bronx, NY 10458, United States. Tel.: +1 718 817 3298; fax: +1 718 817 3236.

[2] Present address: Max-Planck-Institut für Informtik Department 1: Algorithms and Complexity, D-66123 Saarbrücken, Germany.

0747-7171/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jsc.2011.08.021

$\{p \in \mathbb{R}^2 : f(p) = 0\}$). The approximation P must be (1) "topologically correct" and (2) "ε-close" to $S \cap B_0$. We use the standard interpretation of requirement (2), that $d(P, S \cap B_0) \leq \varepsilon$ where $d(\cdot, \cdot)$ is the Hausdorff distance between compact sets. In recent years, it has become accepted (Boissonnat et al., 2006) to interpret requirement (1) to mean that P is isotopic to $S \cap B_0$, which we denote by $P \approx S \cap B_0$. This means that we not only require that P and $S \cap B_0$ are homeomorphic, but also require that they are embedded in $\mathbb{R}^2$ "in the same way". This means that the two embeddings can be continuously deformed to each other, e.g., if $S \cap B_0$ consists of two disjoint ovals, these can be embedded in $\mathbb{R}^2$ as two ovals exterior to each other, or as two nested ovals. Isotopy, but not homeomorphism, requires P to respect this distinction. There is a stronger notion of isotopy called **ambient isotopy** (see the definition in Section 4). We use this stronger notion in this paper (but, for simplicity, we still say "isotopy"). See Boissonnat et al. (2006, p. 183) for a discussion of the connections between ambient and plain isotopy. In this paper, we focus mainly on topological correctness since achieving ε-closeness is not an issue for our particular subdivision approach (cf. Boissonnat et al. (2006, pp. 213–4)). This amounts to setting $\varepsilon = \infty$.

We may call the preceding problem the **2-D implicit meshing problem**. The term "meshing" comes from the corresponding problem in 3-D: given $\varepsilon > 0$ and an implicit surface $S : f(X, Y, Z) = 0$, we want to construct a triangular mesh M such that $d(M, S) \leq \varepsilon$ and $M \approx S$. It is interesting (see Burr et al. (in preparation)) to identify the 1-D meshing with the well-known problem of real root isolation and refinement for a real function $f(X)$.

The **algebraic approach** and the **numerical approach** constitute two extremes of a spectrum among the approaches to most computational problems on curves and surfaces. Algebraic methods can clearly solve most problems in this area, e.g., by an application of the general theory of cylindrical algebraic decomposition (CAD) (Basu et al., 2003). Purely algebraic methods, however, are generally not considered practical, even in the plane (e.g., Hong (1996) and Seidel and Wolpert (2005)), but efficient solutions have been achieved for special cases such as intersecting quadrics in 3-D (Schoemer and Wolpert, 2006). At the other end of the spectrum, the numerical approaches emphasize approximation and iteration. An important class of such algorithms is the class of **subdivision algorithms** which can be viewed as a generalization of binary search. Such algorithms are practical in two senses: they are easy to implement and their complexity is more adaptive with respect to the input instance (Yap, 2006). Another key feature of subdivision algorithms is that they are "localized", meaning that we can restrict our computation to some region of interest.

Besides the algebraic and numerical approaches, there is another approach that might be called the **geometric approach** in which we postulate an abstract computational model with certain (geometric) primitives (e.g., shoot a ray or decide if a point is in a cell). When implementing these geometric algorithms, one must still choose an algebraic or numerical implementation of these primitives. Implementations can also use a hybrid of algebraic and numerical techniques.

Unfortunately, numerical methods seldom have global correctness guarantees. The most famous example is the Marching Cube algorithm (Lorensen and Cline, 1987). Many authors have tried to improve the correctness of subdivision algorithms (e.g., Stander and Hart (1997)). So far, such efforts have succeeded under one of the following situations:

- (A0) Requiring niceness assumptions such as being non-singular or Morse.
- (A1) Invoking algebraic techniques such as resultant computations or manipulations of algebraic numbers.

It is clear that (A0) should be avoided. Generally, we call a method "complete" if the method is correct without any (A0) type restrictions. But many incomplete algorithms (e.g., Marching cube) are quite useful in practice. We want to avoid (A1) conditions because algebraic manipulations are harder to implement and such techniques are relatively expensive and non-adaptive (Yap, 2006). The complete removal of (A0) type restrictions is the major open problem faced by purely numerical approaches to meshing. Thus, Boissonnat et al. (2006, p. 187) state that "*meshing in the vicinity of singularities is a difficult open problem and an active area of research*". Most of the techniques described in their survey are unable to handle singularities. It should be evident that this open problem has an implicit requirement to avoid the use of (A1) techniques.

For example, the subdivision meshing algorithm of Plantinga (2006) and Plantinga and Vegter (2004) requires the non-singularity and the boundedness of curves and surfaces, (A0) assumptions. The subdivision algorithm of Seidel and Wolpert (2005) requires[3] the computation of resultants, an (A1) technique. We thus classify (Seidel and Wolpert, 2005) as a hybrid approach that combines numerical and algebraic techniques. Prior to our work, we are not aware of any meshing algorithm that can handle singularities without resorting to resultant computations. In general, hybrid methods offer considerable promise (e.g., Hong (1996)). This is part of a growing trend to employ numerical techniques to speed up algebraic computations.

Some of our recent work addresses the above (A0)/(A1) concerns: in Yap (2006), we gave a complete numerical approach for determining tangential Bezier curve intersections; in Cheng et al. (2009), we numerically solve zero-dimensional triangular systems without any "regularity" requirements on the systems; in Burr et al. (in preparation), we provide numerical root isolation in the presence of multiple zeros; and Burr et al. (2009) provides one of the first non-probabilistic adaptive analyses of an evaluation-based real root isolation algorithm. These last two papers address the 1-D analog of the Plantinga & Vegter Algorithm. The philosophy behind all of these papers is the design and analysis of complete numerical methods based on approximations, iteration and adaptive methods. Topological exactness is achieved using suitable algebraic bounds, ranging from classical root separation bounds to evaluation bounds and geometric separation bounds. We stress that the worst-case complexity of adaptive algorithms ought not to be the chief criterion for evaluating the usefulness of these algorithms: for the majority of inputs, these algorithms terminate fast. For some experimental results for these types of algorithms see Lin and Yap (2009), Kamath (2010), Plantinga (2006) and Plantinga and Vegter (2004). Zero bounds are only used as stopping criteria for iteration in the algorithms, and simple estimates for them can be computed easily. Computing such bounds does not mean that we compute resultants, even though their justification depends on resultant theory. The present paper continues this line of investigation.

The recent collection (Boissonnat et al., 2006, Chapter 5) reviews the current algorithmic literature in meshing in 2- and 3-D: the subdivision approach is represented by the Plantinga & Vegter Algorithm as well as by Snyder's earlier approach based on parametrizability (Snyder, 1992b,a). The subdivision algorithm of Plantinga & Vegter is remarkable in the following sense: even though it is globally isotopic, it does not guarantee isotopy of the curve within each cell of the subdivision. In contrast, Snyder's subdivision approach (Snyder, 1992b,a) requires the correct isotopy type in each cell. Indeed, because of this, the algorithm is incomplete (Boissonnat et al., 2006, p. 195).

Among geometric approaches to meshing, we have the point sampling approach as represented by Boissonnat and Oudot (2006) and Cheng et al. (2004), the Morse theory approach as represented by Stander and Hart (1997) and Boissonnat et al. (2004) and the sweepline approach (Mourrain and Técourt, 2005). Note that the sweepline approach naturally corresponds to the algebraic operation of projection; therefore, its implementation is often purely algebraic. The idea of the sampling approach is to reduce meshing of a surface S to computing the Delaunay triangulation of a sufficiently dense set of sample points on S (Boissonnat et al., 2006, p. 201–213). To obtain such sample points, Cheng et al. (2004) need a primitive operation that amounts to solving a system of equations involving f and its derivatives. Boissonnat and Oudot (2006) need a primitive for intersecting the surface with a Voronoi edge. These sample points are algebraic, and implementing the primitives exactly would require strong algebraic techniques. But exact implementation does not seem to be justified for these applications, and so we are faced with an implementation gap that shows well-known non-robustness issues. For restrictions and open problems in sampling approaches, see Boissonnat et al. (2006, p. 227–229). In contrast, the computational primitives needed by subdivision approaches work directly with bigfloats, with modest requirements on f.

This paper presents a purely numerical subdivision method for meshing algebraic curves with isolated singularities. In a certain sense, this is the most general geometric situation because, by Proposition 1, reduced algebraic curves have only isolated singularities. Our starting point is the

[3] Their paper is subtitled "Exploiting a little more Geometry and a little less Algebra" which speaks to our concerns with (A1).

algorithm of Plantinga & Vegter (Plantinga and Vegter, 2004; Plantinga, 2006) for implicit meshing of curves. It is important to understand the computational model of Plantinga & Vegter which is also used in this paper. Two capabilities are assumed with regard to $f(X, Y)$:

- (i) Sign evaluation of $f(p)$ at dyadic points p.
- (ii) f is C^1 and we can evaluate the interval analogs (i.e., box functions) of f, $\frac{\partial f}{\partial X}$, $\frac{\partial f}{\partial Y}$ on dyadic intervals.

Note that the Marching Cube algorithm only requires capability (i). Let the **class PV** denote the set of all real functions $f : \mathbb{R}^2 \to \mathbb{R}$ for which capabilities (i) and (ii) are available. Many common functions of analysis belong to PV. In addition, the class of elementary functions for which the zero problem is decidable (see Yap (2007)) is also in PV by Du and Yap (2006). Thus, the approach of Plantinga & Vegter admits a more general setting than algebraic curves.

1.1. Overview of paper

- Section 2 establishes some basic terminology and recalls facts about the singularities of algebraic sets.
- In Sections 3 and 4, we extend the Plantinga & Vegter Algorithm to compute an isotopic approximation of the curve $S = f^{-1}(0)$ restricted to a "nice region" that need not be simply connected. S may have singularities outside the region of interest and we only need $f \in PV$.
- In Section 5, we provide the algebraic evaluation bounds necessary for meshing singular curves.
- In Section 6, we provide a subdivision method to isolate all the singularities of a square-free integer polynomial $f(X, Y)$.
- In Section 7, given a box containing an isolated singularity p, we provide a method to compute the branching degree of p.
- In Section 8, we finally present the overall algorithm to compute the isotopic polygonal approximation.
- We conclude in Section 9.

2. Basic terminology and algebraic facts

Let $\mathbb{F} := \mathbb{Z}\left[\frac{1}{2}\right] = \{m2^n : m, n \in \mathbb{Z}\}$ be the set of **dyadic numbers**. All of our numerical computations are performed in a straightforward manner using $\mathbb{F}$. There are many well-known implementations of arithmetic on such numbers, and, in this case, they are known as **bigfloats**. In short, our computational model is not based on some abstract capability whose implementation may reveal gaps that lead to well-known non-robustness issues.

For $S \subseteq \mathbb{R}$, let $\square S$ be the set of closed intervals $[a, b]$ with endpoints in S, i.e., $a, b \in S$. We write $\square S^n$ for $(\square S)^n$. In particular, $\square\mathbb{F}$ is the set of dyadic intervals, and $\square\mathbb{R}^n$ is the set of n-boxes. The **width** of an interval $I = [a, b]$ is $w(I) := b - a$. The width and diameter, respectively, of an n-box $B = \prod_{i=1}^{n}[a_i, b_i]$ is $w(B) := \min\{b_i - a_i : i = 1, \ldots, n\}$ and $d(B) := \max\{b_i - a_i : i = 1, \ldots, n\}$. The boundary of a set $S \subseteq \mathbb{R}$ is denoted ∂S. If $f : \mathbb{R}^n \to \mathbb{R}$ and $S \subseteq \mathbb{R}$, then $f(S) := \{f(x) : x \in S\}$. A function $\square f : \square\mathbb{F}^n \to \square\mathbb{F}$ is a **box function** for f provided (i) $f(B) \subseteq \square f(B)$ and (ii) if $B_0 \supseteq B_1 \supseteq \cdots$ with $\lim_i B_i = p$ then $\lim_i \square f(B_i) = f(p)$. We regard the limit of intervals in terms of the limits of their endpoints. We say $f \in PV$ if (1) $f \in C^1$ (has continuous first derivatives), (2) there is an algorithm to determine $\mathtt{sign}(f(p))$ for $p \in \mathbb{F}^n$ and $\square f$ and (3) the corresponding functions for the derivatives of f are computable in $\square\mathbb{F}$. In this paper, we only consider box functions for the two-dimensional case.

We only consider boxes of the form $B = I \times J$ where I, J are dyadic intervals. As in Plantinga & Vegter, all our boxes will be squares, i.e., $w(I) = w(J) = w(B)$. Our algorithms work in the slightly more general setting where all boxes have aspect ratios at most 2 (see Lin and Yap (2009) for a proof). The boundary ∂B of B is divided into four **sides** and four **corners**. Note that the 'sides/corners' terminology for boxes should not be confused with the 'edges/vertices' terminology which we reserve for the straightline graph $G = (V, E)$ which is the approximation to our curve. We **split** a box B by subdividing it into 4 subboxes of equal widths. These subboxes are the **children** of B and each has

width $\frac{1}{2}w(B)$. Starting with B_0, the child–parent relationships obtained by an arbitrary sequence of splits yields a **quadtree** rooted at B_0. Two distinct boxes B, B' of a quadtree are **neighbors** if their boundaries overlap: $B \cap B'$ is a line segment, but not a single point. Segments of the form $B \cap B'$ are called **interior segments**. If B is a box whose side s is part of the boundary of B_0, then we call s a **boundary segment** and B a **boundary box**. Moreover, each side of a box is divided into one or more segments.

Some of the regions of interest in this paper will not be squares and may not even be simply connected. To ensure that our algorithm continues to work in this case, we insist that the region comes from a subdivision, i.e., we insist that there exists a square B_0, a subdivision of B_0 and a collection of boxes in the subdivision such that the region of interest is the union of this collection of boxes. Although it is not necessary for our algorithms, in most implementations, it is easiest to maintain a subdivision of B_0 where each box is labeled as either region or complement. The union of the boxes labeled "region" will be written R_0 (or $R_0(T)$ when we stress the subdivision tree) and is the region of interest. We reserve the notation R_0 for non-square regions of interest and use B_0 for square regions of interest.

Algebraic facts. Let $\mathbb{D}$ be a unique factorization domain (UFD) and $f, g \in \mathbb{D}[\mathbf{X}] = \mathbb{D}[X_1, \ldots, X_n]$ where $\mathbf{X} = (X_1, \ldots, X_n)$. We say f, g are **similar** if there exist $a, b \in \mathbb{D} \setminus \{0\}$ such that $af = bg$, and we write this relationship as $f \sim g$. Otherwise, f and g are **dissimilar**. The **square-free part** of f is defined as

$$\mathtt{SqFree}(f) := \frac{f}{\mathtt{GCD}(f, \partial_{X_1} f, \ldots, \partial_{X_n} f)} \tag{1}$$

where ∂_{X_i} indicates differentiation with respect to X_i. f is said to be **square-free** if $\mathtt{SqFree}(f) = f$. From (1) we see that computing $\mathtt{SqFree}(f)$ from f involves only rational operations of $\mathbb{D}$. As the gradient of f is $\nabla f = (\partial_{X_1} f, \ldots, \partial_{X_n} f)$, we may also write $\mathtt{GCD}(f, \nabla f)$ for $\mathtt{GCD}(f, \partial_{X_1} f, \ldots, \partial_{X_n} f)$. See Yap (2000, Chap. 2) for standard conventions concerning GCD.

Let k be an algebraically closed field. For $S \subseteq k[\mathbf{X}] = k[X_1, \ldots, X_n]$, let $\textsc{Zero}(S) := \{p \in k^n : f(p) = 0 \text{ for all } f \in S\}$ denote the **zero set** of S. A zero set is also known as a **variety**. The **singular points** of $\textsc{Zero}(f)$ are defined, when f is square-free, to be the points where $\nabla \mathtt{SqFree}(f) = 0$.

In one dimension, it is well-known that a square-free polynomial $f \in \mathbb{Z}[X]$ has no singularities (i.e., multiple zeros). We now recall two generalizations of this result that will be necessary in the remainder of the paper. See Hartshorne (1977), Cox et al. (1992) and Harris (1992) for similar results.

Proposition 1 (*Harris, 1992, Ex.14.3*)**.** *The singular points of any variety form a proper subvariety.*

This proposition is critical in our paper, because it implies that if $f \in \mathbb{R}[X, Y]$ is square-free, then the singular points are a proper subvariety of a union of curves and hence must be a finite set of points. Thus, we only need to handle isolated singularities.

Proposition 2 (*Algebraic Sard Lemma Harris, 1992, Prop.14.4*)**.** *Let $f : X \to Y$ be any surjective regular map of varieties defined over a field k of characteristic 0. Then there exists a nonempty open subset $U \subseteq Y$ such that for any smooth point $p \in f^{-1}(U) \cap X_{sm}$ in the inverse image of U, the differential df_p is surjective.*

Here, X_{sm} denotes the set of smooth points of variety X. The open sets refer to the Zariski topology. The condition that the differential df_p is surjective is equivalent to insisting that the Jacobian of f has the same rank as the dimension of Y. The situation that we consider is for a function $f : \mathbb{R}^2 \to \mathbb{R}$ where f is a polynomial. In this case, since the image is one dimensional, the condition that df_p is surjective reduces to the condition that $\nabla f(p) \neq 0$. Every point in $\mathbb{R}^2 = X$ is smooth and $\mathbb{R} \setminus U$ is only a finite set. Hence, there are only a finite number of level sets, parametrized by h, where $\textsc{Zero}(f(X, Y) - h)$ has a singular point.

3. Algorithm of Plantinga & Vegter

First, we recall the Plantinga & Vegter Algorithm: Given $\varepsilon > 0$, a bounded and nonsingular curve $f^{-1}(0)$ for $f : \mathbb{R}^2 \to \mathbb{R}$ and a bounding box $B_0 \in \square\mathbb{F}^2$, they compute a polygonal curve P.

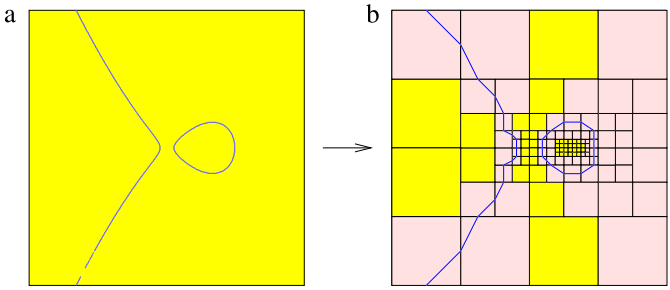


Fig. 1. (a) Original curve $Y^2 - X^2 + X^3 + 0.2 = 0$, (b) isotopic approximation.

The correctness statement is as follows: *if* $f^{-1}(0) \subseteq B_0$, *then* P *is an* ε*-approximation to the curve* $S = f^{-1}(0)$, *i.e.,* $d(P, S) \leq \varepsilon$ *and* $P \approx S$. For simplicity, they focus on topological correctness: $P \approx S$, since it is easy to refine the subdivision to achieve $d(P, S) \leq \varepsilon$. The curves in Fig. 1 are output from our implementation of the Plantinga & Vegter Algorithm as reported in Lin and Yap (2009). The value of ε is small in Fig. 1(a), while $\varepsilon = \infty$ in Fig. 1(b).

The algorithm is based on two simple predicates on boxes B:

- Predicate $C_0(B)$ holds if $0 \notin \square f(B)$.
- Predicate $C_1(B)$ holds if $0 \notin \left(\square \frac{\partial f}{\partial X}(B)\right)^2 + \left(\square \frac{\partial f}{\partial Y}(B)\right)^2$.

These predicates are easily implemented for $f \in PV$. If $C_0(B)$ holds, then the curve S does not intersect B. If $C_1(B)$ holds, then the gradient of f is never zero in B and the gradient vectors approximately point in the same direction. Note that if B satisfies C_1, then, by recalling the parent–child relationship, any child of B also satisfies C_1.

The input box B_0 is a dyadic square and the output will be an undirected graph $G = (V, E)$ where each vertex $v \in V$ is a dyadic point, $v \in \mathbb{F}^2$. In fact, G represents a straightline planar graph that is a polygonal approximation of S.

The algorithm has 3 phases, where Phase i ($i = 1, 2, 3$) is associated with a queue Q_i containing boxes. Initially, $Q_1 = \{B_0\}$, and $Q_2 = Q_3 = \emptyset$. When Q_i is empty, proceed to Phase $i + 1$.

- PHASE 1: SUBDIVISION. While Q_1 is non-empty, remove some B from Q_1, and perform the following: If $C_0(B)$ holds, discard B. If $C_1(B)$ holds, insert B into Q_2. Otherwise, split B into four subboxes and insert them into Q_1.
- PHASE 2: BALANCING. This phase "balances" the subdivision, where a subdivision is **balanced** if the widths of any two neighboring boxes differ by at most a factor of 2. Queue Q_2 is a min-priority queue, where the width of a box serves as its priority. While Q_2 is non-empty, remove the min-element B from Q_2, and perform the following: For each B-neighbor B' in Q_2 with width more than twice the width of B, remove B' from Q_2 and split B'. Insert each child B'' of B' into Q_2 provided $C_0(B'')$ does not hold. B'' might be a new neighbor of B and B'' might be split subsequently. Finally, when every neighbor of B is at most twice the width of B, insert B into Q_3.
- PHASE 3: CONSTRUCTION. This phase constructs the graph $G = (V, E)$. Initially, the boxes in Q_3 are unmarked. While Q_3 is non-empty, remove any B from Q_3 and mark it. Now construct the set $V(B)$ of its vertices. For each B-neighbor B', if B' is marked, retrieve any vertex v on the side $B \cap B'$, and put v into $V(B)$. If B' is unmarked, evaluate the sign of $f(p)f(q)$ where p, q are endpoints of the segment $B \cap B'$. If $f(p)f(q) < 0$, create a vertex $v = (p + q)/2$ for the graph G and put v into $V(B)$. Note that if $f(p) = 0$ for any corner p, treat $f(p)$ as positive; in effect, this is an infinitesimal perturbation at p. It can be shown that $|V(B)| \in \{0, 2, 4\}$. If $|V(B)| = 2$, put the edge (p, q) into G to connect the vertices in $V(B)$. If $|V(B)| = 4$, it can be shown that one side of B contains two vertices a, b. Introduce two edges into E to connect each of a, b to the remaining two vertices. The requirement that these two edges are non-crossing ensures that the connection is unique (see Plantinga and Vegter (2004) and Plantinga (2006)).

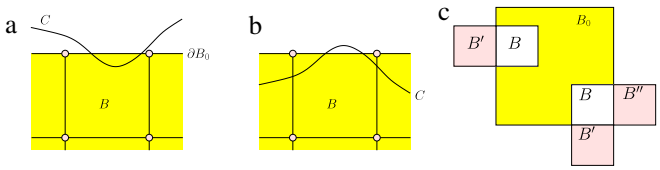


Fig. 2. (a) Incursion, (b) excursion, (c) boundary boxes and their complements.

The output graph $G = (V, E)$ can be viewed as a straightline graph, decomposed into a collection $P = P(G)$ of closed polygons. In the following, we simply use G in place of P as the polygonal approximation.

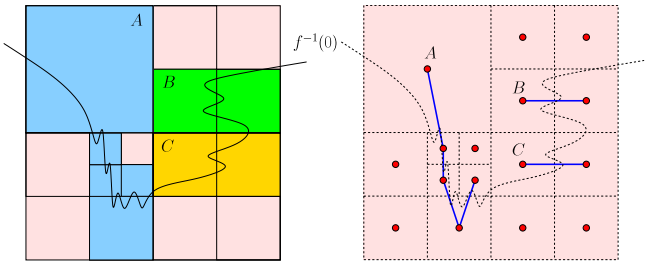
4. Extension of Plantinga & Vegter

4.1. The naïve extension of Plantinga & Vegter

The correctness statement of Plantinga & Vegter requires B_0 to be a bounding box for the curve $f^{-1}(0)$. The power of subdivision methods comes from their ability to adaptively analyze local data. Choosing the initial box to be a bounding box gives up this power of local analysis; also, by insisting that the curve is bounded, the types of possible curves are severely restricted. As a first attempt, one could naïvely attempt to run the Plantinga & Vegter Algorithm starting with an arbitrary box and then ask the question "In what sense is the output G correct?" Intuitively, G should be isotopic to $f^{-1}(0) \cap B_0$, but Plantinga & Vegter did not discuss this issue. The algorithm certainly cannot handle the case when the curve $S = f^{-1}(0)$ has tangential but non-crossing intersections (Yap, 2006) with ∂B_0. If we assume that there are only transversal intersections, we still face two problems: if the curve S (locally) enters and exits ∂B_0 by visiting only one box $B \subseteq B_0$, the above algorithm would fail to detect this small component. Such an error is called an undetected **incursion**, as illustrated in Fig. 2(a). Conversely, the curve S might escape undetected from B_0. Such an error is called an undetected **excursion**, as illustrated in Fig. 2(b). These errors cause the final approximation to be incorrect, since $S \cap B_0$ may not have the same number of components as G. Incursions account for components of $S \cap B_0$ that are undetected by the Plantinga & Vegter Algorithm, and excursions account for connected components of G that are not connected in $S \cap B_0$. A single box may contain several incursions and excursions. The excursion case is more troubling because there is no guarantee that C_1 will hold in the complement of B_0; this means that the connected components of G which are not connected in $S \cap B_0$ might, in fact, not even be connected in S. If we choose B_0 large enough, such errors cannot arise, but this approach gives up the power of adaptivity and localization which subdivision methods possess. If S has singularities, making B_0 large may not be an option. In this paper, we avoid any "largeness" assumption on B_0.

The preceding issues arise because the Plantinga & Vegter Algorithm focuses only on the parity of the endpoints of an edge of the subdivision. They prove that multiple intersections can be removed by applying a suitable isotopy to move small features into a neighboring box. We reproduce their results here in order to examine them. In the following, we fix a balanced quadtree T rooted at a box B_0. Let $V(T)$ denote the collection of boxes at the leaves of T. These boxes constitute a partition of B_0. We also assume that $C_0(B)$ or $C_1(B)$ holds at each $B \in V(T)$. The subdivision constructed through the Plantinga & Vegter Algorithm has this property. The results in this section do not need the tree T to be balanced because this constraint is only used in the construction of the isotopic curve.

In Phase 3 of the Plantinga & Vegter Algorithm, we noted that if $f^{-1}(0)$ passes though a corner of the subdivision, then they treat that point as positive. This can be done via an infinitesimal perturbation of f: we call this $\widetilde{f}$ the **standard perturbation** of f with respect to T. More precisely, $\widetilde{f}$ is defined to be a function that agrees with f everywhere except in infinitesimal neighborhoods of corners of boxes where f is 0; in such neighborhoods, $\widetilde{f}$ is positive at the corner of the box. Clearly, the definition of $\widetilde{f}$ depends on T. In general, $\widetilde{S} = \widetilde{f}^{-1}(0)$ is isotopic to $S = f^{-1}(0)$, but in the case where B_0 is not a bounding box for the curve, it can happen that $\widetilde{S} \cap B_0$ is not isotopic to $S \cap B_0$. This difference is discussed more thoroughly in Section 4.3. All of our correctness statements are about $\widetilde{S}$ and not S.



(a) Quadtree T with curve $f^{-1}(0)$. (b) Dual Monotone Graph H_f.

Fig. 3. A quadtree T and its dual graph induced by $f^{-1}(0)$.

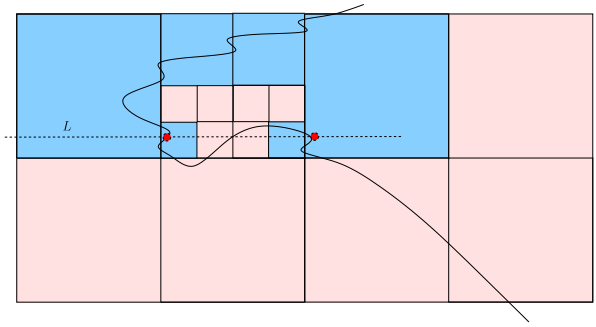


Fig. 4. An X-monotone region that is not globally parametrizable in the Y-direction.

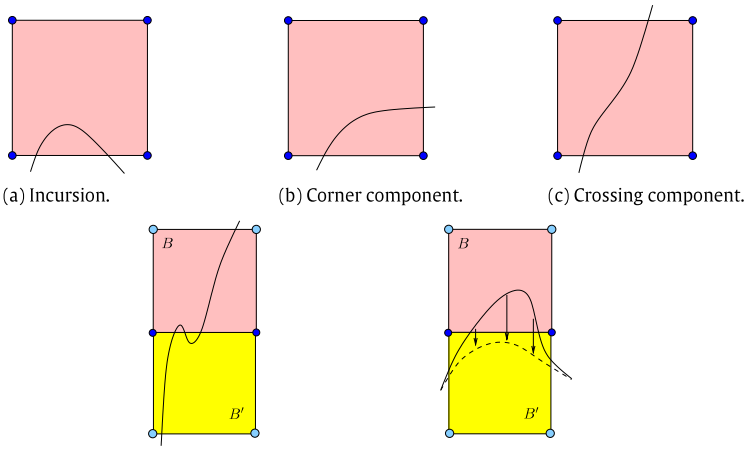
However, for the sake of simplicity in this section, we continue to use the notations S and f instead of $\widetilde{S}$ and $\widetilde{f}$.

Consider the **dual graph** $H = H(T)$ whose vertex set is $V(T)$ and edges connect pairs of neighboring boxes. Thus, $B, B' \in V(T)$ are connected by an edge in H iff $s = B \cap B'$ is a segment, but not a point. We are interested in the subgraph $H_f = H_f(T)$ of H in which B, B' are connected by an edge iff the segment $s = B \cap B'$ intersects the curve $f^{-1}(0)$ more than once. We further focus on the non-trivial connected components of H_f, i.e., those components consisting of more than one box. In Fig. 3, the graph H_f has three non-trivial connected components (A, B, C). The union of all of the boxes corresponding to a non-trivial connected component of H_f is called a **monotone region**, and H_f itself will be known as the **monotone graph**. In Fig. 3(a), we have filled each monotone region (A, B, C) with a distinct tint.

Let us explain the monotone terminology. Each monotone region M can be uniquely classified as a Y**-monotone** or X**-monotone** region in the following sense: we say M is Y**-monotone relative to** f if for all $B, B' \subseteq M$ where (B, B') is an edge of H_f, then $B \cap B'$ is a segment orthogonal to the Y-direction (i.e., s is horizontal). Moreover,

$$0 \notin \frac{\partial f}{\partial Y}(B \cup B'). \tag{2}$$

We call such a rectangular region $B \cup B'$ a **local neighborhood** of the monotone region. Thus (2) implies that the curve $f^{-1}(0)$ restricted to each local neighborhood of a Y-monotone region is parametrizable in the X-direction (i.e., each vertical line intersects the curve at most once). This is a "local property" because the curve $f^{-1}(0)$ might not be parametrizable in the X-direction in the entire Y-monotone region. See Fig. 4 for an example of a X-monotone region in which the curve $f^{-1}(0)$ is not globally parametrizable in the Y-direction: a horizontal line L intersects the curve twice in the monotone region. The curve is parametrizable, however, for connected components of L with respect to the X-monotone region. The definition of a X-monotone regions is similar, after exchanging the roles of X & Y and the roles of horizontal & vertical.



(a) Incursion. (b) Corner component. (c) Crossing component.

(d) Two crossing components. (e) New crossing component.

Fig. 5. Types of components and their isotopic transformation in a local neighborhood.

Let $\mathcal{C}$ be a connected component of $f^{-1}(0) \cap B$. We classify $\mathcal{C}$ into three types: if both endpoints of $\mathcal{C}$ lie on one side of B, then $\mathcal{C}$ is an incursion as discussed above. If the endpoints of $\mathcal{C}$ lie on two adjacent sides of B, it is called a **corner component**. Finally, if the endpoints of $\mathcal{C}$ lie on two opposite sides of B, it is called a **crossing component**. These three kinds of components are illustrated in Fig. 5(a–c). Note that there can be no other types of components in B. In particular, B cannot contain an isolated component since $C_1(B)$ holds. The next lemma also restricts the number of crossing components.

Let us define the notion of isotopy used in this paper. An **isotopy** is a continuous function $I : \mathbb{R}^2 \times [0, 1] \to \mathbb{R}^2$ such that the following two conditions hold: (1) $I(\cdot, 0)$ is the identity and (2) $I(\cdot, t)$ is a homeomorphism for each $t \in [0, 1]$. If, in addition, there is a set $B \subseteq \mathbb{R}^2$ such that for all $t \in [0, 1]$, $I(\cdot, t)$ acts as the identity on the complement of B, i.e., $I(p, t) = p$ for all $p \notin B$, then we call I an **isotopy on** B, and write the isotopy as $I : B \times [0, 1] \to B$. Two sets $S, S' \subseteq \mathbb{R}$ are said to be **isotopic**, denoted $S \approx S'$, if there exists an isotopy I such that $I(S, 1) = S'$. As promised in the introduction, our notion of isotopy is the "ambient" sort (Boissonnat et al., 2006, p. 183). Observe that if I is an isotopy on B, then for any continuous function $f : B \to \mathbb{R}$, $f^{-1}(0)$ and $(f \circ I^{-1}(\cdot, t))^{-1}(0)$ are isotopic. We state without proof the following lemma which is somewhat implicit in Plantinga and Vegter (2004) and Plantinga (2006):

Lemma 3. *Let (B, B') be an edge of the monotone graph H_f. Assume their shared segment $s = B \cap B'$ is horizontal.*

1. *$B \cup B'$ is Y-monotone relative to f*
2. *There is at most one crossing component in each of B and B'. If there is one crossing component in each of B and B' then these two components are part of one single crossing component of $B \cup B'$.*
3. *Let $f^{-1}(0)$ intersect s at two consecutive points p_1 and p_2. Then p_1 and p_2 are connected by a component $X(p_1, p_2)$ of the curve $f^{-1}(0)$ that lies entirely within B or entirely within B'.*
4. *There is an isotopy I on $B \cup B'$ that reduces the number of intersections (counted with multiplicities) on s so that $(f \circ I(\cdot, 1)^{-1})^{-1}(0)$ intersects s in 2 fewer times than $f^{-1}(0)$, and $B \cup B'$ remains Y-monotone relative to $f \circ I(\cdot, t)^{-1}$ for all t.*

Part (2) of this Lemma follows from the fact that $C_1(B \cup B')$ holds. The situation where B and B' both have a crossing component as illustrated in Fig. 5(d). The isotopy I on $B \cup B'$ constructed in part (4) cannot decrease the number of crossing components, but may increase the number by one, as illustrated in Fig. 5(e). The general idea is to repeatedly apply the transformation given by such an isotopy I until the curve $f^{-1}(0)$ composed with the inverses of the isotopies intersects the segment s at most once (see Lin and Yap (2009) for a generalization under a weaker predicate than C_1). The isotopies can easily be chosen to preserve monotonicity by keeping the appropriate coordinate

fixed throughout the isotopy, i.e., the X-coordinate can be kept fixed for an isotopy on a Y-monotone regions. This prepares us for an induction because after each isotopic transformation, monotonicity is preserved. To develop the appropriate global correctness statement, we formulate the sense in which these isotopies interact:

Lemma 4. *Let H be the dual graph to the subdivision given by the tree T, i.e., there is an edge connecting boxes B_1 and B_2 iff they are neighbors. Let H_f be the subgraph of H where B_1 and B_2 are connected by an edge of H and $f^{-1}(0)$ intersects their shared segment more than once. Let D be a connected component of H_f, and R be the union of the boxes appearing in D. There exists an isotopy I on R such that if $f_1 = f \circ I(\cdot, 1)^{-1}$ then $f^{-1}(0) \cap R \approx f_1^{-1}(0) \cap R$ and $f_1^{-1}(0)$ intersects all segments between boxes of D at most once.*

Note, in particular, that boxes which correspond to isolated components of H_f are boxes where $f^{-1}(0)$ intersects each of its sides at most once. The upshot of this argument may be summarized by the following theorem, implicit in Plantinga & Vegter:

Theorem 5. *Let T be the subdivision tree at the end of the Plantinga & Vegter Algorithm. Then there exists an isotopy I on B_0 such that $\widetilde{f}^{-1}(0) \cap B_0 \approx f_1^{-1}(0) \cap B_0$, where $f_1 := \widetilde{f} \circ I(\cdot, 1)^{-1}$. Moreover, $f_1^{-1}(0)$ intersects each interior segment of T at most once.*

A function f_1, as given by the conditions of this theorem, is said to be **normalized** relative to T. It is important to realize that when B_0 is not a bounding box, the normalized curve $f_1^{-1}(0)$ may intersect the boundary segments of T more than once. The reason for this is that we can only apply the transformation of Lemma 3 to interior segments; we cannot apply the transformation to boundary segments without changing the topology of the restriction of the curve to B_0. Therefore, since Plantinga & Vegter focus entirely on the parity across segments, the graph G constructed by the Plantinga & Vegter Algorithm may not be isotopic to $f_1^{-1}(0) \cap B_0$. We note that although the isotopies constructed here might not preserve the C_1 condition since they are only guaranteed to preserve monotonicity, it will be useful to recall that these boxes were derived from ones where C_1 did hold.

One solution for ensuring that G satisfies $G \approx S \cap B_0$ is to ensure that S intersects each boundary segment at most once. A sufficient condition is that either C_0 or C_1 holds on the edge, using the corresponding properties for the one-dimensional version of Plantinga & Vegter's algorithm (cf. Burr et al. (in preparation)). Achieving this sufficiency condition requires that we determine the topology of S on the boundary of B_0, including tangential intersections. This solution is also inefficient because it may require frequent subdivisions near the boundary to resolve fine features; in higher dimensions, this solution must also be recursively applied to lower dimensional boxes. These issues also arise in Snyder's approach (Snyder, 1992b).

4.2. The enlarged region solution

We now provide an alternative solution which is closer to the spirit of the Plantinga & Vegter approach of exploiting isotopy. This solution avoids determining the exact boundary topology as well as making any largeness assumptions on B_0. We wish to slightly enlarge B_0 so that incursions and excursions no longer occur. The basic idea is that, in addition to subdividing B_0, we find a slightly larger region B_0' which is the union of B_0 with a "collar" of squares around B_0. We construct a straightline approximation G^+ that will be isotopic to the curve restricted to some expansion B_0^+ of B_0 into this collar. This is weaker than saying G^+ is isotopic to $f^{-1}(0) \cap B_0$, but it has two major advantages: it allows our algorithm to terminate faster and it does not require $f^{-1}(0)$ to intersect ∂B_0 in any special way.

Call a box $B \subseteq B_0$ a **boundary box** if ∂B intersects ∂B_0. Let B be such a box. If B does not share a corner with B_0, then it has a unique **complementary box** B' such that B' has the same width as B, the interiors of B' and B_0 are disjoint, and $\partial B' \cap \partial B_0 = \partial B \cap \partial B_0$. B and B' are called **partners** of each other. If B shares a corner with B_0, then it determines two complementary boxes B', B''. See Fig. 2(c). We insist that the collar is formed from complementary boxes, all of which satisfy either C_0 or C_1. It is easy to adapt Plantinga & Vegter's algorithm so that this situation occurs. With this property, excursions from B_0 or incursions into B_0 can be limited to the collar region.

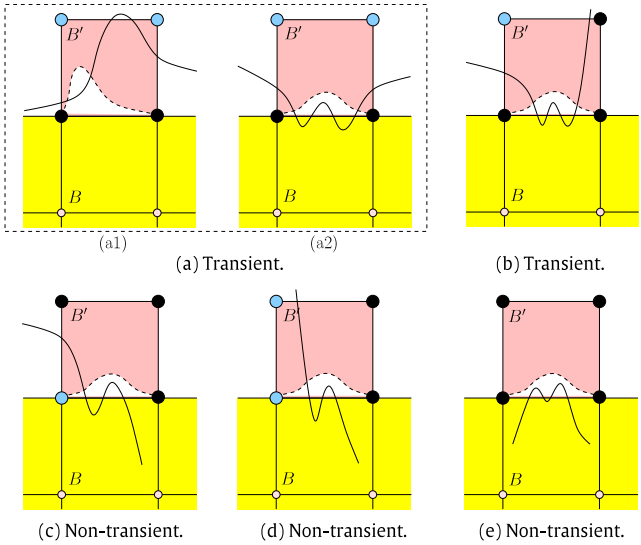


Fig. 6. Classification of complementary boxes according to sign of f at the corners.

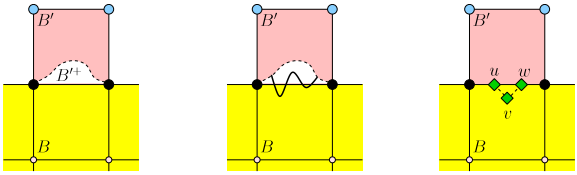
Let B be a boundary box with complementary box B'. If the complementary box satisfies C_0, then the curve does not intersect it. Therefore, there can be no incursions into or excursions from B. If the complementary box satisfies C_1 and there is an excursion, then the isotopy presented in Plantinga & Vegter shows that the approximation constructed by the naïve Plantinga & Vegter Algorithm is correct in $B \cup B'^{+}$, for some region $B \subseteq B'^{+} \subseteq B' \cup B$. The case that is left to consider is when there is an incursion. This case is harder because the isotopy constructed by Plantinga & Vegter would remove the component of the incursion, but that would result in an error. In the spirit of Plantinga & Vegter's algorithm, we consider the sign pattern of the corners of complementary boxes.

Up to symmetry and sign flips, B' can be put into one of five types based solely on the sign of f at the corners of B', (a)–(e). This is illustrated in Fig. 6. Note that the "alternate sign pattern" does not appear because this pattern cannot be C_1 as shown by Plantinga & Vegter.

We draw two instances of types (a) to indicate the two possible "dispositions" of the curve $\widetilde{f}^{-1}(0)$ in B': In (a1) the curve makes no incursion into B, but (a2) represents at least one incursion into B. These two dispositions give rise to distinct isotopy types for the curve $\widetilde{f}^{-1}(0) \cap B$. Similarly, type (b) has two possible dispositions (but we only indicate the case where there is an incursion). Types (c), (d), and (e) do not have analogous dispositions. Because of this difference in dispositions, we further classify types (a) and (b) as **transient**; the other types are called **non-transient**.

We first provide some intuition on this operation: we show how the complementary boxes are used to yield a correctness statement. Suppose B' is a complementary box whose partner is B. Let the straightline graph $G = (V, E)$, when restricted to B, be denoted $G \cap B$. Note that $G \cap B$ has at most two edges. We would like to claim that $\widetilde{f}^{-1}(0) \cap B$ is isotopic to $G \cap B$. This is evidently false for types (a) and (b) because of the two dispositions discussed above; but even for type (c), this claim can be false because $\widetilde{f}^{-1}(0) \cap B$ may have an arbitrary number of components due to incursions/excursions, as illustrated in Fig. 6(c).

To remedy the situation for types (a) and (b), we first expand B into a slightly larger region $B^{+} \subseteq B \cup B'$. The actual expansion $B'^{+} := B^{+} \setminus B$ is represented in Figs. 6 and 7(a, b) by the white region demarcated by a dashed curve and we call B'^{+} a **cap**. Up to isotopy, the incursions into B^{+} in types (a1) and (a2) are equivalent to a single incursion of the curve $\widetilde{f}^{-1}(0)$ into B^{+} from B', as shown in Fig. 7(b). In the graph $G = (V, E)$, we represent this incursion by adding three vertices u, v, w to V, with u and w on the side $B \cap B'$, and with v between u and w, but slightly to the interior of B. We also add the edges (u, v), (v, w) to E. This is illustrated in Fig. 7(c). Let G^{+} denote the **augmented straightline graph** with these additional vertices and edges for transient complementary boxes.



(a) Extending B by a "cap" B^+. (b) $f^{-1}(0) \cap B^+$. (c) Isotopic approximation.

Fig. 7. Representing the transient type: extension of a boundary box by a cap.

Types (a) and (b) are important because they include the situation where $\widetilde{f}^{-1}(0)$ makes a non-crossing tangential intersection with the boundary of B_0. Detecting this situation is expensive, and provides a main motivation for exploiting isotopy. There is a similar expansion B^+ of B into B' for type (c), as indicated by dashed curves in Fig. 6(c). In this case, we do not need to augment the graph $G = (V, E)$ with any new vertices or edges, because G already contains an edge representing this component in B.

We capture the above intuitive explanations as a lemma:

Lemma 6. *Assume that $\widetilde{f}$ is normalized relative to a quadtree T. Let $B \subseteq B_0$ be a boundary box and let B' be its partner (a complementary box) that satisfies C_1 or C_0. Then there exists a region B^+ such that:*
(i) *B^+ is an expansion of B into B':*

$$B \subseteq B^+ \subseteq B \cup B'$$

and the boundary of B^+ consists of two connected parts $\partial_1(B^+)$ and $\partial_2(B^+)$: $\partial_1(B^+)$ is the union of three non-boundary sides of B and $\partial_2(B^+)$ is either the fourth side of B or a curve in the interior of B' (in Fig. 6, it is a dashed curve). In particular $B^+ \cap \partial(B') \subseteq B \cap B'$.
(ii) *The augmented straightline graph $G^+ = (V^+, E^+)$, when restricted to B, is isotopic to $\widetilde{f}^{-1}(0) \cap B^+$, i.e.,*

$$G^+ \cap B \approx \widetilde{f}^{-1}(0) \cap B^+.$$

Moreover, the isotopy I on B^+ that witnesses this graph isotopy can be chosen to be the identity on $\partial_1(B^+)$.

Proof. Note first that even though $\widetilde{f}$ has been normalized, it came from the situation where all boxes of T satisfy C_0 or C_1. This, in turn, implies that the properties of Lemma 3 apply to incursions and excursions and that the isotopies chosen in the normalization process can be chosen to be the identity on these incursions and excursions. Note that $G^+ \cap B$ has between 0 and 3 components. This is because at most two edges can appear in $G \cap B$ according to the Plantinga & Vegter construction rules, and, in types (a) and (b), the augmented graph G^+ has an additional component. We now consider each type in turn, and, in each case, define the expansion B^+ of B. These expansions are illustrated in Fig. 6(a)–(e).

For type (e), only an excursion is possible. If there is an excursion, we can expand B into B^+ to ensure that $\widetilde{f}^{-1}(0)$ never intersects $\partial_2(B^+)$.

For types (c) and (d), the $\widetilde{f}^{-1}(0)$ intersects $B \cap B'$ at least once. If it intersects $B \cap B'$ multiple times, we can expand B into B^+ to ensure that $\widetilde{f}^{-1}(0)$ intersects the boundary $\partial_2(B^+)$ exactly once.

For types (a) and (b), there are two sides of B' where the curve $\widetilde{f}^{-1}(0)$ intersects at least once. Moreover, there is a unique connected component X of $\widetilde{f}^{-1}(0) \cap (B' \cup B)$ that connects these two sides (see Fig. 6(a, b)). Recall that we say X has two possible dispositions: either X intersects the side $B \cap B'$ or it does not. In either case, we can expand B into B^+ so that X intersects $\partial_2(B^+)$ exactly twice. This component $X \cap B^+$ is represented by the augmented edges (u, v), (v, w) in E^+. □

We now present the **Extended Plantinga & Vegter Algorithm**. It has 3 phases that parallel the algorithm in Section 2. Phase i (for $i = 1, 2, 3$) works off queues Q_i and Q_i', transferring boxes into Q_{i+1} and Q_{i+1}'. The queue Q_i' holds complementary boxes while Q_i holds regular boxes. Initially, Q_1 contains B_0 and Q_1' contains the four complementary boxes to B_0.

- PHASE 1: SUBDIVISION. While Q_1 is non-empty, remove some B from Q_1, and perform the following: If $C_0(B)$ holds B is discarded. If $C_1(B)$ holds, and also $C_1(B')$ or $C_0(B')$ holds for every complementary box B' of B, then (a) insert B into Q_2, and (b) for each complementary box B' that

satisfies C_1 but not C_0, insert B' into Q_2'. Otherwise, split B into four subboxes which are inserted into Q_1 and **half-split** every complementary box B' (this means that we split it into four children and consider the two children which intersect ∂B_0 and put the two children into Q_1').

- PHASE 2: BALANCING. The balancing of boxes in Q_2 is done as in Phase 2 of Section 3. Note that boxes are inserted into Q_3 by this process. Next, we perform an analogous while-loop on Q_2': while Q_2' is non-empty, remove any B' from Q_2'. If B' and its partner are different sizes, we then **half-split** B' and put the children of B' into Q_2' provided C_0 does not hold. Otherwise, we place B' into Q_3'.
- PHASE 3: CONSTRUCTION. First, perform Phase 3 of Section 3 which constructs a graph $G = (V, E)$ from the boxes in queue Q_3. Next we augment this graph using queue Q_3': for each $B' \in Q_3'$, if B' is a transient type, we insert three vertices and two edges into the graph G as described above. The resulting straightline graph is denoted $G^+ = (V^+, E^+)$.

4.3. Weak correctness statement

We are ready to prove the correctness of this Extended Plantinga & Vegter Algorithm. We define B_0' to be the union of B_0 with all of the complementary boxes B' that were placed into Q_3'.

Before we formulate the correctness statement, we will discuss the implications of using an infinitesimal perturbation. In most cases the perturbation does not change the topology, e.g., when the curve intersects a corner in the interior of B_0, the topology in B_0 does not change. This is the reason that in Plantinga (2006) and Plantinga and Vegter (2004), the curve $\widetilde{S} = \widetilde{f}^{-1}(0)$ is isotopic to $S = f^{-1}(0)$. In our setting, $\widetilde{S} \cap B_0$ may no longer be isotopic to $S \cap B_0$. For instance, if S either makes a non-crossing tangential intersection with the boundary ∂B_0 at the corner of a box B of T or passes through a corner of B_0, this intersection is an isolated component of $S \cap B_0$; we could lose or enlarge this component in $\widetilde{S} \cap B_0$, depending on the sign of f nearby. To correctly analyze this case would require a more delicate consideration of the boundary, and, in particular, how complementary boxes of neighbors interact. Our correctness statement is therefore about $\widetilde{S} \cap B_0$, and not about $S \cap B_0$.

We regard the use of $\widetilde{f}$ as a reasonable compromise because (1) $\widetilde{f}$ is an infinitesimal perturbation of f; (2) $\widetilde{f}$ is easy to implement and is an effective perturbation (comparing favorably to the common alternative of, say, a randomized perturbation); (3) we know exactly when $\widetilde{S} \cap B_0$ deviates from $S \cap B_0$ (i.e., we encounter a zero at a box corner in the boundary of B_0); (4) it is a very simple solution to what would otherwise be serious complications arising from various degeneracies (e.g., when S contains a horizontal or vertical component); and finally, (5) its use is consistent with the exploitation of isotopy in Plantinga & Vegter.

Theorem 7 (*Weak Correctness*). *Let the curve $S = f^{-1}(0)$ be non-singular in the box B_0, and $G^+ = (V^+, E^+)$ be the augmented straightline graph constructed by the Extended Plantinga & Vegter Algorithm. Then there exists a region B_0^+ isotopic to B_0,*

$$B_0 \subseteq B_0^+ \subseteq B_0',$$

such that

$$G^+ \approx \widetilde{f}^{-1}(0) \cap B_0^+.$$

Proof. Let T be the quadtree at the end of the Extended Plantinga & Vegter Algorithm and $\widetilde{f}$ be the standard perturbation of f relative to T. Using Theorem 5, we have

$$\widetilde{f}^{-1}(0) \cap B_0 \approx f_1^{-1}(0) \cap B_0, \tag{3}$$

where f_1 is a normalization of $\widetilde{f}$ relative to T. Let B_0^+ be the union of B_0 with all of the caps constructed using Lemma 6. Since the isotopy constructed in Theorem 5 acts as the identity outside of B_0, it follows from (3) that:

$$\widetilde{f}^{-1}(0) \cap B_0^+ \approx f_1^{-1}(0) \cap B_0^+. \tag{4}$$

By a direct consequence of Theorem 5, we have $G^+ \cap B \approx f_1^{-1}(0) \cap B$ for each interior box B of T. For each boundary box B of T, Lemma 6 shows that $G^+ \cap B \approx f_1^{-1}(0) \cap B^+$, where B^+ is the union of B and its cap.

Now the composition of all the isotopies involved gives the isotopy $G^+ \approx f_1^{-1}(0) \cap B_0^+$, which, when combined with (4), yields the desired result. □

REMARKS:

1. Under the conditions of this theorem, we call G^+ a **weak isotopic approximation** to $f^{-1}(0) \cap B_0$. In fact, once there is a collar around B_0 where in each complementary box either C_0 or C_1 holds, there is a spectrum of similar algorithms that allow for various correctness statements. This extension algorithm lies at one end of this spectrum, while the naïve extension, which requires $\widetilde{f}^{-1}(0)$ to intersect each boundary segment at most once lies at the other end. The algorithms along this spectrum reflect a trade-off between using isotopy on the boundary and topological correctness within B_0. A slightly different approach which is based on an algorithm that lies between these two extremes was used in our ISSAC 2008 Proceedings version; that solution is in the middle of this spectrum: there, we required $f^{-1}(0)$ to only intersect ∂B_0 transversally. Then we can explicitly distinguish between the two dispositions in the transient boxes (types (a) or (b)). To do this, we require an iteration at each transient box. This application is a main reason why these boxes are called transient; after a finite number of additional subdivisions, the boxes will become non-transient. Finally, the algorithm in ISSAC 2008 augments G with an edge if and only if this iteration detects an incursion. We prefer the current solution because it is non-trivial to ensure that $f^{-1}(0)$ only intersects ∂B_0 transversally and because it uses isotopy as much as possible.
2. We can make the collar $B_0' \setminus B_0$ around B_0 as narrow as desired: to ensure a perturbation bound of $\varepsilon > 0$, it is sufficient that each of the complementary boxes where C_0 does not hold has width at most $\varepsilon/4$. This is easily done by modifying the balancing phase of the algorithm.
3. We have assumed that complementary boxes have the same width as their partner. An alternative, possibly more efficient, approach is to allow complementary boxes to have widths less than their partners. Call these "subcomplementary boxes". A boundary box B can have many subcomplementary boxes. We can do half-splits of subcomplementary boxes as long as their widths are greater than ε. However, to ensure topological correctness, it becomes necessary to insist that a box has only a limited number of certain types of "subcomplementary boxes". For instance, a boundary box B may be restricted to have at most one subcomplementary box of types (c) or (d), and this occurs under strict conditions (we leave the details to the reader).

4.4. Extension to non-simply connected regions

It is essential in our applications later to extend the above refinements to non-simply connected regions. Recall that we only consider regions R_0 which come from subdivisions. To extend Theorem 7 to these regions, we note two simple modifications:

(I) A complementary box B' of a boundary box $B \subseteq R_0$ may intersect the interior of R_0 or other complementary boxes. Thus, Phase 1 must split such boundary boxes B sufficiently. Such interference checks can be checked during the subdivision phase.

(II) The region R_0 can have concave corners. A complementary box B' at a concave corner has two partners $B \subseteq R_0$. Relative to each partner B, we classify B' into one of 5 types as in Fig. 6(a)–(e). This is illustrated in Fig. 8(i)–(iv). In Fig. 8(i), for instance, the box B' is Type (a) (hence transient) relative to the indicated box B, but it is type (d) (hence non-transient) relative to the other partner $\widehat{B}$. In Fig. 8(ii), the corner complementary box B' is type (b) (hence transient) relative to both B and $\widehat{B}$. Similarly, the other two cases seen in Fig. 8(iii)–(iv) have dual classifications. We modify our augmentation of the graph $G = (V, E)$ as follows: for each complementary box at a concave corner, we consider its classification relative to each choice of partner B: if the classification is transient, as before, we add three vertices u, v, w and edges (u, v), (v, w) to G on the side of $B' \cap B$.

5. Evaluation bounds

For any function f, define its **evaluation bound** to be

$$\mathrm{EV}(f) := \inf\{|f(p)| \ : \ f(p) \neq 0, \nabla f(p) = 0\}.$$

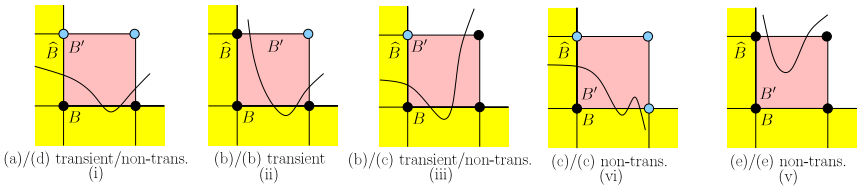


Fig. 8. Classification of complementary boxes at concave corners.

Such bounds were used in Cheng et al. (2009) and Burr et al. (2009). From Proposition 2, we see that $\{f(p) : \nabla f(p) = 0\}$ is a finite set and therefore $\mathrm{EV}(f) > 0$. However there is no explicit bound readily available. The main objective of this section is to provide such a bound:

Theorem 8. *If $f \in \mathbb{Z}[X, Y]$ has degree d and $\|f\| < 2^L$ then $-\lg \mathrm{EV}(f) = O(d^2(L + d \log d))$. More precisely,*

$$\mathrm{EV}(f)^{-1} \leq \max\left\{[d^6 2^{L+2d+11}]^{d^2-1}, [d^{3d+8} 2^{3L+5d}]^d\right\}.$$

Before giving the proof, we provide some definitions and preparation. Let $h = \sum_{i+j=0}^{d} a_{ij} X^i Y^j \in \mathbb{C}[X, Y]$. $\|h\|_k := \sqrt[k]{\sum_{i+j=0}^{d} |a_{ij}|^k}$ will denote the k-norm of h, where we use $k = 1, 2$. We just write $\|h\|$ for $\|h\|_\infty := \max |a_{ij}|$, denoting the height of h.

Now consider f, as in the statement of the theorem, as a function on $\mathbb{C}^2$. As input parameters in our bounds, we use d and L where $\deg f \leq d$ and $\|f\| < 2^L$. Let f_x, f_y denote the derivatives of f. We may write

$$\textsc{Zero}(f_x, f_y) = \bigcup_i U_i \cup \bigcup_j V_j$$

where U_i are the 1-dimensional irreducible components and V_j are the 0-dimensional irreducible components. On each component U_i, one can show that the function f is constant. E.g. $f = (xy+1)^2 - 1$, $f_x = 2(xy + 1)y$ and $f_y = 2(xy + 1)x$. Then $U_1 = \{xy + 1 = 0\}$ and $V_1 = \{(0, 0)\}$. The function f is equal to -1 on U_1 and 0 on V_1.

Let $g := \mathtt{GCD}(f_x, f_y)$ and also

$$\overline{f_x} := f_x/g, \qquad \overline{f_y} := f_y/g.$$

Clearly, we have

$$\textsc{Zero}(f_x, f_y) = \textsc{Zero}(g) \cup \textsc{Zero}(\overline{f_x}, \overline{f_y}).$$

Since $\mathtt{GCD}(\overline{f_x}, \overline{f_y}) = 1$, we conclude that $\textsc{Zero}(\overline{f_x}, \overline{f_y})$ has no 1-dimensional components. Conversely, the hyper-surface $\textsc{Zero}(g)$ has no 0-dimensional components as a subvariety of $\mathbb{C}^2$. This proves:

Lemma 9.

$$\textsc{Zero}(g) = \bigcup_i U_i, \qquad \textsc{Zero}(\overline{f_x}, \overline{f_y}) = \bigcup_j V_j.$$

We now view the ring $\mathbb{Z}[X, Y] \simeq \mathbb{Z}[Y][X] \simeq \mathbb{Z}[X][Y]$ in three alternative ways: a bivariate polynomial f in X and Y can be written as $f = f(X, Y)$, $f = f(Y; X)$ or $f = f(X; Y)$ to indicate these three views respectively. As a member of $\mathbb{Z}[X, Y]$, the coefficients of $f(X, Y)$ are elements of $\mathbb{Z}$. But $f = f(Y; X)$ is a member of $\mathbb{Z}[Y][X]$ whose coefficients are elements of $\mathbb{Z}[Y]$. The leading coefficient and degree of f are likewise affected by these views: $lc(f(Y; X)) \in \mathbb{Z}[Y]$ but $lc(f(X, Y)) \in \mathbb{Z}$, $d = \deg(f(X, Y))$ is the largest total degree of f while $\deg(f(Y; X))$ is the largest power of X occurring in f.

We use Mahler's basic inequality (Yap, 2000, p. 351) that if $p \in \mathbb{Z}[X, Y]$ and $p|q$ then

$$\|p(X, Y)\|_1 \leq 2^D \|q(X, Y)\|_1 \tag{5}$$

where $D = \deg(q(X; Y)) + \deg(q(Y; X))$. This implies:

$$\|g(X, Y)\|_1 \leq 4^{d-1} d^3 2^L, \qquad \|\overline{f_x}(X, Y)\|_1 \leq 4^{d-1} d^3 2^L. \tag{6}$$

since $g|f_x, \overline{f_x}|f_x$, and $\|f_x\|_1 \leq d^2 \|f_x\| \leq d^2 \cdot d2^L$, $\deg(f_x(X; Y)) + \deg(f_x(Y; X)) \leq 2d - 2$. The bound then follows from (5).

Let $h(X)$ be the leading coefficient of $g(X; Y)$. Since $h(X)$ has degree $\leq d - 1$, there is an integer $x_0 \in \{0, 1, \ldots, d - 1\}$ such that $h(x_0) \neq 0$. Intersect $\textsc{Zero}(g)$ with the line $X = x_0$. We claim that this line cuts each non-vertical component U_i in a finite but non-zero number of points. For proof, let $g = \prod_i g_i^j$ where $\textsc{Zero}(g_i) = U_i$. Setting $d_i := \deg g_i(Y; X)$, we see that the vertical components correspond to $d_i = 0$. Then $lc(g(X; Y)) = \prod_i lc(g_i(X; Y))$ and $h(x_0) = lc(g(x_0; Y)) \neq 0$ iff for all i, $lc(g_i(x_0; Y)) \neq 0$. But in this case, $g_i(x_0; Y)$ is a nonzero polynomial of degree d_i in $\mathbb{Z}[Y]$, and, therefore, has exactly d_i solutions in $\mathbb{C}$.

Write $f_0(Y) := f(x_0, Y)$ and $g_0(Y) = g(x_0, Y)$. From (6):

$$\|g_0\|_1 \leq d^d \|g(X, Y)\|_1 \leq 4^{d-1} d^{d+3} 2^L.$$

It is also easy to see that

$$\|f_0\| \leq d^{d+1} 2^{L+1}.$$

Suppose $\beta \in \textsc{Zero}(g_0) \setminus \textsc{Zero}(f_0)$. We want a lower bound on $|f_0(\beta)|$. For this purpose, we use an evaluation bound from Burr et al. (2009, Theorem 13(b)):

Proposition 10 (*Evaluation Bound Burr et al., 2009*)**.** *Let $\phi(x), \eta(x) \in \mathbb{C}[x]$ be complex polynomials of degrees m and n. Let $\beta_1, \ldots, \beta_n$ be all the zeros of $\eta(x)$. Suppose there exists relatively prime $F, H \in \mathbb{Z}[x]$ such that $F = \phi\widetilde{\phi}$, $H = \eta\widetilde{\eta}$ for some $\widetilde{\phi}, \widetilde{\eta} \in \mathbb{C}[x]$. If the degrees of $\widetilde{\phi}$ and $\widetilde{\eta}$ are $\widetilde{m}$ and $\widetilde{n}$, then*

$$\prod_{i=1}^{n} |\phi(\beta_i)| \geq \frac{1}{lc(\widetilde{\eta})^m \cdot ((m+1)\|\phi\|)^{\widetilde{n}} M(\widetilde{\eta})^m \cdot \left((\widetilde{m}+1)\|\widetilde{\phi}\|\right)^{n+\widetilde{n}} M(H)^{\widetilde{m}}}.$$

Here the Mahler measure $M(h)$ of a polynomial $h \in \mathbb{C}[x]$ with zeros $\alpha_1, \ldots, \alpha_n$, is defined as $M(h) := |lc(h)| \prod_{|\alpha_i| \geq 1} |\alpha_i|$. We shall choose the polynomials in Proposition 10 as follows:

$$\phi := f_0, \qquad H := \frac{g_0}{\mathrm{GCD}(f_0, g_0)}.$$

Moreover, let $\widetilde{\phi} := 1$, $\eta(Y) := Y - \beta$ and $\widetilde{\eta} := H/\eta \in \mathbb{C}[x]$. Hence

$$m \leq d, \quad n = 1, \quad \widetilde{m} = 0, \quad \widetilde{n} \leq d - 1.$$

Also

$$lc(\widetilde{\eta}) = lc(H) = lc(g_0) \leq \|g_0\| \leq \|g_0\|_1.$$

Further,

$$M(\widetilde{\eta}) \leq M(H) \leq \|H\|_1 \leq 2^d \cdot \|g_0\|_1.$$

Finally, an application of Proposition 10 gives

$$\begin{aligned} |f_0(\beta)|^{-1} &\leq lc(\widetilde{\eta})^d \cdot ((d+1)\|f_0\|)^{d-1} \cdot M(\widetilde{\eta})^d \\ &< [lc(\widetilde{\eta}) \cdot d\|f_0\| \cdot M(\widetilde{\eta})]^d \qquad (\text{as } (d+1)^{d-1} \leq d^d \text{ for } d \geq 2) \\ &\leq \left[\|g_0\|_1 \cdot dd^{d+1} 2^L \cdot 2^d \|g_0\|_1\right]^d \\ &\leq \left[d^{3d+8} 2^{3L+5d}\right]^d. \end{aligned} \tag{7}$$

(7) is a lower bound on $|f(p)|$ where p lies in a non-vertical component U_i. By considering $g(Y;X)$, the same bound applies for $|f(p)|$ when p lies in a vertical component U_i.

Now, we obtain a lower bound for $f(p)$ with $p \in \textsc{Zero}(\overline{f_x}, \overline{f_y})$. Consider the system $\Sigma \subseteq \mathbb{Z}[X,Y,Z]$ where

$$\Sigma = \{Z - f(X,Y), \overline{f_x}(X,Y), \overline{f_y}(X,Y)\}.$$

The zeros $(\xi_1,\xi_2,\xi_3) = (x,y,f(x,y)) \in \mathbb{C}^3$ of Σ satisfy $\xi_3 = f(\xi_1,\xi_2)$. Since Σ is a zero-dimensional system, we may apply the multivariate zero bound in Yap (2000, p. 350). This bound says that

$$|\xi_3|^{-1} < (2^{3/2}NK)^D 2^{8(d-1)}$$

where $N = \binom{1+2(d-1)}{3}$, $D = d^2 - 1$ and

$$K = \max\{\sqrt{3}, \|\overline{f_x}\|_2, \|\overline{f_y}\|_2, \|Z - f(X,Y)\|_2\}.$$

We have $\|Z - f(X,Y)\|_2 \le (d+1)2^{L+1}$. From (6), we see that $K \le 4^d d^3 2^L$. Using the bound $N < 2d^3$, we obtain

$$|\xi_3|^{-1} < [2^2 \cdot 2d^3 \cdot 4^d d^3 2^L]^{d^2-1} \cdot 2^{8(d^2-1)} < [d^6 2^{L+2d+11}]^{d^2-1}. \tag{8}$$

Now Theorem 8 easily follows from (7) and (8).

6. Isolating singular points

In the remainder of this paper, we assume that $f \in \mathbb{Z}[X,Y]$ and allow the curve $S = f^{-1}(0)$ to have singular points, except on the boundary ∂B_0. We would like to use the Extended Plantinga & Vegter Algorithm to compute an isotopic approximation to $\textsc{Zero}(f)$ when f has only isolated singularities. Since the Plantinga & Vegter Algorithm does not terminate near singular points, it is necessary to first isolate the singular points from the rest of B_0.

We use the auxiliary function $F = f^2 + f_X^2 + f_Y^2$. Finding the singular points of $f^{-1}(0)$ amounts to locating and isolating the zeros of this non-negative function. We use a simple mountain pass theorem (Jabri, 2003) adapted to B_0 to ensure our algorithm isolates the zeros.

Theorem 11. *Suppose that $F \ge 0$ on B_0, and that $F > 0$ on ∂B_0. Then for any two distinct roots p, q of F in B_0, there exists a continuous path $\gamma : [0,1] \to B_0$ connecting p and q which satisfies the following:*

- *γ minimizes $M_\gamma := \max_{x\in[0,1]} F(\gamma(x))$ among all paths connecting p and q in B_0.*
- *γ contains a point y such that either $\nabla F(\gamma(y)) = 0$ or $\gamma(y) \in \partial B$.*

This can be proved using path deformation and the compactness of B_0, or it can be seen as a simple application of the topological mountain pass theorem presented in Jabri (2003). Because of this theorem, distinct zeros of F within B_0 are separated by barriers of height $\epsilon = \min(\mathrm{EV}(F), \min F(\partial B_0))$. This leads us to the following multistep process to localize these zeros. The goal is to find a small rectangle with diameter less than some δ around each zero.

- PHASE 0: DETERMINING ϵ. Initialize ϵ to any lower bound on $\mathrm{EV}(F)$. Also, initialize Q_0 to be $\{B_0\}$ and Q_1 to be empty. Q_0 will contain boxes S that intersect ∂B_0 and have $0 \in \square F(S)$; these boxes will then be subdivided in order to compute a lower bound on $F(\partial B_0)$. While Q_0 is non-empty, remove a square S from it and evaluate $\square F(S)$. If $\square F(S) > 0$ we push S into queue Q_1 and also update ϵ to $\min\{\epsilon, \min \square F(S)\}$. If $0 \in \square F(S)$, subdivide S and push the children of S which intersect ∂B_0 into Q_0, and the others into Q_1. When Q_0 is empty, we stop and fix the current value of ϵ for the remainder of the algorithm.
- PHASE 1: INITIAL SUBDIVISION. Initialize queue Q_2 to be empty.

  While there is an S in Q_1, remove it and evaluate $\square F(S)$. If $\square F(S) > \epsilon/2$, then discard S. Else if $\square F(S) < \epsilon$, push S into Q_2. Else subdivide S and push its children into Q_1.

  Once Q_1 is empty, group the elements of Q_2 into connected regions A_i ($i \in I$). Each A_i contains at most one root, since otherwise, there would be a path connecting the roots within A_i. The value

of F along this path would be less than ϵ, contradicting the mountain pass theorem. Let C be the region $B_0 \setminus \cup_i A_i$. Note by Step 0 that F is greater than $\epsilon/2$ on C and that $\partial B_0 \subseteq C$, i.e., each A_i doesn't intersect ∂B_0.

- PHASE 2: REFINEMENT. For each A_i ($i \in I$), initialize queue $Q_{2,i}$ with all squares $S \in A_i$. So long as neither terminating condition 1 nor 2 (below) holds, we perform the following: for each S in $Q_{2,i}$, if $0 \in \square F(S)$, subdivide S and push its children into $Q_{2,i}$. If $0 \notin \square F(S)$, discard S. We terminate when either of the following two conditions are met:

  1. $Q_{2,i}$ is empty, in which case there isn't a zero in A_i.
  2. A_i', the contents of $Q_{2,i}$ satisfy all of the following:
     (a) $\square F(S) < \epsilon/2$ for some $S \in A_i'$
     (b) R_i, the smallest rectangle containing A_i', lies within the region covered by the original A_i.
     (c) The diameter of R_i is less than δ.

It is clear from the definition of F that this step will halt. We claim that each R_i contains exactly one root. In Step 1, we showed that A_i contains at most one root. To see that R_i contains a root, take a point of A_i' where $\epsilon/2$, then follow the path of steepest descent to reach a zero of F. Because F is less than $\epsilon/2$ on this curve, the curve cannot pass through the region C to reach any other R_j or to leave B_0. Therefore there must be a zero within A_i. It is in R_i because our conditions ensure that F is positive on $A_i \setminus R_i$.

7. Determining the degree of singular points

The following standard result from Krantz and Parks (1992) and Łojasiewicz (1991) describes the global structure of zero sets:

Proposition 12 (*Zero Structure*). *Let f be a real analytic function. Then* Zero(f) *can be decomposed into a finite union of pieces homeomorphic to* $(0, 1)$*, pieces homeomorphic to* S^1*, and singular points.*

In our current situation, the pieces which are homeomorphic to $(0, 1)$ are smooth open subsets of the irreducible components of Zero(f).

Viewing Zero(f) as a multigraph H, the branching **degree** of a singular point is its degree as a vertex of H. We now determine such degrees. Let δ_3 be a separation bound between singular points, so if p and q are two distinct singular points of Zero(f), then the distance between p and q is at least δ_3. Let δ_4 be a separation bound so that if r is a point on Zero(f) such that $\nabla f(r)$ is parallel to the line between r and a singular point p, then the distance between p and r is at least δ_4. If s is on Zero(f) so that the distance between s and a singular point p is smaller than either δ_3 or δ_4, then by following the paths Zero(f) away from s, one of the paths strictly monotonically approaches p until it reaches p and the other path locally strictly monotonically recedes from p. Yap (2006, Cor. 11) provided an explicit bound on δ_3 as a function of degree and height of $f(X, Y)$:

$$\delta_3 \geq \min\left\{(16^{d+2}256^L 81^{2d} d^5)^{-d}, (2^{8L+21}3^{8d})^{-2}\right\}.$$

7.1. Lower bound on δ_4

To derive an explicit bound on δ_4, we consider the following 6 polynomials in $\mathbb{Z}[z, p_x, p_y, q_x, q_y]$:

$$I_f := \{f(p), f(q), f_x(q), f_y(q), (p-q)_x f_y(p) - (p-q)_y f_x(p), z^2 - \|p-q\|^2\} \tag{9}$$

where $p = (p_x, p_y)$ and $q = (q_x, q_y)$ are points on Zero(f). q is a singular point of Zero(f). Moreover, define "$(p-q)_x$" to mean $p_x - q_x$, so that the equation $(p-q)_x f_y(p) - (p-q)_y f_x(p) = 0$ implies that $(p-q)$ is parallel to $\nabla f(p)$. The equation $z^2 - \|p-q\| = 0$ implies that z is the distance between p and q.

Consider the projection Π_z[Zero(I_f)] of the zeros of I_f onto the z-coordinate. Then δ_4 is $\inf\{|z| : z \in \Pi_z[\text{Zero}(I_f)], z \neq 0\}$. We obtain a lower bound on δ_4 using the following general theorem from Brownawell and Yap (2009):

Proposition 13. *Let $I := (P_1, , \ldots, , P_m) \subseteq A := \mathbb{Z}[X_1, \ldots, X_n]$. Let $\mathfrak{P}$ be an isolated prime component of I whose projection onto the first coordinate, $\Pi_1(\textsc{Zero}(\mathfrak{P}))$, is a finite set. If $\overline{\zeta} = (\zeta_1, , \ldots, , \zeta_n) \in \textsc{Zero}(\mathfrak{P})$ and $\zeta_1 \neq 0$, then*

$$|\zeta_1| \geq ((n+1)^2 e^{n+2})^{-n(n+1)D^{n-k}} (k^{n-k-1} mH)^{-(n-k)D^{n-k-1}},$$

where

- $\dim \mathfrak{P} = k$,
- $H \geq \text{Height}(P_i)$, *and*
- $D \geq \deg(P_i)$, $i = 1, , \ldots, , m$.

If $\Pi_z[\textsc{Zero}(I_f)]$ is a finite set, then we use the bounds $n = 5$, $m = 6$, $k \leq 4$, $D = \max\{2, d\}$, and $H \leq d2^{L+1}$ in applying this theorem to get the following bound:

$$\delta_4 \geq (6^2 e^7)^{-30D^5} (4^4 \cdot 6 \cdot d2^{L+1})^{-5D^4}.$$

Combining the two cases for the value of D gives

$$\delta_4 \geq \min\left\{(6^2 e^7)^{-30d^5} (4^4 \cdot 6 \cdot d2^{L+1})^{-5d^4}, (6^2 e^7)^{-30\cdot 2^5} (4^4 \cdot 6 \cdot d2^{L+1})^{-5\cdot 2^4}\right\}.$$

It remains to show that $\Pi_z[\textsc{Zero}(I_f)]$ is a finite set. We prove this in the following lemma:

Lemma 14. *$\Pi_z[\textsc{Zero}(I_f)]$ is a finite set.*

Proof. Without loss of generality, we apply a translation so that we can assume that q is at the origin. To show that this image is a finite set, we show that $\textsc{Zero}(f(p), p_x f_y(p) - p_y f_x(p))$ is contained in finitely many circles centered at the origin. Then, the possible values of z are the radii of these circles, of which there are finitely many.

By Proposition 12, we know that each component of $\textsc{Zero}(f)$ is either a single point or a smooth one-dimensional manifold without boundary. Since there are finitely many components which are a single point, these components are contained within finitely many circles centered at the origin.

Take any one-dimensional component M and let r, s be two points in M and $\gamma : [0, 1] \to M$ a smooth path from $\gamma(0) = r$ to $\gamma(1) = s$ in M. Also, define $\rho : \mathbb{C}^2 \to \mathbb{C}$ by $\rho(x, y) = x^2 + y^2$. For any $p \in M$, we note that since M is smooth, the tangent line of M (or equivalently $\textsc{Zero}(f)$) at p is perpendicular to $\nabla(f) = (f_x(p), f_y(p))$ and this gradient is non-zero because p is a smooth point of $\textsc{Zero}(f)$. In addition, since $p_x f_y(p) - p_y f_x(p)$, we know that $(p_x, p_y) \parallel (f_x(p), f(y))$. Also note that $\nabla\rho = (2x, 2y) = 2 \cdot (x, y)$.

Now, we consider the square of the distance between γ and the origin, $\rho(\gamma(t))$. Taking the derivative of this function gives

$$\frac{d}{dt}\rho(\gamma(t)) = (\nabla\rho)(\gamma(t)) \cdot \gamma'(t) = 2\gamma(t) \cdot \gamma'(t) = 0.$$

Since this dot product is always zero, it implies the distance from the origin to points $\gamma(t)$ on M is constant. Therefore, M is contained in a unique circle centered at the origin. □

To find the degree of a singular point, assume that we have two boxes $B_1 \supsetneq B_2$ where the diameter of B_1 is less than both δ_3 and δ_4, B_2 contains a singular point of f and there is some radius $r > 0$ such that a circle of radius r centered at any point inside B_2 (including the singular point q) must lie entirely within the annulus $B_1 \setminus B_2$, see Fig. 9(a). Note this condition is satisfied if B_1 is at least 5 times larger than B_2 and B_2 is centered in B_1, which is the typical situation we consider below. See Fig. 9(b). Furthermore, to apply our Extended Plantinga & Vegter Algorithm of Section 4, we ensure that $B_1 \setminus B_2$ comes from a subdivision.

Now, there are 3 types of components (other than isolated points) in $\textsc{Zero}(f) \cap (B_1 \setminus \text{int}(B_2))$: (1) images of $[0, 1]$ both of whose endpoints are on ∂B_1, (2) images of $[0, 1]$ both of whose endpoints are on ∂B_2, and (3) images of $[0, 1]$ with one endpoint on each of ∂B_1 and ∂B_2. These three types are illustrated in Fig. 9(a).

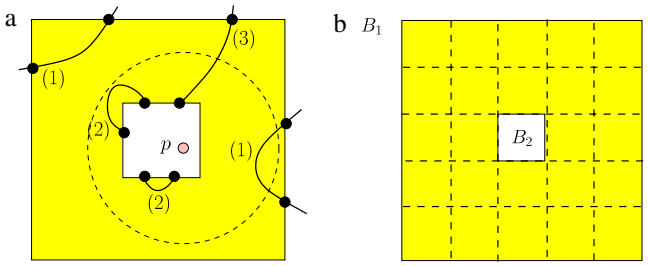


Fig. 9. Annular region $B_1 \setminus B_2$ with singularity q and the three types (1), (2) and (3) of components.

Let s be a point on *any* of these components, then traveling along $\textsc{Zero}(f)$ in one direction must lead to the singular point and the other direction must leave the neighborhood (be further than $\min\{\delta_3, \delta_4\}$) of the singular point. For, if not, then there must be a point r on $\textsc{Zero}(f) \cap B_1$ such that $\nabla f(r)$ is in the same direction as the line between r and the singular point, which is impossible since the width of B_1 is smaller than δ_4. Now, any piece defined by Proposition 12 which reaches the singular point exits the neighborhood of the singular point and the only way to leave the neighborhood is by way of a type (3) component. In addition, each piece includes either one or two components of type (3), and it only includes two components if both endpoints reach the singular point. This follows from the choice of δ_4. This shows:

Lemma 15. *The degree of the singular point in B_2 is the number of components of type* 3.

8. Overall algorithm

We now put all the above elements together to find a weak isotopic approximation to the algebraic curve $S = f^{-1}(0)$ within a region R_0, coming from a subdivision, where $f(X, Y) \in \mathbb{Z}[X, Y]$ has only isolated singularities. We first find the singularities of the curve S in R_0. Using the technique of Section 5, we can isolate the singularities p_i ($i = 1, 2, \ldots$) into disjoint boxes B_i. We assume the width of the B_i's is at most $\min\{\delta_3, \delta_4\}/6$. Let B_i' be the box of width 5 times the width of B_i, and concentric with B_i; we further assume $B_i' \subseteq R_0$. Note that these combinations are chosen to ensure that we have the typical situation in Section 7. Now we proceed to run the Extended Plantinga & Vegter Algorithm on the nice region $R^* := R_0 \setminus \bigcup_i B_i$, yielding a polygonal approximation G. We directly incorporate the technique of Section 7 into the following argument. If p_i is the singular point in B_i, then the degree of p_i is equal to the number of type (3) components in $G \cap (B_i' \setminus B_i)$. We connect these components directly to p_i, and discard any type (2) components. This produces the desired isotopic approximation.

We remark that we have not discussed the ε-approximation step because this is relatively easy to achieve in the Plantinga & Vegter approach. We only have to make sure that each subdivision box that contains a portion of the polygonal approximation G has width at most $\varepsilon/4$ since the result of the Plantinga & Vegter Algorithm only deforms the original curve at most one cell away.

9. Conclusion

This paper presents the first complete numerical subdivision algorithm for meshing an implicit algebraic curve that has only isolated singularities. This solves an open problem in the exact numerical approaches to meshing in 2-D (Boissonnat et al., 2006, p. 187). We pose three challenges:
(a) A worst case complexity bound for our procedure is possible by computing a lower bound for the size of a box in the final partition. However, this may not be the best way to measure the complexity of this algorithm because this algorithm is adaptive: the amount of work performed is not uniform over the initial region. We would like to provide adaptive bounds, similar to the integral analysis in Burr et al. (2009) for 1-D problems.

(b) In 3-D, a square-free integer polynomial $f(X, Y, Z)$ could have 1-dimensional singularities. We pose the problem of designing a purely numerical subdivision algorithm to handle 1-dimensional singularities.
(c) The practical implementation of an adaptive algorithm handling singularities, even based on our outline, must handle many important details. Computational experience is invaluable for future research into singularity computation.

Acknowledgements

The first author was partially supported by NSF Grant DMS-0701578. The third author was partially supported by the DoE under contract DE-FG02-00ER25053. The fourth author was supported by NSF Grant CCF-0728977.

References

Basu, Saugata, Pollack, Richard, Roy, Marie-Françoise, 2003. Algorithms in Real Algebraic Geometry. In: Algorithms and Computation in Mathematics, Springer.
Boissonnat, J.-D., Oudot, S., Provably good sampling and meshing of Lipschitz surfaces. In: Proc. 22nd ACM Symp. on Comp. Geometry, Sedona, Arizona, 2006, pp. 337–346.
Boissonnat, J.-D., Cohen-Steiner, D., Mourrain, B., Rote, G., Vegter, G., 2006. Meshing of surfaces. In: Boissonnat, J.-D., Teillaud, M. (Eds.), Effective Computational Geometry for Curves and Surfaces. Springer (Chapter 5).
Boissonnat, Jean-Daniel, Cohen-Steiner, David, Vegter, Gert, Isotopic implicit surfaces meshing. In: ACM Symp. Theory of Comput., 2004, pp. 301–309.
Brownawell, W.D., Yap, Chee K., 2009. Lower bounds for zero-dimensional projections. In: Proc. 34th Int'l Symp. Symbolic and Algebraic Comp., ISSAC'09, KIAS, Seoul, Korea, July 28–31, pp. 79–86.
Burr, Michael, Krahmer, Felix, Yap, Chee, Continuous amortization: a non-probabilistic adaptive analysis technique. In: Electronic Colloquium on Computational Complexity (ECCC), TR09 (136), December 2009. URL: http://eccc.hpi-web.de/report/2009/136/.
Burr, Michael, Sharma, Vikram, Yap, Chee, Evaluation-based root isolation (in preparation).
Cheng, Jin-San, Gao, Xiao-Shan, Yap, Chee K., 2009. Complete numerical isolation of real zeros in general triangular systems. In: Proc. Int'l Symp. Symbolic and Algebraic Comp., ISSAC'07, Waterloo, Canada, July 29–Aug. 1. Journal of Symbolic Computation 44 (7), 768–785. http://doi.acm.org/10.1145/1277548.1277562.
Cheng, S.-W., Dey, T.K., Ramos, E.A., Ray, T., Sampling and meshing a surface with guaranteed topology and geometry. In: Proc. 20th ACM Symp. on Comp. Geometry, 2004, pp. 280–289.
Cox, D., Little, J., O'Shea, D., 1992. Ideals, Varieties and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra. Springer-Verlag, New York.
Du, Zilin, Yap, Chee, Uniform complexity of approximating hypergeometric functions with absolute error. In: Sung-il Pae, Hyungju Park (Eds.), Proc. 7th Asian Symp. on Computer Math., ASCM 2005, 2006, pp. 246–249.
Harris, Joe, 1992. Algebraic Geometry. Springer-Verlag, New York.
Hartshorne, Robin, 1977. Algebraic Geometry. Springer-Verlag, New York.
Hong, H., 1996. An efficient method for analyzing the topology of plane real algebraic curves. Mathematics and Computers in Simulation 42, 571–582.
Jabri, Youssef, 2003. The Mountain Pass Theorem: Variants, Generalizations and Some Applications. In: Encyclopedia of Mathematics and its Applications, Cambridge University Press.
Kamath, Narayan, Subdivision algorithms for complex root isolation: empirical comparisons. Master's Thesis, Oxford University, August 2010.
Krantz, S.G., Parks, H.R., 1992. A Primer of Real Analytic Functions. Birkhäuser Verlag, Basel, ISBN: 3-7643-2768-5.
Lin, Long, Yap, Chee, Adaptive isotopic approximation of nonsingular curves: the parametrizability and non-local isotopy approach. In: Proc. 25th ACM Symp. on Comp. Geometry, Aarhus, Denmark, June 8–10, 2009, pp. 351–360. Accepted for Special Issue of SoCG 2009 in DCG.
Łojasiewicz, S., 1991. Introduction to Complex Analytic Geometry. Birkhäuser Verlag, Basel, ISBN: 3-7643-1935-6, Translated from the Polish by Maciej Klimek.
Lorensen, W.E., Cline, H.E., 1987. Marching cubes: a high resolution 3D surface construction algorithm. In: Stone, Maureen C. (Ed.), Computer Graphics, SIGGRAPH '87 Proceedings, vol. 21. pp. 163–169.
Mourrain, Bernard, Técourt, J.-P., Isotopic meshing of a real algebraic surface. Technical Report RR-5508, INRIA, Sophia-Antipolis, France, February 2005. Also, electronic proceedings, MEGA 2005.
Plantinga, Simon, Certified Algorithms for Implicit Surfaces, Ph.D. Thesis, Groningen University, Institute for Mathematics and Computing Science, Groningen, Netherlands, December 2006.
Plantinga, Simon, Vegter, Gert, 2004. Isotopic approximation of implicit curves and surfaces. In: Proc. Eurographics Symposium on Geometry Processing. ACM Press, New York, pp. 245–254.
Schoemer, Elmar, Wolpert, Nicola, 2006. An exact and efficient approach for computing a cell in an arrangement of quadrics. Comput. Geometry: Theory and Appl. 33, 65–97.
Seidel, Raimund, Wolpert, Nicola, On the exact computation of the topology of real algebraic curves. In: Proc. 21st ACM Symp. on Comp. Geometry, Pisa, Italy, 2005, pp. 107–116.
Snyder, J.M., 1992a. Generative Modeling for Computer Graphics and CAD: Symbolic Shape Design using Interval Analysis. Academic Press.

Snyder, J.M., 1992b. Interval analysis for computer graphics. SIGGRAPH Computer Graphics 26 (2), 121–130.
Stander, Barton T., Hart, John C., Guaranteeing the topology of an implicit surface polygonalization for interactive meshing. In: Proc. 24th Computer Graphics and Interactive Techniques, 1997, pp. 279–286.
Yap, Chee K., 2000. Fundamental Problems of Algorithmic Algebra. Oxford University Press.
Yap, Chee K., Complete subdivision algorithms, I: Intersection of Bezier curves. In: 22nd ACM Symp. on Comp. Geometry, July 2006, pp. 217–226.
Yap, Chee K., 2007. Is it Really Zero?. In: KIAS Newsletter. Spring (34).