Certified Computation of Planar Morse-Smale Complexes of Smooth Functions

(extended abstract)

Amit Chattopadhyay[1]

[1] Applied Mechanics and Mathematics (MEMA)
Université Catholique de Louvain
B-1348, Louvain-la-Neuve, Belgium
amit.chattopadhyay@uclouvain.be

Gert Vegter[2]

[2] Johann Bernoulli Institute of Mathematics and Computer Science
University of Groningen
Groningen, The Netherlands
G.Vegter@rug.nl

Chee K. Yap[3]

[3] Courant Institute of Mathematical Sciences
New York University
New York, New York, USA
yap@cs.nyu.edu

Abstract

The Morse-Smale complex is an important tool for global topological analysis in various problems of computational geometry and topology. Algorithms for Morse-Smale complexes have been presented in case of piecewise linear manifolds. However, previous research in this field does not provide certified methods in the case of smooth functions. In the current paper we use interval methods to compute a topologically correct approximation of Morse-Smale complex of smooth functions of two variables. The algorithm can also compute approximations that are arbitrarily close to the true geometric complex. Preliminary implementation results are reported.

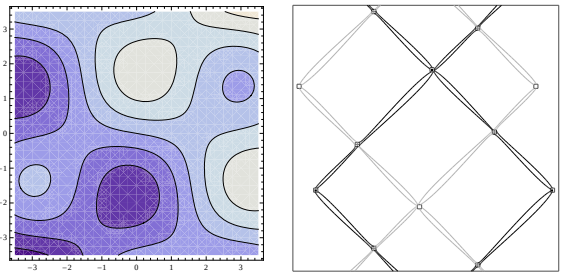
To Appear, ACM Symposium on Computational Geometry (SoCG) 2012.

Keywords: Morse-Smale Complex, Certified Computing, Exact Numerical Computation, Interval Arithmetic.

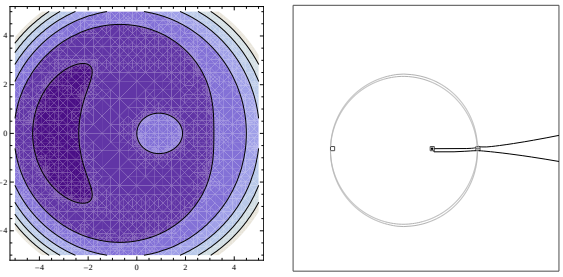
February 17, 2012

1 Introduction

Geometrical shapes occurring in the real world are often extremely complex. To analyze them, one associates a sufficiently smooth scalar field with the shape, e.g., a density function or a function interpolating gray values. Using this function, topological and geometrical information about the shape may be extracted, e.g., by computing its *Morse-Smale complex*. The cells of this complex are maximal connected sets consisting of orthogonal trajectories of the contour lines—curves of steepest ascent—with the same critical point of the function as origin and the same critical point as destination. The leftmost plots in Figures 1(a) and 1(b) illustrate the level sets of such a density function h, and the rightmost pictures the Morse-Smale complex of h as computed by the algorithm in this paper. This complex reveals the global topology of the shape. Recently, the Morse-Smale complex has been successfully applied in different areas like molecular shape analysis, image analysis, data and vector field simplification, visualization and detection of voids and clusters in galaxy distributions [6, 10].



(a) Contour plot (left) and Morse-Smale complex (right) of $h(x, y) = \cos x \ \sin y + 0.2\,(x + y)$ inside box $[-3.5, 3.5] \times [-3.5, 3.5]$. CPU-time: 11 seconds.



(b) Contour plot (left) and Morse-Smale Complex (right) of $h(x, y) = 10x - \frac{13}{2}(x^2 + y^2) + \frac{1}{3}\,(x^2 + y^2)^2$ inside box $[-5, 5] \times [-5, 5]$. CPU-time: 0.5 seconds.

Figure 1: Contour plots of Morse-Smale functions, and their Morse-Smale complexes.

¶1. Problem statement. A Morse function $h : \mathbb{R}^2 \to \mathbb{R}$ is a real-valued function with *non-degenerate critical points* (i.e., critical points with non-singular Hessian matrix). Non-degenerate critical points are *isolated*, and are either maxima, or minima, or saddle points. They correspond to singular points of the *gradient vector field* ∇h of h, of type sink, source or saddle, respectively. Regular integral curves of the gradient vector field ∇h are orthogonal trajectories of the regular level curves of h. We are interested in the configuration of integral curves of the gradient vector field. A *stable (unstable) separatrix* of a saddle point is the set of all regular points whose forward (backward) integral curve emerges from the saddle point. A non-degenerate saddle point has two stable and two unstable separatrices. A *Morse-Smale function* is a Morse function whose stable and unstable separatrices are disjoint. In particular, the unstable separatrices flow into a sink, and separate the unstable regions of two sources. Similarly, the stable separatrices emerge from a source, and separate the stable regions of two sinks. The corresponding gradient vector field is called a *Morse-Smale system* (MS-system). The *Morse-Smale complex* (MS-complex for short) is a complex consisting of all singularities, separatrices and the open regions forming their complement, of the *MS-system*. In other words, a cell of the MS-complex is a maximal connected subset of the domain of h consisting of points whose integral curves have the same origin and destination. See also [7, 8, 19][1]. The MS-complex describes the global structure of a Morse-Smale function.

Existing algorithms for MS-complexes [7, 8] compute the complex of a piecewise linear function on a piecewise linear manifold, or, in other words, of a discrete gradient-like vector field. When h is an analytic function, we cannot use these algorithms without first creating a piecewise linear approximation $\widetilde{h}$. However, the topology of the MS-complex of $\widetilde{h}$ is not guaranteed to be topologically equivalent to that of the corresponding smooth vector field. The topological correctness depends on how close the approximation $\widetilde{h}$ is to h. Here topological correctness means the computation of a complex, which is *topologically equivalent* (homeomorphic, or even isotopic) and *ε-close* to the 'true'

[1] These notions are also defined more precisely in Appendix 7, which is not part of this extended abstract, but added for convenience of the reviewers.

Morse-Smale complex (for some a priory specified $\varepsilon > 0$). As far as we know this problem has never been rigorously studied. Therefore the Main Problem of this paper is *to compute a piecewise-linear complex that is isotopic to the MS-complex of a smooth Morse-Smale function h.* In short, we seek an exact computation in the sense of the Exact Geometric Computation (EGC) paradigm [14]. Note that it is unclear whether many fundamental problems from analysis are exactly computable in the EGC sense. In particular, the current state-of-the-art in EGC does not (yet) provide a good approach for coping with degenerate situations, and, in fact, we it might well this paradigm needs to be extended to incorporate degeneracies. Therefore, we have to *assume* that the gradients we start out with are Morse-Smale systems. However, *generic* gradients are Morse-Smale systems [19], so the presence of degenerate singularities and of saddle-saddle connections is exceptional. Note that in restricted contexts, like the class of polynomial functions, absence of degenerate critical points (the first, and local, Morse-Smale condition) can be detected. However, even (most) polynomial gradient systems cannot be integrated explicitly, so absence of saddle-saddle connections (the second, and global, Morse-Smale condition) cannot be detected with current approaches. Detecting such connections even in a restricted context remains a challenging open problem.

¶2. Overview. We present a solution for the Main Problem which produces output such as illustrated in Figures 1(a)-(b). In particular, our algorithm produces:

- (arbitrarily small) isolated *certified boxes* each containing a unique saddle, source or sink;
- *certified initial and terminal intervals* (on the boundary of saddleboxes), each of which is guaranteed to contain a unique point corresponding of a stable or unstable separatrix;
- disjoint *certified funnels (strips)* around each separatrix, each of which contains exactly one separatrix, and as close to the separatrix as desired.

The construction of the MS-complex of a gradient system ∇h consists of two main steps: constructing disjoint certified boxes for its singular points, and constructing disjoint certified strips (funnels) enclosing its separatrices. Singular points of the gradient system are computed by solving the system of equations $h_x(x, y) = 0, h_y(x, y) = 0$. This is a special instance of the more general problem of solving a *generic system* of two equations $f(x, y) = 0$, $g(x, y) = 0$. Here, generic means that the Jacobi matrix at any solution is non-singular (geometrically it means the two curves $f = 0$ and $g = 0$ intersect transversally). Section 3 presents a method to compute disjoint isolating boxes for all solutions of such generic systems of two equations in two unknowns. In Section 4 these boxes are refined further. Saddle-boxes are augmented with four disjoint intervals in their boundary, one for each intersection of the boundary with the stable and unstable separatrices of the enclosed saddle point. We also show that these intervals can be made arbitrarily small, which is crucial in the second stage of the algorithm. Sink- and source-boxes are refined by computing boxes—not necessarily axis-aligned—around the sink or source on the boundaries of which the gradient system is transversal (pointing into the sink-box and out of the source-box). This implies that all integral curves reaching (emerging from) such a refined sink-box (source-box) lie inside this box beyond (before) the points of intersection.

Having constructed isolating boxes for the critical points of the MS-function h, we isolate the two unstable (stable) separatrices of each saddle point by tracing them in forward (backward) direction to the sinks (sources) that are their destination (origin). Reliable techniques for this tracing step are not readily available. In the simplest case where h is a polynomial (with rational coefficients), isolating the critical points can be done using known algebraic algorithms. But even in this restricted context integral curves are non-algebraic, so computing isolating sets for separatrices is a problem. To do so, we turn the Euler method for tracing integral curves into a certified algorithm, applying quantitative error bounds on the Euler method from [13] to construct isolating strips for all separatrices. Section 5 describes the second stage of the algorithm, in which isolating strips (funnels) for the stable and unstable separatrices are constructed. The boundary curves of funnels enclosing an unstable separatrix are polylines with initial point on a saddle box and terminal point on a sink box.

The gradient vector field is transversally pointing inward at each point of these polylines. The initial points of the polylines are connected by the unstable interval through which the separatrix leaves its saddle box, and, hence, enters the funnel. The terminal points of these polylines lie on the boundary of the same sink-box. See also Figure 6. Given this direction of the gradient system on the boundary of the funnel, the unstable separatrix enters the sink-box and tends to the enclosed sink, which is its destination. Although the width of the funnel may grow exponentially in the distance from the saddle-box, this growth is controlled in a known way. We exploit the computable (although very conservative) upper bound on this growth rate to obtain funnels that isolate separatrices from each other, and, hence, form a good approximation of the Morse-Smale complex together with the source- and sink-boxes. These upper bounds are also used to prove that the algorithm, which may need several subdivision steps, terminates. We have implemented this algorithm, using Interval Arithmetic. Section 7 presents sample output of our algorithm.

Remark 1. In this paper, the domain $\mathbb{D}$ of h is a finite union of axis-aligned dyadic boxes. Furthermore, we (have to) assume that all stable and unstable separatrices of the saddle points are transversal to the boundary. Computationally this means that any sufficiently close approximation of these separatrices is transversal to the boundary as well. □

¶3. Related Work. Milnor [17] provides a basic set-up for Morse theory. The survey paper [5], focussing on geometrical-topological properties of real functions, gives an excellent overview of recent works on MS-complexes. Originally, Morse theory was developed for smooth functions on smooth manifolds. Banchoff [4] introduced the equivalent definition of critical points on polyhedral surfaces. Many of the recent developments on MS-complexes are based on this definition. A completely different discrete version of Morse theory is provided by Foreman [9].

Different methods for computation. In the literature there are two different method for computing the Morse-Smale complexes: (a) boundary based approaches and (b) region based approaches. Boundary based methods compute boundaries of the cells of the MS-complex, i.e., the integral curves connecting a saddle to a source, or a saddle to a sink [23, 3, 8]. On the other hand, watershed algorithms for image segmentation are considered as region based approaches [16]. Edelsbrunner et.al [8] computes the Morse-Smale complex of piecewise linear manifolds using a paradigm called *Simulation of Differentiability*. In higher dimensions they give an algorithm for computing Morse Smale complexes of piecewise linear 3-manifolds [7].

Morse-Smale complexes have also been applied in shape analysis and data simplification. Computing MS-complexes is strongly related to vector field visualization [11]. In a similar context, designing vector fields on surfaces has been studied for many graphics applications [25]. Cazals et.al. [6] applied discrete Morse theory to molecular shape analysis.

This paper contributes to the emerging area of Exact Numerical Algorithms for geometric problems [24]. Recent algorithms of this genre (e.g., [20, 15]) are numerical subdivision algorithms based on interval function evaluation and sign evaluation.

2 Preliminaries

Our computational model[2] has two simple foundations: (1) BigFloat packages and (2) interval arithmetic (IA) [18]. Together, these are able to deliver efficient and guaranteed results in the implementation of our algorithms. A BigFloat package is a software implementation of exact ring $(+, -, \times)$ operations, division by 2, and exact comparisons, over the set $\mathbb{F} = \{m2^n : m, n \in \mathbb{Z}\}$ of *dyadic numbers*. In practice, we can use IEEE machine arithmetic as a filter for BigFloat computation to speed up many computation. *Range functions* form a basic tool of IA: given any function

[2]See Appendix 7 for additional details.

$F : \mathbb{R}^m \to \mathbb{R}^n$, a *range function* $\square F$ for F computes for each m-dimensional interval I (i.e., an m-box) an n-dimensional interval $\square F(I)$, such that $F(I) \subset \square F(I)$. A range function is said to be *convergent* if the diameter of the output interval converges to 0 when the diameter of the input interval shrinks to 0. Convergent range functions exist for the basic operators and functions, so all range functions are assumed to be convergent. Moreover, we assume that the sign of functions can be evaluated exactly at dyadic numbers. All our boxes are *dyadic boxes*, meaning that their corners have dyadic coordinates. A useful technique here is the interval form of the implicit function theorem (see Appendix 7.)

3 Isolating boxes for singularities of gradient fields

As a first step towards the construction of the Morse-Smale complex of h we construct disjoint isolating boxes for the singular points of ∇h. To this end, we first show how to compute isolating boxes for the solutions of a generic system of two equations in two unknowns, which are confined to a bounded domain in the plane, which is a finite union of dyadic boxes. Applying this general method to the case in which the two equations are defined by the components of the gradient vector field ∇h we obtain isolating boxes for the singularities of this gradient field.

¶4. Certified solutions of systems of equations We consider a system of equations

$$f(x, y) = 0, \quad g(x, y) = 0, \tag{1}$$

where f and g are C^1-functions defined on a bounded axisparallel box $\mathbb{D} \subset \mathbb{R}^2$ with *dyadic vertices*. Furthermore, we assume that the system has only non-degenerate solutions, i.e., the Jacobian determinant $\frac{\partial(f,g)}{\partial(x,y)}\Big|_{(x_0,y_0)}$ is non-zero at a solution (x_0, y_0). Geometrically this means that the curves Z_f and Z_g, given by $f(x, y) = 0$ and $g(x, y) = 0$, respectively are regular near a point of intersection (x_0, y_0) and the intersection is transversal. Since $\mathbb{D}$ is compact, the system has finitely many solutions in $\mathbb{D}$.

Our goal is to construct a collection of axis-aligned boxes $\mathbb{B}_1, \ldots, \mathbb{B}_m$ and $\mathbb{B}'_1, \ldots, \mathbb{B}'_m$ such that (i) box $\mathbb{B}_i$ is concentric with and strictly contained in $\mathbb{B}'_i$, (ii) the boxes $\mathbb{B}'_i$ are disjoint, (iii) each solution of (1) is contained in one of the boxes $\mathbb{B}_i$, and (iv) each box $\mathbb{B}'_i$ contains exactly one solution (contained inside the enclosed box $\mathbb{B}_i$). The box pair $(\mathbb{B}_i, \mathbb{B}'_i)$ is *certified*: $\mathbb{B}_i$ contains a solution, and $\mathbb{B}'_i$ provides positive clearance to other solutions.
In fact, the sequence of boxes will satisfy the following stronger conditions: 1. The curves Z_f and Z_g each intersect the boundary of $\mathbb{B}'_i$ transversally in two points.
2. There are disjoint intervals $\mathbb{I}_i^0(f)$, $\mathbb{I}_i^0(g)$, $\mathbb{I}_i^1(f)$ and $\mathbb{I}_i^1(g)$ in the boundary of $\mathbb{B}'_i$ (in this order), where the first and the third interval each contain one point of intersection of Z_f and $\partial\mathbb{B}'_i$, and the second and fourth interval each contain one point of intersection of Z_g and $\partial\mathbb{B}'_i$.
3. The (interval) Jacobian determinant of f and g does not vanish on $\mathbb{B}'_i$, i.e., $0 \notin \square\frac{\partial(f,g)}{\partial(x,y)}(\mathbb{B}'_i)$.

¶5. Construction of certified box pairs. We first subdivide the domain $\mathbb{D}$ into equal-sized boxes $\mathbb{I}$ (called grid-boxes), until all boxes satisfy certain conditions to be introduced now. For a (square, axis-aligned) box $\mathbb{I}$ let $N_\varrho(\mathbb{I})$ be the box obtained by multiplying box $\mathbb{I}$ from its center by a factor of $1 + \varrho$, where $\frac{1}{2} \le \varrho \le 1$. We also denote $N_1(\mathbb{I})$ by $N(\mathbb{I})$; this is the box formed by the union of $\mathbb{I}$ and its eight neighbor grid-boxes. The algorithm subdivides $\mathbb{D}$ until all grid-boxes $\mathbb{I}$ satisfy $C_0(\mathbb{I}) \vee \big(C_1(\mathbb{I}) \wedge C_2(\mathbb{I})\big)$, where the clauses $C_i(\mathbb{I})$, $i = 0, 1, 2$, are the following predicates:

$$C_0(\mathbb{I}): \quad 0 \notin \square f(\mathbb{I}) \vee 0 \notin \square g(\mathbb{I})$$

$$C_1(\mathbb{I}): \quad 0 \notin \square\frac{\partial(f,g)}{\partial(x,y)}(N(\mathbb{I}))$$

$$C_2(\mathbb{I}): \quad C_2(\mathbb{I}, f) \wedge C_2(\mathbb{I}, g), \text{ where } C_2(\mathbb{I}, f) = \langle \square\frac{\nabla f}{||\nabla f||}(N(\mathbb{I})), \square\frac{\nabla f}{||\nabla f||}(N(\mathbb{I}))\rangle \ge \cos\frac{\pi}{30}.$$

If $C_0(\mathbb{I})$ holds, box $\mathbb{I}$ does not contain a solution, so it is discarded. The second predicate guarantees that $N(\mathbb{I})$ contains at most one solution. This is a consequence of Proposition 8. Condition $C_2(\mathbb{I})$ is a *small angle variation condition*, guaranteeing that the variation of the unit normals of the curves $Z_f \cap N(\mathbb{I})$ and $Z_g \cap N(\mathbb{I})$ do not vary too much, so these curves are regular, and even 'nearly linear' (the unit normal of Z_f is the normalized gradient of f).

Remark 2. Condition $C_1(\mathbb{I})$ implies that there is a computable positive lower bound $\alpha(\mathbb{I})$ on the angle between $\nabla f(p)$ and $\nabla g(q)$ where p, q range over the box $\mathbb{I}$. More precisely, to compute $\alpha(\mathbb{I})$, we first compute a lower bound L on the quantity $\frac{\partial(f,g)}{\partial(x,y)} \frac{1}{\|\nabla f\| \cdot \|\nabla g\|}$. This L may be obtained by an interval evaluation of this quantity at $N(\mathbb{I})$; note that $L > 0$ iff condition $C_1(\mathbb{I})$ holds. We define $\alpha(\mathbb{I})$ as $\arcsin(L)$. □

Our algorithm will construct disjoint certified boxes surrounding a box $\mathbb{I}$. Since the surrounding boxes $N(\mathbb{I})$ and $N(\mathbb{J})$ of disjoint boxes $\mathbb{I}$ and $\mathbb{J}$ may intersect, we consider smaller surrounding boxes for a grid box $\mathbb{I}$. If $N(\mathbb{J}) \cap \mathbb{I} = \emptyset$, then $N_{1/2}(\mathbb{I})$ and $N_{1/2}(\mathbb{J})$ have disjoint interiors. This is a key observation with regard to the correctness of our algorithm.

Lemma 3. *Let $\mathbb{I}$ be a box such that conditions $\neg C_0(\mathbb{I})$, $C_1(\mathbb{I})$ and $C_2(\mathbb{I})$ hold. Let d be the length of its edges, and let $\frac{1}{2} \leq \varrho \leq 1$.*
1. If Z_f intersects $\mathbb{I}$, it intersects the boundary of $N_\varrho(\mathbb{I})$ transversally at exactly two points. At a point of intersection of Z_f and an edge e of $\partial N_\varrho(\mathbb{I})$ the angle between Z_f and e is at least $\frac{1}{15}\pi$.
2. If $\mathbb{I}$ contains a point of intersection of Z_f and Z_g, then the points of intersection of Z_f and $\partial N_\varrho(\mathbb{I})$ are at distance at least $2\varrho d \tan\frac{1}{2}\alpha(\mathbb{I})$ from the points of intersection of Z_g and $\partial N_\varrho(\mathbb{I})$. On $\partial N_\varrho(\mathbb{I})$, the points of intersection with Z_f and with Z_g are alternating.

If f and g are linear functions, the previous result is easy to check. The proof of the general case uses the small angle variation of ∇f and ∇g. Note that this Lemma already establishes the existence and isolation of a solution of (1) in case we can certify that Z_f and Z_g intersect $\mathbb{I}$.

¶6. Towards an algorithm. For each grid-box $\mathbb{I}$, the algorithm calls one of the following:
• DISCARD($\mathbb{I}$), if it decides that $\mathbb{I}$ does not contain a solution. It marks box $\mathbb{I}$ as processed.
• REPORTSOLUTION($\mathbb{I}$). It returns the certified pair $(N_{1/2}(\mathbb{I}), N(\mathbb{I}))$, and marks all boxes contained in $N(\mathbb{I})$ as processed.

In the latter case a solution is found inside $N_{1/2}(\mathbb{I})$, but, as will become clear later, it may not be contained in the smaller box $\mathbb{I}$. In view of $C_1(\mathbb{I})$ none of the grid-boxes in $N(\mathbb{I})$ contain a solution different from the one reported, so they are marked as being processed.

Decisions are based on evaluation of the signs of f and g at the vertices of the grid-boxes (or at certain dyadic points on edges of grid-boxes). An edge of a box is called *bichromatic* (*monochromatic*) for f if the signs of the value of f at its vertices are opposite (equal).

Algorithm, case 1: $\mathbb{I}$ has a bichromatic edge for f and a bichromatic edge for g. Then Z_f and Z_g intersect $\mathbb{I}$, and, according to Lemma 3, part 1, both curves intersect the boundary of $N_{1/2}(\mathbb{I})$ transversally in exactly two points. For each of the two points in $\partial N_{1/2}(\mathbb{I})$ the algorithm computes an isolating interval—called an f-interval—on $\partial N_{1/2}(\mathbb{I})$ of length $\frac{1}{2} d \tan\frac{1}{2}\alpha(\mathbb{I})$. The two g-intervals are computed similarly. If the f- and g-intervals are not interleaving, there is no solution of (1) in box $\mathbb{I}$—even though there may be a solution in $N_{1/2}(\mathbb{I})$—and DISCARD($\mathbb{I}$) is called. This follows from Lemma 3, part 2. If the intervals are interleaving, then there is a point of intersection inside $N_{1/2}(\mathbb{I})$, so the algorithm calls REPORTSOLUTION($\mathbb{I}$).

Algorithm, case 2: $\mathbb{I}$ contains no bichromatic edge for f (g), and at least one bichromatic edge for g (f, respectively). We only consider the case in which all edges of $\mathbb{I}$ are monochromatic for f. Then the algorithm also evaluates the sign of f at the vertices of the box $N_{1/2}(\mathbb{I})$. If $N_{1/2}(\mathbb{I})$ has no *disjoint* bichromatic edges (as in the fourth and fifth configuration of

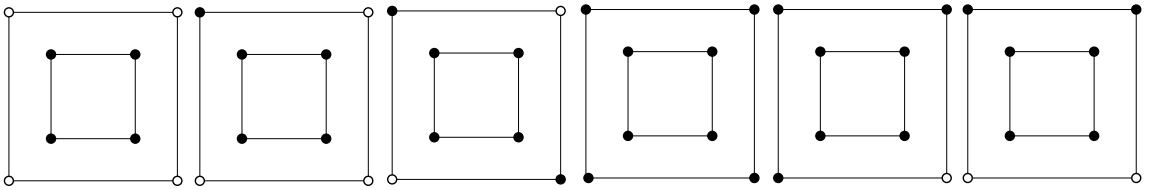


Figure 2: Sign patterns of the box $N_{1/2}(\mathbb{I})$ enclosing the grid-box $\mathbb{I}$ with monochromatic edges for f. The three leftmost configurations are ruled out by the small normal variation condition $C_2(\mathbb{I})$. The fourth, fifth and sixth configuration are all possible, but only in the sixth situation Z_f may intersect the inner box.

Figure 2), the isocurve Z_f does not intersect $\mathbb{I}$, so the algorithm calls DISCARD($\mathbb{I}$). To deal with the remaining case, in which $N_{1/2}(\mathbb{I})$ has two *disjoint* bichromatic edges (as in the sixth configuration in Figure 2) we need to evaluate the sign of f at certain dyadic points of these bichromatic edges.

To make this more precise, let e be the edge of $\mathbb{I}$ closest to the (monochromatic) edge of $N_{1/2}(\mathbb{I})$ at whose vertices the sign of f is the opposite of the sign of f at the vertices of $\mathbb{I}$ (this is the bottom edge of the outer box in the sixth configuration of Figure 2). Let let l and r be the points of intersection of the line through e and the edges of the surrounding box $N_\varrho(\mathbb{I})$, perpendicular to e. See also Figure 3. The intervals on the boundary of this surrounding box with length $\frac{1}{4}(1+\varrho)d$, centered at l and r, respectively, are denoted by $L_\varrho(e)$ and $R_\varrho(e)$, where d is the length of the edges of $\mathbb{I}$, and $\frac{1}{2} \le \varrho \le 1$.

Lemma 4. *Let $\mathbb{I}$ be a box such that $\neg C_0(\mathbb{I})$, $C_1(\mathbb{I})$ and $C_2(\mathbb{I})$ hold, and let e be one of its edges. Let $\frac{1}{2} \le \varrho \le 1$.*
1. If Z_f intersect an edge e of the boundary of $\mathbb{I}$ in at least two points, then it transversally intersects $\partial N_\varrho(\mathbb{I})$ in exactly two points, one in each of the intervals $L_\varrho(e)$ and $R_\varrho(e)$.
2. If Z_f intersects $\partial N_\varrho(\mathbb{I})$ in the intervals $L_\varrho(e)$ and $R_\varrho(e)$, then these intersections are transversal, and Z_f intersects $\partial N_\varrho(\mathbb{I})$ at exactly two points, one in each of these intervals.

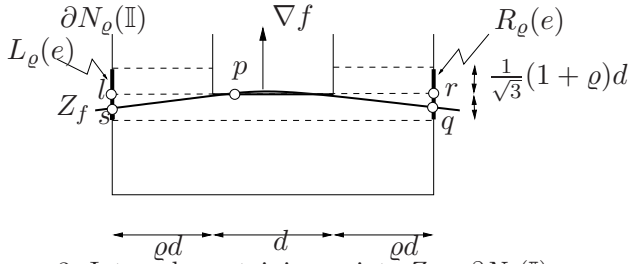


Figure 3: Intervals containing points $Z_f \cap \partial N_\varrho(\mathbb{I})$.

Algorithm, case 2 (cont'd). By evaluating the signs of f at the (dyadic) endpoints of the interval $L_\varrho(e)$ and $R_\varrho(e)$ the algorithm decides whether they contain a point of intersection with Z_f. If at least one of these intervals is disjoint from Z_f, then DISCARD($\mathbb{I}$) is called. Otherwise, the algorithm computes isolating f- and g-intervals of length $\frac{1}{2}d\tan\frac{1}{2}\alpha(I)$. As in case 1, the algorithm calls REPORTSOLUTION($\mathbb{I}$) if these intervals are interleaving, and DISCARD($\mathbb{I}$) otherwise.

Algorithm, case 3: all edges of $\mathbb{I}$ are monochromatic for both f and g. Again, let e be the (unique) edge of $\mathbb{I}$ closest to the edge of $N_{1/2}(\mathbb{I})$ which is monochromatic for f, at whose vertices the sign of f is the opposite of the sign of f at the vertices of $\mathbb{I}$. Edge e' of $\mathbb{I}$ is defined similarly for g.
Case 3.1: $e = e'$. In this case Z_f or Z_g does not intersect $\mathbb{I}$. Indeed, if Z_f intersects $\mathbb{I}$, it intersects e in at least two points, so there is a point $p \in Z_f$ at which $\nabla f(p)$ is perpendicular to e. Condition $C_1(\mathbb{I})$ guarantees that ∇g is nowhere parallel to $\nabla f(p)$, so Z_g does not intersect e, and, hence, does not intersect $\mathbb{I}$. Therefore, DISCARD($\mathbb{I}$) is called.
Case 3.2: $e \neq e'$. If Z_f does not intersect $L_\varrho(e)$ or $R_\varrho(e)$, or if Z_g does not intersect $L_\varrho(e')$ or $R_\varrho(e')$, then, as in case 2, the algorithm calls DISCARD($\mathbb{I}$). Otherwise, $L_\varrho(e)$ or $R_\varrho(e)$ are isolating f-intervals which are disjoint from the isolating g-intervals $L_\varrho(e')$ or $R_\varrho(e')$. If e and e' are perpendicular, then these f- and g-intervals are interleaving, and, hence, REPORTSOLUTION($\mathbb{I}$) is called. Otherwise, there is no solution in $\mathbb{I}$, so DISCARD($\mathbb{I}$) is called.

¶7. Refinement: disjoint surrounding boxes. We would like distinct isolating boxes $\mathbb{I}, \mathbb{J}$ to have disjoint surrounding boxes $N(\mathbb{I}), N(\mathbb{J})$. There is a simple way to ensure this: we just use the predicate $C_1(N(\mathbb{I}))$ to instead of $C_1(\mathbb{I})$ in the above subdivision process. Then, if the interior of $N(\mathbb{I}) \cap N(\mathbb{J})$ is non-empty, we can discard any one of $\mathbb{I}$ or $\mathbb{J}$.

4 Construction of source-, saddle- and sink-boxes

Using the previous method we construct disjoint isolating boxes $\mathbb{B}'_1, \ldots, \mathbb{B}'_m$ for the singularities of ∇h, which are refined further to facilitate the construction of the MS-complex.

Each saddlebox is augmented by computing four disjoint intervals in its boundary, one for each intersection of a stable or unstable separatrix with the box boundary. See the leftmost picture in Figure 4. These *isolating separatrix intervals* can be made *arbitrarily small*. These intervals are constructed as the intersection of disjoint stable and unstable wedges along the boundary of which the gradient vector field is pointing outward and inward, respectively. Therefore, these wedges enclose the stable and unstable separatrices, respectively. In the full paper we show how the width of these wedges can be made arbitrarily small, provided the saddlebox satisfies some additional (computable) criteria.

Inside each source- and sink-box we construct a box on the boundary of which the ∇h is pointing outward or inward, respectively. These boxes are contained in the source- and sinkboxes constructed in the previous section, but are not necessarily axes-aligned. They are aligned with approximations of the (orthogonal) eigenvectors of the Hessian of h at the center of the source- or sinkbox, however. See the rightmost picture in Figure 4.

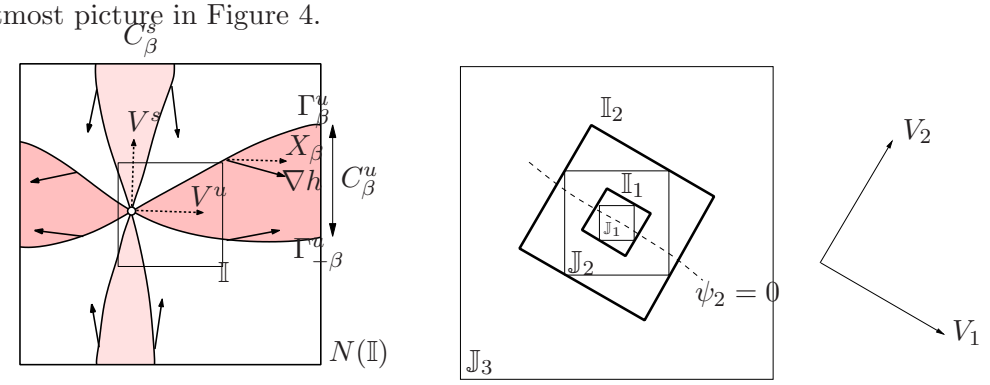


Figure 4: **Left:** The unstable wedge C^u_β enclosing the unstable separatrix, and the stable wedge C^s_β enclosing the stable separatrix. The gradient vector field ∇h, represented by solid arrows, is transversally pointing inward along the boundary of the unstable wedge, and outward along the boundary of the stable wedge. At points of the unstable wedge boundary the vector field X_β is (anti)parallel to the unstable eigenvector V^u of the Hessian of h at the center of the box, so ∇h makes an angle $-\beta$ with V^u or $-V^u$ there. **Right:** Sink-/source-boxes, nearly aligned with the eigenvectors of the Hessian matrix of h at the center of the box, at the boundaries of which ∇h is transversal (pointing inward/outward).

5 Isolating funnels around separatrices

We continue by constructing certified *funnels* around separatrices inside the region $\mathbb{D}^* = \mathbb{D} \setminus \cup_{i=1}^m \mathbb{B}'_i$, where $\mathbb{B}'_1, \ldots, \mathbb{B}'_m$ are the boxes isolating the singular points of ∇h. If the forward orbits of the endpoints of an unstable separatrix interval have the same sink of ∇h as destination, these forward orbits bound a region around the unstable separatrix leaving the saddle box via this unstable segment. This region is called a *funnel* for the separatrix (this terminology is borrowed from [13]). In this section we show how to construct a (slightly larger) certified funnel for each separatrix, the boundaries of which are two polylines, each starting from an endpoint of the corresponding separatrix interval in the saddle box of the separatrix. These polylines are integral curves of ∇h through a point p by enclosing it between (approximations of) the forward orbits of X_ϑ and $X_{-\vartheta}$ through p, where

$X_{\pm\vartheta}$ is the vector field obtained by rotating the vector field $X = \nabla h$ over an angle $\pm\vartheta$, i.e., $X_\vartheta = \begin{pmatrix} \cos\vartheta & -\sin\vartheta \\ \sin\vartheta & \cos\vartheta \end{pmatrix} \begin{pmatrix} h_x \\ h_y \end{pmatrix}$. See Figure 5 (Right).

¶8. Small angle variation. We subdivide the region $\mathbb{D}^*$ into square boxes over which the angle variation of ∇h is at most ϑ, where ϑ is to be determined later. If $X = (f, g)$ is a vector field on $\mathbb{R}^2$, then the angle variation over a regular curve Γ is given by [2, Section 36.7]: $\int_\Gamma \frac{g\,df - f\,dg}{f^2+g^2}$. If $X = \nabla h$, this angle variation is equal to

$$\int_\Gamma \frac{(h_x h_{xy} - h_y h_{xx})\,dx + (h_x h_{yy} - h_y h_{xy})\,dy}{h_x^2 + h_y^2}.$$

The next result gives an upper bound for the size of grid-boxes such that the angle variation of ∇h over any box in $\mathbb{D}^*$ is less than ϑ. It also provides a construction of a 'box-wise' funnel around integral curves of ∇h.

Lemma 5. *Let C_0 be a constant such that $\max_{\mathbb{D}^*}\left(\left|\frac{h_x h_{xy} - h_y h_{xx}}{h_x^2 + h_y^2}\right|, \left|\frac{h_x h_{yy} - h_y h_{xy}}{h_x^2 + h_y^2}\right|\right) \le C_0$, and let $0 < \vartheta < \frac{\pi}{2}$, and let the width w of grid-boxes in $\mathbb{D}^*$ satisfy $w \le \frac{\vartheta}{2C_0\sqrt{2}}$. Then*

1. The angle variation of ∇h over any gridbox is less than ϑ.
2. Let p be a point on an edge of a gridbox, and let q_ϑ ($q_{-\vartheta}$) be the point on the boundary of the gridbox into which $X_\vartheta(p)$ ($X_{-\vartheta}(p)$) is pointing, such that the line segment pq_ϑ ($pq_{-\vartheta}$) has direction $X_\vartheta(p)$ ($X_{-\vartheta}(p)$). See Figure 5. Then ∇h is pointing rightward along the (directed) line segment pq_ϑ and leftward along $pq_{-\vartheta}$.
3. The function h is strictly increasing on the line segments pq_ϑ and $pq_{-\vartheta}$.

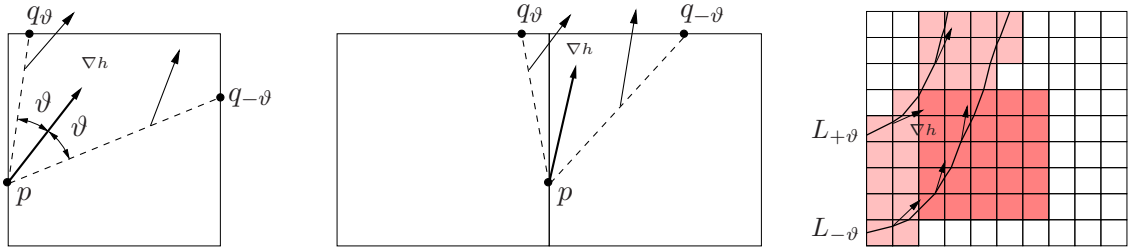


Figure 5: **Left:** The orientation of ∇h with respect to $X_\vartheta(p)$ does not change over a grid box. **Right:** A funnel formed by two polylines covered by two vertical ε-strips, one ε-box and three horizontal ε-strips. Here ε is six times the width of a grid box. $L_{+\vartheta}$ intersects the ε-box in a single edge, which is quasivertical but not quasihorizontal. All edges of $L_{-\vartheta}$ inside the ε-box are quasivertical as well (and some of them are also quasihorizontal). ∇h points inward, and the enclosed separatrix (not shown) runs from left to right.

The third property will be crucial for constructing fences around separatrices that cross each grid-box at most once.

¶9. Fencing in the separatrices. For each isolating unstable separatrix interval $\mathbb{J}$ on the boundary of a saddle box we construct two polylines (called *fences*) $L_{-\vartheta}(\mathbb{J})$ and $L_\vartheta(\mathbb{J})$ as follows. The initial points of these polylines are the endpoints q_- and q_+ of $\mathbb{J}$, where q_- comes before q_+ in the counterclockwise orientation of the boundary of the saddle box. The polyline $L_\vartheta(\mathbb{J})$ is uniquely defined by requiring that its vertices $q_+ = p_0, p_1, \ldots, p_n$ ly on grid edges, with the property that (i) The line segment $p_{i-1}p_i$, $0 < i \le n$, lies in a (closed) grid box of $\mathbb{D}^*$, and has direction $X_\vartheta(p_{i-1})$, and (ii) p_n, the last vertex, lies on the boundary of $\mathbb{D}^*$. The polyline $L_{-\vartheta}(\mathbb{J})$ is defined similarly with respect to q_- and $X_{-\vartheta}$. If the endpoints of the fences $L_\vartheta(\mathbb{J})$ and $L_{-\vartheta}(\mathbb{J})$ lie on the same connected component of the boundary of $\mathbb{D}^*$, then these fences split $\mathbb{D}^*$ into two connected regions. See Figure 6. In this case, the region containing the separatrix interval $\mathbb{J}$ in its boundary is called the *funnel* of $\mathbb{J}$ (with angle ϑ) denoted by $F_\vartheta(\mathbb{J})$. Its boundary consists of $\mathbb{J}$, the two fences $L_\vartheta(\mathbb{J})$ and $L_{-\vartheta}(\mathbb{J})$, and a curve

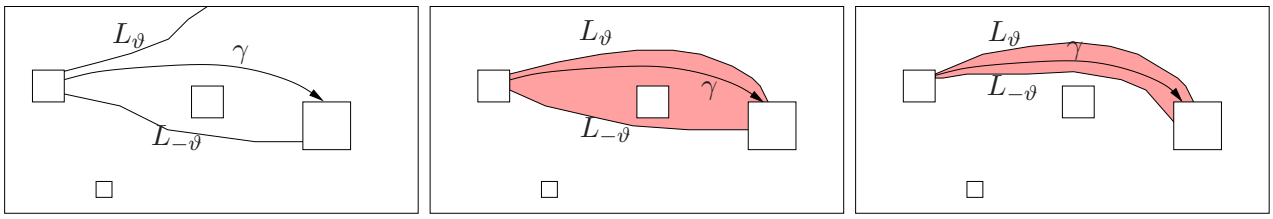


Figure 6: Fences around a separatrix γ. If the fences end in the same connected component of the boundary of $\mathbb{D}^*$, *and* enclose a simply connected region in $\mathbb{D}^*$, then this region isolates the separatrix from the source-, sink- and saddle-boxes (rightmost picture).

$\mathbb{J}^*$ on the boundary of $\partial\mathbb{D}$ connecting the endpoints of these fences. The fences $L_{\pm\vartheta}(\mathbb{J})$ start from the endpoints of $\mathbb{J}$. New line segments are added until a fence hits the boundary of $\mathbb{D}^*$, or until the width of the enclosed funnel becomes larger than a positive number ε (to be specified later). To define the width of a funnel more precisely, we distinguish quasihorizontal and quasivertical parts of a funnel, and show that the width of a funnel does not increase substantially at transitions between these quasihorizontal and quasivertical parts. A nonzero vector $v = (v_1, v_2)$ is called *quasihorizontal* if $|v_2| \leq 2|v_1|$, and *quasivertical* if $|v_1| \leq 2|v_2|$. Note that each nonzero vector is quasihorizontal, quasivertical, or both. A *horizontal ε-strip* is the union of a sequence of boxes where successive boxes share a vertical edge, such that the horizontal edge of the rectangle thus obtained has length at most ε. A *vertical ε-strip* is defined similarly. An *ε-box* is a square box with edge length at most ε which is the union of a number of boxes.

The construction of a funnel is illustrated in Figure 5 (Right). If both inital segments of $L_{\pm\vartheta}$ are quasihorizontal (quasivertical), the width of the funnel (under construction) is measured in the vertical (horizontal) direction. As long as the width is less than ε, the funnel crosses the next vertical (horizontal) ε-strip by extending $L_{\pm\vartheta}$ with one or two line segments. As soon as one of the new segments of L_ϑ or $L_{-\vartheta}$ ceases to be quasihorizontal (quasivertical), the funnel is extended through a square box of minimal width. Upon exit of this square box the edges of L_ϑ are both quasivertical (quasihorizontal). See Figure 5 (Right). This observation, together with a standard result from Dynamical Systems Theory [13, Theorem 4.4.1] on the growth of the deviation of the Euler approximation from the integral curve it approximates, yields the following (rather conservative) upper bound on the *growth of the funnel width* of a separatrix having M transitions between quasihorizontal and quasivertical parts (which are bounded by singularity-boxes or ε-boxes).

Lemma 6. *Let T be the (computable) edge length of a bounding square of the domain $\mathbb{D}$ of h, and let M be the total number of quasihorizontal and quasivertical parts of the polylines bounding a separatrix funnel. There are computable constants*[3] *ϑ_0, c_1, c_2 and C, which* ***only depend on*** *$\mathbb{D}^*$* ***and*** *h, such that the width of the funnel does not exceed $(c_1\vartheta + c_2 w)\, e^{CMT}$, provided $\vartheta \leq \vartheta_0$. In particular, this width is at most ε if $c_1\vartheta + c_2 w \leq \varepsilon\, e^{-CMT}$.*

6 Construction of the MS-complex

¶10. The algorithm. The construction of the MS-complex of h is a rather straightforward application of the preceding results. It uses a parameter M, the (a priori unknown) number of transitions (at ε-boxes) between quasihorizontal and quasivertical parts of a funnel. Let T be the edge length of a bounding square of the domain $\mathbb{D}$ of h. Then the algorithm performs the following steps.

Step 1. Construct certified isolating boxes $\mathbb{B}'_1, \ldots, \mathbb{B}'_m$ for the singularities of ∇h (cf Section 3).

Step 2. Let $\mathbb{D}^*$ be the closure of $\mathbb{D} \setminus (B'_1 \cup \cdots \cup \mathbb{B}'_m)$. Compute the constants ϑ_0, c_1, c_2 and C, which depend only on h and $\mathbb{D}^*$. Set ε to the minimum of the width of the source-, sink- and sinkboxes.

[3] See Appendix 7—not part of the actual paper—for a precise definition of these constants, and for a proof of Lemma 6.

Step 3. Let ϑ and w be such that $w \leq \frac{\vartheta}{2C_0\sqrt{2}}$, $\vartheta \leq \min(\frac{\pi}{40}, \vartheta_0)$, and $c_1\vartheta + c_2 w \leq \varepsilon\, e^{-CMT}$ (cf Lemma 6). Subdivide $\mathbb{D}^*$ until all gridboxes have maximum width w. For each saddle box, compute four separatrix intervals on its boundary, of width at most w.

Step 4. For each stable and unstable separatrix interval do the following. Start the computation of a funnel for a separatrix by setting M to a small number M_0 (say 4). Compute the fences $L_{-\vartheta}$ and L_ϑ, keeping track of the width of the enclosed funnel under construction and of the number m of transitions between quasihorizontal and quasivertical parts of this funnel.

If *the width of the funnel exceeds ε or the number of transitions m exceeds M*, then abort the computation of the current funnel, discard all funnels constructed so far, set M to twice its current value and goto Step 3.

If *the funnel intersects an already constructed funnel, or a source- or sinkbox on which it does not terminate* (i.e., if only one of its fences intersects this box), then set ε to half its current value, discard all funnels constructed so far, and goto Step 3.

If *the funnel intersects a saddlebox $\mathbb{B}_i'$*, then decrease the size of $\mathbb{B}_i'$ by a factor of two via *subdivision*, discard all funnels constructed so far, set ε to half its current value, and goto Step 2. (Note that $\mathbb{D}^*$ gets larger, so the constants in Step 2 have to be recomputed.)

Otherwise, the fences end on the same component of the boundary of $\partial\mathbb{D}^*$. The enclosed funnel is simply connected, and does not intersect any of the funnels constructed so far. Add the funnel to the output, and reset M to M_0 (and repeat until all separatrices have been processed).

¶11. Termination. Since the gradient field ∇h is a $2D$ Morse-Smale system, its separatrices are disjoint. Their intersections with $\mathbb{D}^*$ are compact, and have positive distance (although this distance is not known a priori). In the main loop of the algorithm, the maximal funnel width ε is bisected if funnels intersect, and saddleboxes intersected by the funnel are subdivided, so after a finite number of iterations of the main loop its value is less than half the minimum distance between any pair of distinct separatrices, and funnels stay clear from saddleboxes (apart from the one containing the origin or destination of the enclosed separatrix).

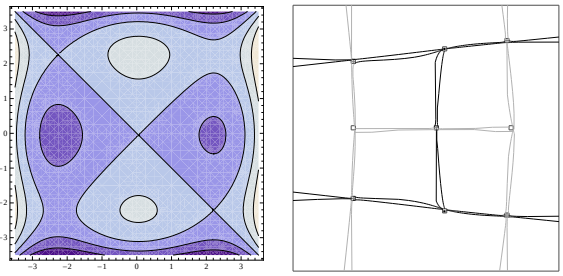
Separatrices that intersect $\partial\mathbb{D}$ do so transversally, cf Remark 1. Therefore, after a finite number of subdivision steps, both fences around such separatrices will intersect $\partial\mathbb{D}$ transversally. Hence, eventually all funnels become disjoint, at which point the algorithm terminates after returning a topologically correct MS-complex for ∇h.

7 Implementation and experimental results

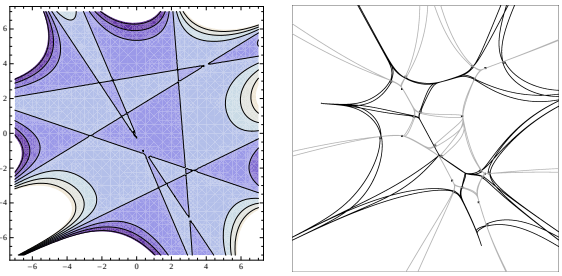
The algorithm has been implemented using the `Boost` library [1] for IA. All experiments have been performed on a 3GHz Intel Pentium 4 machine under Linux with 1 GB RAM using the `g++` compiler, version 3.3.5. Figures 1(a)-(b) and 7(a)-(b) depict the output of our algorithm, for several Morse-Smale functions. In our implementation the parameter θ, used in the construction of separatrix-funnels, is $\frac{\pi}{30}$, which is larger than the theoretical bound given by Lemma 6. The algorithm halves this angle several times, depending on the input, until the funnels are simply connected, mutually disjoint, and connect a saddle-box to a source-box (for stable separatrices) or sink-box (for unstable separatrices), in which case a topologically correct MS-complex has been computed. Each of the funnels with deep black boundaries contains an unstable separatrix, whereas a funnel with light black boundaries contains a stable separatrix. The CPU-time for computing a MS-system increases with the number of critical points and the complexity of the vector field, as indicated in the captions of the figures.

References

[1] `Boost` interval arithmetic library. `http://www.boost.org`.



(a) Contour plot (left) and MS-Complex (right) of $h(x,y) = -10\,x^2 + x^4 + 10\,y^2 - y^4 + x + xy^2$, on the box $[-4, 3.5] \times [-4, 3.5]$. CPU-time: 11 seconds.



(b) Contour plot (left) and MS-complex (right) of a product of seven linear functions, on the box $[-7, 7] \times [-7, 7]$. CPU-time: 11.5 minutes.

Figure 7: Contour plots of MS-functions and their Morse-Smale complexes.

[2] V.I. Arnol'd. *Ordinary Differential Equations*. Universitext. Springer-Verlag, New York, Heidelberg, Berlin, 2006.

[3] C.L. Bajaj and D.R. Schikore. Topology preserving data simplification with error bounds. *Comput. Graph.*, 22(1):3–12, 1998.

[4] T. F. Banchoff. Critical points and curvature for embedded polyhedral surfaces. *Amer. Math. Month.*, 77:475–485, 1970.

[5] S. Biasotti, L. De Floriani, B. Falcidieno, P. Frosini, D. Giorgi, C. Landi, L. Papaleo, and M. Spagnuolo. Describing shapes by geometrical-topological properties of real functions. *ACM Computing Surveys*, 40(4):12.1–12:87, 2008.

[6] F. Cazals, F. Chazal, and T. Lewiner. Molecular shape analysis based upon the Morse-Smale complex and the Connolly function. In *In SCG 2003: Proceedings of the 19th Annual Symposium on Computational Geometry*, pages 351–360, ACM Press, New York, NY, 351-360, 2003.

[7] H. Edelsbrunner, J. Harer, V. Natarajan, and V. Pascucci. Morse-Smale complexes for piecewise linear 3-manifolds. In *Proc. 19th Ann. Sympos. Comput. Geom.*, pages 361–370, 2003.

[8] H. Edelsbrunner, J. Harer, and A. Zomorodian. Hierarchical Morse-Smale complexes for piecewise linear 2-manifolds. *Discrete Comput. Geom*, 30:87–107, 2003.

[9] R. Forman. Morse theory for cell complexes. *Adv. Math.*, 134:90–145, 1998.

[10] A. Gyulassy, P. Bremer, B. Hamann, and V. Pascucci. A practical approach to Morse-Smale complex computation. *IEEE Transactions on Visualization and Computer*, 14:1619–1626., 2008.

[11] J. L. Helman and L. Hesselink. Visualizing vector field topology in fluid flows. *IEEE Computer Graphics and Applications*, 11(3):36–46, 1991.

[12] M. W. Hirsch and S. Smale. *Differential Equations, Dynamical Systems, and Linear Algebra*. Academic Press, 1974.

[13] J.H. Hubbard and B.H. West. *Differential Equations. A Dynamical Systems Approach. Part I*, volume 5 of *Texts in Applied Mathematics*. Springer Verlag, New York, Heidelberg, Berlin, 1991.

[14] C. Li, S. Pion, and C. Yap. Recent progress in Exact Geometric Computation. *J. of Logic and Algebraic Programming*, 64(1):85–111, 2004. Special issue on "Practical Development of Exact Real Number Computation".

[15] Long Lin and Chee Yap. Adaptive isotopic approximation of nonsingular curves: the parametrizability and non-local isotopy approach. In *Proc. 25th ACM Symp. on Computational Geometry*, pages 351–360, June 2009. Aarhus, Denmark, Jun 8-10, 2009. To Appear, Special Issue of SoCG 2009 in DCG.

[16] F. Meyer. Topographic distance and watershed lines. *Signal Process.*, 38:113–125, 1994.

[17] J. Milnor. *Morse Theory.* Princeton University Press, 1968. Based on lecture notes by M. Spivak and R. Wells.

[18] R.E. Moore. *Interval Analysis.* Prentice-Hall., 1996.

[19] J. Palis and W. de Melo.

[20] Simon Plantinga and Gert Vegter. Isotopic approximation of implicit curves and surfaces. In *Proc. Eurographics Symposium on Geometry Processing*, pages 245–254, New York, 2004. ACM Press.

[21] J.M. Snyder. *Generative Modeling for Cimputer Graphics and CAD. Symbolic Shape Design Using Interval Analysis.* Academic Press Professional, Inc., San Diego, CA, USA, 1992.

[22] J.M. Snyder. Interval analysis for computer graphics. *SIGGRAPH Comput. Graph.*, 26(2):121–130, 1992.

[23] S. Takahashi, T. Ikeda, Y. Shinagawa, and I. Fujishiro. Algorithms for extracting correct critical points and constructing topological graphs from discrete geographic elevation data. *Comput. Graph. For.*, 14(3):181–192, 1995.

[24] Chee K. Yap. In praise of numerical computation. In Susanne Albers, Helmut Alt, and Stefan Näher, editors, *Efficient Algorithms*, volume 5760 of *Lecture Notes in Computer Science*, pages 308–407. Springer-Verlag, 2009. Essays Dedicated to Kurt Mehlhorn on the Occasion of His 60th Birthday.

[25] E. Zhang, K. Mischaikow, and G. Turk. Vector field design on surfaces. *ACM Transactions on Graphics*, 25(4):1294–1326, 2006.

APPENDIX A: Preliminaries

¶12. Morse function. A function $h : \mathcal{D} \subset \mathbb{R}^2 \to \mathbb{R}$ is called a *Morse function* if all its critical points are non-degenerate. The *Morse lemma* [17] states that near a non-degenerate critical point a it is possible to choose local co-ordinates x, y in which h is expressed as $h(x, y) = h(a) \pm x^2 \pm y^2$. Existence of these local co-ordinates implies that non-degenerate critical points are isolated. The number of minus signs is called the *index* $i_h(a)$ of h at a. Thus a two variable Morse function has three types of non-degenerate critical points: minima (index 0), saddles (index 1) and maxima (index 2).

¶13. Integral curves. An integral curve $\mathbf{x} : I \subset \mathbb{R} \to \mathcal{D}$ passing through a point p_0 on $\mathcal{D}$ is a unique maximal curve satisfying:

$$\begin{cases} \dot{\mathbf{x}}(t) = & \nabla h(\mathbf{x}(t)) \\ \mathbf{x}(0) = & p_0 \end{cases} \tag{2}$$

for all $t \in I$. Integral curves corresponding to the gradient vector field of a smooth function $h : \mathcal{D} \to \mathbb{R}$ has the following properties:

1. Two integral curves are either disjoint or same.
2. The integral curves cover all the points of $\mathcal{D}$.
3. The integral curves of the gradient vector field of h form a partition of $\mathcal{D}$.
4. The integral curve $\mathbf{x}(t)$ through a critical point p_0 of h is the constant curve $\mathbf{x}(t) = p_0$.
5. The integral curve $\mathbf{x}(t)$ through a regular point p of h is injective, and if $\lim_{t\to\infty} \mathbf{x}(t)$ or $\lim_{t\to-\infty} x(t)$ exists, it is a critical point of h. This implies integral curves corresponding to gradient vector field are never closed curves.
6. The function h is strictly increasing along the integral curve of a regular point of h.
7. Integral curves are perpendicular to the regular level sets of h.

¶14. Stable and unstable manifolds. Consider the integral curve $x(t)$ passing through a point p. If the limit $\lim_{t\to\infty} \mathbf{x}(t)$ exists, it is called the ω-limit of p and is denoted by $\omega(p)$. Similarly, $\lim_{t\to-\infty} \mathbf{x}(t)$ is called the α-limit of p and is denoted by $\alpha(p)$ – again provided this limit exists. The stable manifold of a singular point p is the set $W^s(p) = \{q \in \mathcal{D} \mid \omega(q) = p\}$. Similarly, the unstable manifold of a singular point p is the set $W^u(p) = \{q \in \mathcal{D} \mid \alpha(q) = p\}$. Here we note that both $W^s(p)$ and $W^u(p)$ contain the singular point p itself [12].

Now, the stable and unstable manifolds of a saddle point are 1-dimensional manifolds. A stable manifold of a saddle point consists of two integral curves converging to the saddle point. Each of these integral curves (not including the saddle point) are called the stable separatrices of the saddle point. Similarly, an unstable manifold of a saddle point consists of two integral curves diverging from the saddle point and each of these integral curves (not including the saddle point) are called the unstable separatrices of the saddle point.

¶15. The Morse-Smale complex. A Morse function on $\mathcal{D}$ is called a Morse-Smale (MS) function if its stable and unstable separatrices are disjoint. In particular, a Morse-Smale function on a two-dimensional domain has no integral curve connecting two saddle points, since in that case a stable separatrix of one of the saddle points would coincide with an unstable separatrix of the other saddle

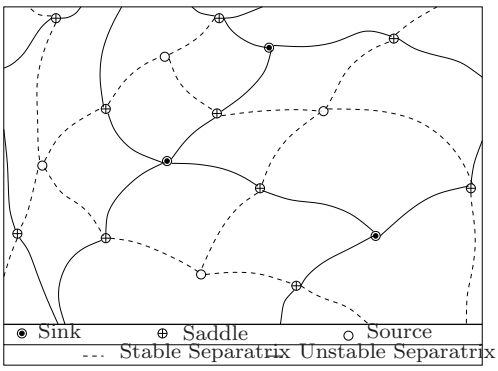


Figure 8: Morse-Smale complex

point. The MS-complex associated with a MS-function h on $\mathcal{D}$ is the subdivision of $\mathcal{D}$ formed by the connected components of the intersections $W^s(p) \cap W^u(q)$, where p, q range over all singular points of h.
The regions of the MS-complex are characterized by the following result.

Lemma 7 (Quadrangle Lemma ([8]))**.**
If $\mathcal{D} = \mathbb{R}^2$, then each region of the MS-complex is a quadrangle with vertices of index $0, 1, 2, 1$, in this order around the region.

¶16. Stability of equilibrium points. We note that a gradient vector field of a MS-function $h : \mathcal{D} \to \mathbb{R}$ can have three kinds of equilibria or *singular points*, namely, sinks (corresponding to maxima of h), saddles (saddles of h) and sources (corresponding to minima of h). These singular points can be distinguished based on the local behavior of the integral curves around those points. Locally, a sink has a neighborhood, which is a stable 2-manifold. Similarly, locally a source has a neighborhood, which is an unstable 2-manifold. Locally, a saddle has a stable 1-manifold and an unstable 1-manifold crossing each other at the saddle point. A sink is called a stable equilibrium point, where as a source or a saddle is called unstable equilibrium point. We note that, a source corresponding to a MS-function h is a sink corresponding to the function $-h$.

Interval implicit function theorem

To introduce a useful tool from IA, we recall some notation for *interval matrices*. An $n \times n$ interval matrix $\square M$ is defined as

$$\square M = \{M | M_{ij} \in \square M_{ij}, i, j \in \{1 \dots n\}\}$$

Also note that we write

$$0 \notin \det \square M$$

if there exists no matrix $M \in \square M$ such that $\det M = 0$.
Consider a system of equations:

$$f(x, y) = 0, \quad g(x, y) = 0, \tag{3}$$

and let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be the function with components f and g.
Let $I = I_x \times I_y$ be a 2D-interval (box) in $\mathbb{R}^2$. Define the *interval Jacobian determinant* $\square J_{x,y} F$ as the 2×2 interval determinant given by

$$\square J_{x,y} F(I) = \begin{vmatrix} \square \dfrac{\partial f}{\partial x}(I) & \square \dfrac{\partial f}{\partial y}(I) \\ \square \dfrac{\partial g}{\partial x}(I) & \square \dfrac{\partial g}{\partial y}(I) \end{vmatrix}$$

The following theorem gives a necessary condition guaranteeing that there is at most one solution of the system (3) in I.

Theorem 8. *(**Interval Implicit Function Theorem, Snyder [21, 22]**) Let $F : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a C^1-map. If $I \subset \mathbb{R} \times \mathbb{R}$ is a box for which*

$$0 \notin \square J_{x,y} F(I),$$

then the system (3) has at most one solution in I.

APPENDIX B: The constants in Lemma 6

A gridbox is called *quasihorizontal* (*quasivertical*) if it contains a point at which ∇h is quasihorizontal (quasivertical). Again, a gridbox may be both quasihorizontal and quasivertical.

An integral curve of ∇h in a quasihorizontal gridbox $[x_0, x_1] \times [y_0, y_1]$ is the graph of a function $x \mapsto y(x)$, where $y(x)$ is a solution of the differential equation

$$\begin{aligned} y'(x) &= F(x, y(x)), \\ y(x_0) &= y_0, \end{aligned} \tag{4}$$

where $F(x, y) = \dfrac{h_y(x, y)}{h_x(x, y)}$. Here x ranges over the full interval $[x_0, x_1]$ if $y_0 \le y(x) \le y_1$. Otherwise, the range of x is restricted to a suitable maximal subinterval $[\xi_0, \xi_1]$, such that $(\xi_0, y(\xi_0))$ and $(\xi_1, y(\xi_1))$ are points on the boundary of the gridbox. Similarly, a trajectory of X_ϑ in a quasihorizontal gridbox $[x_0, x_1] \times [y_0, y_1]$ is the graph of a function $x \mapsto y(x)$, where $y(x)$ is a solution of the differential equation

$$\frac{dy}{dx} = F_\vartheta(x, y), \tag{5}$$

with

$$F_\vartheta(x, y) = \frac{h_x(x, y)\, \sin\vartheta + h_y(x, y)\, \cos\vartheta}{h_x(x, y)\, \cos\vartheta - h_y(x, y)\, \sin\vartheta}.$$

Similarly, a trajectory of $X_{-\vartheta}$ is the graph of a function $y \mapsto x(y)$, where $x(y)$ is a solution of the differential equation

$$\frac{dx}{dy} = G_\vartheta(x, y),$$

with

$$G_\vartheta(x, y) = \frac{1}{F_\vartheta(x, y)} = \frac{h_x(x, y)\, \cos\vartheta - h_y(x, y)\, \sin\vartheta}{h_x(x, y)\, \sin\vartheta + h_y(x, y)\, \cos\vartheta}.$$

Here y ranges over the full interval $[y_0, y_1]$ if $x_0 \le x(y) \le x_1$, or a suitable maximal subinterval otherwise.

The union of all quasihorizontal gridboxes in $\mathbb{D}^*$ is denoted by ${\mathbb{D}^*}_{\mathrm{qh}}$, and the union of all quasivertical gridboxes by ${\mathbb{D}^*}_{\mathrm{qv}}$.

Even though the width of a funnel may grow exponentially in the number of grid boxes it traverses, this growth is controlled. To this end, we introduce several *computable* constants that only depend on the function h and (the size of) its domain $\mathbb{D}^*$. Let A_{qh}, A_{qv}, B_{qh}, B_{qv}, C_{qh} and C_{qv} be positive constants such that

$$\max_{(x,y)\in{\mathbb{D}^*}_{\mathrm{qh}}} |F(x, y)| \le A_{\mathrm{qh}}, \qquad \max_{(x,y)\in{\mathbb{D}^*}_{\mathrm{qv}}} |G(x, y)| \le A_{\mathrm{qv}},$$

$$\max_{(x,y)\in{\mathbb{D}^*}_{\mathrm{qh}}} \left|\frac{\partial F}{\partial x}(x, y)\right| \le B_{\mathrm{qh}}, \qquad \max_{(x,y)\in{\mathbb{D}^*}_{\mathrm{qv}}} \left|\frac{\partial G}{\partial y}(x, y)\right| \le B_{\mathrm{qv}},$$

$$\max_{(x,y)\in{\mathbb{D}^*}_{\mathrm{qh}}} \left|\frac{\partial F}{\partial y}(x, y)\right| \le C_{\mathrm{qh}}, \qquad \max_{(x,y)\in{\mathbb{D}^*}_{\mathrm{qv}}} \left|\frac{\partial G}{\partial x}(x, y)\right| \le C_{\mathrm{qv}}.$$

Note that

$$F_\vartheta(x,y) - F(x,y) = \frac{(h_x(x,y)^2 + h_y(x,y)^2)\,\sin\vartheta}{h_x(x,y)^2\,\cos\vartheta - h_x(x,y)\,h_y(x,y)\,\sin\vartheta}.$$

Let $M^{(1)}_{\mathrm{qh}}$ be a dyadic number such that

$$\max_{(x,y)\in\mathbb{D}^*{}_{\mathrm{qh}}} \left|\frac{h_y(x,y)}{h_x(x,y)}\right| \le M^{(1)}_{\mathrm{qh}}. \tag{6}$$

Take $\vartheta_{\mathrm{qh}} \in (0, \frac{1}{2}\pi)$ such that $\tan\vartheta_{\mathrm{qh}} \le \dfrac{1}{2M^{(1)}_{\mathrm{qh}}}$. Finally, let $M^{(2)}_{\mathrm{qh}}$ be a constant such that

$$\max_{(x,y)\in\mathbb{D}^*{}_{\mathrm{qh}}} \left|\frac{h_x(x,y)^2 + h_y(x,y)^2}{h_x(x,y)^2\cos\vartheta_{\mathrm{qh}}}\right| \le M^{(2)}_{\mathrm{qh}}. \tag{7}$$

Taking $M_{\mathrm{qh}} = \dfrac{M^{(2)}_{\mathrm{qh}}}{2M^{(1)}_{\mathrm{qh}}}$, we have, for $|\vartheta| \le \vartheta_{\mathrm{qh}}$:

$$\max_{(x,y)\in\mathbb{D}^*{}_{\mathrm{qh}}} |F_\vartheta(x,y) - F(x,y)| \le M_{\mathrm{qh}}\sin\vartheta. \tag{8}$$

Similarly, there are (computable) constants M_{qv} and ϑ_{qv} such that

$$\max_{(x,y)\in\mathbb{D}^*{}_{\mathrm{qv}}} |G_\vartheta(x,y) - G(x,y)| \le M_{\mathrm{qv}}\sin\vartheta, \tag{9}$$

for $|\vartheta| \le \vartheta_{\mathrm{qv}}$. Finally, let the constants c_0, c_1 and ϑ_0 be defined by

$$c_0 = 2\max(C_{\mathrm{qh}} + A_{\mathrm{qh}}B_{\mathrm{qh}}, C_{\mathrm{qv}} + A_{\mathrm{qv}}B_{\mathrm{qv}}) \tag{10}$$

$$c_1 = \max(\frac{1}{2M_{\mathrm{qh}}K_{\mathrm{qh}}}, \frac{1}{2M_{\mathrm{qv}}K_{\mathrm{qv}}}) \tag{11}$$

$$\vartheta_0 = \max(\vartheta_{\mathrm{qh}}, \vartheta_{\mathrm{qv}}). \tag{12}$$

The next result provides an upper bound for the growth of the funnel width along a quasihorizontal part of its bounding polylines. We assume that the funnel runs from left to right, so its initial points are on the line with smallest x-coordinate. If the funnel runs from right to left, a similar result is obtained.

Lemma 9. *Let $y_{\vartheta,w}, y_{-\vartheta,w} : [a,b] \to [c,d]$ be the piecewise linear functions the graphs of which are quasihorizontal parts of the polylines L_ϑ and $L_{-\vartheta}$ for a grid with edge length w, respectively. Let Δ be an upper bound for the distance of the initial points of these polylines, i.e.,*

$$|y_{\vartheta,w}(a) - y_{-\vartheta,w}(a)| \le \Delta.$$

Then the width of the fence, bounded by L_ϑ and $L_{-\vartheta}$, is bounded:

$$|y_{\vartheta,w}(x) - y_{-\vartheta,w}(x)| \le \Delta\, e^{C_{\mathrm{qh}}(x-a)} + (c_0 w + c_1\sin\vartheta)\,\frac{e^{C_{\mathrm{qh}}(x-a)} - 1}{C_{\mathrm{qh}}}.$$

Proof. Let $y_{\pm\vartheta}(x)$ be the exact solution of the rotated system with initial condition $y_{\pm\vartheta}(a)$. In particular, $|y_\vartheta(a) - y_{-\vartheta}(a)| \le \Delta$. Then (8) implies

$$\left|\frac{dy_{\pm\vartheta}}{dx}(x) - F(x, y_{\pm\vartheta}(x))\right| = \left|F_{\pm\vartheta}(x, y_{\pm\vartheta}(x) - F(x, y_{\pm\vartheta}(x))\right| \le M_{\mathrm{qh}}\sin\vartheta.$$

Therefore, according to the Fundamental Inequality [13, Theorem 4.4.1] we have

$$|y_\vartheta(x) - y_{-\vartheta}(x)| \le \Delta\, e^{C_{\rm qh}(x-a)} + \frac{2M_{\rm qh}\sin\vartheta}{C_{\rm qh}}\,(e^{C_{\rm qh}(x-a)} - 1). \tag{13}$$

The interval $[a,b]$ is subdivided into a finite number of subintervals of length at most w, where the endpoints correspond to the x-coordinates of the breakpoints of the fences L_ϑ and $L_{-\vartheta}$. Let $y_{\vartheta,w}$ be the Euler approximation to the ordinary differential equation (5). Its graph is (a quasihorizontal) part of the fence L_ϑ. Theorem 4.5.2 in [13] gives the following explicit bound for the error in Euler's method:

$$|y_{\vartheta,w}(x) - y_\vartheta(x)| \le w\,\frac{B_{\rm qh} + A_{\rm qh}C_{\rm qh}}{C_{\rm qh}}\,(e^{C_{\rm qh}(x-a)} - 1). \tag{14}$$

We get a similar inequality for the Euler approximation $y_{-\vartheta,w}$ of $y_{-\vartheta}$. Combining (13) and (14), and using (10) and (11), yields

$$\begin{aligned}|y_{\vartheta,w}(x) - y_{-\vartheta,w}(x)| &\le \Delta\, e^{C_{\rm qh}(x-a)} + \\ &\qquad 2(w(B_{\rm qh} + A_{\rm qh}C_{\rm qh}) + M_{\rm qh}\sin\vartheta)\,\frac{e^{C_{\rm qh}(x-a)} - 1}{C_{\rm qh}} \\ &= \Delta\, e^{C_{\rm qh}(x-a)} + (c_0 w + c_1\sin\vartheta)\,\frac{e^{C_{\rm qh}(x-a)} - 1}{C_{\rm qh}}.\end{aligned}$$

□

A similar result holds for quasivertical trajectories. Next we need to control the increase of the funnel width upon transition from a quasihorizontal to a quasivertical part its bounding polylines (or from a quasivertical to a quasihorizontal part).

¶17. Transitions: bounded increase of funnel width. Transition from a quasihorizontal to a quasivertical, or from a quasivertical to a quasihorizontal part of the funnel takes place at an ε-box. If the width of the funnel at the 'entry' of the box is less than the width w of a grid box, then the width may increase, but it will not be greater than $2w$ at the exit. This is made more precise by the following result.

Lemma 10. *Let $\mathbb{J}$ be a ε-box, where, moreover, the initial points p and q of $L_\vartheta \cap \mathbb{J}$ and $L_{-\vartheta} \cap \mathbb{J}$, respectively, are on the boundary of the gridboxes containing the vertices of edge e of $\mathbb{J}$. If the distance between p and q is at least w, then the distance between the terminal points $\overline{p}$ and $\overline{q}$ of $L_\vartheta \cap \mathbb{J}$ and $L_{-\vartheta} \cap \mathbb{J}$, respectively, is less than the distance of p and q. If the distance between p and q is less than w, then the distance of $\overline{p}$ and $\overline{q}$ is at most $2w$.*

Proof. Assume that the first edge of L_ϑ is quasivertical, but not quasihorizontal. Edge e of $\mathbb{J}$ is then vertical. Assume that this polyline consists of a single edge, namely the line segment $p\overline{p}$. Let β_+ be the angle between $p\overline{p}$ and edge e, then $\arctan\frac{1}{2} \le \beta_+ \le \vartheta + \arctan\frac{1}{2}$. Let β_- be the angle between the line segment $q\overline{q}$ and edge e, then β_- is inbetween the smallest and largest slope of any edge of $L_{-\vartheta}$. Since the angle variation of X over $\mathbb{J}$ is less than $\frac{\pi}{20}$, the angle β_- is greater than $\beta_+ - \frac{\pi}{20}$. Let a be the distance of p to the nearest vertex of e, then $a \le w$. If $d \ge w$, the distance $\overline{d}$ between $\overline{p}$ and $\overline{q}$ satisfies

$$\begin{aligned}\overline{d} &= (d+a)\,\tan\beta_- - a\tan\beta_+ \\ &\le d\,\tan(\beta_+ + \tfrac{\pi}{20}) + a\,(\tan(\beta_+ + \tfrac{\pi}{20}) - \tan\beta_+) \\ &< \tfrac{3}{4}d + \tfrac{1}{4}a \\ &\le d,\end{aligned}$$

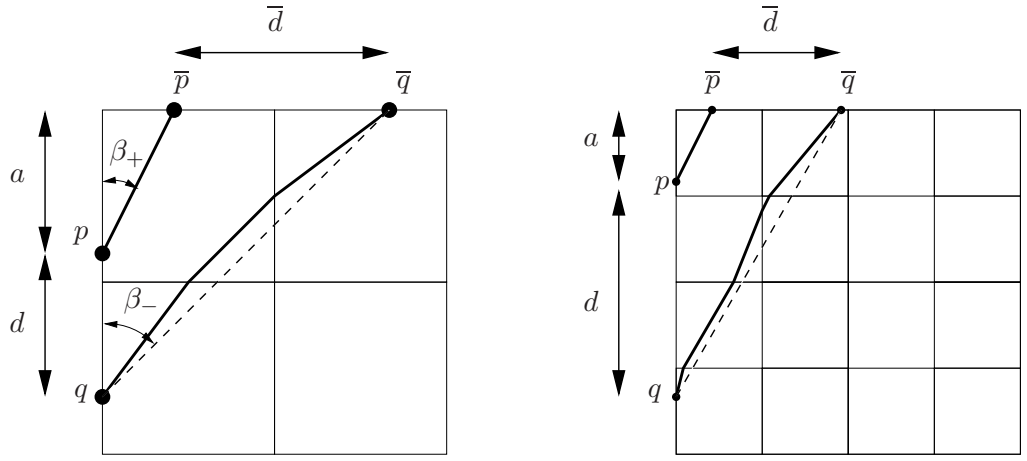


Figure 9: The distance between the two polylines upon entry and exit of a box. Left: If the distance d between the initial points p and q of the polylines is less than the edge-length w of a grid box, then the distance $\overline{d}$ between the terminal points $\overline{p}$ and $\overline{q}$ is less than $2w$. Right: Otherwise, the distance $\overline{d}$ between the terminal points is less than d.

since $a \le w \le d$. Here we used $\tan\beta_+ \le \frac{1}{2}$ to get

$$\tan(\beta_+ + \tfrac{\pi}{20}) = \frac{\tan\beta_+ + \tan\frac{\pi}{20}}{1 - \tan\beta_+ \tan\frac{\pi}{20}} \le \frac{\frac{1}{2} + \tan\frac{\pi}{20}}{1 - \frac{1}{2}\tan\frac{\pi}{20}} \le \tfrac{3}{4}.$$

Since $\arctan\frac{1}{2} - \frac{\pi}{40} \le \arctan\frac{1}{2} - \vartheta \le \beta_+ \le \arctan\frac{1}{2}$, a short computation shows that $\tan(\beta_+ + \frac{1}{20}\pi) - \tan\beta_+ < \frac{1}{4}$.

If $d < w$, then q lies in the same gridbox as p, or in a gridbox adjacent to it. Then it is easy to see that $\overline{p}$ lies in the same grid box as p, and $\overline{q}$ also lies in this box, or in a box adjacent to it. Therefore, $\overline{d} \le 2w$ in this case.
If $L_{-\vartheta}$ consists of a single edge, then the argument is similar. □

Lemmas 9 and 10 provide the following result on the upper bound on the funnel width of a separatrix with M transitions between quasihorizontal and quasivertical parts.

Corollary 11. *Let T be the maximum of the width and the height of the box $\mathbb{B}$ (the domain of the function h), and let M be the total number of quasihorizontal and quasivertical parts of the polylines bounding a separatrix funnel. Let $C = \max(C_{\mathrm{qh}}, C_{\mathrm{qv}})$ and let $D = \min(C_{\mathrm{qh}}, C_{\mathrm{qv}})$. Then the width of the funnel does not exceed*

$$(c_1 \sin\vartheta + c_2 w)\,\frac{e^{CMT} - 1}{D},$$

where $c_2 = 2 + \dfrac{c_0}{C}$, with c_0 and c_1 given by (10) and (11), respectively.

In particular, this width is at most ε if

$$c_1 \sin\vartheta + c_2 w \le \frac{D}{e^{CMT} - 1}\,\varepsilon. \tag{15}$$

Proof. Let $\mathbb{D}^* \subset [a,b] \times [c,d]$, then $T \le \max(b-a, d-c)$. There are $M-1$ transitions from quasihorizontal to quasivertical parts of the funnel, each occurring at an ε-box. Let Δ_0 be the width of the initial separatrix interval, and let $\Delta_1, \dots, \Delta_{M-1}$ be the width of the funnel at the entry of the corresponding boxes, in other words, Δ_k is the width at the end of the k-th part of the funnel. Using induction, we will prove that, for $k = 1, \dots, M$:

$$\Delta_k \le 2w\, e^{kCT} + \frac{c_0 w + c_1 \sin\vartheta}{D}\,(e^{kCT} - 1). \tag{16}$$

Since (16) for $k = M$ is the claim of the lemma, this is all we have to prove.

So assume (16) holds for $k = n-1$. If $\Delta_{n-1} > w$, the initial width of the n-th part of the funnel does not exceed Δ_{n-1}, cf Lemma 10. Assume that the n-th part of the funnel is quasihorizontal, then Lemma 9 implies that the width of this part at a point with horizontal coordinate x is at most

$$\Delta_{n-1}\, e^{C_{\rm qh}(x-a)} + (c_0 w + c_1 \sin\vartheta)\, \frac{e^{C_{\rm qh}(x-a)} - 1}{C_{\rm qh}},$$

so in particular, since $D \le C_{\rm qh} \le C$ and $0 \le x - a \le T$:

$$\Delta_n \le \Delta_{n-1}\, e^{AT} + \frac{c_0 w + c_1 \sin\vartheta}{D}\, (e^{CT} - 1).$$

Therefore, (16) holds for $k = n$. If $\Delta_{n-1} \le w$, then the initial width of the n-th part of the funnel is at most $2w$, cf Lemma 10. Therefore, Lemma 9 implies that the width of this part at a point with horizontal coordinate x is at most

$$2w\, e^{C_{\rm qh}(x-a)} + (c_0 w + c_1 \sin\vartheta)\, \frac{e^{C_{\rm qh}(x-a)} - 1}{C_{\rm qh}},$$

so in particular

$$\begin{aligned}\Delta_n &\le 2w\, e^{CT} + \frac{c_0 w + c_1 \sin\vartheta}{D}\, (e^{CT} - 1) \\ &\le 2w\, e^{nCT} + \frac{c_0 w + c_1 \sin\vartheta}{D}\, (e^{nCT} - 1),\end{aligned}$$

which is (16) for $k = n$. □