Computational Geometry 48 (2015) 311–332

Contents lists available at ScienceDirect

Computational Geometry: Theory and Applications

www.elsevier.com/locate/comgeo

Local, smooth, and consistent Jacobi set simplification

Harsh Bhatia [a,b,*], Bei Wang [a], Gregory Norgard [c], Valerio Pascucci [a], Peer-Timo Bremer [a,b]

[a] *Scientific Computing and Imaging (SCI) Institute, University of Utah, Salt Lake City, USA*
[b] *Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, Livermore, USA*
[c] *Numerica Corporation, Loveland, USA*

ARTICLE INFO

Article history:
Received 8 May 2013
Accepted 25 October 2014
Available online 31 October 2014

Keywords:
Jacobi set
Comparison of scalar functions
Simplification
Morse functions
Smoothness

ABSTRACT

The relation between two Morse functions defined on a smooth, compact, and orientable 2-manifold can be studied in terms of their Jacobi set. The Jacobi set contains points in the domain where the gradients of the two functions are aligned. Both the Jacobi set itself as well as the segmentation of the domain it induces, have shown to be useful in various applications. In practice, unfortunately, functions often contain noise and discretization artifacts, causing their Jacobi set to become unmanageably large and complex. Although there exist techniques to simplify Jacobi sets, they are unsuitable for most applications as they lack fine-grained control over the process, and heavily restrict the type of simplifications possible.

This paper introduces the theoretical foundations of a new simplification framework for Jacobi sets. We present a new interpretation of Jacobi set simplification based on the perspective of domain segmentation. Generalizing the cancellation of critical points from scalar functions to Jacobi sets, we focus on simplifications that can be realized by smooth approximations of the corresponding functions, and show how these cancellations imply simultaneous simplification of contiguous subsets of the Jacobi set. Using these extended cancellations as atomic operations, we introduce an algorithm to successively cancel subsets of the Jacobi set with minimal modifications to some user-defined metric. We show that for simply connected domains, our algorithm reduces a given Jacobi set to its *minimal configuration*, that is, one with no birth–death points (a birth–death point is a specific type of singularity within the Jacobi set where the level sets of the two functions and the Jacobi set have a common normal direction).

1. Introduction

In scientific modeling and simulation, one often defines multiple functions, e.g., temperature and pressure, species distributions, etc. on a common domain. Understanding the relation between such functions is crucial for data exploration and analysis. The *Jacobi set* [8] of two scalar functions provides an important tool for such analysis, as it describes points where the gradients of the two functions are aligned, and thus segments/partitions the domain into regions based on relative gradient orientation. A variety of interesting physical phenomena such as the interplay between salinity and temperature of water in oceanography [1] and the critical paths of gravitational potentials of celestial bodies [8] (similar to the Lagrange

* Corresponding author at: Scientific Computing and Imaging (SCI) Institute, University of Utah, Salt Lake City, USA.
E-mail address: hbhatia@sci.utah.edu (H. Bhatia).

http://dx.doi.org/10.1016/j.comgeo.2014.10.009
0925-7721/

points in astrophysics) can be modeled using Jacobi sets. In data analysis and image processing, Jacobi sets have been used to compare multiple scalar functions [13], as well as to express the paths of critical points over time [3,8], silhouettes of objects [17], and ridges in image data [33].

However, Jacobi sets can be extremely detailed, such that their complexity may impede or even prevent a meaningful analysis (e.g., refer to Fig. 1). Often, one is not interested in the fine-scale details, e.g., minor silhouette components due to surface roughness, but rather in more prevalent features such as significant protrusions. Jacobi sets are also highly sensitive to noise, which further leads to undesired artifacts such as small loops and zig-zag patterns. Finally, the most common algorithm to compute Jacobi sets [8,33] is designed for piecewise linear functions defined on triangulations, and is well known to introduce a large number of discretization artifacts that can skew the analysis. The natural answer to these problems is a controlled simplification of a Jacobi set by ranking and ultimately removing portions of it in order of importance.

Some previous techniques exist that can be broadly classified into direct and indirect Jacobi set simplification. *Indirect* techniques [3,25] simplify the underlying functions in a hope to obtain a structurally and geometrically simpler Jacobi set, which poses several problems. First, especially in the case of two nontrivial functions, changing either of them can introduce a large number of complex changes in the Jacobi set. These changes are difficult to predict and track, and the simplified Jacobi set is typically recomputed, which, however, can quickly become costly. Second, the Jacobi set encodes the relation between two functions and therefore simplifying one function may not actually simplify the Jacobi set. For example, two functions with complex gradient flows, which are similar in terms of relative orientation, define a small and simple Jacobi set. In this case, smoothing the gradient flow of either of the functions can introduce significant additional complexity into the Jacobi set. Finally, creating an appropriate metric to rank potential simplification steps can be challenging as small changes relative to traditional function norms, such as L_2 or L_∞, may induce large changes in the Jacobi set and vice versa.

Alternatively, *direct* simplification aims to identify and remove "unimportant" portions of the Jacobi set, and subsequently, to determine the necessary changes in the corresponding functions. Such techniques are designed to reduce the complexity of a Jacobi set measured by a user-defined metric. The first step [31] proposed in this direction views the Jacobi set as the zero level set of a complexity measure [13] and removes components of the level set (which correspond to loops of the Jacobi set) in order of their hyper-volume. However, this strategy is limited to removing entire loops of the Jacobi set, whereas much of the complexity of the Jacobi set is due to small undulations in the level sets of the functions causing zig-zag patterns (e.g., refer to Fig. 1). Such features are not addressed by a loop removal, which limits the usability of this approach, since one can find cases where loops should be combined rather than removed (e.g., refer to Fig. 14(d)).

While our proposed work also aims at "denoising" the Jacobi set, we focus on allowing more general and flexible changes during simplification. Our method could be classified as a *hybrid* technique that combines both direct simplification (i.e., by identifying and removing portions of the Jacobi set) and indirect simplification (i.e., by removing critical points of the underlying functions). The goal of our proposed work is to reduce a given Jacobi set to its minimal configuration, i.e., one that does not contain a specific type of singularity called birth–death points (where the level sets of the two functions and the Jacobi set have a common normal direction), while maintaining smoothness in the simplified functions. We present a new perspective to the domain segmentation created by the Jacobi set, based on which, the presented simplification procedure can be guided in a controlled manner, and performed hierarchically on portions of the Jacobi set that can be ranked by any user-defined importance metric. For example, in this work, we consider the gradient-based complexity measure [13] between the two functions as our ranking criteria.

Contributions To overcome the current limitations in Jacobi set simplification, we introduce the theoretical foundations of a new simplification framework for the Jacobi set of two Morse functions defined on a common smooth, compact, and orientable 2-manifold without boundary. By extending the notion of critical point cancellations in scalar fields to Jacobi sets, we perform simplifications that can be realized by smooth approximations of the corresponding functions. Based on a user-defined metric, we then rank these operations and progressively simplify the Jacobi set. Our framework provides

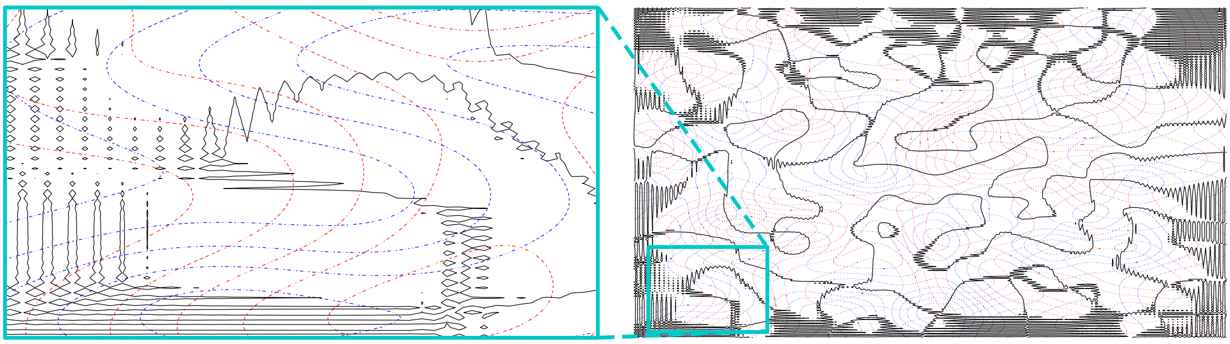


Fig. 1. Right: The Jacobi set (black) of two Morse functions (red and blue) can have a complex structure with a large number of loops segmenting the domain into fine regions. Left: A zoomed in view shows the noise (e.g., artifacts such as small loops and zig-zag patterns) in the structure that makes it difficult to perform any meaningful analysis.

fine-grained control over a very general set of possible simplifications and allows the combination of loops and the removal of zig-zag patterns in addition to the traditional loop removal. In particular,

- We present a new interpretation of Jacobi set simplification based on the perspective of domain segmentation. With simplification of this segmentation in mind, we introduce the notion of *local* pairings of points in the Jacobi set that can be canceled. These pointwise cancellations are then extended to contiguous subdomain bounded by segments of the Jacobi set, referred to as *Jacobi regions*, which are simplified simultaneously in a *consistent* manner. To obtain *smooth* realization of the simplification, the modification of Jacobi regions is extended to a collection of adjacent regions, referred to as *Jacobi sequences*. Each such sequence is a contiguous subdomain ranked by a user-defined metric and is simplified as one atomic operation;
- We introduce a simplification algorithm (or a class of simplification algorithms based on various ranking metrics) that constructs and successively cancels Jacobi sequences. Our approach cancels critical points of both functions, removes and/or combines loops, straightens the Jacobi set by removing zig-zag patterns, and always reduces the number of BD points;
- We show that for simply connected domains, our algorithm reduces a given Jacobi set to its minimal configuration (i.e., one without any BD points), whereas for non-simply connected domains, we discuss some fundamental challenges in Jacobi set simplification;
- We disprove a previous claim on the minimal Jacobi set for manifolds with arbitrary genus, and show that for domains with even genus there always exist function pairs that create a single loop in the Jacobi set.

This paper focuses on the theoretical foundations, construction, and properties of the algorithm and its required elements. The implementation details and other practical considerations associated with domains with boundary are forthcoming.

2. Background

This section presents the relevant background on Morse theory [26,28] and Jacobi sets [8]. In the following, let $\mathbb{M}$ be a smooth, compact, and orientable 2-manifold without boundary.

Morse function Given a smooth function $f : \mathbb{M} \to \mathbb{R}$, a point $x \in \mathbb{M}$ is called a *critical point* if the gradient of f at x equals zero ($\nabla f(x) = 0$), and the value of f at x is called a *critical value*. All other points are *regular points* with their function values being *regular values*. A critical point is *nondegenerate* if the Hessian, i.e., the matrix of second partial derivatives at the point, is invertible. Given a nondegenerate critical point p of f, the Morse lemma [26, Theorem 1.11, p. 8] states that f can be represented as a standard form by choosing appropriate local coordinates with p at the origin. The number of negative signs in this standard formulation is called the *index* of p. For functions of two variables, there exist three standard forms – one each for a minimum (index 0), a saddle (index 1), and a maximum (index 2).

A given smooth function f is a *Morse function* if (a) all its critical points are nondegenerate and (b) all its critical values are distinct. Since a Morse function can be modified without changing its critical points in such a way that the critical points of the modified Morse function take distinct values [26, Theorem 2.34, p. 69], one can also choose to exclude condition (b) from the definition of Morse functions, e.g., as the definition given by Matsumoto [26, Definition 2.15, p. 43]. However, for consistency, we adapt the definitions used in the initial paper on Jacobi set [8] and the book by Edelsbrunner [10, p. 153].

Critical point cancellation The Morse theory provides a way of simplifying Morse functions by canceling their critical points in pairs. According to Matsumoto [26, Theorem 2.30, p. 64], for a Morse function f on a compact manifold $\mathbb{M}$, there exists a gradient-like vector field[1] for f. Given a closed manifold $\mathbb{M}$, a Morse function f on $\mathbb{M}$, and a gradient-like vector field $\vec{V}$ for f, suppose that all critical points of f are arranged in ascending order of their critical values, i.e., $\cdots p_{i-1}, p_i \cdots$, where $f(p_{i-1}) < f(p_i)$. Then, the Morse cancellation theorem [30] (summarized by Milnor as the first cancellation theorem [29, Theorem 5.4, p. 48] and by Matsumoto [26, Theorem 3.28, p. 120]) states that by altering $\vec{V}$, it is possible to cancel two critical points p_{i-1} and p_i when they satisfy two conditions: (1) the index of p_i is one larger than the index of p_{i-1}; and (2) the boundaries of the *lower and upper disks*[2] of p_i and p_{i-1}, respectively, intersect transversely at a single point. For such a pair (p_{i-1}, p_i) of critical points, the Morse cancellation theorem guarantees that there exists a simpler Morse function $\tilde{f}$ that contains all critical points of f except p_i and p_{i-1}. For brevity, we call a pair of critical points that satisfies these two conditions a *cancelable pair*, as illustrated in Fig. 2.

A large number of pairing criteria (i.e., which introduce a particular ordering of cancelable pairs) have been explored to simplify Morse functions, e.g., local geometric measures [6,7], persistence [16], and data values [23,39]. These criteria can be applied to simplifications of topological structures of Morse functions (e.g., contour tree [6,7], Reeb graph [19,22,34], and

[1] A gradient-like vector field [26, p. 63] is a vector field $\vec{V}$ defined for a Morse function f such that: (1) away from the critical points of f, f increases in the direction of $\vec{V}$; and (2) in the local neighborhood of the critical points of f, $\vec{V} = \nabla f$ when f has the standard form.

[2] Informally, the lower/upper disk of a critical point p is the set of points whose integral lines in a gradient-like vector field converge to p as time goes to $+\infty/-\infty$. For a precise definition, the reader should refer to the book by Matsumoto [26, p. 112].

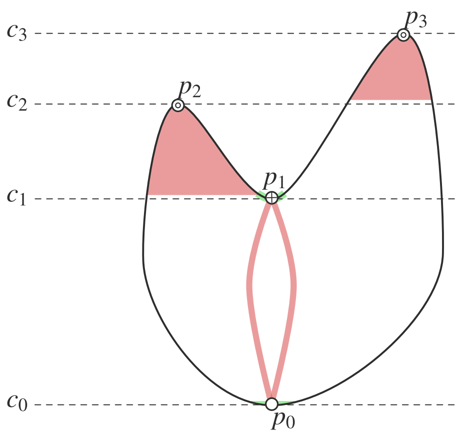


Fig. 2. Illustration of a cancelable pair for a height function f defined on a 2-manifold, where $c_i = f(p_i)$. The lower disks (red shaded regions) of the maxima (i.e., p_2 and p_3), saddle (i.e., p_1), and minimum (i.e., p_0) are diffeomorphic to 2-disks, 1-disks, and 0-disks, respectively, whereas their upper disks (green shaded regions) are diffeomorphic to 0-disks, 1-disks, and 2-disks, respectively. The pair (p_1, p_2) is cancelable since the boundary of the lower disk of p_2 and the boundary of the upper disk of p_1 intersect transversally at a single point. For the pair (p_3, p_1), the boundaries of their disks do not intersect, so they are not cancelable. Whereas, the pair (p_0, p_1) is not cancelable because the boundaries of their disks intersect at two points, respectively.

Morse–Smale complex [21]). In the Morse–Smale complex setting, the basic idea is that, by using the nondegeneracy conditions of Smale [38], one can modify a Morse–Smale function such that a pair of neighboring critical points collapse/cancel, resulting in a Morse function without these critical points. For example, a saddle can be canceled with a neighboring maximum/minimum by modifying a gradient-like vector field along its *ascending/descending 1-manifold*.[3]

Jacobi set Given a *generic pair*[4] of two Morse functions $f, g : \mathbb{M} \to \mathbb{R}$, their Jacobi set $\mathbb{J} = \mathbb{J}(f, g) = \mathbb{J}(g, f)$ is the closure of the set of points where their gradients are linearly dependent [8], i.e.,

$$\mathbb{J} = \mathrm{cl}\left\{x \in \mathbb{M} \;\middle|\; \nabla f(x) + \lambda \nabla g(x) = 0 \text{ or } \nabla g(x) + \lambda \nabla f(x) = 0\right\}. \tag{1}$$

See Fig. 3(a) for an example. The sign of λ for each x is called its *alignment*, as it defines whether the two gradients are aligned or anti-aligned. By definition, the Jacobi set contains the critical points of both f and g. Let $g^{-1}(t)$ represent the *level set* of g for $t \in \mathbb{R}$, and $f_t := f\,|_{g^{-1}(t)}: g^{-1}(t) \to \mathbb{R}$ represent the restriction of f on $g^{-1}(t)$. Then, the Jacobi set can equivalently be defined as the closure of the set of critical points of f_t for all regular values t of g [8],

$$\mathbb{J} = \mathrm{cl}\{x \in \mathbb{M} \mid x \text{ is a critical point of } f_t\}. \tag{2}$$

The critical points of f_t are also referred to as the *restricted critical points* of f (with respect to g). Symmetrically, $\mathbb{J} = \mathrm{cl}\{x \in \mathbb{M} \mid x \text{ is a critical point of } g_s\}$, where g_s is the restriction of g on the level set of g, $g_s := g\,|_{f^{-1}(s)}: f^{-1}(s) \to \mathbb{R}$.

Note that the Jacobi set of two functions is a 1-manifold embedded in $\mathbb{M}$ [8]. Furthermore, if $\mathbb{M}$ is a 2-manifold, generic level sets of f and g are 1-manifolds, and thus, the restricted functions f_t and g_s are 1D functions in general. Without loss of generality, the following discusses only the restricted function f_t. Note that f_t is a Morse function almost everywhere.[5] There exist three types of degeneracies where f_t is not Morse for some $t \in \mathbb{R}$: (a) t is a critical value of g, then the level set $g^{-1}(t)$ contains a singularity and is not a 1-manifold; (b) two or more critical points in f_t share the same function value; and (c) f_t contains an inflection point (a degenerate critical point). These degeneracies play an important role in our discussion of Jacobi set simplification. For example, each restricted critical point along $\mathbb{J}$ is an extrema of f_t for some $t \in \mathbb{R}$. As t varies, maxima and minima of f_t can approach each other and ultimately merge at an inflection point. In the context of Jacobi sets, the inflection point is called a *birth–death (BD) point*, illustrated in Fig. 3(a). Alternatively, traveling along $\mathbb{J}$, critical points of f_t switch their criticality (from maximum to minimum or vice versa) at BD points. Furthermore, the restricted functions f_t switch criticality at critical points of g (but not at critical points of f). Similarly, the alignment of restricted critical points switches at critical points of both f and g.

Comparison measure There exist several other descriptions of Jacobi sets [8,11,13,31]. One particularly useful description is in terms of the *comparison measure*, κ [13], which is a gradient-based metric to compare two functions. It plays a significant

[3] An ascending/descending manifold of a critical point p is the set of all points whose integral lines in a gradient-like vector field converge to p as time goes to $+\infty/-\infty$. Note that an ascending manifold surrounding a critical point is a super-set of its lower disk, and a descending manifold is a super-set of its upper disk.

[4] A pair of two functions whose critical points do not overlap.

[5] Given a function f_t, the set of points x where the function is not Morse is a finite set of measure zero.

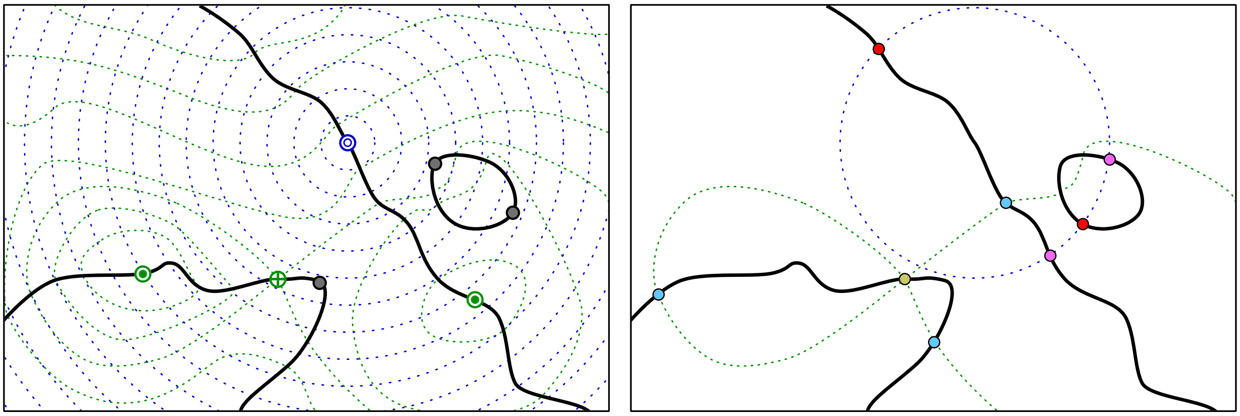


Fig. 3. The Jacobi set (solid black) of two functions (whose level sets are in blue and green) is the closure of the set of restricted critical points of the two functions. (a) The BD points are shown in gray and critical points in blue and green, respectively. (b) The level set neighbors of red points along the blue level set are marked in pink and vice versa. Along the green level set, there exist three level set neighbors of the saddle (yellow) and are marked in cyan.

role in assigning an importance to subsets of a Jacobi set in terms of the underlying functions f and g by measuring the relative orientation of their gradients. For a domain Ω,

$$\kappa = \kappa(\Omega) = \frac{1}{\mathrm{Area}(\Omega)} \int\limits_{x\in\Omega} \kappa_x \, \mathrm{d}x = \frac{1}{\mathrm{Area}(\Omega)} \int\limits_{x\in\Omega} \left\| \nabla f(x) \times \nabla g(x) \right\| \mathrm{d}x,$$

where, dx is the area element at x, and $\mathrm{Area}(\Omega) = \int_{x\in\Omega} \mathrm{d}x$. Here, $\kappa_x = \|\nabla f(x) \times \nabla g(x)\|$ represents the limit of κ to a single point, and the Jacobi set is its 0-level set [13,25,31].

Level set neighbors By definition, every point $v \in \mathbb{J}$ is a critical point of $f_t : g^{-1}(t) \to \mathbb{R}$ for some t. Given two critical points u and v of f_t, we refer to them as *level set neighbors* if along the corresponding level set $g^{-1}(t)$, they are situated next to each other. In other words, there exists an oriented curve between u and v along the level set $g^{-1}(t)$ such that no other critical points of f_t lie in its interior. For a given point $v \in \mathbb{J}$, its level set neighbors, denoted as $n_g(v)$, lie in $\mathbb{J} \cap g^{-1}(g(v))$. Level set neighbors are illustrated along two level sets intersecting with three curves of the Jacobi set in Fig. 3(b). Note that $u \in n_g(v)$ implies that $v \in n_g(u)$. Generically, $|n_g(v)| \le 2$, however, for an extremum of g, $|n_g(v)| = 0$, and for a saddle of g, $|n_g(v)| \le 4$. Such a definition can be extended to smooth curves in $\mathbb{J}$. Two smooth parametrized curves $\alpha, \beta : (a,b) \to \mathbb{M}$ in $\mathbb{J}$ are level set neighbors if $\alpha(t)$ and $\beta(t)$ are level set neighbors in $g^{-1}(t)$ for all $t \in (a,b)$. For simplicity in notations, for such level set neighbors, we choose a and b to be function values of g, i.e., $g(\alpha(a)) = a$ and $g(\alpha(b)) = b$. We further define their *bounded region*, denoted by $R_{(a,b)}(\alpha,\beta)$, as the open subset of $\mathbb{M}$ bounded by curves α, β, and level sets of g that pass through their end points, i.e., $g^{-1}(a)$ and $g^{-1}(b)$.[6] These constructs can be symmetrically defined with respect to the level sets of f.

3. Related work

The topology of scalar fields is usually described through constructs such as Reeb graph [36] and contour trees [5,7,40]. To understand the relation between multiple scalar fields, one can extract and compare these constructs from individual fields [23]. Alternatively, there exist techniques that define similar descriptors for multiple functions, such as Reeb space [14] and joint contour nets [4]. The focus of this paper is a third class of techniques that uses the Jacobi set, and is particularly useful to study the relation between two or more functions directly [13].

However, as illustrated in Fig. 1, the Jacobi set may contain a number of components that represent noise, degeneracies, or insignificant features in the data. As a result, Jacobi set simplification is both necessary and desirable. Bremer et al. [3] use the Jacobi set to track the critical points of a 2-dimensional time-varying function $f : \mathbb{M} \times \mathbb{R} \to \mathbb{R}$, where time is represented as $g : \mathbb{M} \times \mathbb{R} \to \mathbb{R}$ and $g(x,t) = t$. The Jacobi set $\mathbb{J} = \mathbb{J}(f,g)$ is therefore the trajectory of the critical points of f_t as time varies. To simplify the Jacobi set, Bremer et al. use the Morse–Smale complex [12,15,21] of f_t at discrete time-steps to pair critical points, cancel pairs below a persistence [9,10,16] threshold, and remove small components of the Jacobi set that lie entirely within successive time-steps. This method, however, is difficult to extend to a general setting for two reasons. First,

[6] In the case where $\alpha, \beta : (a,b) \to \mathbb{J}$ are subsets of some larger parametrized curves $\alpha', \beta' : (a',b') \to \mathbb{J}$, i.e., α, β are the restriction of α', β' to $(a,b) \subseteq (a',b')$, i.e., $\alpha = \alpha'|_{(a,b)}$ and $\beta = \beta'|_{(a,b)}$, we denote their bounded region as $R_{(a,b)}(\alpha',\beta')$.

only one function, f, is simplified and the other is assumed to be trivial. Second, only a small, discrete number of f_t are simplified and all intermediate changes are ignored. Tracking singularities of a function over time and determining the type of topological changes required for simplification is also addressed by Gingold and Zorin [20]. They identify the need for explicit control on generic topological changes during simplification and filtering, by working with standard filters such as Laplacian smoothing, sharpening, and anisotropic diffusion. The modifications to the function are then adjusted to prevent disallowed changes, which may depend upon the problem at hand.

Luo et al. [25] propose an algorithm to compute the Jacobi set of a point cloud representing surfaces embedded in $\mathbb{R}^3$. The Jacobi set is considered as the 0-level set of κ_x, which is computed by approximating the gradients ∇f and ∇g. Reducing the number of eigenvectors used in the gradient approximation, therefore, corresponds to a simpler Jacobi set after recomputation. This technique is the foremost example of an indirect simplification in which f and g are smoothed, leading to some (unpredictable) changes in $\mathbb{J}$. Instead, our approach aims at identifying and removing unimportant portions of $\mathbb{J}$ by determining how f and g must be modified correspondingly.

Nagaraj and Natarajan [31] consider the simplification of the Jacobi set as the reduction in the number of components of $\mathbb{J}$ with minimal change to the relationship between the two functions, quantified by κ_x. Considering surfaces embedded in $\mathbb{R}^3$, the authors construct the Reeb graph [37] of κ_x, and associate a percentage of κ as offset cost with each critical point and 0-level set point in the Reeb graph. A greedy strategy is then applied to modify components in the Jacobi set with low offset costs until a threshold is reached. However, this technique can remove only entire loops of $\mathbb{J}$, which significantly restricts its flexibility. For example, there exist cases where $\mathbb{J}$ is highly complex, yet contains a single loop. Fig. 4(b) shows such an example where no loops can be canceled to simplify $\mathbb{J}$.

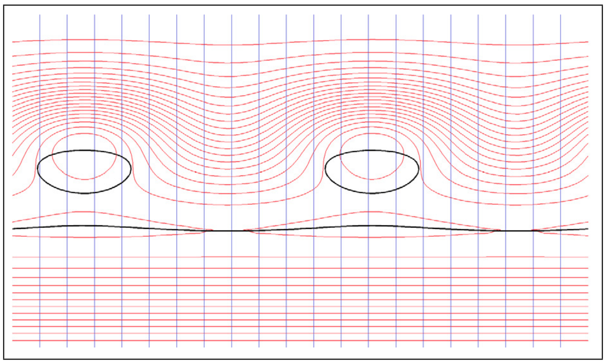
There is also a large amount of work on singularities of plane maps, which are related to Jacobi sets [18,24,27]. As pointed out by Edelsbrunner et al. [13], the interaction between two Morse functions $f, g : \mathbb{M} \to \mathbb{R}$ can be described by the Jacobian of a smooth function $\phi = (f, g) : \mathbb{M} \to \mathbb{R}^2$. In particular, the Jacobi set $\mathbb{J}$ is defined as the set of points where the Jacobian matrix J of partial derivatives of ϕ does not have full rank,

$$\mathbb{J} = \{x \in \mathbb{M} \mid \operatorname{rank}(J(x)) < 2\}. \tag{3}$$

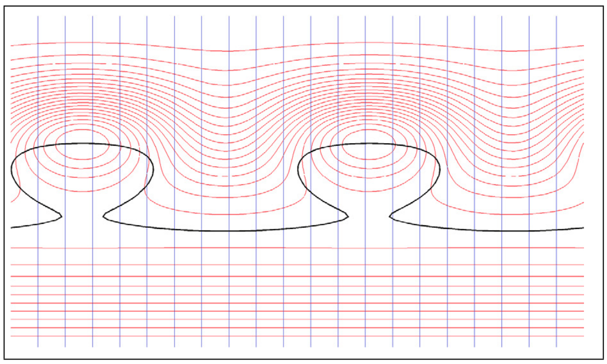
In other words, the Jacobi set corresponds to the singularities of plane maps in singularity theory. They consist of smooth disjoint curves called *folds* whose points have rank one, and discrete set of points on the fold curves called *cusp points* (i.e., these cusp singularities are the same as the birth–death points) whose rank restricted to ϕ is zero [27]. One interesting observation that may deserve further investigation is that our simplification algorithm described in the subsequent sections could potentially lead to a Jacobi set modification that coincides with three procedures for describing homotopies between generic maps that change the structures of the folds and cusps, namely, cancellation of two cusps by running them together, introduction of two cusps by a twist, and exchange of cusps, as detailed by Millett [27]. Furthermore, a direct consequence of [27, Theorem 1] is that for an oriented manifold $\mathbb{M}$, there exists a simplified Jacobi set without any BD points.

4. Jacobi set simplification – An overview

This paper introduces a new procedure for Jacobi set simplification. The proposed scheme removes a given set of points from $\mathbb{J}$ by understanding the required changes in f and/or g. The goal is to obtain a Jacobi set with fewer BD points, by making the gradients of the underlying functions more similar. In the following, we describe simplification of $\mathbb{J}$ that modifies f with respect to the level sets of g, but all concepts apply symmetrically to modifications of g with respect to f.



(a)



(b)

Fig. 4. (a) The Jacobi set (black) of two functions (whose level sets are in red and blue) contains three loops, and can be simplified using existing techniques [31]. (b) Adding a small perturbation to the red function creates a Jacobi set with a single loop, which cannot be simplified further using existing techniques. The above example suggests that existing simplification techniques are not robust against such perturbations and hence not suited for practical applications in data analysis due to the presence of noise and/or measurement errors.

In practice, we consider simplifications that modify either f or g, and typically interleave operations acting on one or the other.

Since the Jacobi set $\mathbb{J}(f,g)$ is defined as the closure of the restricted critical points of f, or the critical points of f_t for regular values $t \in \mathbb{R}$, where f_t is a 1D function, it is natural to simplify $\mathbb{J}$ by canceling critical points in f_t. Therefore, we remove pairs of restricted critical points to construct continuous simplified function $\overline{f_t}$, such that no other critical points of f_t are affected. To obtain a smooth approximation f_t^* of the simplified function $\overline{f_t}$, the modification can be extended to allow an ϵ-slope for the modified function. The region of influence of this cancellation is the region where $f_t \neq f_t^*$, as shown in green shaded region of Fig. 5(a). In order to perform these cancellations, we must first define a scheme for pairing restricted critical points. Section 5.1 discusses in detail the choice of our pairing scheme, and the procedure for carrying out such cancellations.

Although cancellation of restricted critical points produces smooth simplified restricted functions f_t^* as shown in Fig. 5(a), performing a single such cancellation creates a discontinuity across the level set $g^{-1}(t)$. In order to obtain smoothness across level sets, we must simultaneously cancel more than one contiguous pair of restricted critical points, called the *Jacobi regions*. Since understanding Jacobi regions requires understanding adjacent restricted functions, in the interest of brevity, we will denote the set of restricted functions f_t for $t \in [a,b]$ as $f_{[a,b]}$. For example, consider two Jacobi regions R_1 and R_2 existing between the level sets $g^{-1}(a)$, $g^{-1}(c)$, and $g^{-1}(b)$ as shown in Fig. 5(b) (left). These Jacobi regions represent contiguous pairs of restricted critical points of $f_{[a,c]}$ and $f_{[c,b]}$ shown as red and blue lines, respectively. A smooth simplification f^* that cancels all critical points in $f_{[a,c]}$ can be obtained by modifying f in the corresponding shaded region. The construction, properties, and cancellation of the Jacobi regions are discussed in detail in Section 5.2.

The cancellation of Jacobi regions, however, creates smooth functions only at the interior of the regions, and the discontinuities are pushed to their boundaries. In order to create globally smooth simplified functions f^*, we must further cancel a sequence of adjacent regions at the same time, e.g., canceling R_1 and R_2 at the same time as shown in Fig. 5(b) (right). We show that any discontinuities can be avoided by local modifications if these *Jacobi sequences* start and end with BD points, and discuss their construction and cancellation in Section 5.3.

In summary, the entire Section 5 focuses on a simplification scheme that extends the concept of critical point cancellation in scalar functions to Jacobi sets. The defining characteristic of a *valid simplification* is the removal of pairs of restricted critical points in $\mathbb{J}$ in a *local*, *smooth*, and *consistent* manner.

Definition 4.1 *(Valid simplification).* Let V be a set that contains n pairs of curves of the Jacobi set that are level set neighbors, i.e., $V = \{R_i\} = \{(\alpha_i, \beta_i)\}_{i=1}^n$, such that each $R_i = (\alpha_i, \beta_i)$ represents a pair of level set neighbors in $f_{[a_{i-1},a_i]} := \{f_t \mid t \in [a_{i-1}, a_i] \subseteq \mathbb{R}\}$. Removing V from $\mathbb{J}$ is considered a *valid simplification* if it is

1. *local*: There exists a continuous function $\overline{f}_{[a_0,a_n]} = \{\overline{f}_t\}$ for $t \in [a_0, a_n]$ containing all (restricted) critical points of $f_{[a_0,a_n]}$ except for the ones included in V;
2. *smooth*: There exists a smooth function $f_\epsilon^* : \mathbb{M} \to \mathbb{R}$ such that $\| f^*_{\epsilon\,[a_0,a_n]} - \overline{f}_{[a_0,a_n]} \|_\infty < \epsilon$, for any $\epsilon > 0$; and
3. *consistent*: $\mathbb{J}(f_\epsilon^*, g) = \mathbb{J}(f,g)$ for all x with $g(x) \in (-\infty, a_0) \cup (a_n, \infty)$, and $f_\epsilon^*(x) = f(x)$ for all x with $g(x) \in (-\infty, a_0 - \epsilon] \cup [a_n + \epsilon, \infty)$ for any $\epsilon > 0$.

According to this definition, given any $\epsilon > 0$, the corresponding simplified function must satisfy conditions 2 and 3 to qualify as a valid simplification. Furthermore, note that although the simplified function $f_\epsilon^*(x)$ is defined with respect to a given ϵ, since there must exist a valid simplified function for any $\epsilon > 0$, for brevity, the rest of this article omits the subscript, and denotes the simplified function as $f^*(x)$. Referring to Fig. 5(a), it is important to note that the locality conditions implies that the modification in f_t must not impact any restricted critical points other than u and v. In Fig. 5(b), this means that for any level set of g (vertical line), the red and blue shaded regions must not touch any portion of the Jacobi set other than the ones shown in red and blue, respectively. Notice that whereas locality is associated with continuous function $\overline{f_t}$, the second condition requires f^* to be smooth along as well as across level sets. In order to create such a smooth f^*, the locality condition must be relaxed within a small neighborhood around the canceled Jacobi region. Furthermore, the consistency condition requires that no portions of the Jacobi set outside $[a,b]$ are modified; thus, the

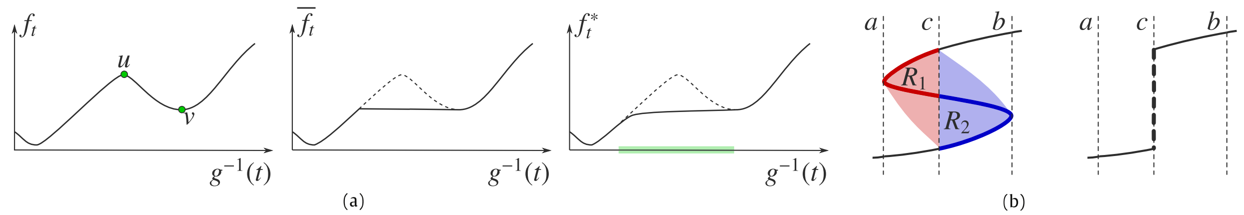


Fig. 5. (a) Cancellation of a pair of critical points in the 1-dimensional function f_t (left) gives a continuous simplified function $\overline{f_t}$ (middle). A smooth simplification can be obtained as f_t^* (right) where the cancellation modifies f_t in the shaded region (green) of its domain. (b) (Left) Valid simplification of R_1 and R_2 must construct smooth f^* by modifications in the corresponding shaded regions only. a, b, and c represent level sets of g. (Right) The simplified Jacobi set.

consistency condition implies locality across level sets. Whereas the locality condition is obtained by defining a special pairing function, a smooth and consistent simplification can be performed when the Jacobi sequence begins and ends with BD points (as in Fig. 5(b)). Also, notice that $\mathbb{J}(f^*, g) \not\subset \mathbb{J}(f, g)$, since new points (dashed line) may be added to the Jacobi set to connect the existing curves.

Unfortunately, as detailed in Section 5.3, the saddles of g present unresolvable discontinuities in the pairings that can obstruct the construction of Jacobi sequences. Consequently, the simplification scheme discussed above may not be able to progress. To handle such cases, we use a conventional critical point cancellation technique in 2D to cancel a saddle of g with its nearby maximum or minimum. As discussed in Section 6, this procedure simplifies (reduces the number of) alignment switches in the Jacobi set, while introducing only minor and tractable structural changes.

The technique discussed in Section 5 is a direct simplification where carefully identified pieces of a Jacobi set are removed. On the other hand, the critical point cancellation discussed in Section 6 is a type of indirect simplification. Using the techniques discussed in the two sections, Section 7 presents a hybrid algorithm for simplifying Jacobi sets that can be guided by an arbitrary metric. The simplification of a Jacobi sequence discussed in Section 5 always removes at least two BD points. On the other hand, although the critical point cancellation may not remove BD points, it facilitates formation of new Jacobi sequences, and thus, makes progress towards the final goal. We provide correctness proofs for the algorithm, and show that for simply connected domains, this algorithm obtains the minimum configuration of Jacobi sets. On the other hand, for non-simply connected domains, we discuss current challenges and list them as future work.

5. Cancellation of restricted critical points in f

This section details the procedure of canceling restricted critical points in $\mathbb{J}$ to obtain simplified functions. Starting with the simplification of 1D restricted functions, we discuss the cancellation of entire segments of $\mathbb{J}$ by canceling Jacobi sequences.

5.1. Pairing and cancellation of restricted critical points

Since restricted critical points of f_t must be canceled in pairs, we need a mechanism to define such pairings. The *topological persistence pairing* [16,41] seems an obvious choice, where critical points are paired and removed in order of *persistence*. However, since persistence pairing is assigned globally, restricted critical points, which are not level set neighbors, may be paired. These pairs cannot be canceled without violating the locality condition, which prevents most simplifications. Therefore, we instead use a localized variant of persistence pairing that guarantees that each point on the Jacobi set is paired with one of its level set neighbors as described below.

Given a nondegenerate restricted critical point $v \in f_t$ and its two level set neighbors $u, w \in n_g(v)$, the goal is to understand how f_t can be modified in a local neighborhood surrounding v, in order to cancel v with either u or w. Consider, e.g., v_3 shown in Fig. 6. One can lower v_3 to the level of v_4 canceling (v_3, v_4), but cannot lower it to the level of v_2 as this would impact v_4, and thus become a nonlocal simplification. In general, each restricted critical point can be canceled with only one of its level set neighbors in this fashion, and we call such a neighbor its *partner*.

Definition 5.1 *(Local pairing).* The relation between a restricted critical point and its partner can be described through a *local pairing function*, $\mu : \mathbb{J} \to \mathbb{J}$, such that for every $v \in \mathbb{J}$, its *partner* $\mu(v)$ is defined as

- v, if v is a degenerate critical point of f_t or a critical point of g; or
- an arbitrary element in the set $\{u \mid \mathrm{argmin}_{u \in n_g(v)} \|f_t(u) - f_t(v)\|\}$ otherwise.

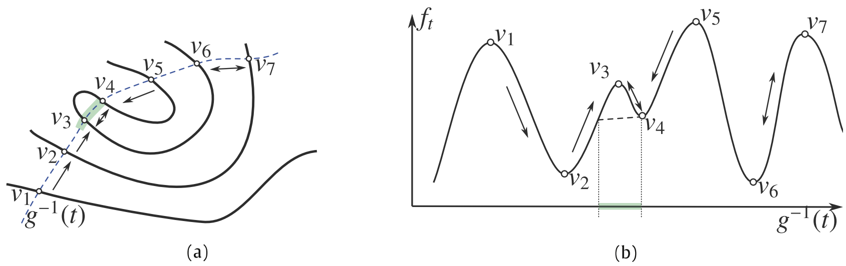


Fig. 6. Illustration of restricted critical points and local pairings. (a) A Jacobi set $\mathbb{J}$ (black solid lines) intersects a level set $g^{-1}(t)$ (blue dashed line), and (b) the corresponding restricted function f_t is shown. Local pairings among the restricted critical points in f_t are indicated by arrows. The pair (v_3, v_4) can be canceled by lowering the maximum v_3 to match the value of v_4 (black dashed line in (b)). For the cancellation, its region of influence along the level set is shown in green (in both (a) and (b)). It is the subset of domain of the function (i.e., the level set) where the function value is modified, which is shown as a thick region only for the purpose of illustration.

Intuitively, every nondegenerate restricted critical point v is paired with one of its level set neighbors u with minimal difference in function value. Then (v, u) is referred to as a *(local) pair*. Notice that, $\mu(v) = u$ does not imply $\mu(u) = v$. Traveling along a Jacobi curve, the discontinuities of $\mu(v)$ reflect a change in partner for v. Since BD points and extrema of g are paired to themselves, μ is continuous at such points. Fig. 6 indicates the pairings between restricted critical points as directed arrows pointing from v to its partner $\mu(v)$.

It can be verified that the local pair is a cancelable pair as defined in Section 2. Therefore, the Morse cancellation theorem guarantees that a simplification that removes the local pairs of critical points of f_t exists. We can perform such a cancellation by moving a critical point to the level of its partner to obtain a continuous simplified function $\overline{f_t}$, as shown in Figs. 5(a) and 6(b). Finally, an ϵ-slope can be introduced in $\overline{f_t}$, while still maintaining locality, to create a smooth and monotonic f_t^*. The Weierstrass approximation theorem guarantees that such a smooth f_t^* always exists for any $\epsilon > 0$. Consequently, for a pair (v, u), a cancellation where v is moved to the level of u always guarantees locality. Notice in Fig. 6(b) that the pair (v_5, v_6) could also be canceled locally by bringing both points to a function value between v_4 and v_7. In general, one can potentially bring both points to a common intermediate value for a local cancellation. However, such cancellations may not admit valid simplification steps for reasons explained in Section 5.2, and therefore are not considered. From now on, a cancellation induced by a pair (v, u) always implies a procedure that moves v to the level of u.

5.2. Construction and cancellation of Jacobi regions

The cancellation of a pair of restricted critical points creates a smooth restricted function f_t^*. However, the function f^* is still discontinuous across the level set $g^{-1}(t)$, since the neighboring restricted functions are unchanged. Hence, canceling a single pair of restricted critical points in isolation introduces unwanted discontinuities, and therefore violates the smoothness condition of a valid simplification. Instead, one can extend these cancellations to adjacent restricted functions, which, however violates the consistency condition by modifying f outside the interval $[t_0, t_0]$. For example, consider the scenario shown in Fig. 7. Canceling $(u, v) \in f_{t_0}$ creates a discontinuous simplified function. This modification can be extended to an adjacent region $f_{[t_0-\epsilon, t_0+\epsilon]}$, allowing the creation of a smooth function f^* at t_0 that cancels (u, v). However, since $\mathbb{J}$ is now modified beyond the level set $g^{-1}(t_0)$, it is no longer a consistent simplification.

Instead, we must cancel connected sets of neighboring restricted critical points that are paired "consistently". To understand their construction, we define *switch points* as the set of points in $\mathbb{J}$ where μ is not continuous, and *boundary points* as the points that are switch points, BD points, or critical points of g. Then, the Jacobi set $\mathbb{J}$ can be decomposed into a set of nonoverlapping *Jacobi segments*, which are maximal open subsets of $\mathbb{J}$ separated by boundary points. By definition, restricted critical points within the interior of Jacobi segments are *consistently paired*, meaning that μ is continuous at the interior of Jacobi segments. As a result, μ induces a pairing between segments. Finally, we define *image points* as the level set neighbors of boundary points. Together, the boundary points and the image points decompose the Jacobi set into pieces α_i that have mutually consistent pairing, meaning that μ is continuous at the interior of α_i as well as their respective partners. Given two such maximal subsets of Jacobi segments that are level set neighbors parametrized as $\alpha, \beta : (a, b) \to \mathbb{J}$, we call their bounded region $R_{(a,b)}(\alpha, \beta)$ a *Jacobi region*. Referring to Fig. 8(a), we point out that the inclusion of image points during this decomposition is important, as it ensures that the segments of a Jacobi region are consistently paired, since $\mu(x)$ is continuous for all $x \in (\alpha(a), \alpha(b))$ and their partners, all $x \in (\mu(\alpha(a)), \mu(\alpha(b)))$, i.e., all $x \in (\beta(a), \beta(b))$. Similar to the pointwise cancellation, the entire segment α can be moved to the level of β to cancel both the segments. Fig. 8(b) shows boundary and image points, Jacobi segments, and Jacobi regions as pairings between them for a typical Jacobi set.

There exist various classes of Jacobi regions with different implications for the simplification process. A Jacobi region is called *regular* if its closure does not contain BD points or critical points of g. Regular regions have four "corners" made up of two switch and two image points, e.g., R_5, R_8 in Fig. 8(b). With slight abuse of notation, we denote a corner as $\alpha(a) = \lim_{t \to a} \alpha(t)$. We further identify special but not mutually exclusive types of regions shown in Fig. 9: (a) *BD internal regions* where α and β share at least one BD point, i.e., $\alpha(a) = \beta(a)$ and/or $\alpha(b) = \beta(b)$; (b) *BD side region* where α and/or β are bounded by a BD point but $\alpha(x) \neq \beta(x)$, for all $x \in [a, b]$; (c) *BD external region* where the boundary of the region contains a BD point but neither α nor β does; (d) *Saddle region* where the boundary of the region contains a saddle of g but neither α nor β does; and (e) *Extremal region* containing an extremum of g.

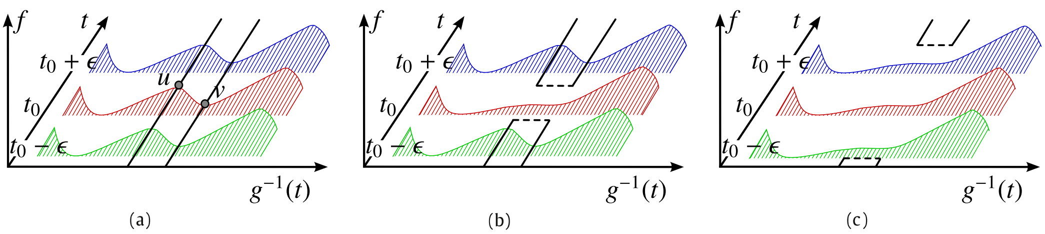


Fig. 7. Cancellation of a pair of restricted critical points $(u, v) \in f_t$. (a) The original f_t's and the Jacobi set (in black). (b) Canceling (u, v) in f_t in isolation creates a discontinuity across $t = t_0$, and hence is invalid. (c) Extending the cancellation to $f_{[t_0-\epsilon, t_0+\epsilon]}$ creates a smooth f^*, but the cancellation is inconsistent since $\mathbb{J}$ outside $[t_0, t_0]$ is modified.

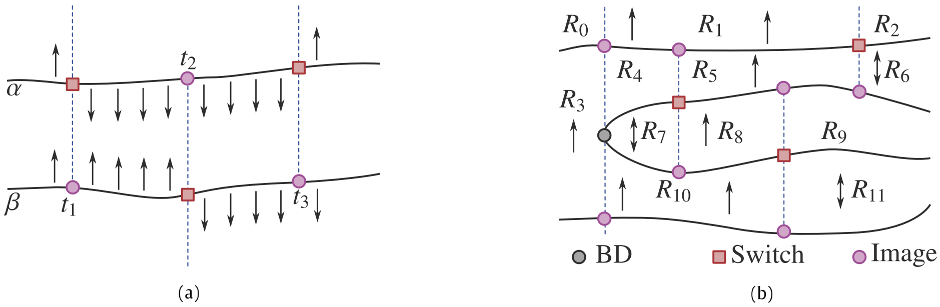


Fig. 8. (a) The pairing function μ is continuous for Jacobi segments (separated by switch points), e.g., $(\alpha(t_1), \alpha(t_3))$, $(\beta(t_1), \beta(t_2))$, and $(\beta(t_2), \beta(t_3))$. However, Jacobi regions denote the segments obtained when μ is continuous on both α and β, e.g., $R_{(t_1,t_2)}(\alpha, \beta)$ or $R_{(t_2,t_3)}(\alpha, \beta)$. (b) A standard configuration showing some Jacobi regions along with their boundary points.

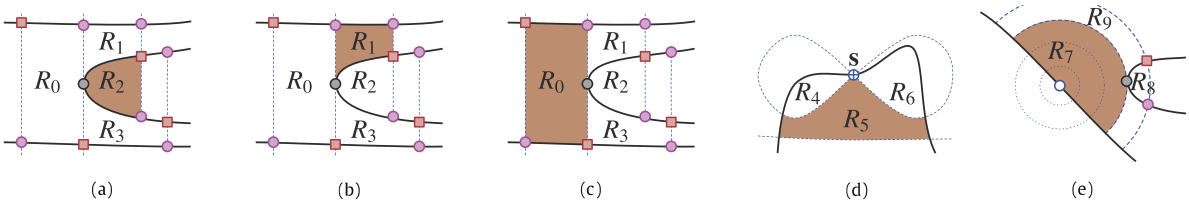


Fig. 9. Special Jacobi regions: (a) BD internal; (b) BD side; (c) BD external; (d) saddle; and (e) extremal region.

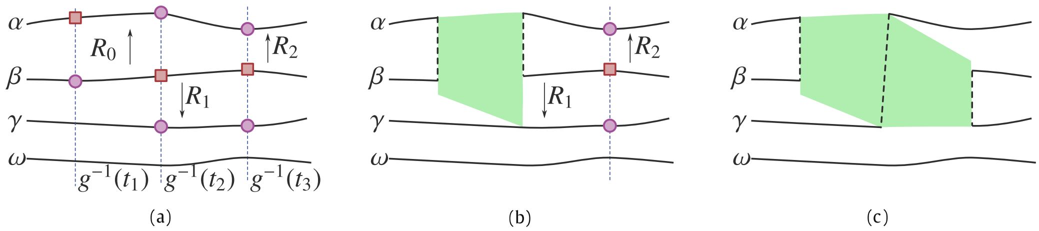


Fig. 10. Cancellation of $R_0 = R_{(t_1,t_2)}(\alpha, \beta)$ and $R_1 = R_{(t_2,t_3)}(\beta, \gamma)$ with regions of influence shown in green. The black dashed lines represent the points added to the Jacobi set to connect the existing loops.

By construction, Jacobi segments are paired consistently within each region. Except for extremal regions, which contain only two restricted critical points per level set, boundary segments of a Jacobi region $R_{(a,b)}(\alpha, \beta)$ (such that $\mu(\alpha(t)) = \beta(t)$) can be canceled by setting $\overline{f}_t(\alpha(t)) = f_t(\beta(t))$ for all $t \in (a, b)$, and imposing ϵ-slopes to create smooth and monotonic f_t^*. Adding a slope creates a simplified function f^* that is smooth and contains no critical points within the interior of the region. As shown for region R_0 in Fig. 10(b), this cancellation modifies f only within a small neighborhood around R_0 still bounded by $g^{-1}(t_1)$ and $g^{-1}(t_2)$. We call the modified region the *region of influence* of the corresponding cancellation and point out that it does not contain portions of $\mathbb{J}$ not part of R_0, and thus satisfies the consistency condition.

This modification creates a continuous $\overline{f}$ and a smooth f^* in the region, but $\overline{f}$ is still discontinuous at the boundary, and constructing a corresponding smooth f^* requires a nonlocal change. However, consider the cancellation of R_1 following the cancellation of R_0 as shown in Fig. 10(c). Since $\beta(t_2)$ is a switch point, $f_t(\alpha(t_2)) = f_t(\gamma(t_2))$. Consequently, the region of influence of R_1 matches that of R_0 at their shared boundary along $g^{-1}(t_2)$. As illustrated in Fig. 10(c) and Fig. 11 and described subsequently, canceling R_1 after R_0 removes the discontinuity along $g^{-1}(t_2)$, and creates a smooth simplification covering the interval (t_1, t_3). In general, given two regular regions $R_{(t_1,t_2)}(\alpha, \beta)$ and $R_{(t_2,t_3)}(\beta, \gamma)$ sharing a switch point $\beta(t_2)$, there always exists a simplification on the interval (t_1, t_3) that creates a smooth function along $g^{-1}(t_2)$, and removes the two Jacobi regions as well as their shared switch and image points from $\mathbb{J}$. Note that performing smooth cancellations across switch points would not be possible if the cancellation of R_0 modified both α and β. Therefore, we modify the values of either α or β, but not both (as pointed out in Section 5.1).

In order to obtain a valid $\mathbb{J}(f^*, g)$ consisting of closed loops, the simplification must also reconnect the portions of $\mathbb{J}(f, g)$ rendered disconnected due to the cancellations. For a continuous simplification, this connection can be made within a single restricted function. However, a smooth simplification demands modifications that cannot be confined within the level set containing the switch point. For example, as shown in Fig. 10(c), the segments α and γ are connected using a new parametrized curve $\xi(t)$ (middle dotted line) for $t \in (t_2 - \epsilon, t_2 + \epsilon)$. To understand the construction of $\xi(t)$, without loss of generality, assume $\alpha(t)$ and $\gamma(t)$ to be (restricted) maxima. The corresponding restricted functions in $[t_2 - \epsilon, t_2 + \epsilon]$

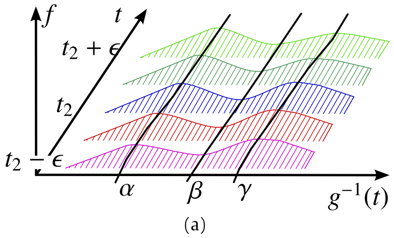


Fig. 11. The existing Jacobi curves may be connected smoothly by adding restricted critical points shown along the dashed curve. (a) Zoom-in view of the restricted functions in the neighborhood of $g^{-1}(t_2)$ from Fig. 10(a). (b) A continuous simplification can be obtained by creating a transition within a single level set. (c) However, for a valid simplification, a smooth transition $\xi(t)$ must be made by modifying the restricted functions in $[t_2 - \epsilon, t_2 + \epsilon]$.

are shown in Fig. 11. For cancellation of restricted critical points, $\beta(t)$ is moved towards $\alpha(t)$ for $t < t_2$, and towards $\gamma(t)$ for $t > t_2$. Fig. 11(b) shows the restricted function when the (continuous but not smooth) transformation is made within a single level set. However, to obtain a smooth transition, the simplification must also modify γ in $(t_2 - \epsilon, t_2)$, and α in $(t_2, t_2 + \epsilon)$. As shown in Fig. 11(c), the maxima $\gamma(t)$ and $\alpha(t)$ in the corresponding ranges are spatially shifted towards $\beta(t)$ to create a restricted maximum along the level set $g^{-1}(t_2)$ in place of the original switch point $\beta(t_2)$. Alternately, by construction, $f^*_{t_2}$ is smooth and monotonically decreasing in both the spatial intervals $(\beta(t_2), \alpha(t_2))$ and $(\beta(t_2), \gamma(t_2))$, and therefore, a restricted maximum is created in place of $\beta(t_2)$. In summary, $\xi(t_2)$ is a restricted maximum of $f^*_{t_2}$ that spatially overlaps with the canceled switch point $\beta(t_2)$. Once again following the definition of a switch point, we note that such a transition can always be created at switch points. Therefore, for simplicity in the rest of the figures, we assume smooth transitions and illustrate them as vertical lines (along a single level set).

Finally, since the Jacobi set remains unchanged outside of $[t_1, t_3]$, consistency is maintained, therefore, producing a valid simplification. The following lemma uses the properties discussed above to show that the simplified function does not create new critical points.

Lemma 5.1. *Cancellation of two adjacent Jacobi regions, $R_{(t_1,t_2)}(\alpha, \beta)$ and $R_{(t_2,t_3)}(\beta, \gamma)$ sharing a switch point $\beta(t_2)$ along their common level set $g^{-1}(t_2)$, does not create new critical points in the simplified function f^*.*

Proof. Recall that, by construction, f^* is smooth in (t_1, t_3), and contains no critical points in (t_1, t_2), and (t_2, t_3), i.e., within the interior of the two regions. It remains to prove that the restricted extrema $\xi(t_2)$ of $f^*_{t_2}$ cannot be a critical point of f^*.

Since $\beta(t_2)$ is a switch point, by definition, μ is discontinuous across $\beta(t_2)$. Without loss of generality, assume that $\mu(\beta(t_2 - \epsilon)) = \alpha(t_2 - \epsilon)$ and $\mu(\beta(t_2 + \epsilon)) = \gamma(t_2 - \epsilon)$ (as shown in Figs. 10 and 11). We know that the smooth restricted functions f^*_t can be constructed for any $\epsilon > 0$. For appropriately chosen values of ϵ, the simplification can ensure that $\xi(t_2)$ is a regular point along the curve ξ, i.e., $f^*(\xi(t_2 - \epsilon)) < f^*(\xi(t_2)) < f^*(\xi(t_2 + \epsilon))$. Therefore, $\xi(t_2)$ cannot be an extremum of f^*.

Next, we show by contradiction that $\xi(t_2)$ cannot be a saddle of f^*. Assuming $\xi(t_2)$ to be a saddle, there must exist another parametrized curve $\eta(t)$ along which $\xi(t_2)$ is a restricted minimum. By the properties of saddle point, $\eta(t)$ must be locally orthogonal to $g^{-1}(t_2)$ at $\xi(t_2)$, and therefore, must intersect the level set $g^{-1}(t_2 - \epsilon)$. However, since the restricted maximum $\xi(t_2 - \epsilon)$ is lower than $\xi(t_2)$, it follows that in the local neighborhood of $\xi(t_2)$, there does not exist any point for $t < t_2$ that is higher than $\xi(t_2)$, as shown in Fig. 12. Therefore, $\xi(t_2)$ cannot be a restricted minimum along $\eta(t)$, producing a contradiction. □

5.3. Construction and cancellation of Jacobi sequences

As discussed above, one can construct (partially) valid simplifications by simultaneously canceling adjacent Jacobi regions. In this section, we describe how to assemble *Jacobi sequences* as ordered sets of regions that allow a valid simplification. Formally, we call two Jacobi regions *adjacent* if they share a boundary point, and we use the function value of g to induce

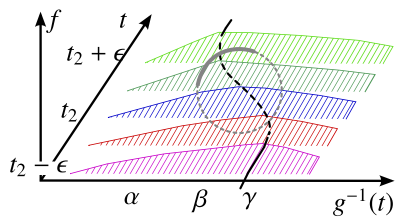


Fig. 12. Analyzing the local neighborhood of $\xi(t_2)$, it can be seen that it is not a critical point of f^*. The upper star in the local neighborhood (where function value is higher than that of $\xi(t_2)$) is marked in bold gray, whereas the lower star is in dashed gray.

an ordering among adjacent regions. To construct a sequence that admits a valid simplification, it is important to understand (a) where such a sequence may start or end, and (b) how to construct its corresponding simplification $\overline{f}$ and ultimately f^*.

Following the discussion in Section 5.2, we claim that valid sequences are naturally bounded by BD internal regions because at a BD point, the region of influence shrinks to a single point and any arbitrary small interval outside the BD point allows the construction of a smooth f^*. More specifically, consider a sequence of Jacobi regions covering the interval (a, b) that starts and ends with BD internal regions, and contains only regular regions otherwise. Given the discussion above, for any $\epsilon > 0$ we can create a smooth f^* covering the interval $(a - \epsilon, b + \epsilon)$ canceling all restricted critical points in the closure of the sequence. By construction f^* is local, smooth, and consistent, and thus forms a valid simplification.

Furthermore, note that BD external, BD side, extremal, and saddle regions can never be part of a valid simplification. Refer to Fig. 9 and notice that it is not possible to continue across the BD point for BD external and BD side regions, since the discontinuity across the level set of BD point cannot be removed locally. A similar argument holds for a saddle region, whose cancellation leaves unresolvable discontinuity around the saddle. Finally, an extremal region cannot be canceled since all level sets inside the region already contain only two restricted critical points, and cannot be simplified further.

As a result, valid sequences are comprised of only regular regions and BD internal regions, where they must begin and end with a BD internal region. Therefore, all sequences are seeded at BD internal regions and constructed by progression into adjacent regions monotonically in g until another BD internal region is encountered, at which point the sequence is considered complete. Due to the ordering imposed on adjacent regions, a sequence cannot form loops.

On the other hand, if during its construction, a sequence encounters any of the regions that cannot be simplified, it is considered invalid and discarded. Although such regions can invalidate some sequences, the progress of the simplification does not stop. If no valid sequence exists due to the presence of saddle and/or extremal regions, we perform a conventional 2D critical point cancellation in g to create new sequences, which introduces only minor structural changes to the Jacobi set, and can be done independent of any sequence cancellation. Section 6 discusses saddle cancellation in detail. Again, the BD external or BD side regions may invalidate some sequences. However, in such a case, we can always seed a new sequence from the corresponding BD internal region.

From Section 5.2, we know that a region $R_{(a,b)}(\alpha, \beta)$, such that $\mu(\alpha(t)) = \beta(t)$ for all $t \in (a, b)$, can be canceled by moving the segment α to the level of β, that is, by setting $\overline{f}(\alpha(t)) = f(\beta(t))$. However, if the region is *mutually paired*, meaning $\mu(\alpha(t)) = \beta(t)$ and $\mu(\beta(t)) = \alpha(t)$, one can move either α or β, which provides flexibility in sequence construction, as one can smoothly transition from moving α to moving β. Since valid simplification requires cancellation of adjacent regions in which the same segment can be moved to its respective partners, it follows that one can potentially cancel either of the two adjacent regions after canceling R. For example, consider Fig. 13 where regions R_0 and R_1 are already a part of a Jacobi sequence. For cancellation in R_0, the segment β is moved to match the value of α. For cancellation in R_1, we can either move β towards γ, or switch segments by smoothly transitioning from moving β to moving γ. The former leads to the sequence $\{R_0, R_1, R_2\}$ where β is moved to its respective partners in all regions, and the latter leads to $\{R_0, R_1, R_3\}$ where β and γ are moved in R_0 and R_3, respectively, whereas a transition between moving β and moving γ is performed in R_1.

We point out that the construction and cancellation of Jacobi sequences of different lengths can handle general forms of structural changes to the Jacobi set, some of which are shown in Fig. 14. Although the figure shows examples of some short Jacobi sequences, there can exist substantially longer Jacobi sequences as well.

Starting with a Morse function f, the cancellation described in this section creates a simplified function f^* while ensuring that no new critical points are created. Thus, Lemma 5.1 leads to the following corollary.

Corollary 5.1. *Cancellation of a Jacobi sequence results in a simplified function f^* that is Morse.*

5.4. Ordering the cancellations

In order to obtain a hierarchy on the simplification process, and to distinguish noise from features, we need to define a metric to measure the importance of Jacobi regions/Jacobi sequences, and the amount of modification needed for each simplification step. Choosing a metric enables a controlled, fine-grained simplification of a Jacobi set by ranking and ultimately removing portions of it in order of importance. Although the choice of the metric is flexible, we choose a gradient-based

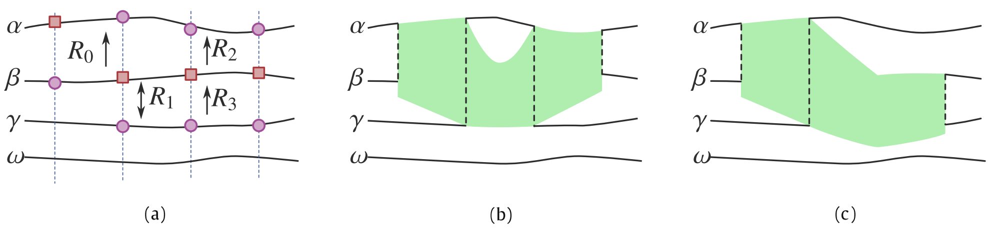
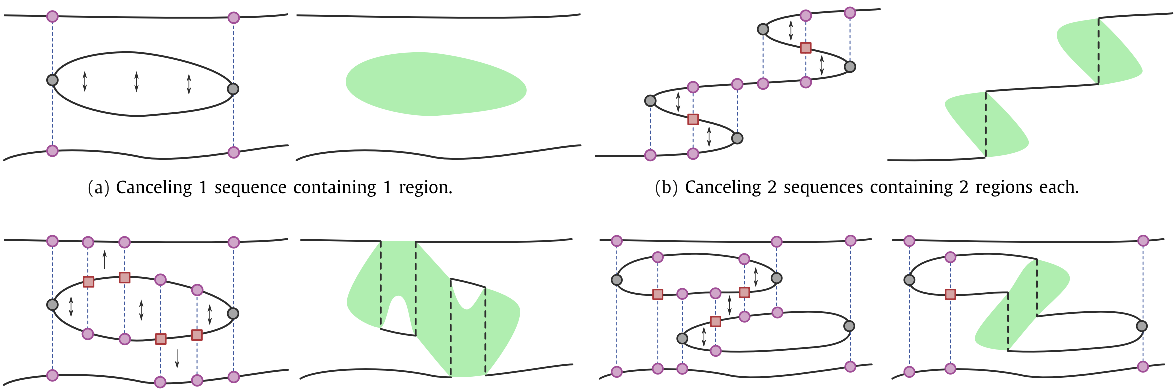


Fig. 13. Mutually paired regions offer a choice of the segment to be moved. (a) Original configuration, where R_0 can be canceled by moving β towards α. Subsequently, R_1 can be canceled by: (b) moving β to γ leading to the sequence $\{R_0, R_1, R_2\}$. (c) smoothly transitioning between moving β to moving γ leading to sequence $\{R_0, R_1, R_3\}$.



(a) Canceling 1 sequence containing 1 region.

(b) Canceling 2 sequences containing 2 regions each.

(c) Canceling 1 sequence containing 5 regions.

(d) Canceling 1 sequence containing 3 regions.

Fig. 14. The proposed simplification can perform general operations such as canceling of loops (a) and (c), straightening of zig-zags (b), and merging of loops (d). Pairings (marked by arrows) and the corresponding regions of influence (green shaded regions) are shown only for the regions that are canceled.

metric capable of measuring the relative variation between the two functions inside a region, i.e., the *comparison measure* $\kappa(R)$ (see Section 2). Our choice is inspired by the fact that the cancellation of a region creates a *flat* f^* in its interior, i.e., $\|\nabla f^*\| \leq \epsilon$. An alternative formulation of κ [32], by rewriting it as an integral over the Jacobi set, is

$$\kappa(R) = \frac{1}{2\,\mathrm{Area}(R)} \int\limits_{v \in \mathbb{J}} \left|2f(v) - f(u) - f(w)\right| \cdot \left\|\nabla g(v)\right\| \mathrm{d}v, \tag{4}$$

where $u, w \in n_g(v)$. Therefore, $\kappa(R)$ for every region R can be computed by integrating over its bounding segments. The modification needed to cancel a Jacobi sequence is the sum of modifications of all regions in the sequence. The Jacobi sequences are simplified in the increasing order of κ. The comparison measure, κ, as an importance metric, ranks the Jacobi sequences based on the amount of modification required in aligning the gradients of the two functions. We could simplify the Jacobi set up to a predetermined threshold κ^* (for the κ measure). When the simplification terminates, all Jacobi sequences with a measure up to κ^* are considered topological noise and have been simplified.

In practice, one can have a long Jacobi sequence whose κ is lower than that of a much shorter sequence. Such a situation represents a case where the two functions are more dissimilar in the shorter sequence than in the longer one, and therefore, we choose to simplify the longer sequence first. Alternatively, one could choose the length of Jacobi sequences as the ranking criteria, which may result in different simplification sequences.

6. Cancellation of critical points in *g*

As discussed in Section 5.3, no valid simplification sequence of f can cancel a critical point of g. However, there may exist configurations such that all Jacobi sequences of f contain critical points of g and all sequences in g contain critical points of f. In this case, there exists no valid sequence and the Jacobi set cannot be simplified through the cancellation procedure discussed in the previous sections. Therefore, in order to make progress, we use critical point cancellations based on Morse theory to remove pairs of critical points from either function.

For a 2D Morse function g, a cancelable pair of critical points can be either a pair of minimum and saddle, or a pair of saddle and maximum. Without loss of generality, we discuss only a saddle-maximum pair, but all concepts apply symmetrically to a minimum-saddle pair. Given a cancelable saddle-maximum $(\mathbf{s}, \mathbf{m}_1)$ pair of g, this section discusses their cancellation, and the resulting impact on the Jacobi set and the associated comparison measure. In particular, the section is divided into five parts. First, we describe how one can construct a smooth function g^* that differs from g only in an arbitrary small neighborhood of the super-level set around $\mathbf{m}_1$; second, we show that $\mathbb{J}(f, g) = \mathbb{J}(f, g^*)$ except for an ϵ-neighborhood around $\mathbf{s}$; third, we discuss how the cancellation affects Jacobi segments and regions; fourth, we identify the modification needed for this cancellation; and finally, we show that for a 2-sphere (closed simply-connected 2-manifold), we can always cancel all saddles of g.

Critical points pair cancellation To understand the cancellation $\mathbf{s}$ with $\mathbf{m}_1$, we refer to the Morse cancellation theorem discussed in Section 2. In particular, as illustrated by Milnor [29, Fig. 5.2, p. 49], critical points can be canceled by altering the neighborhood of a trajectory of a gradient-like vector field that connects $\mathbf{s}$ with $\mathbf{m}_1$. Assuming a chosen trajectory, first we discuss the cancellation. Later, we will elaborate on the rationale of how to choose the trajectory along which the cancellation must be performed.

Fig. 15 illustrates this cancellation by changing the gradient-like vector field to "re-route" the gradient lines in the neighborhood of the chosen trajectory, such that the direction of the chosen trajectory is inverted. Fig. 16 shows the same

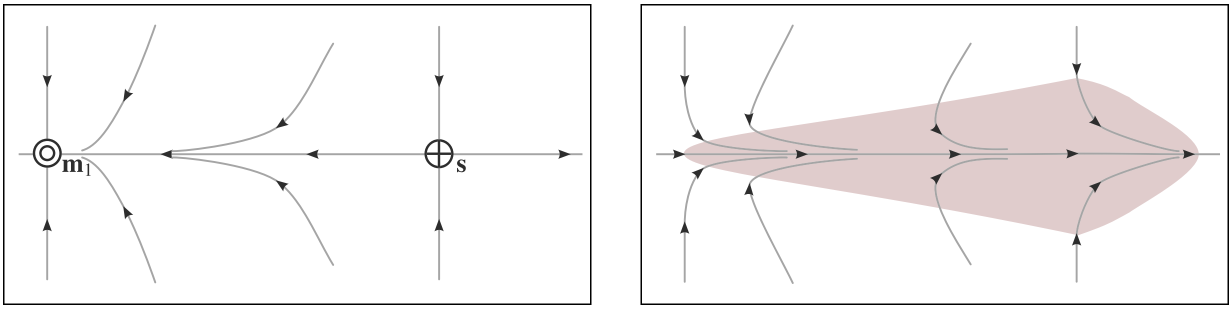
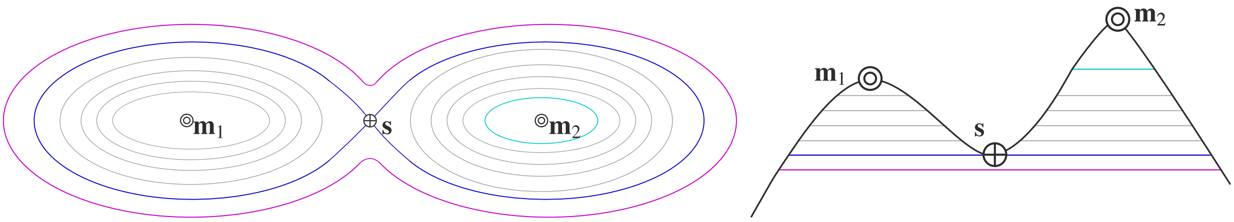
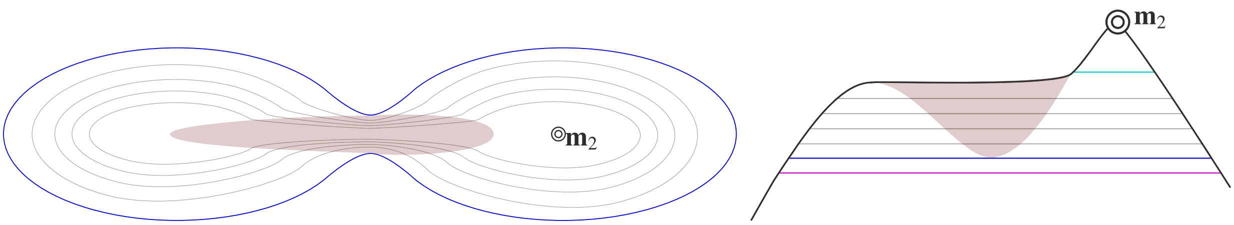


Fig. 15. Cancellation of a saddle-maximum pair $(\mathbf{s}, \mathbf{m}_1)$. The initial and the final configurations of a gradient-like vector field are shown on the left and on the right, respectively. The function g is modified along a chosen trajectory connecting $\mathbf{m}_1$ with $\mathbf{s}$. The final function, g^* does not contain the saddle $\mathbf{s}$, and the maximum $\mathbf{m}_1$. The modification (from g to g^*) is confined only to the shaded regions.



(a) Initial configuration of the function g.



(b) Final configuration of the function g^*.

Fig. 16. Cancellation of a saddle-maximum pair $(\mathbf{s}, \mathbf{m}_1)$. The top views of the function are shown on the left with colored level sets. The side views of the function are shown on the right. The saddle $\mathbf{s}$ is raised to the level of the maximum $\mathbf{m}_1$, and the two 2-disks on either side of the saddle are merged to form a single 2-disk.

cancellation by highlighting the level sets of the function instead. In particular, we can raise the saddle $\mathbf{s}$ slightly above the value of $\mathbf{m}_1$ such that $g(\mathbf{m}_1) < g^*(\mathbf{s}) < g(\mathbf{m}_1) + \epsilon$. This cancellation assures that g and g^* are different only between the level sets $g^{-1}(g(\mathbf{s}) - \epsilon)$ and $g^{-1}(g(\mathbf{m}_1) + \epsilon)$. Assuming the initial function g, the Morse cancellation theorem guarantees that the simplified function is also Morse.

Jacobi set geometry As mentioned above, the construction and choice of the trajectory, along which the cancellation is performed, plays an important role in the cancellation, and depends upon the configuration of the Jacobi set with respect to the critical points pair (i.e., cancelable pair) under consideration. Here, we discuss the rationale for making such a choice. Generically, there exist three different configurations of Jacobi sets in the neighborhood of a critical point pair as shown in Fig. 17. The most common configuration is a Jacobi set connecting $\mathbf{m}_1$ and $\mathbf{s}$ (Fig. 17(a)). In this case we can define the trajectory along $\mathbb{J}$, which guarantees that $\nabla g^*(x)$ and $\nabla g(x)$ are aligned, for all $x \in \mathbb{J}(f, g)$. It follows that $x \in \mathbb{J}(f, g)$ implies $x \in \mathbb{J}(f, g^*)$. Furthermore, denoting the region of modification (i.e., highlighted region in Fig. 15(right)) as C, for all $x \notin C$

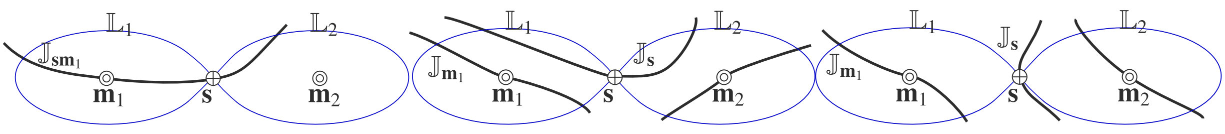


Fig. 17. Different cases of Jacobi set connectivity for a cancelable saddle-maximum pair $(\mathbf{s}, \mathbf{m}_1)$. The saddle and the maximum may be parts of by (a) the same Jacobi set component $\mathbb{J}_{\mathbf{s},\mathbf{m}_1}$, or (b) and (c) separate Jacobi components $\mathbb{J}_\mathbf{s}$ and $\mathbb{J}_{\mathbf{m}_1}$, respectively. $\mathbb{L}_1$ and $\mathbb{L}_2$ are super level sets of g surrounding $\mathbf{m}_1$ and $\mathbf{m}_2$, respectively.

we have $\nabla g^*(x) = \nabla g(x)$, which implies $\mathbb{J}(f, g) = \mathbb{J}(f, g^*)$ for $\mathbb{M} \setminus C$, meaning that Jacobi set outside of C is not modified. However, with g^* as defined above, additional Jacobi set loops may be created during the above process (that is, there may exist additional points $x \in C$ with $x \in \mathbb{J}(f, g^*)$ but $x \notin \mathbb{J}(f, g)$). Since the Jacobi set outside of C is not modified, these extra Jacobi loops must be isolated Jacobi set components entirely contained in C. As such, they must form a valid cancellation sequence and can be removed using the approach discussed in Section 4.

The situations shown in Figs. 17(b) and 17(c) follow a similar argument except that it cannot be guaranteed that $\mathbb{J}(f, g) = \mathbb{J}(f, g^*)$ around $\mathbf{s}$. However, since this portion of the Jacobi set enters and exits C exactly once, there must exist a g^* that connects the entry and exit points with a single line of the Jacobi set containing no BD points. Therefore, $\mathbb{J}(f, g) \neq \mathbb{J}(f, g^*)$ only in a small neighborhood around $\mathbf{s}$.

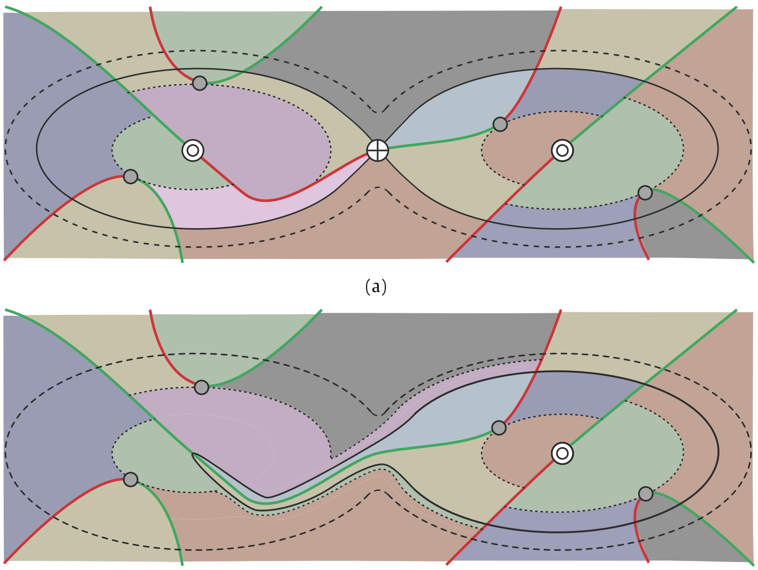
Modifications in Jacobi segments and Jacobi regions To understand how the Jacobi segments and Jacobi regions are affected by this cancellation, we refer once again to Fig. 16. Initially, the level sets $g^{-1}(t)$ for $g(\mathbf{s}) < t < g(\mathbf{m}_1) + \epsilon$ have two components, each of which is diffeomorphic to a 2-disk. Due to the cancellation, the two components of these level sets merge to create a single 2-disk. The pairings in f_t must be recomputed along the modified level sets, and new regions may be created. To illustrate the modifications in pairings and Jacobi regions, we give an example of such cancellation in Fig. 18. The figure shows that most of the regions are unaffected. Only the regions that include the region C as described above are modified, and are extended along the new level sets.

Modification needed for the cancellation We note that $|\nabla g^*(x)| = O(\epsilon)$ for all $x \in \mathbb{L}_1$ (where $\mathbb{L}_1$ is the super level set surrounding $\mathbf{m}_1$). Then, the comparison measure of $\mathbb{L}_1$ after cancellation, κ^*, is given by

$$\kappa^*(\mathbb{L}_1) = \frac{\int_{\mathbb{L}_1} \|\nabla f(x) \times \nabla g^*(x)\| \, \mathrm{d}x}{\mathrm{Area}(\mathbb{L}_1)} = O(\epsilon)$$

Note that κ^* is independent of both the difference in the function values of $\mathbf{s}$ and $\mathbf{m}_1$, $\mathbb{L}_1$. Thus, the amount of perturbation introduced by this cancellation is approximately $\lim_{\epsilon \to 0}(\kappa - \kappa^*) = \kappa$.

Simplification of g on a 2-sphere A Morse function g defined on a 2-sphere must have at least one minimum and one maximum [26, Theorem 3.35, p. 128]. We claim that on simply connected domains, we can always cancel all saddles of g by removing cancelable pairs of critical points until this simplest configuration (i.e., a Morse function with one minimum, one maximum, and zero saddle) is obtained. To prove this claim, we first present other important results regarding how saddles of g may be connected with its extrema.



(a)

(b)

Fig. 18. Effect of saddle cancellation on Jacobi segments and Jacobi regions. In addition to the level sets of BD points (dotted), level sets $g^{-1}(g(\mathbf{s}))$ (solid) and $g^{-1}(g(\mathbf{s}) - \epsilon)$ (dashed) are shown for reference. The Jacobi set is shown as a red–green curve, with color representing the criticality (i.e., restricted maximum or minimum). Jacobi regions corresponding to (a) the original Jacobi set $\mathbb{J}(f, g)$, and (b) the Jacobi set after cancellation, $\mathbb{J}(f, g^*)$, are shown in different colors.

Definition 6.1 *(Singly and doubly connected pairs).* A saddle-maximum pair $(\mathbf{s}, \mathbf{m})$ is called *singly connected* if there exists exactly one ascending 1-manifold of $\mathbf{s}$ that connects it with $\mathbf{m}$. The pair is *doubly connected* when both ascending 1-manifolds of $\mathbf{s}$ connect it with $\mathbf{m}$.

We note that a doubly connected pair is not cancelable since the boundaries of the lower disk of $\mathbf{m}$ and the upper disk of $\mathbf{s}$ intersect twice (once for each ascending 1-manifold), and hence the Morse cancellation theorem cannot be applied. On the other hand, if a saddle is singly connected to two distinct maxima $\mathbf{m}_1$ and $\mathbf{m}_2$, then it forms a cancelable pair with the lower of the two maxima, say $\mathbf{m}_1$ because the boundaries of the lower disk of $\mathbf{m}_1$ and the upper disk of $\mathbf{s}$ intersect transversally at a single point. For example, in Fig. 2, (p_1, p_0) is a doubly connected pair while (p_1, p_2) is singly connected.

Lemma 6.1. *Any saddle of a function defined on a 2-sphere may form at most one doubly connected pair. Therefore, each saddle must form at least two singly connected pairs.*

Proof. Without loss of generality, assume that a given saddle $\mathbf{s}$ forms a doubly connected pair with a minimum $\mathbf{m}_0$. See Fig. 2 for an example, by setting $\mathbf{s} := p_1$ and $\mathbf{m}_0 := p_0$. Equivalently, both descending 1-manifolds of $\mathbf{s}$ are connected to $\mathbf{m}_0$. Therefore, $\mathbb{M}$ can be cut along this closed loop splitting it into two simply connected 2-manifolds with boundary, say $\mathbb{M}_1$ and $\mathbb{M}_2$. From the Morse lemma, it follows that the ascending and descending 1-manifolds of $\mathbf{s}$ are locally orthogonal. Therefore, the two ascending 1-manifolds cannot both lie in the same piece (either $\mathbb{M}_1$ or $\mathbb{M}_2$). Since each ascending 1-manifold must connect to a maximum, $\mathbf{s}$ must be connected to two different maxima $\mathbf{m}_1$ and $\mathbf{m}_2$ in $\mathbb{M}_1$ and $\mathbb{M}_2$, respectively, which guarantees that $\mathbf{s}$ cannot form a second doubly connected pair, and must form at least two singly connected pairs (with $\mathbf{m}_1$ and $\mathbf{m}_2$). See Fig. 2 for an example, by setting $\mathbf{m}_1 := p_2$ and $\mathbf{m}_2 := p_3$. □

Using the lemmas proven above, the final result is stated as the following lemma. The existence of the simplest Morse function on closed connected manifolds was proved by Matsumoto [26, Theorem 3.35, p. 128] – such a function contains one minimum and one maximum. The following lemma shows that this simplest Morse function on a 2-sphere can be obtained using the cancellations discussed in this section.

Lemma 6.2. *All saddles of a given Morse function defined on a 2-sphere can be removed by successively removing cancelable pairs of saddles and extrema.*

Proof. Let g be a Morse function defined on a 2-sphere $\mathbb{M}$. Assuming that g contains saddles, Lemma 6.1 guarantees that each saddle forms at least two singly connected pairs (either with maxima, or with minima). If none of these singly connected pairs is cancelable, it implies that there exist at least one other critical point whose function value lies between the function values of the pair. However, it is possible to rearrange the critical points smoothly to make the pair cancelable [26, Lemma 3.26, p. 115]. Therefore, one of these two singly connected pairs becomes a cancelable pair, and can be removed along with the corresponding extremum. Applying this procedure successively, it is possible to remove all saddles of g as parts of cancelable pairs with extrema of g. □

7. Summary and correctness

Given the discussion on the cancellation of restricted critical points of f_t, and critical points of g, we now summarize the complete procedure to simplify a given Jacobi set. So far, the discussion has focused on modifying f with respect to the level sets of g. However, we may wish to interleave the modifications of either of these functions with respect to the other. Thus, to simplify the Jacobi set, we need to identify all Jacobi sequences with respect to both – the level sets of f and the level sets of g.

Step 1. Identify all possible simplification steps with respect to the level sets of g/f, by creating all possible Jacobi sequences, and identifying all cancelable saddle-extremum pairs.
- Compute the pairings between restricted critical points and identify the switch points.
- Create Jacobi segments by decomposing $\mathbb{J}$ into subsets bound by the BD points in $\mathbb{J}$, critical points of g/f, switch points, and their images.
- Create Jacobi regions using the pairings induced on the segments, and compute their κ.
- Create Jacobi sequences $\{S\}_g$ and $\{S\}_f$ by seeding them at BD internal regions, and propagating monotonically into adjacent regions in a depth-first manner, and compute its κ.
- Identify all cancelable saddle-extremum pairs $\{P\}_g$ and $\{P\}_f$, and compute their κ.

Step 2. Store all sequences $\{S\}_f$ and $\{S\}_g$, and all pairs of saddle-extremum pairs $\{P\}_f$ and $\{P\}_g$ into a common list $\mathcal{L}$, ordered by their κ, the amount of modification needed for their cancellation.

Step 3. Select the element (S or P) with the lowest κ from $\mathcal{L}$, perform its cancellation, recompute the pairings in $\mathbb{J}$, and create corresponding Jacobi regions.

Step 4. Remove from $\mathcal{L}$ all the existing sequences that cease to exist due to this cancellation, and identify and add to $\mathcal{L}$ any new sequences containing the newly created regions.
Step 5. Repeat steps 3 and 4 until the Jacobi set reaches its simplest possible configuration under our definition of validity or a user-defined threshold is achieved.

Correctness and termination In order to prove the correctness of this simplification scheme, we remind the reader that every valid simplification step ensures that the resulting function f^* (or g^*) is Morse (by Corollary 5.1 and the Morse cancellation theorem), and the simplified Jacobi set reflects the Jacobi set of the simplified functions. Therefore, we have the following corollary:

Corollary 7.1. *The simplified functions f^* and g^* are Morse, and the simplified Jacobi set is a valid Jacobi set $\mathbb{J}(f^*, g^*)$.*

By construction, the algorithm terminates when no other pair of restricted critical points can be canceled through a valid simplification, and no other cancelable critical points can be canceled through critical point cancellation.

As a reminder, the purpose of choosing a metric (e.g., κ) is to assign an ordering to the simplification process in order to perform controlled denoising of the Jacobi set. Instead of κ, a different and a more application-relevant metric may also be used for this purpose. Furthermore, if the application does not require such a fine control, one may not use any metric for the purpose of ranking, but rather proceed by canceling any valid Jacobi sequence or cancelable pair. The correctness guarantees given in Corollaries 5.1 and 7.1, and Lemmas 6.2 and 8.1 will still be applicable for such a naive procedure.

8. Obtaining the simplest configuration

In order to simplify Jacobi set, it is important to understand its simplest possible configuration, as defined below.

Definition 8.1 *(Minimal Jacobi set).* The *minimal Jacobi set* $\mathbb{J}(f, g)$ is a Jacobi set that contains no birth–death (BD) points.

Section 8.1 discusses an important property of the minimal Jacobi set in terms of the number of loops it may contain. Previously, Bennett et al. [2] defined a minimal Jacobi set with respect to the number of loops, and suggested that for a domain with genus γ, the minimal Jacobi set has $\gamma + 1$ loops. We disprove this claim by constructing Jacobi sets that contain at least one and at most two loops. In particular, for manifolds with an even genus, a Jacobi set with a single loop exists. Furthermore, instead of defining the minimal Jacobi set with respect to the number of loops, we define it as the one containing no BD points. If a Jacobi set contains no BD points, then as shown in Lemma 8.1, it implies that such a Jacobi set must contain at least one and at most two loops. However, the reverse is not true, i.e., even if a given Jacobi set contains a single loop, it may still contain BD points. Therefore, it may be simplified further, and should not be considered minimal. A related concept to our definition of the minimal Jacobi set is the notion of minimal contour given by Pignoni [35].

Next, Section 8.2 shows that for simply connected domains (where the genus $\gamma = 0$), our algorithm achieves this minimal configuration. Unfortunately, for non-simply connected domains ($\gamma > 1$), there exist nonminimal configurations that cannot be simplified through local modifications. Section 8.3 discusses these challenges on non-simply connected domains.

8.1. Minimal Jacobi sets

A property of the minimal Jacobi set in terms of the number of loops is established in Lemma 8.1. As a proof, we construct Jacobi sets containing two loops on a (single-) torus, and a single loop on a double-torus. Since a manifold of even genus is homeomorphic to a connected sum of double-tori, and a manifold of odd genus is homeomorphic to a connected sum of double-tori and a (single-) torus, a similar construction procedure can be applied to show that there exist functions f and g such that $\mathbb{J}(f, g)$ has a single loop for an even genus, and two loops for an odd genus. Recall $\mathbb{M}$ is a smooth, compact, and orientable 2-manifold without boundary.

Lemma 8.1. *The minimal Jacobi set $\mathbb{J}(f, g)$ on a manifold $\mathbb{M}$ of genus γ contains at least one and at most two loops.*

Proof. In the case when $\gamma = 0$, it is easy to see that there exist f and g that create only a single loop. For example, imagine a sphere embedded into $\mathbb{R}^3$ centered at the origin. Then, the functions $f(x, y, z) = x$ and $g(x, y, z) = z$ will create such a Jacobi set.

For $\gamma > 0$, $\mathbb{M}$ is homeomorphic to a connected sum of γ tori. Such a surface can be constructed as the union of bent and straight cylinders as shown in Fig. 19. Imagine each piece embedded into $\mathbb{R}^3$ with $g(x, y, z) = z$, the height function. Defining $f(x, y, z) = x$ creates a Jacobi set that follows the silhouette and creates $\gamma + 1$ loops. However, along a straight cylinder we can smoothly transition to $f(x, y, z) = -x$ (and the reverse), which winds the Jacobi set around the cylinder in a half turn (Fig. 19(e)–(f)). Combining these twisted cylinders, one can reconnect the default $\gamma + 1$ loops. The Jacobi sets for the torus and double-torus are also shown in Fig. 20 without the gluing cylinders.

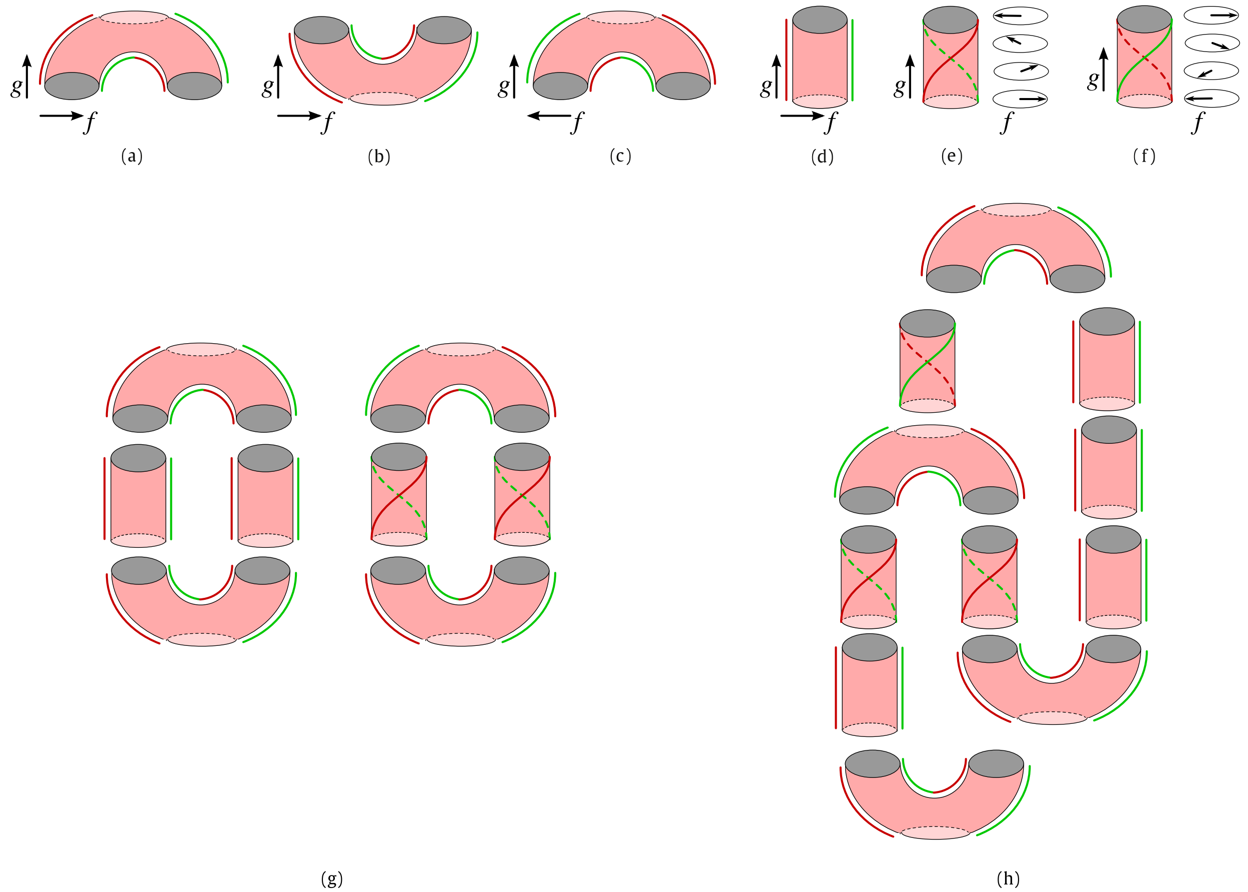


Fig. 19. A single- (g) and double-torus (h) can be constructed by gluing together six smaller pieces (a) to (f). The arrows indicate the directions of functions f and g, the colors of the Jacobi loops denote criticality of f_t, and the dashed lines denote part of the loop on the back side of the manifold. In (e) and (f), f is smoothly changed from top to bottom such that its gradient is inverted. This operation rotates the Jacobi loop between left and right of the corresponding pieces. When the pieces are glued together, this rotation makes it possible for a single Jacobi loop to connect all the critical points of f and g for a double-torus. However, for a single-torus, it simply interchanges the connectivity of the two loops.

In particular, as shown in Fig. 19(h) for a double-torus, we can connect all pieces into a single loop. Clearly, as shown Fig. 21, combining double-tori creates functions with a single Jacobi loop for all surfaces with an even genus. However, for a single-torus, the same technique simply intertwines two loops (Fig. 19(g)). Nevertheless, treating a surface with an odd genus as one with an even genus plus a torus, there must exist f and g that create only two loops, which proves the lemma. □

We conjecture that for surfaces with an uneven genus, two loops is the minimal configuration as the recombinations must come in pairs but currently there exists no proof. On the other hand, similar scenarios have been treated by several authors [18,24,27]. Therefore making more concrete connections between our conjecture and the singularity theory is an interesting future direction.

8.2. Simplification of a Jacobi set on simply connected domains

To show that our simplification can obtain the minimal configuration on simply connected domains, we first argue that if two BD points are connected by a Jacobi loop, there always exists a valid sequence that removes both BD points from the Jacobi set. Next, assuming that g contains only two extrema on a simply connected domain, there exists only a single configuration where the Jacobi set may contain BD points that are not connected by the same Jacobi loop (shown in Fig. 22). We prove that these BD points can also be canceled using a valid Jacobi sequence.

Lemma 8.2. *If $\mathbb{M}$ is a simply connected domain, and two BD points, u and v, are connected by a Jacobi loop such that no critical points of g or other BD points are between them (within the loop), then there exists a sequence of Jacobi regions connecting u with v that forms a valid simplification.*

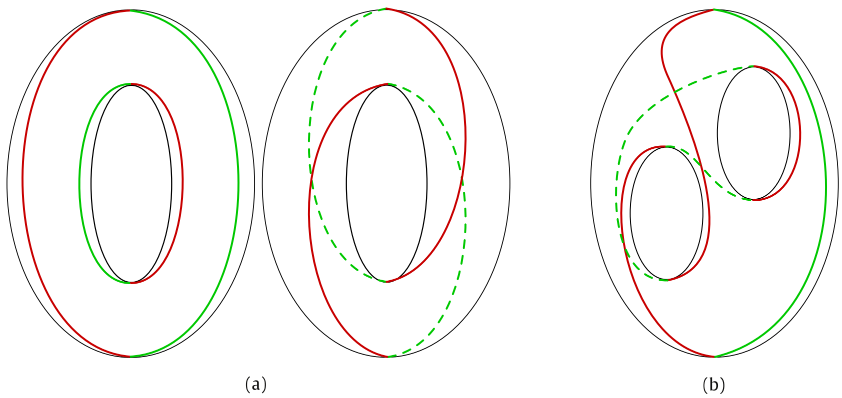


Fig. 20. (a) The Jacobi set on a single-torus ($\gamma = 1$) contains two loops, with two possible configurations. (b) In the case of a double-torus ($\gamma = 2$), a configuration with a single Jacobi loop is feasible. The color of the Jacobi loops denote criticality of f_t, the dashed line denotes the loop on the back side of the torus.

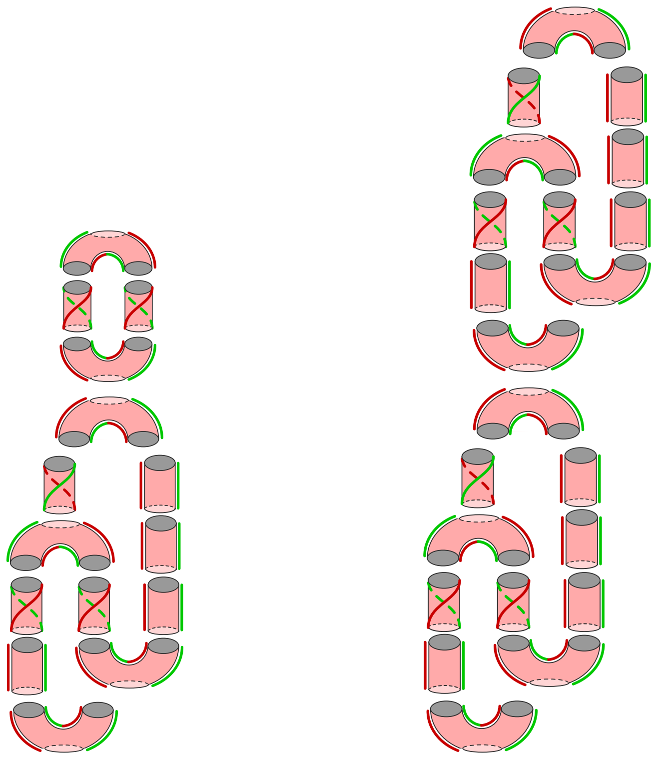


Fig. 21. A three-torus (left) constructed as connected sum of a single-torus and a double-torus ($\mathbb{T}^3 = \mathbb{T}\#\mathbb{T}^2$), and a four-torus (right) constructed as connected sum of two double-tori ($\mathbb{T}^4 = \mathbb{T}^2\#\mathbb{T}^2$). Clearly, all critical points on the four-torus can be connected by the single loop. On the other hand, for a three-torus, one needs two Jacobi loops.

Proof. Let $t_0, t_1 \in \mathbb{R}$ denote the function values of g at the two BD points connected by a Jacobi loop, that is, $t_0 = g^{-1}(u)$ and $t_1 = g^{-1}(v)$, and without loss of generality, assume $t_0 < t_1$. The BD points create and destroy two restricted critical points. Since the restricted functions $f_{t_0-\epsilon}$ and $f_{t_1+\epsilon}$ are Morse, they must contain at least two restricted critical points. It follows that for all $t \in (t_0, t_1)$, f_t has at least four restricted critical points. As a result, each point on the Jacobi set connecting u with v is paired and can be canceled with its partner. Since at a BD point, $\mathbb{J}$ is always mutually paired on the "inside" (of the BD internal region), there must exist a valid sequence or Jacobi regions connecting u with v. Some possible configurations for this scenario are shown in Figs. 14(a), 14(b), and 14(c). $\square$

To prove the main result, we note that on a simply connected domain, all saddles of f and g can be removed either through simplifying the Jacobi set or through direct cancellations, such that only a single minimum and a single maximum

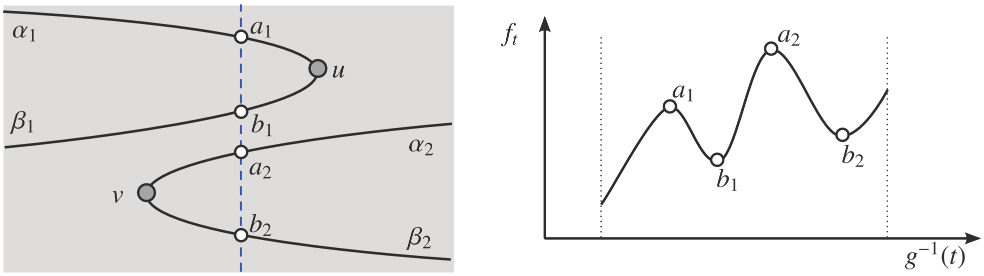


Fig. 22. A pair of BD points must always be connected by a Jacobi sequence. Consider a level set $g^{-1}(t)$ between the level sets containing the BD points u and v. Since the corresponding f_t is periodic, both pairs of restricted critical points (a_1, b_1), and (a_2, b_2) cannot be mutually paired. Thus, there must exist a region $R_{[t_1,t_2]}(\alpha_1, \beta_2)$, or $R_{[t_1,t_2]}(\alpha_2, \beta_1)$ leading to a valid sequence connecting u and v.

remain. In this case, any potentially remaining BD points must form a valid sequence of regions, as no critical points exist that may block a sequence from being formed.

Lemma 8.3. *If $\mathbb{M}$ is a simply connected domain ($\gamma = 0$), the algorithm reduces a Jacobi set to its minimal configuration – a single loop without birth–death points.*

Proof. Without loss of generality, we suppose f and g contain no saddles, since for simply connected domain, all saddles can be canceled. Following Euler's characteristic, this supposition implies that both f and g contain a single minimum and a single maximum. As a result, the level sets of g can be seen as a collection of vertical lines periodic at ∞ as shown in Fig. 22. Following Lemma 8.2, all BD points connected by Jacobi loops can be removed through valid cancellations. Nevertheless, there can exist two BD points, say u and v, that are not connected to each other through a Jacobi set curve. We note that along a Jacobi set curve, the criticality of the restricted critical points of f_t switches at BD points and the critical points of g. Therefore, out of the two curves containing u and v, one must form a loop with the maximum of g, and other with the minimum of g. Both loops must overlap since each f_t must have at least two restricted critical points.

Referring to Fig. 22, assume that there does not exist a sequence connecting the two BD points. This assumption means that the curve α_2 is never paired with β_1, and α_1 is never paired with β_2, otherwise pairing switches exist along β_1 or α_2, implying that there exists a Jacobi sequence connecting u and v. Equivalently, it follows that α_1 is always mutually paired with β_1 and α_2 is always mutually paired with β_2, which will be shown to create a contradiction. Assume that a_1 and a_2 are maxima (i.e., local maxima for the function restricted to the level set), and b_1 and b_2 are minima (i.e., local minima for the function restricted to the level set). If α_1 is mutually paired with β_1 then $f(b_1) > f(b_2)$ and $f(a_1) < f(a_2)$. (Remember that the level sets are periodic.) However, α_2 mutually paired with β_2 implies $f(a_2) < f(a_1)$, which gives a contradiction, and hence proves the lemma. $\square$

8.3. *Simplification of a Jacobi set on non-simply connected domains*

On simply connected domains, we showed that our simplification scheme can obtain the minimal Jacobi set configuration. Here, we discuss the fundamental problems due to the topology of non-simply connected domains, and how they impact our simplification algorithm. For non-simply connected domains, there exist saddles that cannot be removed even through conventional critical point cancellation (e.g., based on the Morse theory). These saddles can block the construction of Jacobi sequences, such that no more (valid) Jacobi sequences may be formed leading to a premature termination of the algorithm (without eliminating all BD points). Thus, our algorithm may terminate without achieving the minimal Jacobi set.

To contrive such an example, we start with the minimal Jacobi set on a torus $\mathbb{T}$ with $f(x, y, z) = x$ and $g(x, y, z) = z$, as shown in Fig. 20(a). The function f can then be changed along the outer silhouette of the torus, using a sinusoidal kernel that replaces the restricted maxima with a valley and restricted minima with a ridge. For each f_t, this operation replaces one restricted critical point by three, thus creating two extra Jacobi loops. Since the function must stay smooth, the kernel must go to zero at the critical points of g, where the restricted critical points of f_t switch criticality.

To understand this Jacobi set, recall that a torus is constructed as the product of two circles. If θ and ϕ denote the polar angles of the two circles, then the torus can be parametrized as $\mathbb{T}(\theta, \phi)$. Fig. 23 shows the level sets and critical points of the two functions (in red and blue) on the θ–ϕ plane along with the Jacobi set (in black). Clearly, there exist four loops in the Jacobi set. The saddles on $\mathbb{J}_1$ and $\mathbb{J}_3$ also act as BD points. Any sequences that are seeded at the BD points always get stuck at the saddles and hence, no valid sequence is possible.

Consequently, the proposed algorithm cannot simplify this Jacobi set further, since locality is an integral property of our simplification. However, going forward, we envision more general and global simplifications steps, which modify more than two loops of a Jacobi set simultaneously. Such simplifications will be able to handle difficult cases for non-simply connected domains such as the one discussed above.

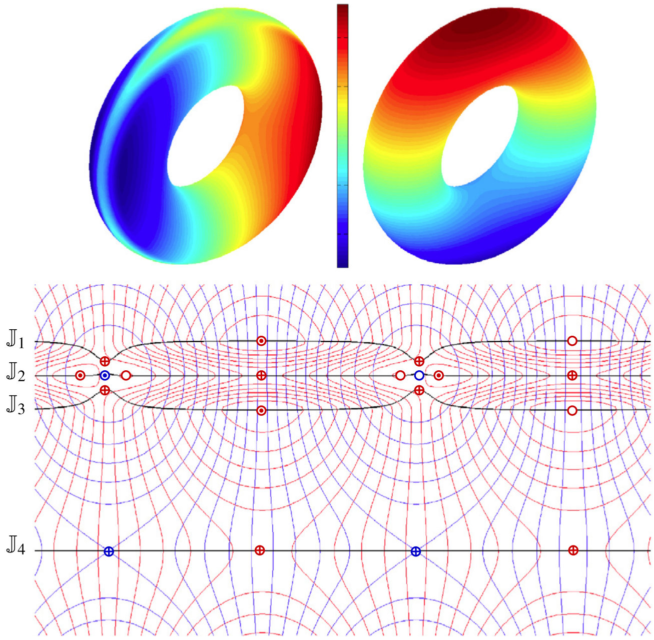


Fig. 23. (Top) Functions f (left) and g (right) are defined on a torus, $\mathbb{T}(\theta, \phi)$. (Bottom) The level sets and critical points of f and g are shown in red and blue, respectively, along with the Jacobi set in black, on the θ–ϕ plane. Since the domain is periodic, the four Jacobi loops are closed. Although there exist BD points on $\mathbb{J}_3$ and $\mathbb{J}_4$ (coinciding with saddles of f), the algorithm can not find a valid Jacobi sequence due to the presence of irremovable saddles. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

9. Discussion and future work

In this paper, we introduce a technique for Jacobi set simplification, aimed at achieving local, smooth, and consistent modifications to the underlying functions. Guided by a user-defined metric, our technique offers fine control over the simplification process, and is widely applicable in many data analysis applications. The presented procedure performs cancellations in the increasing order of κ, and can be seen as a greedy strategy. Nevertheless, the underlying idea of generating a hierarchical representation of features in the data by choosing minimal modification with respect to a chosen ranking criterion is similar to other existing simplification schemes. As future work, we would like to explore other ranking criteria for cancellations.

For simply connected domains, we show that (irrespective of the choice of the metric) this procedure can reduce a given Jacobi set to its minimal configuration (i.e., one with no BD points). We note that there may exist other procedures to obtain the minimal Jacobi set, potentially requiring even less modification (with respect to κ) in the underlying functions. The main contribution of our work is to provide ideas and theoretical constructs (such as Jacobi regions and Jacobi sequences) based on a domain-segmentation perspective, to arrive at a minimal configuration of the Jacobi set during simplification, and to provide fine-grained control over the simplification process based on some form of ranking criteria. In other words, the presented algorithm is not aimed at providing a global optimal solution in terms of minimizing the accumulated κ measure, but rather, it uses κ as a ranking criterion to guide the detailed simplification process that separates topological features from noise. We present a first step towards understanding and obtaining the minimal Jacobi set by proposing meaningful constructs whose theoretical foundations represent pairings between restricted critical points. Based on domain segmentation, this paper, for the first time, also highlights the importance of understanding the domain topology in Jacobi set simplification. Whereas the algorithm reduces a Jacobi set to its minimal configuration for simply connected domains, there exist cases where this is not possible for non-simply connected domains. There is a need to further understand such cases in more detail, and we suspect that one may need global simplification operations that can help obtain the simplest Jacobi set for non-simply connected domains. We wish to explore such cases and extend our simplification scheme to address them.

Lastly, the focus of the current work is a detailed discussion of the various elements of the simplification for smooth functions on smooth domains without boundary. A practical implementation of the presented simplification scheme for discrete functions defined on domains with boundaries requires addressing additional concerns, such as degeneracies, numerical instabilities, memory and running time efficiencies, etc. A detailed discussion of the discrete adaptation of the simplification scheme with practical applications and results is forthcoming.

Acknowledgements

We thank Attila Gyulassy for insightful discussions during the early stage of this work. We are also thankful to the anonymous reviewers whose feedback helped us improve this article. This work is supported in part by BNSF CISE ACI-0904631, NSG IIS-1045032, NSF EFT ACI-0906379, DOE/NEUP 120341, DOE/Codesign P01180734, DOE/SciDAC DESC0007446, CCMSC DE-NA0002375 and DE-EE0004449. This work was performed under the auspices of the US Department of Energy (DOE) by Lawrence Livermore National Laboratory (LLNL) under contract DE-AC52-07NA27344. LLNL-JRNL-662444.

References

[1] T.P. Barnett, D.W. Pierce, R. Schnur, Detection of anthropogenic climate change in the world's oceans, Science 292 (5515) (2001) 270–274.
[2] J. Bennett, V. Pascucci, K. Joy, Genus oblivious cross parameterization: robust topological management of inter-surface maps, in: Proceedings of the Pacific Conference on Computer Graphics and Applications, 2007, pp. 238–247.
[3] P.-T. Bremer, E.M. Bringa, M.A. Duchaineau, A. Gyulassy, D. Laney, A. Mascarenhas, V. Pascucci, Topological feature extraction and tracking, J. Phys. Conf. Ser. 78 (2007).
[4] H. Carr, D. Duke, Joint contour nets: computation and properties, in: Proceedings of the IEEE Pacific Visualization Symposium, 2013, pp. 161–168.
[5] H. Carr, J. Snoeyink, U. Axen, Computing contour trees in all dimensions, Comput. Geom. 24 (2) (2003) 75–94, special issue on the Fourth {CGC} Workshop on Computational Geometry.
[6] H. Carr, J. Snoeyink, M. van de Panne, Simplifying flexible isosurfaces using local geometric measures, in: IEEE Visualization, 2004, pp. 497–504.
[7] H. Carr, J. Snoeyink, M. van de Panne, Flexible isosurfaces: simplifying and displaying scalar topology using the contour tree, Comput. Geom. 43 (1) (2010) 42–58, special issue on the 14th Annual Fall Workshop.
[8] H. Edelsbrunner, J. Harer, Jacobi sets of multiple Morse functions, in: F. Cucker, R. DeVore, P. Olver, E. Süli (Eds.), Foundations of Computational Mathematics, Minneapolis 2002, Cambridge University Press, 2002, pp. 37–57.
[9] H. Edelsbrunner, J. Harer, Persistent Homology – A Survey, Contemporary Mathematics, vol. 453, American Mathematical Society, 2008, pp. 257–282.
[10] H. Edelsbrunner, J. Harer, Computational Topology: An Introduction, American Mathematical Society, 2010.
[11] H. Edelsbrunner, J. Harer, A. Mascarenhas, V. Pascucci, J. Snoeyink, Time-varying Reeb graphs for continuous space–time data, Comput. Geom. 41 (3) (2008) 149–166.
[12] H. Edelsbrunner, J. Harer, V. Natarajan, V. Pascucci, Morse–Smale complexes for piecewise linear 3-manifolds, in: Proceedings of the Nineteenth Annual Symposium on Computational Geometry, 2003, pp. 361–370.
[13] H. Edelsbrunner, J. Harer, V. Natarajan, V. Pascucci, Local and global comparison of continuous functions, in: Proceedings of the Conference on Visualization, 2004, pp. 275–280.
[14] H. Edelsbrunner, J. Harer, A. Patel, Reeb spaces of piecewise linear mappings, in: Proceedings of the Twenty-Fourth Annual Symposium on Computational Geometry, 2008, pp. 242–250.
[15] H. Edelsbrunner, J. Harer, A.J. Zomorodian, Hierarchical Morse–Smale complexes for piecewise linear 2-manifolds, Discrete Comput. Geom. 30 (1) (2003) 87–107.
[16] H. Edelsbrunner, D. Letscher, A.J. Zomorodian, Topological persistence and simplification, Discrete Comput. Geom. 28 (4) (2002) 511–533.
[17] H. Edelsbrunner, D. Morozov, A. Patel, The stability of the apparent contour of an orientable 2-manifold, in: V. Pascucci, X. Tricoche, H. Hagen, J. Tierny (Eds.), Topological Methods in Data Analysis and Visualization, in: Mathematics and Visualization, Springer, Berlin, Heidelberg, 2011, pp. 27–41.
[18] J.M. Éliašberg, On singularities of folding type, Math. USSR, Izv. 4 (5) (1970) 1119.
[19] X. Ge, I. Safa, M. Belkin, Y. Wang, Data skeletonization via Reeb graphs, in: J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, K. Weinberger (Eds.), Advances in Neural Information Processing Systems 24, vol. 24, Curran Associates, Inc., 2011, pp. 837–845.
[20] Y.I. Gingold, D. Zorin, Controlled-topology filtering, Comput. Aided Des. 39 (8) (2007) 676–684.
[21] A. Gyulassy, V. Natarajan, V. Pascucci, B. Hamann, Efficient computation of Morse–Smale complexes for three-dimensional scalar functions, IEEE Trans. Vis. Comput. Graph. 13 (6) (2007) 1440–1447.
[22] W. Harvey, O. Rübel, V. Pascucci, P.-T. Bremer, Y. Wang, Enhanced topology-sensitive clustering by Reeb graph shattering, in: R. Peikert, H. Hauser, H. Carr, R. Fuchs (Eds.), Topological Methods in Data Analysis and Visualization II, in: Mathematics and Visualization, Springer, Berlin, Heidelberg, 2012, pp. 77–90.
[23] M. Hilaga, Y. Shinagawa, T. Kohmura, T.L. Kunii, Topology matching for fully automatic similarity estimation of 3D shapes, in: Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, 2001, pp. 203–212.
[24] T. Kálmán, Stable maps of surfaces into the plane, Topol. Appl. 107 (3) (2000) 307–316.
[25] C. Luo, I. Safa, Y. Wang, Approximating gradients for meshes and point cloud via diffusion metric, Comput. Graph. Forum 28 (5) (2009) 1497–1508.
[26] Y. Matsumoto, An Introduction to Morse Theory, American Mathematical Society, 2002, translated from Japanese by Kiki Hudson and Masahico Saito.
[27] K.C. Millett, Generic smooth maps of surfaces, Topol. Appl. 18 (2–3) (1984) 197–215.
[28] J. Milnor, Morse Theory, Princeton University Press, 1963.
[29] J. Milnor, Lectures on the h-Cobordism Theorem, Princeton Mathematical Notes, Princeton University Press, 1965.
[30] M. Morse, Bowls of a nondegenerate function on a compact differentiable manifold, in: Differential and Computational Topology (Symposium in Honor of Marston Morse), vol. 75, Princeton University Press, 1965, pp. 81–104.
[31] S. Nagaraj, V. Natarajan, Simplification of Jacobi sets, in: V. Pascucci, X. Tricoche, H. Hagen, J. Tierny (Eds.), Topological Data Analysis and Visualization, in: Mathematics and Visualization, Springer, Berlin, Heidelberg, 2011, pp. 91–102.
[32] V. Natarajan, Topological analysis of scalar functions for scientific data visualization, Ph.D. thesis, Department of Computer Science, Duke University, 2004.
[33] G. Norgard, P.-T. Bremer, Ridge-valley graphs: combinatorial ridge detection using Jacobi sets, Comput. Aided Geom. Des. 30 (6) (2012) 597–608.
[34] V. Pascucci, G. Scorzelli, P.-T. Bremer, A. Mascarenhas, Robust on-line computation of Reeb graphs: simplicity and speed, in: Proceedings of SIGGRAPH, 2007, ACM Trans. Graph. 26 (3) (2007) 58.1–58.9.
[35] R. Pignoni, Projections of surfaces with a connected fold curve, Topol. Appl. 49 (1) (1993) 55–74.
[36] G. Reeb, Sur les points singuliers d'une forme de Pfaff complètement intégrable ou d'une fonction numérique, C. R. Math. Acad. Sci. 222 (1946) 847–849.
[37] Y. Shinagawa, T.L. Kunii, Constructing a Reeb graph automatically from cross sections, IEEE Comput. Graph. Appl. 11 (6) (1991) 44–51.
[38] S. Smale, On the structure of manifolds, Am. J. Math. 84 (3) (1962) 387–399.
[39] S. Takahashi, G.M. Nielson, Y. Takeshima, I. Fujishiro, Topological volume skeletonization using adaptive tetrahedralization, in: Proceedings Geometric Modeling and Processing, 2004, pp. 227–236.
[40] M. van Kreveld, R. van Oostrum, C.L. Bajaj, V. Pascucci, D.R. Schikore, Contour trees and small seed sets for isosurface traversal, in: Proceedings of the Thirteenth Annual Symposium on Computational Geometry, 1997, pp. 212–220.
[41] A.J. Zomorodian, G. Carlsson, Computing persistent homology, Discrete Comput. Geom. 33 (2) (2005) 249–274.