

Optimal Topological Simplification of Discrete Functions on Surfaces

Ulrich Bauer · Carsten Lange · Max Wardetzky

Received: 30 October 2010 / Revised: 21 February 2011 / Accepted: 22 March 2011 /
Published online: 27 April 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract Given a function f on a surface and a tolerance $\delta > 0$, we construct a function f_δ subject to $\|f_\delta - f\|_\infty \leq \delta$ such that f_δ has a minimum number of critical points. Our construction relies on a connection between discrete Morse theory and persistent homology and completely removes homological noise with persistence $\leq 2\delta$ from the input function f . The number of critical points of the resulting simplified function f_δ achieves the lower bound dictated by the stability theorem of persistent homology. We show that the simplified function can be computed in linear time after persistence pairs have been computed.

Keywords Discrete Morse theory · Persistent homology · Topological denoising

1 Introduction

Measured data and functions constructed from measured data suffer from omnipresent noise introduced during the measuring process. Separating relevant information from noise is therefore a widely considered problem.

Taking a topological point of view, we regard noise as a source of critical points. Indeed, even arbitrarily small amounts of noise (with respect to the supremum norm)

U. Bauer (✉) · M. Wardetzky
Institute for Numerical and Applied Mathematics, University of Göttingen, Lotzestr. 16–18,
37083 Göttingen, Germany
e-mail: bauer@math.uni-goettingen.de

M. Wardetzky
e-mail: wardetzky@math.uni-goettingen.de

C. Lange
Department of Mathematics and Computer Science, Freie Universität Berlin, Arnimallee 6,
14195 Berlin, Germany
e-mail: clange@math.fu-berlin.de

may give rise to an arbitrarily large number of critical points. We may hence interpret critical points that can be *eliminated* by small perturbations as being caused by noise. Consequently, we consider the following optimization problem:

Problem (Topological simplification on surfaces) Given a function f on a surface and a real number $\delta > 0$, find a function f_δ subject to $\|f_\delta - f\|_\infty \leq \delta$ such that f_δ has a minimum number of critical points.

Of course, in order to make this problem precise, one needs to choose a class of admissible functions that provides a meaningful notion of critical points. We present a solution to this problem for the class of *discrete pseudo-Morse functions*, which generalizes the class of discrete Morse functions introduced by Forman [15]. Our notion of pseudo-Morse functions incorporates both piecewise linear functions and pixel data.

The *Bottleneck Stability Theorem* [6], a fundamental result in the theory of *persistent homology* [11, 36], provides a lower bound on the number of critical points for any δ -perturbation of a given input function (see Corollary 15):

Stability bound Let f be a discrete pseudo-Morse function, and let $\delta \geq 0$. Then the number of critical points of any pseudo-Morse function f_δ with $\|f_\delta - f\|_\infty \leq \delta$ is bounded from below by the number of those critical points of f that have persistence $> 2\delta$.

Our main result is a constructive proof of the tightness of the stability bound on surfaces (see Theorem 16), providing an optimal solution to the topological simplification problem:

Theorem (Tightness of the stability bound) *On a combinatorial surface, the stability bound is tight for every discrete pseudo-Morse function f and for every $\delta \geq 0$.*

A similar statement does not hold in higher dimensions or for non-manifold 2-complexes, see Sect. 6.5.

1.1 Overview

Our solution to the problem of topological simplification on surfaces relies on a synthesis of two powerful theories of combinatorial and computational topology: *discrete Morse theory* and *persistent homology*.

Discrete Morse theory [15, 16], developed by Forman in the context of CW complexes, provides combinatorial equivalents of several core concepts of classical Morse theory, such as Morse functions, gradient vector fields, critical points, and a cancellation theorem for the elimination of pairs of critical points from a vector field. The discrete theory maintains the intuition of its classical counterpart while enabling simple and explicit constructions that are considerably more involved in the smooth setting.

Persistent homology [11, 36] quantifies topological features of a function. It identifies pairs of birth and death of homology classes at critical points (*persistence pairs*) and provides a quantitative notion of the significance (*persistence*) of such pairs.

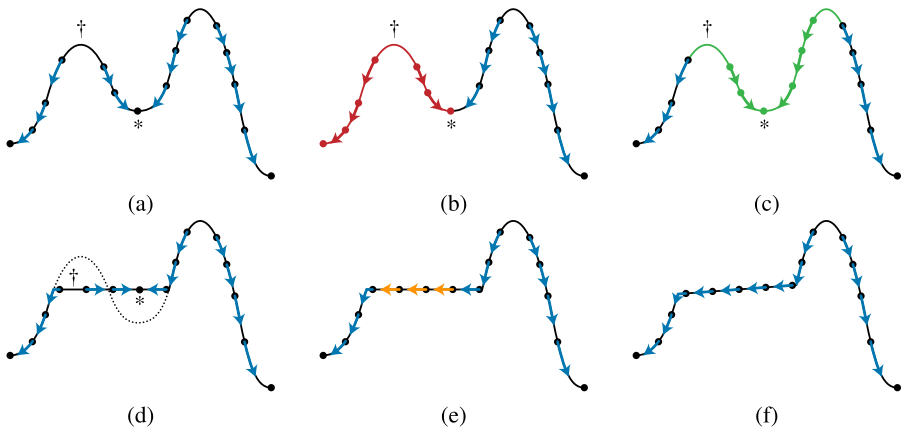


Fig. 1 Cancellation of critical points in a nutshell. Starting with the graph of a function together with the directions of its gradient vector field (a), the values of the descending set (b) of the upper critical point (marked with †) and of the ascending set (c) of the lower critical point (marked with *) are cut off at the average value of the two critical points, creating a plateau (d). The old gradient directions are still consistent with the new function. The gradient vector field can now be reversed along the path between the critical points, eliminating the pair (e). The resulting function has a plateau, but can be perturbed slightly to become non-degenerate (f)

In both theories, complementary pairs of critical points of indices $(i, i + 1)$ play a central role. However, whereas (discrete) Morse theory makes statements about the *homotopy type* of the sublevel sets of a function, persistence theory is concerned with their *homology*.

In the following, we provide a brief overview of our method and contributions.

Canceling a Single Pair of Critical Points from a Function Forman [15] describes a simple method for eliminating a pair of critical points from a discrete *vector field*. Modifying a *function* accordingly requires a slight extension of Forman’s method. We first observe that a discrete gradient vector field in the sense of Forman induces a partial order on the cells of the underlying complex, giving rise to the notion of *ascending* and *descending sets* (in analogy to the notion of ascending and descending manifolds in the classical theory). Building on these concepts, we describe a canonical method for eliminating a pair of critical points from a discrete Morse function. Our construction complements Forman’s cancellation method for discrete gradient vector fields; in particular, it is applicable to functions on general CW complexes (Sect. 4.1). An informal description of our cancellation method is depicted in Fig. 1.

In order to cancel a single pair of critical points whose values differ by 2δ , our method perturbs the function by δ in the supremum norm, which is the minimum required for canceling such a pair (see Fig. 1). To achieve this minimum, function values have to be modified on the ascending and descending sets of the canceled pair of critical points. Since these sets may contain cells of *any* dimension, other *critical* values might also have to be changed in this process.

Degenerate Functions Morse theory, in any of its variations, fundamentally relies on the assumption that critical points are non-degenerate. This condition prevents

the theory from being directly applicable to arbitrary input functions. Additionally, in our construction, the canonical function arising from canceling a single pair of critical points has a plateau (see Fig. 1) and is hence a degenerate function in the sense of discrete Morse theory. This necessitates a method to deal with degenerate functions. To do so, we devise a symbolic perturbation scheme based on discrete gradient vector fields, which allows to treat the degenerate case in much the same way as the generic case by introducing the larger class of *pseudo-Morse functions* (Sect. 2.2). In order to obtain a notion of critical points for pseudo-Morse functions, we work with an *explicit* gradient vector field consistent with the function, which coincides with the usual discrete gradient vector field in the non-degenerate case. Our symbolic perturbation scheme can be interpreted as providing a Morse function that is consistent with the given gradient vector field and is arbitrarily close to the given pseudo-Morse function.

Multiple Cancellations In principle, multiple pairs of critical points can be eliminated by sequentially applying the above (single pair) cancellation method. However, as a prerequisite for cancellation, the above method relies on the existence of a unique gradient path between the pair of critical points to be canceled. For every step in a cancellation sequence, this poses the problem of finding an admissible pair of critical points that satisfies this prerequisite. It is natural to ask whether such admissible cancellation pairs can be related to persistence pairs in some way. In particular, can every persistence pair eventually be eliminated by a cancellation sequence?

For the case of surfaces, the answer is indeed affirmative *if* persistence pairs are canceled in a *nested* order. Concretely, we show that a persistence pair (σ, τ) can be canceled after all persistence pairs $(\tilde{\sigma}, \tilde{\tau})$ with $f(\sigma) < f(\tilde{\sigma}) < f(\tilde{\tau}) < f(\tau)$ have been removed (assuming, by symbolic perturbation, that all critical points have distinct values). This result is established by introducing a certain hierarchy on the persistence pairs with indices $(0, 1)$, which by duality extends to the pairs with indices $(1, 2)$, see Sect. 3.3.

We note that in general this statement does neither hold for manifolds of dimension greater than two nor for non-manifold 2-complexes.

Tightness of the Stability Bound Based on a nested sequence of persistence pairs, our cancellation method is capable of removing all pairs with persistence $\leq 2\delta$ without removing other critical points. In view of the topological simplification problem, it remains to show that the corresponding sequential cancellations do not violate the δ -tolerance constraint. Indeed, we show that a nested cancellation sequence of persistence pairs leads to a function that matches the stability bound (Theorem 16).

The main difficulty in proving this result stems from the fact that even canceling a single persistence pair might affect an arbitrary number of other critical points. As a consequence, it is not possible to *independently* apply the cancellation for persistence pairs of indices $(0, 1)$ and $(1, 2)$. This stands in contrast to previous related methods [2, 13], which modify the function independently on two disjoint subsets of the surface, corresponding to the cancellation of persistence pairs of indices $(0, 1)$ and $(1, 2)$, respectively. These methods do not alter the values of critical points that are not canceled, at the cost of giving up optimality by having to modify the function by 2δ instead of δ .

Efficient Solution Our construction of sequential cancelations is convenient for establishing tightness of the stability bound but yields an algorithm with quadratic time complexity. Devising a more efficient algorithm requires a slightly different approach.

Our (inefficient) sequential construction amounts to *alternating* between two processes: (i) canceling persistence pairs from a discrete vector field and (ii) adapting the function correspondingly. In order to obtain an efficient solution, we perform these processes *separately*: In a first step, we compute a fully simplified vector field from the persistence pairs. In a second step, we construct a simplified function from the input function and the simplified vector field resulting from the first step. Both steps can be performed in time $\mathcal{O}(n)$ using simple graph traversal methods (Sect. 5). Note that computing persistence pairs takes time $\mathcal{O}(\text{sort}(n))$ [2]. Here n denotes the number of cells of the surface and $\mathcal{O}(\text{sort}(n))$ denotes the time complexity of sorting n numbers.

The existence of an efficient algorithm for the topological simplification problem is perhaps surprising in view of the fact that the problem is NP-hard when restricted to simplexwise linear functions on a triangulated surface, which follows from extending a result by Gray et al. [17]; see [4] for details. Here, a *simplexwise* linear function on a triangulated surface is a function that is linear on each simplex of a *fixed* triangulation, while a *piecewise* linear function is only required to be linear on each simplex of *some* triangulation of the same surface. In particular, a simplexwise linear function on a subdivision of a triangulation is piecewise linear but not necessarily simplexwise linear with respect to the original triangulation. The emphasis on simplexwise linear as opposed to just piecewise linear functions is crucial here: a multiple saddle can be split into several non-degenerate saddles by an arbitrarily small perturbation (in the supremum norm) in the space of piecewise linear functions, but not in the subspace of simplexwise linear functions. This emphasizes the important role of discrete Morse theory: the hardness of the problem in the simplexwise linear setting arises from the possibility that the input contains multiple saddles, which is excluded by definition in discrete Morse theory. Going from simplexwise linear functions to discrete Morse functions can be interpreted as splitting multiple saddles (Sect. 2.3).

Energy Minimization of Simplified Functions The solution to the topological simplification problem is not unique in general: both the δ -constraint and the simplified discrete gradient vector field impose a set of linear inequalities on the simplified function, so the solution set is a convex polytope. This additionally allows to minimize a suitable convex energy functional. We employ this technique to remove artifacts from the initial solution and to improve the similarity to the input function (Sect. 6.4).

1.2 Related Work

Topological simplification of functions within a δ -tolerance constraint has been considered before by Edelsbrunner et al. [13] and Attali et al. [2]. The problem considered there differs from ours by a seemingly small but significant detail: in [2, 13] the critical points of the input function f that are *not* eliminated are additionally assumed to exactly maintain their original values. This restriction has significant consequences: while it allows for eliminating all critical points of f with persistence $\leq \delta$,

it does not in general allow to eliminate all critical points with persistence $\leq 2\delta$; an example is given in [13]. Hence, under this restriction it is not always possible to match the stability bound.

The methods presented in [2, 13] can be interpreted as variants of the so-called *carving* method proposed by Soille [33] in the context of terrain simplification. Similarly, there is another popular method for removing extrema from terrains, called *filling* or *flooding* [1, 9, 20]. A combination of both methods has been proposed in [34]. Our method of canceling critical points from a function can be interpreted as a combination of carving and flooding in the realm of discrete Morse theory.

Apart from the above mentioned works, persistent homology provides the basis for several other methods for computing and simplifying multi-scale structures derived from a function. In their original paper on persistent homology, Edelsbrunner et al. [11] discuss the closely related problem of simplifying a *filtration* of a simplicial complex. Edelsbrunner et al. [12] and Gyulassy et al. [18] consider simplification of *cell decompositions* (Morse–Smale complexes) resulting from a given gradient vector field. Unfortunately, a simplified Morse–Smale complex does not directly give rise to a simplified function. Indeed, simplifying a Morse–Smale complex is closely related to simplifying a discrete gradient vector field.

Several statements of this article can also be transferred to the setting of Morse–Smale complexes. For example, Theorem 13 can be used to show that the successive simplification of a Morse–Smale complex on a surface proposed by Edelsbrunner et al. [12] is always possible. This extends the Adjacency Lemma in [12], which shows a necessary but not sufficient condition for the successive cancelation of persistence pairs.

The problem of constructing discrete *gradient vector fields* (as opposed to functions) that minimize the number of critical points *without* constraints is addressed by Lewiner et al. [28] for surfaces and by Joswig and Pfetsch [21] for complexes of arbitrary dimension. King et al. [23] were the first to propose the combination of persistence with discrete Morse theory to simplify the gradient vector field of an input function on a 3-dimensional simplicial complex. Their method has quadratic time complexity and produces a simplified discrete gradient vector field but not a function. It does not aim at optimality since in dimension 3 not every persistence pair can be canceled.

2 Discrete Morse Theory

Classical (smooth) Morse theory [30] relates the critical points of a generic smooth real-valued function on a manifold to the global topology of that manifold. Forman [15, 16] carried over the main ideas of Morse theory to a combinatorial setting. We briefly review some important notions and results here that are used throughout this article, together with some extensions to Forman’s theory that provide important tools for our results.

A CW complex \mathcal{K} is a topological space constructed inductively: starting with a discrete set \mathcal{K}_0 of 0-cells, the n -skeleton \mathcal{K}_n is formed by attaching n -cells (closed n -dimensional balls) by continuous maps $\mathbb{S}^{n-1} \rightarrow \mathcal{K}_{n-1}$ from their boundary to the

$(n - 1)$ -skeleton. The set of cells of \mathcal{K} is denoted by K . Throughout this article, we consider only finite CW complexes. Whenever a cell $\tau \in K$ is attached to a cell σ (i.e., $\sigma \subset \partial\tau$, where $\partial\tau$ denotes the boundary of τ), we call σ a *face* of τ ; a face of codimension 1 is called a *facet*. If all attaching maps are homeomorphisms, \mathcal{K} is called a *regular CW complex*. A regular CW complex whose underlying space is a topological 2-manifold is called a *combinatorial surface*. We refer to [19, 29] for details on CW complexes.

2.1 Discrete Vector Fields

One of the central concepts of discrete Morse theory is that of a *discrete vector field*—a purely combinatorial analogue of a classical vector field.

Definition (Discrete vector field, critical cell [15, 16]) A *discrete vector field* V on a regular CW complex \mathcal{K} is a set of pairs of cells $(\sigma, \tau) \in K \times K$, with σ a facet of τ , such that each cell of K is contained in at most one pair of V . A cell $\sigma \in K$ is *critical* with respect to V if σ is not contained in any pair of V . The dimension of a critical cell is also called its *index*.

A pair (σ, τ) in a discrete vector field V can be visualized as an arrow from σ to τ (as in Fig. 2). A critical cell is also called a *critical point*.

In the following, we consider an important subclass of vector fields in which the arrows do not form closed paths. This can be made precise using the concept of *V-paths*.

Definition (*V-path* [16]) Let V be a discrete vector field. A *V-path* from a cell σ_0 to a cell σ_r is a sequence $\Gamma = (\sigma_0, \tau_0, \sigma_1, \dots, \tau_{r-1}, \sigma_r)$ of cells such that for every $0 \leq i \leq r - 1$:

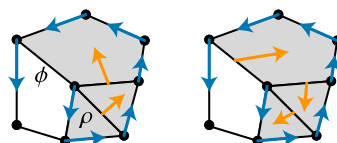
$$\sigma_i \text{ is a facet of } \tau_i \text{ with } (\sigma_i, \tau_i) \in V \quad \text{and} \quad \sigma_{i+1} \text{ is a facet of } \tau_i \text{ with } (\sigma_{i+1}, \tau_i) \notin V.$$

A *V-path* Γ is called *closed* if $\sigma_0 = \sigma_r$ and *nontrivial* if $r > 0$. We call $\dim \sigma_0$ the *dimension of* Γ .

By a *V-path* from $\partial\rho$ to ϕ we mean a *V-path* from a facet of ρ to ϕ (see Fig. 2 for an example).

Definition (Discrete gradient vector field [16]) A discrete vector field V is a *discrete gradient vector field* if it contains no nontrivial closed *V-paths*.

Fig. 2 Reversing a gradient vector field along the unique path from $\partial\rho$ to ϕ produces a gradient vector field in which the 1-cell ϕ and the 2-cell ρ are no longer critical



The main technique for reducing the number of critical points is that of *reversing* a gradient vector field V along a V -path (see Fig. 2 for an example). It provides a discrete analogue of Morse's cancellation theorem [32]:

Theorem 1 (Forman [15]) *Let ϕ and ρ be two critical cells of a gradient vector field V with exactly one V -path Γ from $\partial\rho$ to ϕ . Then there is a gradient vector field \tilde{V} obtained by reversing V along the path Γ . The critical cells of \tilde{V} are exactly the critical cells of V apart from $\{\phi, \rho\}$. Moreover, $V = \tilde{V}$ except along the path Γ .*

Gradient vector fields on combinatorial surfaces have additional properties that do not hold in higher dimensions. The following property is readily checked using the fact that a 1-cell is only attached to at most two 0-cells, and at most two 2-cells are attached to a 1-cell:

Lemma 2 *Two V -paths of dimension 0 cannot branch at a common cell, and two V -paths of dimension 1 cannot merge (except at their last cell).*

Corollary 3 *Let ρ be a critical 1-cell of a discrete vector field V on a combinatorial surface. Then there are at most two V -paths from $\partial\rho$ to critical 0-cells, each starting at one of the two 0-cells in $\partial\rho$. Similarly, there are at most two V -paths from facets of critical 2-cells to ρ .*

2.2 Pseudo-Morse Functions and Symbolic Perturbation

As in smooth Morse theory, a discrete gradient vector field can be understood as the gradient of some non-degenerate function in the following sense:

Definition (Discrete Morse function [15]) A function $f : K \rightarrow \mathbb{R}$ on the cells of a regular CW complex K is a *discrete Morse function* if there is a gradient vector field V_f such that whenever σ is a facet of τ then

$$(\sigma, \tau) \notin V_f \quad \text{implies} \quad f(\sigma) < f(\tau) \quad \text{and} \quad (\sigma, \tau) \in V_f \quad \text{implies} \quad f(\sigma) \geq f(\tau).$$

V_f is called the *gradient vector field of f* .

In contrast to simplexwise linear functions, which are determined by their function values at the vertices, discrete Morse functions take values on cells of *any* dimension.

The gradient vector field of a discrete Morse function encodes only the *sign* of the difference between function values, not the difference itself. Therefore a discrete gradient vector field does not uniquely determine a discrete Morse function, but for every discrete Morse function f there is exactly one gradient vector field V_f .

In order to be able to treat non-generic input functions, it is useful to consider a more general class of functions, which we call *pseudo-Morse functions*. Pseudo-Morse functions substitute the strict inequality in the definition of Morse functions by a weak one.

Definition (Pseudo-Morse function, consistency) A function $f : K \rightarrow \mathbb{R}$ on the cells of a regular CW complex \mathcal{K} is a *discrete pseudo-Morse function* if there is a gradient vector field V such that whenever σ is a facet of τ then

$$(\sigma, \tau) \notin V \text{ implies } f(\sigma) \leq f(\tau) \quad \text{and} \quad (\sigma, \tau) \in V \text{ implies } f(\sigma) \geq f(\tau).$$

In this case, we call f and V *consistent*.

Note that a gradient vector field V consistent with a pseudo-Morse function f is not unique in general. The following lemma provides a useful characterization of discrete pseudo-Morse functions.

Lemma 4 *Let $f : K \rightarrow \mathbb{R}$ be a function on the cells of a regular CW complex \mathcal{K} and let V be a gradient vector field on \mathcal{K} . Then f is a discrete pseudo-Morse function consistent with V if and only if for every $\epsilon > 0$ there exists a discrete Morse function $f_\epsilon : K \rightarrow \mathbb{R}$ with $\|f_\epsilon - f\|_\infty \leq \epsilon$ such that V is the gradient vector field of f_ϵ .*

Proof Assume that f is a pseudo-Morse function consistent with a gradient vector field V . There exists a discrete Morse function g whose gradient vector field V_g is precisely given by V [15]. Let G be the maximum absolute value of g . Given $\epsilon > 0$, for each cell σ define $f_\epsilon(\sigma) := f(\sigma) + \epsilon \frac{g(\sigma)}{G}$. Then it is straightforward to check that f_ϵ is a discrete Morse function with gradient vector field V and $\|f_\epsilon - f\|_\infty \leq \epsilon$.

On the other hand, assume that for every $\epsilon > 0$ there is a discrete Morse function $f_\epsilon : K \rightarrow \mathbb{R}$ consistent with V and $\|f_\epsilon - f\|_\infty \leq \epsilon$. Choose ϵ such that for every $\phi, \rho \in K$ with $f(\phi) \neq f(\rho)$ we have $\epsilon < \frac{|f(\phi) - f(\rho)|}{2}$. In this case, one easily verifies that f is a pseudo-Morse function consistent with V . □

The previous lemma provides a *symbolic perturbation* scheme based on gradient vector fields in order to allow for non-generic (degenerate) input functions. Starting with a pseudo-Morse function f , we can choose a consistent gradient vector field V , which may not be unique. Lemma 4 asserts that there is a discrete Morse function f_ϵ arbitrarily close to f and consistent with V . Therefore we can work with f as if it were a discrete Morse function with gradient vector field V . In particular, we use Lemma 4 to consider critical points associated to a pseudo-Morse function by choosing a consistent gradient vector field.

This first symbolic perturbation scheme is not sufficient for all our purposes; the definition of persistence pairs (given later in Sect. 3) not only requires a gradient vector field but also a total order on the critical cells, which again is not always uniquely defined by a pseudo-Morse function f and a consistent gradient vector field V . We now derive a second perturbation scheme that meets these requirements.

Since a gradient vector field imposes certain inequality constraints on the functions consistent with it, we can ask how these inequalities affect the relation between the function values of any two cells. We observe that any discrete gradient vector field gives rise to a strict partial order on the set of cells:

Definition (Induced partial order) Let V be a discrete gradient vector field and consider the relation \leftarrow_V defined on K such that whenever σ is a facet of τ then

$$(\sigma, \tau) \notin V \quad \text{implies} \quad \sigma \leftarrow_V \tau \quad \text{and} \quad (\sigma, \tau) \in V \quad \text{implies} \quad \sigma \rightarrow_V \tau.$$

Let \prec_V be the transitive closure of \leftarrow_V . Then \prec_V is called the (*strict*) *partial order induced by V* .

The interpretation of this partial order is that for *any* pseudo-Morse function f consistent with V and any two cells ϕ and ρ , the relation $\phi \prec_V \rho$ implies $f(\phi) \leq f(\rho)$. The relation \leftarrow_V is the *covering relation* of \prec_V , i.e., $\phi \leftarrow_V \rho$ implies $\phi \prec_V \rho$ and there is no ψ with $\phi \prec_V \psi \prec_V \rho$. The covering relation of a partial order forms a directed acyclic graph called the *Hasse diagram*. We use the convention that the edges are oriented as suggested by \leftarrow_V . The Hasse diagram H_V of \prec_V is obtained from the Hasse diagram of the face relation on K by inverting the orientation of all edges corresponding to pairs $(\sigma, \tau) \in V$, as described by Chari [5]. H_V has the property that $\phi \prec_V \rho$ if and only if there is a directed path from ρ to ϕ . Note that $\sigma \leftarrow_V \tau$ implies $f(\sigma) \leq f(\tau)$, i.e., both the arrow visualizing $(\sigma, \tau) \in V$ and the arrow symbolizing $\sigma \leftarrow_V \tau$ point toward a (weakly) *decreasing* function value of f .

Assume we are given a pseudo-Morse function f consistent with a gradient vector field V . On the one hand we have the induced partial order \prec_V . On the other hand the function f canonically induces a strict partial order \prec_f given by $\phi \prec_f \rho \Leftrightarrow f(\phi) < f(\rho)$. Since the two orders \prec_f and \prec_V are compatible by assumption (there are no two cells (ϕ, ρ) with $\phi \prec_V \rho$ and $\phi \succ_f \rho$), we can merge them into a strict partial order $\prec_{f,V}$ (the transitive closure of $(\prec_f \cup \prec_V) \subset K \times K$). A linear extension of this order is now called consistent with both f and V :

Definition (Consistent total order) Let V be a discrete gradient vector field V consistent with a discrete pseudo-Morse function f . Then a strict total order \prec is called *consistent* with (f, V) if it is a linear extension of \prec_f and \prec_V .

Such a total order \prec gives rise to a canonical injective function $K \rightarrow \mathbb{N}$ that is a discrete Morse function and consistent with V . If we use this function as the function g in the proof of Lemma 4 to construct f_ϵ , then f_ϵ is an injective discrete Morse function with gradient vector field V and the total order induced by f_ϵ is \prec again. We thus obtain a second symbolic perturbation scheme for situations where a total order on the cells is required.

We make use of this concept in the following definition. A classical object of study in smooth Morse theory is the *sublevel set* $\{x \in M : f(x) \leq t\}$ of a function $f : M \rightarrow \mathbb{R}$ on a manifold M . In the discrete theory, the analogous object is the *level subcomplex*, and the equivalent construction using our second symbolic perturbation scheme is the *order subcomplex*:

Definition (Level subcomplex [15], order subcomplex) Let f be a pseudo-Morse function on a regular CW complex \mathcal{K} . Let the *carrier* of a subset $L \subset K$ be the smallest subcomplex of \mathcal{K} containing all of L . Then for $t \in \mathbb{R}$, the *level subcomplex*

is

$$\mathcal{K}(t) = \text{carrier} \left(\bigcup_{\rho \in K: f(\rho) \leq t} \rho \right).$$

Similarly, let $<$ be a strict total order on the cells K of a regular CW complex \mathcal{K} . Then for a cell $\sigma \in K$, the *order subcomplex* is

$$\mathcal{K}(\sigma) = \text{carrier} \left(\bigcup_{\rho \in K: \rho \leq \sigma} \rho \right).$$

Like in the smooth theory, the homotopy type of level subcomplexes changes only at critical cells. The statement can trivially be rephrased for order subcomplexes:

Theorem 5 (Forman [15]) *Let V be a gradient vector field on \mathcal{K} and let $<$ be a linear extension of $<_V$. If ρ and ψ are two cells such that $\rho < \psi$ and there is no critical cell ϕ with respect to V such that $\rho < \phi \leq \psi$, then $\mathcal{K}(\psi)$ collapses to $\mathcal{K}(\rho)$.*

The order subcomplexes provide a finer filtration of the complex \mathcal{K} (by single cells or pairs of cells) than the level subcomplexes, in particular if f is degenerate. This turns out to be useful when working with persistent homology in Sect. 3.

2.3 Piecewise Linear Functions and Discrete Morse Functions

In this section we discuss a canonical relationship between discrete and piecewise linear (PL) Morse theory. As it turns out, it is possible to translate statements from one setting to the other seamlessly. Equivalent constructions have been used in [2, 23, 31].

Assume that \mathcal{K} is a simplicial complex. Let f_{PL} be a simplexwise linear function on \mathcal{K} and let f_0 be its restriction to the 0-skeleton of \mathcal{K} . The function f_0 inductively gives rise to a discrete pseudo-Morse function f in the following way. For each 0-cell α , let $f(\alpha) = f_0(\alpha)$. For a cell τ with $\dim \tau > 0$, let $f(\tau)$ be the maximum value of f on any facet of τ . The function f can easily be seen to be pseudo-Morse since it is consistent with the empty vector field $V = \emptyset$ (all cells are critical). Note that any level subcomplex of f coincides with the induced subcomplex of \mathcal{K} on the corresponding sublevel set of f_0 . This induced subcomplex, in turn, is homotopy equivalent to the corresponding sublevel set of f_{PL} [26, 31]. This means that from a Morse-theoretic point of view, the PL function f_{PL} and the pseudo-Morse function f are equivalent. We conclude:

Theorem 6 *Let f_{PL} be a simplexwise linear function on a simplicial complex \mathcal{K} . Then there is a canonical pseudo-Morse function f on \mathcal{K} such that for every $t \in \mathbb{R}$ the sublevel set $\{x \in \mathcal{K} : f_{\text{PL}}(x) \leq t\}$ is homotopy equivalent to the level subcomplex $\mathcal{K}(t)$.*

Vice versa, we can interpret any discrete pseudo-Morse function f on a regular CW complex \mathcal{K} as a simplexwise linear function $f_{\text{sd}} : |\text{sd } \mathcal{K}| \rightarrow \mathbb{R}$ on the underlying

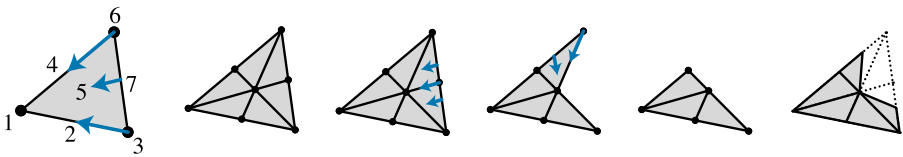


Fig. 3 Illustration to Theorem 7. The level subcomplex $\mathcal{K}(t)$ (left) is homotopy equivalent to the sublevel set $\{x \in |\text{sd } \mathcal{K}| : f_{\text{sd}}(x) \leq t\}$ (right), with $t = 5$. The homotopy equivalence is shown using a sequence of simplicial collapses of the barycentric subdivision $\text{sd } \mathcal{K}$

space of the barycentric subdivision $\text{sd } \mathcal{K}$. The *barycentric subdivision* of a regular CW complex \mathcal{K} is the *order complex* of the face relation, i.e., the *abstract simplicial complex* $\text{sd } \mathcal{K}$ whose vertices are the cells of \mathcal{K} and whose simplices are the totally ordered subsets of \mathcal{K} with regard to the face relation. The underlying space $|\text{sd } \mathcal{K}|$ is homeomorphic to \mathcal{K} [29]. The function f_{sd} is assumed to linearly interpolate the values of f at the vertices of $|\text{sd } \mathcal{K}|$ inside each simplex of $|\text{sd } \mathcal{K}|$. Again, the sublevel sets of f_{sd} are homotopy equivalent to the corresponding level subcomplexes of f :

Theorem 7 *Let f be a pseudo-Morse function on a simplicial complex \mathcal{K} . Then f induces a simplexwise linear function f_{sd} on $|\text{sd } \mathcal{K}|$ such that for every $t \in \mathbb{R}$ the level subcomplex $\mathcal{K}(t)$ is homotopy equivalent to the sublevel set $\{x \in |\text{sd } \mathcal{K}| : f_{\text{sd}}(x) \leq t\}$.*

Proof Let V be a discrete gradient vector field on \mathcal{K} that is consistent with f and let $<$ be a total order consistent with (f, V) . Let $K(t)$ and $K(\rho)$ denote the cells of the level and order subcomplexes $\mathcal{K}(t)$ and $\mathcal{K}(\rho)$, respectively. Let $\Delta(U)$ denote the induced subcomplex of $\text{sd } \mathcal{K}$ on a vertex set U . The induced subcomplex $\Delta(K(t))$ is easily seen to be identical to $\text{sd } \mathcal{K}(t)$. Let $F(t) = \{\phi \in K : f(\phi) \leq t\} \subset K(t)$. We now show that $\Delta(K(t))$ collapses simplicially onto $\Delta(F(t))$. See Fig. 3 for an example.

Let $\sigma \in K(t) \setminus F(t)$ and let σ_- denote its predecessor with respect to $<$. We write $\Delta(\rho)$ for $\Delta(\{\phi \in K : \phi < \rho\})$. We first show that $\Delta(\sigma)$ collapses onto $\Delta(\sigma_-)$. It follows from the definition of a level subcomplex that a cell σ with $f(\sigma) > t$ can only be contained in $K(t)$ if it is a *free face* of some cell τ with $f(\tau) \leq t$, i.e., σ has no other cofaces than τ . Consequently, we have $\tau < \sigma$, and for every simplex $S \in \Delta(\sigma)$ with $\sigma \in S$ and $\tau \notin S$ the simplex $T = S \cup \{\tau\}$ is also contained in $\Delta(\sigma)$. Hence, these pairs (S, T) constitute a discrete gradient vector field W on $\Delta(\sigma)$ such that exactly the simplices containing σ (the *vertex star* of σ) are non-critical. This vector field W provides a simplicial collapse of $\Delta(\sigma)$ onto $\Delta(\sigma_-)$ by applying Theorem 5 with an arbitrary linear extension of $<_W$. By repeatedly applying this argument, we find that $\Delta(K(t))$ collapses onto $\Delta(F(t))$. This implies that the underlying spaces $|\Delta(K(t))|$ and $|\Delta(F(t))|$ are homotopy equivalent.

Finally, let f_{sd} be the simplexwise linear extension of f from the vertices of $\text{sd } \mathcal{K}$ to the whole complex. Recall that $|\Delta(F(t))|$ is homotopy equivalent to the sublevel set $\{x \in |\text{sd } \mathcal{K}| : f_{\text{sd}}(x) \leq t\}$, see [26, 31]. The claim now follows. \square

This equivalence allows us to translate back and forth between piecewise linear functions and pseudo-Morse functions, and to use theorems of piecewise linear Morse theory in the context of discrete Morse theory.

In a similar fashion, a discrete pseudo-Morse function can be constructed from a function defined on the set of 2-cells of a combinatorial surface by defining $f(\sigma)$ as the minimum value of all cells that contain σ as a facet. This can be used to construct discrete pseudo-Morse functions from functions defined on cubical grids, such as pixel images, by interpreting each pixel as a 2-cell. The resulting level subcomplexes correspond to the cubical complexes extracted from images as described by Kaczynski [22]. Vice versa, a pseudo-Morse function on a cubical complex can be interpreted as a function defined on a subdivided grid. This construction has been used in the examples in Sect. 6.

Note that starting with a PL function and constructing a pseudo-Morse function consistent with the empty vector field means that initially all cells are considered critical. King et al. [23] propose to construct an initial discrete gradient vector field with critical cells corresponding to the critical vertices (in the PL sense, see [3, 14, 24]) of a (non-degenerate) input PL function instead.

We do not consider the problem of finding such an initial gradient vector field separately, since this is a special case of the topological simplification problem discussed in Sect. 4 with $\delta = 0$. In this case, the problem reduces to minimizing the number of critical points among all gradient vector fields consistent with the input function. We discuss the simplification of a gradient vector field in Sects. 3.3 and 5.3. As it turns out, a solution to the topological simplification problem can be found independently of the gradient vector field consistent with the input function. Therefore we do not require a simplified initial gradient vector field but use the empty vector field instead.

3 Persistent Homology of Discrete Morse Functions

The notions of persistent homology and persistence pairs were introduced in [6, 11, 36] in order to investigate the change of the homology groups in a filtration of a topological space (a nested sequence of subspaces). This concept can naturally be applied to discrete pseudo-Morse functions. The following definitions are concerned with cellular homology with coefficients in an arbitrary field F ; however, in all our statements we assume \mathbb{Z}_2 coefficients when working with non-orientable surfaces. For definitions we refer to [19, 29]. We write $H_d(\mathcal{K})$ as a shorthand for the d^{th} homology group $H_d(\mathcal{K}; F)$ of \mathcal{K} and $H_*(\mathcal{K}) = \bigoplus_d H_d(\mathcal{K})$.

Convention and Notation Throughout Sect. 3 we consider a pseudo-Morse function f consistent with a gradient vector field V on a regular CW complex \mathcal{K} and a strict total order $<$ consistent with (f, V) .

3.1 Birth, Death, and Persistence Pairs

As a consequence of Theorem 5, the homology groups of order subcomplexes change only at critical cells of V . Let σ and τ be critical cells such that $\sigma < \tau$ and consider the inclusion map $i^{\sigma, \tau} : \mathcal{K}(\sigma) \hookrightarrow \mathcal{K}(\tau)$ between the order subcomplexes with regard to the total order $<$. This map induces a homomorphism

$i_*^{\sigma, \tau} : H_*(\mathcal{K}(\sigma)) \rightarrow H_*(\mathcal{K}(\tau))$ between homology groups. For every cell ρ , let ρ_- denote its predecessor with regard to \prec . Now consider the sequence

$$H_*(\mathcal{K}(\sigma_-)) \rightarrow H_*(\mathcal{K}(\sigma)) \rightarrow H_*(\mathcal{K}(\tau_-)) \rightarrow H_*(\mathcal{K}(\tau))$$

of induced homomorphisms. Here we allow for the cases $\sigma = \tau_-$ and $\mathcal{K}(\sigma_-) = \emptyset$ (if σ is the first cell in \prec).

Definition (Positive/negative cell, birth/death, persistence pair [11]) We say that σ is a *positive cell* and a class $h \in H_*(\mathcal{K}(\sigma))$ is *born at σ* (or *created by σ*) if

$$h \notin \text{im}(i_*^{\sigma_-, \sigma}).$$

Moreover, we say that τ is a *negative cell* and a class $h \in H_*(\mathcal{K}(\sigma))$ that is born at σ *dies entering τ* (or *gets merged by τ*) if there is a class $\tilde{h} \in H_*(\mathcal{K}(\sigma_-))$ such that

$$i_*^{\sigma, \tau_-}(h) \notin \text{im}(i_*^{\sigma_-, \tau_-}) \quad \text{but} \quad i_*^{\sigma, \tau}(h) = i_*^{\sigma_-, \tau}(\tilde{h}) \in \text{im}(i_*^{\sigma_-, \tau}).$$

If there exists a class h that is born at σ and dies entering τ , then (σ, τ) is a *persistence pair*. The difference $f(\tau) - f(\sigma)$ is called the *persistence* of (σ, τ) .

Note that in this definition we always have $\dim \tau = \dim \sigma + 1$. On combinatorial surfaces, the only possible cases for $(\dim \sigma, \dim \tau)$ are $(0, 1)$ or $(1, 2)$.

3.2 Duality and Persistence

For any closed combinatorial surface \mathcal{K} , there is an associated *dual complex* \mathcal{K}^* , a combinatorial surface homeomorphic to \mathcal{K} whose i -cells correspond to $(2 - i)$ -cells of \mathcal{K} [19]. A discrete pseudo-Morse function f on \mathcal{K} gives rise to a discrete pseudo-Morse function f^* on \mathcal{K}^* via $\sigma^* \mapsto -f(\sigma)$ [15].

Moreover, the persistence pairs of dimension $(1, 2)$ for \mathcal{K} correspond to the persistence pairs of dimension $(0, 1)$ for the dual complex \mathcal{K}^* (with $\tau^* \prec \sigma^* \Leftrightarrow \sigma \prec \tau$) when using \mathbb{Z}_2 coefficients or if \mathcal{K} is orientable [2, 7, 10].

The homology groups $H_0(\mathcal{K}(\rho_i))$ (generated by the connected components of $\mathcal{K}(\rho_i)$), and hence the persistence pairs of dimension $(0, 1)$, are determined solely by the 1-skeleton of \mathcal{K} , also called the (*primal*) *graph* of \mathcal{K} . Consequently, the persistence pairs of dimension $(1, 2)$ are determined by the 1-skeleton of \mathcal{K}^* , called the *dual graph*. This means that all persistence pairs of a surface can be determined in terms of Morse functions on graphs.

In order to treat surfaces with boundary, we employ the canonical construction of attaching an additional 2-cell (with function value ∞) to each boundary component. In this way we obtain a closed surface having the same sequence of order subcomplexes (up to the additional cells) and hence the same persistence pairs as the original surface.

3.3 The Persistence Hierarchy and Sequential Cancellations

Persistence pairs on surfaces carry a certain hierarchical structure that allows us to establish a connection to the cancellation theorem of discrete Morse theory. The main result of this section is that persistence pairs on surfaces can always be canceled sequentially if the order of cancellations respects this hierarchy.

Definition (Parent, child, persistence hierarchy) On a combinatorial surface \mathcal{K} , let (σ, τ) be a persistence pair with $\dim \sigma = 0$, and let $[\sigma] \in H_0(\mathcal{K}(\sigma))$ be the class created by σ (considered as a 0-cycle). Let $\tilde{\sigma}$ be the unique cell creating the class $[\tilde{\sigma}] \in H_0(\mathcal{K}(\tau))$ into which $[\sigma]$ gets merged by τ , i.e., $[\tilde{\sigma}] \notin \text{im}(i_*^{\tilde{\sigma}, \tau})$ and $[\tilde{\sigma}] = i_*^{\sigma, \tau}([\sigma])$. Then $\tilde{\sigma}$ is called the *parent* of σ (in the *persistence hierarchy*), and σ is called the *child* of $\tilde{\sigma}$. The transitive closure of the child relation is called *descendant*.

Let (σ, τ) and $(\tilde{\sigma}, \tilde{\tau})$ be two persistence pairs. If either $\dim \sigma = \dim \tilde{\sigma} = 0$ and $\tilde{\sigma}$ is the parent of σ or $\dim \tau = \dim \tilde{\tau} = 2$ and $\tilde{\tau}^*$ is the parent of τ^* (with regard to the persistence hierarchy on the dual complex), then we also call the pair $(\tilde{\sigma}, \tilde{\tau})$ the *parent* of (σ, τ) and (σ, τ) the *child* of $(\tilde{\sigma}, \tilde{\tau})$. The following definition and lemma justify this nomenclature:

Definition (Nested pairs) On a combinatorial surface \mathcal{K} , let (σ, τ) and $(\tilde{\sigma}, \tilde{\tau})$ be two persistence pairs. We say that (σ, τ) is *nested* in $(\tilde{\sigma}, \tilde{\tau})$ if $\tilde{\sigma} < \sigma < \tau < \tilde{\tau}$.

Lemma 8 Let (σ, τ) be a descendant of $(\tilde{\sigma}, \tilde{\tau})$ in the persistence hierarchy. Then (σ, τ) is nested in $(\tilde{\sigma}, \tilde{\tau})$.

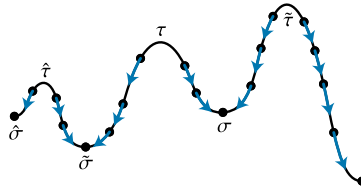
Proof Without loss of generality, assume $\dim \sigma = 0$; otherwise, by duality, the argument can be applied to (τ^*, σ^*) instead of (σ, τ) .

By definition of the persistence hierarchy, the class $[\sigma]$ gets merged into the class $[\tilde{\sigma}] \in H_0(\mathcal{K}(\tau))$ created by $\tilde{\sigma}$. This implies that $\tilde{\sigma} < \sigma$. It also implies that the class created by $\tilde{\sigma}$ has not been merged by any cell of $\mathcal{K}(\tau)$, hence $\tau < \tilde{\tau}$. \square

We now turn our attention to the sequential cancellation of persistence pairs. Note that the cancellation theorem (Theorem 1) applies to vector fields, which only provide a *partial order* on the cells, while the notion of persistence is based on a *total order*. After canceling a persistence pair, the new vector field is no longer consistent with the initial total order. It is important to keep in mind that we only talk about persistence pairs of the *initial* total order $<$, which is consistent with (f, V) ; we do not consider a new total order after applying a cancellation (which would complicate things considerably). Applying several cancellations results in a sequence of simplified vector fields:

Definition (Persistence cancellation sequence) A *persistence cancellation sequence* is a sequence of gradient vector fields (V_0, V_1, \dots, V_n) such that each V_i is constructed from V_{i-1} by canceling a persistence pair (σ_i, τ_i) using Theorem 1.

Fig. 4 The persistence hierarchy. Both (σ, τ) and $(\hat{\sigma}, \hat{\tau})$ are children of, and hence nested in, $(\tilde{\sigma}, \tilde{\tau})$. Only (σ, τ) needs to be canceled before $(\tilde{\sigma}, \tilde{\tau})$ can be canceled



A persistence cancellation sequence is called *nested* if in this construction every pair (σ_i, τ_i) nested in another pair (σ_j, τ_j) is canceled first, i.e.,

$$\sigma_j < \sigma_i < \tau_i < \tau_j \implies i < j.$$

A persistence cancellation sequence is called a δ -persistence cancellation sequence if exactly those persistence pairs are canceled that have persistence $\leq \delta$.

A persistence pair (σ, τ) can be canceled from a vector field as soon as all descendants have been canceled (compare also to Edelsbrunner et al. [12] for the existence part of the following statement in a special case):

Lemma 9 *On a combinatorial surface \mathcal{K} , let (V_0, V_1, \dots, V_i) be a persistence cancellation sequence. Assume that a persistence pair (σ, τ) has not been canceled in the sequence but that every descendant of (σ, τ) has been canceled. Then there exists a V_i -path from $\partial\tau$ to σ and this path is unique.*

Assume further that every persistence pair nested in (σ, τ) has been canceled. If there is a unique V_i -path from $\partial\tau$ to another cell $\tilde{\sigma} \neq \sigma$ that is critical in V_i then we have $\sigma > \tilde{\sigma}$.

Figure 4 shows that the condition is sufficient but not necessary. The proof of Lemma 9 relies on a few auxiliary lemmas and is given after Lemma 12.

Lemma 10 *Let (V_0, V_1, \dots, V_i) be a persistence cancellation sequence and let (σ, τ) be a persistence pair with $\dim \sigma = 0$ that has not been canceled in the sequence. Let \mathcal{C} be the connected component of the subcomplex $\mathcal{K}(\tau_-)$ containing σ , and let C denote the cells of \mathcal{C} . Then every $(\phi, \rho) \in V_i$ with $\dim \phi = 0$ satisfies $\phi \in C \Leftrightarrow \rho \in C$.*

Proof The claim is shown by induction over i . The base case follows from consistency of the total order $<$ with (f, V) . Consider the cancellation of a persistence pair (σ_i, τ_i) . If $\dim \sigma_i \neq 0$, the pairs in V_i of dimensions $(0, 1)$ stay unchanged and the claim immediately follows from the induction hypothesis. Now assume $\dim \sigma_i = 0$. We show that the claim holds for every $(\phi, \rho) \in V_i \setminus V_{i-1}$.

The non-critical cells of the vector field $V_i \setminus V_{i-1}$ are τ_i and the cells on the V_{i-1} -path $(\phi_0, \rho_0, \phi_1, \dots, \rho_{r-1}, \phi_r)$ from $\phi_0 \in \partial\tau_i$ to $\phi_r = \sigma_i$. By the induction hypothesis we have $\phi_k \in C \Leftrightarrow \rho_k \in C$. Because \mathcal{C} is a subcomplex, we also have $\rho_{k-1} \in C \Rightarrow \phi_k \in C$ (with $\rho_{-1} = \tau_i$). Moreover, if $\sigma_i \in C$, then σ_i is a descendant of σ and by Lemma 8 (σ_i, τ_i) is nested in (σ, τ) , implying that σ_i and τ_i are in the

same connected component of $\mathcal{K}(\tau_-)$. Hence we also have $\sigma_i \in C \Rightarrow \tau_i \in C$. Consequently, either all or none of the non-critical cells of $V_i \setminus V_{i-1}$ are contained in C and the claim immediately follows. \square

We also require the notion of the restriction of a vector field to a subcomplex:

Definition (Restriction of a vector field to a subcomplex) Let V be a discrete vector field on \mathcal{K} and let $\tilde{\mathcal{K}}$ be a subcomplex of \mathcal{K} with cells \tilde{K} . The restriction of V to $\tilde{\mathcal{K}}$ is $\tilde{V} = V \cap (\tilde{K} \times \tilde{K})$, i.e., the pairs of cells in V that are both in \tilde{K} .

As a direct consequence of this definition, new critical cells may arise when restricting a vector field to a subcomplex:

Lemma 11 Let \tilde{V} be the restriction of a discrete vector field V on \mathcal{K} to a subcomplex $\tilde{\mathcal{K}}$. The critical d -cells of \tilde{V} are exactly the critical d -cells of V that are contained in $\tilde{\mathcal{K}}$ if and only if each pair $(\sigma, \tau) \in V$ with $\dim \sigma = d$ satisfies $\sigma \in \tilde{K} \Leftrightarrow \tau \in \tilde{K}$.

Moreover, we use the following fact:

Lemma 12 Let V be a discrete gradient vector field V on \mathcal{K} with only one critical 0-cell σ . Then there is a V -path from every 0-cell $\tilde{\sigma}$ to σ .

Proof Each V -path of dimension 0 ending at a non-critical cell $\tilde{\sigma} \neq \sigma$, $(\tilde{\sigma}, \tilde{\tau}) \in V$, can be extended by $\tilde{\tau}$ and the unique 0-cell $\hat{\sigma} \in \partial \tilde{\tau}$, $\hat{\sigma} \neq \tilde{\sigma}$. Since K is finite and V does not contain nontrivial closed paths, the extension will eventually end up at σ . \square

Proof of Lemma 9 Without loss of generality, assume $\dim \sigma = 0$; otherwise, by duality, the argument can be applied to (τ^*, σ^*) instead of (σ, τ) .

Let \mathcal{C} be the connected component of the subcomplex $\mathcal{K}(\tau_-)$ corresponding to the homology class $[\sigma] \in H_0(\mathcal{K}(\tau_-))$ created by σ . Apart from σ , every 0-cell in \mathcal{C} that is critical in V is a descendant of σ . By assumption, all descendants of σ have been canceled, and hence σ is the only 0-cell in \mathcal{C} that is critical in V_i . By Lemmas 10 and 11, σ is also the only critical 0-cell in the restriction of V_i to \mathcal{C} . By Lemma 12, there is a V_i -path to σ from every 0-cell in \mathcal{C} , in particular from exactly one of the two 0-cells in $\partial \tau$, since $\partial \tau \cap \mathcal{C}$ contains exactly one cell. By Lemma 2, this path is unique.

Now assume that every persistence pair nested in (σ, τ) has been canceled and there is a unique V_i -path from $\partial \tau$ to another cell $\tilde{\sigma} \neq \sigma$ that is critical in V_i . By assumption, $\tilde{\sigma}$ is not a descendant of σ , meaning that $\tilde{\sigma}$ and σ are in different connected components of $\mathcal{K}(\tau_-)$. Moreover, the component $\tilde{\mathcal{C}} \neq \mathcal{C}$ of $\mathcal{K}(\tau_-)$ containing $\tilde{\sigma}$ is created by $\tilde{\sigma}$, because otherwise there would be a persistence pair $(\tilde{\sigma}, \tilde{\tau})$ nested in (σ, τ) . This persistence pair would have been canceled in V_i , contradicting the assumption that $\tilde{\sigma}$ is critical. Since τ is paired with σ and merges $\tilde{\mathcal{C}}$ and \mathcal{C} , we know that σ is a descendant of $\tilde{\sigma}$ and $\sigma \succ \tilde{\sigma}$. \square

As a consequence of Lemma 9, we can construct a sequence of cancelations to eliminate all persistence pairs below a certain persistence threshold:

Theorem 13 *Let f be a pseudo-Morse function on a combinatorial surface \mathcal{K} and let $\delta \geq 0$. Then there exists a nested δ -persistence cancellation sequence.*

Proof If the subsequence $(V_0, V_1, \dots, V_{i-1})$ satisfies the assumptions of Lemma 9 for some persistence pair (σ_i, τ_i) , we can use Theorem 1 to construct V_i from V_{i-1} . A canonical choice satisfying these assumptions is given by canceling the persistence pairs (σ_i, τ_i) with persistence $\leq \delta$ according to the order $<$ on the negative cells, i.e., $\tau_i < \tau_{i+1}$ for every i . The claim follows by induction. \square

3.4 The Stability Bound

Cohen-Steiner et al. [6] studied properties of *persistence diagrams*, which are a representation of the value pairs $(f(\sigma), f(\tau))$ corresponding to the persistence pairs (σ, τ) of a function f . Here we use $\mathbb{R} = \mathbb{R} \cup \{-\infty, \infty\}$.

Definition (Persistence diagram [6]) The persistence diagram $D(f)$ of a pseudo-Morse function f is the multiset on $\mathbb{C} \times \overline{\mathbb{R}}^2$ containing one instance of $(f(\sigma), f(\tau))$ for each persistence pair (σ, τ) of f , one instance of $(f(\sigma), \infty)$ for each unpaired positive cell σ , and each point on the diagonal with countably infinite multiplicity.

The main result of [6] is the *Bottleneck Stability Theorem* for persistence diagrams: if two functions are close in the supremum norm, then their persistence diagrams are also close with respect to the following metric:

Definition (Bottleneck distance) Let X and Y be two multisets of $\overline{\mathbb{R}}^2$. The *bottleneck distance* is $d_B(X, Y) := \inf_{\gamma} \sup_{x \in X} \|x - \gamma(x)\|_{\infty}$, where γ ranges over all bijections from the individual elements of X to the individual elements of Y .

Here we assume $(a, \infty) - (b, \infty) = (a - b, 0)$, $(a, \infty) - (b, c) = (a - b, \infty)$, and $\|(a, \infty)\|_{\infty} = \infty$ for $a, b, c \in \mathbb{R}$. By the *individual elements* of a multiset S we mean the disjoint union of $m(e)$ instances of each $e \in S$, where $m(e)$ is the multiplicity of e .

Due to the correspondence between piecewise linear functions and discrete pseudo-Morse functions (see Sect. 2.3), the statement translates as follows using the language of discrete Morse theory:

Theorem 14 (Cohen-Steiner et al. [6]) *Let $f, g : K \rightarrow \mathbb{R}$ be two discrete pseudo-Morse functions. Then the respective persistence diagrams satisfy*

$$d_B(D(f), D(g)) \leq \|f - g\|_{\infty}.$$

Note that the bottleneck distance provides a metric on the persistence diagrams of pseudo-Morse functions on \mathcal{K} , in particular, $d_B(D(f), D(g)) = 0$ if and only if $D(f) = D(g)$. Therefore, in contrast to the *persistence pairs*, the *persistence diagram*

of a discrete pseudo-Morse function f is well-defined; in particular, it is independent of the total order \prec chosen and even independent of the gradient vector field V consistent with f . Theorem 14 provides a lower bound on the number of critical points among all pseudo-Morse functions f_δ with $\|f_\delta - f\|_\infty \leq \delta$:

Corollary 15 (Stability bound) *Let f be a discrete pseudo-Morse function and let $\delta \geq 0$. For any pseudo-Morse function f_δ with $\|f_\delta - f\|_\infty \leq \delta$, the number of critical points of f_δ is bounded from below by the number of those critical points of f that have persistence $> 2\delta$.*

Proof Let D and D_δ be the persistence diagrams of f and f_δ , respectively. By Theorem 14 we have $d_B(D, D_\delta) \leq \delta$. This means that there is a bijection γ between D and D_δ with $\|p - \gamma(p)\|_\infty \leq \delta$ for all $p \in D$. Let

$$p = (p^*, p^\dagger) = (f(\sigma), f(\tau)) \in D$$

represent a persistence pair (σ, τ) of f with persistence $p^\dagger - p^* > 2\delta$. Letting

$$q = (q^*, q^\dagger) := \gamma(p),$$

this implies that $p^* + \delta \geq q^*$ and $p^\dagger - \delta \leq q^\dagger$. Together with $p^\dagger - p^* > 2\delta$, this yields $q^\dagger - q^* > 0$. Hence there must be a persistence pair of f_δ corresponding to each persistence pair of f with persistence $> 2\delta$. Since moreover $\|p - \gamma(p)\|_\infty \leq \infty$ implies that γ must map the unpaired critical cells of f exactly to those of f_δ , the claim follows. □

4 Function Simplification Guided by Discrete Gradient Vector Fields

We are interested in functions that achieve the lower bound of Corollary 15:

Definition (Perfect δ -simplification) *Let f be a pseudo-Morse function on a combinatorial surface \mathcal{K} and let $\delta \geq 0$. A perfect δ -simplification of f is a pseudo-Morse function f_δ such that $\|f_\delta - f\|_\infty \leq \delta$ and the number of persistence pairs of f_δ equals the number of those persistence pairs of f that have persistence $> 2\delta$.*

In this section, we prove the following central result:

Theorem 16 (Tightness of the stability bound) *Let f be a discrete pseudo-Morse function on a combinatorial surface and let $\delta \geq 0$. Then there exists a perfect δ -simplification of f . In particular, the stability bound is tight for every discrete pseudo-Morse function f on a combinatorial surface and for every $\delta \geq 0$.*

The proof of Theorem 16 is constructive and hence leads to an algorithm. The corresponding construction is outlined in Sect. 4.1. However, the resulting algorithm has a running time that is quadratic in the input size. We present a more efficient

algorithm in Sect. 5. Nevertheless, our proof of correctness for this algorithm will make use of the fact that Theorem 16 has already been established.

Using Lemma 4, the result of Theorem 16 can also be stated for (non-degenerate) discrete Morse functions (in a slightly different form, because only critical points with persistence $< 2\delta$ can be eliminated within a tolerance of δ in the set of discrete Morse functions):

Corollary 17 *Given a discrete Morse function f on a surface and $\delta > 0$, there exists a discrete Morse function f_δ such that $\|f_\delta - f\|_\infty < \delta$ and the number of critical points of f_δ equals the number of those critical points of f that have persistence $\geq 2\delta$.*

Convention and Notation Throughout this section we consider a given pseudo-Morse function f consistent with a gradient vector field V on a combinatorial surface \mathcal{K} , a strict total order $<$ consistent with (f, V) , and a nested 2δ -persistence cancellation sequence (V_0, \dots, V_n) with $V_0 = V$. Moreover, we let $<_j := <_{V_j}$ denote the partial order induced by V_j .

4.1 The Plateau Function

For every V_i in the cancellation sequence, we inductively define a pseudo-Morse function f_i consistent with V_i , see Fig. 1 for an illustration. By assumption we start with a pseudo-Morse function $f_0 := f$ consistent with $V_0 := V$. Suppose that we have constructed a pseudo-Morse function f_{i-1} consistent with V_{i-1} . Let (σ, τ) be the persistence pair that is canceled in the construction of V_i from V_{i-1} using Theorem 1.

$$m_i = \frac{f(\sigma) + f(\tau)}{2} \quad \text{and} \quad f_i(\rho) := \begin{cases} m_i & \text{if } \rho \succeq_{i-1} \sigma \text{ and } f_{i-1}(\rho) < m_i \\ & \text{or } \rho \preceq_{i-1} \tau \text{ and } f_{i-1}(\rho) > m_i, \\ f_{i-1}(\rho) & \text{otherwise.} \end{cases}$$

This means that the *ascending set* $\{\rho : \rho \succeq_{i-1} \sigma\}$ of σ in V_{i-1} is raised to at least the value m_i , and analogously the *descending set* $\{\rho : \rho \preceq_{i-1} \tau\}$ of τ is lowered. Hence, f_i creates a local *plateau* at the value m_i . The following lemma is a direct consequence of the way we construct f_i from f_{i-1} and the fact that f_i is constant along the path from $\partial\tau$ to σ . It can be proven using a straightforward induction argument.

Lemma 18 *The plateau function f_i is consistent with both V_{i-1} and V_i .*

Note that the construction of the plateau function does not depend on the properties of combinatorial surfaces but can be applied to regular CW complexes of *arbitrary dimensions*. Moreover, it is not restricted to the cancellation of persistence pairs: whenever we have a pseudo-Morse function f consistent with a gradient vector field V and \tilde{V} is constructed from V by a cancellation using Theorem 1, we can obtain a plateau function \tilde{f} that is consistent with both V and \tilde{V} .

4.2 Checking the Constraint

It remains to show that the plateau construction above produces admissible functions for the topological simplification problem, i.e., that each of the functions f_i satisfies the δ -constraint.

Lemma 19 *Each plateau function f_i satisfies $\|f_i - f\|_\infty \leq \delta$.*

Proof We show the statement by induction. The base case is trivial since $f_0 = f$.

Let (σ, τ) be the persistence pair that is canceled when constructing V_i from V_{i-1} . We show that the δ -constraint is neither violated by increasing the value of any cell ρ in the ascending set of σ in V_{i-1} , nor by decreasing the value of any cell in the descending set of τ . Since $f_i(\rho) = f_{i-1}(\rho)$ for all cells ρ not treated in these two cases, the claim follows.

We first show $|f_i(\rho) - f(\rho)| \leq \delta$ for any cell $\rho \succeq_{i-1} \sigma$ with $f_{i-1}(\rho) < m_i$. By induction hypothesis we have a lower bound $f_{i-1}(\rho) \geq f(\rho) - \delta$. By the construction of f_i , the value of ρ is increased: $f_i(\rho) = m_i > f_{i-1}(\rho)$. Therefore, the lower bound remains valid after step i :

$$f_i(\rho) > f_{i-1}(\rho) \geq f(\rho) - \delta.$$

To show the upper bound $f_i(\rho) \leq f(\rho) + \delta$, we first use $f(\tau) - f(\sigma) \leq 2\delta$ to obtain

$$f_i(\rho) = m_i = \frac{f(\sigma) + f(\tau)}{2} \leq \frac{f(\sigma) + (f(\sigma) + 2\delta)}{2} = f(\sigma) + \delta.$$

This is almost the desired inequality except that the right hand side contains $f(\sigma)$ instead of $f(\rho)$. To finish the proof, it therefore suffices to show that $f(\sigma) \leq f(\rho)$. This, in turn, is a consequence of the facts that, according to the upcoming Lemma 21, $\sigma \prec_{i-1} \rho$ implies $\sigma \prec \rho$, and that \prec is consistent with (f, V) .

It remains to show $|f_i(\rho) - f(\rho)| \leq \delta$ for any cell $\rho \preceq_{i-1} \tau$ with $f_{i-1}(\rho) > m_i$. The proof of this statement is analogous to the above. \square

In order to prove Lemma 21, we first need to investigate how the reversal of a gradient vector field may change the induced partial order (see Fig. 5 for an example):

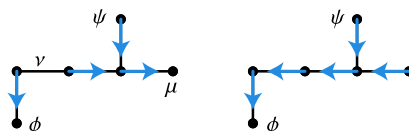


Fig. 5 Example illustrating Lemma 20. *Left:* gradient vector field W (before reversing the path from ∂v to μ). *Right:* gradient vector field \tilde{W} (after path reversal). Note that we have the new relation $\phi \prec_{\tilde{W}} \psi$ (corresponding in this example to a \tilde{W} -path from ψ to ϕ). In the example, the conclusion $\phi \preceq_W v$ and $\mu \preceq_W \psi$ of Lemma 20 is reflected by the two W -paths from ∂v to ϕ and from ψ to μ , respectively

Lemma 20 *Let μ, ν, ϕ, ψ be not necessarily disjoint cells of a regular CW complex \mathcal{K} . Let W and \tilde{W} be two gradient vector fields such that the cells μ and ν are critical for W . Assume further that there is a unique W -path from $\partial\nu$ to μ , and \tilde{W} is constructed by reversing W along this path. If $\phi \not\prec_W \psi$ and $\phi \prec_{\tilde{W}} \psi$, then $\phi \leq_W \nu$ and $\mu \leq_W \psi$.*

Proof By definition of the induced partial order, $\phi \prec_{\tilde{W}} \psi$ implies that there exists a sequence (ρ_1, \dots, ρ_k) with $\rho_1 = \phi$, $\rho_k = \psi$ and $\rho_i \leftarrow_{\tilde{W}} \rho_{i+1}$ for all $1 \leq i \leq k - 1$. Here either ρ_i is a facet of ρ_{i+1} or ρ_{i+1} is a facet of ρ_i , and we therefore also have either $\rho_i \leftarrow_W \rho_{i+1}$ or $\rho_i \rightarrow_W \rho_{i+1}$. But since $\phi \not\prec_W \psi$, there exists a smallest index j such that $\rho_j \rightarrow_W \rho_{j+1}$. Since the relations \leftarrow_W and $\leftarrow_{\tilde{W}}$ differ only along the W -path from $\partial\nu$ to μ (including ν), it follows that the cells ρ_j and ρ_{j+1} are contained in this W -path. Hence we have $\rho_j \leq_W \nu$. Moreover, by the choice of j we have $\phi = \rho_1 \leq_W \rho_j$. Therefore we conclude that $\phi \leq_W \nu$. By an analogous argument one also shows that $\mu \leq_W \psi$. □

Lemma 21 *Let (V_0, \dots, V_n) be a nested persistence cancellation sequence and let (σ, τ) be a persistence pair of \prec with σ and τ critical cells of V_i . Then for any $\rho \in K$,*

- (a) $\rho \succ_i \sigma$ implies $\rho \succ \sigma$, and
- (b) $\rho \prec_i \tau$ implies $\rho \prec \tau$.

Proof We only present the proof of part (a), which is done again by induction: We show that $\rho \succ_i \sigma$ implies $\rho \succ \sigma$ for all $0 \leq i \leq n$. Part (b) can be shown analogously.

The base case $i = 0$ is trivial since \succ is a linear extension of \succ_0 . Assume that $\rho \succ_i \sigma$. If $\rho \succ_{i-1} \sigma$, the claim follows directly from the induction hypothesis. Hence we assume that $\rho \not\prec_{i-1} \sigma$. Let $(\tilde{\sigma}, \tilde{\tau})$ be the persistence pair that is canceled when constructing V_i from V_{i-1} ; this implies $\tilde{\sigma} \prec_{i-1} \tilde{\tau}$. From Lemma 20 with $(W, \tilde{W}) = (V_{i-1}, V_i)$ and $(\mu, \nu, \phi, \psi) = (\tilde{\sigma}, \tilde{\tau}, \sigma, \rho)$, we infer that $\sigma \leq_{i-1} \tilde{\tau}$ and $\tilde{\sigma} \leq_{i-1} \rho$. This has two consequences:

- (i) $\sigma \prec_{i-1} \tilde{\tau}$ (since σ is critical in V_i while $\tilde{\tau}$ is not), and
- (ii) $\tilde{\sigma} \leq \rho$ (by the induction hypothesis).

To finish the proof of the claim, by (ii) it suffices to show that $\sigma \prec \tilde{\sigma}$. We proceed by case analysis on the dimensions of $\tilde{\sigma}$ and σ . Since these two cells are paired by assumption, they have dimension less than 2.

Case 1 ($\dim \sigma = 1, \dim \tilde{\sigma} = 0$): This case cannot occur since reversing the V_{i-1} -path from the 1-cell $\tilde{\tau}$ to the 0-cell $\tilde{\sigma}$ does not change the ascending set of any critical 1-cell (and in particular σ), contradicting $\rho \not\prec_{i-1} \sigma$ and $\rho \succ_i \sigma$.

Case 2 ($\dim \sigma = 0, \dim \tilde{\sigma} = 1$): First assume $\tau \prec \tilde{\tau}$. If additionally $\tilde{\sigma} \prec \sigma$, this contradicts the assumption that the cancellation sequence is nested and (σ, τ) is canceled after $(\tilde{\sigma}, \tilde{\tau})$. Therefore $\tau \prec \tilde{\tau}$ implies $\sigma \prec \tilde{\sigma}$.

Now assume $\tau \succ \tilde{\tau}$. This means that σ creates a connected component that is not yet merged in $\mathcal{K}(\tilde{\tau})$. Since $\sigma \prec_{i-1} \tilde{\tau}$ by (i), there is a sequence (ρ_1, \dots, ρ_k) with $\rho_1 = \sigma$, $\rho_k = \tilde{\tau}$, and $\rho_j \leftarrow_{V_{i-1}} \rho_{j+1}$ for all $1 \leq j \leq k - 1$. For each ρ_j we trivially have $\rho_j \prec_{i-1} \tilde{\tau}$ and hence $\rho_j \prec \tilde{\tau}$ by the induction hypothesis, implying that

$\rho_j \in \mathcal{K}(\tilde{\tau})$. Moreover, since either ρ_j is a facet of ρ_{j+1} or ρ_{j+1} is a facet of ρ_j , we know that all ρ_j , and in particular σ and $\tilde{\tau}$, are in the same connected component of $\mathcal{K}(\tilde{\tau})$. In an analogous way one shows that $\tilde{\sigma}$ and $\tilde{\tau}$, and hence σ and $\tilde{\sigma}$, are in one and the same connected component. Since we know that σ created that component, it follows that $\sigma < \tilde{\sigma}$.

Case 3 ($\dim \sigma = \dim \tilde{\sigma} \in \{0, 1\}$): The relation $\sigma <_{i-1} \tilde{\tau}$ from (i) above implies the existence of a V_{i-1} -path from $\partial \tilde{\tau}$ to σ . We will show by contradiction that this path must be unique. To see this, assume that there are two V_{i-1} -paths from $\tilde{\tau}$ to σ . Without loss of generality, assume that $\dim \sigma = \dim \tilde{\sigma} = 0$ (and hence $\dim \tilde{\tau} = 1$); otherwise, by duality the following argument can be applied to $\sigma^*, \tilde{\tau}^*$ instead of $\tilde{\tau}, \sigma$. By Corollary 3, each of the 0-cells in $\partial \tilde{\tau}$ must belong to exactly one of the two V_{i-1} -paths from $\partial \tilde{\tau}$ to σ . Now by a similar argument as in Case 2 above, we see that each cell of these two V_{i-1} -paths is contained in the same connected component of $\mathcal{K}(\tilde{\tau}_-)$ as σ . But since $\tilde{\tau}$ is a negative 1-cell, the two 0-cells in its boundary belong to different connected components of $\mathcal{K}(\tilde{\tau}_-)$, a contradiction.

Hence, there is a unique V_{i-1} -path from $\partial \tilde{\tau}$ to σ . Lemma 9 asserts that $\tilde{\sigma}$ is the largest cell (with regard to $<$) with a unique V_{i-1} -path from $\partial \tilde{\tau}$ to $\tilde{\sigma}$. Since $\sigma \neq \tilde{\sigma}$, we obtain $\sigma < \tilde{\sigma}$. □

Proof of Theorem 16 According to Theorem 13 there exists a nested 2δ -persistence cancellation sequence (V_0, V_1, \dots, V_n) for the gradient vector field $V_0 = V$ consistent with the pseudo-Morse function f . Let f_n be the plateau function corresponding to V_n . Since f_n is consistent with V_n by Lemma 18 and $\|f_n - f\|_\infty \leq \delta$ by Lemma 19, it is a perfect δ -simplification. □

5 An Efficient Algorithm

The definition of the plateau function in Sect. 4.1 canonically leads to an algorithm that runs in time quadratic in the input size. In this section we present a method for computing a perfect δ -simplification in time dominated by the computation of persistence pairs, i.e., $\mathcal{O}(\text{sort}(n))$, where $n = |K|$ is the number of cells of \mathcal{K} . Apart from this computation, all steps of our algorithm take linear time $\mathcal{O}(n)$. We stress that pre- and post-processing steps, like conversion from and to PL functions, also require only linear time $\mathcal{O}(n)$.

The algorithm can be summarized as follows. First, persistence pairs are computed using a variant of Kruskal’s algorithm for minimum spanning trees. Next, the persistence pairs are used to construct a simplified gradient vector field by graph traversals of both the primal and dual 1-skeleton. In a third step, the simplified vector is used to compute the simplified function by a graph traversal on the Hasse diagram of the partial order induced by the simplified vector field.

5.1 Defining a Consistent Total Order

Assume we are given a pseudo-Morse function f consistent with a discrete gradient vector field V as input. Let $<_T$ be a linear extension of $<_V$. If f is constructed

from data given as a PL or piecewise constant function as explained in Sect. 2.3, then V can be assumed to be the empty vector field (all cells are critical), meaning that $\phi \prec_V \rho$ if and only if ϕ is a face of τ . In this case, \prec_T can be constructed by choosing an arbitrary total order on the d -cells for each d and letting $\phi \prec_T \rho$ whenever $\dim \phi < \dim \rho$.

We define a total order \prec by combining \prec_f (the order induced by f) and \prec_T lexicographically: we have $\phi \prec \rho$ if and only if either

- (a) $f(\phi) < f(\rho)$, or
- (b) $f(\phi) = f(\rho)$ and $\phi \prec_T \rho$.

Notice that \prec is consistent with (f, V) by construction.

5.2 Computing Persistence Pairs

Recall that the persistence pairs of dimension $(0, 1)$ are determined solely by the 1-skeleton of \mathcal{K} . Therefore, persistence pairs can be computed by applying a variant of Kruskal's algorithm [25] for finding a minimum spanning tree to both the primal and the dual 1-skeleton [2, 11]. Let G be the 1-skeleton of \mathcal{K} and let $M(G)$ be the minimum spanning tree of G (using the total order \prec for determining the edge weights, which implies uniqueness of $M(G)$). Kruskal's algorithm for computing $M(G)$ initializes a graph T with the vertices of G , sweeps over the edges of G in order \prec , adds to T every edge of G that does not create a 1-cycle, and returns the final graph T . Note that the set of edges of $M(G)$ consists of all negative 1-cells together with all 1-cells τ such that $(\sigma, \tau) \in V$ for some σ ; all other 1-cells create a cycle in T . When encountering a negative 1-cell, we compute the persistence of the corresponding dimension $(0, 1)$ pair by storing for each connected component of the intermediate graph T the 0-cell that created it. Clearly we obtain all dimension $(0, 1)$ persistence pairs this way. Simultaneously, we construct the subgraph $M_\delta(G)$ of $M(G)$ by removing the negative 1-cells with persistence $> 2\delta$. In an analogous way, for the dual 1-skeleton G^* we can compute the minimum spanning tree $M(G^*)$ and obtain the subgraph $M_\delta(G^*)$ together with all dimension $(1, 2)$ persistence pairs.

Kruskal's algorithm has a time complexity of $\mathcal{O}(n \log n)$ when using comparison-based sorting. Assuming that the function values are represented by a small (i.e., $\mathcal{O}(\log n)$) word size, Attali et al. [2] point out that persistence pairs on a graph can be computed in linear time $\mathcal{O}(n)$.

5.3 Extracting the Gradient Vector Field

We now explain how to construct a simplified gradient vector field V_δ . To this end, we traverse (using depth-first search) each of the connected components of the primal graph $M_\delta(G)$ (constructed in Sect. 5.2) from the 0-cell that created the component. During this traversal, whenever we encounter an edge (1-cell) ψ that connects a previously visited vertex (0-cell) ρ with an unvisited vertex ϕ , we add (ϕ, ψ) to the gradient vector field V_δ . This construction takes $\mathcal{O}(n)$ time.

We perform an analogous traversal for the dual graph $M_\delta(G^*)$. Again, whenever we encounter an edge ψ^* that connects a visited vertex ρ^* with an unvisited vertex ϕ^*

(i.e., ψ is a 1-cell and ρ, ϕ are 2-cells of the original complex), we add (ψ, ϕ) to the gradient vector field V_δ . Note that the final V_δ results from both the primal and dual traversals and is a vector field on \mathcal{K} .

Theorem 22 *The gradient vector field V_δ is identical to the final vector field V_n of a 2δ -persistence cancelation sequence (V_0, \dots, V_n) .*

Proof First observe that if $(\sigma, \tau) \in V_n$ and $\dim \sigma = 0$, then both σ and τ are cells of $M_\delta(G)$ since all non-critical cells of V_n either are non-critical in V as well or have persistence $\leq 2\delta$ (with respect to f and \prec). Moreover, the 0-cells creating a connected component of $M_\delta(G)$ are the only critical 0-cells of V_n (by definition) and of V_δ (by construction). Since $M_\delta(G)$ is a forest, the pairs $(\sigma, \tau) \in V_n$ with $\dim \sigma = 0$ are uniquely defined by this property. Thus the dimension $(0, 1)$ pairs of V_n and V_δ coincide. By applying the dual argument to $M_\delta(G^*)$, the statement follows. \square

5.4 Constructing the Simplified Function

Finally, we construct a function f_δ (different from the plateau function defined in Sect. 4.1) that is consistent with the simplified gradient vector field V_δ . Consider the Hasse diagram $H := H_{V_\delta}$ of the strict partial order \prec_{V_δ} as described in Sect. 2.2. We visit the vertices K of H in a linear extension of \prec_{V_δ} . The problem of finding a linear extension of a partial order is also called *topological sorting* and can be solved using depth-first search on H [8]. At each visited cell σ , we define $f_\delta(\sigma)$ as the minimum value that satisfies the lower bound $f_\delta(\sigma) \geq f(\sigma) - \delta$ and renders f_δ consistent with V_δ , i.e.,

$$f_\delta(\sigma) = \max\left(f(\sigma) - \delta, \max_{\rho \prec_{V_\delta} \sigma} f_\delta(\rho)\right).$$

The construction of f_δ also takes $O(n)$ time.

5.5 Correctness of the Algorithm

Theorem 23 *The function f_δ constructed using the above algorithm is a perfect δ -simplification of f .*

Proof By construction f_δ is consistent with V_δ . At the same time, by Theorem 22, V_δ is the final vector field of a 2δ -persistence cancelation sequence. Therefore, by the definition of a perfect δ -simplification, it only remains to show that the constraint $\|f_\delta - f\|_\infty \leq \delta$ is satisfied. The lower bound $f_\delta \geq f - \delta$ is satisfied by construction. It thus remains to show the upper bound $f_\delta \leq f + \delta$.

Observe that the set of all perfect δ -simplifications consistent with V_δ is defined by a set of linear inequalities: the upper and lower bounds on the function values given by $f \pm \delta$, and the inequalities imposed by consistency with V_δ . Therefore, the set of δ -simplifications is a convex polyhedron $P \subset \mathbb{R}^n$ with $n = |K|$. The polyhedron P is bounded since it is a subset of the product of intervals $\prod_{\sigma \in K} [f(\sigma) - \delta, f(\sigma) + \delta]$.

From Theorem 16, we know that P is not empty. We now show that f_δ is contained in P .

First, consider the (unbounded) convex polyhedron \tilde{P} defined by the lower bound $f_\delta \geq f - \delta$ and the inequalities induced by V_δ . By construction, f_δ is contained in \tilde{P} . Moreover, again by construction, f_δ minimizes the function value of any cell among all functions in \tilde{P} . In other words, for any function \tilde{f} in $P \subset \tilde{P}$, we have $\tilde{f} \geq f_\delta$. This implies the upper bound $f_\delta \leq \tilde{f} \leq f + \delta$. \square

6 Discussion

6.1 Computational Results

We implemented the algorithm of Sect. 5 in C++. For a complex with over 4 million cells (the cubical complex for a 1025×1025 pixel image), we obtained a running time of about 15 seconds for computing a perfect δ -simplification on a 2.4 GHz Intel Core 2 Duo laptop.

6.2 Symmetrizing the Algorithm

The method described in Sect. 5 assigns to each cell the smallest possible value. As a consequence, the output function differs from the input function f even if the input function is already a perfect δ -simplification. Moreover, the method is not symmetric in the sense that we obtain an output function which *maximizes* the values if we apply the algorithm to the function $-f$ on the dual complex and return the negative of the simplified function. Since both the minimal and maximal solutions are points of a convex polyhedron as explained in Sect. 5.5, we can take the component-wise arithmetic mean to obtain another perfect δ -simplification.

With this modification, if the input function f is already a perfect δ -simplification, then the minimal solution is given by $f - \delta$, while the maximal solution equals $f + \delta$, so the arithmetic mean of both solutions returns f again as desired.

6.3 Flooding and Carving Artifacts

The methods presented in the present article can be seen as combinations of the so-called *carving* and *flooding* approaches. Consequently, they also inherit some characteristics of these methods that may not always be desirable in practical applications (see Fig. 6).

Carving methods [2, 13, 33] cancel a pair of critical cells by changing only the descending or ascending set of the 1-cell (saddle). This results in a noticeable thin path being carved in the function. On the other hand, modifying only the descending or ascending set of extrema, i.e., lowering maxima and raising minima, produces regions with constant function value; this is called *filling* or *flooding* [9, 20]. Although this effect is less disturbing, it might appear unnatural in certain applications. In the next section, we propose a way to remedy both kinds of artifacts.

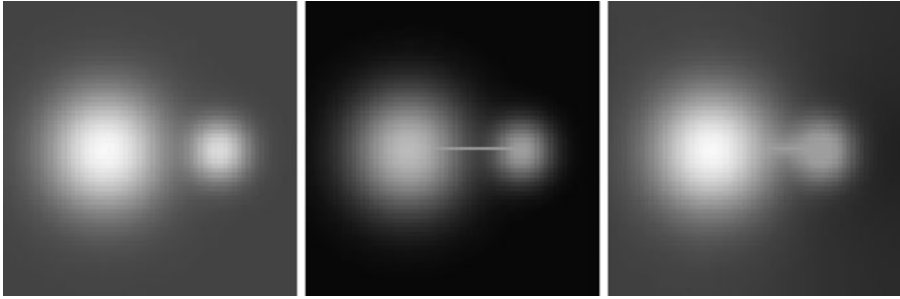


Fig. 6 Visualization of simplification artifacts. Function values are indicated by gray levels. *Left:* Original function. *Middle:* Function obtained by the algorithm of Sect. 5. Note the bright path joining the two spots. *Right:* Function obtained after constraint energy minimization according to Sect. 6.4. While the simplified topological structure is maintained, the visual appearance is closer to the original function

6.4 Combining Topological Simplification and Energy Methods

As mentioned in Sect. 5.5, the set of perfect δ -simplifications consistent with the simplified gradient vector field V_δ is a convex polyhedron P . Hence, the presented method can be combined with energy minimization methods, since the polyhedron P can be used as the feasible region for an arbitrary convex optimization problem. For example, we used the interior point solver Ipopt [35] to minimize (a discretization of) the Dirichlet energy of the difference $f_\delta - f$ in order to obtain a function f_δ that looks as similar as possible to the input function f (see Figs. 6 and 7). Alternatively, we minimized the Dirichlet energy of the simplified function itself in order to obtain smooth contour lines. Note, however, that solving a constraint optimization problem is much more expensive than finding an initial perfect δ -simplification. Computing the function shown in Fig. 7 took about an hour, while the initial simplification took only five seconds to compute.

6.5 A Counterexample for General 2-Complexes

The example of Fig. 8 shows that a perfect δ -simplification may not exist on a non-manifold 2-dimensional cell complex. For the sake of simplicity, the example is given for a non-regular CW complex; it is straightforward to rephrase this example using a regular CW complex by subdividing the cells. The complex consists of two 0-cells ζ and γ with $f(\zeta) = f(\gamma) = 0$, three 1-cells a , b , and c with $f(a) = 1$, $f(b) = 2$, and $f(c) = 0$, and two 2-cells A and B with $f(A) = 2$ and $f(B) = 3$. Note that the complex is not manifold since it is not locally Euclidean at the 1-cell b . A canonical ordering leads to the persistence pairs (a, A) , (b, B) , and (γ, c) . To obtain a perfect δ -simplification for $\delta = 0.5$, one would need to set $f_\delta(b) = f_\delta(B) = 2.5$ and $f_\delta(a) = f_\delta(A) = 1.5$. The corresponding simplified gradient vector field would be $V_\delta = \{(a, A), (b, B)\}$. But since b is a facet of A , we must have $f_\delta(b) \leq f_\delta(A)$. Hence, we cannot cancel both (a, A) and (b, B) at the same time. This constellation also appears in [11] under the name *conflict of type (1, 2)*.

Since such a 2-complex can also appear as a level subcomplex of an n -manifold CW complex for $n \geq 3$ (e.g., a triangulated 3-ball), the example also shows that a

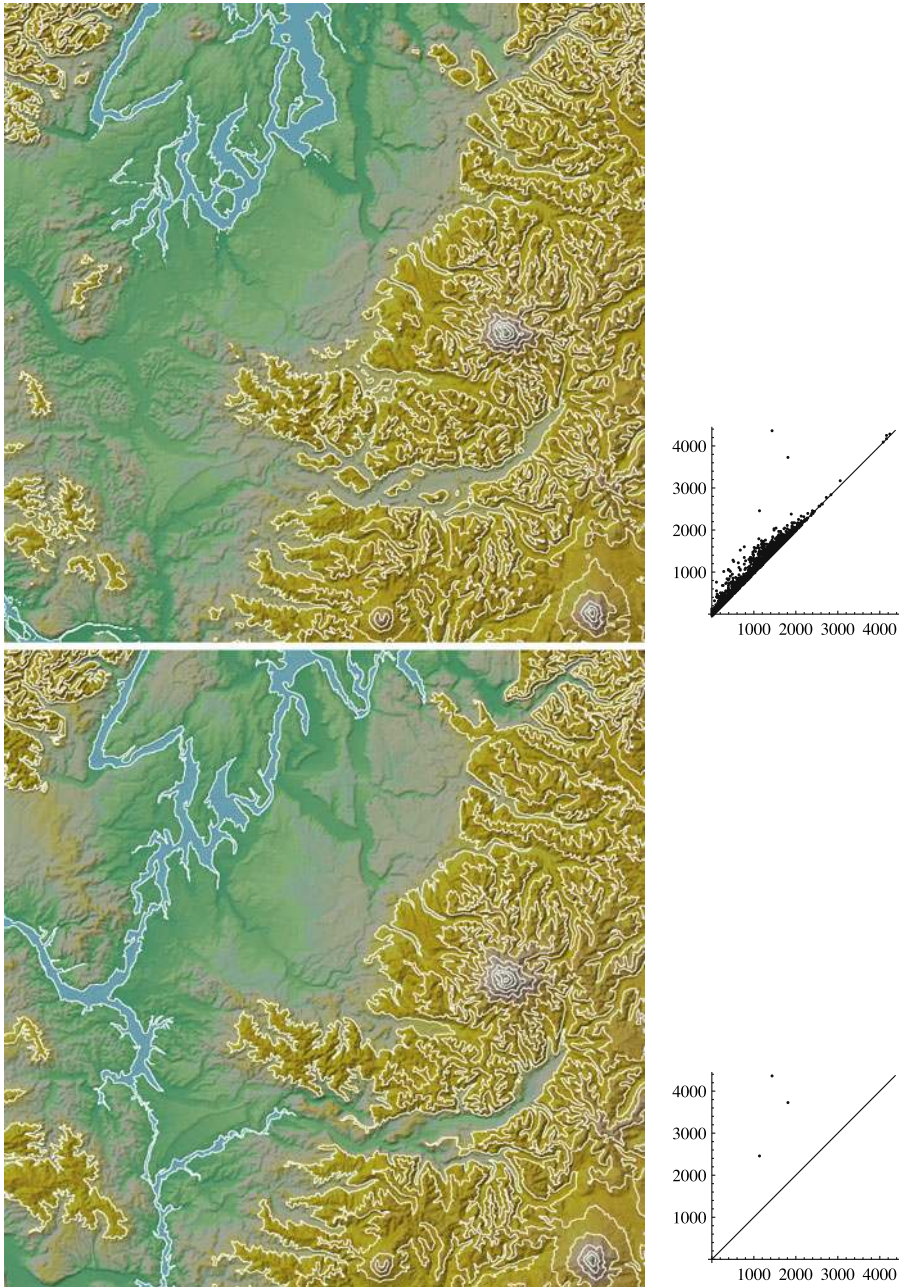
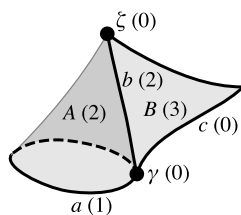


Fig. 7 *Top:* Topographic map of elevation data set “Puget Sound” [27], showing the region around Tacoma. *Contour lines* shown every 500 meters. Elevation data are converted from a 512×512 grid into a pseudo-Morse function on 1050625 cells. 33120 critical cells have persistence > 0 (persistence diagram shown on the *right*). *Bottom:* Simplified elevation function obtained after constraint energy minimization according to Sect. 6.4 with $\delta = 500$ meters. The function has one minimum, three saddles, and three maxima

Fig. 8 A discrete Morse function on a 2-complex that does not have a perfect δ -simplification. The function values of the cells are indicated in parentheses



perfect δ -simplification does not always exist for functions on manifolds. Thus the topological simplification problem is more difficult in higher dimensions.

6.6 Removing Local Extrema from Functions on Manifolds

As a concluding remark, we want to mention that the constructions and proofs presented in this article can also be adapted to the problem of minimizing the number of local extrema of a pseudo-Morse function within a δ -tolerance on any d -dimensional manifold CW complex.

Problem (Extrema simplification on manifolds) Given a pseudo-Morse function f on a regular manifold CW complex and a real number $\delta \geq 0$, find a function f_δ subject to $\|f_\delta - f\|_\infty \leq \delta$ such that f_δ has a minimum number of local extrema.

By a local extremum we mean a critical cell of dimension 0 or d of a gradient vector field consistent with f . We obtain the following theorem, which can be proven by a straightforward modification of the proof of Theorem 16.

Theorem 24 *Given a pseudo-Morse function f on a finite regular closed manifold CW complex and a real number $\delta \geq 0$, there exists a pseudo-Morse function f_δ such that $\|f_\delta - f\|_\infty \leq \delta$ and the number of local extrema of f_δ equals the number of those local extrema of f that have persistence $> 2\delta$. This number is minimal.*

Note that in the case $d = 2$ this problem is equivalent to the topological simplification problem by the following argument. Let c_i denote the number of critical cells of dimension i . Since the Euler characteristic $\chi = c_0 - c_1 + c_2$ is a topological invariant and we have $c_0 + c_1 + c_2 = 2(c_0 + c_2) - \chi$, the number of critical points is minimal if and only if the number of extrema is minimal.

Acknowledgements Thanks to Yassin Sabih, Henrik Schumacher, and the anonymous referees for providing valuable feedback. The “Puget Sound” data set used in Fig. 7 is taken from the Large Geometric Models Archive of the Georgia Institute of Technology. The original elevation data are obtained from The United States Geological Survey (USGS), made available by The University of Washington.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Agarwal, P.K., Arge, L., Yi, K.: I/O-efficient batched union-find and its applications to terrain analysis. In: SCG '06: Proceedings of the Twenty-Second Annual Symposium on Computational Geometry, pp. 167–176. ACM, New York (2006)
2. Attali, D., Glisse, M., Hornus, S., Lazarus, F., Morozov, D.: Persistence-sensitive simplification of functions on surfaces in linear time. Preprint (2008)
3. Banchoff, T.: Critical points and curvature for embedded polyhedra. *J. Differ. Geom.* **1**(3–4), 245–256 (1967)
4. Bauer, U.: Persistence in discrete Morse theory. PhD thesis, University of Göttingen (2011)
5. Chari, M.K.: On discrete Morse functions and combinatorial decompositions. *Discrete Math.* **217**(1–3), 101–113 (2000)
6. Cohen-Steiner, D., Edelsbrunner, H., Harer, J.: Stability of persistence diagrams. *Discrete Comput. Geom.* **37**(1), 103–120 (2007)
7. Cohen-Steiner, D., Edelsbrunner, H., Harer, J.: Extending persistence using Poincaré and Lefschetz duality. *Found. Comput. Math.* **9**(1), 79–103 (2008)
8. Cormen, T.H., Leiserson, C.E., Rivest, R.L., Stein, C.: Introduction to Algorithms, 3rd edn. MIT Press, Cambridge (2009)
9. Danner, A., Møhlhave, T., Yi, K., Agarwal, P.K., Arge, L., Mitasova, H.: Terra-Stream: from elevation data to watershed hierarchies. In: GIS '07: Proceedings of the 15th Annual ACM International Symposium on Advances in Geographic Information Systems, pp. 1–8. ACM, New York (2007)
10. de Silva, V., Morozov, D., Vejdemo-Johansson, M.: Dualities in persistent (co)homology. Unpublished manuscript (2010)
11. Edelsbrunner, H., Letscher, D., Zomorodian, A.: Topological persistence and simplification. *Discrete Comput. Geom.* **28**(4), 511–533 (2002)
12. Edelsbrunner, H., Harer, J., Zomorodian, A.: Hierarchical Morse–Smale complexes for piecewise linear 2-manifolds. *Discrete Comput. Geom.* **30**(1), 87–107 (2003)
13. Edelsbrunner, H., Morozov, D., Pascucci, V.: Persistence-sensitive simplification functions on 2-manifolds. In: SCG '06: Proceedings of the Twenty-Second Annual Symposium on Computational Geometry, pp. 127–134. ACM, New York (2006)
14. Eells, J., Kuiper, N.: Manifolds which are like projective planes. *Publ. Math. Inst. Hautes Études Sci.* **14**(1), 5–46 (1962)
15. Forman, R.: Morse theory for cell complexes. *Adv. Math.* **134**(1), 90–145 (1998)
16. Forman, R.: A user's guide to discrete Morse theory. *Sémin. Lothar. Comb.* **B48c**, 1–35 (2002)
17. Gray, C., Kammer, F., Löffler, M., Silveira, R.I.: Removing local extrema from imprecise terrains. Preprint (2010). [arXiv:1002.2580](https://arxiv.org/abs/1002.2580)
18. Gyulassy, A., Natarajan, V., Pascucci, V., Hamann, B.: Efficient computation of Morse–Smale complexes for three-dimensional scalar functions. *IEEE Trans. Vis. Comput. Graph.* **13**(6), 1440–1447 (2007)
19. Hatcher, A.: Algebraic Topology. Cambridge University Press, Cambridge (2002)
20. Jenson, S.K., Domingue, J.O.: Extracting topographic structure from digital elevation data for geographic information system analysis. *Photogramm. Eng. Remote Sens.* **54**(11), 1593–1600 (1988)
21. Joswig, M., Pfetsch, M.E.: Computing optimal Morse matchings. *SIAM J. Discrete Math.* **20**(1), 11–25 (2006)
22. Kaczynski, T., Mischaikow, K., Mrozek, M.: Computational Homology, 1st edn. Applied Mathematical Sciences, vol. 157. Springer, Berlin (2004)
23. King, H., Knudson, K., Mramor, N.: Generating discrete Morse functions from point data. *Exp. Math.* **14**(4), 435–444 (2005)
24. Kosinski, A.: Singularities of piecewise linear mappings. I Mappings into the real line. *Bull. Am. Math. Soc.* **68**(2), 110–114 (1962)
25. Kruskal, J.B.: On the shortest spanning subtree of a graph and the traveling salesman problem. *Proc. Am. Math. Soc.* **7**(1), 48–50 (1956)
26. Kühnel, W.: Triangulations of manifolds with few vertices. In: Triccerri, F. (ed.) *Advances in Differential Geometry and Topology*, pp. 59–114. World Scientific, Singapore (1990)
27. Large Geometric Models Archive. Georgia Institute of Technology. Available from: http://www.cc.gatech.edu/projects/large_models/
28. Lewiner, T., Lopes, H., Tavares, G.: Optimal discrete Morse functions for 2-manifolds. *Comput. Geom.* **26**(3), 221–233 (2003)

29. Lundell, A.T., Weingram, S.: *The Topology of CW Complexes*. Van Nostrand-Reinhold, New York (1969)
30. Milnor, J.: *Morse Theory*. *Annals of Mathematics Studies*, vol. 51. Princeton University Press, Princeton (1963)
31. Morozov, D.: *Homological illusions of persistence and stability*. PhD thesis, Duke University (2008)
32. Morse, M.: The elimination of critical points of a non-degenerate function on a differentiable manifold. *J. Anal. Math.* **13**(1), 257–316 (1964)
33. Soille, P.: Morphological carving. *Pattern Recognit. Lett.* **25**(5), 543–550 (2004)
34. Soille, P.: Optimal removal of spurious pits in grid digital elevation models. *Water Resour. Res.* **40**(12), W12509+ (2004)
35. Wächter, A., Biegler, L.T.: On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Math. Program.* **106**(1), 25–57 (2006)
36. Zomorodian, A., Carlsson, G.: Computing persistent homology. *Discrete Comput. Geom.* **33**(2), 249–274 (2005)