On the definition of the homological critical value

Dejan Govc

Received: 28 May 2014 / Accepted: 26 January 2015 / Published online: 10 February 2015
© Tbilisi Centre for Mathematical Sciences 2015

Abstract We point out that there is a problem with the definition of the homological critical value as defined by Cohen-Steiner, Edelsbrunner and Harer in their seminal paper on stability of persistence diagrams. Under that definition, their critical value lemma in fact fails. We provide two counterexamples and a definition (due to Bubenik and Scott) we feel should be preferred. We also prove a version of the critical value lemma that remains valid under the original definition.

Keywords Persistent homology · Homological critical value · Definition · Counterexample

Mathematics Subject Classification 55N99 · 54G20 · 54E45 · 68W30

1 Introduction

The homological critical value is an important concept used in the theory of persistent homology. As noticed already by Bubenik and Scott in [1], the widely cited definition of this concept given by Cohen-Steiner et al. in [2] is problematic. To provide some justification for their own version of the definition, Bubenik and Scott [1, footnote of Definition 4.3] present a counterexample to the critical value lemma of [2] under the
original definition of [2]. Their counterexample, however, is a discontinuous function. Using elementary topological methods, our aim is to show that continuous (or even smooth) counterexamples can be constructed as well, which leads us to conclude that the definition in [2] is inadequate even in nice settings. An extended exposition of our results is available in [3].

1.1 Definitions and conventions

We need to carefully distinguish between two definitions of the homological critical value. This requires us to establish some new terminology. We also describe some of the context along the way.

In [2], the authors state the following definition of the homological critical value:

**Definition** [2] Let $X$ be a topological space and $f$ a real function on $X$. A *homological critical value* of $f$ is a real number $a$ for which there exists an integer $k$ such that for all sufficiently small $\epsilon > 0$ the map $H_k(f^{-1}(-\infty, a - \epsilon]) \to H_k(f^{-1}(-\infty, a + \epsilon])$ induced by inclusion is not an isomorphism.

Here $H_k$ denotes the $k$-th singular homology (possibly with coefficients in a field, as is usually assumed in the context of persistent homology). The main purpose of this definition is to establish the critical value lemma, which is stated in [2] as follows:

**Lemma 1.1** [2] Suppose the function $f : X \to \mathbb{R}$ has no homological critical values on the closed interval $[x, y]$. Then the inclusion $f^{-1}(-\infty, x] \hookrightarrow f^{-1}(-\infty, y]$ induces isomorphisms on all homology groups.

As mentioned, under the above definition, the critical value lemma is in fact false. This problem is solved easily by replacing the above definition with the following one, suggested by Bubenik and Scott [1, footnote of Definition 4.3], which we state in a slightly less general form as follows:

**Definition** Let $X$ be a topological space and $f$ a real-valued function on $X$. A real number $a$ is a *homological regular value* of the function $f$ if there exists an $\epsilon > 0$ such that for each pair of real numbers $x < y$ on the interval $(a - \epsilon, a + \epsilon)$, the inclusion $f^{-1}(-\infty, x] \hookrightarrow f^{-1}(-\infty, y]$ induces isomorphisms on all homological groups. A real number $a$ that is not a homological regular value of $f$ is called a *homological critical value* of $f$.

For this definition, the critical value lemma does indeed hold, as Bubenik and Scott show (in a more general form) in [1, Lemma 4.4]. This leaves us with two non-equivalent definitions of the homological critical value. To distinguish between the two, we adopt the following:

**Convention** Homological critical values from the first (problematic) definition [2] will be called *symmetric homological critical values* of $f$. Homological critical values from the second definition will simply be called *homological critical values*, i.e. we accept the second definition as the preferred one. A real number that is not a symmetric homological critical value of $f$ will be called a *symmetric homological regular value*. 
2 Counterexamples

We will now show that under the symmetric definition of the homological critical value, the critical value lemma fails.

2.1 An easy counterexample

The following counterexample is inspired by the discontinuous counterexample given by Bubenik and Scott in [1]. Define a topological space $X \subseteq \mathbb{R}^2$ (equipped with the subspace topology) as follows:

$$X = \{0\} \times [-1, 1] \cup (0, 1) \cup \{1\} \times [0, 1].$$

(Throughout this paper, we adopt the convention that $\times$ binds more strongly than $\cup$.) Define a function $f : X \to \mathbb{R}$ on this space by the formula $f(x, y) = y$, i.e. the height function. The various sublevel sets $f^{-1}(-\infty, a]$ look as follows:

Note that $X$ looks exactly like the sublevel set pictured for $a > 0$, but with a square instead of a rectangle. The sublevel sets for $a < -1$ are empty.

It can be seen that $f^{-1}(-\infty, a]$ is contractible for $a \in (-1, 0) \cup (0, 1)$ and has two path components for $a = 0$. Therefore the numbers $a \in (-1, 0) \cup (0, 1)$ are not homological critical values. The number $a = 0$ is a homological critical value, but is not a symmetric homological critical value. This means that there are no symmetric homological critical values in $[-\frac{1}{2}, 0]$ and yet, $f^{-1}(-\infty, -\frac{1}{2}] \leftrightarrow f^{-1}(-\infty, 0]$ does not induce an isomorphism on $H_0$. As [2] uses the symmetric definition to define the homological critical value, this contradicts the critical value lemma as stated there. Note that it is just as easy to make the critical value lemma fail on $H_1$ (See [3]).

2.2 A compact smooth counterexample

Now, we will construct a compact smooth manifold embedded in $\mathbb{R}^4$ which, together with its height function, is a counterexample to the critical value lemma as stated in
First, define $\vartheta : \mathbb{R} \to \mathbb{R}$ by $\vartheta(x) = e^{-\frac{1}{x^2}}$ for $x > 0$ and $\vartheta(x) = 0$ otherwise. This well-known smooth function is very useful when defining various correction terms. Next, we examine the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x, y) = \begin{cases} e^{-\frac{1}{x^2}}(y - \sin \pi x)^2; & x \neq 0, \\ 0; & x = 0. \end{cases}$$

This is a smooth function with particularly interesting sublevel sets: $f^{-1}(-\infty, 0]$ is a variant of the topologist’s sine curve. We will show that it has three path components and a single connected component. On the other hand, the sublevel set $f^{-1}(-\infty, a]$ is contractible for every $a > 0$.

In fact, these sublevel sets may be computed explicitly. Suppose $a \geq 0$. Solving $f(x, y) \leq a$ for $y$, we immediately obtain:

$$f^{-1}(-\infty, a] = \left\{ (x, y) \in \mathbb{R}^2 \mid \sin \frac{\pi}{x} - \sqrt{ae^{\frac{1}{x^2}}} \leq y \leq \sin \frac{\pi}{x} + \sqrt{ae^{\frac{1}{x^2}}} \text{ or } x = 0 \right\}.$$ 

In particular,

$$f^{-1}(-\infty, 0] = \left\{ (x, y) \in \mathbb{R}^2 \mid y = \sin \frac{\pi}{x} \text{ or } x = 0 \right\}.$$ 

For example, the sublevel sets $f^{-1}(-\infty, a]$ for $a = 0$ and $a = 0.01$ look as follows:

To make this into another counterexample, we need to somehow introduce contractible sublevel sets for $a < 0$ into the situation. We would also like to make the counterexample compact, so it would help if we could make the sublevel sets compact. To achieve this, we introduce some smooth correction terms and define $h : \mathbb{R}^2 \to \mathbb{R}$ as follows:

\[ \text{Springer} \]
\[ h(x, y) = f(x, y) - \vartheta \left( \frac{1}{4} - (x - 1)^2 \right) + x^2 \vartheta (x^2 - 4) + y^2 \vartheta (y^2 - 4). \]

The minimum value of this function is \( h(1, 0) = -e^{-16} \). We will prove that the first correction ensures that the sublevel sets for \(-e^{-16} < a < 0\) are contractible, as required, and the other two ensure that all sublevel sets are compact. (Each of them ensures that the sublevel set is bounded in the direction of the corresponding coordinate axis.) The sublevel sets for \( a > 0 \) remain contractible, and the sublevel set for \( a = 0 \) remains a (slightly different) variant of the topologist’s sine curve.\(^1\) For example, the sublevel sets \( h^{-1}(-\infty, a] \) for \( a = 0 \) and \( a = 0.01 \) look as follows:

Let us examine these claims more closely. Our main tool is the following lemma:

**Lemma 2.1** Suppose \( X \) is contractible and \( A \subseteq X \times \mathbb{R} \) has the property that \( A \cap (\{x\} \times \mathbb{R}) \) is contractible (i.e. an interval) for all \( x \in X \). Suppose there is a continuous function \( s : X \to \mathbb{R} \), whose graph \( \Gamma_s \) is contained in \( A \). Then \( A \) is contractible.

**Proof** We may define a deformation retraction \( R : A \times I \to A \) of \( A \) to \( \Gamma_s \) by \( R((x, y), t) = (x, (1-t)y + ts(x)) \). But \( \Gamma_s \) is homeomorphic to \( X \), which is contractible. \( \square \)

**Proposition 2.2** The sublevel set \( h^{-1}(-\infty, a] \) is contractible for \( a \in [-e^{-16}, \infty) \setminus \{0\} \). The sublevel set \( h^{-1}(-\infty, 0] \) is connected, but not path connected. It consists of three contractible path components.

**Proof** Define \( q(x) = \sin \frac{\pi}{x} \) for \( x \neq 0 \) and \( q(0) = 0 \). For \( x \in \mathbb{R} \), define \( h_x : \mathbb{R} \to \mathbb{R} \) by \( h_x(y) = h(x, y) \). Observe that, for a fixed \( x \), \( f(x, y) \) is decreasing in \( y \) on \((-\infty, q(x)]\) and increasing on \([q(x), \infty)\). Similarly, \( y^2 \vartheta (y^2 - 4) \) is decreasing on \((-\infty, -2]\), equal to 0 on \([-2, 2]\) and increasing on \([2, \infty)\). So, \( h_x \) is decreasing on \((-\infty, q(x)]\) and increasing on \([q(x), \infty)\). Therefore, for each \( a \in \mathbb{R} \), the set \((\{x\} \times \mathbb{R}) \cap h^{-1}(-\infty, a] \) is either empty or contractible.

\(^1\) Despite the appearance, the set \( h^{-1}(-\infty, 0] \) is not a curve above \([\frac{1}{2}, \frac{3}{2}]\), since it is slightly fattened.
Now, let $W = \{ (x, y) \in \mathbb{R}^2 \mid y = q(x) \}$ be the graph of the function $q$ and let $p_1 : \mathbb{R}^2 \to \mathbb{R}$ be the projection $p_1(x, y) = x$. We claim that $J(a) := p_1(h^{-1}(\infty, a])$ is contractible for each $a \geq -e^{-16}$. If $(x, y) \in h^{-1}(\infty, a)$, then $(x, q(x)) \in h^{-1}(\infty, a]$, since $h_x$ achieves its minimum at $q(x)$. Therefore, if $h|_W$ is the restriction of $h$ to $W$, we have:

$$J(a) = p_1 \left( (h|_W)^{-1}(\infty, a) \right).$$

For $(x, y) \in W$, the function $h$ simplifies to

$$h(x, y) = -\vartheta \left( \frac{1}{4} - (x - 1)^2 \right) + x^2 \vartheta(x^2 - 4).$$

Since this expression is decreasing on $(-\infty, 1]$ and increasing on $[1, \infty)$, the contractibility of $J(a)$ is established.

To prove the contractibility of $h^{-1}(\infty, a]$ for $a \in [-e^{-16}, \infty) \setminus \{0\}$ by Lemma 2.1, it remains to find an appropriate continuous function $s : J(a) \to \mathbb{R}$. For $a \in [-e^{-16}, 0)$, such a function is provided by $q|_{J(a)}$; the interval $J(a)$ does not contain 0 in this case because $h(0, 0) = 0 > a$.

For $a > 0$, the discontinuity of $q$ at zero prevents us from using $q|_{J(a)}$. Luckily, a small modification turns out to work: let $L = \{ 0 \} \times [-2, 2]$. Since $h|_L \equiv 0$, there is an open neighborhood $U$ of $L$ such that $h|_U \leq a$. By the tube lemma [6, chapter 3, Lemma 5.8], we may assume that $U$ is of the form $(-c, c) \times (-2 - c, 2 + c)$ for some $c > 0$. Therefore we may use the function $s : J(a) \to \mathbb{R}$ defined by

$$s(x) = \begin{cases} q(x); & x \notin (-c, c), \\ \frac{c}{2} q(c); & x \in [-c, c]. \end{cases}$$

Note that the graph of $s$ is contained in $h^{-1}(\infty, a]$ since $h(x, \frac{c}{2} q(c)) \leq a$ holds for $x \in [-c, c]$.

Finally, for the set $h^{-1}(\infty, 0]$, no such function $s$ exists. Indeed, $h^{-1}(\infty, 0]$ deformation retracts to $([-2, 0] \times [-2, 2]) \cap f^{-1}(\infty, 0]$. This is because $([1/2, 3/2] \times \mathbb{R}) \cap h^{-1}(\infty, 0]$ deformation retracts to $([1/2, 3/2] \times \mathbb{R}) \cap f^{-1}(\infty, 0]$ by the construction from the proof of Lemma 2.1, where the required continuous function is provided by $q\left|_{\left[\frac{1}{2}, \frac{3}{2}\right]}\right.$ This deformation retraction can then be extended simply by keeping all other points stationary. This completes the proof.  

Making the counterexample compact is easy now. Composing stereographic projection $p : S^2 \setminus \{N\} \to \mathbb{R}^2$ (where $N = (0, 0, 1)$ is the north pole) with the function $h$, we obtain a function $h \circ p : S^2 \setminus \{N\} \to \mathbb{R}$. We can modify it into a smooth function $H : S^2 \to \mathbb{R}$ such that the sublevel sets $h^{-1}(\infty, a]$ and $H^{-1}(\infty, a]$ are homeomorphic for all $a \leq 1$.

To do this, observe that the set $K = (h \circ p)^{-1}(\infty, 1]$ is a compact subset of $S^2 \setminus \{N\}$. Let $U$ be an open neighborhood of $N$ in $S^2$ such that the closure $\overline{U}$ is contained in $S^2 \setminus K$. Then, there exists a smooth function $\eta : S^2 \to [0, 1]$ whose
restrictions satisfy $\eta|_U \equiv 0$ and $\eta|_K \equiv 1$. (See e.g. [5, Chapter 2].) This allows us to define

$$H(w) = \begin{cases} (h \circ p)(w)\eta(w) + (1 - \eta(w)); & w \neq N, \\ 1; & w = N, \end{cases}$$

which has the desired properties.

Using the standard tricks from smooth manifold theory, we can now make $H$ into a height function: the formula $i(x, y, z) = (x, y, z, H(x, y, z))$ defines a smooth embedding $i: S^2 \to \mathbb{R}^4$. The desired height function $F: i(S^2) \to \mathbb{R}$ is then defined by $F(x, y, z, w) = w$. We have therefore found a compact smooth manifold embedded in $\mathbb{R}^4$ and a smooth height function on it that contradicts the critical value lemma.

3 Positive results

In this section, we note that there are, nevertheless, positive statements which can be made under the symmetric definition of the homological critical value. First we prove the following lemma, where $i: X \to Y$ denotes the inclusion map of spaces $X \subseteq Y$.

**Lemma 3.1** Suppose $A \subseteq C \subseteq B \subseteq D$ are topological spaces such that the inclusions $i_B^A$ and $i_D^C$ induce isomorphisms on $k$-th homology. Then $i_B^C \circ i_C^A$ and $i_D^B \circ i_B^D$ also induce isomorphisms on $k$-th homology.

**Proof** The maps factor as $i_B^C = i_B^D \circ i_B^C$ and $i_B^C = i_B^C \circ i_B^D$. Since these maps induce isomorphisms, we may conclude that $i_B^D$ and $i_B^C$ induce monomorphisms and that $i_B^D$ and $i_B^C$ induce epimorphisms. In particular, $i_B^D$ induces an isomorphism. Since $i_B^D$ and $i_B^C$ both induce isomorphisms, so does $i_B^D$ by the first factorisation. By the second factorisation, $i_B^C$ induces an isomorphism. Finally, composing $i_C^A$, $i_B^C$ and $i_B^D$ yields the result for $i_B^A$.

This allows us to prove the following symmetric version of the critical value lemma:

**Theorem 3.2** Suppose $f: X \to \mathbb{R}$ is continuous and has no symmetric homological critical values on the interval $[x, y)$. Then the inclusion $f^{-1}(-\infty, x) \hookrightarrow f^{-1}(-\infty, y)$ induces isomorphisms on all homology groups.

So, for continuous functions, the lemma holds if we examine open sublevel sets rather than closed ones. To see that continuity is necessary here, consider the function $f: D^2 \to \mathbb{R}$, defined by $f(z) = t$ if $z = e^{2\pi it}$ for some $t \in [0, 1)$ and $f(z) = 1$ otherwise. Despite the fact that this function has no symmetric homological critical values in the interval $[\frac{1}{2}, 1]$, the inclusion $f^{-1}(-\infty, \frac{1}{2}) \hookrightarrow f^{-1}(-\infty, 1)$ fails to induce an isomorphism on $H_1$, as it maps a contractible set into $S^1$. The crucial property of continuous functions is that their closed sublevel sets are closed and open sublevel sets are open. Open sublevel sets can therefore be used to form open covers, allowing us to exploit compactness.
Proof of the theorem We may work with homology in a fixed dimension, say with $H_k$. Associate the inclusion $i^b_a: f^{-1}(\infty, a] \hookrightarrow f^{-1}(\infty, b]$ to each open interval $(a, b)$. Note that if two such intervals $(a, b)$ and $(c, d)$ overlap, and the associated inclusions $i^b_a$ and $i^d_c$ induce isomorphisms on $H_k$, then the inclusion associated to the interval $(a, b) \cup (c, d)$ also induces an isomorphism by Lemma 3.1.

Now, let $\delta \in (0, y - x)$. Since there are no symmetric homological critical values on $[x, y - \delta]$, we may choose for each $z \in [x, y - \delta]$ an $\epsilon(z) \in (0, \frac{\delta}{2})$ such that $i^z_{z-\epsilon(z)}$ induces isomorphisms on $k$-th homology. By compactness, finitely many intervals of the form $(z - \epsilon(z), z + \epsilon(z))$ already cover $[x, y - \delta]$. By the first paragraph, the inclusion associated to the union of these intervals induces an isomorphism. Therefore, there exist $\epsilon, \eta \in (0, \frac{\delta}{2})$ such that the inclusion $i^{y-\eta}_{x-\epsilon}$ induces an isomorphism.

Since this construction works for arbitrary $\delta \in (0, y - x)$, we may choose strictly decreasing sequences $\epsilon_n \to 0$ and $\eta_n \to 0$ such that the maps $i^{y-\eta_n}_{x-\epsilon_n}$ induce isomorphisms. Since $x - \epsilon_n < x - \epsilon_{n+1} < y - \eta_n < y - \eta_{n+1}$, the maps $i^{y-\eta_{n+1}}_{y-\eta_n}$ and the map $i^{y-\eta_n}_{x-\epsilon}$ induce isomorphisms by Lemma 3.1. These maps form a direct system of isomorphism-inducing inclusions and $\bigcup_{n=1}^{\infty} f^{-1}(\infty, y - \eta_n] = f^{-1}(\infty, y)$ holds. Since $f$ is continuous, the sets $f^{-1}(\infty, y - \eta_n)$ form an open cover of $f^{-1}(\infty, y)$, so each compact subset of $f^{-1}(\infty, y)$ is contained in some $f^{-1}(\infty, y - \eta_n)$. This allows us to use [4, Proposition 3.33] and conclude that $f^{-1}(\infty, x - \epsilon_1) \to f^{-1}(\infty, y)$ induces an isomorphism.

By the same argument, $f^{-1}(\infty, x - \epsilon_n) \to f^{-1}(\infty, y)$ induces an isomorphism for each $n \in \mathbb{N}$, so the maps $i^{y-\eta_{n+1}}_{x-\epsilon_n}$ also form a direct system of isomorphism-inducing inclusions. Reasoning as in the previous paragraph, [4, Proposition 3.33] allows us to conclude that the inclusion $f^{-1}(\infty, x - \epsilon_1) \to f^{-1}(\infty, x)$ induces an isomorphism. Since $f^{-1}(\infty, x - \epsilon_1) \to f^{-1}(\infty, y)$ also induces an isomorphism, so does $f^{-1}(\infty, x) \to f^{-1}(\infty, y)$, concluding our proof. □

Remark As Primož Škraba points out, an easy consequence of this result is that if $\text{hfs} X > 0$ (where we redefine the concept of $\text{hfs} X$ from [2] as the infimum of positive homological critical values$^2$ of the distance function $d^X: \mathcal{M} \to \mathbb{R}$ defined by $d^X(y) = d(y, X)$) then the inclusion $(d^X)^{-1}(\infty, \epsilon) \to (d^X)^{-1}(\infty, \text{hfs} X)$ induces isomorphisms$^3$ on all homology groups for all $\epsilon \in (0, \text{hfs} X)$.

To see this, observe that $(d^X)^{-1}(\infty, \epsilon) \to (d^X)^{-1}(\infty, \text{hfs} X)$ induces an isomorphism on $k$-th homology by the above theorem.$^4$ Since there are no homological critical values in $[\frac{\epsilon}{2}, \epsilon]$, the inclusion $(d^X)^{-1}(\infty, \frac{\epsilon}{2}) \to (d^X)^{-1}(\infty, \epsilon]$ induces an isomorphism by the ordinary critical value lemma. (See [1, Lemma 4.4].) Now, use the fact that

$$(d^X)^{-1}(\infty, \frac{\epsilon}{2}) \subseteq (d^X)^{-1}(\infty, \epsilon) \subseteq (d^X)^{-1}(\infty, \epsilon] \subseteq (d^X)^{-1}(\infty, \text{hfs} X),$$

and we are done by Lemma 3.1.

---

$^2$ Rather than positive symmetric homological critical values as in [2].

$^3$ For the concept of $\text{hfs} X$ from [2], this only holds for $(d^X)^{-1}(\infty, \epsilon) \to (d^X)^{-1}(\infty, \text{hfs} X)$.

$^4$ This is because the absence of homological critical values in an interval also implies the absence of symmetric ones.

© Springer
4 Conclusions and future work

We have shown that the definition of the homological critical value as stated in [2] is problematic even in nice settings. We have presented two counterexamples to the critical value lemma under that definition. An essential feature of these was an instantaneous (zero-persistent) change occurring in homology. We have also seen that a positive result remains valid for open sublevel sets of continuous functions. This suggests that the open sublevel sets behave nicely when it comes to homology and that it might be possible to study them using nice behaviour of inclusions along particular sequences, whereas before, we needed a whole interval where the behaviour was nice.

Acknowledgments The author would like to thank the anonymous reviewer for all the valuable comments and Primož Škraba, Jaka Smrekar and João Pita Costa for numerous helpful discussions and suggestions.

References

1. Bubenik, P., Scott, J.A.: Categorification of persistent homology (2012). arXiv:1205.3669v1
2. Cohen-Steiner, D., Edelsbrunner, H., Harer, J.: Stability of persistence diagrams. Discret. Comput. Geom. 37, 103–120 (2007)
3. Govc, D.: On the definition of homological critical value (2013). arXiv:1301.6817v1
4. Hatcher, A.: Algebraic Topology. Cambridge University Press, Cambridge (2002)
5. Lee, J.M.: Introduction to Smooth Manifolds. Springer, New York (2003)
6. Munkres, J.: Topology, a First Course. Prentice Hall, NJ (1975)