ON UNIMODAL ADDITIVE DECOMPOSITIONS OF PROBABILITY DENSITIES IN THE PLANE

DAN BURGHELEA AND BORIS SHAPIRO

ABSTRACT. Inspired by the recent articles [2, 3] of Y. Baryshnikov and R. Ghrist, we discuss below the problem of additive decomposition of a smooth non-negative compactly supported in $\mathbb{R}^2$ function as a sum of smooth non-negative unimodal functions. We give an upper bound for the minimal number of such summands in terms of topology of the function under consideration.

1. INTRODUCTION

Motivation and background: For many special classes of probability distributions, the number of their modes (i.e., local maxima) has been a subject of intensive studies for several decades, see e.g. [1, 9, 15, 16, 18] and references therein. This important characteristic of a probability distribution is far from being its only natural qualitative feature. In a similar line of study, Y. Baryshnikov and R. Ghrist formulated the following fundamental question, see [2, 3] and also [5].

Problem 1. *Given a smooth probability density function f in $\mathbb{R}^k$, find/estimate the minimal number ℓ such that f can be represented as*

$$f = \sum_{i=1}^{\ell} f_i, \tag{1}$$

where each f_i is a smooth, non-negative and unimodal function.

They were motivated by the previous research and statistical applications. A very naive explanation of the origin and importance of Problem 1 is as follows. In applications a statistical variable with a unimodal probability density f_i is considered as a single *phenomenological factor* affecting a more complicated statistical variable with probability distribution f. The goal of Problem 1 is to find the minimal number of such factors whose total impact (i.e. the sum of their densities) gives f. (In what follows, statements labeled by letters are borrowed from the existing literature while statements labeled by numbers are hopefully new.)

Recall that, by definition, a continuous non-negative function $\phi : \mathbb{R}^k \to [0, \infty)$ is called *unimodal* if for every $t \in [0, \infty)$, the set

$$\Omega_t = \{(x_1, \ldots, x_k) \in \mathbb{R}^k | \phi(x_1, \ldots, x_k) \geq t\}$$

is contractible. In [2, 3], the above number ℓ was called the *unimodal category* of f and denoted by $ucat(f)$. Observe that Problem 1 is a topological question since an additive decomposition of a function and the property of unimodality are presented under homeomorphisms of the source space. For $k = 1$, Y. Baryshnikov and R. Ghrist were able to explicitly calculate $ucat(f)$ as follows.

Given a smooth non-negative $f : \mathbb{R} \to [0, \infty)$ with isolated maxima and minima, consider an open interval (a, b), where a and b are two consecutive minima of f. We say that (a, b) is the *interval of forced maximum* for f if $f(a) + f(b) - f_{max} < 0$, where f_{max} is the (unique) local maximum of f in (a, b).

Proposition A (comp. with Proposition 10 of [2])**.** If (a, b) is an interval of forced maximum, then f can not be represented on (a, b) as the sum of an arbitrary number of smooth nonnegative functions where each of the summands is either strictly decreasing or strictly increasing on (a, b).

As a direct corollary of Proposition A one obtains the main result of [2], see Theorem 11 in loc. cit.

Theorem A. The unimodal category $ucat(f)$ equals the maximal number of non-intersecting intervals of forced maximum of f.

While the fact that the latter number is not less than $ucat(f)$ is obvious from Proposition A, their coincidence follows from a nice explicit construction shown in Fig. 1 borrowed from [5] with the kind permission of the authors.

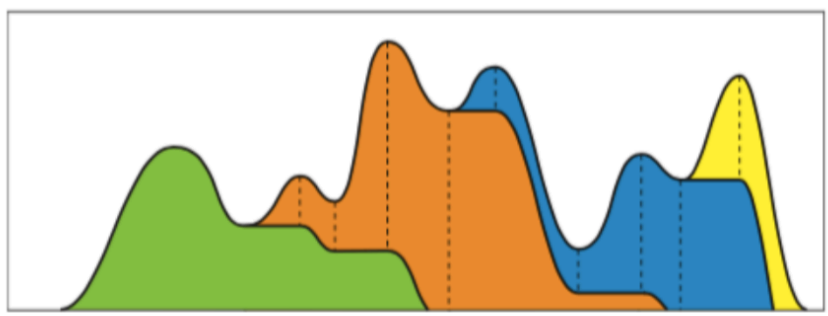


FIGURE 1. Finding $ucat$ for smooth univariate densities.

Observe that Theorem A implies the following.

Corollary A. In the univariate case, $ucat(f)$ is smaller than or equal to the number of local maxima of the function $f : \mathbb{R} \to [0, \infty)$.

So far attempts to find a satisfactory analog of Theorem A in higher dimensions have not been very sucessfull. Observe, in particular, that Corollary A fails when the number of variables is greater than one. In 2014 some partial progress in the calculation of $ucat(f)$ for functions in the plane has been reported in [14]. Namely, the authors were able to calculate $ucat(f)$ for $f : \mathbb{R}^2 \to [0, \infty)$ for which the so-called Morse-Smale graph of f is a tree. EXPAND... In recent paper [11, 12] D. Govc extended Theorem A to the case of functions on a circle and also disproved the so-called monotonicity conjecture of Baryshnikov-Ghrist in the case of general functions in the plane. The most recent results related to the unimodal category can be found in [13, 4] which apparently exhaust all the currently available information on this problem.

The purpose of the paper is to present several advances in Problem 1 in the case of compactly supported generic probability densities in the plane. Being quite non-trivial, the

case of two variables is still conceptually easier than the general case of many variables since for two variables one can encode a generic smooth probability density with compact support (considered up to a diffeomorphism of the plane) by a certain graph with labelled vertices, see the notion of the Reeb graph below. Such encoding is no longer possible for more than two variables. The best possible imaginable result would have been an algorithm which, given a density function f (or its Reeb graph), will calculate its $ucat(f)$ and/or produce its optimal additive unimodal decomposition (1) similar to the process illustrated in Fig. 1 which can be easily transformed into an exact algorithm. However we were not able to achieve this goal here.

The structure of the paper is as follows. In § 2 we collect the preliminary information on probability densities and Reeb graphs to be used later....

Acknowledgements. The second author wants to acknowledge the financial support of his research by the Swedish Research council through the grant 2016-04416.

2. Preliminaries

2.1. **Densities.** In the present paper, we are interested in smooth compactly supported probability densities in $\mathbb{R}^2$. Since any continuous compactly supported function in $\mathbb{R}^2$ is integrable using scaling we can without loss of generality drop the condition $\iint_{\mathbb{R}^2} f = 1$. From now on, we will assume that our functions are defined on a disk $D^2 \subset \mathbb{R}^2$.

Definition 1. (i) A *density* on D^2 is a continuous tame function [1] $f : D^2 \to \mathbb{R}_{\geq 0}$ with finitely many critical values (i.e. values $t \in \mathbb{R}_{\geq 0}$ for which the level $f^{-1}(t)$ changes its homotopy type) and $\partial D^2 \subseteq f^{-1}(0)$. A density is called *generic* if $f^{-1}((0,\infty)) =$ interior (D^2).

(ii) A density is called *unimodal* if it satisfies the additional property that for any $t \geq 0$, $f^{-1}([t,\infty))$ is either contractible or empty. In particular, a unimodal density has exactly one local maximum (which also has to be its global maximum).

(iii) A *Morse density* is a density which is a Morse function and has 0 as a regular value. In particular, each connected component of a regular level of a Morse density is a simple closed curve and each component of a critical level is either a simple closed curve, or a curve with finitely many nodes below, or just a single point, see example in Fig. 2. (The number of nodes on some connected components is equal to the number of critical points of index 1 belonging to it.) Each connected component of a level set which coincides with a single point is either a critical point of index 0 or a critical point of index 2.

(iv) A Morse density is called *generic* if it attains different critical values at different critical points. In particular, all connected components of a critical level of a generic Morse density except one are smooth simple closed curves and the only singular component is either a single point or a smooth closed curve with exactly one node. Such a curve is homeomorphic to the base point union of two circles, see examples in Fig. 3 below. (Generic Morse densities are very close to *excellent Morse functions* used in [17].)

(v) Two densities $f : D^2 \to \mathbb{R}_{\geq 0}$ and $g : D^2 \to \mathbb{R}_{\geq 0}$ with the same sequence of critical values $c_0 < c_1 < \cdots c_r$ are called *equivalent* if there exists an homeomorphism $\theta : D^2 \to D^2$ and a homeomorphism $h : \mathbb{R} \to \mathbb{R}$ preserving the critical values and such that $h \cdot f \cdot \theta = g$.

[1] *tame* means that $f^{-1}(t)$ is an ANR

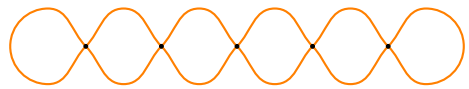


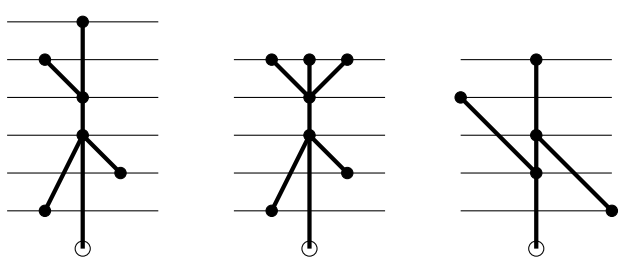
FIGURE 2. Connected component of a level curve with five nodes.

Type 1 Type 2

FIGURE 3. Two topological types of smooth closed curves with one node.

2.2. **Reeb graphs.** [2] This notion will play an important role in our considerations.

Definition 2. A *Reeb graph* (Γ, λ) is a finite graph Γ equipped with a continuous map $\lambda : \Gamma \to \mathbb{R}_{\geq 0}$ with the following properties:
(i) Γ is a rooted tree whose root O corresponds to $\lambda^{-1}(0)$;
(ii) the restriction of λ to each edge is a homeomorphism of this edge onto its image, see Fig. 4.



a) Non-generic Reeb graphs b) Generic Reeb graph

FIGURE 4. Non-generic Reeb graphs versus generic. (Small empty circles are the roots of the Reeb graphs).

Remark 1. In what follows, when we will be talking about the vertices of Γ we will be always excluding its root $O \in \Gamma$ because it corresponds to a regular value of the function λ.

Definition 3. A Reeb graph (Γ, λ) is called *generic* if for each critical value $t > 0$, its critical level $\lambda^{-1}(t)$ contains exactly one vertex of Γ and the valency of each vertex of Γ is either 1 or 3, see Fig 4.

Definition 4. The values of λ at the vertices of Γ (excluding the root) are called the *critical values* of the Reeb graph (Γ, λ).

[2] In the old Soviet literature Reeb graphs are called *Kronrod's graphs* after the Soviet mathematician Alexander Kronrod who apparently defined them earlier than Georges Reeb.

Definition 5. For each vertex x of a generic Reeb graph (Γ, λ), denote by $n_-(x)$ (resp. by $n_+(x)$) the number of edges of Γ incident to x such that the restriction of λ to each such edge takes values smaller than $\lambda(x)$ (resp. larger than $\lambda(x)$.)

For each vertex x of a generic Reeb graph (Γ, λ), define its *index* $i(x)$ as:

$$i(x) = \begin{cases} 0, & \text{if } n_+(x) = 1 \text{ and } \mathrm{n}_-(\mathrm{x}) = 0; \\ 1, & \text{if } n_+(x) + n_-(x) > 1; \\ 2, & \text{if } n_-(x) = 1 \text{ and } \mathrm{n}_+(\mathrm{x}) = 0. \end{cases}$$

Definition 6. Two Reeb graphs (Γ_1, λ_1) and (Γ_2, λ_2) are called *equivalent* if and only if $\Gamma_1 \simeq \Gamma_2$ and λ_1 has the same critical values as λ_2. In particular, there exists a homeomorphism $\theta : \Gamma_1 \to \Gamma_2$ preserving both vertices and edges such that $\lambda_1 = \lambda_2 \cdot \theta$.

Remark 2. Observe that any Reeb graph (Γ, λ) is equivalent to $(\widetilde{\Gamma}, \widetilde{\lambda})$ having the property that the restriction of $\widetilde{\lambda}$ to each edge of $\widetilde{\Gamma}$ is linear. Both $(\widetilde{\Gamma}, \widetilde{\lambda})$ and the homeomorphism θ are uniquely determined.

Definition 7. To each density $f : D^2 \to \mathbb{R}_{\geq 0}$, one associates its Reeb graph $R(f) = (\Gamma, \lambda(f))$ defined using the diagram:

$$\begin{tikzcd} D^2 \arrow[r, "\tilde{f}"] \arrow[d] & \mathbb{R}_{\geq 0} \\ \Gamma = D^2/\sim \arrow[ur, "\lambda(f)"'] & \end{tikzcd}$$

where $x \sim y$ if and only if (i) $f(x) = f(y)$, and (ii) x and y lie in the same connected component of the level set $f^{-1}(f(x)) = f^{-1}(f(y))$.

Remark 3. One can easily check that the associated Reeb graph $R(f)$ as in Definition 7 is indeed a Reeb graph according to Definition 2 with vertices corresponding to the critical values of f. Moreover for generic Morse densities, the index of each vertex of $R(f)$ as in Definition 5 coincides with the index of the critical point corresponding to the respective criticial value.

Lemma 1. *The Reeb graph $R(f)$ corresponding to a generic Morse function f is generic.*

Proof. We need to show that a generic Morse function f gives rise to the generic graph $R(f)$. Indeed, the set of critical values of a (generic) Morse function f coincides with that of its Reeb graph $R(f)$. The fact that to each critical value c of a generic Morse function f corresponds the unique critical point $p(c)$ implies that to each critical value of $R(f)$ corresponds its unique vertex. Further, since all critical points of f are simple, the only phenomena which can occur with the topology of the level set when the parameter passes the critical value are as follows. If c is the critical value of f of index 0, i.e. a local maximum, then a small oval is born when the increasing level parameter t passes c. In this case the corresponding vertex $v(c)$ of its Reeb graph has valency 1 and index 0. If c is the critical value of f of index 1, i.e. the corresponding critical point p a saddle, then either two ovals of the level curve are glued together into a single oval or one oval of the level curve splits into 2 ovals. In this case the corresponding vertex $v(c)$ of its Reeb graph has valency 3 and index 1. Finally, if c is the critical value of f of index 2, i.e. a local minimum, then a small oval disappears when t passes c. In this case the corresponding vertex $v(c)$ of its Reeb graph has valency 1 and index 0. $\square$

We will call Reeb graphs arising from Morse densities *realizable*. Let us now describe the additional conditions required for the converse of Lemma 1 to be true, i.e. for the realizability of generic Reeb graphs.

Definition 8. Take a generic Reeb graph (Γ, λ). Since Γ is a rooted tree there exists a unique path connecting the root O with a given vertex v of Γ. In this way Γ attains its canonical orientation "pointing away from its root". Obviously, each trivalent vertex gets one incoming edge and two outgoing. The canonical orientation of an edge is called *positive* if the value of λ at its out-vertex is bigger than at its in-vertex and *negative* otherwise. In other words, the function λ induces another orientation of each of Γ in the direction of its increase. An edge is *positive* if both its orientations coincide and *negative* if they are opposite.

We say that a generic Reeb graph (Γ, λ) is *balanced* if at each its trivalent vertex v at least one of the outgoing edges has the same sign of orientation as the incoming edge, i.e. if the incoming edge has positive orientation then one or both outgoing edges have positive orientation. (Similarly, if the incoming edge has a negative orientation.) Trivalent vertices of a balanced generic Reeb graph with both outgoing edge having the same sign of their orientations as the incoming edge will be called *type* 1*-vertices*. Trivalent vertices of a balanced generic Reeb graph with one outgoing edge having the same sign of its orientation as the incoming edge and one having the opposite sign will be called *type* 2*-vertices*.

Fig. 5 a) represent balanced generic Reeb graphs while Fig. 5 b) shows a non-balanced graph; in the latter case the region bounded by the dotted circle explains why.

Lemma 2. *If a generic Morse function f has a critical point of index 1 with the critical value c such that the singular connected component of the level curve $f = c$ has type 1 (resp. type 2), see Fig. 3, then the corresponding vertex of its Reeb graph $R(f)$ has type 1 (resp. type 2).*

Proof. Indeed, for the singular connected component of type 1, the direction of the gradient outside the two ovals continues inside the ovals, i.e. if the function was increasing (resp. decreasing) towards the ovals it continues to increase (resp. decrease) inside them as well which means that the orientations of the incoming edge of $R(f)$ at the vertex $v(c)$ is the same as those of the outgoing edges. On the other hand, for the singular connected component of type 2, the direction of the gradient outside the bigger ovals continues inside this ovals, but is opposite to that inside the smaller oval. This means that the orientations of the incoming edge of $R(f)$ at the vertex $v(c)$ is the same as the orientation of the outgoing edge corresponding to the bigger oval, but is opposite to the orientation of the outgoing edge corresponding to the smaller oval. □

Theorem 3. *Generic Morse functions are in $1-1$-correspondence with generic balanced Reeb graphs.*

Proof. The $\Rightarrow$ -implication is settled in the above Lemmas 1 and 2.

Now let us settle the $\Leftarrow$ -implication claiming that a generic balanced Morse graph (Γ, λ) gives rise to a generic Morse function f such that $R(f) \simeq (\Gamma, \lambda)$ and that such f is defined uniquely up to diffeomorphism. Given a balanced trivalent tree Γ we can uniquely restore the nested configuration of curves of type 1 and 2 and points starting with the largest oval corresponding to the root O which contains the rest of it. To obtain it we draw an oval corresponding to O and up in the tree. (Each edge corresponds to a cylinder consisting of a family of closed simple curves.) Reaching a trivalent vertex we draw the curve of type

1 in the respective oval if the vertex is of type 1 and the curve of type 2 if the vertex is of type 2. □

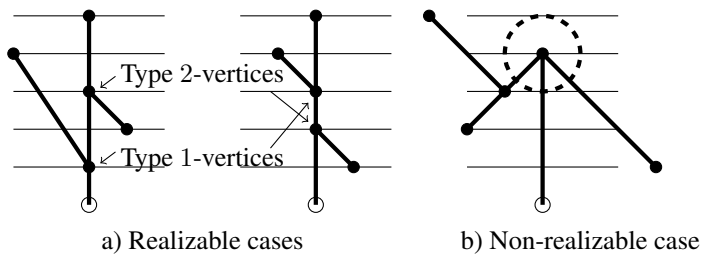


FIGURE 5. Realizable generic Reeb graphs versus non-realizable.

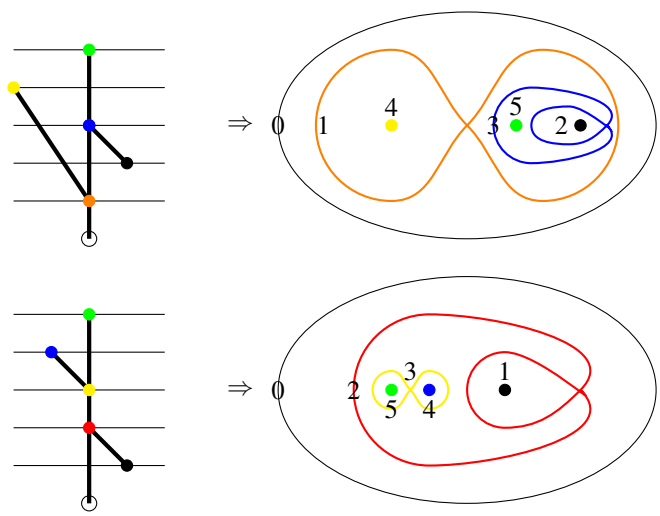


FIGURE 6. Construction of a nested family of type 1 and 2 curves.

Remark 4. Observe that Theorem 3 is not quite true for non-generic Morse densities. There are some miniscule subtleties. For any two embedded graphs in the sphere with all vertices of even degree, if they have same Euler characteristics, but not diffeomorphic, one can have two functions for which these graphs are the preimages of saddle critical value, and all other critical points (max and mins are the same). Then the Reeb graphs will be the same as well, but the functions are not equivalent. WHat about the case when Γ is a trivalent tree?

2.3. **Reeb graphs and bar codes.** INTRODUCE BAR CODES...

Lemma 4. *In terms of the set $\mathcal{B}_0(f)$ of zero-bar codes, Property P for $R(f)$ is equivalent to the following:*

If $0 < c_1 < c_2 < \cdots < c_N$ is the sequence of all critical values of $R(f)$ and one denotes by $ind(c_i)$ the index of the unique critical point corresponding to the critical value c_i, then

(1) *each bar code has multiplicity one;*
(2) *there are no open bar codes and only one closed bar code* $[0, c_N]$*;*
(3) *if* $[c_i, c_j)$ *is a closed-open bar code then* $ind(c_i) = 0, j = i + 1$ *and* $ind(c_j) = 1$ *and the multiplicity of* $[c_i, c_{i+1})$ *is one;*
(4) *if* $(c_i, c_j]$ *is an open-closed bar code then* $ind(c_j) = 2, ind(c_i) = 1$ *and the multiplicity of* $(c_i, c_j]$ *is one.*

Proof. BLA □

Remark 5. Note that in view of the Poincaré duality, the set $\mathcal{B}_1(f)$ of 1-bar codes is in bijective correspondence with $\mathcal{B}_0(f)$. The 1-bar code which corresponds to the 0-bar code obtained from this one by changing "[" into "(" and "]" into ")".

Each collection of bar codes $\mathcal{B}_0(f)$ or $\mathcal{B}_1(f)$ determines the generic Morse density up to an equivalence. (Most likely this happens for non-generic Morse densities and Reeb graphs as well).

The calculation of bar codes is "computer friendly" i.e. achievable by an effective computer implementable algorithm.

Proposition 5. *Assertion (ii) of Theorem 3 is equivalent to the coincidence of the collections of* 0*-bar codes or to the coincidence of the collections of* 1*-bar codes.*

Proof. BLA □

3. Main results

3.1. Symmetric representatives of generic Morse densities.

Theorem 6. Any generic Morse density $f(x, y)$ on D^2 is equivalent to a generic Morse density $g(x, y)$ with the following properties:
(i) The density $g(x, y)$ is *y-symmetric*, i.e. $g(x, y) = g(x, -y)$;
(ii) The connected components of a critical level are : circles with center on the $x-$axis, a pair of two tangent circles whose centers and point of tangency located on the $x-$axis, or one point located on the $x-$axis;
(iii) The lines $y(t) = c$ intersect transversally or tangentially any connected component of a critical level of the density.

See illustration in Fig. 5.

Proof. The proof is based on the fact that any realisable Reeb graph can be realised by a density which satisfies (i) – (iii) above and on the fact that generic Morse densities with the same Reeb graph are equivalent. □

Remark 6. A symmetric representative of a generic Morse density will be called a *symmetric generic Morse density*. Observe that besides being symmetric such a function has all its critical points on the symmetry axis. Additionally notice that restriction of a unimodal density in the plane to an arbitrary line is, in general, not unimodal. However, the restriction of a unimodal symmetric density to its symmetry axis is still unimodal. (Indeed, for a symmetric unimodal Morse density $\phi(x, y)$ the domain $\Omega_k := \{(x, y)|\phi(x, y) \geq K > 0\}$ (if nonempty) is contractible and symmetric. Therefore its intersection with the x-axis is contractible as well.)

Definition 9. Two equivalent symmetric generic Morse densities are called *symmetrically equivalent* if one can find a symmetric diffeomorphism sending one to the other.

The next statement is straight-forward.

Lemma 7. *The number of symmetrically non-equivalent forms of a given generic Morse density equals* $2^{\sharp(saddles)-1}$*, where* $\sharp(saddles)$ *is the number of critical points of index* 1*=number of trivalent vertices of the Reeb graph.*

Proof. (Sketch) This number equals the number of different planar embeddings of the rooted tree modulo the global change $x \mapsto -x$. □

Definition 10. Given a symmetric generic Morse function, we can consider the minimal number of summands in its additive decomposition into symmetric unimodal densities. We denote this number by $ucat^{sym}(f)$. Obviously,

$$ucat(f) \leq ucat^{sym}(f).$$

The latter inequality can be strict.

3.2. **Generic Morse densities without local minima.** Here we consider a special class of generic Morse densities which have no local minima. We will call them *min-less*. It is easy to characterize such functions.

Lemma 8. *A generic Morse density* f *has no local minima if and only all its connected singular curve are of type* 1*, see Fig. 3. Equivalently, all trivalent vertices of* $R(f)$ *are of type* 1.

Proof. (Sketch) Indeed, If f has no local minima, then there is no closed domain $D \subset \mathbb{R}^2$ bounded by the oval $\mathcal{O}$ of a singular curve such that near $\mathcal{O}$ f decreases inside D. On the other hand, if a connected singular curve of type 2 is present this necessarily happens. And conversely, If we have a smooth function such that it has a connected closed level curve on such that this function decreases close to the boundary on the inside, then this function must have a minimum inside this curve. □

Lemma 9. *Each min-less generic Morse density* $f(x, y)$ *has a symmetric form* $\widetilde{\phi}(x, y)$ *which is unimodal on each vertical line* $x = K$.

We call such symmetric generic densities *ridge-like*.

Proof. (Sketch) Assume that $\phi(x, y)$ is a symmetric generic Morse density without local minima. To start with, observe that, for any $x_0 \in \mathbb{R}$, the restriction $f(x_0, y)$ has a local maximum at $y = 0$. Consider the vector field $\mathcal{V}_\phi = -\text{grad}\,\phi$. All trajectories of $\mathcal{V}_\phi$ starting in the open upper (resp. lower) halfplane stay in the respective halfplane and tend to infinity. We can use these trajectories as the lines $x = const$ for a new function $\widetilde{\phi}(x, y)$. WRITE THIS BETTER! □

Lemma 10. *For any ridge-like symmetric density* $\phi(x, y)$*, one has*

$$ucat^{sym}(\phi(x, y)) = ucat(\phi(x, 0)).$$

Proof. Indeed if we have a symmetric additive unimodal decomposition of $\phi(x, y)$, then its restriction to $y = 0$ gives us a unimodal decomposition of $\phi(x, 0)$. (Recall that the restriction of a unimodal symmetric density to its symmetry axis is unimodal. Therefore, $ucat^{sym}(\phi(x, y)) \geq ucat(\phi(x, 0))$. In the opposite direction, consider a ridge-like function $\phi(x, y)$ and let

$$\phi(x, 0) = \phi_1(x) + \phi_2(x) + \cdots + \phi_k(x)$$

be the optimal unimodal additive decomposition of its restriction $\phi(x,0)$ as shown in Fig. 1. Let us show how to extend this decomposition to the unimodal symmetric decomposition

$$\phi(x,y) = \phi_1(x,y) + \phi_2(x,y) + \cdots + \phi_k(x,y)$$

such that $\phi_j(x,0) = \phi_j(x)$, for $j = 1,\ldots,k$ which concludes the proof. To obtain such an extension look again at Fig. 1. We start with the leftmost unimodal function in the decomposition shown in green. On the interval (x_0, x_1) on which this function coincides with the $\phi(x,0)$, we set $\phi_1(x,y) = \phi(x,y)$ for all y. On the interval x_1, x_2 on which $\phi_1(x) < \phi(x)$, we set

$$\phi_1(x,y) = max\{\phi(x,y) - \phi(x,0) + \phi_1(x), 0\},$$

see Fig. 7. Now subtracting the function $\phi(x,y)$ from $\phi(x,y)$, we can use induction. □

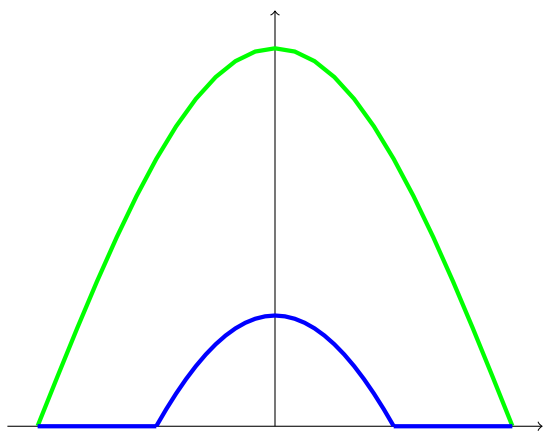


FIGURE 7. Modifying the ridge-like function in the y-direction.

Problem 2. *Is it true that for a min-less generic density $f(x,y)$,*

$$ucat(f) = \min_{\phi \in Sym(f)} ucat(\phi(x,0)),$$

where $Sym(f)$ is the set of symmetric densities equivalent to $f(x,y)$?

Fig. 8 shows the topographic map, the Reeb graph and the collection $\mathcal{B}_0(f)$ for a generic Morse density f on a disk.

3.3. **Special decompositions and estimates of $ucat$.** The calculation/estimation of $ucat(f)$ of a given density f will be simplified by the following two special type of decompositions.

First decomposition. Suppose f is a generic Morse density with a critical value c and let C be the singular connected component of $f^{-1}(c)$. i.e., the one which contains the critical point x. Suppose that $ind(x) = 1$ and C is of type 1, i.e. $C = C_1 \cup C_2$, each C_i being a simple closed curve lying outside each other. Then $x = C_1 \cap C_2$. This implies that f take values larger than c in the interior of each of the closed curve C_1 and C_2. Denote by D_1 resp. D_2 the domains bounded by C_1 resp. C_2.

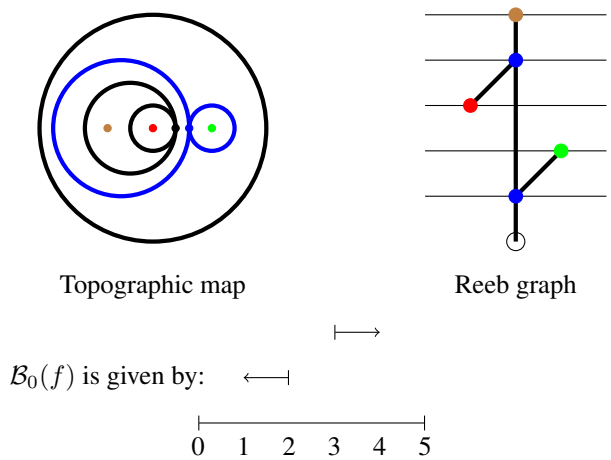


FIGURE 8. The topographic map, the Reeb graph and the collection of bar codes $\mathcal{B}_0(f)$.

Proposition 11. Under the above hypotheses one can produce two generic Morse densities f_1 and f_2 with $f = f_1 + f_2$ enjoying the following properties:

(i) f_1 equal to f outside an arbitrary small neighborhood of D_2 and no critical points in this neighborhood;

(ii) f_2 equal to $f - c$ on the interior of D_2 and zero outside D_2:

(iii) f_1 and f_2 are arbitrary closed to the tame densities f_1' and f_2', where

$$f_1' = \begin{cases} f(x), & \text{for } x \in D^2 \setminus D_2, \\ c, & \text{for } x \in D_2; \end{cases}$$

and

$$f_2' = \begin{cases} f(x) - c, & \text{for } x \in D_2, \\ 0, & \text{for } x \in D^2 \setminus D_2. \end{cases}$$

Proof. Indeed, observe that, by construction, we have an obvious additive decomposition $f = f_1' + f_2'$ where f_1' and f_2' are non-negative functions. Say smth about their small perturbation producing f_1 and f_2... □

In particular, the critical points (critical values) of f_1 are the same as those of f which are located in $D^2 \setminus D_2$ and the critical points (critical values) of f_2 are the same as those of f located in the interior of D_2.

Clearly then $Cr(f) = Cr(f_1) \sqcup Cr(f_2) \sqcup x$ and then

$$ucatf \leq ucatf_1 + ucatf_2$$

This implies

Corollary 1. If the generic Morse density has no critical points of index 0 then $ucatf \leq n_1 + 1 = n_2$.

This inequality is sharp however not so good as the one provided by Observation 1.

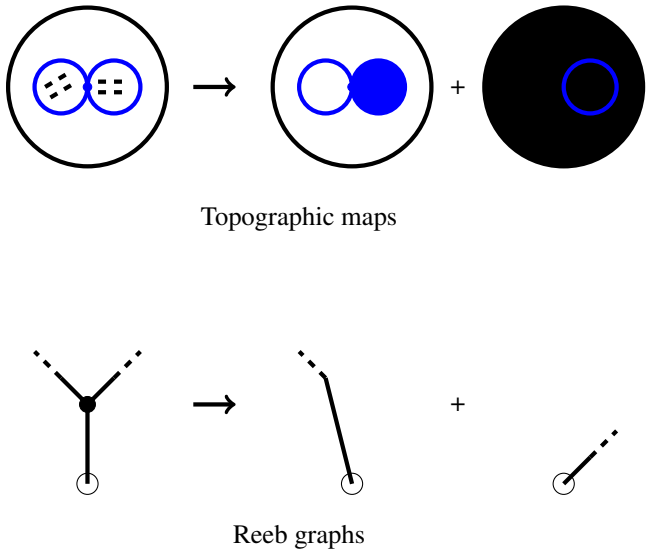


FIGURE 9. Illustration of the first decomposition

Remark 7. One derives the Morse densities f_1 and f_2 by arbitrary small perturbations from the tame densities f_1' and f_2' which are not exactly Morse. Note that f_1' is Morse outside ${f_1'}^{-1}(c) = C_1(c) \cup D_2$ and ${f_2'}^{-1}(0) = D^2 \setminus D_2$ which retracts by deformation to ∂D^2.

Fig. 9 shows the topographic maps of f, f_1', f_2'. The modifications of f_1' and f_2' producing f_1 and f_2 is left to the reader and are basically obtained by small deformations of piecewise smooth functions into generic smooth functions. ONE NEEDS MORE WORK HERE!

Second decomposition.

Suppose f is a generic Morse density with the critical points y of index 0 and x of index 1 having consecutive critical values $c' < c$. Assume that for and $\epsilon' > 0$, no other critical values exist in the interval $[c' - \epsilon', c + \epsilon']$. Denote by $y \sqcup C'$ the critical level $f^{-1}(c')$ (shown in heavy black in Fig. 10) and by $C_1 \sqcup C_2$ (shown in heavy blue in Fig. 10) the critical level $f^{-1}(c) = C_1 \cup C_2$, C_1 and C_2 simple closed curves tangent at x with $C_1 \cap C_2 = x$ bounding the domain D_1 and D_2 with $D_1 \subset D_2$ and K a pass (shown in red in Fig. 10) from y to x which intersects any level $f^{-1}(t), t \in (c', c]$ transversally in only one point. Let $\mathcal{U}$ be an arbitrary small neighborhood of K inside the domain bounded by C, see Fig. 10 below.

Proposition 12. Under these hypotheses one can provide the generic Morse density f_1 and the smooth unimodal density f_2 with $f = f_1 + f_2$ and satisfying the following properties:

(i) $f = f_1$ on $D^2 \setminus \mathcal{U}$ and has no critical points inside $\mathcal{U}$; in particular f_1 has the same critical values less c' and c and the same critical points less y and x as f.

(ii) f_2 is a unimodal smooth map with support a contractible domain with smooth boundary inside $\mathcal{U}$ (see $\overline{U}$ in Fig. 11). When restricted to this domain f_2 is a Morse function with x the only critical point of index 2.

Proof. The construction of f_1 and f_2 are explained below using Fig. 10 and Fig. 11.

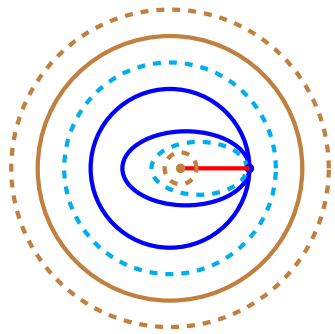


FIGURE 10. Illustration of the second decomposition

In view of Theorem 2 by changing the density up to equivalence one can suppose the following, see Fig. 10):

- the density is $y-$symmetric with the lines $Y = t$ transveral or tangential to the connected components of the critical levels of the density and all critical points located on the $x-$axis,
- the regular level $f^{-1}(c' - \epsilon)$ is the doted brown circle and the critical level $f^{-1}(c')$ is $f^{-1}(c') = \{y \sqcup C'\}$ where $y = (0, 0)$ and C' the brown circle,
- the critical level $f^{-1}(c)$ is $f^{-1}(c) = C_1 \cup C_2$ with C_1 and C_2 the tangent blue circles at $x = C_1 \cap C_2$ and the regular levels $f^{-1}(c - \epsilon)$ the two disjoint simple closed curves $C_1(\epsilon)$ and $C_2(\epsilon)$ collored in dotted light blue, (THEY DO NOT TOUCH THE BLUE CIRCLES AS MIGHT APPEAR IN THIS PICTURE)
- K is the segment $[x, y]$ coloured in red and $\mathcal{U}$ an open neighborhood of K,
- ϵ is so small that $\mathcal{U}$ contains a neighborhood of the segment $[y, z]$ (z is the intersection of the $x-$axis with C_2').

One choses a close tubular neighborhood $\overline{U}$ of the segment $[y, z]$ in the closed domain $D_2(\epsilon)$ whose boundary is the curve $C_2(\epsilon)$ (cf. Fig. 11.) whose boundary $\overline{L} = L_{11} \cup L_{1,2} \cup L_{2,1} \cup L_{1,3} \cup L_{3,1} \cup L_{1,4} \cup L_{4,1} \cup L_{4,4}$ with

- L_{11} (shown in green), subset of the level $f^{-1}(\epsilon)$,
- $L_{1,2}$, $L_{2,1}$ (shown in light red) $L_{1,3}L_{3,1}$ (shown in dotted blue), $L_{1,4}$, $L_{4,1}$ (shown in light dotted red), all on the lines $Y = \epsilon$ and $Y = -\epsilon$
- $L_{4,4}$(shown in light blue), neighborhood of z in the level $f^{-1}(c - \epsilon)$.

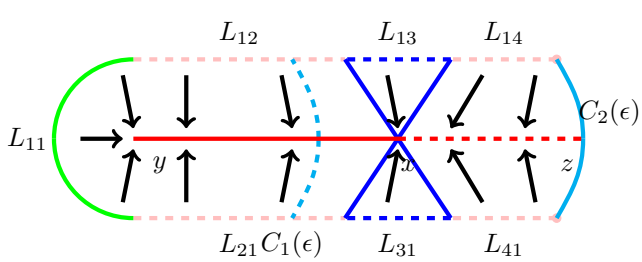


FIGURE 11. The tubular neighborhood $\bar{u}$.

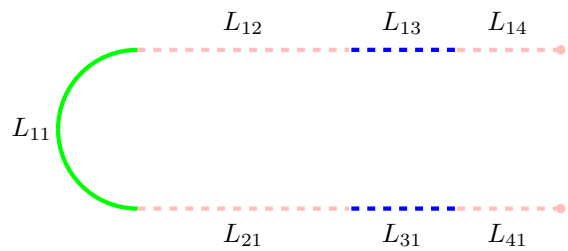


FIGURE 12. The curve L.

Write $\overline{L} = L \cup L_{4,4}$ with

$$L = L_{11} \cup L_{1,2} \cup L_{2,1} \cup L_{1,3} \cup L_{3,1} \cup L_{1,4} \cup L_{4,1}.$$

One can regard $\overline{U}$ as the "mapping cylinder" of the map $\omega : L \to \overline{K} = [y, z]$, map which sends:
(a) L_{11} to y, $L_{1,3} \cup L_{3,1}$ to x;
(b) $L_{1,2}$ and $L_{2,1}$ homeomorphicaly onto the segment $K = [x, y]$;
(c) $L_{1,4}$ and $L_{4,1}$ homeomorphially onto the segment $\overline{K} = [x, z]$.

More precisely one has the surjective map $\theta : L \times [0, 1] \to \overline{U}$ such that:

- θ restricted to $L \times [0, 1)$ is a diffeomorphism onto $\overline{U} \setminus \overline{K}$ and restricted to $L \times 0$ is the inclusion $L \subset \overline{U}$,
- θ restricted to $L \times 1$ is equal to ω,
- θ restricted to any $u \times [0, 1]$ is linear and a homeomorphism onto its image. and one can choose ω s.t. the density f is strictly decreasing along any ray $\theta(u \times [0.1])u \in L$, from $f((\theta(u \times 0))$ to $f(\omega(u))$.

Observe that for a smooth family of decreasing maps $h_u(t) : [0.1] \to \mathbb{R}_{\geq 0}$ from $a(u)$ to $b(u)$ with $a(u) > b(u)$, one can find two smooth families $h_{1u}(t)$ and $h_{2u}(t)$, the first decreasing from $a(u)$ to 0 and the second increasing from 0 to $b(u)$ such that

$$h_u(t) = h_{1u}(t) + h_{2u}(t).$$

Based on this observation applied to $h_u(t) = f(\theta(u \times t))$ one changes the density f on $\overline{U}$ into f_1' by using $h_{1u}(t)$ on each ray $\theta(u \times [0, 1])$.

A careful choice of $h_{1u}(t)$ can make f' a Morse function outside $f'^{-1}(c-\epsilon) = f^{-1}(c-\epsilon) \cup \overline{K}$ which retracts by deformation to $f^{-1}(c - \epsilon)$. One considers f_2' defined by f on $\theta(L \times 1)$ and by $h_{2u}(t)$ on any ray $\theta(u, [0, 1])$and $zero$ away from $\overline{U}$. This is clearly unimodal.

It is easy to see that by arbitrary little perturbation of f_1' and f_2' (by adding to f_1' a small unimodal smooth map with support in a small neighborhood of $\overline{K}$, so small that does not introduces new critical points when added to f_1' and subtracted from f_2' still keeps the difference unimodal, one ends up with the densities f_1 and f_2 as stated. □

Observation 1. As a consequence of Theorem 6, Proposition **??** and Proposition 12, an upper bound estimate for $ucat\, G$ of any realizable Reeb graph G (or for a generic Morse density) has as upper bound the $ucat(G_0)$, G_0 of the graph obtained by deleting the vertices of index zero and adjacent edges plus the number of vertices of index zero.

This is not yet the best we can do except in the case of no critical points of index zero.

3.4. **A test for** $ucat(f) = 2$**.** Let us begin with the example of a unimodal density $g : D^2(1+\epsilon) \to \mathbb{R}_{\geq 0}$ defined as follows:

(2)
$$g(x,y) := \begin{cases} (M/4)(3-x), & \text{if } \mathrm{x}^2+\mathrm{y}^2 \leq 1; \\ (M/4\epsilon)(3-x/\sqrt{x^2+y^2})(1+\epsilon-\sqrt{x^2+y^2}), & \text{if } 1 \leq \mathrm{x}^2+\mathrm{y}^2 \leq 1+\epsilon. \end{cases}$$

Observe that the restriction of $g(x,y)$ to $D^2(1)$ is the linear function $\rho(x,y) = M/4(3-x)$ and on $D^2(1+\epsilon) \setminus D^2(1)$ it is decreasing linearly along each radius corresponding to $\theta \in [0, 2\pi]$ from the value $(M/4)(3-\cos\theta)$ attained at the point $(cos\theta, sin\theta)$ to zero attained at the point $(1+\epsilon)\cos\theta, (1+\epsilon)\sin\theta)$.

Fig. 13 below indicates the levels for $3M/4$ (red), $M/2$ (blue) and $M/4$ (green). The corresponding set $g(x,y) \geq const$ are the domains bounded by these curves. This map is unimodal, with maximal value M and any other unimodal Morse density with the same maximal value is equivalent to g. The point where g reaches its maximum is $(-1, 0)$.

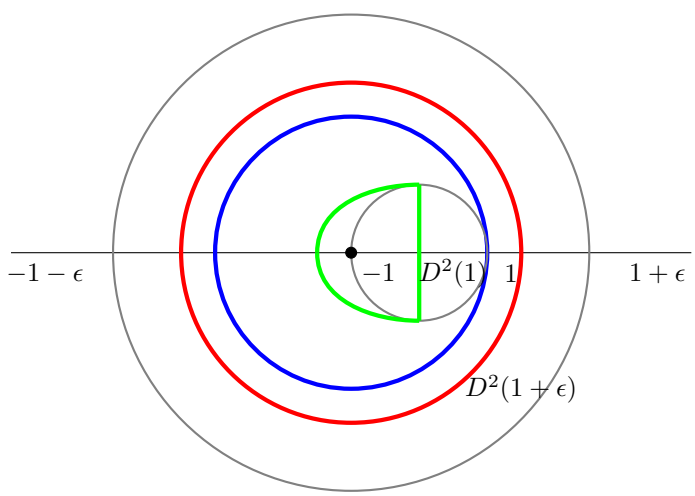


FIGURE 13. Illustration.

Consider now the density $\tilde{f} : D^2(1+\epsilon) \to \mathbb{R}_{\geq 0}$ derived from the smooth map $f : D^2(1) \to [M, \infty)$ which satisfies the following properties:

- $f^{-1}(M) = \partial D^2(1)$ with M regular value,
- $(f'_x + M/4)^2 + (f'_y)^2 \neq 0$ for any $(x,y) \in D^2$

defined by the following formula:

(3)
$$\tilde{f}(x,y) := \begin{cases} f(x,y), & \text{if } \mathrm{x}^2+\mathrm{y}^2 \leq 1 \\ (M/\epsilon)(1+\epsilon-\sqrt{x^2+y^2}), & \text{if } 1 \leq \mathrm{x}^2+\mathrm{y}^2 \leq 1+\epsilon \end{cases}$$

and set
$$h(x,y) := \tilde{f}(x,y) - g(x,y).$$

Proposition 13. The map h is a continuous unimodal density with maximal value equal to $M/2$ attained at $(1, 0)$.

PROOF?

First observe that the restriction of h to $D^2(1)$ is a smooth map equal to $f(x, y) + M/4(x - 3)$ which is a smooth map with no critical values on D^2 in view of (2) above and whose restriction to the boundary $\partial D^2(2)$ at the point $(\cos\theta, \sin\theta)$ given by $M/4(1+\cos\theta)$ whose maximal value is taken at $(1, 0)$

The restriction of h to $D^2(1 + \epsilon) \setminus D^2(1)$ is decreasing linearly along each radius corresponding to $\theta \in [0.2\pi]$ from $M/4(3 - \cos\theta)$ to 0. Then the over level of any value t is therefore contractible. Fig. 14 shows the graphs of the restrictions of $g, \tilde{f}, h$ to the $x-$axis.

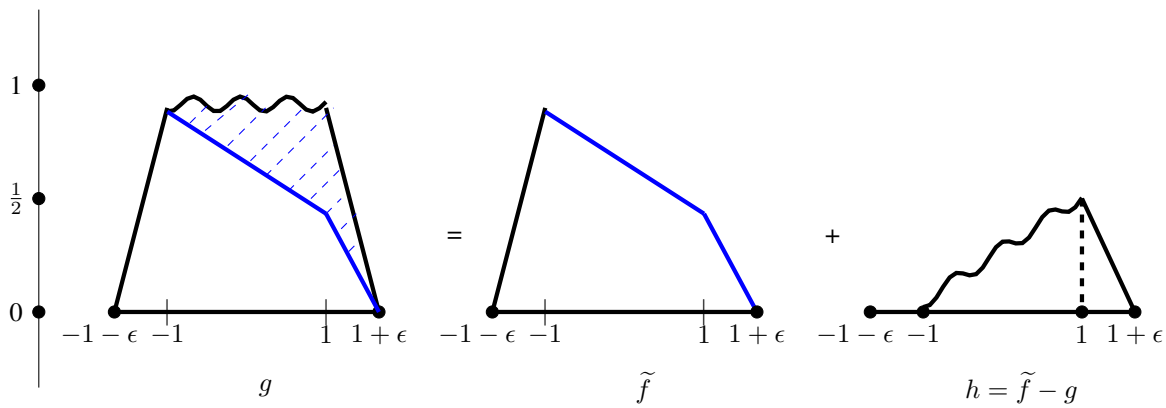


FIGURE 14. Decomposition into two summands

In particular $\tilde{f} = g + h$ both unimodal hence $ucat(\tilde{f}) = 2$. As a consequence we have $M/4(3-\cos\theta)$ value taken on $(\cos\theta, \sin\theta)$ to zero attained at $(1+\epsilon)\cos\theta, (1+\epsilon)\sin\theta)$.

The above discussion implies that for a density $f : D^2 \to \mathbb{R}_{\geq 0}$ with the property that if $\omega(f) := \sup_{x^2+y^2\leq 1}(f_x^2 + f_y^2)^{1/2} > 0$ and $\omega(f) < c_1$ the smallest critical value then $ucat(f) \leq 2$.

4. OUTLOOK

1. What about multidimensional case?

REFERENCES

[1] C. Amendola, A. Engström, and C. Haase, *Maximum number of modes of Gaussian mixtures*, Information and Inference: A Journal of the IMA, iaz013, https://doi.org/10.1093/imaiai/iaz013.

[2] Y. Baryshnikov and R. Ghrist, *Unimodal Category and Topological Statistics*, Proc. NOLTA, (2011), available from http://ect.bell-labs.com/who/ymb/

[3] Y. Baryshnikov and R. Ghrist, *Topological Unimodal Decompositions*, available from http://ect.bell-labs.com/who/ymb/

[4] Y. Baryshnikov and R. Ghrist, *Minimal Unimodal Decompositions on Trees*, arXiv: 1806.09673v1, LicenseCC BY 4.0.

[5] Y. Baryshnikov, *Topological unimodal decomposition* slides of a talk MASS, September 07,

[6] V. Benci, K. A. De Rezende, *Maximal Poincare polynomials and minimal Morse functions,* Proc. AMS, **129(2)**, (2001), 3511–3518.
[7] D. Burghelea, *New topological invariants for real- and angle-valued maps; an alternative to Morse-Novikov theory,* World Scientific Publishing Co. Pte. Ltd, 2017
[8] D. Burghelea and T. K. Dey, *Persistence for circle valued maps. Discrete Comput. Geom.*. **50** (2013), 69–98.
[9] M. Carreira–Perpinan and C. Williams, *On the number of modes of a Gaussian mixture*, in Scale-Space Methods in Computer Vision, Lecture Notes in Comput. Sci., vol. 2695, 2003, 625–640.
[10] O. Cornea, G. Lupton, J. Oprea, and D. Tanre, *Lusternik - Schnirelmann Category*, Amer. Math. Soc., 2003.
[11] D. Govc, *Unimodal category and the monotonicity conjecture*, arXiv: 1709.06547.
[12] D. Govc, *Unimodal category*, PhD thesis, Univ. of Ljubljana, 2017.
[13] St. Huntsman, *Topological Mixture Estimation*, arXiv: 1712.04487, Proceedings of the 35th International Conference on Machine Learning, PMLR 80:2088–2097, 2018.
[14] L. Hickok, J. Villatoro, and X. Wang, *Unimodal category of* 2*-dimensional distributions*, Report of the Research Experience Program, UIUC, 2014.
[15] I. Kakiuchi, *Unimodality conditions of the distribution of a mixture of two distributions,* Math. Sem. Notes Kobe Univ., 9, 1981, 315–325.
[16] J. Kemperman, *Mixtures with a limited number of modal intervals,* Ann. Statist. 19, 1991, 2120–2144.
[17] L. Nicolaescu, *Counting Morse functions on the two sphere*, Compositio Math. 144 (2008) 1081–1106.
[18] S. Ray and B. Lindsay, *The Topography Of Multivariate Normal Mixtures*, Annals of Statistics, Vol. 33, No. 5, 2005, 2042–2065.
[19] B. Shapiro, *Personal communication.*

DEPARTMENT OF MATHEMATICS, OHIO STATE UNIVERSITY, 231 W 18TH AVENUE, COLUMBUS, OH 43210, USA
E-mail address: burghele@math.ohio-state.edu

DEPARTMENT OF MATHEMATICS, STOCKHOLM UNIVERSITY, S-10691, STOCKHOLM, SWEDEN
E-mail address: shapiro@math.su.se