# Indefinite Morse 2–functions; broken fibrations and generalizations - DRAFT

David T. Gay, Robion Kirby *

*Euclid Lab, 428 Kimball Rd, Iowa City, IA 52245*
*Department of Mathematics,University of Iowa, Iowa City, IA 52242*
and
*University of California, Berkeley, CA 94720*

Email: `d.gay@euclidlab.org` and `kirby@math.berkeley.edu`

**Abstract**

*This is a DRAFT, comments are welcome. There may or may not be sections added later on maps to surfaces other than disks and spheres and on a collection of interesting examples.*

We study generic smooth maps from smooth manifolds to surfaces, which we call Morse 2–functions. In the absence of definite folds (in which case we say that the Morse 2–function is indefinite), these are natural generalizations of broken (Lefschetz) fibrations. We prove existence and uniqueness results for indefinite Morse 2–functions mapping to the disk and the 2–sphere. By "uniqueness", we mean that we show how to eliminate definite folds from generic homotopies between indefinite Morse 2–functions, so that a set of moves, remaining indefinite, suffices to go between two homotopic indefinite Morse 2–functions. We also pay close attention to connectedness of fibers, and are able to extend our existence and uniqueness results to indefinite Morse 2–functions with connected fibers.

**AMS Classification numbers** Primary: 57M50

Secondary: 57R17

**Keywords:** broken fibration, Morse function, Cerf theory, definite fold

*Supported in part by South African National Research Foundation Focus Area grant FA2007042500033 and United States National Science Foundation grant EMSW21-RTG

## 1 Introduction

We are interested in generalizing Morse theory, the study of generic maps from arbitrary smooth manifolds to 1–manifolds, to the study of generic maps to 2–manifolds. In particular, this paper will generalize various existence and uniqueness results about Morse functions without critical points of minimal or maximal index, which we call *indefinite Morse functions*.

For example, any nontrivial homotopy class of maps from a closed, connected manifold to $S^1$ can be represented by a $S^1$–valued indefinite Morse function, and any connected cobordism between two nonempty manifolds has an indefinite Morse function to $I = [0, 1]$. Generic homotopies connecting Morse functions are Morse for all but finitely many times, at which times cancelling pairs of critical points appear or disappear, and we will generalize the fact [9] that any two indefinite $I$–valued Morse functions on a fixed cobordism can be connected by a generic homotopy $f_t$ such that, at all intermediate times when $f_t$ is Morse, $f_t$ is indefinite. In other words, if we start and end without index 0 or $m = \dim(M)$ critical points, we never need to introduce index 0 or $m$ critical points in between.

The motivation for generalizing this line of results to consider maps to two dimensions comes from the importance of Lefschetz fibrations in complex and symplectic geometry and from the new ideas around broken Lefschetz fibrations [3, 6, 11, 12]. Lekili [10] shows that if two broken Lefschetz fibrations on a 4–manifold can be connected by an indefinite generic homotopy then they can be related by a finite sequence of elementary moves, staying in the world of broken Lefschetz fibrations. Thus our uniqueness result for indefinite Morse 2–functions over disks and spheres (see below) was developed in the hope that it will help, using Lekili's results, to show that the potential invariants of 4–manifolds proposed by Perutz are in fact invariants. However, the existence and uniqueness results we give can also be set in the context of singularity theory.

We will give precise definitions later, but we begin with the essential ideas needed to understand the statements of our main results: Our generalization of a Morse function is a *Morse 2–function*, which is a suitably generic smooth map to a 2–manifold. The local behaviour of a Morse 2–function on an $n$–manifold is $(t, x_1, \ldots, x_{n-1}) \mapsto (t, f_t(x_1, \ldots, x_{n-1}))$, where $f_t$ is a generic homotopy of Morse functions, as described above; the singular set is 1–dimensional and its image is immersed in the 2–manifold base except at cusps, where $f_t$ fails to be Morse. A non-cusp point in the singular set is called a *fold point* and an arc of fold points is called a *fold*. A minimal or maximal index critical point of a

Morse function is a *definite* critical point (because the quadratic form giving the local model there is definite), and a fold is *definite* if the corresponding Morse singularities of $f_t$ are definite. A Morse function or Morse 2–function is *indefinite* if it has no definite critical points or folds, respectively. (At the end of this introduction we provide some commentary and history on related terminology and results involving the adjectives "broken", "wrinkled" and "Lefschetz".) A generic homotopy between Morse 2–functions will be a Morse 2–function for all but finitely many times, at which times one of a short list of moves occurs. A generic homotopy between Morse functions or Morse 2–functions is *indefinite* if it is indefinite at all the times when it is a Morse function or a Morse 2–function, respectively.

**Theorem 1.1** *Given any smooth, compact, connected $n$–manifold $X^n$ with nonempty boundary, with $n \geq 3$, and given an indefinite Morse function $g\colon \partial X \to S^1$, there exists an indefinite Morse 2–function $G\colon X \to D^2$ with $G|_{\partial X} = g$. If the level sets of $g$ are all connected, then we can arrange that all fibers of $G$ are connected. Given two indefinite Morse 2–functions $G_0, G_1\colon X \to D^2$ with $G_0|_{\partial X} = G_1|_{\partial X}$, if $n \geq 4$ then there exists an indefinite generic homotopy $G_s\colon X \to D^2$ between $G_0$ and $G_1$ with $G_s|_{\partial X} = G_0|_{\partial X}$ for all $s \in I$. If neither $G_0$ or $G_1$ have any disconnected fibers, then the same is true for all $G_s$.*

**Theorem 1.2** *Given any smooth, closed, connected $n$–manifold $X^n$, with $n \geq 3$, and any framed codimension–2 submanifold $F^{n-2} \subset X$, there exists an indefinite Morse 2–function $G\colon X \to S^2$ with $G^{-1}(\text{n.p.}) = F^{n-2}$ with its given framing. If $F$ is connected then we can arrange that all fibers of $G$ are connected. Thus we can realize any homotopy class of maps to $S^2$ by an indefinite Morse 2–function with connected fibers. Furthermore, given two homotopic indefinite Morse 2–functions $G_0, G_1\colon X \to S^2$, if $n \geq 4$ then there exists an indefinite generic homotopy $G_s\colon X \to S^2$ connecting $G_0$ to $G_1$. Again, if neither $G_0$ or $G_1$ have any disconnected fibers, then the same is true for all $G_s$.*

We think of these results as existence and uniqueness theorems for Morse 2–functions, since the existence of an indefinite generic homotopy between two indefinite Morse 2–functions tells us that the two Morse 2–functions can be connected by a sequence of indefinite Morse 2–functions each related to the next by one of a finite collection of elementary moves.

The existence part of Theorem 1.2 is originally due to Saeki [13]. The uniqueness part of Theorem 1.2, when $n = 4$, was originally proved by Williams [15], without the assertion about connectedness of fibers. In section 5 we will also state versions of these theorems paying attention to sections, which is an important issue if one wishes to relate fibration-type structures to pencil-type structures. Because our results start with mapping to the disk, our techniques will naturally extend to surfaces other than $D^2$ and $S^2$, although Saeki [13] does exhibit obstructions to existence of indefinite Morse 2–functions over non-simply connected surfaces, so that there do need to be some hypotheses for existence.

If we remove the adjective "indefinite" from the above theorems (and the assertions about connectedness of fibers), the theorems become simply the facts that Morse functions, Morse 2–functions, and generic homotopies between them are, in fact, generic. Although in the above discussion we simply stated that Morse functions, Morse 2–functions and the homotopies we are calling "generic" are actually generic, in fact the definitions we prefer are in terms of local models, and the fact that maps and homotopies with these local models are generic (and in fact stable) is a standard result in singularity theory.

To prove Theorem 1.1, we spend most of our time proving a variant of the theorem, expressed in terms of Morse 2–functions over the square $I^2 = I \times I$. Here the natural structure on the $n$–dimensional domain $X$ of a Morse 2–function $G\colon X \to I^2$ is that of a cobordism with sides from $M_0$ to $M_1$, where $M_0$ is an $(n-1)$–dimensional cobordism from $F_{00}$ to $F_{01}$ and $M_1$ is an $(n-1)$–dimensional cobordism from $F_{10}$ to $F_{11}$, with $F_{00} \cong F_{10}$ and $F_{01} \cong F_{11}$. We ask that this cobordism structure should be mapped to the cobordism structure on $I^2$ as a cobordism from $I$ to $I$, and the boundary data comes in the form of $I$–valued Morse functions on $M_0$ and $M_1$.

Then the idea is to separate out the nontrivial topology of $X$ (encoded in an indefinite Morse function on $X$) from the problem of interpolating between two Morse functions on an $(n-1)$–dimensional submanifold of $X$, and then reduce to Cerf theoretic statements. In particular, we (re)prove the facts that: (1) connected cobordisms between nonempty manifolds support indefinite $I$–valued Morse functions and, (2) two such indefinite $I$–valued Morse functions on the same cobordism are homotopic through an indefinite generic homotopy. The first of those facts is completely standard, and the second is a key part of [9]. Then we go one step further to prove that two indefinite generic homotopies on the same cobordism between the same two indefinite $I$–valued Morse functions are homotopic through an indefinite generic homotopy of homotopies. In other words, we prove the key Cerf theory result of [9] with two parameters instead of just one. We also need to add a little extra detail to the first two cases regarding the issue of arranging for certain framed submanifolds (attaching maps for handles) to lie in level sets of the Morse functions involved.

Here is the outline of the rest of the paper: First we will give a selection of basic definitions, and standard results from singularity theory stated in our context and in terms of these definitions. We will then show how to extend the ideas from [9] to prove the Cerf theoretic results about Morse functions. Then we will collect the extra ingredients needed to prove results about $I^2$–valued Morse 2–functions. Finally the proofs of the main theorems become more or less immediate.

**Remark 1.3** Here we outline some of the motivating history behind these ideas and some of the terminology that has been used. Auroux, Donaldson and Katzarkov [3] proved that 4–manifolds equipped with near-symplectic forms could, after blowing up, be presented as what they called *singular Lefschetz fibrations* over $S^2$. Here we have Lefschetz-type singularities as well as indefinite folds with images parallel to the equator in $S^2$. Such fibrations became known as *broken Lefschetz fibrations* in [6], which gave a purely topological proof that all 4–manifolds supported *achiral* broken Lefschetz fibrations. The achirality condition was shown to be unnecessary in [5], [10] and [1]. In particular, both [5] and [10] related broken Lefschetz fibrations to generic smooth maps to dimension 2, and Lekili [10] introduced the notion of a *purely wrinkled fibration* for what we would call an $S^2$–valued indefinite Morse 2–function on a 4–manifold, and showed how to move between broken Lefschetz fibrations and purely wrinkled fibrations. This set up the problem (the uniqueness problem) of showing that two homotopic purely wrinkled fibrations could be connected by a generic homotopy staying in the purely wrinkled setting, i.e. remaining indefinite.

Perutz [11, 12] set up a Floer-theoretic framework in which one might produce invariants from broken Lefschetz fibrations (which he called, simply, *broken fibrations*). This motivated the uniqueness problem posed by Lekili, since the uniqueness result is needed to show that Perutz's Lagrangian matching invariant is actually a smooth invariant. Our uniqueness result was first proved by Williams [15], but without making assertions about connectedness of fibers, and it turns out that disconnected fibers, especially nullhomologous $S^2$ components of fibers, cause analytical problems in Perutz's framework.

Of course there is an extensive history behind this paper in the world of singularity theory, which is too long to present, and an extensive history in complex algebraic geometry in the study of honest Lefschetz fibrations. A purely topological precedent lies in the study of round handles; see [2] and [4], for example.

**Remark 1.4** *Advice to the reader:* It might be easiest to read this paper on a first pass ignoring the issues related to maintaining connected fibers. In many of the proofs, this issue is dealt with towards the end of the proof as an addendum, and often adds a level of technical detail that might be more confusing than necessary.

**Remark 1.5** Unless otherwise stated, all manifolds are oriented and all diffeomorphisms and local diffeomorphisms preserve orientations.

## 2 Definitions and basic results

We begin with the usual definition of Morse functions in terms of local models, as a warm-up to the succeeding definitions, and also add a few slightly less standard terms to this setting.

**Definition 2.1** The *standard index $k$ Morse model in dimension $m$* is the function $\mu_k^m(x_1, \ldots, x_m) = -x_1^2 - \ldots - x_k^2 + x_{k+1}^2 + \ldots + x_m^2$. When the ambient dimension $m$ is understood we will write $\mu_k$ instead of $\mu_k^m$. We will also abbreviate $\mu_k(x_1, \ldots, x_m)$ as $\mu_k(\mathbf{x})$.

**Definition 2.2** Given an $m$–manifold $M$ and a 1–manifold $N$, a smooth function $g\colon M \to N$ is *locally Morse* if there exist coordinates in a neighborhood of each critical point $p$ together with coordinates in a neighborhood of $g(p)$ with respect to which $g(x_1, \ldots, x_m) = \mu_k(\mathbf{x})$, where $k$ is the *index* of $p$. A *Morse function* is a proper map $g\colon M \to N$ which is locally Morse with the additional property that distinct critical points map to distinct critical values. When $g$ is a Morse function from $M$ to $I = [0, 1]$ we imply that $M$ is given as a cobordism from $F_0$ to $F_1$ and that $g^{-1}(0) = F_0$ and $g^{-1}(1) = F_1$.

It is a standard fact that Morse functions are stable and generic. Next we will discuss homotopies and homotopies of homotopies between Morse functions, and also make similar statements that homotopies satisfying certain properties are stable and generic. These facts are only slightly less standard, and are discussed in many different references on singularity theory and Cerf theory. Probably the most comprehensive reference for the facts we mention is [8]. To see these results in the more general context of singularity theory, look at [14]. For a modern exposition explicitly in a low dimensional setting, which also explains much of the motivation for this paper, we recommend [10].

We want to discuss homotopies $g_t\colon M \to N$ between Morse functions which are not necessarily Morse at intermediate times, in which case it is useful to discuss also the associated function $G\colon I \times M \to I \times N$ defined by $G(t,p) = (t, g_t(p))$, and its singular locus $Z_G$, the trajectory of the critical points of $g_t$. Given an $m$–manifold $M$ and a 1–manifold $N$ with two Morse functions $g_0, g_1\colon M \to N$, we are interested in homotopies $g_t\colon M \to N$ satisfying the following properties: The functions $g_t$ should be Morse for all but finitely many values of $t$ and, at those values $t_*$ when $g_{t_*}$ is not Morse exactly one of the following events should occur, possibly with the $t$ parameter reversed (Figure 1 illustrates these by drawing the Cerf graphic $G(Z_G)$ for a typical generic homotopy):

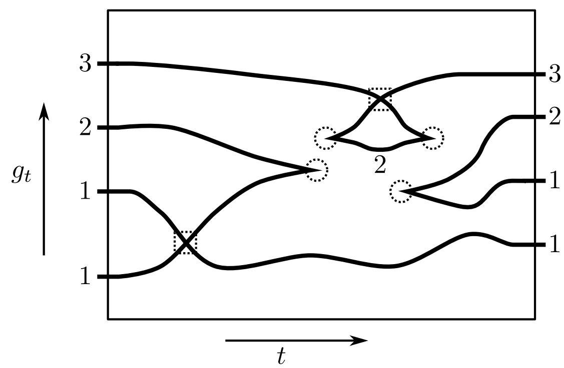


Figure 1: Example of the Cerf graphic for a generic homotopy between Morse functions, indices of critical points labelled with integers, birth and death cusps indicated by dotted circles and critical value crossings indicated by dotted squares.

(1) Two critical values cross at $t_*$: More precisely, $g_{t_*}$ is locally Morse but not Morse, and $Z_G \cap ([t_* - \epsilon, t_* + \epsilon] \times M)$ is a collection of arcs on which $G$ is an embedding except for exactly one transverse double point where the images of two arcs cross. For future reference we call this event a *1–parameter crossing*, or just a *crossing*.

(2) A pair of cancelling critical points are born: For all $t \in [t_* - \epsilon, t_* + \epsilon]$, $g_t$ is Morse outside a ball, and inside that ball there are coordinates on domain and range (possibly varying with $t$) with respect to which $g_t(x_1, \ldots, x_m) = -x_1^2 - \ldots - x_k + x_{k+1}^3 - (t - t_*)x_{k+1} + x_{k+2}^2 + \ldots x_m^2$, with no other critical values near 0. Thus for $t \neq t_*$, $g_t$ is Morse, but for $t < t_*$ there are no critical points in this ball, and for $t > t_*$ there are two critical points of index $k$ and $k+1$ in this ball. Note that here $G$ is injective on $Z_G \cap ([t_* - \epsilon, t_* + \epsilon] \times M)$, and $Z_G \cap ([t_* - \epsilon, t_* + \epsilon] \times M)$ is a collection of arcs all but one of which have end points at $t_* - \epsilon$ and $t_* + \epsilon$

and are smoothly embedded via $G$, and one of which has both end points at $t_*+\epsilon$ and is mapped via $G$ to a semicubical cusp in $[t_*-\epsilon, t_*+\epsilon]\times N$. For future reference we call this a 1*–parameter birth singularity* (or *death singularity* when $t$ is reversed).

It is a standard fact that homotopies satisfying these properties are stable and generic, so for this reason:

**Definition 2.3** We call a homotopy $g_t\colon M\to N$, with $g_0$ and $g_1$ Morse, a *generic homotopy between Morse functions* if $g_t$ satisfies the properties listed in the preceding paragraph.

We distinguish the above from the following:

**Definition 2.4** An *arc of Morse functions* is a homotopy $g_t$ which is Morse for all $t$.

Next we discuss homotopies $g_{s,t}\colon M\to N$ between generic homotopies between Morse functions, which are not necessarily generic homotopies for certain fixed values of $s$. In this case it is useful to consider the associated functions $G_s\colon I\times M\to I\times N$ defined by $(t,p)\mapsto(t,g_{s,t}(p))$ and $\mathcal{G}\colon I\times I\times M\to I\times I\times N$ defined by $(s,t,p)\mapsto(s,t,g_{s,t}(p))$, and their singular loci $Z_{G_s}\subset I\times M$ and $Z_{\mathcal{G}}\subset I\times I\times M$. Given an $m$–manifold $M$ and a 1–manifold $N$, with one generic homotopy $g_{0,t}\colon M\to N$ between Morse functions $g_{0,0}$ and $g_{0,1}$ and another generic homotopy $g_{1,t}\colon M\to N$ between Morse functions $g_{1,0}$ and $g_{1,1}$, we are interested in connecting these through a 2–parameter family $g_{s,t}\colon M\to N$, with $s,t\in I$, satisfying the following conditions:

(1) $g_{s,0}$ is an arc of Morse functions from $g_{0,0}$ to $g_{1,0}$ and $g_{s,1}$ is an arc of Morse functions from $g_{0,1}$ to $g_{1,1}$.

(2) For all but finitely many fixed values of $s$, $g_{s,t}$ is, in the parameter $t$, a generic homotopy between the Morse functions $g_{s,0}$ and $g_{s,1}$.

(3) At those values $s_*$ when $g_{s_*,t}$ is not a generic homotopy there is a single value $t_*$ such that $g_{s_*,t}$ is a generic homotopy for $t\in[0,t_*)$ and for $t\in(t_*,1]$.

(4) At each of these points $(s_*,t_*)\in I\times I$ exactly one of the following events occurs, possibly with either the $s$ or $t$ parameter reversed (some of which are illustrated in figures below by drawing sequences of Cerf graphics $G_s(Z_{G_s})$):

(a) (This event is not particularly important to us but we list it for completeness.) The function $g_{s_*,t_*}$ is locally Morse but the 1–parameter family $g_{s_*,t}$ does not meet the requirements to be a generic homotopy because exactly two of the events listed in Definition 2.3 occur simultaneously at $t = t_*$. For example, a birth singularity may happen at the same time $t_*$ as a crossing. This phenomenon should be transverse in the obvious sense; for example, for $s < s_*$, the birth might happen before the crossing, and for $s > s_*$, the birth would then happen after the crossing. We call this event a *2–parameter coincidence*.

(b) The function $g_{s_*,t_*}$ is locally Morse but the 1–parameter family $g_{s_*,t}$ does not meet the requirements to be a generic homotopy because the singular locus $Z_{G_{s_*}} \cap ([t_* - \epsilon, t_* + \epsilon] \times M)$ is mapped into $I \times N$ via $G_{s_*}$ with a single non-transverse quadratic double point at $t = t_*$. However, we require here that the singular locus $Z_{\mathcal{G}} \cap ([t_* - \epsilon, t_* + \epsilon] \times [s_* - \epsilon, s_* + \epsilon] \times M)$ is a collection of disjoint squares and is mapped into $I \times I \times N$ via $\mathcal{G}$ with a single arc of transverse double points. In other words, the image of $Z_{G_s}$ in $I \times N$ changes via a Reidemeister-II type move at $s = s_*$. See Figure 2; we call this event a *Reidemeister-II fold crossing*.

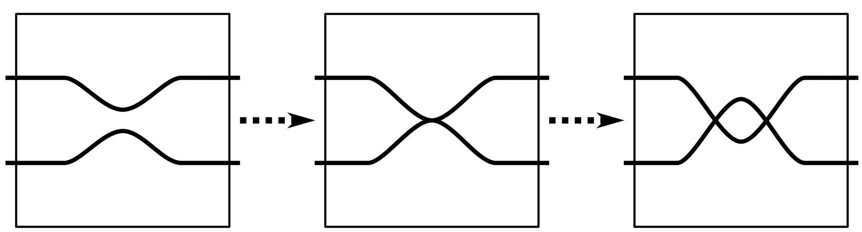

Figure 2: Non-transverse double point in the singular locus for a generic homotopy between generic homotopies between Morse functions. Note that in general the indices of the two critical points involved can be arbitrary.

(c) The function $g_{s_*,t_*}$ is locally Morse but the 1–parameter family $g_{s_*,t}$ does not meet the requirements to be a generic homotopy because the singular locus $Z_{G_{s_*}} \cap ([t_* - \epsilon, t_* + \epsilon] \times M)$ is mapped into $I \times N$ via $G_{s_*}$ with a single transverse triple point. However, we require here that the singular locus $Z_{\mathcal{G}} \cap ([t_* - \epsilon, t_* + \epsilon] \times [s_* - \epsilon, s_* + \epsilon] \times M)$ is a collection of disjoint squares and is mapped into $I \times I \times N$ via $\mathcal{G}$ with three arcs of double points which meet transversely at the triple point. In other words, the image of $Z_{G_s}$ in $I \times N$ is modified via a Reidemeister-III type move. See Figure 3; we call this event a *Reidemeister-III fold crossing.*

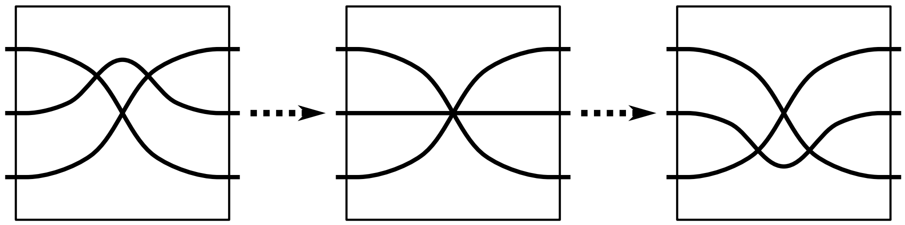

Figure 3: Transverse triple point in the singular locus for a generic homotopy between generic homotopies between Morse functions. Again, the indices involved can be arbitrary.

(d) The 1–parameter family $g_{s_*,t}$ fails to be a generic homotopy because a birth (or death) occurs at time $t_*$ at a point $p \in M$ at the same value as another Morse critical point $q$; $g_{s_*,t_*}(p) = g_{s_*,t_*}(q)$. In other words, $G_{s_*}$ maps $Z_{G_{s_*}}$ into $I \times M$ in such a way that a non-transverse double point occurs between a cusp and a non-cusp point. However, here we require that the 1–dimensional cusp locus $C_\mathcal{G} \subset I \times I \times M$ and the 2–dimensional singular locus $Z_\mathcal{G} \subset I \times I \times M$ are mapped into $I \times I \times N$ via $\mathcal{G}$ with a transverse intersection at this point. See Figure 4; we call this event a *cusp-fold crossing.*

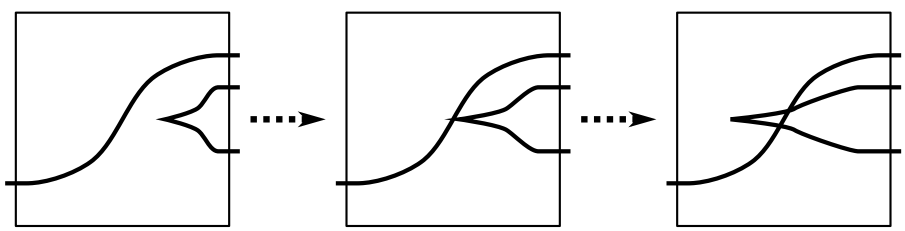

Figure 4: Non-transverse double point involving a cusp, occurring in a generic homotopy between generic homotopies between Morse functions. The only constraint on indices is that coming from the cusp, namely that the two critical points born at the cusp are of successive index.

(e) The function $g_{s_*,t_*}$ is Morse away from a point $p \in M$, and in neighborhoods of $p$ and $g_{s_*,t_*}(p)$ we have coordinates with respect to which, for $|s - s_*| < \epsilon$ and $|t - t_*| < \epsilon$, $g_{s,t}$ is given by $g_{s,t}(x_1, \ldots, x_m) = -x_1^2 - \ldots - x_k^2 + x_{k+1}^3 + (t - t_*)^2 x_{k+1} - (s -$

$s_*)x_{k+1} + x_{k+2}^2 + \ldots + x_m^2$. Furthermore, for these $(s,t)$ there are no other singularities of $g_{s,t}$ in the inverse image of a small neighborhood of $g_{s_*,t_*}(p)$. Geometrically, this is the birth of a pair of cusps joined in an "eye" shape, involving a birth and a death of a pair of cancelling critical points. See Figure 5; we call this event an *eye birth singularity* (or *death* when $s$ is reversed).

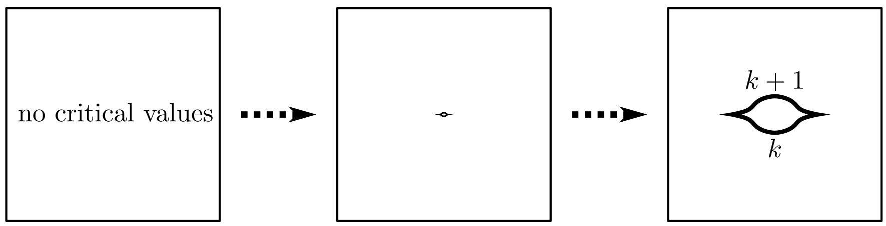


Figure 5: The birth of a birth–death pair involving a pair of cancelling critical points of successive index.

(f) The function $g_{s_*,t_*}$ is Morse away from a point $p \in M$, and in neighborhoods of $p$ and $g_{s_*,t_*}(p)$ we have coordinates with respect to which, for $|s - s_*| < \epsilon$ and $|t - t_*| < \epsilon$, $g_{s,t}$ is given by $g_{s,t}(x_1, \ldots, x_m) = -x_1^2 - \ldots - x_k^2 + x_{k+1}^3 - (t - t_*)^2 x_{k+1} - (s - s_*)x_{k+1} + x_{k+2}^2 + \ldots + x_m^2$. Furthermore, for these $(s,t)$ there are no other singularities of $g_{s,t}$ in the inverse image of a small neighborhood of $g_{s_*,t_*}(p)$. Here a death and a birth of a cancelling pair merge together, so that afterwards there is no cancellation. See Figure 6; we call this event a *merge singularity* (or *unmerge* when $s$ is reversed).

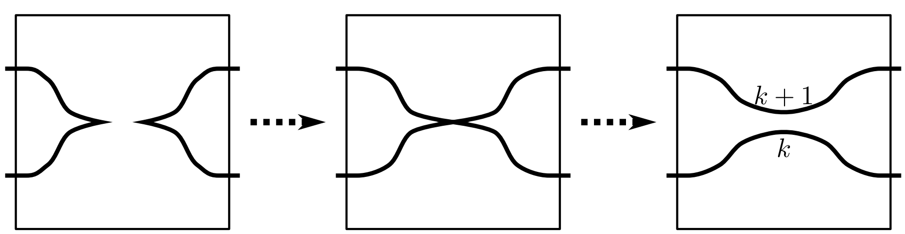


Figure 6: The merge of a death–birth pair involving a pair of cancelling critical points of successive index.

(g) The function $g_{s_*,t_*}$ is Morse away from a point $p \in M$, and in neighborhoods of $p$ and $g_{s_*,t_*}(p)$ we have coordinates with respect to which, for $|s - s_*| < \epsilon$ and $|t - t_*| < \epsilon$, $g_{s,t}$ is given by $g_{s,t}(x_1, \ldots, x_m) = -x_1^2 - \ldots - x_k^2 + x_{k+1}^4 + (t - t_*)x_{k+1}^2 - (s - s_*)x_{k+1} + x_{k+2}^2 + \ldots + x_m^2$. Furthermore, for these $(s,t)$ there are no other singularities of $g_{s,t}$ in the inverse image of a small neighborhood of $g_{s_*,t_*}(p)$. This singularity is known as a swallowtail. See Figure 7; we call this event a *swallowtail birth singularity* (or *death* when $s$ is reversed).

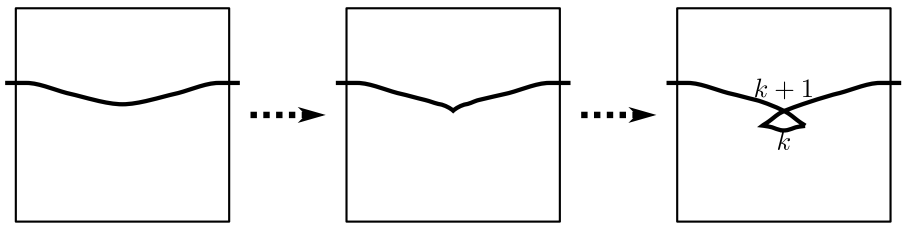


Figure 7: Birth of a swallowtail.

Note that, besides the coincidence event, we have two types of events: 2–*parameter crossings* (Reidemeister-II's, Reidemeister-III's and cusp-fold's) and 2–*parameter singularities* (eye births and deaths, merges and unmerges, and swallowtail births and deaths). (As a technical point, note also that, in the definitions of the 2–parameter singularities, the coordinates in which the homotopy of homotopies takes on the standard models may vary with $s$ and $t$, and also the parametrization of $t$ may depend on $s$.)

It is also standard that such homotopies of homotopies are generic and stable, and so for this reason:

**Definition 2.5** A homotopy $g_{s,t}$ between generic homotopies $g_{0,t}$ and $g_{1,t}$ is a *generic homotopy of homotopies* if it satisfies the properties described above. If $g_{0,0} = g_{1,0}$ and $g_{0,1} = g_{1,1}$, we can also ask that $g_{s,0} = g_{0,0}$ and $g_{s,1} = g_{0,1}$ for all $s$, in which case we say that $g_{s,t}$ is a generic homotopy *with fixed endpoints*.

Again, we distinguish this from the following:

**Definition 2.6** An *arc of generic homotopies* is a homotopy of homotopies $g_{s,t}$ which, for each fixed value of $s$, is a generic homotopy in the parameter $t$.

**Definition 2.7** Given an $n$–manifold $X$ and a 2–manifold $\Sigma$, a smooth proper map $G\colon X \to \Sigma$ is a *Morse 2–function* if for each $q \in \Sigma$ there is a compact neighborhood $S$ of $q$ with a diffeomorphism $\psi\colon S \to I\times I$ and a diffeomorphism $\phi\colon G^{-1}(S) \to I\times M$, for an $(n-1)$–manifold $M$, such that $\psi\circ G\circ\phi^{-1}\colon I\times M \to I \times I$ is of the form $(t,p) \mapsto (t, g_t(p))$ for some generic homotopy between Morse functions $g_t\colon M \to I$. A singular point for $G$ is called a *fold point* if the homotopy used to model $G$ at that point can actually be taken to be Morse, and is called a *cusp point* if the homotopy has a birth or death at that point. An arc of fold points is called a *fold*. When $\Sigma$ is given as a cobordism between 1–manifolds $N_0$ and $N_1$ then $X$ should be given as a cobordism between $(n-1)$–manifolds $M_0$ and $M_1$, with $G^{-1}(N_i) = M_i$ and with $G|_{M_i}\colon M_i \to N_i$ a Morse function. When $\Sigma$ is given as a cobordism between cobordisms (in particular, when $\Sigma = I^2$, a cobordism from $I$ to $I$, with $I$ being a cobordism from $\{0\}$ to $\{1\}$), then $X$ should also be given as such a relative cobordism, with all the cobordism structure preserved by $G$. For us the structure of a relative cobordism includes an explicit product structure on the sides, and this should also be respected by $G$. In particular, there should be no critical points along the side of the cobordism.

**Remark 2.8** The important thing to understand here is that Morse 2–functions look locally like generic homotopies between Morse functions, but that there is no global time direction. Note that the *index* of a fold is not well defined, but that if we choose a transverse direction to the fold, and consider local models $(t,p) \mapsto (t, g_t(p))$ in which the second coordinate in the range is given by this transverse direction, then we do have a well defined index. In figures, we will indicate this by drawing a small arrow transverse to the fold and labelling it with the index. If, however, we are drawing a Cerf graphic, then it is understood that the transverse direction is up, and we will label folds (arcs of critical points) with indices without indicating the arrow. We illustrate these conventions in Figure 8, which show the images of the singular locus for, on the left, a hypothetical Morse 2–function mapping to a genus 2 surface and, on the right, a generic homotopy between Morse functions.

**Definition 2.9** A 1–parameter family $G_s\colon X \to \Sigma$ is a *generic homotopy between Morse 2–functions* if, for each $q \in \Sigma$ and each $s_* \in I$ there is an $\epsilon > 0$ and a compact neighborhood $S$ of $q$ with a diffeomorphism $\psi\colon S \to I \times I$ and a 1–parameter family of diffeomorphisms $\phi_s\colon G_s^{-1}(S) \to I \times M$, for an $(n-1)$–manifold $M$ and for $|s - s_*| < \epsilon$, such that $\psi\circ G_s\circ\phi_s^{-1}\colon I\times M \to I\times I$ is of the form $(t,p) \mapsto (t, g_{s,t}(p))$ for some generic homotopy of homotopies $g_{s,t}\colon M \to I$. Generic homotopies of Morse 2–functions $G_s\colon X \to \Sigma$ are expected to be constant (independent of $s$) on $\partial X$.

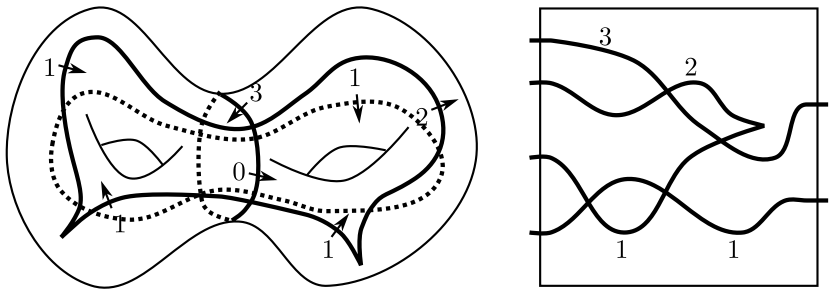


Figure 8: Morse 2–functions versus Cerf graphics, and index labelling conventions. On the left, we are mapping from a 4–manifold to a genus 2 surface. On the right, we are illustrating a generic homotopy between $I$–valued Morse functions on a 3–manifold.

Again, although our terminology is not standard, it is a standard fact that Morse 2–functions and generic homotopies of Morse 2–functions are stable and generic. This is mostly explained in Section 4 and Appendix A of [10].

We will be interested, for most of this paper, in the special case of Morse 2–functions mapping to $I^2$, seen as a cobordism from $\{0\} \times I$ to $\{1\} \times I$. We use coordinates $(t, z)$ on $I^2$, i.e. $t$ is the horizontal axis. Here it is useful to impose one extra genericity condition:

**Definition 2.10** Suppose $X^n$ is a cobordism from $M_0$ to $M_1$, with each $M_i$ a cobordism from $F_{i0}$ to $F_{i1}$. A *square Morse 2–function* on $X$ is a Morse 2–function $G\colon X \to I^2$, respecting the cobordism structure, such that the projection onto the horizontal axis, $t \circ G\colon X \to I$, is itself a Morse function.

It is not hard to see that, amongst Morse 2–functions mapping to $I^2$ respecting the cobordism structures on domain and range, square Morse 2–functions are generic and stable.

**Definition 2.11** A Morse function is *indefinite* if there are no critical points of minimal or maximal index, i.e. no critical points with the local model $(x_1, \ldots, x_m) \mapsto \pm(x_1^2 + \ldots x_m^2)$. A generic homotopy, or generic homotopy of homotopies, of Morse functions is *indefinite* if it is indefinite at all parameter values at which it is Morse. A Morse 2–function, resp. generic homotopy of Morse 2–functions, is *indefinite* if it can always be locally modelled, as in the definition, by an indefinite generic homotopy, resp. generic homotopy of homotopies.

The following definition will be useful when we want to make assertions about the connectedness of fibers:

**Definition 2.12** A Morse function $g\colon M \to I$ is *ordered* if, given two critical points $p, q \in M$ with indices $i, j$, respectively, if $i < j$ then $g(p) < g(q)$. A generic homotopy or generic homotopy of homotopies is ordered if it is ordered at all parameter values at which it is Morse. A Morse function (or homotopy or homotopy of homotopies) is *almost ordered* if, whenever $i < j+1$, we have $g(p) < g(q)$.

(Note that it is not clear how to generalize this definition to Morse 2–functions.) We leave the proof of the following observation to the reader:

**Lemma 2.13** *Consider a Morse function $g\colon M^m \to I$, with $M$ a cobordism from $F_0$ to $F_1$, and with $F_0$ and $F_1$ both connected. If $m \geq 3$ and $g$ is indefinite and ordered then all level sets of $g$ will be connected. If $m \geq 4$ and $g$ is indefinite and almost ordered then the level sets will all be connected.*

# 3 Theorems about $I$–valued Morse functions on cobordisms

Throughout this section, we are given the following data:

(1) A connected $m$–dimensional cobordism $M$ from $F_0 \neq \emptyset$ to $F_1 \neq \emptyset$. We also assume $m \geq 2$ to avoid very-low-dimensional confusion.

(2) A collection $L_1, \ldots, L_p$ (possibly empty) of closed manifolds with $\dim(L_i) = l_i < m/2$ and with mutually disjoint embeddings $\phi_i\colon [-\epsilon, \epsilon] \times B^{m-1-l_i} \times L_i \hookrightarrow (M \setminus \partial M)$, for some small $\epsilon > 0$. Note that if $l_i < m/2$ then $l_i < m-1$.

(3) A collection of values $z_1, \ldots, z_p \in (0,1)$.

In the results that follow we will say that a Morse function $g\colon M \to I$ *is standard with respect to $\phi_i$ at height $z_i$* if $g \circ \phi_i\colon [-\epsilon, \epsilon] \times B^{m-1-l_i} \times L_i \to I$ is of the form $(z, x, p) \mapsto z + z_i$ on some neighborhood of $\{0\} \times \{0\} \times L_i$. The most illuminating example to bear in mind is when $m = 3$ and each $l_i = 1$, and we think of $L_1 \cup \ldots \cup L_p$ as a link in $M$ and of each embedding $\phi_i$ as given by a framing of $L_i$. Then we are interested in Morse functions with $L_i$ in the level set at level $z_i$, with framing coming from a framing in this level set. More generally the $\phi_i$'s are going to be attaching maps for handles, and we will see in the next section the importance of attaching handles along spheres lying in level sets of a Morse function, with tubular neighborhoods interacting well with the Morse function.

**Remark 3.1** In the proofs of the following theorems, we will use generic gradient-like vector fields, and in particular the ascending and descending manifolds of critical points, as a tool to organize local modifications of Morse functions, such as cancellation of critical points. Recall that a gradient-like vector field is a vector field which is transverse to level sets and such that, near each critical point, there are local coordinates with respect to which the Morse function takes the usual form $-x_1^2 - \ldots - x_k^2 + x_{k+1}^2 + \ldots + x_m^2 = \mu_k^m(\mathbf{x})$ and the vector field is the usual Euclidean gradient of this function. For a fixed Morse function $g$, a generic gradient-like vector field is one for which the ascending and descending manifolds meet transversely in intermediate level sets. For a generic homotopy $g_t$ between Morse functions, a generic 1–parameter family of gradient-like vector fields $V_t$ is one for which the 1–parameter families of ascending and descending manifolds intersected with intermediate level sets are transverse in the 1–parameter sense, and for a generic homotopy $g_{s,t}$ between generic homotopies we have the natural notion of a generic 2–parameter family of gradient-like vector fields $V_{s,t}$. (We should also require that, at the non-Morse singularities, the vector field is the usual Euclidean gradient for the standard model of the singularity in local coordinates.) It is clear from the fact that the transversality properties of the ascending and descending manifolds are generic that the associated "genericity" properties of the vector fields are actually generic.

**Theorem 3.2** *There exists an indefinite ordered Morse function $g\colon M \to I$, with critical values not in $\{z_1, \ldots, z_p\}$, which is standard with respect to each of $\phi_1, \ldots, \phi_p$, at heights $z_1, \ldots, z_p$ respectively.*

In the following proof, we will make essential use of the standard lemma that, for a Morse function $g$ with a generic gradient-like vector field, if there is a single gradient flow line from a critical point $q$ of index $k+1$ down to a critical point $p$ of index $k$, then the two critical points can be cancelled. More precisely, there exists a generic homotopy $g_t$, with $g_0 = g$, with exactly one death singularity at $g_{1/2}$ involving $p$ and $q$, and no other birth or death singularities. Also note that no other critical values need to move if there is a regular level set between $p$ and $q$ such that the descending manifold for $q$ and the ascending manifold for $p$ avoid all other critical points on their way to this level set. If this is not the case then other critical points may need to move "out of the way" to facilitate the crossing. See Figure 9 for an illustration of these ideas. To be even more precise, we make the following definitions:

**Definition 3.3** Given a critical point $q$ for the Morse function $g\colon M \to I$ with a gradient-like vector field (not necessarily generic), and any value $z < g(q)$, the

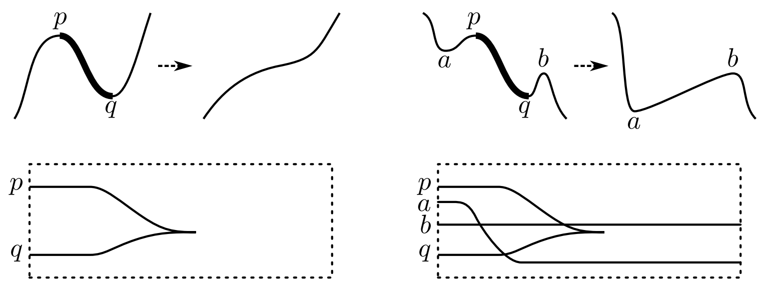


Figure 9: Cancellation of critical points $p$ and $q$ in one dimension, with accompanying Cerf graphics. In the first example, no other critical points need to move, but in the second example the critical point $a$ needs to drop below $p$ before $q$ and $p$ can cancel.

*descending complex for* $q$ *down to height* $z$ is the union of descending manifolds in $[z, g(q)]$ for all critical points $q' \in g^{-1}[z, q]$ which are connected to $q$ by a chain of descending gradient flow lines, including, of course, the descending manifold for $q$ itself.

**Definition 3.4** Given a Morse function $g\colon M \to I$ with a gradient-like vector field and critical points $q$ of index $k+1$ and $p$ of index $k$, with $g(q) > g(p)$, we say that $q$ *cancels* $p$ if there is a unique gradient flow line from $q$ to $p$.

Then we can summarize the standard cancellation lemma (without proof) as follows:

**Lemma 3.5** *Given $g\colon M \to I$ with a generic gradient-like vector field, if critical point $q$ cancels critical point $p$, let $C$ be the descending complex for $q$ down to height $g(p)$. Note that $p \in C$. Then there is a generic homotopy $g_t$ between Morse functions, with $g_0 = g$, which is independent of $t$ outside an arbitrarily small neighborhood of $C$, passes through exactly one death singularity at $g_{1/2}$ in which $q$ and $p$ cancel, and has no other birth or death singularities. If the descending complex for $q$ is actually just the descending disk, then no other critical values will change; otherwise all the critical points in this descending complex besides $q$ and $p$ will have to move below $p$ before the death occurs.*

**Proof of Theorem 3.2** First define $g\colon M \to I$ on the submanifolds $\phi_i([-\epsilon, \epsilon] \times B^{m-1-l_i} \times L_i)$ by $\phi_i^{-1}$ followed by projection to $[-\epsilon, \epsilon]$ followed by translation by $z_i$. This places $\phi_i(L_i)$ at height $z_i$ as desired.

There is no obstruction to extending this map to a Morse function $g\colon M \to I$. (Issues of smoothness at the boundary of $\phi_i([-\epsilon, \epsilon]) \times B^{m-1-l_i} \times L_i)$ are easily avoided either by use of tubular neighborhood theorems or by slightly shrinking the image of $\phi_i$.) To make $g$ indefinite and ordered, first choose a generic gradient-like vector field, so that we can construct arguments using gradient flow lines and ascending and descending manifolds. Indefiniteness is easily achieved because each critical point of index 0 (or $m$) must be cancelled by a critical point of index 1 (or $m-1$), since $M$ is connected. (If, for an index 0 critical point $p$, there was no such cancelling index 1 critical point, then there could not be a path from $p$ to $F_0$, and $M$ would not be connected.) The descending complex $C$ for an index 1 critical point is 1–dimensional, and will thus miss a neighborhood of the $\phi_i(L_i)$'s by genericity of the gradient-like vector field. Here we use the fact that $\dim(L_i) = l_i < m/2 \leq m-1$, so that each $\phi_i(L_i)$ has positive codimension in the level set $g^{-1}(z_i)$. Thus we can cancel all the index 0 and $m$ critical points without modifying $g$ near the $\phi_i(L_i)$'s.

To arrange that $g$ is ordered, suppose that $p$ and $q$ are critical points of index $j$ and $k$, respectively, with $g(p) < g(q)$, and with no critical values in $(g(p), g(q))$. If the descending manifold $D_q$ for $q$ and the ascending manifold $A_p$ for $p$ are disjoint then there is a generic homotopy supported in a neighborhood of $D_q \cap g^{-1}[g(p), g(q)]$ which lowers $g(q)$ below $g(p)$ (without creating any new critical points). In this case there is also a generic homotopy supported in a neighborhood of $A_p \cap g^{-1}[g(p), g(q)]$ which raises $g(p)$ above $g(q)$. If $j \geq k$ this disjointness can always be arranged by a generic choice of gradient-like vector field, and thus we can get $g$ to be ordered, using either raising or lowering homotopies each time we need to switch the relative order of two critical points. However, if we are not careful, we may mess up the behavior of $g$ near $\phi_i(L_i)$. To avoid this, we need to make sure that either $D_q$ or $A_p$ misses each $\phi_i(L_i)$, which can be done by generically choosing the gradient-like vector field, as long as $\dim(L_i) = l_i < m/2$. (To see this, count dimensions in the level set $F$ containing $\phi_i(L_i)$ and note that we are asking for either $l_i < j$ or $l_i + k < m$; if $l_i \geq j \geq k$ then $l_i + k \leq 2l_i < m$.) □

**Theorem 3.6** *Given two indefinite Morse functions $g_0, g_1 \colon M \to I$ there exists an indefinite generic homotopy $g_t \colon M \to I$ from $g_0$ to $g_1$. If $m \geq 3$ and both $g_0$ and $g_1$ are standard with respect to each $\phi_i$ at height $z_i$, then we can arrange that, for all $t$, $g_t$ is standard with respect to each $\phi_i$ at height $z_i$. If $g_0$ and $g_1$ are ordered then we can arrange that $g_t$ is ordered. (If both the ordering condition and the condition about respecting the $\phi_i$'s hold, with $m \geq 3$, then we can arrange that both stated conditions hold for $g_t$.) In all cases we can arrange that all the births occur before the critical point crossings and that all the deaths occur after the critical point crossings.*

The proof of this theorem (and the one to follow about homotopies of homotopies) is in essentially the same spirit as the proof of the preceding theorem; we just need the right cancellation lemmas to get rid of definite critical points over time and we need to count dimensions to see that we can order critical points appropriately and avoid the submanifolds $\phi_i(L_i)$ as we modify the homotopy.

For the cancellation lemmas, we need to articulate conditions under which we can pass through eye, unmerge or swallowtail singularities to simplify the homotopy. When discussing a generic homotopy $g_t\colon M \to I$, we will frequently use the Cerf graphic to organize our argument; recall that this is the image in $I \times I$ of the critical points of $g_t$ under the map $G\colon (t,p) \mapsto (t, g_t(p))$. We will use the term "$k$–fold" to refer to an arc of index $k$ critical points in $I \times M$. If we label a $k$–fold $P$, then for a fixed time $t$, $P_t$ will refer to the index $k$ critical point on $P$ at time $t$. We will also fix a generic 1–parameter family of gradient-like vector fields so that we may refer to gradient flow lines for each $g_t$. (Here "generic" means that the 1–parameter families of descending manifolds intersect transversely in level sets, which means that handle slides occur at isolated times, with lower index critical points never sliding over higher index critical points.) We will say that a $(k+1)$–fold $Q$ "cancels" a $k$–fold $P$ over a time interval $A$ if $Q_t$ cancels $P_t$ for every time $t \in A$.

In the following three lemmas, suppose that $g_t\colon M \to I$ is a generic homotopy between Morse functions $g_0$ and $g_1$ with a generic 1–parameter family of gradient-like vector fields. We leave the proofs to the reader; the basic idea, as with the proof of the standard cancellation lemma (Lemma 3.5), is as follows: We first push extraneous critical points on a descending complex down far enough so that we only need to work with a descending disk and an ascending disk. Then we use the uniqueness of a gradient flow line to find coordinates on a neighborhood of the gradient flow line in which the gradient flow line itself lies in one coordinate axis, and in the other coordinates the Morse function has the usual Morse model. Then the cancellation occurs entirely in the coordinate chart containing the gradient flow line.

**Lemma 3.7** *Consider a $k$–fold $P$ and a $(k+1)$–fold $Q$ over a time interval $[t_0, t_1]$ such that $Q$ cancels $P$ over all of $[t_0, t_1]$. Then for some arbitrarily small $\delta > 0$ there is a generic homotopy between homotopies $g_{s,t}$, with $g_{0,t} = g_t$, passing through a single unmerge singularity at $s = 1/2$ and no other 1–parameter singularities, independent of $s$ for $t \in [0, t_0] \cup [t_1, 1]$, such that, with respect to $g_{1,t}$, the cancelling pair $Q_t$ and $P_t$ die at $t = t_0 + \delta$ and are reborn at $t = t_1 - \delta$. For each $t \in [t_0, t_1]$, $g_{s,t}$ is independent of $s$ outside an arbitrarily small neighborhood of the descending complex for $Q_t$. Also note* *that, with respect to* $g_{1,t}$*,* $Q_t$ *still cancels* $P_t$ *on* $[t_0, t_0 + \delta)$ *and on* $(t_1 - \delta, t_1]$*. Furthermore we can arrange that any other folds that cancelled* $P$ *on* $[t_0, t_0 + \delta]$ *or* $[t_1 - \delta, t_1]$ *will still cancel* $P$ *there.*

**Lemma 3.8** *Consider a* $k$*–fold* $P$ *and a* $(k+1)$*–fold* $Q$ *over a time interval* $[t_0, t_1]$ *such that* $Q$ *cancels* $P$ *over all of* $(t_0, t_1)$*. Also suppose that the critical points* $Q_t$ *and* $P_t$ *are born as a cancelling pair at time* $t_0$ *and die as a cancelling pair at time* $t_1$*. Then for some small* $\delta > 0$ *there is a generic homotopy between homotopies* $g_{s,t}$*, with* $g_{0,t} = g_t$*, passing through a single eye death singularity at* $s = 1/2$ *and no other* $2$*–parameter singularities, independent of* $s$ *for* $t \in [0, t_0 - \delta] \cup [t_1 + \delta, 1]$*, such that, in* $g_{1,t}$*, the cancelling pair* $Q_t$ *and* $P_t$ *have cancelled for all* $t \in [t_0, t_1]$*. For each* $t \in (t_0, t_1)$*,* $g_{s,t}$ *is independent of* $s$ *outside an arbitrarily small neighborhood of the descending complex for* $Q_t$*. For* $t \in \{t_0, t_1\}$*,* $g_{s,t}$ *is independent of* $s$ *outside a neighborhood of the birth/death point.*

**Lemma 3.9** *Consider a* $k$*–fold* $P$ *and two* $(k+1)$*–folds* $Q$ *and* $R$ *over a time interval* $(t_0, t_1)$*. Suppose furthermore that* $Q$ *and* $P$ *are born as a cancelling pair at time* $t_0$*, that* $R$ *and* $P$ *die as a cancelling pair at time* $t_1$*, and that* $Q$ *cancels* $P$ *over* $(t_0, (t_0 + t_1)/2 + \delta)$ *while* $R$ *cancels* $P$ *over* $((t_0 + t_1)/2 - \delta, t_1)$*. Then there is a generic homotopy between homotopies,* $g_{s,t}$*, with* $g_{0,t} = g_t$*, passing through a single swallowtail death singularity at* $s = 1/2$ *and no other* $2$*–parameter singularities, independent of* $s$ *for* $t \in [0, t_0 - \delta] \cup [t_1 + \delta, 1]$ *such that, in* $g_{1,t}$*, the* $k$*–fold* $P$ *has disappeared and the* $(k+1)$*–folds* $Q$ *and* $R$ *have become the same fold. For each* $t \in (t_0, (t_0 + t_1)/2 - \delta)$*,* $g_{s,t}$ *is independent of* $s$ *outside an arbitrarily small neighborhood of the descending complex for* $Q_t$*, while for each* $t \in ((t_0 + t_1)/2 + \delta, t_1)$ *the independence is outside a neighborhood of the descending complex for* $R_t$*, and for* $t \in [(t_0 + t_1)/2 - \delta, (t_0 + t_1)/2 + \delta]$ *we need a neighborhood of the union of the descending complexes for* $Q_t$ *and* $R_t$*. For* $t \in \{t_0, t_1\}$ *the independence is outside a neighborhood of the birth/death points.*

**Proof of Theorem 3.6** As in the proof of Theorem 3.2, there is no difficulty in finding a generic homotopy $g_t \colon M \to I$, and if $g_0$ and $g_1$ are standard with respect to each $\phi_i$ at height $z_i$ then we can make $g_t$ independent of $t$ on these neighborhoods. Arranging for births to happen first and deaths last is standard, using connectedness; $g_t$ is modified via a generic homotopy between homotopies which passes through cusp-fold crossings, moving left-cusps (births) further left (earlier), and right-cusps (deaths) further right (later). However it requires some work to make $g_t$ indefinite and then ordered for all $t \in (0, 1)$.

We will cancel the 0–folds one at a time. Consider a 0–fold $P$ which is born at time $a$ and dies at time $b$. At each time $t \in [a,b]$, there is some index 1 critical point $q$ which cancels $P_t$. Thus there is a sequence $a = t_0 < t_1 < \ldots < t_n = b$ and some $\delta > 0$, giving a covering of $[a,b]$ by intervals $I_1 = [a, t_1 + \delta), I_2 = (t_1 - \delta, t_2 + \delta) \ldots I_n = (t_{n-1} - \delta, b]$, and a sequence of 1–folds $Q^1, \ldots, Q^n$ such that each $Q^i$ cancels $P$ over the interval $I_i$. We do this so that $Q^1$ is the 1–fold born with $P$ as a cancelling 0–1 pair at time $t_0 = a$ and so that $Q^n$ is the 1–fold which dies with $P$ as a cancelling pair at time $t_n = b$. Using the above three lemmas we can then cancel $P$ with the $Q^i$'s: First cancel over the non-overlapping parts of the open intervals $I_2, \ldots, I_{n-1}$ using Lemma 3.7. Then cancel over the overlaps and $I_1$ and $I_n$ using either Lemma 3.8 or Lemma 3.9, depending on whether the two cancelling 1–folds $Q^{i-1}$ and $Q^i$ in the overlap region $(t_i - \delta, t_i + \delta)$ are the same or different. Going from cancelling on the nonoverlapping regions to the overlapping regions requires the extra clauses in Lemma 3.7 to the effect that whatever cancelled $P$ at the beginning still cancels $P$ wherever it has not been killed with Lemma 3.7.

The above argument ignored the issue of the submanifolds $\phi_i(L_i)$. If $m \geq 3$ and both $g_0$ and $g_1$ are standard with respect to each $\phi_i$ at height $z_i$, then we want to arrange that, for all $t$, $g_t$ is standard with respect to each $\phi_i$ at height $z_i$. Since, before cancelling the definite folds, we had arranged for this property to hold, we need to arrange that, in each application of Lemmas 3.7, 3.8 and 3.9, we avoid neighborhoods of each $\phi_i(L_i)$. In other words, all the descending complexes for the cancelling 1–folds down to the level of the cancelled 0–folds should avoid $\phi_i(L_i)$. Counting dimensions we see that this can generically be acheived at all but finitely many times $t$, which are distinct from the times $t_0, t_1, \ldots, t_n$ at which we switch from one cancelling 1–fold to another. If the descending complex for a 1–fold $Q^j$ intersects some $\phi_i(L_i)$ at time $t_*$, with $t_{j-1} < t_* < t_j$, we break $Q^j$ into two 1–folds by introducing a 1–2 swallowtail (passing through a swallowtail singularity) at time $t_*$ along $Q^j$. This is illustrated in Figure 10; we label the two new 1–folds $Q^j_-$ and $Q^j_+$ as indicated in the figure, and observe that we can arrange for the descending complex for $Q^j_-$ to meet $\phi_i(L_i)$ at some time $t_- > t_*$ while the descending complex for $Q^j_+$ meets $\phi_i(L_i)$ at some time $t_+ < t_*$. Then we break the interval $(t_{j-1} - \delta, t_j + \delta)$ into two overlapping intervals $(t_{j-1} - \delta, t_* + \delta)$ and $(t_* - \delta, t_j + \delta)$ and replace the single cancelling 1–fold $Q^j$ with $Q^j_-$ over $(t_{j-1} - \delta, t_* + \delta)$ and $Q^j_+$ over $(t_* - \delta, t_j + \delta)$. (We might, of course, need to decrease $\delta$.)

Note that the above argument required $m \geq 3$ because otherwise the 2–fold in the 1–2 swallowtail is a definite fold.

To arrange that $g_t$ is ordered when $g_0$ and $g_1$ are ordered, we first need to

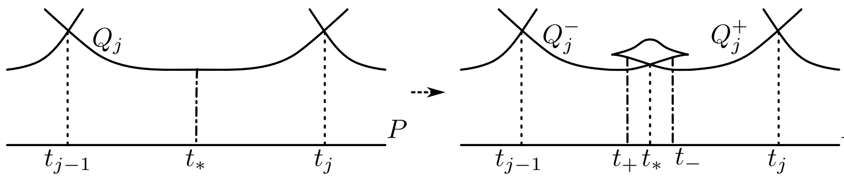

Figure 10: Splitting $Q^j$.

arrange that each birth or death of a cancelling $k$–$(k+1)$ pair occurs above all the other $k$–folds and below all the other $(k+1)$–folds. This is straightforward because such a modification of $g_t$ can be acheived through a generic homotopy supported in a neighborhood of an arc, which can be chosen to be disjoint from the $\phi_i(L_i)$'s. Now the only issue is pulling $k$–folds below $j$–folds when $k < j$ (or pushing $j$–folds above $k$–folds); this can be achieved if we ignore the $\phi_i(L_i)$'s for the same reason that it can be achieved for a fixed Morse function, as in Theorem 3.2, namely that, for a generic 1–parameter family of gradient-like vector fields, $k$–handles will not slide over $j$–handles if $k < j$. However, if we want to avoid modifying $g_t$ near each $\phi_i(L_i)$ we need to be more careful, and to do this we count dimensions again.

Here we need to check that either the 1–parameter descending disk for the $k$–fold $Q$ or the 1–parameter ascending disk for the $j$–fold $P$ misses the $l_i$–dimensional submanifold $\phi_i(L_i)$ in the level set $g^{-1}(z_i)$, which is presumed to be between $Q$ and $P$. The level set is $(m-1)$–dimensional, the descending sphere for $Q$ in the level set is $(k-1)$–dimensional, and the ascending sphere for $P$ is $(m-j-1)$–dimensional. However, because of the parameter $t$, we now want that either $(k-1)+1+l_i < m-1$ or that $(m-j-1)+1+l_i < m-1$, i.e. that $k+l_i < m-1$ or that $l_i < j-1$. If $l_i \geq j-1$, so that $k < j \leq l_i+1$, we have $k \leq l_i$ and thus $k+l_i \leq 2l_i$. Thus we are fine as long as $l_i < (m-1)/2$, but in our initial hypotheses we only assumed that $l_i < m/2$. The only potentially bad case is when $m$ is odd, $k = l_i = (m-1)/2$ and $j = (m+1)/2$. In this case both the ascending sphere for $P$ and the descending sphere for $Q$ will intersect $\phi_i(L_i)$ at discrete times. However, now we simply note that, again by genericity, these times will be distinct for $P$ and $Q$, and so at times when the ascending sphere for $P$ intersects $\phi_i(L_i)$ we lower $Q$ below $P$ while at times when the descending sphere for $Q$ intersects $\phi_i(L_i)$ we raise $P$ above $Q$.

□

**Theorem 3.10** *Suppose that $m \geq 3$. Given two indefinite generic homotopies $g_{0,t}, g_{1,t} \colon M \to I$ between indefinite Morse functions $g_{0,0} = g_{1,0}$ and $g_{0,1} = g_{1,1}$,* *there exists an indefinite generic homotopy of homotopies $g_{s,t}\colon M \to I$ from $g_{0,t}$ to $g_{1,t}$ with fixed endpoints. If $g_{0,t}$ and $g_{1,t}$ are ordered then we can arrange that $g_{s,t}$ is almost ordered. When $m \geq 4$ and $F_0$ and $F_1$ are both connected this guarantees that all level sets of each $g_{s,t}$ are connected. In the case where $m = 3$ and $F_0$ and $F_1$ are both connected, we can do a little extra work to arrange that all level sets of each $g_{s,t}$ are connected, even though the "almost ordered" condition is not sufficient to imply this.*

(Compare Proposition 3.6 in [8], which deals with approximately the same issue, but is about cancelling critical points of arbitrary indices and requires high ambient dimensions. There are many striking similarities between that proof and our proof of Theorem 3.10.)

Note that we could also ask that $g_{s,t}$ behave well on neighborhoods of the $L_i$'s as in the preceding theorems, and presumably there are constraints in terms of the dimensions involved, but we have no need for such a result in this paper.

Although the statement of the theorem does not say this, we will actually be able to modify a given generic homotopy $g_{s,t}$ from $g_{0,t}$ to $g_{1,t}$ through a generic homotopy $g_{r,s,t}$ with yet one more parameter $r \in [0,1]$, with $g_{0,s,t} = g_{s,t}$, so that $g_{1,s,t}$ satisfies the various conditions we want. In doing so we will pass through higher codimension singularities, and so we could have "cancellation" lemmas analogous to Lemmas 3.5, 3.8, 3.7 and 3.9. In our case they would involve the "butterfly singularity" and the "monkey saddle" (or "elliptic umbilic"). However, since each only occurs once in the proof, we just develop them in the course of the proof.

**Proof** There is always a generic homotopy $g_{s,t}\colon M \to I$ rel. boundaries, so the only issue is to make it indefinite, that is, to remove all 0–folds ($m$–folds are treated the same way using $1 - g_{s,t}$).

Instead of the traditional Cerf graphic, we consider a 1–parameter family of Cerf graphics, and in this case the 0–folds form a 2–dimensional immersed surface $\Sigma$, as in the example in Figure 11. In $\Sigma$, with respect to the $s$ direction, there are merges, unmerges, eyes and swallowtails; apart from the swallowtails, these appear as smooth curves with tangents parallel to the $t$ direction.

The first step is to cut $\Sigma$ into pieces, each of which is embedded in $I \times I \times I$. We do this by first cutting $\Sigma$ in the $t$ direction at many fixed $s$ values. Such a cut is done by first making $g_{s,t}$ independent of $s$ in a small $s$–interval $[s_* - \delta, s_* + \delta]$, then applying the technique in the proof of Theorem 3.6 to modify the homotopy $g_{s_*,t}$ to get rid of definite folds at $s_*$, and then noting that the modification

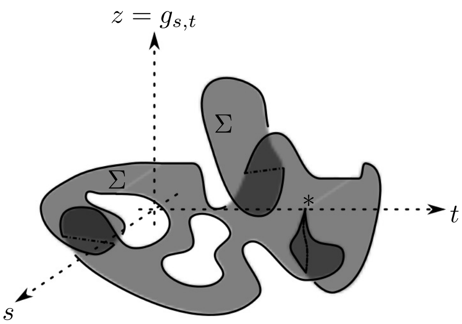


Figure 11: Example of a 2–dimensional surface Σ of 0–folds in a generic 2–parameter homotopy between Morse functions. Note the swallowtail singularity at the point labelled ∗.

is through a generic homotopy between homotopies, which can then be run forward and backward in the $s$ direction as $s$ ranges from $s_* - \delta$ to $s_*$ to $s_* + \delta$. A typical example of the result is illustrated in Figure 12.

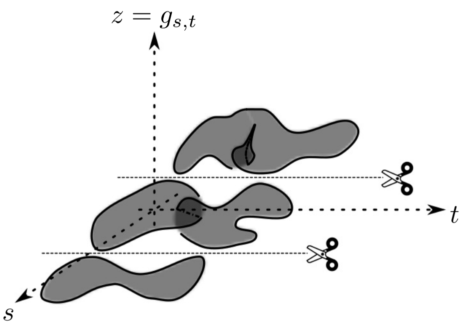


Figure 12: After the first cuts, along constant $s$ slices. Note that components are still not embedded, due to the bad swallowtail.

The components of Σ are not yet embedded because of the possibility that there are births of swallowtails in the middle of Σ. However, in [8], page 199 (see also item 1 on page 194), it is shown exactly how such a swallowtail birth can be extended past the 0–1 cusp and onto the surface of 1–folds by passing through a "butterfly singularity", with the result that the swallowtail cuts Σ into two parts, which intersect each other but no longer have a self-intersection. The change in the movie of Cerf graphics is shown in Figure 13, with accompanying graphs of the 1–dimensional Morse functions.

Now each component $\mathcal{P}$ of Σ is embedded in $I \times I \times I$ and we will eliminate

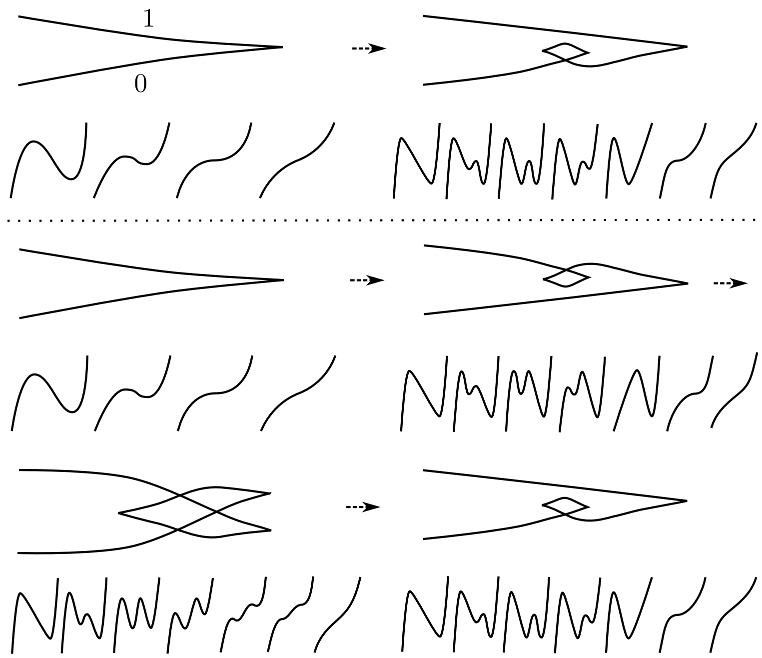


Figure 13: Cutting the bad swallowtail onto the surface of 1–folds using the butterfly singularity. Above the dotted line, the swallowtail occurs in the middle of a 0–fold, leading to surface of 0–folds that is not embedded. Below the line, the swallowtail is born in the middle of a 1–fold, so that the two 0–folds which intersect are in distinct components of the surface of 0–folds.

these components one by one, in much the same way we eliminated individual 0–folds in Theorem 3.6 and individual index 0 critical points in Theorem 3.2. Noting that, for each $(s,t)$, the index 0 critical point $\mathcal{P}_{s,t}$ coming from $\mathcal{P}$ is cancelled by some index 1 critical point, we can cover $\mathcal{P}$ with open disks $\{\mathcal{P}^i\}$ over each of which we have chosen a particular disk $\mathcal{Q}^i$ of cancelling index 1 critical points. We also arrange that every vertex in the nerve of this cover (including vertices on $\partial\mathcal{P}$) has valence 3 and that every edge of the nerve is transverse to constant $s$ slices. Then we can use the ideas in the proof of Theorem 3.6 to eliminate $\mathcal{P}$ at all points in exactly one or two of the open sets of the cover, and we reduce to the case where $\mathcal{P}$ is a union of disjoint triangles each cancelled by three distinct surfaces of 1–folds.

Figure 14 shows a sequence of Cerf graphics, representing the 2–parameter Cerf graphic where three open sets intersect; on the right we show the covering in parameter space, with the nerve and its trivalent vertex. The labels $a$, $b$ and $c$

indicate the 1–handles that cancel in each of the three open sets. At each point in parameter space, we adopt the convention that the closest index 1 critical point to the 0–fold is the cancelling one. This is not necessarily the case at first, but right before the cancellations this will be true. Next, Figure 15 shows the result of cancelling the 0–fold with the appropriate 1–folds away from the overlaps in the cover, with the labels $ab$, $bc$ and $ac$ indicating the pairs of 1–folds which cancel in each region. Finally, Figure 16 shows how this sequence of Cerf graphics is transformed by cancelling the swallowtails that remain when two of the open sets intersect, leaving the 0–fold uncancelled only in a cusped-triangular neighborhood of the trivalent vertex. (The boxed numbers indicate points in parameter space for reference in the next figure to come.)

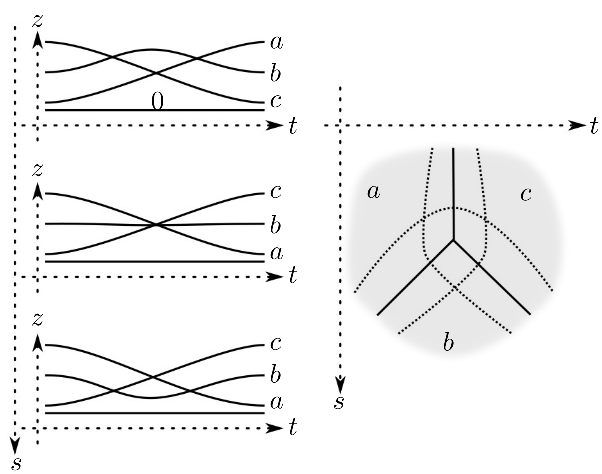


Figure 14: Three 1–folds $a$, $b$ and $c$ cancelling a 0–fold in two parameters.

Because there are three 1–handles cancelling the 0–handle across this triangle, we can construct a local model in which we separate out two local coordinates in the domain in which the cancellations occur and keep the other coordinates as a sum of squares independent of the parameters. In other words, locally $g_{s,t}$ is given by $g_{s,t}(x_1, x_2, x_3, \ldots, x_m) = h_{s,t}(x_1, x_2) + x_3^2 + \ldots + x_m^2$. In Figure 17 we schematically illustrate the handle decomposition corresponding to $h_{s,t}$ at representative points in the $(s,t)$–parameter space, as labelled by circled numbers in Figure 16. In [8], page 202, this is shown to be precisely the southern hemisphere of the $S^2$–boundary of a $B^3$ space of deformations of the monkey saddle $h(x_1, x_2) = x_1^3 - 3x_1x_2^2$. The south pole is visualized as pushing down in the middle of the monkey saddle to create an index 0 critical point with three cancelling index 1 critical points, while the equator is a loop of Morse functions

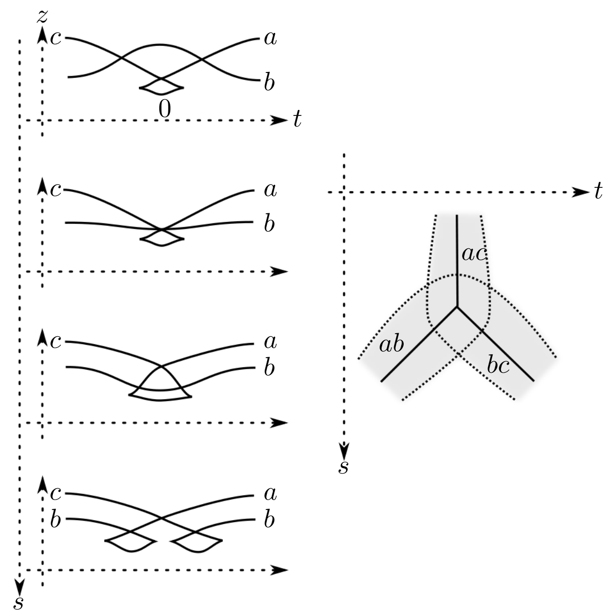


Figure 15: After cancelling the 0–1 pairs away from the overlapping regions.

involving two index 1 critical points that slide over each other three times. To eliminate this 0–fold completely, we simply replace the southern hemisphere of this $S^2$ with the northern hemisphere, which involves pushing the middle of the monkey saddle upwards to create an index 2 critical point. We replace the triangular 0–fold surface with a triangular 2–fold surface, exactly as in [8], page 198, item (iv) and Lemma 4.1. (This is where it is important that $m \geq 3$, so that index 2 is indefinite.)

Do this to each component of $\Sigma$ and, upside down, do the same to the index $m$ critical points and we have an indefinite generic homotopy between homotopies.

Getting $g_{s,t}$ to be almost ordered when $g_{0,t}$ and $g_{1,t}$ are ordered follows the same argument as for a fixed Morse function or a homotopy between Morse functions, except that we now note that for a generic 2–parameter family of gradient-like vector fields the descending manifold for an index $j$ critical point may meet the ascending manifold for an index $j+1$ critical point at isolated points $(s,t)$ in the 2–dimensional parameter space. This is why we can at most ask that, if an index $k$ critical point is above an index $j$ critical point, then $j \leq k+1$. As noted earlier, when $F_0$ and $F_1$ are connected and $m \geq 3$ then ordered implies connected level sets. In fact, for connectedness of level sets we

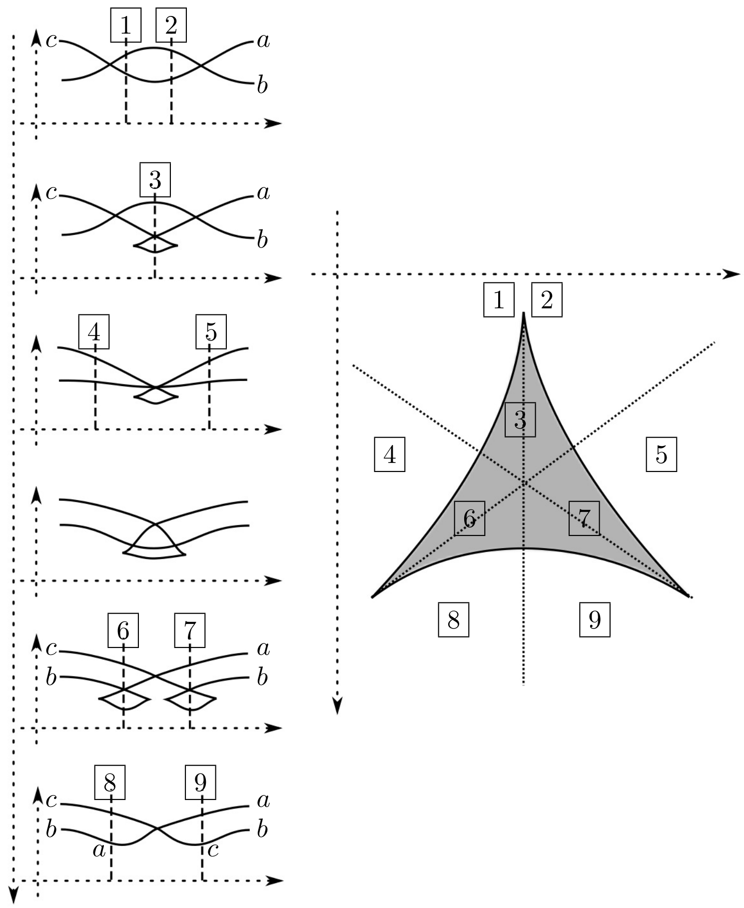


Figure 16: The 2–parameter Cerf graphic after cancelling along the double overlaps but not the triple overlaps. The dotted lines in the figure on the right indicate the points where two 1–folds intersect. Note that the labelling of the 1–folds $a$, $b$ and $c$ is no longer consistent because, as one moves across the bottom Cerf graphic, $c$ becomes $b$ and $b$ becomes $a$, and as one moves across the top Cerf graphic, $c$ becomes $a$.

simply need that all index $m-1$ critical points are above all index 1 critical points, and thus almost ordered implies connected level sets when $m \geq 4$. So, if $m \geq 4$, the proof is complete.

For the remainder of the proof, we assume $m = 3$, in which case level sets are 2–dimensional.

There are no definite folds, so the only way for a level set to become disconnected

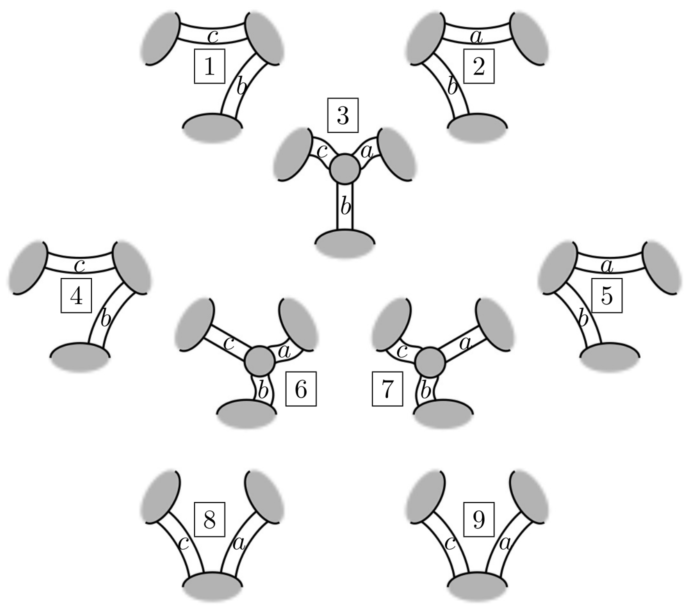


Figure 17: Handle decompositions at various points in the Cerf graphic in Figure 16.

is by adding a 2–handle $H_2$ to a separating circle $C$ in a lower level set. Then this level set would remain disconnected all the way to the top of the Morse function unless a higher 1–handle $H_1$ is attached to the different components of the level set. In the 0– and 1–parameter cases, we can always arrange that the 1–handle is added below the 2–handle in which case the attaching circle $C$ of the 2–handle does not separate. But in the 2–parameter case, the attaching 0–sphere of the 1–handle $H_1$ may go over the 2–handle $H_2$ and not be able to be pulled off. This is illustrated in Figure 18, where one foot of the 1–handle moves around $C_1$ and over the 2–handle $H_2$, as the parameter runs over a 2–disk $D$. At the center of $D$, the foot is stuck at the critical point of $H_2$.

However, it is important to note that this problem occurs at isolated points, so we can pull the 1–handles below the 2–handles everywhere except at small disks exactly like this disk $D$.

To deal with the disconnectedness of the level set over the middle of $D$, above

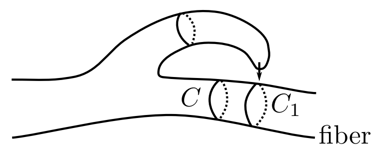


Figure 18: A 1–handle sliding over a 2–handle in a 2–parameter family.

the 2–handle $H_2$ and below the 1–handle $H_1$, we need a "helper" 1–handle $H_1'$, so create a 1–2 cancelling pair over the entire disk $D$ in which the helper 1–handle $H_1'$ is attached to a parallel copy $C'$ of $C$ and over $H_2$, and $H_2'$ is a "tunnel" between $H_1$ and $H_1'$. Now perturb $H_1'$ so that the value of the parameter in $D$ at which $H_1'$ hits the critical point of $H_2$ is different from $0 \in D$, say $d \in D$. Now the fibers are connected except for disjoint disk neighborhoods of 0 and $d$ of radius less than $d/3$. However, in the disk around $0 \in D$, the helper 1–handle can be pulled below $H_2$, so that $H_2$ does not disconnect and then $H_1$ is attached, and lastly $H_2'$ which obviously does not disconnect. By symmetry, the level sets can also be made connected in the disk around $d \in D$.

□

We end this section with a related lemma, of a slightly different flavor, that will be used several times in the following sections to relate connectedness of level sets to the property of being ordered. We have already stated that ordered $I$–valued Morse functions have connected level sets, and the converse is obviously not true. However, we do have:

**Lemma 3.11** *Let $g\colon M \to I$ be an indefinite Morse function which is standard with respect to each $\phi_i$ at height $z_i$, and suppose that all level sets of $g$ are connected. Then there exists a generic homotopy $g_t\colon M \to I$ between Morse functions, with $g_0 = g$, such that $g_1$ is ordered, all level sets of $g_t$ are connected for all $t$, and such that each $g_t$ is standard with respect to each $\phi_i$ at height $z_i$.*

**Proof** In the process of ordering the critical points, we just need to check that we do not create disconnected level sets. Disconnected level sets only come from $(m-1)$–handles attached along separating spheres. We will never be moving $(m-1)$–handles below other handles in the ordering process. If the attaching sphere for an $(m-1)$–handle $H$ is nonseparating, it can only be made separating by attaching another $(m-1)$–handle below $H$. Thus we never need to create disconnected level sets while ordering if all level sets were connected to begin with. □

# 4 Theorems about $I^2$–valued Morse 2–functions on cobordisms between cobordisms

Throughout this section let $X$ be an $n$–dimensional cobordism from $M_0$ to $M_1$, where each $M_i$ is an $(n-1)$–dimensional cobordism from $F_{i0}$ to $F_{i1}$, with $F_{00} \cong F_{10}$ and $F_{01} \cong F_{11}$. Recall that this means that $\partial X$ is equipped with a fixed identification with $-M_0 \cup (I \times F_{00}) \cup (I \times (-F_{01})) \cup M_1$, with $F_{0j} \subset M_0$ identified with $\{0\} \times F_{0j}$ and $F_{1j} \subset M_1$ identified with $\{1\} \times F_{0j}$. Also suppose that we are given indefinite Morse functions $\zeta_0 \colon M_0 \to I$ and $\zeta_1 \colon M_1 \to I$. On $I^2$ we use coordinates $(t,z)$, thinking of $t$ as horizontal and $z$ as vertical.

Our goal in this section is to prove the following two theorems, an existence theorem and a uniqueness theorem for square Morse 2–functions:

**Theorem 4.1** *Suppose that $n \geq 3$. Given any indefinite Morse function $\tau \colon X \to I$ which is projection to $I$ on $I \times F_{00}$ and $I \times (-F_{01})$ in $\partial X$, there exists an indefinite square Morse 2–function $G \colon X \to I^2$ such that $z \circ G|_{M_0} = \zeta_0$, $z \circ G|_{M_1} = \zeta_1$ and $t \circ G = \tau$. If $\zeta_0$, $\zeta_1$ and $\tau$ are ordered then we can arrange that all fibers of $G$ are connected.*

**Theorem 4.2** *Suppose that $n \geq 4$. Given two indefinite square Morse 2–functions $G_0, G_1 \colon X \to I^2$ with $z \circ G_0|_{M_0} = z \circ G_1|_{M_0} = \zeta_0$ and $z \circ G_0|_{M_1} = z \circ G_1|_{M_1} = \zeta_1$, there exists an indefinite generic homotopy $G_s \colon X \to I^2$ between $G_0$ and $G_1$. Let $\tau_0 = t \circ G_0$ and let $\tau_1 = t \circ G_1$. If $\zeta_0$, $\zeta_1$, $\tau_0$ and $\tau_1$ are all ordered Morse functions and if $G_0$ and $G_1$ have all their fibers connected then we can arrange for each $G_s$ to have all fibers connected as well.*

Note that in the preceding section we have already proved these two theorems in the special case that $X = [0,1] \times M$, $M_0 = \{0\} \times M$, $M_1 = \{1\} \times M$, and $\tau = \tau_0 = \tau_1$ is projection to $[0,1]$. This is because a generic homotopy $g_t \colon M \to I$ between Morse functions $g_0, g_1 \colon M \to I$ gives a square Morse 2–function $G \colon X \to I^2$ defined by $G(t,p) = (t, g_t(p))$, with $z \circ G|_{M_i} = g_i$. The key difference between a general square Morse 2–function and one coming from a generic homotopy between Morse functions is that the "Cerf graphic" for a general square Morse 2–function, i.e. the image $G(Z_G) \subset I^2$ of the critical point set, may have vertical tangencies. These vertical tangencies correspond precisely to critical points of the horizontal Morse function $t \circ G\colon X \to I$. In the absence of such vertical tangencies, the horizontal Morse function is trivial and hence $X$ is a product. We exploit these ideas repeatedly in the following proofs.

Before working on the proofs, we spend some time understanding neighborhoods of these critical points. In other words, when $t \circ G\colon X \to I$ has a critical point at $p \in X$, what can we say about $G$ near $p$? We first construct two local models which are illustrated in Figure 19 (recall that $\mu_k^n(\mathbf{x}) = \mu_k^n(x_1, \ldots, x_n) = -x_1^2 - \ldots - x_k^2 + x_{k+1}^2 + \ldots + x_n^2$):

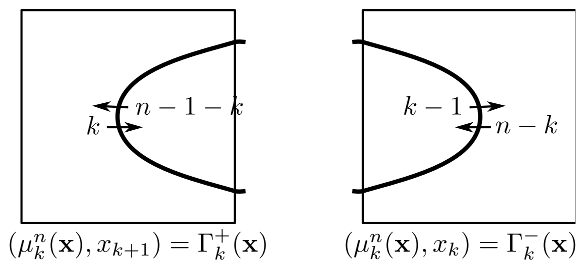


Figure 19: Local models for Morse critical points of index $k$ realized via Morse 2–functions. On the left is a "forward $k$–handle", on the right is a "backward $k$–handle". Again, the arrows indicate the index of the fold in the given direction, not the index of the Morse critical point.

(1) We call the following map a *forward index $k$ critical point*, or a *forward $k$–handle*: $\Gamma_k^+(x_1, \ldots, x_n) = (\mu_k^n(\mathbf{x}), x_{k+1})$. Note that this only makes sense for $0 \le k \le n-1$. By construction $t \circ \Gamma_k^+$ has a critical point of index $k$ at $0$. Reparametrizing the range by $(t, z) \mapsto (t - z^2, z)$ transforms $\Gamma_k^+$ to $(x_1, \ldots, x_n) \mapsto (\mu_k^{n-1}(x_1, \ldots, \hat{x_{k+1}}, \ldots, x_n), x_{k+1})$, showing that $\Gamma_k^+$ really is a (local) Morse 2–function with a single fold $Z_{\Gamma_k^+}$ along the $x_{k+1}$ axis and with the image $\Gamma_k^+(Z_{\Gamma_k^+})$ of this fold equal to the rightward-opening parabola $t = z^2$. The fold is indefinite when $0 < k < n-1$.

(2) This map is a *backward index $k$ critical point*, or a *backward $k$–handle*: $\Gamma_k^-(x_1, \ldots, x_n) = (\mu_k^n(\mathbf{x}), x_k)$. This only makes sense for $1 \le k \le n$. Again, by construction $t \circ \Gamma_k^-$ has a critical point of index $k$ at $0$. In this case, reparametrizing the range by $(t, z) \mapsto (t + z^2, z)$ transforms $\Gamma_k^-$ to $(x_1, \ldots, x_n) \mapsto (\mu_{k-1}^{n-1}(x_1, \ldots, \hat{x_k}, \ldots, x_n), x_k)$, showing that $\Gamma_k^-$ really is a (local) Morse 2–function with a single fold $Z_{\Gamma_k^-}$ along the $x_k$ axis and with the image $\Gamma_k^-(Z_{\Gamma_k^-})$ of this fold equal to the rightward-opening parabola $t = -z^2$. This fold is indefinite when $1 < k < n$.

Note that we could have defined $\Gamma_k^+$ by $\Gamma_k^+(\mathbf{x}) = (\mu_k^n(\mathbf{x}), \pm x_j)$ for any $j \in \{k+1, \ldots, n\}$ and it would still have all the properties listed, and in fact such a definition is equivalent to the one given up to a change of coordinates in the domain. Similarly we could define $\Gamma_k^-(\mathbf{x}) = (\mu_k^n(\mathbf{x}), \pm x_j)$ for any $j \in \{1, \ldots, k\}$. However, $\Gamma_k^+$ and $\Gamma_k^-$ are not in general related by a change of coordinates, even allowing a change of coordinates in the range, because the indices of the folds are different, as can be seen in Figure 19. If we turn a forward $k$–handle backwards, i.e. post-compose $\Gamma_k^+$ with the time-reversal $(t, z) \mapsto (-t, z)$, it becomes a backward $(n-k)$–handle.

Here are some further observations about a forward $k$–handle; the reader can figure out the parallel statements for backward handles:

(1) When $0 < k \leq n-1$, the descending disk $\{x_{k+1} = \ldots = x_n = 0\}$ has image equal to the horizontal line $\{t \leq 0, z = 0\}$. When $k = 0$ there is, of course, no descending disk.

(2) When $0 \leq k < n-1$, the ascending disk $\{x_1 = \ldots = x_k = 0\}$ has image equal to the "interior" of the parabola $\{t \geq z^2\}$. (Of course, when $k = 0$ the ascending disk is the whole domain of the function.) When $k = n-1$, the image of the ascending disk is just the parabola $\{t = z^2\}$.

(3) For $k > 0$, the descending $(k-1)$–sphere $\{x_{k+1} = \ldots = x_n = 0, x_1^2 + \ldots + x_k^2 = R^2\}$ has image equal to the point $(-R^2, 0)$.

(4) For $k > 0$, the attaching region for the $k$–handle, which we identify as the set $\{\mu_k^n(\mathbf{x}) = -R^2, -\epsilon \leq x_{k+1} \leq \epsilon, x_{k+2}^2 + \ldots + x_n^2 = \epsilon^2\} \cong [-\epsilon, \epsilon] \times B^{n-1-k} \times S^{k-1}$ has image equal to the line segment $\{t = -R^2, -\epsilon \leq z \leq \epsilon\}$, and the map to this line segment is simply projection onto $[-\epsilon, \epsilon]$. This is where we first see the relevance of the conditions in the preceding section regarding constructing Morse functions which are standard with respect to embeddings of $[-\epsilon, \epsilon] \times B^{n-1-k} \times S^{k-1}$.

(5) For $k < n-1$, the ascending $(n-k-1)$–sphere $\{x_1 = \ldots = x_k = 0, x_{k+1}^2 + \ldots + x_n^2 = R^2\}$ maps to the line $\{t = R^2\}$ via the standard Morse function on a sphere, with image equal to the interval $\{-R \leq z \leq R\}$. For $k = n-1$ the ascending sphere is two points mapping to $(R^2, \pm R)$.

First we prove existence using these local models:

**Proof of Theorem 4.1** If $\tau$ has no critical points then we use a gradient flow to identify $X$ with $I \times M_0$ such that $\tau(t, p) = t$, and then we see $\zeta_0$ and $\zeta_1$ as two indefinite Morse functions on $M_0$. Then our result follows from Theorem 3.6; we get an indefinite generic homotopy $\zeta_t$ and we let $G(t, p) = (t, \zeta_t(p))$.

Thus if we can now prove the theorem in the case where $\tau$ has exactly one critical point $p \in X$, then we are done. Suppose that $\tau(p) = 1/2$ and that $p$ has index $k \leq n/2$. (If $k > n/2$ then replace $\tau$ with $1-\tau$ and switch $M_0$ and $M_1$.) Choose a gradient-like vector field $V$ for $\tau$, and use this to find an embedding $\phi\colon [-\epsilon, \epsilon] \times B^{n-1-k} \times S^{k-1} \hookrightarrow M_0$ which gives the glueing map for the associated handle with appropriate framing. (We have split the normal direction into a product of $[-\epsilon, \epsilon]$ and $B^{n-1-k}$ as preparation for the use of Theorem 3.2. Note that $k-1$ is the dimension referred to as $l_i$ in the statement of that theorem, and that the dimension of $M_0$ is $m = n-1$. We need to verify that $k-1 < (n-1)/2$, which we do have, because $k-1 \leq (n/2)-1 = (n-2)/2 < (n-1)/2$.) Let $T \subset M_0$ be the image of $\phi$. For some small $\delta > 0$ we can then decompose $X$ into a union of four parts, $X = X_0 \cup X_c \cup H \cup X_1$ with the following properties: (Figure 20 shows where these four parts will sit in $I^2$ and shows what the singular locus $G(Z_G)$ will look like in each part.)

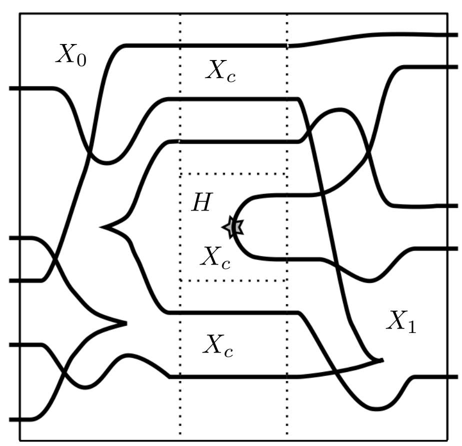


Figure 20: Constructing a Morse 2–function with a single critical point in the horizontal Morse function. This diagram shows the images of the four parts of $X$, $X_0$, $X_1$, $X_c$ and the handle $H$; the image of the critical point is at the star. Note that the images of $X_0$ and $X_1$ are disjoint, while the image of $X_c$ contains the image of $H$, even though $X_c$ and $H$ intersect in $X$ only along their boundaries.

(1) $X_0 = \tau^{-1}[0, 1/2-\delta]$ and is identified, via $V$, with $[0, 1/2-\delta] \times M_0$ in such a way that $\tau|_{X_0}(t,p) = t$.

(2) $X_1 = \tau^{-1}[1/2+\delta, 1]$ and is identified, via $V$, with $[1/2+\delta, 1] \times M_1$ in such a way that $\tau|_{X_1}(t,p) = t$.

(3) $X_c \subset \tau^{-1}[1/2-\delta, 1/2+\delta]$ and is identified, via $V$, with $[1/2-\delta, 1/2+\delta] \times (M_0 \setminus (T \setminus \partial T))$, in such a way that $\tau|_{X_c}(t,p) = t$.

(4) $H$ is the $k$–handle, the union of the forward flow lines for $V$ starting at $T$, together with the ascending manifold of $p$ (using $V$), intersected with $\tau^{-1}[1/2-\delta, 1/2+\delta]$. On $H$ we have coordinates $(x_1, \ldots, x_n)$ with respect to which $\tau(\mathbf{x}) = 1/2+\mu_k^n(\mathbf{x})$ and $V = -x_1\partial_{x_1} - \ldots - x_k\partial_{x_k} + x_{k+1}\partial_{x_{k+1}} + \ldots + x_n\partial_{x_n}$. We choose these coordinates so that the $x_{k+1}$ direction is the $[-\epsilon, \epsilon]$ direction in the attaching region $[-\epsilon, \epsilon] \times B^{n-1-k} \times S^{k-1}$ (and $(x_{k+2}, \ldots, x_n)$ give the $B^{n-1-k}$ directions while the sphere in the $(x_1, \ldots, x_k)$ coordinates gives the $S^{k-1}$ factor).

In order to construct $G$, we first use Theorem 3.2 to construct an indefinite Morse function $\zeta_{1/2-\delta}\colon M_0 \to I$ which is standard with respect to $\phi$ at height $1/2$, so that $\zeta_{1/2-\delta}$ on the attaching region $T$ is of the form $(t, x, p) \mapsto t + 1/2$ (identifying $T$ with $[-\epsilon, \epsilon] \times B^{n-1-k} \times S^{k-1}$ via $\phi$). If we want connected fibers for $G$, make sure that $\zeta_{1/2-\delta}$ is ordered. Now use Theorem 3.6 to construct an indefinite generic homotopy $\zeta_t$ (for $t \in [0, 1/2-\delta]$) connecting $\zeta_0$ to $\zeta_{1/2-\delta}$. Again, make this ordered if we care about connected fibers. Then we let $G\colon X_0 \to [0, 1/2-\delta] \times I$ be $G(t, p) = (t, \zeta_t(p))$, after identifying $X_0$ with $[0, 1/2-\delta] \times M_0$ as above. On $H$, at first just let $G(\mathbf{x}) = \Gamma_k^+(\mathbf{x}) + (1/2, 1/2)$, a forward $k$–handle. This gives the single vertical tangency as part of a horizontal parabola seen in the middle in Figure 20. This fits together smoothly with the definition of $G$ on $X_0$. Now we postcompose with an isotopy of $[1/2-\delta, 1/2+\delta] \times [0,1]$ to make the image of $H$ exactly equal to the square $[1/2-\delta, 1/2+\delta] \times [1/2-\epsilon, 1/2+\epsilon]$, so that $G$ as defined on $X_0$ and $H$ extends smoothly to $X_c \cong [1/2-\delta, 1/2+\delta] \times (M_0 \setminus (T \setminus \partial T))$ via $G(t, p) = (t, \zeta_{1/2-\delta}(p))$. One sees then that these definitions fit together smoothly to define $G$ over $\tau^{-1}[0, 1/2+\delta]$, and that $z \circ G$ then defines an indefinite Morse function $\zeta_{1/2+\delta}$ on $\tau^{-1}\{1/2+\delta\}$, which is identified with $M_1$ via $V$. Finally, we use Theorem 3.6 to construct an indefinite generic homotopy $\zeta_t$ (for $t \in [1/2+\delta, 1]$) connected $\zeta_{1/2+\delta}$ to $\zeta_1$, and we define $G$ on $X_1 \cong [1/2+\delta, 1] \times M_1$ by $G(t, p) = (t, \zeta_t(p))$.

If we arranged that $\zeta_t$ is ordered for $t \in [1/2+\delta, 1]$, we would have completed the proof of the connectedness assertion. This is fine if $\zeta_{1/2+\delta}$ (handed to us by the construction on $X_0 \cup X_c \cup H$) is ordered. There is only one case when $\zeta_{1/2+\delta}$ will not be ordered, and that is when $k = n/2$. Here we have a critical point of index $n/2$ below a critical point of index $n-1-n/2 = n/2-1$. The relevant indices are indicated on the left in Figure 21. However, as long as the level sets of $\zeta_{1/2+\delta}$ are connected, we can start off the homotopy $\zeta_t$, $t \in [1/2+\delta, 1]$, by switching the heights of the two offending critical points as indicated on the right in Figure 21, and then keep everything ordered thereafter. The only case in which the level sets might be disconnected is when $n/2 = n-2$ and $n/2-1 = 1$, i.e. when $n = 4$ and $k = 2$. In this case we should make sure that, in $\zeta_{1/2-\delta}$, the attaching $S^1$ for the 2–handle $H$ does not separate the level set in which it lies. A moment's thought about the proof of Theorem 3.2 shows that this is easy to achieve. □

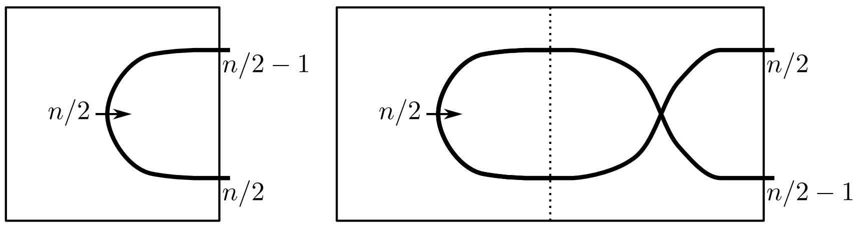


Figure 21: After attaching a handle of index $n/2$, the vertical Morse function will not be ordered, as on the left. The arrow labelled $n/2$ indicates that this fold is of index $n/2$ when looked at in the direction of the arrow. The numbers to the right of each box are the index of the critical points of the vertical Morse function there. In the box on the right, we have switched the two critical points to restore order.

The rest of this section is devoted to proving uniqueness, i.e. the proof of Theorem 4.2. To do this, we first show that our two models, forward and backward handles, are a complete list of local models, in the following sense:

**Lemma 4.3** *Consider a square Morse 2–function $G\colon X \to I^2$ and a critical point $p \in X$ of $t \circ G\colon X \to I$, of index $k$, with $G(p) = (t_p, z_p)$. Suppose we are given standard coordinates $(x_1, \ldots, x_n)$ on a neighborhood $\nu$ of $p$ such that $\tau = t \circ G(x_1, \ldots, x_n) = \mu^n_k(\mathbf{x}) + t_p$. Then there exists an arc $G_s$ of Morse 2–functions (i.e. $G_s$ is Morse for all $s$) supported inside $\nu$, with $G_0 = G$, such that, inside a smaller neighborhood $\nu' \subset \nu$, $G_1(\mathbf{x}) = \Gamma^{\pm}_k(\mathbf{x})$. It will be $\Gamma^+_k$, i.e. a forward $k$–handle, exactly when the point $(t_p, z_p)$ is a local minimum for $t|_{G(Z_G)}$, and it will be $\Gamma^-_k$, a backward $k$–handle, exactly when $(t_p, z_p)$ is a local maximum.*

**Proof** Let $\tau = t \circ G$ and $\zeta = z \circ G$, i.e. $G(\mathbf{x}) = (\tau(\mathbf{x}), \zeta(\mathbf{x}))$. We know that $\tau = \mu^n_k(\mathbf{x}) + t_p$. Because the rank of $DG$ at $p = 0$ must be 1, we know that $\zeta$ is nonsingular at 0 and so, after a small perturbation of $G$ (staying Morse) supported in a neighborhood of 0 we can assume that $\zeta$ is linear of the form $\zeta(x_1, \ldots, x_n) = a_1 x_1 + \ldots a_n x_n + z_p$. Now note that $-a_1^2 - \ldots - a_k^2 + a_{k+1}^2 + \ldots + a_n^2 \neq 0$, because otherwise we would have a 1–dimensional subset of the singular locus of $G$ on which $G$ was constant, which does not happen in any of the local models for a Morse 2–function. Thus there is a linear transformation of $\mathbb{R}^n$ preserving the quadratic form $\mu_k^n$ and taking $(a_1, \ldots, a_n)$ to $(0, \ldots, 0, c, 0, \ldots, 0)$ for some positive $c$, with the $c$ in either the $k$'th or $(k+1)$'st coordinate. Smoothly interpolate from the identity to this linear transformation while going in radially towards the origin in our neighborhood, and this creates an ambient isotopy of the domain. Precomposing with this ambient isotopy and postcomposing with a rescaling isotopy gives the arc of Morse 2–functions $G_s$ with the desired properties.

□

At some point one might wish that a certain forward index $k$ critical point was actually a backward index $k$ critical point, or vice versa. The next lemma addresses this:

**Lemma 4.4** *Suppose that $G\colon X \to I^2$ is a square Morse 2–function and that $p \in G$ is a critical point of $t \circ G$ of index $k$ with local coordinates with respect to which $G(\mathbf{x}) = \Gamma_k^{\pm}(\mathbf{x}) + (t_p, z_p)$. If $2 \leq k \leq n-2$ there exists an indefinite generic homotopy $G_s$ of Morse 2–functions supported inside this coordinate neighborhood, with $G_0 = G$, such that, inside a smaller neighborhood of $p$, we have $G_1(\mathbf{x}) = \Gamma_k^{\mp}(\mathbf{x}) + (t_p, z_p)$. If all fibers of $G$ are connected then we can arrange that all fibers of $G_s$ are connected for all $s$. We can further arrange that $t \circ G_s$ is independent of $s$.*

**Proof** Figure 22 shows how to do this (without the indefinite condition) when $n = 2$ and $k = 1$. For this we should in principle be able to write down an explicit formula, but the illustration probably explains what is going on better. We know that a forward handle has become a backward handle simply because the vertical tangency in the fold locus has changed from being a rightward-opening parabola to a leftward-opening parabola. The homotopy has passed through a swallowtail singularity. To get the higher dimensional version, and indefiniteness, consider Figure 22 to be a picture of what is happening in the $(x_k, x_{k+1})$–plane, and keep the homotopy independent of $s$ in the other coordinates $(x_1, \ldots, x_{k-1})$ and $(x_{k+2}, \ldots, x_n)$. It is easy to see that, for $n \geq 4$ and $2 \leq k \leq n-2$, this is an indefinite deformation and does not disconnect fibers. The way we have drawn it, the critical point moves slightly to the left, but after modifying by a small isotopy we can keep $t \circ G_s$ independent of $s$ throughout the homotopy.

□

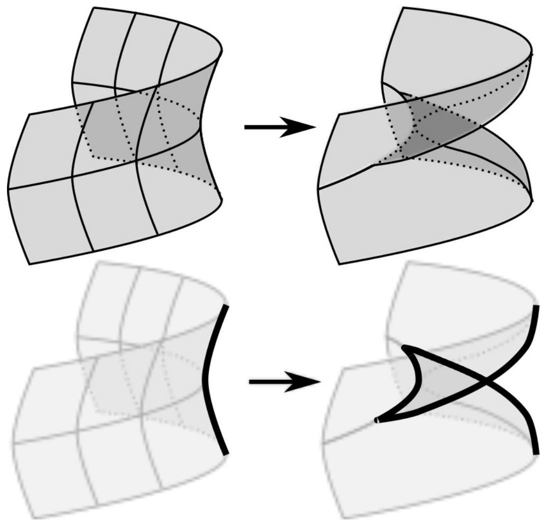

Figure 22: Bending a forward handle to a backward handle. The point is to see that the only change to the projection to the horizontal axis in this deformation is that the critical points moves a little to the left, but that when we look at the map to 2 dimensions, the fold at the critical point changes from opening to the right to opening to the left.

We will want to accompany the lemmas above with a lemma stating that, outside the standard neighborhood of a critical point of $t \circ G$, $G$ can be taken to be "constant in the $t$ direction". We state this precisely as follows:

**Lemma 4.5** *Consider a square Morse 2–function $G\colon X \to I^2$ and an index $k$ critical point $p \in X$ of $\tau = t \circ G$, with local coordinates near $p$ with respect to which $G = \Gamma^+_k(\mathbf{x}) + (t_p, z_p)$. In these local coordinates, let $V$ be the standard gradient vector field for $\tau = t \circ G = \mu^n_k(\mathbf{x})$ and, for any small $\epsilon > 0$, let $H_\epsilon$ be the $k$–handle neighborhood of $p$ obtained by flowing forward along $V$ starting from the attaching region $\{(x_1, \ldots, x_n) \mid \tau(x_1, \ldots, x_n) = -\epsilon + t_p, x^2_{k+1} = \epsilon^2, x^2_{k+2} + \ldots + x^2_n \leq \epsilon^2\} \cong S^{k-1} \times [-\epsilon, \epsilon] \times B^{n-k-1}$, together with the ascending disk, and stopping at $\tau = t_p + \epsilon$. Note that $H_\epsilon \subset \tau^{-1}[t_p - \epsilon, t_p + \epsilon]$; let $H^c_\epsilon$ be the closure of $\tau^{-1}[t_p - \epsilon, t_p + \epsilon] \setminus H_\epsilon$ and $M^c_{t_p-\epsilon} = H^c_\epsilon \cap \tau^{-1}(t_p - \epsilon)$. Using $V$, $H^C_\epsilon$ is naturally identified with $[t_p - \epsilon, t_p + \epsilon] \times M^c_{t_p-\epsilon}$. Then there exists an $\epsilon > 0$ and an arc of Morse 2–functions $G_s$, with $G_0 = G$, which is independent of $s$ inside $H_\epsilon$ and independent of $s$ outside $\tau^{-1}(t_p - 2\epsilon, t_p + 2\epsilon)$, such that, on $H^c_\epsilon$, identified with $[t_p - \epsilon, t_p + \epsilon] \times M^c_{t_p-\epsilon}$, $G_1$ is of the form $(t, x) \mapsto (t, g(x))$ for a fixed Morse function $g = \zeta|_{\tau^{-1}(t_p-\epsilon)}$. In addition, we can arrange that there* *are no critical values of* $g|_{M^c_{t_p-\epsilon}}$ *in* $[z_p - \epsilon, z_p + \epsilon]$.

**Proof** Given the standard model on the handle, the complement of the handle in $\tau^{-1}([t_p - 2\epsilon, t_p + 2\epsilon])$ is a product cobordism and $G$ on this product is identified with an arc of Morse functions $g_t$. A standard homotopy can make $g_t$ independent of $t$ for $t \in [t_p - \epsilon, t_p + \epsilon]$. □

We now present the proof of uniqueness for indefinite square Morse 2–functions as a sequence of steps forward-referencing two further lemmas which will be stated and proved afterwards.

**Proof of Theorem 4.2** We are given two indefinite square Morse 2–functions $G_0, G_1 \colon X \to I^2$ which agree on $M_0$ and $M_1$, i.e. $z \circ G_0|_{M_i} = \zeta_i = z \circ G_1|_{M_i}$, for $i = 0, 1$. Then we need to construct an indefinite generic homotopy $G_s \colon X \to I^2$ between $G_0$ and $G_1$, and we need to address the issue of connected fibers. The steps are as follows:

(1) Let $\tau_0 = t \circ G_0$ and let $\tau_1 = t \circ G_1$. These are indefinite $I$–valued Morse functions. Let $\tau_s \colon X \to I$ be an indefinite generic homotopy from $\tau_0$ to $\tau_1$ such that all the births of cancelling pairs of critical points occur for $s \in [0, 1/4]$ and all the deaths occur for $s \in [3/4, 1]$, and such that $\tau_s$ is independent of $s$ for all $s \in [1/4, 3/4]$. We then construct the desired generic homotopy $G_s$ for $s \in [0, 1/4]$ and for $s \in [3/4, 1]$ such that $t \circ G_s = \tau_s$. (We need to slightly modify $\tau_s$ to acheive this.) This step is carried out in Lemma 4.6. The key outcome of this step is that $t \circ G_{1/4} = t \circ G_{3/4}$, so that when we construct $G_s$ for $s \in [1/4, 3/4]$, we can leave $t \circ G_s$ independent of $s$ and work on $z \circ G_s$.

(2) Now we need to connect $G_{1/4}$ to $G_{3/4}$. Let $\tau = t \circ G_{1/4} = t \circ G_{3/4}$. Our next step is to extend the homotopy $G_s$ to $s \in [1/4, 1/2]$, keeping $t \circ G_s = \tau$ for all $s \in [1/4, 1/2]$ so that, for some $\epsilon > 0$ and for each critical value $t_*$ of $\tau$, $G_{1/2}$ and $G_{3/4}$ agree on $\tau^{-1}([t_* - \epsilon, t_* + \epsilon]) = G^{-1}_{1/2}([t_* - \epsilon, t_* + \epsilon] \times I) = G^{-1}_{3/4}([t_* - \epsilon, t_* + \epsilon] \times I)$. This step is carried out in Lemma 4.7.

(3) Finally, we extend $G_s$ to $s \in [1/2, 3/4]$ to connect $G_{1/2}$ to $G_{3/4}$ as follows: The parts of $X$ where $G_{1/2}$ and $G_{3/4}$ do not yet agree are of the form $X_* = \tau^{-1}([t_* + \epsilon, t'_* - \epsilon])$, for two consecutive critical values $t_* < t'_*$ of $\tau$. But then $X_*$ can be identified with a product $[t_* + \epsilon, t'_* - \epsilon] \times M_*$, where $M_* = \tau^{-1}(t_* + \epsilon)$. Furthermore, with this identification, for $s = 1/2$ and $s = 3/4$, we see that $G_s|_{X_*}$ is of the form $(t, p) \mapsto (t, g_{s,t}(p))$, precisely

because $t \circ G_{1/2} = t \circ G_{3/4} = \tau$. Thus we can use Theorem 3.10 from the preceding section to construct a homotopy $g_{s,t}$ from $g_{1/2,t}$ to $g_{3/4,t}$, and then define $G_s$ for $s \in [1/2, 3/4]$ and $p \in X_* = [t_* + \epsilon, t'_* - \epsilon] \times M_*$ by $G_s(t, p) = (t, g_{s,t}(p))$. Finally, since $G_{1/2}$ and $G_{3/4}$ already agree on $\tau^{-1}([t_* - \epsilon, t_* + \epsilon])$ for critical points $t_*$, we can define $G_s = G_{1/2} = G_{3/4}$ on these subsets, for all $s \in [1/2, 3/4]$, and we are done.

In the above steps we did not address the issue of keeping fibers connected. If $\tau_0$ and $\tau_1$ are ordered then we can arrange for $\tau_s$ to be ordered for all $s$ (Theorem 3.6). In this case, Lemma 4.6 also states that $G_s$ will have all fibers connected for all $s \in [0, 1/4]$ and for all $s \in [3/4, 1]$. Lemma 4.7 then explicitly asserts that, if the fibers of $G_{1/4}$ and $G_{3/4}$ are connected in each $\tau^{-1}([t_* - \epsilon, t_* + \epsilon])$, for $t_*$ a critical point of $\tau$, then we can keep the fibers of $G_s$ connected there when we construct $G_s$ for $s \in [1/4, 1/2]$. Finally, when we use Theorem 3.10 to construct $G_s$ for $s \in [1/2, 3/4]$, we should first use Lemma 3.11 to get each of the Morse function $g_{1/2, t_* + \epsilon} = g_{3/4, t_* + \epsilon}$ and $g_{1/2, t'_* - \epsilon} = g_{3/4, t'_* - \epsilon}$ to be ordered.

□

We now state and prove the two lemmas referenced in the proof above.

**Lemma 4.6** *Given any indefinite Morse 2–function $G\colon X \to I^2$ and an indefinite generic homotopy $\tau_s\colon X \to I$ between Morse functions, with $t \circ G = \tau_0$ and with no deaths of cancelling pairs of critical points, there exists an indefinite generic homotopy $\tau'_s\colon X \to I$, with $\tau'_0 = \tau_0$ and $\tau'_1 = \tau_1$, which is connected to $\tau_s$ by an arc of generic homotopies, and there exists an indefinite generic homotopy of Morse 2–functions $G_s\colon X \to I^2$ with $G_0 = G$ and with $t \circ G_s = \tau'_s$. If $\tau_s$ is ordered and all fibers of $G$ are connected then we can arrange that all fibers of $G_s$ are connected for all $s$.*

**Proof of Lemma 4.6** We will show how to construct generic indefinite homotopies $G_s$ such that $t \circ G_s$ is a generic homotopy between Morse functions which realizes either (1) a desired birth of a cancelling pair or (2) a desired crossing of two critical points. The given $\tau_s$ will then tell us where the births should be and which critical points should cross when. Making the births occur at these points and the right critical points cross at the right time, and pre- and post-composing $G$ with ambient isotopies, we can construct $G_s$ so that $\tau'_s = t \circ G_s$ is connected to $\tau_s$ by an arc of generic homotopies. (To see this, first note that an arc of Morse functions can always be realized by pre- and post-composing with ambient isotopies. This is shown carefully in Proposition ???

in [7]. Then note that births and critical point crossings can be localized by using bump functions to keep given homotopies stationary for short time periods outside standard neighborhoods.)

We deal with these two moves as follows:

(1) The easiest way to arrange a birth is to arrange that, inside the ball in which the birth should occur, $G$ has the form $G(x_1, \ldots, x_n) = (-x_1^2 - \ldots - x_k^2 + x_{k+1} + x_{k+2}^2 + \ldots + x_n^2, x_{k+1})$. This is a fold which is index $k$ looked at from left to right and the image of the fold set is the line $z = t$. We can arrange this via, for example, an eye birth as illustrated in Figure 23; there are two cases, one which is indefinite for $1 \leq k \leq n-3$ and one which is indefinite for $2 \leq k \leq n-2$. Now let $f_s(x)$ be a function which equals $x^3 - sx$ for $x$ in a neighborhood of 0, has no critical points outside that neigborhood for any $s \in [-\epsilon, \epsilon]$, and is linear in $x$ and independent of $s$ outside a slightly larger neighborhood. Finally let $G_s(x_1, \ldots, x_n) = -x_1^2 - \ldots - x_k^2 + f_s(x_{k+1}) + x_{k+2}^2 + \ldots + x_n^2, x_{k+1})$, the result of which is also illustrated in Figure 23. As long as this eye

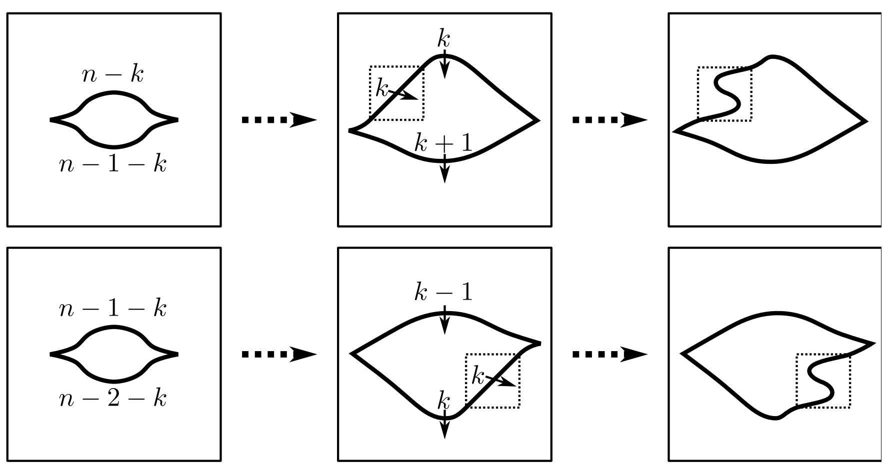


Figure 23: Two ways to realize a birth of a cancelling $k$–$(k+1)$ pair in $\tau_s$ via a homotopy of $G$; the birth occurs inside the dotted square.

birth is indefinite, it will not disconnect fibers.

(2) Perhaps surprisingly, moving one critical point below another is the more subtle move to deal with here. Here we have two critical points $p$ and $q$ with, say, $\tau_0(p) = 3/4$ and $\tau_0(q) = 1/2$ and no other critical values between $1/4$ and $3/4$. Suppose that $p$ has index $k$ and $q$ has index $l$; we

may assume that $k \leq l$ and that $k \leq n/2$. Using Lemmas 4.3 and 4.4, we arrange for local coordinates near $p$ with respect to which $G$ has the form $G = \Gamma_k^+(\mathbf{x}) + (3/4, z_p)$, a forward $k$–handle, and we arrange for a disjoint coordinate system near $q$ with respect to which $G = \Gamma_l^-(\mathbf{x}) + (1/2, z_q)$, a backward $l$–handle. Since there are no critical points between $p$ and $q$, $G^{-1}([1/2+\epsilon, 3/4-\epsilon])$ is diffeomorphic to $[1/2+\epsilon, 3/4-\epsilon] \times M$ for an $(n-1)$–manifold $M$, and via this diffemorphism $G$ has the form $(t,p) \mapsto (t, g_t(p))$ where $g_t$ is an indefinite generic homotopy between Morse functions on $M$. Furthermore, because of the local models we have found for $G$ near $p$ and $q$, we have an embedding $\phi_q$ of a neighborhood of the ascending sphere for $q$ and an embedding $\phi_p$ of a neighborhood of the descending sphere for $p$ in $M$ such that $g_{1/2+\epsilon}$ is standard with respect to $\phi_q$ at height $z_q$ while $g_{3/4-\epsilon}$ is standard with respect to $\phi_p$ at height $z_p$. It is then a straightforward application of Theorems 3.2 and 3.6 to arrange first that $g_t$ is standard with respect to $\phi_p$ at height $z_p$ for $t$ near $1/4$ and $1/2$, and then a straightforward application of Theorem 3.10 to arrange that $g_t$ is standard with respect to $\phi_p$ at height $z_p$ for all intermediate values of $t$. Here we are implicitly using Lemma 4.5 so that we can smoothly connect the behavior of $g_t$ for $t \in [1/2+\epsilon, 3/4-\epsilon]$ to the behavior of $g_t$ for $t \in [1/4+\epsilon, 1/2-\epsilon]$. Having done this, we can easily lower $p$ past $q$. This is illustrated in Figure 24, where "lowering" $p$ really means moving $p$ to the left.

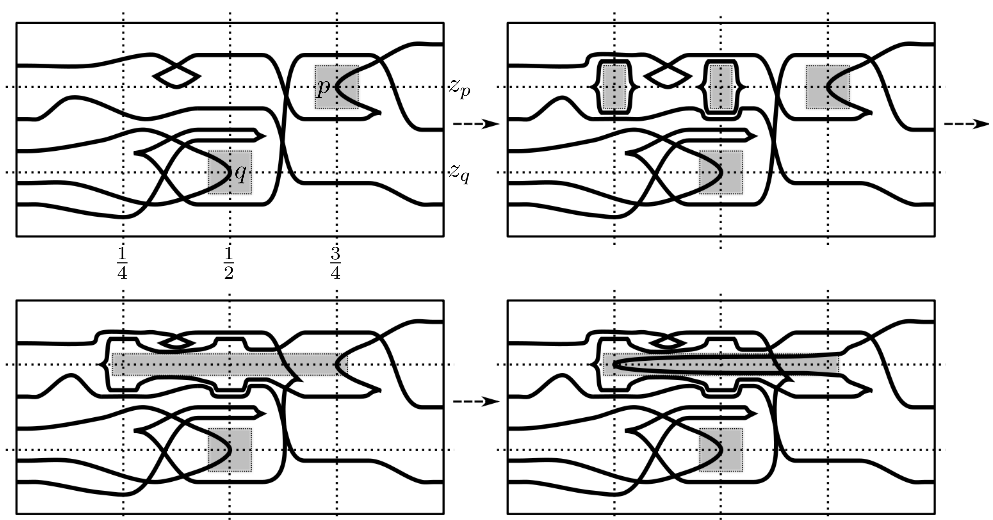


Figure 24: Passing two critical points past each other.

Regarding connectedness, note that the only potential problem arises when the attaching sphere for $p$ has codimension 1 in the fiber, and then we need to make sure that it never separates the fibers as we lower $p$ past $q$. Thus we only need to worry when $k = n - 2$, but since $k \leq n/2$ and $n \geq 4$, the only case of concern is $n = 4$ and $k = 2$. (To a 4–manifold topologist this is, of course, the most interesting case.) Thus, when we apply Theorems 3.2, 3.6 and 3.10 to arrange that $g_t$ is standard with respect to $\phi_p$ at height $z_p$ for all values of $t$ between $1/4$ and $3/4$, we actually need to do a little more: We need to ensure that the descending sphere for $p$ remains nonseparating in its fiber for all $t$. Thus we need a slight improvement on Theorem 3.6 which says that, in this special case $m = n - 1 = 3$ and $l_i = k - 1 = 1$, we can maintain the nonseparating property throughout a generic homotopy between Morse functions. The easiest way to do this is to arrange that a dual circle to the attaching circle also lies in the fiber throughout the homotopy, and since the proof of the relevant part of Theorem 3.6 simply involves counting dimensions and appealing to transversality, this slight improvement is straightforward.

□

The other lemma needed in the proof of Theorem 4.2 is:

**Lemma 4.7** *Given two indefinite Morse 2–functions $G, G' \colon X \to I^2$ such that $z \circ G|_{M_0} = z \circ G'|_{M_0} = \zeta_0$, $z \circ G|_{M_1} = z \circ G'|_{M_1} = \zeta_1$ and $t \circ G = t \circ G' = \tau \colon X \to I$, there exists an $\epsilon > 0$ and an indefinite deformation of Morse 2–functions $G_s \colon X \to I^2$ such that:*

(1) *$t \circ G_s = \tau$ for all $s$, and*

(2) *for each critical value $t_*$ of $\tau$, letting $X_* = \tau^{-1}[t_* - \epsilon, t_* + \epsilon]$, we have $G_1|_{X_*} = G'|_{X_*}$.*

*If $G$ and $G'$ have all fibers connected then we can arrange for $G_s$ to be have all fibers connected as well.*

**Proof** Again, we use Lemma 4.3 to standardize $G = (\tau, \zeta)$ and $G' = (\tau, \zeta')$ near each critical point of $\tau$, so that $\zeta$ and $\zeta'$ are equal in a neighborhood of each critical point. Then use Lemma 4.5 to make $G$ "constant in the $t$ direction" inside each $\tau^{-1}[t_* - \epsilon, t_* + \epsilon]$ but away from the critical point. Then the lemma follows quickly from Theorems 3.6 and 3.10. □

## 5 The main results

Finally we can put the pieces together to prove our main results about indefinite Morse 2–functions over disks, spheres and other surfaces. The result for the disk is a trivial extension of the results for square Morse 2–functions from the preceding section:

**Proof of Theorem 1.1** We are given a connected $n$–manifold $X$ with nonempty boundary, with an indefinite Morse function $g\colon \partial X \to S^1$. Identify $S^1$ with $\partial(I \times I)$ so that there are no critical values in $I \times \{0,1\}$, and then, with respect to this identification, let $M_0 = g^{-1}(\{0\} \times I)$, $M_1 = g^{-1}(\{1\} \times I)$, and, using a gradient-like vector field, identify $g^{-1}(I \times \{0,1\})$ with $I \times g^{-1}(\{0\} \times \{0,1\})$. Then we have $X$ realized as a cobordism from $M_0$ to $M_1$, which are themselves cobordisms, and both $M_0$ and $M_1$ come equipped with indefinite Morse functions $g_0$ and $g_1$, respectively. We apply Theorem 4.1 to this setting to get the extension $G\colon X \to I^2$, and then identify $I^2$ with $D^2$ in a way compatible with the original identification of $S^1$ with $\partial I^2$. This proves the basic existence result.

If the level sets of $g$ are connected, then we carry out the procedure above, but before applying Theorem 4.1, we use Lemma 3.11 to get a generic homotopy from $g_0$ to some $g_0'$ which is ordered, and likewise a generic homotopy from $g_1$ to some $g_1'$ which is ordered, without destroying connectedness of level sets along the way. Then we apply Theorem 4.1 with $g_0'$ and $g_1'$ as input Morse functions, and then tack on the chosen generic homotopies.

The argument to reduce uniqueness to Theorem 4.2 is essentially identical to the existence argument above; the only additional ingredient is the observation that a general Morse 2–function to $I^2$ can always be perturbed by an arbitrarily small perturbation to make it a square Morse 2–function, i.e. to arrange that the projection to the $t$–axis is itself a Morse function. □

The result over $S^2$ follows from the result over $D^2$ together with some subtle variations on the Thom-Pontrjagin construction:

**Proof of Theorem 1.2** Now we are given a closed, connected $X^n$ and a framed submanifold $F^{n-2} \subset X$. The existence of an indefinite Morse 2–function $G\colon X \to S^2$ with $G^{-1}(\text{n.p.}) = F$ then follows as an immediate corollary of Theorem 1.1, by mapping a product neighborhood of $F$ to the arctic cap and extending to all of $S^2$ using Theorem 1.1. If $F$ is connected, then the

$S^1$–valued Morse function which is the input to Theorem 1.1 (just projection $S^1 \times F \to S^1$) has connected level sets, so we get the connectedness of fibers immediately as well.

For the uniqueness result, we are given two homotopic indefinite Morse 2–functions $G_0, G_1 \colon X \to S^2$. Let $F_0 = G_0^{-1}(\text{n.p.})$ and let $F_1 = G_1^{-1}(\text{n.p.})$, with induced framings. Because $G_0$ is homotopic to $G_1$, we know that $F_0$ is framed cobordant to $F_1$. In general this cobordism lies in $I \times X$, but by Lemma 5.1 below we can reduce to the case where there is a connected $(n-1)$–dimensional cobordism $M \subset X$ with $\partial M = -F_0 \cup F_1$ compatible with the given framings on $F_0$ and $F_1$. An indefinite Morse function on $M$ then extends to an indefinite Morse 2–function from a regular neighborhood of $M$ to a disk such that $F_0$ is the (framed) preimage of one point and $F_1$ is the (framed) preimage of a different point in the disk. Extend this using Theorem 1.1 to an indefinite Morse 2–function $G' \colon X \to S^2$. Then the uniqueness part of Theorem 1.1 shows that $G_0$ can be connected to $G'$ via an indefinite generic homotopy and that $G'$ can be connected to $G_1$ via an indefinite generic homotopy, so thus we can get from $G_0$ to $G_1$ via an indefinite generic homotopy. Connectedness of fibers also follows immediately from connectedness of fibers in Theorem 1.1 and the connectedness assertions in Lemma 5.1. □

**Lemma 5.1** *Given two nonempty framed cobordant codimension–2 framed submanifolds $F, F' \subset X$, there is a sequence of framed, nonempty, codimension-2 submanifolds $F = F_0, F_1, ..., F_r = F'$ in X, with $F_1, \ldots, F_{r-1}$ connected, and a sequence of connected, framed, embedded, codimension-1 cobordisms $C_1, C_2, ..., C_r$ in $X$, with each $C_i$ a cobordism from $F_{i-1}$ to $F_i$.*

**Proof** Let $q_n$ be the north pole of $S^2$, let $q_s$ be the south pole and let $A_e$ and $A_w$ be "east" and "west" meridional arcs from $q_n$ to $q_s$, i.e. two halves of a great circle through the poles. Let $G, G' \colon X \to S^2$ be Morse 2–functions with $F = G^{-1}(q_n)$ and $F' = (G')^{-1}(q_n)$ (with framings). Let $G_t$, $t \in I$ be a generic homotopy from $G = G_0$ to $G' = G_1$, which exists because $F$ and $F'$ are framed cobordant. Choose $0 = t_0 < t_1 < \ldots < t_r = 1$, with $r$ even, such that:

(1) For each $G_{t_i}$, both $q_n$ and $q_s$ are regular values, and $G_{t_i}$ is a Morse 2–function.

(2) For each odd $i$, for each $t \in [t_i, t_{i+1}]$, $q_s$ is a regular value of $G_t$.

(3) For each even $i$, for each $t \in [t_i, t_{i+1}]$, $q_n$ is a regular value of $G_t$.

Because $q_s$ remains a regular value throughout each odd interval $[t_i, t_{i+1}]$, its inverse image $G_t^{-1}(q_s)$ moves by an isotopy in each such interval, and likewise for $G_t^{-1}(q_n)$ in each even interval. Thus we can modify $G_t$ by pre-composing with ambient isotopies supported alternately away from the inverse images of $q_n$ and $q_s$ so as to arrange that, on each odd interval, $G_t^{-1}(q_s)$ does not move at all and, on each even interval, $G_t^{-1}(q_n)$ does not move at all.

Ignoring the connectedness assertions, we can now produce the submanifolds and cobordisms in this lemma as follows: Let $F_0 = F = G_{t_0}^{-1}(q_n) = G_{t_1}^{-1}(q_n)$, $F_1 = G_{t_1}^{-1}(q_s) = G_{t_2}^{-1}(q_s)$, $F_2 = G_{t_2}^{-1}(q_n) = G_{t_3}^{-1}(q_n)$, up to $F_{r-1} = G_{t_{r-1}}^{-1}(q_s) = G_{t_r}^{-1}(q_s)$, $F_r = F' = G_{t_r}^{-1}(q_n)$. Then let $C_1 = G_{t_1}^{-1}(A_e)$, $C_2 = G_{t_2}^{-1}(A_w)$, $C_3 = G_{t_3}^{-1}(A_e)$, up to $C_r = G_{t_r}^{-1}(A_w)$.

We now need to make the $F_i$'s and $C_i$'s connected. First, if any $C_i$ has any closed components, discard them. Then if we make each $F_i$ connected, each $C_i$ will automatically be connected.

Note that we can think of $C = C_1 \cup C_2 \cup \ldots \cup C_r$ as an immersed smooth codimension–1 submanifold with nonempty boundary $\partial C = (-F_0) \cup F_r$. ($C_i$ and $C_{i+1}$ join smoothly along $F_i$ because we alternately used the arcs $A_e$ and $A_w$ in their definitions.) Our first step is to modify the $C_i$'s so that $X \setminus C$ is connected. Let $X_0$ be a component of $X \setminus C$ containing some part of $F_0$ in its closure. Choose an arc $a$ connecting $X_0$ to an adjacent component, but intersecting $C$ transversely at two points, with opposite signs. In other words $a$ goes through some $C_i$, to connect the two components, and also through $C_1$ somewhere near $F_0$ from the appropriate side of $C_1$. If $C_i = C_1$, then we can just self-connect sum $C_1$ along this arc, opening up a tunnel between the two components. If $C_i \neq C_1$, then we still perform the connect sum, but we subdivide the connect sum tube into multiple tubes ($I \times S^{n-2}$'s), consecutively considered as components of $C_1, C_2, \ldots, C_i$. This creates new $S^{n-2}$ components of $F_1, \ldots, F_{i-1}$, but also decreases the number of components of $X \setminus C$, and does not change $F_0$ or $F_r$. Repeat this process until $X \setminus C$ is connected.

Now, for each disconnected $F_i$, we can find arcs connecting the components of $F_i$, meeting $C$ only at their endpoints (in $F - i$) and transverse to $C$ at these endpoints. Now perform a self-connect sum of $C$ along each of these arcs, but note that the connect sum tube splits into two halves, one naturally in $C_{i-1}$ and one naturally in $C_i$. The result is that the components of $F_i$ are connected via band-connect sums and the $C_i$'s are modified via boundary connect sums.

□

We end with some observations about sections.

**Theorem 5.2** *If, in the existence part of Theorem 1.1 or Theorem 1.2, respectively, we are given a 2–dimensional submanifold $S$ (diffeomorphic to $D^2$ or $S^2$, respectively) which intersects the fiber $F$ once, we can construct the indefinite Morse 2–function so that $S$ is a section. If, in the uniqueness parts, the two given Morse 2–functions agree on a common section $S$ and they are homotopic via some homotopy that does not move $S$, then we can arrange that the indefinite homotopy we produce remains fixed on $S$. In both cases, these assertions do not destroy the assertions about connectedness of fibers. The same results also hold when $S$ is a union of several disjoint sections.*

**Proof** The idea is to delete a neighborhood of $S$, and prove appropriately relative versions of all the results involved in the proofs of Theorem 1.1 and 1.2. In Theorems 3.2, 3.6 and 3.10, it is straightforward to generalize to setting where the $m$–manifold $M$ is a cobordism between cobordisms, i.e. a cobordism with sides. Then in Theorem 4.1, we use these results applied to both the complement of a neighborhood of $S$ in the $n$–manifold $X$ and the complement of a neighborhood of $S \cap M_t$ in the $(n-1)$–manifolds $M_t$. The same idea extends immediately to Theorem 4.2. These theorems about maps to the square are essentially equivalent to Theorem 1.1, about maps to $D^2$.

As for maps to $S^2$, we now only need to verify that the section $S$ does not interfere with the arguments in Theorem 1.2 related to the Thom-Pontrjagin construction. There is clearly no problem with the existence part. For the uniqueness part, once we have the topological fact that there exists some homotopy not moving $S$ and the fact that $S$ is codimension at least 2, the rest of the construction works exactly the same.

□

**Remark 5.3** Note that, by removing neighborhoods of sections in the above proof, we are essentially working with generalizations of open book decompositions as given boundary conditions, rather than just with $S^1$–valued Morse functions. Thus there are natural reformulations of all of our results in this more general relative setting.

# References


[1] Selman Akbulut and ÇağrıKarakurt. Every 4-manifold is BLF. *J. Gökova Geom. Topol. GGT*, 2:83–106, 2008.

[2] Daniel Asimov. Round handles and non-singular Morse-Smale flows. *Ann. of Math. (2)*, 102(1):41–54, 1975.

[3] Denis Auroux, Simon K. Donaldson, and Ludmil Katzarkov. Singular Lefschetz pencils. *Geom. Topol.*, 9:1043–1114, 2005.

[4] R. I. Baykur and N. Sunukjian. Round handles, logarithmic transforms, and smooth 4-manifolds. *ArXiv e-prints*, September 2010.

[5] R. İnanç Baykur. Existence of broken Lefschetz fibrations. *Int. Math. Res. Not. IMRN*, pages Art. ID rnn 101, 15, 2008.

[6] David T. Gay and Robion Kirby. Constructing Lefschetz-type fibrations on four-manifolds. *Geom. Topol.*, 11:2075–2115, 2007.

[7] David T. Gay, Katrin Wehrheim, and Chris Woodward. Connected Cerf theory. In preparation.

[8] Allen Hatcher and John Wagoner. *Pseudo-isotopies of compact manifolds.* Société Mathématique de France, Paris, 1973. With English and French prefaces, Astérisque, No. 6.

[9] Robion Kirby. A calculus for framed links in $S^3$. *Invent. Math.*, 45(1):35–56, 1978.

[10] Yanki Lekili. Wrinkled fibrations on near-symplectic manifolds. *Geom. Topol.*, 13(1):277–318, 2009. Appendix B by R. İnanç Baykur.

[11] Tim Perutz. Lagrangian matching invariants for fibred four-manifolds. I. *Geom. Topol.*, 11:759–828, 2007.

[12] Tim Perutz. Lagrangian matching invariants for fibred four-manifolds. II. *Geom. Topol.*, 12(3):1461–1542, 2008.

[13] Osamu Saeki. Elimination of definite fold. *Kyushu J. Math.*, 60(2):363–382, 2006.

[14] Gordon Wassermann. Stability of unfoldings in space and time. *Acta Math.*, 135(x):57–128, 1975.

[15] Jonathan Williams. The h-principle for broken Lefschetz fibrations. *Geom. Topol.*, 14:1015–1061, 2010.