Indefinite Morse 2–functions; broken fibrations and generalizations

DAVID T. GAY, ROBION KIRBY *

Euclid Lab, 160 Milledge Terrace, Athens, GA 30606
Department of Mathematics, University of Georgia, Athens, GA 30602
and
University of California, Berkeley, CA 94720

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

Abstract

A Morse 2–function is a generic smooth map from a smooth manifold to a surface. 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 arbitrary compact, oriented surfaces. "Uniqueness" means there is a set of moves which are sufficient to go between two homotopic indefinite Morse 2–functions while remaining indefinite throughout. We extend the existence and uniqueness results to indefinite, Morse 2–functions with connected fibers.

AMS Classification numbers   Primary: 57M50

Secondary: 57R17, 57R45

Keywords: broken fibration, Morse function, Cerf theory, definite fold, elliptic umbilic

1 Introduction

A Morse 2–function on a smooth n–manifold X is a generic smooth map from X to a 2–manifold, just as an ordinary Morse function is a generic smooth map to a 1–manifold. The singularities are *folds* and *cusps*. Folds look locally like $(t, x_1, \ldots, x_{n-1}) \mapsto (t, f(x_1, \ldots, x_{n-1}))$ for a standard Morse singularity f, and cusps look locally like $(t, f(x_1, \ldots, x_{n-1})) \mapsto (t, f_t(x_1, \ldots, x_{n-1}))$ for a standard birth f_t of a cancelling pair of Morse singularities.

We develop techniques for working with Morse 2–functions and generic homotopies between them, paying particular attention to (1) avoiding *definite folds*, in which the modelling function f is a definite Morse singularity, i.e. a local extremum, and (2) guaranteeing connected fibers. When definite folds are avoided, we say that the Morse 2–function (or generic homotopy) is *indefinite*. When fibers are connected, we say that the function or homotopy is *fiber-connected*. The same adjectives also describe Morse functions and their homotopies, when definite singularities (local extrema) are avoided and when level sets are connected.

Let X be a compact, connected, oriented, smooth (C^∞) n–manifold and let Σ be a compact, connected, oriented surface (with possibly empty, possibly disconnected, boundaries).

Theorem 1.1 (Existence) *Let $g : \partial X \to \partial \Sigma$ be an indefinite, surjective, Morse function which extends to a map $G' : X \to \Sigma$. If $n > 2$ and $G'_*(\pi_1(X))$ has finite index in $\pi_1(\Sigma)$, then G' is homotopic rel. boundary to an indefinite Morse 2–function $G : X \to \Sigma$. When $n > 3$, if g is fiber-connected and $G'_*(\pi_1(X)) = \pi_1(\Sigma)$ then we can arrange that G is fiber-connected.*

Theorem 1.2 (Uniqueness) *Let $G_0, G_1 : X \to \Sigma$ be indefinite Morse 2–functions which agree on ∂X and are homotopic rel. boundary. If $n > 3$ then G_0 and G_1 are homotopic through an indefinite generic homotopy G_s. If in addition G_0 and G_1 are fiber-connected then we can also arrange that G_s is fiber-connected.*

These results are analogs of the following facts in ordinary Morse theory:

Theorem 1.3 (Existence) *A compact, connected m–dimensional cobordism M between $F_0 \neq \emptyset$ and $F_1 \neq \emptyset$ supports an indefinite Morse function $g : M \to I = [0,1] = B^1$. If $m > 2$ and F_0 and F_1 are connected then we can also arrange that g is fiber-connected. Any homotopically nontrival map g' from*

2

a closed connected m–manifold M to S^1 is homotopic to an indefinite Morse function $g : M \to S^1$. If $m > 2$ and $g'_*(\pi_1(X)) = \pi_1(S^1)$ then we can also arrange that g is fiber-connected.

Theorem 1.4 (Uniqueness) *In both cases above, two homotopic (rel. ∂) indefinite Morse functions $g_0, g_1 : M \to N^1$, where $N = B^1$ or $N = S^1$, are homotopic through an indefinite generic homotopy g_s, and if $m > 2$ and g_0 and g_1 are fiber-connected then we can arrange that g_s is fiber-connected for all s.*

Remark 1.5 Because our proofs of the above theorems begin with the case of maps to $I \times I$ or I and are completed with Thom-Pontrjagin type arguments, all the homotopies constructed can be taken to be homotopic to given homotopies.

The motivation for generalizing Morse functions and Cerf theory [6] from dimension one to dimension two comes originally from the importance of Lefschetz fibrations in complex and symplectic geometry, and the new ideas around broken Lefschetz fibrations. After LeBrun [14] and Honda [11, 10] showed that a smooth 4-manifold X^4 with $b_2^+ > 0$ has a *near symplectic form* (a closed 2-form ω with $\omega \wedge \omega \geq 0$ and zero only on a smooth embedded 1-manifold Z in X), Auroux, Donaldson and Katzarkov [3] proved that such 4-manifolds are Lefschetz pencils in the complement of Z, where Z mapped onto latitudes of S^2. Gay and Kirby [7], Lekili [15], Baykur [5] and Akbulut and Karakurt [1] extended this result to all smooth, oriented, compact 4-manifolds as broken fibrations (not just pencils).

One aim was to define invariants by counting pseudo-holomorphic curves in X^4 which limit on Z, as had been done in the symplectic case by Taubes [19] and Usher [21]. To do this, Perutz [16, 17] defined his Lagrangian matching invariants for a broken Lefschetz fibration, but to get invariants of the underlying smooth 4-manifold, one needs *moves* between broken Lefschetz fibrations, preserving connectedness of fibers, under which the Lagrangian matching invariants are preserved. Our uniqueness theorem is intended to provide these moves, and thus provide purely topological definitions of invariants.

Note that maps $X^4 \to S^2$ are partitioned according to their homotopy class into the elements of the cohomotopy set $\pi^2(X^4)$, calculated homotopically by Taylor [20] and geometrically by Kirby, Melvin and Teichner [13]. It is not clear how this partitioning relates to known invariants, but all elements are realized by indefinite, fiber connected, Morse 2-functions.

Theorem 1.1, in the case of $\Sigma = S^2$ and without fiber-connectedness, is originally due to Saeki [18], who also pointed out that the finiteness of the index $[G'_*(\pi_1(X)) : \pi_1(\Sigma)]$ is a necessary condition. A short proof of existence for closed X to S^2 is sketched in [8]. A significant step forward in the uniqueness case was provided by Lekili [15] when he reintroduced singularity theory into the subject and showed how to go back and forth between Lefschetz singularities and cusps on fold curves.

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.

Theorem 1.2, when $n = 4$ and without fiber-connectedness, was originally proved by Williams [23]. Theorem 1.3 is completely standard. The B^1–valued (cobordism) case of Theorem 1.4 is an essential ingredient in developing the calculus of framed links for 3–manifolds and thus appears in [12]. It seems that the fiber-connected assertion in the S^1–valued case of Theorem 1.4 is a new result, and was originally posed to us as a question by Katrin Wehrheim and Chris Woodward.

If we remove the adjectives "indefinite" and "fiber-connected" from the above theorems then 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 Theorems 1.1 and 1.2, we spend most of our time on the case where $\Sigma = B^2$, the disk, and in fact think of B^2 as the square $I^2 = I \times I$. Here the natural structure on the n–dimensional domain X of a Morse 2–function $G : X \to I^2$ 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. See Figure 1.

Consider the special case where $X = [0,1] \times M$ and $G(t,p) = (t, g_t(p))$, a generic homotopy between Morse functions $g_0, g_1 : M \to [0,1]$. Then removing definite folds from G is the same as removing definite critical points from g_t,
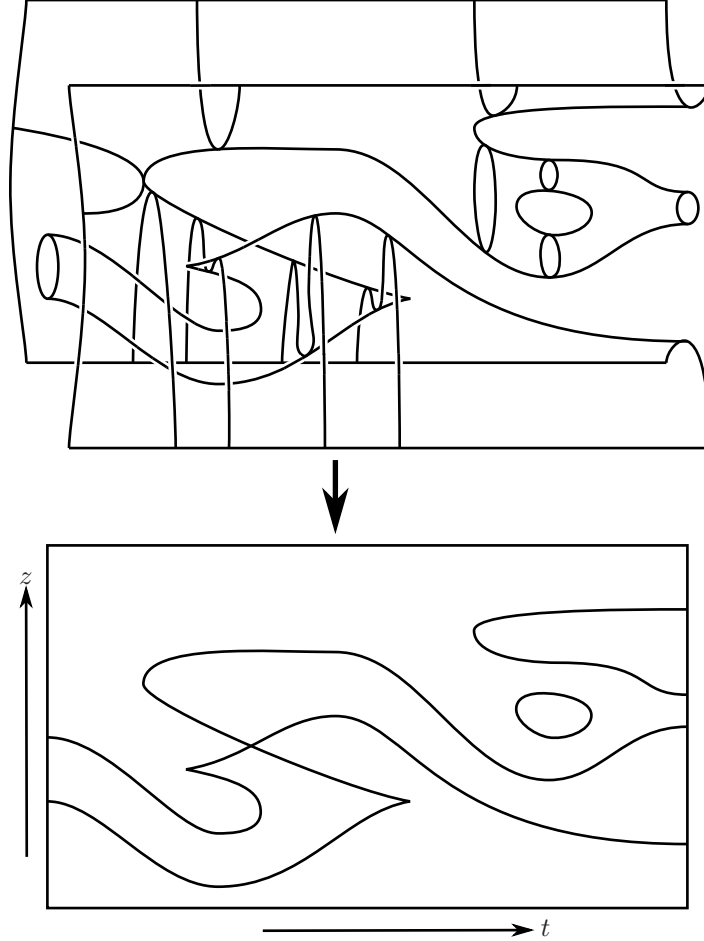
Figure 1: A Morse 2–function on a surface, mapping to the square $I \times I$.

and this is done in Section 4 and also in [12]. Fibers remain connected, and therefore existence is done in this special case. For uniqueness, suppose we have a generic 2–parameter family $g_{s,t}$ between $g_{0,t}$ and $g_{1,t}$, giving a generic homotopy $G_s(t,p) = g_{s,t}(p)$. The 2–dimensional definite folds are shown to be removable in Section 4 by use of singularity theory, in particular, use of the codimension two singularities called the *butterfly* and the *elliptic umbilic*.

For existence in the general case of a cobordism $(X, M_0, M_1) \rightarrow (I \times I, \{0\} \times I, \{1\} \times I)$, choose a Morse function $\tau : X \rightarrow I$ with no definite critical points. We will construct the indefinite Morse 2–function G so that $t \circ G = \tau$, where (t, z) are coordinates on $I \times I$. Choose times t_a and t_b just before and after
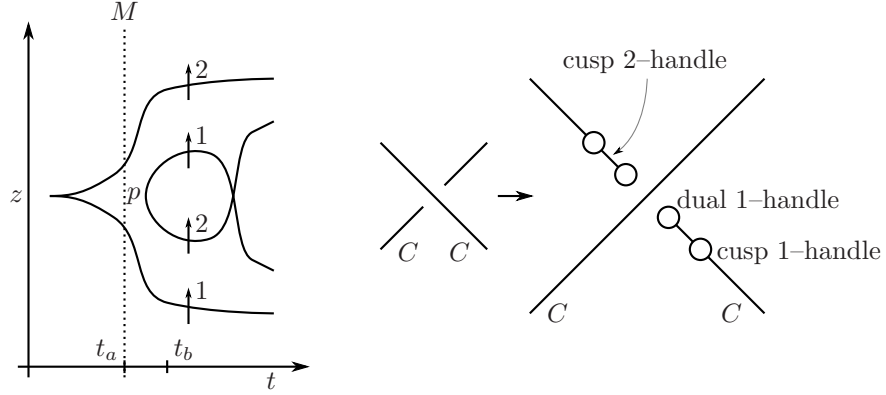
5

Figure 2: Resolving crossings.

a critical value of τ. At the critical point there is a standard function in local coordinates giving the descending sphere. Choose a z–valued Morse function ζ_a on the slice $\tau^{-1}(t_a)$ such that the descending sphere lies in a level set $\zeta_a^{-1}(z_a)$, which will mean that the descending sphere lies in the fiber of G over (t_a, z_a). Away from the local coordinates, we essentially have a product between $\tau^{-1}(t_a)$ and $\tau^{-1}(t_b)$ and the Morse function on $\tau^{-1}(t_a)$ determines a Morse function on $\tau^{-1}(t_b)$.

To see how to find ζ_a, an interesting and illustrative case is that of a (horizontal) Morse function $\tau : X^4 \to R$, a critical point p of index 2, a second Morse function $\zeta : M_a^3 \to R$ where M_a is the level set of τ at t_a just below $\tau(p)$, and the attaching circle C of the descending disk from p lying in M_a, but not in a level set of ζ. But we wish it to lie in a level set, for that will be a fiber of the eventual Morse 2-function. See Figure 2.

First we can isotope C into a region $[-1, 1] \times L$ where L is a level set of M_a with respect to ζ. After generically projecting C into $0 \times L = L$, we get crossings, and these have to be resolved somehow. For each crossing a cusp consisting of a cancelling 1-2–pair of critical points must be created at a lower level of τ, that is, before trying to embed C in F, so that they are available to remove the crossing.

Then to construct ζ on $M_b = \tau^{-1}(t_b)$ near the crossing, we modify the ζ on M_a as follows. The 1-handle is attached first, on either side of the arc of C which is the under crossing. Then C is added, with the crossing being resolved by sending one strand of C over the 1-handle. Next we must add the dual 1-handle of C. This is attached to the 0-sphere bundle (which is the boundary of the normal line bundle to C in F), and this 0-sphere can be placed where

6

convenient. In our case it is as drawn in Figure 2. Finally, the 2-handle of the cusp, which must cancel its 1-handle, goes over each of the 1-handles once, as drawn. This describes how ζ changes locally at each crossing while shifting from M_a to M_b. At each crossing of C in F, a cusp is added and the genus of the fiber F is raised by one.

A similar argument works in higher dimensions, and is needed to cover the case when the dimension of X is even, $2k$, and a k-handle is being attached and needs to have its attaching $k-1$-sphere lie in the fiber whose dimension is $2k-2$. Again, there are isolated points of self intersection of the attaching sphere, and these are resolved by adding a cusp for a cancelling $(k-1)-k$-pair. Then we first add the $(k-1)$-handle from the cusp, next the attaching $k-1$-sphere has one of the two intersecting $(k-1)$-disks go over the $(k-1)$-handle just added, next the dual $(k-1)$-handle, and finally the cusp's k-handle to cancel its partner. At each such step the fiber is changed by connected sum with $S^{k-1} \times S^{k-1}$.

We finish the existence outline by filling in the gaps between the strips around critical values by choosing Cerf graphics without definite folds; i.e. appealing to the existence of generic homotopies between ordinary Morse functions without definite critical points. Thus we get images of the folds in $I \times I$ as in Figure 3.
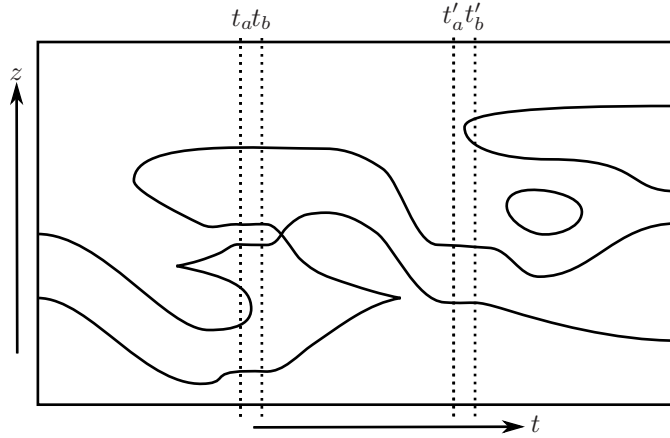

Figure 3: An example illustrating a Cerf graphic in between two critical values of $\tau = t \circ G$.

Now the same ideas are used for uniqueness in Section 5. We are given two maps $G_0, G_1 : (X, M_0, M_1) \to (I \times I, \{0\} \times I, \{1\} \times I)$ and $\tau_0 = t \circ G_0$ and $\tau_1 = t \circ G_1$ are homotopic indefinite Morse functions. Appealing again to the existence of

indefinite homotopies between Morse functions, we get a generic homotopy τ_s from τ_0 to τ_1 with no definite critical points. This homotopy is a sequence of indefinite births, followed by changes of heights of critical values, followed by deaths. We first construct the homotopy G_s for $s \in [0, 1/4] \cup [3/4, 1]$ so as to arrange that $t \circ G_s$ realizes the appropriate births and height changes for $s \in [0, 1/4]$ and the appropriate deaths for $s \in [3/4, 1]$. For example, a birth is achieved by the introduction of an eye followed by a kink, as in Figure 33 (see Section 5). Then we have arranged that $t \circ G_{1/4} = t \circ G_{3/4}$ and we construct G_s for $s \in [1/4, 3/4]$ keeping $t \circ G_s$ fixed. Here we end up appealing again to the existence of indefinite 2–parameter homotopies between ordinary Morse functions, as in Section 4, where now the Morse functions are of the form $z \circ G|_{G^{-1}(\{t\} \times I)}$.

Once the case of image equal to $I \times I$ is done, it is not hard using a *zig-zag* argument to extend the theorems to other surfaces (see Section 6). Extra care with Thom-Pontrjagin type arguments is needed to keep fibers connected.

Although our motivation comes from the $n = 4$ case, the arguments all generalize in a straightforward manner to all dimensions, save occasionally troubles with dimensions ≤ 3. It is hoped that in other dimensions, even 3, the theorems will be useful.

Remark 1.6 In this paper we have assumed X is 0–connected and we have removed definite folds (0– and $(n-1)$–folds) when $n \geq 3$, and removed them in 1–parameter families when $n \geq 4$. One could speculate that, if X is 1–connected, we could remove 0–, 1–, $(n-2)$– and $(n-1)$–folds for $n \geq 5$ (existence) or $n \geq 6$ (uniqueness). In particular, simply connected 5–manifolds would have only 2–folds and no cusps. This speculation would further generalize to k–connected n–manifolds with $n \geq 2k + 3$ (existence) or $n \geq 2k + 4$ (uniqueness).

We would like to thank Denis Auroux, İnanç Baykur, Michael Freedman, Yankı Lekili, Kevin Walker, Katrin Wehrheim and Jonathan Williams for helpful discussions during the preparation of this paper.

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 μ_k instead of μ_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 an oriented 1–manifold N, a smooth function $g : M \to N$ 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 $g : M \to N$ 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 [9]. To see these results in the more general context of singularity theory, look at [22]. For a modern exposition explicitly in a low dimensional setting, which also explains much of the motivation for this paper, we recommend [15].

We want to discuss homotopies $g_t : M \to N$ between Morse functions which are not necessarily Morse at intermediate times, in which case it is useful to discuss also the associated function $G : I \times M \to I \times N$ 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 $g_0, g_1 : M \to N$, we are interested in homotopies $g_t : M \to N$ 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 4 illustrates these by drawing the Cerf graphic $G(Z_G)$ for a typical generic homotopy):

(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*.
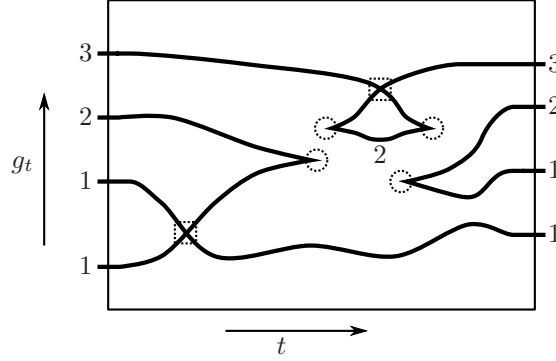
Figure 4: Example of the Cerf graphic for a generic homotopy between Morse functions, indices of critical points labelled with integers, birth and death cusps indicated by dotted circles and critical value crossings indicated by dotted squares.

(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
$$g_t(x_1, \ldots, x_m) = -x_1^2 - \ldots - x_k^2 + x_{k+1}^3 - (t - t_*)x_{k+1} + x_{k+2}^2 + \ldots + x_m^2,$$
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 $g_t : M \to N$, 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.

10

Next we discuss homotopies $g_{s,t} : M \to N$ 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 $G_s : I \times M \to I \times N$ defined by $(t, p) \mapsto (t, g_{s,t}(p))$ and $\mathcal{G} : I \times I \times M \to I \times I \times N$ 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 $g_{0,t} : M \to N$ between Morse functions $g_{0,0}$ and $g_{0,1}$ and another generic homotopy $g_{1,t} : M \to N$ between Morse functions $g_{1,0}$ and $g_{1,1}$, we are interested in connecting these through a 2–parameter family $g_{s,t} : M \to N$, 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 (or has a birth or death in the parameter t) 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 5; we call this event a *Reidemeister-II fold crossing*.
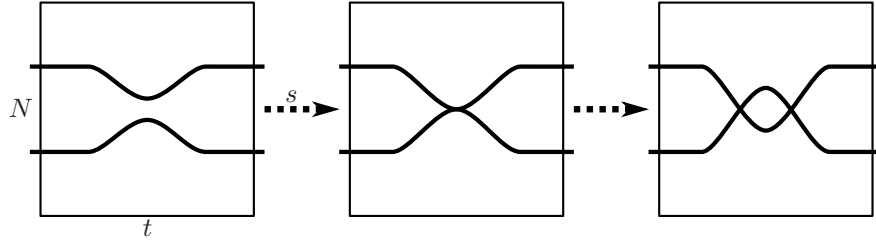

Figure 5: Non-transverse double point in the singular locus for a generic homotopy between generic homotopies between Morse functions. Note that in general the indices of the two critical points involved can be arbitrary.

(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 6; we call this event a *Reidemeister-III fold crossing*.
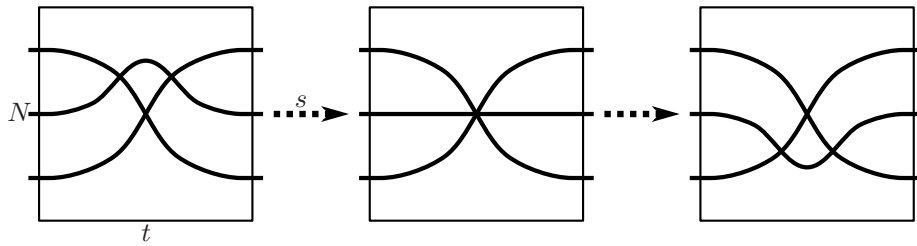

Figure 6: Transverse triple point in the singular locus for a generic homotopy between generic homotopies between Morse functions. Again, the indices involved can be arbitrary.

(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

12

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 $I \times N$ 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 7; we call this event a *cusp-fold crossing*.
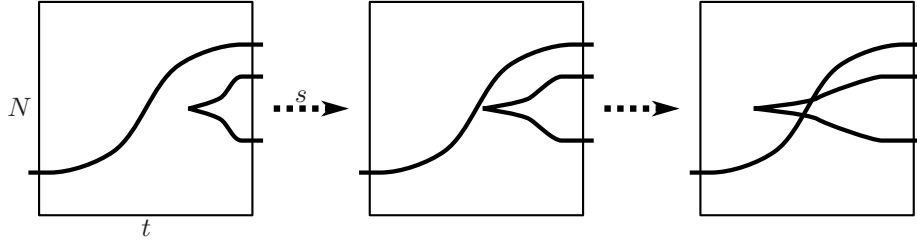

Figure 7: Non-transverse double point involving a cusp, occurring in a generic homotopy between generic homotopies between Morse functions. The only constraint on indices is that coming from the cusp, namely that the two critical points born at the cusp are of successive index.

(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 8; we call this event an *eye birth singularity* (or *death* when s is reversed).

(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 9; we call this event a *merge singularity* (or *unmerge* when s is reversed).
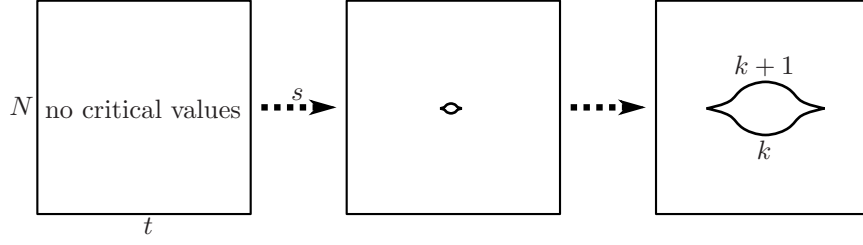
13

Figure 8: The birth of a birth–death pair (eye birth) involving a pair of cancelling critical points of successive index.
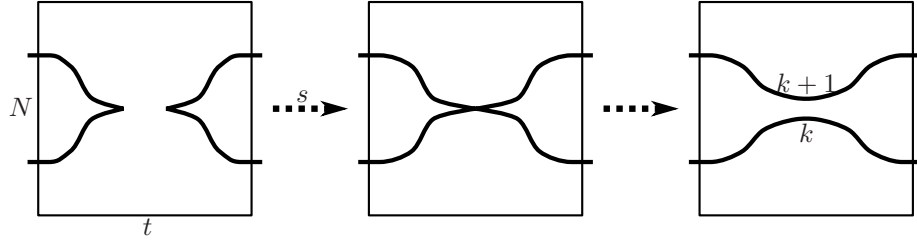
Figure 9: The merge of a death–birth pair involving a pair of cancelling critical points of successive index.

(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 $g_{s,t}(x_1, \ldots, x_m) = -x_1^2 - \ldots - x_k^2 + x_{k+1}^4 - (s-s_*)x_{k+1}^2 + (t-t_*)x_{k+1} + x_{k+2}^2 + \ldots + x_m^2$. 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 10; we call this event a *swallowtail birth singularity* (or *death* when s is reversed).
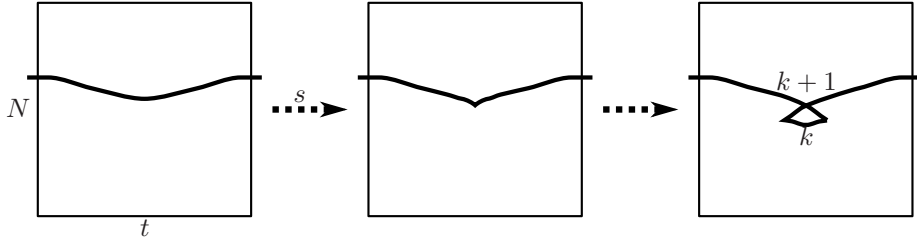
Figure 10: Birth of a swallowtail; an upside down version also occurs.

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 Σ, a smooth proper map $G : X \to \Sigma$ is a *Morse 2–function* if for each $q \in \Sigma$ there is a compact neighborhood S of q with a diffeomorphism $\psi : S \to I \times I$ and a diffeomorphism $\phi : G^{-1}(S) \to I \times M$, for an $(n-1)$–manifold M, such that $\psi \circ G \circ \phi^{-1} : I \times M \to I \times I$ is of the form $(t, p) \mapsto (t, g_t(p))$ for some generic homotopy between Morse functions $g_t : M \to I$. 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 Σ 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 $G|_{M_i} : M_i \to N_i$ a Morse function. When Σ 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 11, 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.
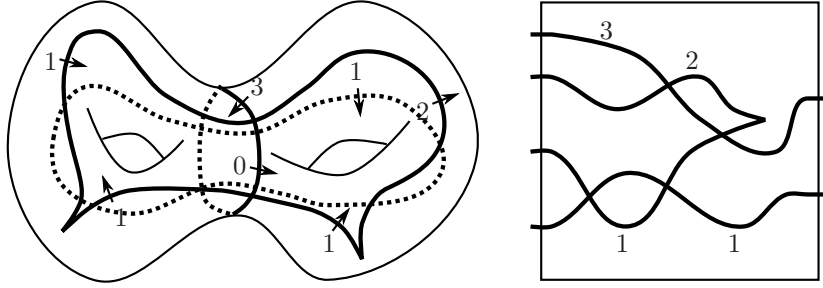

Figure 11: Morse 2–functions versus Cerf graphics, and index labelling conventions. On the left, we are mapping from a 4–manifold to a genus 2 surface. On the right, we are illustrating a generic homotopy between I–valued Morse functions on a 3–manifold.

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

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 [15].

16

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 $G : X \to I^2$, respecting the cobordism structure, with no critical values in $I \times \{0, 1\}$, such that the projection onto the horizontal axis, $t \circ G : X \to I$, is itself a Morse function. In particular, there is a parametrization of the sides $G^{-1}(I \times \{0, 1\})$ as $I \times (F_{00} \amalg F_{10})$ with respect to which the horizontal projection $t \circ G$ restricts as projection to I. Homotopies between square Morse 2–functions are assumed to maintain this last condition.

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 $g : M \to I$ 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 $g : M^m \to I$, 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 An extended example

An important example in dimension 4, first described in Section 8.2 of [3], concerns Morse 2–functions over S^2 with some fibers being torus fibers. Because $\mathrm{Diff}(S^1 \times S^1)$ is not simply connected, a neighborhood $B^2 \times S^1 \times S^1$ of a torus fiber can be removed and glued back in via a nontrivial loop in $\mathrm{Diff}(S^1 \times S^1)$, i.e. by performing a logarithmic transform on the fiber. (See also [4].) Thus we may change the 4–manifold without changing the data of folds, fibers and attaching maps on S^2. The example in [3] involves, in particular, an indefinite Morse 2–function $S^4 \to S^2$ which is homotopically trivial and can be obtained from a definite Morse 2–function $S^4 \to B^4 \to B^2 \hookrightarrow S^2$ by flipping a circle of 0–folds to a circle of 1–folds as illustrated in Figure 16. In this section, we show in detail how this happens in dimension 3, in which case the fiber in question is $S^0 \times S^1$, we show how the nontrivial loop arises, and explain that the example generalizes to arbitrary dimensions $n \geq 3$, with fiber $S^{n-3} \times S^1$. This example also illustrates the important ideas, used throughout the paper, associated with thinking of a disk B^2 in the base as $I \times I$.

We begin with the following simple example of a Morse 2–function $G : S^1 \times \mathbb{R}^2 \to \mathbb{R}^2$: Using cartesian coordinates on $\mathbb{R}^2$ in $S^1 \times \mathbb{R}^2$ and polar coordinates on the range $\mathbb{R}^2$, G is defined by $G(\theta, x_1, x_2) = (1/2 + (x_1^2 + x_2^2)/2, \theta)$. The singular set is a single circle of definite folds at $S^1 \times \{(0,0)\}$ and is embedded via G into $\mathbb{R}^2$ as the circle of radius $1/2$. Figure 12 illustrates this map by showing the image of the fold locus as a dark circle, with paraboloid fibers over rays emanating from the origin, showing clearly that the total space is $S^1 \times \mathbb{R}^2$.

Now let B be the square $[-1,1] \times [-1,1] \subset \mathbb{R}^2$ and let $X = G^{-1}(B)$. Then X is a solid torus seen as a cobordism from $G^{-1}(\{-1\} \times [-1,1])$ to $G^{-1}(\{1\} \times [-1,1])$, both of which are diffeomorphic to $[-1,1] \times S^1$, and G is a Morse 2–function from X to B. The first row in Figure 13 illustrates "vertical slices" of this map, i.e. the inverse images of vertical line segments $\{t\} \times [-1,1]$; the reader should take a moment to reconcile this with Figure 12, which shows the inverse images of rays from the origin. Figures 13 and 14 then illustrate a homotopy G_s beginning with the map $G_0 = G$ described above. Each row of surfaces
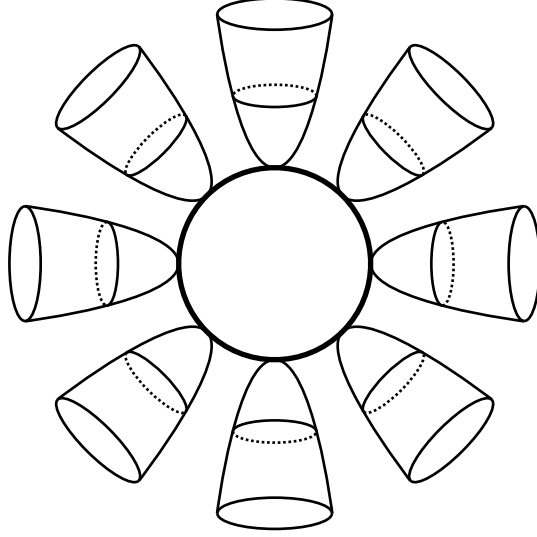
Figure 12: A Morse 2–function $G : S^1 \times \mathbb{R}^2 \to \mathbb{R}^2$ with a single definite fold circle.

illustrates G_s for a fixed s, beginning with G_0 in the top row of Figure 13. We have chosen six representative values of s, hence six rows. (Figure 15 enlarges two regions in Figure 14 just to illustrate the detail there carefully.) In each row (corresponding to a fixed value of s), the surfaces illustrated are each of the form $M_{s,t} = G_s^{-1}(\{t\} \times [-1, 1])$, for nine representative values of t ranging from -1 to 1. Each surface $M_{s,t}$ is drawn embedded in $\mathbb{R}^3$; each embedding is such that the function $G_s|_{M_{s,t}} : M_{s,t} \to \{t\} \times [-1, 1]$ is the height function, projection to the z–axis. Thus the critical locus of each G_s can be seen as the trace of the critical points of each $G_s|_{M_{s,t}}$ as t ranges from -1 to 1.

We can extract from these figures two related sequences of diagrams. The first, Figure 16, illustrates the images of the singular loci in the base $[-1, 1] \times [-1, 1]$ for G_s for each of the six values of s in Figures 13 and 14. The second, Figure 17, indicates how the respective ascending and descending manifolds of the critical points, for the vertical height functions $G_s|_{M_{s,t}}$, intersect the middle-level 1–manifold in each surface $M_{s,t}$ from Figures 13 and 14. Here we have only drawn the diagrams for the final five values of s and for the middle five values of t.

There are several key features to note here:

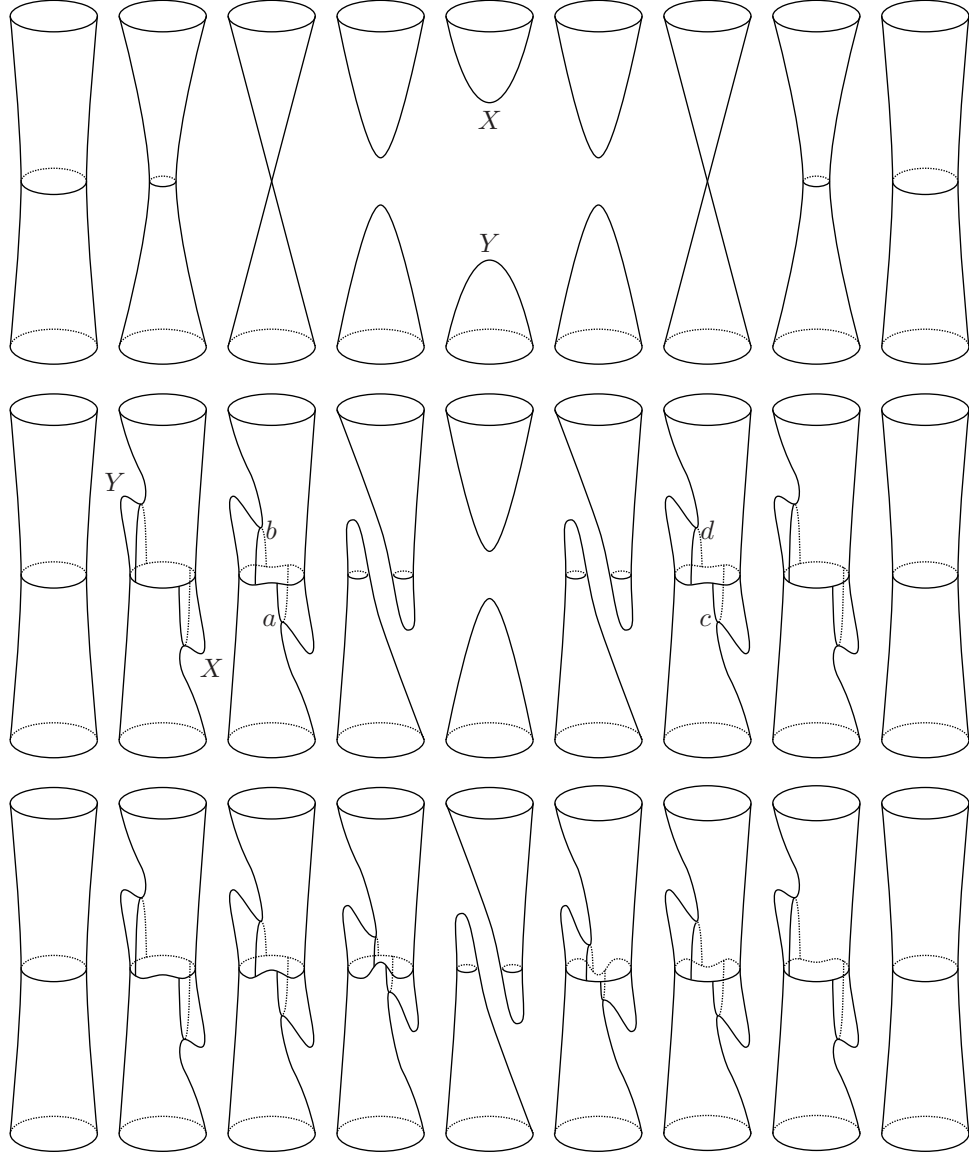(1) For each Morse 2–function G_s, we can consider the function $t \circ G_s$,

Figure 13: The first half of a 1–parameter family of Morse 2–functions on $S^1 \times B^2$. Critical points are labelled for coordination with following figures. From the second row to the third row, we have broken the symmetry of the mid-level circle to prepare for the next move, going to the first row of Figure 14. Looking at the three middle surfaces in the bottom row, we see, from left to right, a 3–dimensional 2–handle attached, surgering the mid-level circle to two circles, followed by a 3–dimensional 1–handle which reattaches the two circles.

20
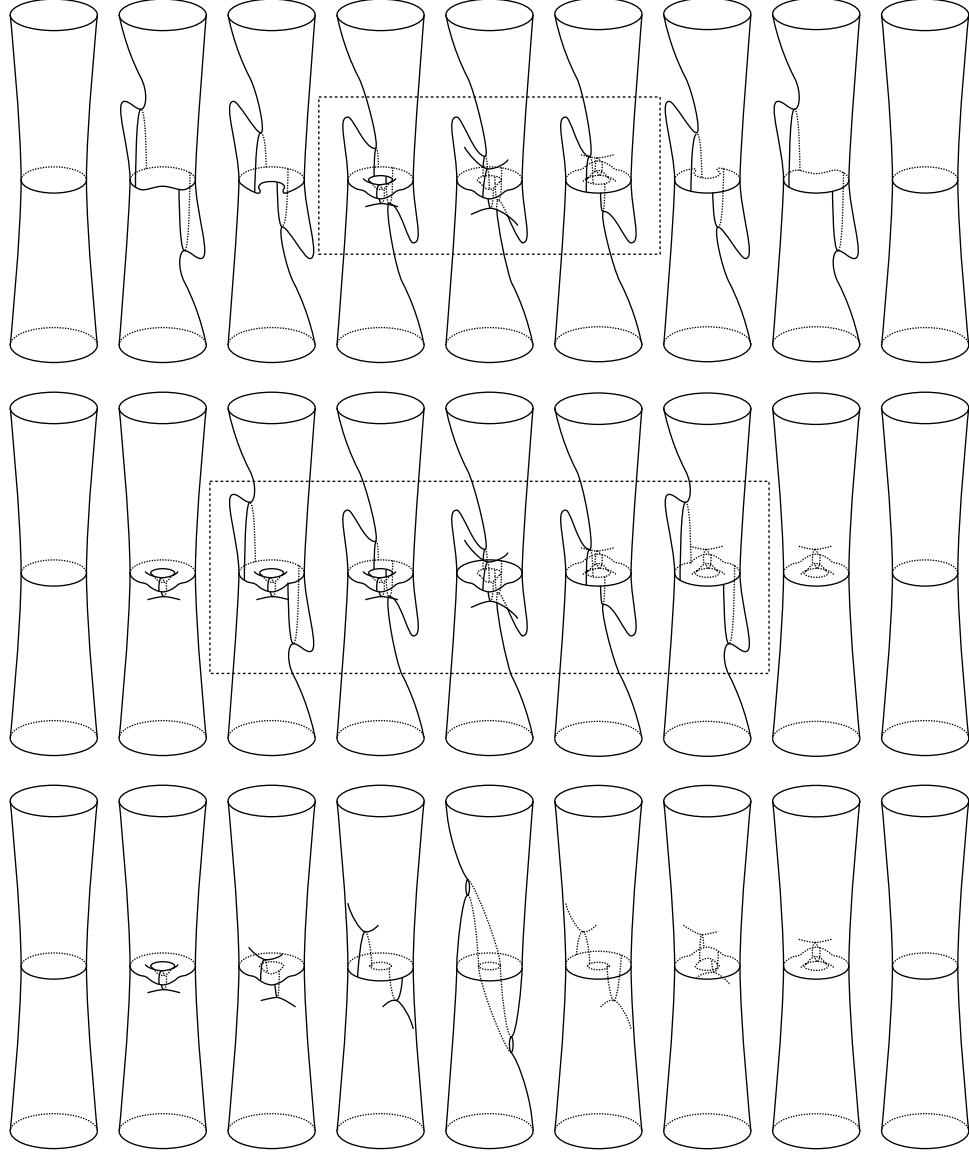
Figure 14: The second half of a 1–parameter family of Morse 2–functions on $S^1 \times B^2$, with definite folds at the beginning and indefinite folds at the end. The two regions enclosed in boxes are shown enlarged in Figure 15. The 3–dimensional 1–handle mentioned in the caption for Figure 13 is seen in the first row between the third and fourth surfaces, while the 2–handle is seen in the first row between the sixth and seventh surface.
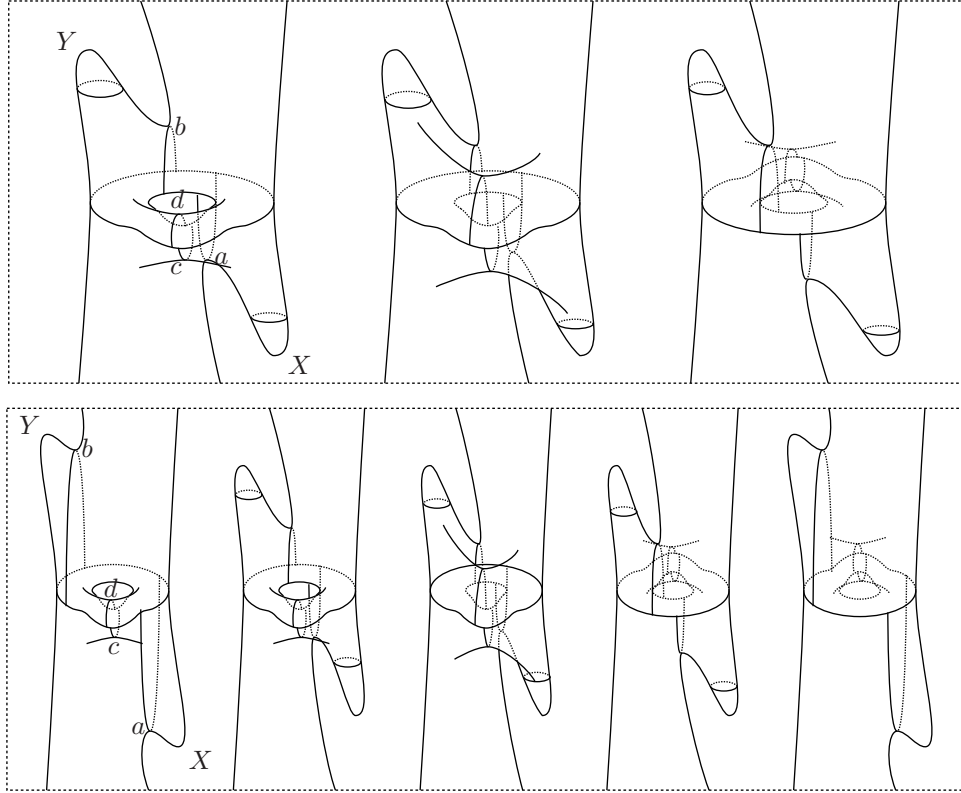
Figure 15: Zooming in on two regions of Figure 14. Critical points are labelled corresponding to labels in Figure 13 and Figure 16. An important point to note here is that, in the second row, going from the first surface to the second, we see two handle slides: b slides over d and a slides under c. And, of course, we have the symmetric slides at the other end of that row.
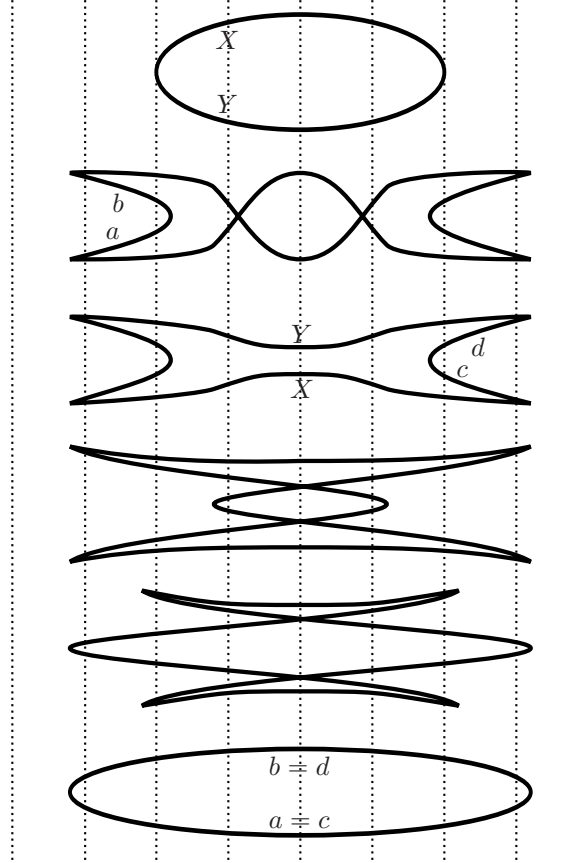
Figure 16: The singular loci of G_s for each of the six values of s corresponding to the rows of Figures 13 and 14. The nine values of t corresponding to the columns are indicated by vertical dotted lines. The correspondence with the critical points in Figures 13 and 14 is indicated by the letter labels. Capital letters X and Y indicate definite folds (2–dimensional 0– and 2–handles) while lower-case letters a, b, c and d indicate indefinite folds (2–dimensional 1–handles). Note that by the end of the homotopy, a and c have become the same fold and b and d have become the same fold; this arises due to the cancellation of the two "swallowtails" at the top and bottom of the preceding singular locus.

23
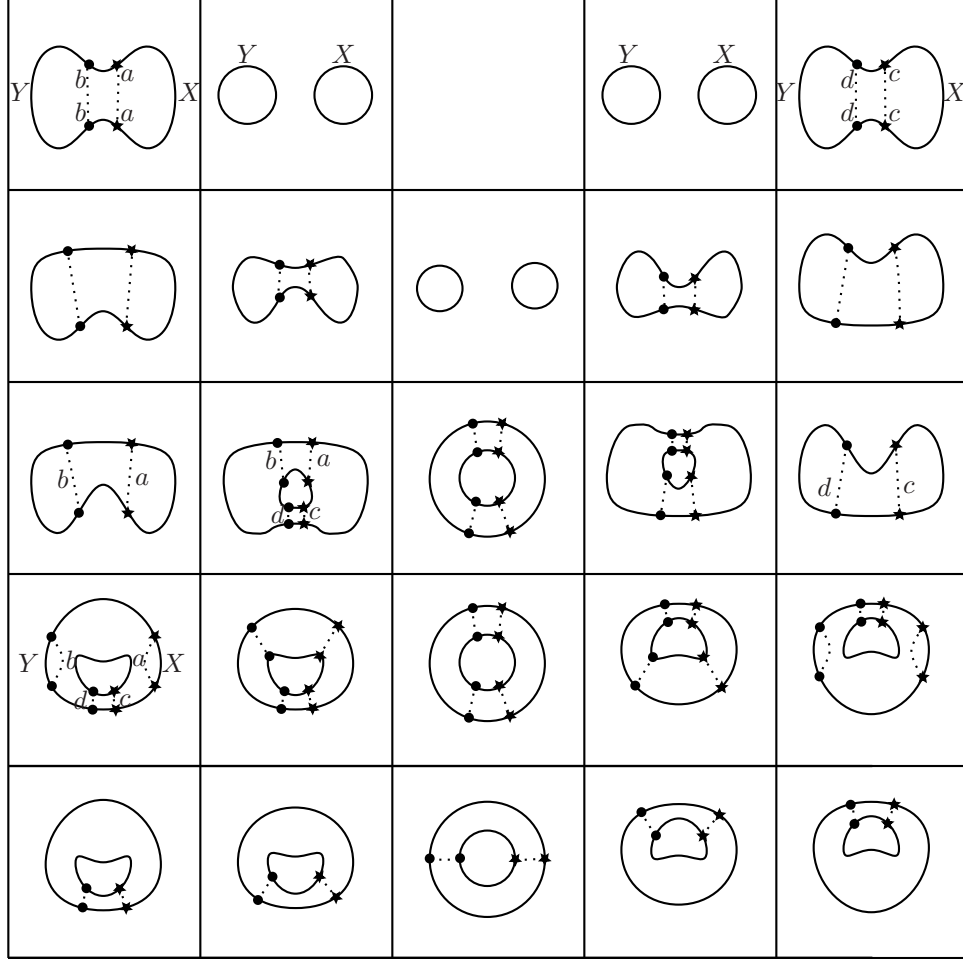
Figure 17: Handle attachment data for a 5×5 block from Figures 13 and 14. In each case we have drawn the middle-level 1–manifold for the surface, and shown where the ascending and descending manifolds of the index 1 critical points (for the vertical height function) intersect this 1–manifold. The circles indicate descending manifolds and the stars indicate ascending manifolds. We have also indicated the ascending manifold for the 0–handle X and the 2–handle Y. Note again the handle slides in the fourth row. By the last row, the 0– and 2–handles have been cancelled, the 1–handles b and d can no longer be distinguished, and the 1–handles a and c can no longer be distinguished.

where $t : \mathbb{R}^2 \to \mathbb{R}$ is projection to the horizontal axis. In our examples, this "horizontal" function is in fact (locally) an ordinary Morse function, which we call the "horizontal Morse function" associated to G_s. The critical points of $t \circ G_s$ occur at precisely the vertical tangencies of the singular loci, as illustrated in Figure 16. These critical points should not be confused with the (2–dimensional) critical points of the vertical Morse function $z \circ G_s|_{M_{s,t}}$ on each surface $M_{s,t}$ in Figures 13 and 14. Looking at how $t \circ G_s$ varies with s, in the beginning, we have a (3–dimensional) critical point of index 2 on the left and a (3–dimensional) critical point of index 1 on the right. By the fourth row, the index 1 critical point has moved to the left and the index 2 critical point to the right.

(2) The first Morse 2–function G_0 has only definite folds, the intermediate functions have both definite and indefinite folds, and the final function G_1 has only indefinite folds.

(3) At G_1, the fiber over points inside the circle of indefinite folds is $S^0 \times S^1$ and, considering rays going out from the center point of the circle, we see, over each such ray, a 2–dimensional 1–handle attached along $S^0 \times \{p\}$, where the point p moves once around the S^1 as the ray rotates once around the center point.

(4) This entire example generalizes to a generic homotopy $G_s : X^n = S^1 \times B^{n-1} \to B^2$. The starting point G_0 is easy to describe in coordinates, exactly as we have done here for the case $n = 3$. The final map G_1 is harder to see in higher dimensions but this example makes it clear that, in the end, we get a circle of indefinite folds with the fiber over points inside the circle being $S^{n-3} \times S^1$. Furthermore, over rays going out from the center we see $(n-1)$–dimensional $(n-2)$–handles attached along $S^{n-3} \times \{p\}$, where p rotates once around S^1 as the ray rotates once around the center point. When $n = 4$, this example was already seen in [3] in their description of a broken fibration of S^4 over S^2. This example also highlights a subtlety involved in reading off information about the total space of a Morse 2–function from data on the base; this subtlety is discussed in more detail in [8].

(5) Such an example can be placed anywhere in a Morse 2–function by adding a cancelling 0–1 round handle pair along any loop in X^n which maps to an embedded circle bounding a disk in the base, so that the image of the 0–fold ends up on the inside of the 1-fold. After that, the round 0–handle can be traded for a round $(n-2)$–handle as we have seen here. Adding this cancelling 0–1 round handle pair and then trading the round 0–handle is a homotopy that starts and ends without definite folds but

25

which passes through definite folds during the homotopy. A worthwhile example for the reader to consider, when following the proofs in this paper, is how to carry out this homotopy without definite folds. (The authors have not done this.)

4 Theorems about I–valued Morse functions on cobordisms

In this section we will prove Theorems 1.3 and 1.4 in the case where the base is the interval I, and we will prove Theorems 1.1 and 1.2 in the case where the total space is $X^n = I \times M^{n-1}$, the base is $I \times I$, and the Morse 2–functions are of the form $G(t, p) = (t, g_t(p))$. In preparation for general Morse 2–functions over $I \times I$ (see Section 5), we will need versions of Theorem 1.3 and 1.4 in which attaching maps for n–dimensional handles lie in level sets of I–valued Morse functions on $(n - 1)$–manifolds.

Throughout this section, we are given the following data:

(1) A connected m–dimensional cobordism M from $F_0 \neq \emptyset$ to $F_1 \neq \emptyset$, where F_0 and F_1 are compact $(m - 1)$–manifolds, possibly with boundary. 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 $\phi_i : [-\epsilon, \epsilon] \times B^{m-1-l_i} \times L_i \hookrightarrow (M \setminus \partial M)$, for some small $\epsilon > 0$. Note that if $l_i < m/2$ then $l_i < m - 1$. Assume the order is such that $l_1 \leq \ldots \leq l_p$.

(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 $g : M \to I$ *is standard with respect to* ϕ_i *at height* z_i if $g \circ \phi_i : [-\epsilon, \epsilon] \times B^{m-1-l_i} \times L_i \to I$ 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 ϕ_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. These can be constructed by starting with a given Morse function, projecting L_i into a level set, and then resolving crossings by stabilizing to increase the genus. More generally the ϕ_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 4.1 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 4.2 *There exists an indefinite ordered Morse function $g : M \to I$, 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. Furthermore, the indices of the critical values and the dimensions l_i of the submanifolds L_i are such that all critical values of index $\leq l_i$ are below z_i while all critical values of index $> l_i$ are above z_i.*

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 18 for an illustration of these ideas.
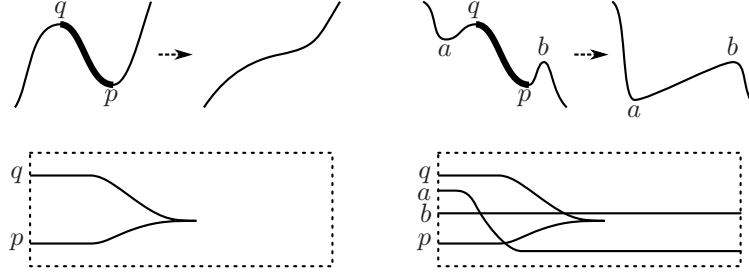
Figure 18: Cancellation of critical points q and p in one dimension, with accompanying Cerf graphics. In the first example, no other critical points need to move, but in the second example the critical point a needs to drop below p before q and p can cancel (or raise b above q).

Definition 4.3 Given a Morse function $g : M \to I$ 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 4.4 *Given $g : M \to I$ with a generic gradient-like vector field, if critical point q cancels critical point p, let C be the closure of the descending manifold for q inside $g^{-1}[g(p), g(q)]$. 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 C is actually just the descending disk, then no other critical values will change; otherwise all the critical points in C besides q and p will have to move below p before the death occurs.*

Proof of Theorem 4.2 First define $g : M \to I$ on the submanifolds $\phi_i([-\epsilon, \epsilon] \times B^{m-1-l_i} \times L_i)$ by ϕ_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 $g : M \to I$. (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 ϕ_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 set 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 $l_i + k \leq 2l_i < m$.)

To arrange that the critical values are ordered nicely with respect to the values z_i as stated, we may need to further raise or lower some critical values, but the same dimension count argument works in this case. $\qquad\square$

The following theorem is about ordered, indefinite generic homotopies on M; if F_0 and F_1 are connected, then ordered and indefinite implies fiber-connected. At the end of this section, in Lemma 4.10, we will discuss fiber-connectedness without the ordered assumption.

Theorem 4.5 *Given two indefinite, ordered Morse functions $g_0, g_1 : M \to I$ there exists an indefinite, ordered generic homotopy $g_t : M \to I$ from g_0 to g_1. If $m \geq 3$ and both g_0 and g_1 are standard with respect to each ϕ_i at height z_i, then we can arrange that, for all t, g_t is standard with respect to each ϕ_i*

at height z_i. We can also 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 $g_t : M \to I$, 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 $G : (t, p) \mapsto (t, g_t(p))$. 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 $g_t : M \to I$ is a generic homotopy between Morse functions g_0 and g_1 with a generic 1–parameter family of gradient-like vector fields. We state conditions under which we can modify g_t so as to reverse the arrows in Figures 8, 9 and 10. We leave the proofs to the reader; the basic idea, as with the proof of the standard cancellation lemma (Lemma 4.4), is as follows: We first push extraneous critical points down far enough so that we only need to work with a single descending disk and 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 4.6 (Unmerge) *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 (see Figure 9)*

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 manifold 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 4.7 (Eye death) *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 (see Figure 8) 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 manifold 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 4.8 (Swallowtail death) *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 (see Figure 10) 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 manifold 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 manifold 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 manifolds 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 4.5 As in the proof of Theorem 4.2, there is no difficulty in finding a generic homotopy $g_t : M \rightarrow I$, and if g_0 and g_1 are standard with respect to each ϕ_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 4.6. Then cancel over the overlaps and I_1 and I_n using either Lemma 4.7 or Lemma 4.8, 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 4.6 to the effect that whatever cancelled P at the beginning still cancels P wherever it has not been killed with Lemma 4.6.

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 ϕ_i at height z_i, then we want to arrange that, for all t, g_t is standard with respect to each ϕ_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 4.6, 4.7 and 4.8, we avoid neighborhoods of each $\phi_i(L_i)$. In other words, all the descending manifolds 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 achieved 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 manifold 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 19; 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 manifold for Q^j_- to meet $\phi_i(L_i)$ at some time $t_- > t_*$ while the descending manifold 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

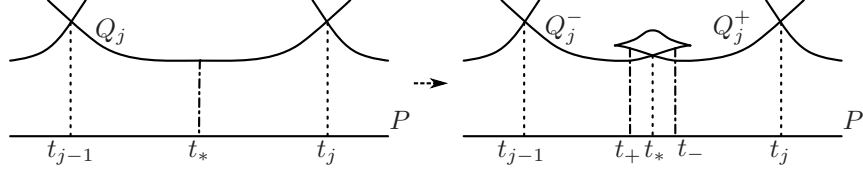$(t_* - \delta, t_j + \delta)$. (We might, of course, need to decrease δ.)

Figure 19: Splitting Q^j.

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 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 achieved 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 4.2, 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 4.9 *Suppose that $m \geq 3$. Given two indefinite, ordered generic homotopies $g_{0,t}, g_{1,t} : M \to I$ between indefinite Morse functions $g_{0,0} = g_{1,0}$ and $g_{0,1} = g_{1,1}$, there exists an indefinite, almost ordered generic homotopy of homotopies $g_{s,t} : M \to I$ from $g_{0,t}$ to $g_{1,t}$ with fixed endpoints. 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 [9], 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 4.9.)

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 4.4, 4.7, 4.6 and 4.8. 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. After discovering the necessary homotopies corresponding to these singularities, we then realized that they also play a central role, for similar reasons, in the work of Hatcher and Wagoner [9].

Proof There is always a generic homotopy $g_{s,t} : M \to I$ rel. boundaries, so the first 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 Σ, as in the example in Figure 20. In Σ, 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.
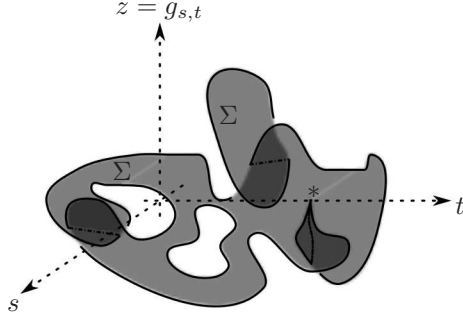
Figure 20: Example of a 2–dimensional surface Σ of 0–folds in a generic 2–parameter homotopy between Morse functions. Note the swallowtail singularity at the point labelled $*$.

The first step is to cut Σ into pieces, each of which is embedded in $I \times I \times I$. We do this by first cutting Σ 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 4.5 to modify the homotopy $g_{s_*,t}$ to get rid of definite folds at s_*, and then noting that the modification 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 21.
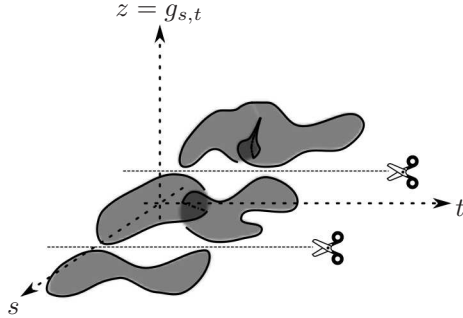

Figure 21: After the first cuts, along constant s slices. Note that components are still not embedded, due to the bad swallowtail.

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 [9], page 199 (see also item 1 on page 194), it is shown exactly how such a swallowtail birth can be

35

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 22, with accompanying graphs of the 1–dimensional Morse functions.
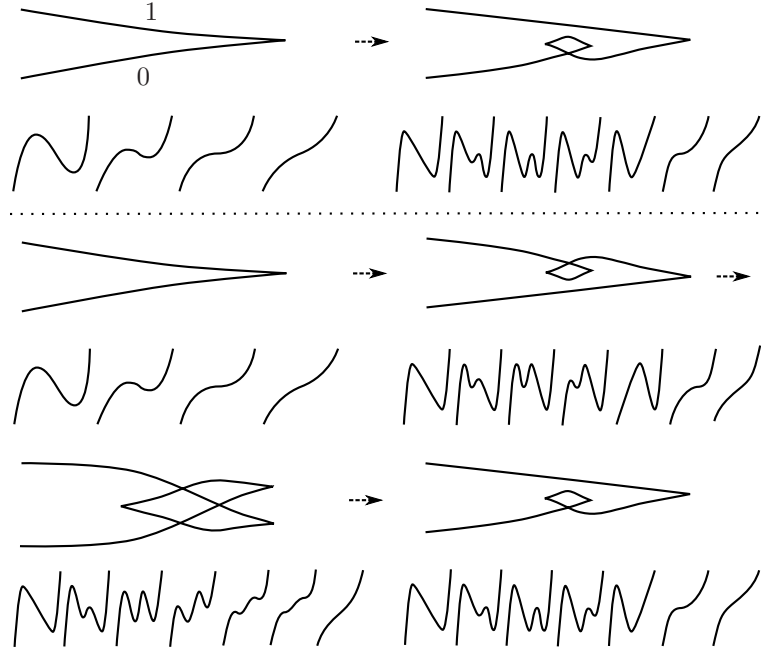

Figure 22: Cutting the bad swallowtail onto the surface of 1–folds using the butterfly singularity. Above the dotted line, the swallowtail occurs in the middle of a 0–fold, leading to surface of 0–folds that is not embedded. Below the line, the swallowtail is born in the middle of a 1–fold, so that the two 0–folds which intersect are in distinct components of the surface of 0–folds.

Now each component $\mathcal{P}$ of Σ is embedded in $I \times I \times I$ and we will eliminate these components one by one, in much the same way we eliminated individual 0–folds in Theorem 4.5 and individual index 0 critical points in Theorem 4.2. 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 4.5 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 23 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 24 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 25 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.)

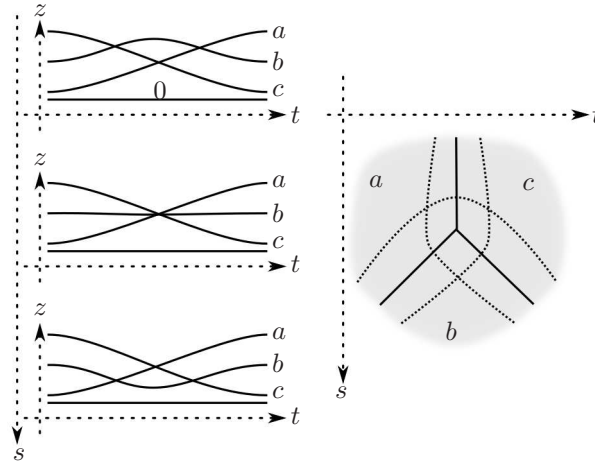Figure 23: Three 1–folds a, b and c cancelling a 0–fold in two parameters.

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 $g_{s,t}$
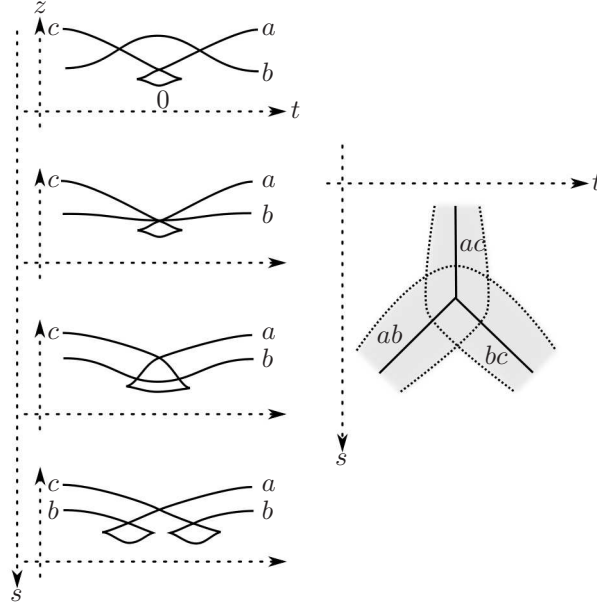
Figure 24: After cancelling the 0–1 pairs away from the overlapping regions.

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 26 we schematically illustrate the handle decomposition corresponding to $h_{s,t}$ at representative points in the (s, t)–parameter space, as labelled by boxed numbers in Figure 25. In [9], 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 $h(x_1, x_2) = x_1^3 - 3x_1 x_2^2$. 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 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 [9], 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 Σ 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
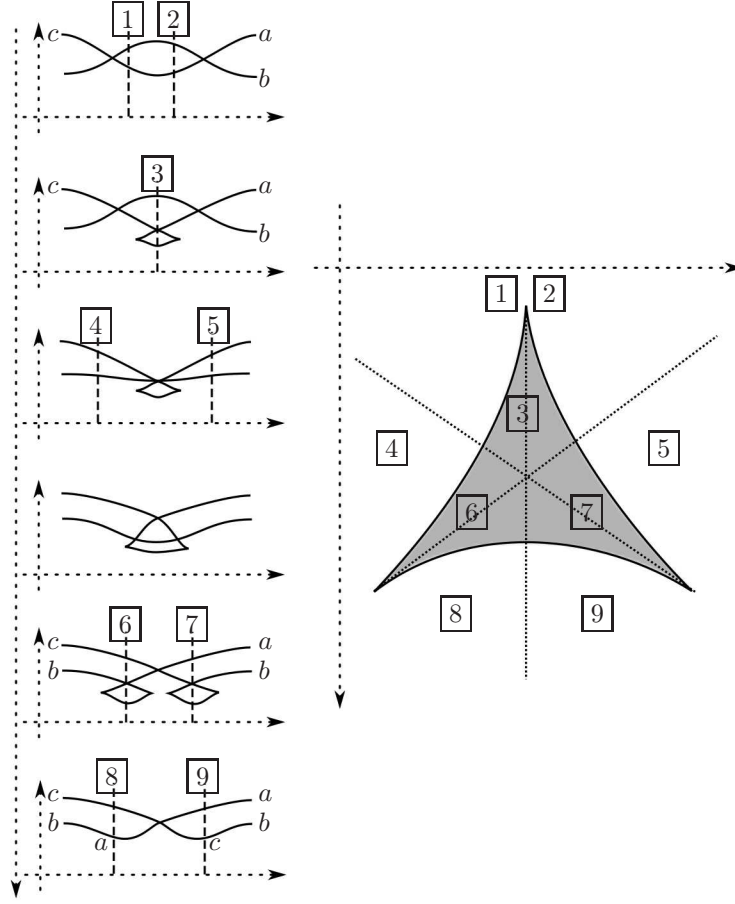
Figure 25: The 2–parameter Cerf graphic after cancelling along the double overlaps but not the triple overlaps. The dotted lines in the figure on the right indicate the points where two 1–folds intersect. Note that the labelling of the 1–folds a, b and c is no longer consistent because, as one moves across the bottom Cerf graphic, c becomes b and b becomes a, and as one moves across the top Cerf graphic, c becomes a.

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
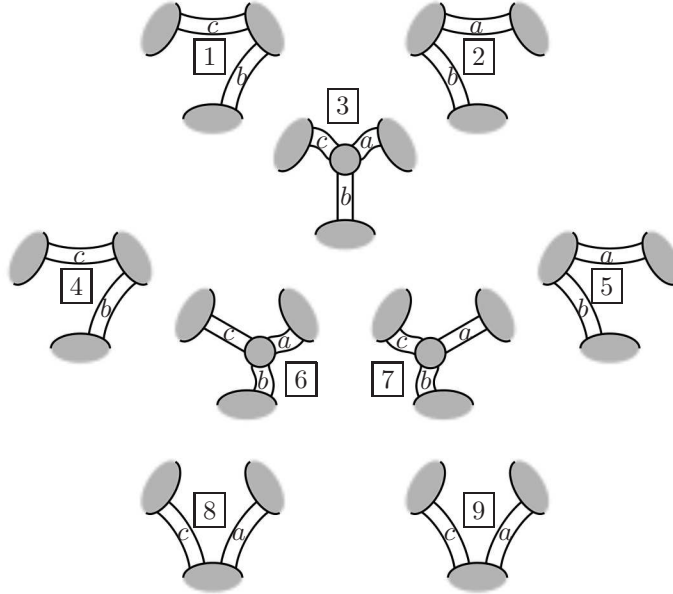
Figure 26: Handle decompositions at various points in the Cerf graphic in Figure 25.

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 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 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 27, 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.
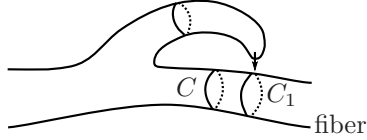

Figure 27: A 1–handle sliding over a 2–handle in a 2–parameter family.

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 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 $d \in D$.

$\square$

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 4.10 *Let $g : M \to I$ be an indefinite Morse function which is standard with respect to each ϕ_i at height z_i, and suppose that all level sets of g are*

connected. Then there exists a generic homotopy $g_t : M \to I$ 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 ϕ_i at height z_i, with the critical values also ordered with respect to the heights z_i as in the statement of Theorem 4.2.

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. ◻

5 Theorems about I^2–valued Morse 2–functions on cobordisms between cobordisms

Throughout this section let X be a connected n–dimensional cobordism from M_0 to M_1, where each M_i is a nonempty $(n-1)$–dimensional cobordism from F_{i0} to F_{i1}, with $F_{00} \cong F_{10}$ and $F_{01} \cong F_{11}$, with each F_{ij} closed and nonempty. Recall that this means that ∂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 $\{1\} \times F_{0j}$ (see Figure 1). Also suppose that we are given indefinite, ordered Morse functions $\zeta_0 : M_0 \to I$ and $\zeta_1 : M_1 \to I$. On I^2 we use coordinates (t, z), thinking of t as horizontal and z as vertical.

Note that we are not at this point assuming that the M_i's or the F_{ij}'s are connected. However, the assumption that each ζ_i is indefinite implies that each component of M_i is a connected cobordism between nonempty components of the F_{ij}'s.

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 5.1 *Suppose that $n \geq 3$. Given any indefinite, ordered Morse function $\tau : X \to I$ which is projection to I on $I \times F_{00}$ and $I \times (-F_{01})$ in ∂X, there exists an indefinite square Morse 2–function $G : X \to I^2$ such that $z \circ G|_{M_0} = \zeta_0$, $z \circ G|_{M_1} = \zeta_1$ and $t \circ G = \tau$. If $n \geq 4$ and each M_i and F_{ij} is connected, then we can arrange that G is fiber-connected.*

Theorem 5.2 *Suppose that $n \geq 4$. Given two indefinite square Morse 2–functions $G_0, G_1 : X \to I^2$ with $z \circ G_0|_{M_0} = z \circ G_1|_{M_0} = \zeta_0$ and $z \circ G_0|_{M_1} = z \circ G_1|_{M_1} = \zeta_1$, and such that both $\tau_0 = t \circ G_0$ and $\tau_1 = t \circ G_1$ are ordered, there exists an indefinite generic homotopy $G_s : X \to I^2$ between G_0 and G_1 such that $\tau_s = t \circ G_s$ is ordered (and τ_s restricted to $I \times F_{0i}$ is projection onto I, as in Definition 2.10). If in addition G_0 and G_1 are fiber-connected then we can arrange that G_s is fiber-connected.*

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 $g_t : M \to I$ between Morse functions $g_0, g_1 : M \to I$ gives a square Morse 2–function $G : X \to I^2$ 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 $t \circ G : X \to I$. 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 $t \circ G : X \to I$ 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 28 (recall that $\mu_k^n(\mathbf{x}) = \mu_k^n(x_1, \dots, x_n) = -x_1^2 - \dots - x_k^2 + x_{k+1}^2 + \dots + x_n^2$):

$$(\mu_k^n(\mathbf{x}), x_{k+1}) = \Gamma_k^+(\mathbf{x}) \qquad (\mu_k^n(\mathbf{x}), x_k) = \Gamma_k^-(\mathbf{x})$$
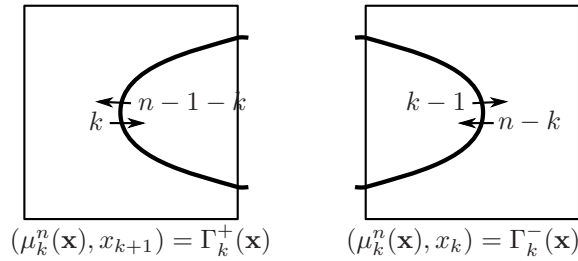
Figure 28: Local models for Morse critical points of index k realized via Morse 2–functions. On the left is a "forward k–handle", on the right is a "backward k–handle". Again, the arrows indicate the index of the fold in the given direction, not the index of the Morse critical point.

(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 Γ_k^+ to $(x_1,\ldots,x_n) \mapsto (\mu_k^{n-1}(x_1,\ldots,\widehat{x_{k+1}},\ldots,x_n), x_{k+1})$, showing that Γ_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 Γ_k^- to $(x_1,\ldots,x_n) \mapsto (\mu_{k-1}^{n-1}(x_1,\ldots,\widehat{x_k},\ldots,x_n), x_k)$, showing that Γ_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 leftward-opening parabola $t = -z^2$. This fold is indefinite when $1 < k < n$.

Note that we could have defined Γ_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, Γ_k^+ and Γ_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 28. If we turn a forward k–handle backwards, i.e. post-compose Γ_k^+ with the time-reversal $(t,z) \mapsto (-t,z)$, it becomes a backward $(n-k)$–handle.

Here are some further observations about forward k–handles, illustrated below in Figure 29; the reader can figure out the parallel statements for backward handles:

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

(2) When $0 \le k < n-1$, the ascending disk $\{x_1 = \ldots = x_k = 0\}$ has image equal to the "interior" of the parabola $\{t \ge z^2\}$. (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 $0 < k \le n-1$, 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 $0 < k \leq n - 1$, the attaching region for the k–handle, which we identify as the set $\{\mu_k^n(\mathbf{x}) = -R^2, -\epsilon \leq x_{k+1} \leq \epsilon, x_{k+2}^2 + \ldots + x_n^2 \leq \epsilon^2\} \cong S^{k-1} \times [-\epsilon, \epsilon] \times B^{n-1-k}$, 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 $(R^2, \pm R)$.
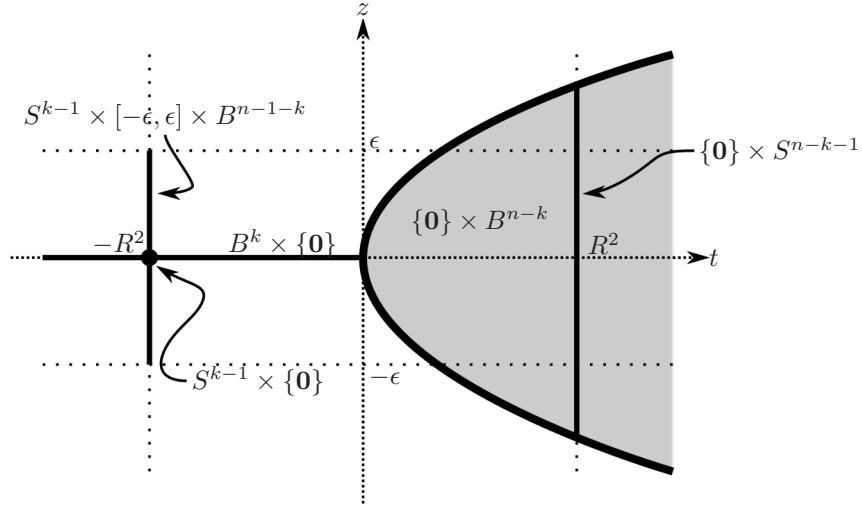

Figure 29: Images of ascending and descending disks and spheres, and attaching region, for a forward k-handle $B^k \times B^{n-k}$. The parabola $t = z^2$ is the image of the fold.

First we prove existence using these local models:

Proof of Theorem 5.1 If τ 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 ζ_0 and ζ_1 as two indefinite Morse functions on M_0. Then our result follows from Theorem 4.5; we get an indefinite generic homotopy ζ_t and we let $G(t, p) = (t, \zeta_t(p))$.

Thus if we can now prove the theorem in the case where τ 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 τ with $1 - \tau$ and switch M_0 and M_1.)

Choose a gradient-like vector field V for τ, and use this to find an embedding $\phi : [-\epsilon, \epsilon] \times B^{n-1-k} \times S^{k-1} \hookrightarrow M_0$ 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 4.2. 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 ϕ. 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 30 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.)
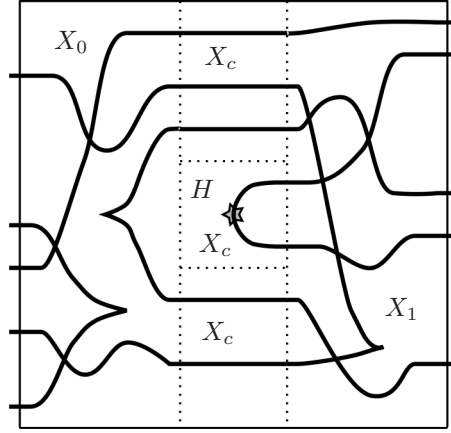
Figure 30: Constructing a Morse 2–function with a single critical point in the horizontal Morse function. This diagram shows the images of the four parts of X, X_0, X_1, X_c and the handle H; the image of the critical point is at the star. Note that the images of X_0 and X_1 are disjoint, while the image 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 4.2 to each component of M_0 to construct an indefinite, ordered Morse function $\zeta_{1/2-\delta} : M_0 \to I$ which is standard with respect to ϕ 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 ϕ), with critical values of index $\leq k$ at heights $< 1/2$ and critical values of index $\geq k+1$ at heights $> 1/2$. Now use Theorem 4.5 to construct an indefinite, ordered generic homotopy ζ_t (for $t \in [0, 1/2 - \delta]$) connecting ζ_0 to $\zeta_{1/2-\delta}$. Then we let $G : X_0 \to [0, 1/2 - \delta] \times I$ 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 30. 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 4.5 to construct an indefinite generic homotopy ζ_t (for $t \in [1/2 + \delta, 1]$) connecting $\zeta_{1/2+\delta}$ to ζ_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 ζ_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 are two different cases when $\zeta_{1/2+\delta}$ will not be ordered. The first case 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 31. However, as long as the level sets of $\zeta_{1/2+\delta}$ are connected, we can start off the homotopy ζ_t, $t \in [1/2 + \delta, 1]$, by switching the heights of the two offending critical points as indicated on the right in Figure 31, 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 4.2 shows that this is easy to achieve.

The second case when $\zeta_{1/2+\delta}$ will not be ordered is when $n - 1 - k > k + 1$, in which case the upper of the two new critical points, which has index $n - 1 - k$, may have critical points of lower index above it. In this case start off the homotopy ζ_t, $t \in [1/2 + \delta, 1]$, by lifting this critical point above those of lower index, and then proceed as above. The only case when this could conceivably cause any connectedness problems is when $k = 1$ and $n - k - 1 = n - 2$, but adding a 1–handle and its dual $(n - 2)$–handle to a fiber that is already connected cannot disconnect the fiber. $\square$
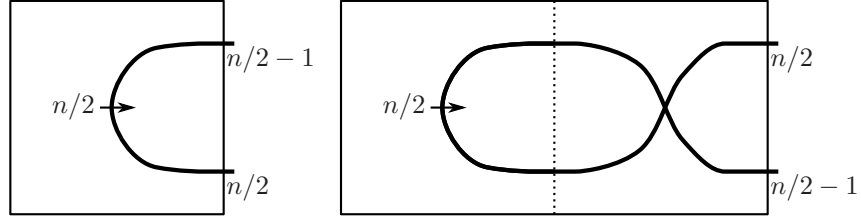

Figure 31: After attaching a handle of index $n/2$, the vertical Morse function will not be ordered, as on the left. The arrow labelled $n/2$ indicates that this fold is of index $n/2$ when looked at in the direction of the arrow. The numbers to the right of each box are the index of the critical points 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 5.2. 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 5.3 *Consider a square Morse 2–function $G : X \to I^2$ and a critical point $p \in X$ of $t \circ G : X \to I$, of index k, with $G(p) = (t_p, z_p)$. Suppose we are given standard coordinates $(x_1, \ldots, x_n)$ on a neighborhood ν of p such that $\tau = t \circ G(x_1, \ldots, x_n) = \mu_k^n(\mathbf{x}) + t_p$. Then there exists an arc G_s of Morse 2–functions (i.e. G_s is Morse for all s) supported inside ν, with $G_0 = G$, such that, inside a smaller neighborhood $\nu' \subset \nu$, $G_1(\mathbf{x}) = \Gamma_k^\pm(\mathbf{x}) + (t_p, z_p)$. It will be Γ_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 Γ_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 $\tau = \mu_k^n(\mathbf{x}) + t_p$. Because the rank of DG at $p = 0$ must be 1, we know that ζ is nonsingular at 0 and so, after a small perturbation of G (staying Morse) supported in a neighborhood of 0 we can assume that ζ 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 μ_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.

$\square$

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 5.4 *Suppose that $G : X \to I^2$ 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 32 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 32 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.
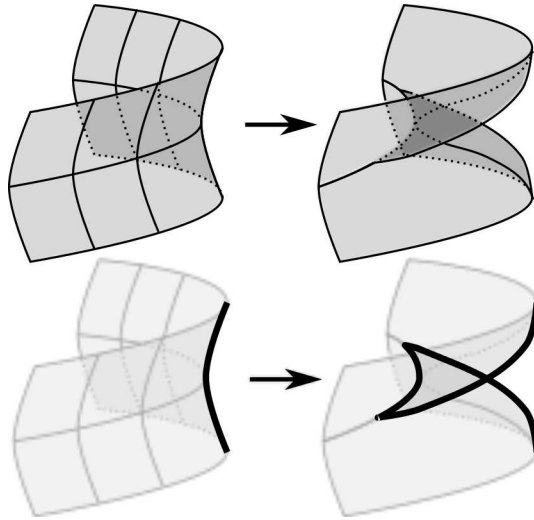

Figure 32: Bending a forward handle to a backward handle. The point is to see that the only change to the projection to the horizontal axis in this deformation is that the critical points moves a little to the left, but that when we look at the map to 2 dimensions, i.e. projection to the page exactly as in this figure, the fold at the critical point changes from opening to the right to opening to the left. This figure is in fact nothing more than the standard homotopy that introduces a swallowtail.

$\square$

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 5.5 *Consider a square Morse 2–function $G : X \to I^2$ 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 $\tau = t \circ G = \mu_k^n(\mathbf{x}) + t_p$ and, for any small $\epsilon > 0$, let H_ϵ be the k–handle neighborhood of p obtained by flowing forward along V starting from the attaching region $\{(x_1, \ldots, x_n) \mid \tau(x_1, \ldots, x_n) = -\epsilon + t_p, x_{k+1}^2 = \epsilon^2, x_{k+2}^2 + \ldots + x_n^2 \leq \epsilon^2\} \cong S^{k-1} \times [-\epsilon, \epsilon] \times B^{n-k-1}$, 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_ϵ^c be the closure of $\tau^{-1}[t_p - \epsilon, t_p + \epsilon] \setminus H_\epsilon$ and $M_{t_p - \epsilon}^c = H_\epsilon^c \cap \tau^{-1}(t_p - \epsilon)$. Using V, H_ϵ^c is naturally identified with $[t_p - \epsilon, t_p + \epsilon] \times M_{t_p - \epsilon}^c$. 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_ϵ and independent of s outside $\tau^{-1}(t_p - 2\epsilon, t_p + 2\epsilon)$, such that, on H_ϵ^c, identified with $[t_p - \epsilon, t_p + \epsilon] \times M_{t_p - \epsilon}^c$, 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 $g|_{M_{t_p - \epsilon}^c}$ 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 5.2 We are given two indefinite square Morse 2–functions $G_0, G_1 : X \to I^2$ 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 $G_s : X \to I^2$ 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, and we are given that they are ordered. Let $\tau_s : X \to I$ be an indefinite, ordered generic homotopy from τ_0 to τ_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 τ_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 τ_s to achieve this.) This step is carried out in Lemma 5.6. 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 τ, $G_{1/2}$ and $G_{3/4}$ agree on $\tau^{-1}([t_* - \epsilon, t_* + \epsilon]) = G_{1/2}^{-1}([t_* - \epsilon, t_* + \epsilon] \times I) = G_{3/4}^{-1}([t_* - \epsilon, t_* + \epsilon] \times I)$. This step is carried out in Lemma 5.8.

(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 τ. 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 4.9 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 when G_0 and G_1 are fiber-connected. We have already arranged for τ_s to be ordered for all s. In this case, Lemma 5.6 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 5.8 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 value of τ, 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 4.9 to construct G_s for $s \in [1/2, 3/4]$, we should first use Lemma 4.10 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.

$\square$

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

Lemma 5.6 *Given any indefinite Morse 2–function $G : X \to I^2$ and an indefinite, ordered generic homotopy $\tau_s : X \to I$ between Morse functions, with $t \circ G = \tau_0$ and with no deaths of cancelling pairs of critical points, there exists an indefinite, ordered generic homotopy $\tau'_s : X \to I$, with $\tau'_0 = \tau_0$ and $\tau'_1 = \tau_1$, which is connected to τ_s by an arc of generic homotopies, and there exists an indefinite generic homotopy of Morse 2–functions $G_s : X \to I^2$ with*

$G_0 = G$ and with $t \circ G_s = \tau_s'$. If G is fiber-connected then we can arrange that G_s is fiber-connected for all s.

Proof of Lemma 5.6 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 τ_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 τ_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 because we can post-compose with an isotopy so that the critical values are constant, then pre-compose so that the critical points are constant, then pre-compose again to arrange that the homotopy is constant on neighborhoods of the critical points, and finally integrate a time-like vector field to get the full isotopy. 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 33; 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 33. As long as this eye birth is indefinite, it will not disconnect fibers.

(2) Since τ_s is ordered, we only need to switch critical values of the same index. Lemma 5.7 below deals with more general critical values switches, but applies in particular to this case. (We have separated that lemma from the proof of this lemma because we will need the more general case in the following section.)
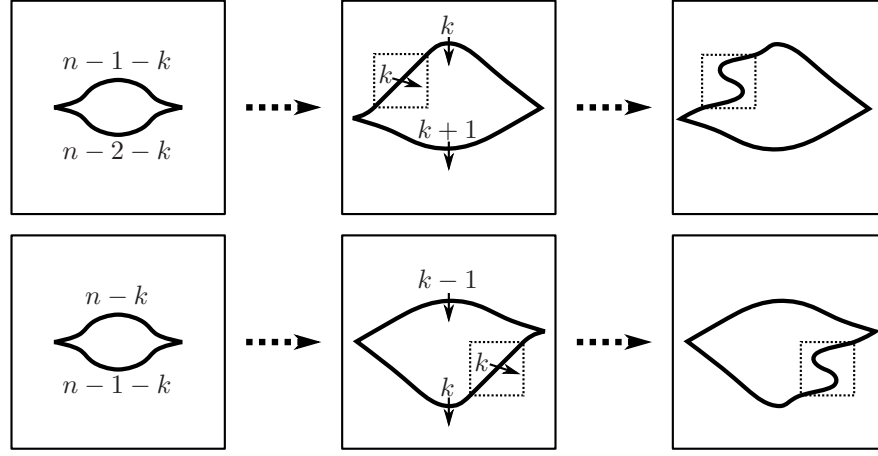
Figure 33: Two ways to realize a birth of a cancelling k–$(k+1)$ pair in τ_s via a homotopy of G; the birth occurs inside the dotted square.

$\square$

Lemma 5.7 *Given an indefinite Morse 2–function $G\colon X \to I^2$, let $\tau = t \circ G$. Consider two critical points p and q of τ, with p having index $k \leq n/2$, with q having index $l \geq k$, with $\tau(q) < \tau(p)$ and with no other critical values between $\tau(q)$ and $\tau(p)$. Fix a generic gradient-like vector field for τ. Then there exists an indefinite generic homotopy G_s with $G_0 = G$ such that $\tau_s = t \circ G_s$ agrees with τ outside a neighborhood of the descending manifold for p and, inside this neighborhood, no new critical points are born and $\tau_s(p)$ decreases monotonically from $\tau(p)$ to $\tau(q) - \epsilon$ for some small $\epsilon > 0$.*

Proof Suppose that $\tau(p) = 2/3$ and $\tau(q) = 1/3$ and no other critical values between $1/3$ and $2/3$. Using Lemmas 5.3 and 5.4, we arrange for local coordinates near p with respect to which G has the form $G = \Gamma_k^+(\mathbf{x}) + (2/3, z_p)$, a forward k–handle, and we arrange for a disjoint coordinate system near q with respect to which $G = \Gamma_l^\pm(\mathbf{x}) + (1/3, z_q)$, a forward or backward l–handle. We further use Lemma 5.5 to arrange that Γ is constant in the t direction outside a neighborhood of q for $t \in [1/3 - \epsilon, 1/3 + \epsilon]$. First we consider the case that q is a a backward l–handle.

Since there are no critical points between p and q, $\tau^{-1}([1/3 + \epsilon, 2/3 - \epsilon])$ is diffeomorphic to $[1/3 + \epsilon, 2/3 - \epsilon] \times M$ for an $(n-1)$–manifold M (a connected cobordism between non-empty manifolds), and via this diffeomorphism G has
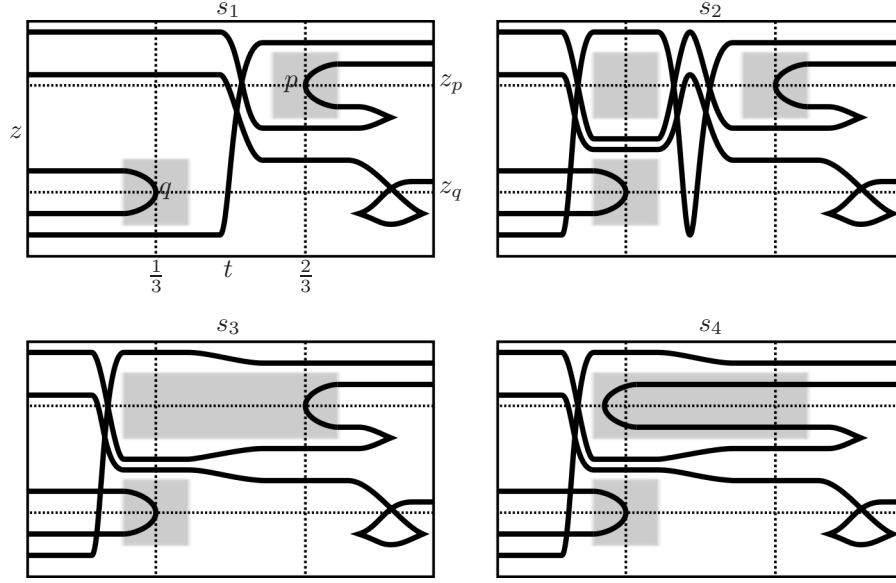
54

Figure 34: Passing two critical points past each other; the homotopy is illustrated at four successive values of s. Shaded squares are regions where g_t is standard with respect to ϕ_p or ϕ_q. To pass from s_1 to s_2, we modify g_t in a neighborhood of $t = 1/3$ so that the framed attaching sphere for p lies in a level set at $t = 1/3$. Going from s_2 to s_3, we modify g_t between $t = 1/3$ and $t = 2/3$ so that this attaching sphere lies in a level set on the whole interval $[1/3, 2/3]$.

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 ϕ_q of a neighborhood of the ascending sphere for q and an embedding ϕ_p of a neighborhood of the descending sphere for p in M such that $g_{1/3+\epsilon}$ is standard with respect to ϕ_q at height z_q while $g_{2/3-\epsilon}$ is standard with respect to ϕ_p at height z_p. It is then a straightforward application of Theorems 4.2 and 4.5 to arrange first that g_t is standard with respect to ϕ_p at height z_p for t near $1/3$, and then a straightforward application of Theorem 4.9 to arrange that g_t is standard with respect to ϕ_p at height z_p for all intermediate values of t. Here we use the t–independence of G outside a neighborhood of q, as given by Lemma 5.5 so that we can smoothly connect the behavior of $g_{1/3+\epsilon}$ to the behavior of $g_{1/3-\epsilon}$. Having done this, we can easily lower p past q. This is illustrated in Figure 34, where "lowering" p really means moving p to the left.

55

In the case that q is a forward l–handle, we use the same argument as above but look first at $M' = \tau^{-1}(1/3 - \epsilon)$, with a vertical Morse function $g_{1/3-\epsilon}\colon M' \to I$ which is standard with respect to the descending manifold for q at height z_q. We modify this to be standard with respect to the descending manifolds for both p and q, using the t–independence of G outside a neighborhood of q for $t \in [1/3 - \epsilon, 1/3 + \epsilon]$, so that again we get $g_{1/3+\epsilon}\colon M \to I$ to be standard with respect to the descending manifold for p and proceed as above.

Regarding fiber-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 4.2, 4.5 and 4.9 to arrange that g_t is standard with respect to ϕ_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 4.5 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, i.e. that the union of two circles intersecting transversely at one point stays in the fiber. Since the proof of the relevant part of Theorem 4.5 simply involves counting dimensions and appealing to transversality, this slight improvement is straightforward.

$\square$

The other lemma needed in the proof of Theorem 5.2 is:

Lemma 5.8 *Given two indefinite Morse 2–functions $G, G' : X \to I^2$ 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 $t \circ G = t \circ G' = \tau : X \to I$, there exists an $\epsilon > 0$ and an indefinite deformation of Morse 2–functions $G_s : X \to I^2$ such that:*

(1) *$G_0 = G$,*

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

(3) *for each critical value t_* of τ, 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 have all fibers connected as well.

Proof Again, we use Lemma 5.3 to standardize $G = (\tau, \zeta)$ and $G' = (\tau, \zeta')$ near each critical point of τ, so that ζ and ζ' are equal in a neighborhood of each critical point. Then use Lemma 5.5 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 Theorem 4.5. □

6 The main results

In this section we prove the theorems stated in the introduction. We will apply Thom-Pontrjagin type arguments to reduce to the case of maps to disks, so first we show quickly how the results of the preceding two sections immediately give our main theorems in the case when we are mapping to B^1 or B^2.

Proof of Theorem 1.3 for maps to B^1 This is exactly Theorem 4.2, with ordered implying fiber-connected when $m > 2$. □

Proof of Theorem 1.4 for maps to B^1 Given two indefinite Morse functions $g_0, g_1 \colon M^m \to B^1$, we need to construct an indefinite generic homotopy g_t connecting them, and if $m \geq 3$ and g_0 and g_1 are fiber-connected we need to arrange that g_t is fiber-connected. Apply Lemma 4.10 to homotope g_0 and g_1 to be ordered, without destroying fiber-connectedness if they are given as fiber-connected. Then apply Theorem 4.5. □

Proof of Theorem 1.1 for maps to B^2 Here we are given an indefinite Morse function $g \colon \partial X^n \to S^1$ and we need to construct an indefinite Morse 2–function $G \colon X \to B^2$ with $G|_{\partial X} = g$. When $n \geq 4$ and g is fiber-connected we need to arrange that G is fiber-connected. Arbitrarily identify S^1 with ∂I^2 so that g has no critical values in $I \times \{0, 1\} \subset \partial I^2$, and then use this identification to realize X as a cobordism from $M_0 = g^{-1}(\{0\} \times I)$ to $M_1 = g^{-1}(\{1\} \times I)$. Apply Lemma 4.10 to begin the construction of G on a collar neighborhood of ∂X so as to reduce to the case where g is ordered on M_0 and M_1, without destroying fiber-connectedness if we had it to begin with. This gives us ζ_0 and ζ_1 as in the preceding section, in preparation for Theorem 5.1. Also apply Theorem 4.2 to X to produce the desired indefinite, ordered Morse function τ. Finally apply Theorem 5.1 to produce G. □

Proof of Theorem 1.2 for maps to B^2 Now we are given two indefinite Morse 2–functions $G_0, G_1 \colon X^n \to B^2$ which agree on ∂X, and, assuming that $n \geq 4$, we need to construct an indefinite generic homotopy G_s connecting them. When G_0 and G_1 are fiber-connected we need to arrange that G_s is fiber-connected. In order to reduce to Theorem 5.2, we first proceed as in the previous proof, identifying B^2 with I^2 and applying Lemma 4.10 to get ζ_0 and ζ_1 ordered as in Theorem 5.2. Now we need to arrange that $\tau_0 = t \circ G_0$ and $\tau_1 = t \circ G_1$ are ordered. We do this with Lemma 5.7, switching critical values so as to order them by index, making sure to always move critical points of index $\leq n/2$ to the left and critical points of index $> n/2$ to the right. Now we can apply Theorem 5.2. $\qquad\square$

Now we can consider more interesting topology in our target spaces. For existence of indefinite, fiber-connected S^1–valued Morse functions, for example, we need to arrange for one fiber to be connected, and then we can cut open along that fiber and reduce to the B^1–valued case. These proofs are essentially a refinement of the Thom-Pontrjagin construction, in which maps to S^n and homotopies between them are constructed and characterized by specifying the fiber over a point. However, we need to be more careful than in the basic Thom-Pontrjagin construction because of the fact that we need to arrange for connected fibers, and because we want to construct homotopies which connect fibers without introducing extraneous components along the way.

In the traditional Thom-Pontrjagin construction for maps from S^n to S^k, a framed codimension–k submanifold of S^n determines a map by sending the submanifold to the north pole and the framed normal k–disk bundle to S^k, with boundary and the complement of the normal disk bundles all mapping to the south pole. A framed cobordism in $I \times S^n$ between two such framed submanifolds determines a homotopy between their maps in the same way.

In our setting, the framed cobordisms are easy to see but we require more control on the associated homotopies than we would get by applying the standard Thom-Pontrjagin arguments. In particular, if we want to construct a homotopy which connects two components of a disconnected fiber, we will need to choose an arc connecting the components and then construct a homotopy with support in a neighborhood of the arc, such that, during the homotopy, extraneous components are not introduced.

We begin with the existence and uniqueness results for indefinite, fiber-connected, Morse functions. As mentioned earlier, we already have these results when mapping to I, so we focus now on maps to S^1.
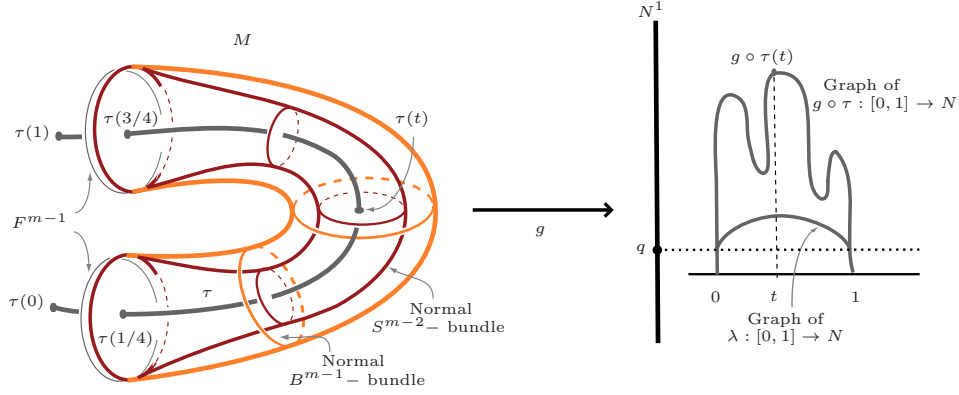
Figure 35: Connecting components of $g^{-1}(q)$.

Proposition 6.1 *Given a Morse function $g : M^m \to S^1$ which is surjective on π_1, with M connected, and given a regular value $q \in S^1$, if $g^{-1}(q)$ is disconnected then we can find two components of $g^{-1}(q)$ connected by an arc $a \subset M$ which intersects $g^{-1}(q)$ only at it's endpoints and which is disjoint from either $g^{-1}[q - \epsilon, q)$ or $g^{-1}(q, q + \epsilon]$, for some small $\epsilon > 0$.*

Proof Begin with an arc connecting any two components of $g^{-1}(q)$, meeting $g^{-1}(q)$ transversely, intersecting $g^{-1}(q)$ with opposite signs at its endpoints, and projecting to a homotopically trivial loop in S^1. (Here we use π_1–surjectivity.) If this arc hits $g^{-1}(q)$ anywhere in its interior, either take an innermost arc intersecting distinct components with opposite signs at its endpoints, or, if an innermost arc starts and ends in the same component, shortcut in the obvious way, reducing the number of intersections. □

We will use such arcs to construct homotopies that connect components of $g^{-1}(q)$. For the basic idea, referring to Figure 35, consider a given Morse function $g : M^m \to N^1$, with regular value q and suppose the fiber $F = g^{-1}(q)$ is disconnected. Choose an arc $\tau : [0, 1] \to M$ transverse to F, intersecting F only at $\tau(1/4)$ and $\tau(3/4)$, connecting two components of F. (If τ intersects F at more points, choose a shorter arc.) Note that τ has a normal B^{m-1}–bundle with boundary S^{m-2}–bundle. We wish to construct a homotopy g_t starting at $g_0 = g$, supported inside this B^{m-1}–bundle, such that, watching the level sets $g_t^{-1}(q)$ as t ranges from 0 to 1, we see two fingers stick out from F following τ, meeting at $\tau(1/2)$, and merging to achieve a surgery on the 0–sphere $\{\tau(1/4), \tau(3/4)\}$. In other words, we connect the two components of

59

F with a tube along τ, without introducing any extraneous components during the homotopy.

To visualize the homotopy g_t, choose coordinates for the normal B^{m-1}–bundle to the arc $\tau([0,1])$ so that g is projection on the t coordinate, that is, each normal B^{m-1} maps to a point of N. Then g_t should equal g_0 on the boundary S^{m-2} of each normal B^{m-1} and should push the interior of this B^{m-1} down N until the center of B^{m-1} has gone past q. If the center is below q, then a normal S^{m-2} will map to q, and these spheres will make up the cylinder $I \times S^{m-2}$ which connects the two components of the fiber.

In order not to create extra components during the homotopy g_t, the interiors of the normal B^{m-1}'s need to be pushed across q in turn, starting monotonically at $t = 1/4$ and $t = 3/4$, and finally at $t = 1/2$. To organize this, choose a parabolic arc λ as drawn in Figure 35 and first push all the centers of the normal B^{m-1}'s, i.e. $\tau(t)$, down to λ, and then push the centers down as though λ was being translated down.

The details are as follows:

Lemma 6.2 *Suppose that $g : M^m \to N^1$ is a Morse function and that $\tau : I \to M$ is a smooth embedded path avoiding the critical points of g such that $\gamma = g \circ \tau$ is homotopically trivial (rel. endpoints) in N. Let $q \in N$ be a regular value for both $g : M \to N$ and $\gamma : [0,1] \to N$, with $\gamma^{-1}(q) = \{1/4, 3/4\}$. Suppose that $\Gamma : I \times I \to N$ is a homotopy, with $\Gamma(0,x) = \gamma(x)$, such that q is regular for Γ and such that $\Gamma^{-1}(q)$ is an arc $a \subset I \times I$ from $(0, 1/4)$ to $(0, 3/4)$ and such that projection onto the first factor of $I \times I$ restricts to a as a Morse function with a single index 1 critical point at $(1/2, 1/2)$. Then, inside any neighborhood of $\tau(I)$ there is an embedding $\overline{\tau} : B^{m-1} \times I \to M$, with $\overline{\tau}(\mathbf{0}, x) = \tau(x)$ such that, letting $\overline{\gamma} = g \circ \overline{\tau} : B^{m-1} \times I \to N$, we have a homotopy $\overline{\Gamma} : I \times B^{m-1} \times I \to N$ satisfying the following properties:*

(1) *For any $(0, \mathbf{y}, x) \in I \times B^{m-1} \times I$, $\overline{\Gamma}(0, \mathbf{y}, x) = \overline{\gamma}(\mathbf{y}, x)$.*

(2) *For any $(t, \mathbf{0}, x) \in I \times B^{m-1} \times I$, $\overline{\Gamma}(t, \mathbf{0}, x) = \Gamma(t, x)$.*

(3) *For any $(t, \mathbf{y}, x) \in I \times ((S^{m-2} \times I) \cup (B^{m-1} \times \{0, 1\}))$, $\overline{\Gamma}(t, \mathbf{y}, x) = \overline{\Gamma}(0, \mathbf{y}, x) = \overline{\gamma}(\mathbf{y}, x)$.*

(4) *The point $q \in N$ is a regular value for $\overline{\Gamma}$.*

(5) *$\overline{\Gamma}^{-1}(q)$ is a m–dimensional submanifold of $I \times B^{m-1} \times I$ on which the projection onto the first factor of $I \times B^{m-1} \times I$ has a single Morse critical point of index 1 at $(1/2, \mathbf{0}, 1/2)$. This is illustrated in Figure 36.*

60

(6) *This submanifold $\overline{\Gamma}^{-1}(q)$ intersects the boundary of $I \times B^{m-1} \times I$ as*

$$(\{0\} \times B^{m-1} \times I) \cap \overline{\Gamma}^{-1}(q) = \{0\} \times B^{m-1} \times \{1/4, 3/4\}$$

and

$$(I \times S^{m-2} \times I) \cap \overline{\Gamma}^{-1}(q) = I \times S^{m-2} \times \{1/4, 3/4\}.$$

(7) *In particular, this implies that $(\{1\} \times B^{m-1} \times I) \cap \overline{\Gamma}^{-1}(q)$ is properly embedded in $\{1\} \times B^{m-1} \times I$ and diffeomorphic to $S^{m-2} \times [1/4, 3/4]$, only meeting the boundary of $\{1\} \times B^{m-1} \times I$ at $S^{m-2} \times \{1/4, 3/4\}$.*
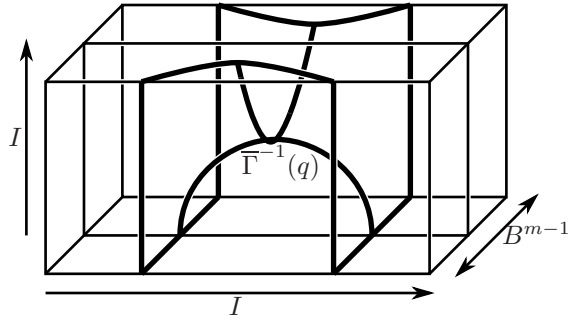

Figure 36: The model for $\overline{\Gamma}^{-1}(q)$ in the statement of Lemma 6.2.

Proof Let $\overline{\tau}$ be a parametrization of a small neighborhood of $\tau([0,1])$ such that $\overline{\gamma}^{-1}(q) = B^{m-1} \times \{1/4, 3/4\}$. Let H be a model saddle hypersurface in $I \times B^{m-1} \times I$ satisfying the behavior given in the statement of the lemma for $\overline{\Gamma}^{-1}(q)$. Then, the constraints given for $\overline{\Gamma}$, where we ask that $\overline{\Gamma}^{-1}(q) = H$, completely determine $\overline{\Gamma}$ on the following subset C of $I \times B^{m-1} \times I$:

$$C = (\{0\} \times B^{m-1} \times I) \cup (I \times S^{m-2} \times I) \cup H \cup (I \times \{\mathbf{0}\} \times I) \cup (I \times B^{m-1} \times \{0, 1\})$$

There is a natural extension of this prescribed behavior of $\overline{\Gamma}$ on C to a smooth function on a regular neighborhood ν of C so that q becomes a regular value and such that $\overline{\Gamma}(\partial\nu)$ avoids q. Then note that $I \times B^{m-1} \times I$ deformation retracts onto C, so there is a continuous extenstion of $\overline{\Gamma}$ to all of $I \times B^{m-1} \times I$, which can be smoothed and made generic without changing the behaviour on ν, and without introducing any components of $\overline{\Gamma}^{-1}(q)$ outside ν.

$\square$

Existence: Proof of Theorem 1.3 We have already proved this for the case of maps to B^1. Now we consider the case of maps to S^1. We are given

a homotopically nontrivial map $g' : M \to S^1$ and we want to produce an indefinite Morse function $g : M \to S^1$ which is homotopic to g'. To simplify notation, we will simply refer to all our maps to S^1 as g, modifying g in stages by homotopies. So we begin with $g = g'$. Because g is homotopically nontrivial, we can lift to a finite cover of S^1 so that the induced map on π_1 is surjective, and thus reduce to the case $g_*(\pi_1(M)) = \pi_1(S^1)$. Fiber-connectedness is not preserved under post-composition with a covering space map, but we are not asking for fiber-connectedness unless g is π_1–surjective to begin with.

So now assume g is π_1–surjective and Morse. Identify S^1 with $\mathbb{R}/\mathbb{Z}$ so that 0 is a regular value. We want to homotope g to arrange that $g^{-1}(0)$ is connected. (The new g should still be Morse and 0 should still be regular.) Then our theorem will reduce to the cobordism case, Theorem 4.2, by cutting M open along $g^{-1}(0)$.

Choose some $\epsilon > 0$ such that there are no critical values of g in $[-\epsilon, 0]$. Choose two components of $g^{-1}(0)$ as in Proposition 6.1, and, extending the arc produced in the proposition, choose a path $\tau : [0, 1] \to M$ connecting them in M, but starting and ending in $g^{-1}(-\epsilon)$ and passing through the two components at $\tau(1/4)$ and $\tau(3/4)$, respectively, such that $\gamma = g \circ \tau$ is homotopic to $0 \in \pi_1(S^1, -\epsilon)$. We can then choose a homotopy $\Gamma : I \times I \to N$ as in the hypotheses of Lemma 6.2, with 0 being the regular value q in the lemma. Then the embedding $\overline{\tau}$ and the homotopy $\overline{\Gamma}$ gives us a homotopy g_t defined as the identity outside the image of $\overline{\tau}$ and as $g_t(p) = \overline{\Gamma}(t, \overline{\tau}^{-1}(p))$ for p in the image of $\overline{\tau}$. The effect on $g^{-1}(0)$ is to replace two B^{m-1} neighborhoods in $g^{-1}(0)$ of the two points $\tau(1/4)$ and $\tau(3/4)$ with a tube diffeomorphic to $S^{m-2} \times I$, and thus the two components get connected. Repeating, we connect all the components of $g^{-1}(0)$.

Note that the final homotopy is a concatenation of homotopies each supported in a neighborhood of an arc in M. Thus we could just as well have performed all the homotopies at the same time, as long as we can choose all the relevant arcs to be disjoint from the beginning, which we can do if $m \geq 3$. This point of view is more useful in the next proof.

$\square$

Uniqueness: Proof of Theorem 1.4 Again, we have already proved this in the case of homotopies between maps to B^1, so now we consider the case of homotopies between maps to S^1. We are given two homotopic indefinite Morse functions $g_0, g_1 : M \to S^1$ and we need to produce an indefinite homotopy g_t between them; when g_0 and g_1 are fiber-connected then g_t should also be fiber

connected for all t. Throughout the rest of this proof suppose we are trying to prove both indefiniteness and fiber-connectedness. Assuming that g_0 and g_1 are indefinite and fiber-connected implies that both maps are π_1–surjective. The proof without asking for fiber-connectedness follows by lifting to an appropriate cover.

Again, identify S^1 with $\mathbb{R}/\mathbb{Z}$ and assume that 0 and $1/2$ are regular values of both g_0 and g_1. We first choose any generic homotopy g_t, and we will show how to modify g_t so as to arrange that, for each t, both $g_t^{-1}(0)$ and $g_t^{-1}(1/2)$ are connected.

In the proof of Theorem 1.3 we saw how to do this for a fixed t, for either of the level sets $g_t^{-1}(0)$ or $g_t^{-1}(1/2)$. The construction depended on a choice of arcs in M based at the level set in question, and then produced a homotopy supported in a neighborhood of those arcs and loops. For a fixed t, it is just as easy to do this for both level sets simultaneously, since the arcs may be taken to be disjoint (recall that $m \geq 3$, where m is the dimension of M). Even if one of the level sets contains a single Morse critical point, the same construction works since we can just keep the arcs and loops away from that critical point, so we can really do this for each t. Now note that the arcs and the guiding homotopies in S^1 which are used for a fixed value $t = t_0$ will work for all t in some short interval $[t_0 - \epsilon, t_0 + \epsilon]$, modulo modifying the arcs by small ambient isotopies near their endpoints. Cover $[0, 1]$ by a finite collection of such short intervals, using disjoint arcs and loops where the intervals overlap, and use bump functions to patch together the homotopies over the whole of $[0, 1]$.

We only need to check that, in the middle of one of these homotopies, we have not introduced new components of $g_t^{-1}(0)$ or $g_t^{-1}(1/2)$, since now that we are patching homotopies together with bump functions, we do see the intermediate stages in the homotopy. In fact we do not introduce new components because, in Lemma 6.2, $\overline{\Gamma}^{-1}(q)$ has a single critical point of index 1 with respect to projection onto the first factor of $I \times B^{m-1} \times I$, so for each t, $(\{t\} \times B^{m-1} \times I) \cap \overline{\Gamma}^{-1}(q)$ either has two components (for $t < 1/2$) or one component (for $t \geq 1/2$), but never more.

So now assume that, for each t, both $g_t^{-1}(0)$ and $g_t^{-1}(1/2)$ are connected. Generically we can also assume that, at those times t when g_t is not Morse, the non-Morse singularities do not occur in $g_t^{-1}(0)$ or $g_t^{-1}(1/2)$, and that, for every time t, either 0 or $1/2$ is a regular value of g_t.

The zig-zag argument: If g_0 and g_1 agreed on some level set L, say, $g_0^{-1}(0) = g_1^{-1}(0)$, then we could cut open M along L and obtain two Morse functions to
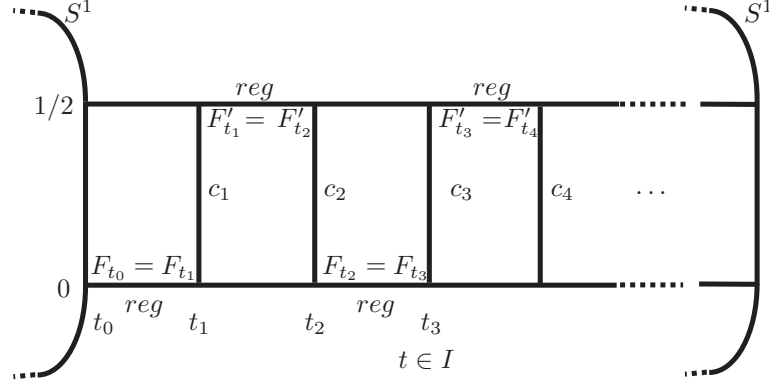
Figure 37: The zig-zag argument.

I, thus reducing to the already proven Theorem 4.5. The goal now is to arrange for this to be true for g_{t_i} and $g_{t_{i+1}}$ for $0 = t_0 < t_1 < \ldots < t_{2n} = 1$.

As in Figure 37, let the t_i's be times chosen such that 0 is regular for g_t for each $t \in [t_0, t_1] \cup [t_2, t_3] \cup \ldots \cup [t_{n-2}, t_{n-1}] \cup \{t_n = 1\}$ and $1/2$ is regular for g_t for each $t \in \{0 = t_0\} \cup [t_1, t_2] \cup [t_3, t_4] \cup \ldots \cup [t_{n-1}, t_n]$.

Let $F_{t_i} = g_{t_i}^{-1}(0)$ and let $F'_{t_i} = g_{t_i}^{-1}(1/2)$. Note that F_{t_0} is isotopic to F_{t_1} because 0 is regular for each g_t, $t \in [t_0, t_1]$. Similarly, $F_{t_{2i}}$ is isotopic to $F_{t_{2i+1}}$. We can assume they are, in fact, equal. The same works for $F'_{t_{2i+1}}$ and $F'_{t_{2i+2}}$.

The vertical cobordism $C_i = g_{t_i}^{-1}[0, 1/2]$ between F_{t_i} and F'_{t_i} is connected because the top, $g_{t_i}^{-1}(1/2)$, is connected, as is the bottom, and because $g_{t_i} : M \to S^1$ is surjective on π_1. (For, given $p, q \in C_i$, there is an arc joining them in M which intersects F_{t_i} and F'_{t_i} algebraically 0 each, and now connectedness of F_{t_i} and F'_{t_i} allows the arc to be replaced by an arc in C_i.)

Now we construct, by Theorem 4.2, an intermediate function $h_{t_1} : M \to S^1$ such that $h_{t_1}^{-1}[0, 1/2] = C_1$ and h_{t_1} is indefinite and fiber-connected.

Then g_0 and h_{t_1} agree on the level set $F_{t_0} = g_0^{-1}(0)$ and thus are joined by a fiber-connected homotopy. Also h_{t_1} and g_{t_2} agree on the level set $F'_{t_2} = g_{t_2}^{-1}(1/2)$ so they are also joined by a fiber-connected homotopy. Therefore $g_0 = g_{t_0}$ and g_{t_2} are joined by the desired homotopy which replaces g_t, $t \in [t_0, t_2]$, while g_t remains the same on $[t_2, t_n = 1]$.

Now use the same argument on the interval $[t_2, t_4]$ with an intermediate h_{t_3}, and replace g_t on $[t_2, t_4]$ by a fiber-connected homotopy. Iterating this up and down zig-zag construction finishes the proof.
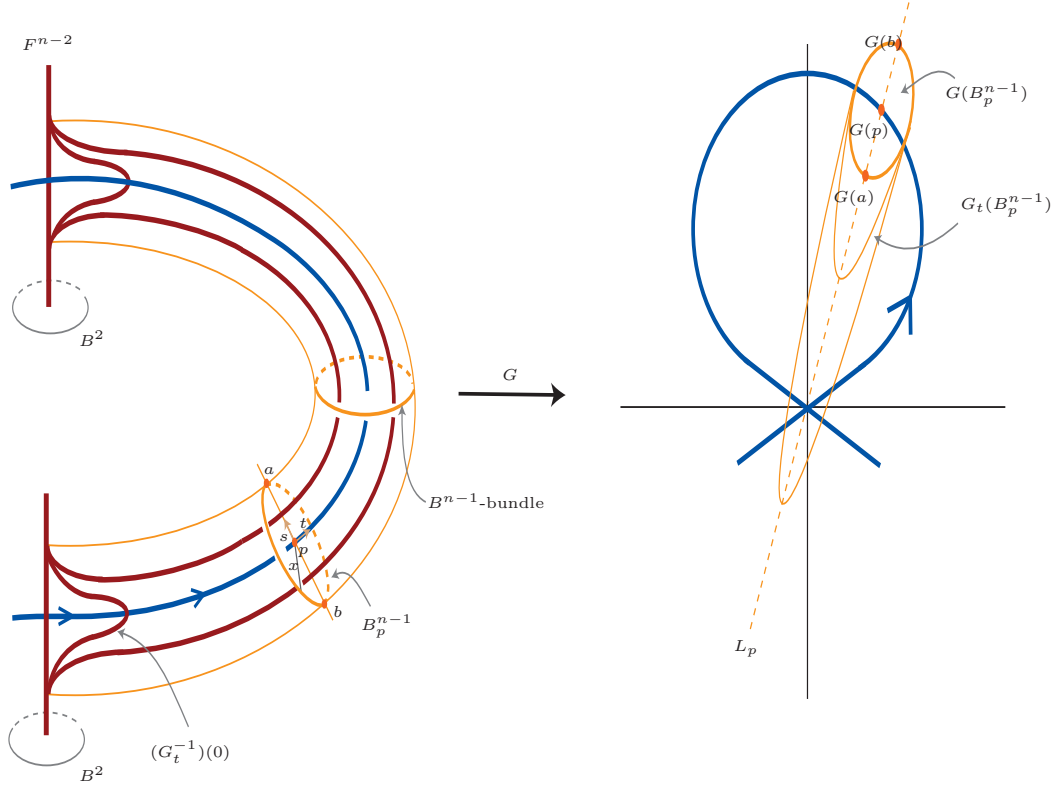
Figure 38: Connecting components of a codimension–2 fiber $F = G^{-1}(q)$

We now need versions of Proposition 6.1 and Lemma 6.2 for Morse 2-functions $G\colon X^n \to \Sigma^2$. If a fiber $F^{n-2} = G^{-1}(q)$ is not connected, then we can find an arc α in X connecting two components, meeting the components transversely only at two points (here "transverse" is in the sense of linearly independent tangent spaces), and whose image $G(\alpha)$ is immersed in the plane about q, as in Figure 38.

If α meets the fiber more than twice, perturb α slightly rel. endpoints to eliminate the extra points of intersection. And if the image $G(\alpha)$ is not as simple as in Figure 38, then there is a smaller curve, inside $g(\alpha)$, and an immersed crescent between them, and G can be homotoped without changing $G^{-1}(q)$ so that the image of α is as drawn.

As before, we wish to homotope G so that we see two fingers push out from the fiber, eventually meeting each other and morphing into a cylinder $I \times S^{n-3}$.

The idea for the homotopy is quite similar to the previous case for M mapping to N^1.

At a point p on the arc α, there is a normal B_p^{n-1} which is projected onto a 2-ball centered at $G(p)$. The boundary of B_p^{n-1} should be fixed throughout the homotopy, but the interior of the image should be pushed along the line L_p, a line from $G(p)$ through q, with $G(p)$ moving along L_p and the rest being dragged along. The two fingers will meet when $G(p)$ reaches q and they morph in the cylinder as $G(p)$ goes beyond q.

The details follow.

Lemma 6.3 *Suppose that $G : X^n \to \Sigma^2$ is a Morse 2–function and that $\tau : B^1 \times I \to X$ is a smooth embedding avoiding the critical points of G, such that $\gamma = G \circ \tau : B^1 \times I \to \Sigma$ is (locally) a Morse 2–function, with no critical points on $B^1 \times \{0, 1\}$. Suppose that $q \in \Sigma$ is a regular value of G and of γ, such that $\gamma^{-1}(q) = \{(0, 1/4), (0, 3/4)\}$, and that there exists a homotopy (rel. boundary) $\Gamma : I \times B^1 \times I \to \Sigma$ with $\Gamma(0, y, x) = \gamma(y, x)$, with q regular for Γ, such that $\Gamma^{-1}(q)$ is an arc from $(0, 0, 1/4)$ to $(0, 0, 3/4)$ on which projection to the first factor of $I \times B^1 \times I$ is Morse with a single index 1 critical point at $(1/2, 0, 1/2)$. Then, inside any neighborhood of $\tau(B^1 \times I)$ there is an embedding $\overline{\tau} : B^{n-2} \times B^1 \times I \to X$, with $\overline{\tau}(\mathbf{0}, y, x) = \tau(y, x)$, such that, letting $\overline{\gamma} = G \circ \overline{\tau} : B^{n-2} \times B^1 \times I \to \Sigma$, we have a homotopy $\overline{\Gamma} : I \times B^{n-2} \times B^1 \times I \to \Sigma$ satisfying the following properties:*

(1) *For any $(0, \mathbf{z}, y, x) \in I \times B^{n-2} \times B^1 \times I$, $\overline{\Gamma}(0, \mathbf{z}, y, x) = \overline{\gamma}(\mathbf{z}, y, x)$.*

(2) *For any $(t, \mathbf{0}, y, x) \in I \times B^{n-2} \times B^1 \times I$, $\overline{\Gamma}(t, \mathbf{0}, y, x) = \Gamma(t, y, x)$.*

(3) *For any $(t, \mathbf{z}, y, x) \in I \times \partial(B^{n-2} \times B^1 \times I)$, $\overline{\Gamma}(t, \mathbf{z}, y, x) = \overline{\Gamma}(0, \mathbf{z}, y, x) = \overline{\gamma}(\mathbf{z}, y, x)$.*

(4) *The point $q \in \Sigma$ is a regular value for $\overline{\Gamma}$.*

(5) *$\overline{\Gamma}^{-1}(q)$ is a $(n-1)$–dimensional submanifold of $I \times B^{n-2} \times B^1 \times I$ on which the projection onto the first factor of $I \times B^{n-2} \times B^1 \times I$ has a single Morse critical point of index 1 at $(1/2, \mathbf{0}, 0, 1/2)$.*

(6) *This submanifold $\overline{\Gamma}^{-1}(q)$ intersects the boundary of $I \times B^{n-2} \times B^1 \times I$ as*

$$(\{0\} \times B^{n-2} \times B^1 \times I) \cap \overline{\Gamma}^{-1}(q) = \{0\} \times B^{n-2} \times \{(0, 1/4), (0, 3/4)\}$$

and

$$(I \times \partial(B^{n-2} \times B^1 \times I)) \cap \overline{\Gamma}^{-1}(q) = I \times S^{n-3} \times \{(0, 1/4), (0, 3/4)\}$$

(7) In particular, this implies that $(\{1\} \times B^{n-2} \times B^1 \times I) \cap \overline{\Gamma}^{-1}(q)$ is properly embedded in $\{1\} \times B^{n-2} \times B^1 \times I$ and diffeomorphic to $S^{n-3} \times [1/4, 3/4]$, only meeting the boundary of $\{1\} \times B^{n-2} \times B^1 \times I$ at $S^{n-3} \times \{(0, 1/4), (0, 3/4)\}$.

Proof There are no essential differences between the proof of this lemma and the proof of Lemma 6.2. One constructs $\overline{\tau}$ as a neighborhood of $\tau(B^1 \times I)$. Then one constructs $\overline{\Gamma}$ first on a neighborhood ν of a compact subset $C \subset I \times B^{n-2} \times B^1 \times I$ containing the appropriate inverse image of q and the parts of the boundary on which the behavior of $\overline{\Gamma}$ is prescribed. Then one extends to all of $I \times B^{n-2} \times B^1 \times I$ using the fact that $I \times B^{n-2} \times B^1 \times I$ deformation retracts onto ν. This is easy to see by restricting attention to $I \times A \times B^1 \times I$ where A is a ray from 0 to ∂B^{n-2} in B^{n-2}. $\qquad\square$

Existence: Proof of Theorem 1.1 Here we are given a compact, connected, oriented n–manifold X and a 2–manifold Σ and an indefinite, surjective Morse function $g : \partial X \to \partial \Sigma$ which extends to a map $G' : X \to \Sigma$. We wish to homotope G' rel. boundary to an indefinite Morse 2–function $G : X \to \Sigma$, and perhaps also arrange fiber-connectedness. Again, we will drop the primes and simply refer to all our maps to Σ as G, constructing or modifying G in stages.

The base case is when Σ is the disk B^2, which we have already addressed. Now we reduce all other cases to this base case. The given map G can be homotoped to a Morse 2–function; we need to make it indefinite if $[G_*(\pi_1(X)) : \pi_1(\Sigma)] < \infty$ and fiber-connected if $G_*(\pi_1(X)) = \pi_1(\Sigma)$. If $[G_*(\pi_1(X)) : \pi_1(\Sigma)] < \infty$ we can lift to a finite cover of Σ and reduce to the case that G is π_1–surjective.

First suppose that Σ and X are closed, in which case there is no boundary condition g. Then choose a regular value q for G and let $F = G^{-1}(q)$. We want to homotope G to arrange that F is connected, after which we will work on the complement of a disk neighborhood of q in Σ and thus reduce to the case of nonempty boundaries. (Note that removing a disk or regular values from Σ and its inverse image from X preserves π_1–surjectivity because removing the disk from Σ adds a generator to $\pi_1(\Sigma)$, a loop around the disk, and this loop lifts since the inverse image of the disk is just $D^2 \times F$ for a smooth fiber F.)

Choose a collection of disjoint arcs based in F which connect the components of F and which map, by G, to homotopically trivial loops in Σ based at $q = G(F)$. In addition, each of these arcs should be homotopically trivial via a homotopy which avoids q, except at its endpoints of course. (Here we use

both π_1–surjectivity of G and π_1–surjectivity of $G|_{X \setminus F}$.) We can extend each such arc to an embedding τ of $B^1 \times I$ as in Lemma 6.3. (To arrange that the homotopy Γ has the desired properties one needs to arrange that the two points of intersection of $\tau(B^1 \times I)$ with the fiber F have opposite signs.) Then Lemma 6.3 shows us how to construct the desired homotopy on a neighborhood of each arc, connecting the components of F without creating new components.

Now suppose that $\partial \Sigma \neq \emptyset \neq \partial X$. Now we want to arrange that there is a collection of properly embedded arcs $a_1, \dots, a_n \subset \Sigma$, transverse to the image of the folds and avoiding the cusps, which cut Σ into a disk, such that the fiber over each point in each arc is connected. Once we do this, the problem reduces to the case $\Sigma = B^2$. Again, it is important to observe that removing a properly embedded arc from Σ and its inverse image from X does not destroy π_1–surjectivity.

Consider a single properly embedded, non-separating arc $a \subset \Sigma$. By π_1–surjectivity, there exists a properly embedded arc $b \subset X$ with $G(\partial b) = \partial a$ and with $G(b)$ homotopic rel. ∂ to a. We can arrange that b avoids the singular locus of G in X and that $G|_b$ is an immersion. We want $G|_b$ to be an embedding. At any given double point q of $G(b)$, i.e. points $q_1, q_2 \in b$ with $G(q_1) = q = G(q_2)$, we can apply the argument from above, which we used to connect fiber components in the $\partial X = \partial \Sigma = \emptyset$ case, to the path $[q_1, q_2] \subset b$ to shorten $G(b)$ (via a homotopy of G) at the double point. Repeating this process we can eliminate all double-points and get $G|_b$ to be an embedding, and we now replace a with $a = G(b)$. Now, we can arrange for $G^{-1}(a)$ to be connected by connecting the other components to the component containing $G^{-1}(\partial a)$, using the technique of the previous paragraph at specific points along a. Finally, now that $G^{-1}(a)$ is connected, we look at $G|_{G^{-1}(a)} : G^{-1}(a) \to a$ as a Morse function and use Theorem 1.3 to homotope G so that $G|_{G^{-1}(a)}$ is fiber-connected.

We can now cut Σ along a and X along $G^{-1}(a)$, reducing $-\chi(\Sigma)$. Now pick a new arc in the new Σ and perform the same construction, and iterate until Σ has been reduced to a disk. $\qquad\square$

Uniqueness: Proof of Theorem 1.2 Now we are given a compact, connected, oriented n–manifold X and a 2–manifold Σ and two indefinite Morse 2–functions $G_0, G_1 : X \to \Sigma$ which agree on ∂X and are homotopic rel ∂. We need to construct an indefinite generic homotopy G_s between them, which is fiber-connected when G_0 and G_1 are fiber-connected. Begin with an arbitrary generic homotopy G_s and we will modify this in stages, referring to it as G_s
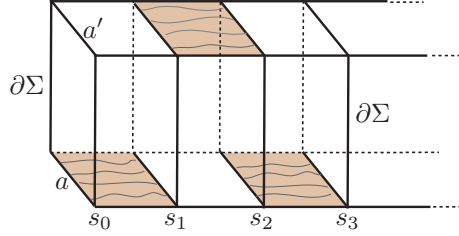
Figure 39: The zig-zag argument for Morse 2–functions

before and after each modification. We also reduce to the fiber-connected case by lifting to a cover, as in preceding proofs.

As in the preceding proof, the base case is when $\Sigma = B^2$, which we have addressed. Now suppose Σ is closed; we wish to reduce to the case with boundary, which we will prove last. Exactly as in the proof of Theorem 1.4, we can use Lemma 6.3 to arrange that, for some fixed point q which is regular for both G_0 and G_1, and for each $s \in [0,1]$, $G_s^{-1}(q)$ is connected. To get connectedness, simply use disjoint arcs to connect components of $G_s^{-1}(q)$ over short overlapping time intervals $s \in [s_{i-1}, s_{i+1}]$, and patch the homotopies coming from Lemma 6.3 together using bump functions. The key point is that the homotopies coming from Lemma 6.3 do not introduce new components while connecting the two given components. Do this also for another point p which is regular for both G_0 and G_1 and arrange that, for every s, either p or q is regular for G_s. Choose an embedded arc a in Σ from p to q. As in the proof of the $\partial\Sigma \neq \emptyset$ case of Theorem 1.1, immediately preceding this proof, G_s can be homotoped so that, for all s, $G_s^{-1}(a)$ is connected, with connected fibers over each point in a.

The zig-zag argument: We now finish the Σ closed case, using the as yet unproved $\partial\Sigma \neq \emptyset$ case, with a zig-zag argument almost identical to the previous one for maps to S^1.

In Figure 39, the t–coordinate is now labelled by s, with $0 = s_1 < s_2 < \ldots < s_{2m} = 1$. The points 0 and 1/2 on S^1 are replaced by p and q in Σ, and the interval $[0, 1/2]$ is replaced by the arc a. The argument proceeds as before, except for the technical point that it is not enough to have G_0 and the intermediate function h_{s_1} agree on the fiber $F_{s_0} = G_0^{-1}(p)$, but they must also agree on a normal 2–disk bundle, so that the closed complement maps $\Sigma \setminus \mathrm{int}(D^2)$. This is easily arranged because p and q are regular values.

Now suppose $\partial\Sigma \neq \emptyset \neq \partial X$. This is proved by induction on $-\chi(\Sigma)$, with the case $\Sigma = B^2$ already proved in Theorem 5.2.

69

As in the preceding proof, the proof of Theorem 1.1, we find an arc b in X with $\partial b \subset \partial X$ which projects to an embedded, non-separating arc a in Σ with $\partial a \subset \partial \Sigma$, with a transverse to the image of the fold locus for G_s at all but finitely many times s. As in the proof of Theorem 1.1, we can arrange that $G_s^{-1}(a)$ is connected with connected fibers for each $s \in [0,1]$. Let a' be a parallel copy of a, and also arrange that $G_s^{-1}(a')$ is connected with connected fibers for each $s \in [0,1]$. Then a and a' are opposite sides of a rectangle R with the other two sides in $\partial \Sigma$.

Now, in Figure 39, we set up the usual zig-zag argument where the vertical interval $[0, 1/2]$ in Figure 37 is replaced by R, with 0 replaced by a and $1/2$ replaced by a'.

As before, choose $0 = s_0 < s_1 < \ldots < s_{2m} = 1$, so that G_s is transverse to a for $s \in [s_{2k}, s_{2k+1}]$ and G_s is transverse to a' for $s \in [s_{2k+1}, s_{2k+2}]$, for all $k \in \{0, \ldots, m-1\}$. It follows that $G_{s_0}^{-1}(a)$ is isotopic (and hence can be taken to be equal) to $G_{s_1}^{-1}(a)$, that $G_{s_1}^{-1}(a') = G_{s_2}^{-1}(a')$, etc. Note that $G_{s_1}^{-1}(R)$ is a connected cobordism between the connected submanifolds $G_{s_1}^{-1}(a)$ and $G_{s_1}^{-1}(a')$, on each of which G_{s_1} is fiber-connected. Thus we can construct an intermediate Morse 2–function H_{s_1} on X which agrees with G_{s_1} outside the interior of $G_{s_1}^{-1}(R)$ and is fiber-connected over R. Then G_0 agrees with H_{s_1} over a, so there exists a fiber-connected homotopy between G_0 and H_{s_1} (because $-\chi(\Sigma \setminus a) < -\chi(\Sigma)$ and the inductive hypothesis holds). Similarly H_{s_1} and G_{s_2} agree over a', so they are homotopic with connected fibers. Thus G_0 and G_{s_2} are homotopic via a fiber-connected homotopy and this argument can be iterated over each interval $[s_{2i}, s_{2i+2}]$, to finish the proof.

$\square$

References

[1] Selman Akbulut and Çağrı Karakurt. 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. *J. Topol.*, 2012; doi: 10.1112/jtopol/jts027.

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

[6] Jean Cerf. La stratification naturelle des espaces de fonctions différentiables réelles et le théorème de la pseudo-isotopie. *Inst. Hautes Études Sci. Publ. Math.*, (39):5–173, 1970.

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

[8] David T. Gay and Robion Kirby. Fiber connected, indefinite Morse 2–functions on connected n–manifolds. *Proc. Natl. Acad. Sci. USA*, 108:8122-8125, 2011. To appear.

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

[10] Ko Honda. Local properties of self-dual harmonic 2-forms on a 4-manifold. *J. Reine Angew. Math.*, 577:105–116, 2004.

[11] Ko Honda. Transversality theorems for harmonic forms. *Rocky Mountain J. Math.*, 34(2):629–664, 2004.

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

[13] Robion Kirby, Paul Melvin, and Peter Teichner. Cohomotopy sets of 4–manifolds. *Proceedings of the Freedman Fest, 2011*, 161–190. *Geom. Topol. Monogr. 18*, Math. Sci. Publ., Berkeley, CA, 2012.

[14] Claude LeBrun. Yamabe constants and the perturbed Seiberg-Witten equations. *Comm. Anal. Geom.*, 5(3):535–553, 1997.

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

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

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

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

[19] Clifford Henry Taubes. *Seiberg Witten and Gromov invariants for symplectic 4-manifolds*, volume 2 of *First International Press Lecture Series*. International Press, Somerville, MA, 2000. Edited by Richard Wentworth.

[20] L. R. Taylor. The principal fibration sequence and the second cohomotopy set. *Proceedings of the Freedman Fest, 2011*, 235–251. *Geom. Topol. Monogr. 18*, Math. Sci. Publ., Berkeley, CA, 2012.

[21] Michael Usher. The Gromov invariant and the Donaldson-Smith standard surface count. *Geom. Topol.*, 8:565–610, 2004.

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

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