Morse theory for manifolds with boundary

Citation for published version:

Digital Object Identifier (DOI):
10.2140/agt.2016.16.971

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Algebraic and Geometric Topology
MORSE THEORY FOR MANIFOLDS WITH BOUNDARY

MACIEJ BORODZIK, ANDRÁS NÉMETHI, AND ANDREW RANICKI

ABSTRACT. We develop Morse theory for manifolds with boundary. Beside standard and expected facts like the handle cancellation theorem and the Morse lemma for manifolds with boundary, we prove that under suitable connectedness assumptions a critical point in the interior of a Morse function can be moved to the boundary, where it splits into a pair of boundary critical points. As an application, we prove that every cobordism of connected manifolds with boundary splits as a union of left product cobordisms and right product cobordisms.

1. Introduction

For some time now, Morse theory has been a very fruitful tool in the topology of manifolds. One of the milestones was the $h$-cobordism theorem of Smale [Sm2], and its Morse-theoretic exposition by Milnor [Mi1, Mi2]. Recently, Morse theory has become even more popular, for two reasons. In the first instance, on account of its connections with Floer homology, see e.g. [Sa, Wi, Ni, KM]. Secondly, the stratified Morse theory developed by Goresky and MacPherson [GM]. In the last 20 years Morse theory has also had an enormous impact on the singularity theory of complex algebraic and analytic varieties.

Despite much previous interest in Morse theory, there still remain uncharted territories. Morse theory for manifolds with boundary is a particular example. The theory was initiated by Kronheimer and Mrowka in [KM], and there is also a recent paper of Laudenbach [La] devoted to the subject. Our paper is a further contribution.

In this paper we prove some new results in the Morse theory for manifolds with boundary. Beside some standard and expected results, like the boundary handle cancellation theorem (Theorem 5.1) and the topological description of passing critical points on the boundary (using the notions of right and left half-handles introduced in Section 2) we discover a new phenomenon. An interior critical point can be moved to the boundary and there split into two boundary critical points. In particular, if we have a cobordism of manifolds with boundary, then under a natural topological assumption we can find a Morse function which has only boundary critical points. We use this result to prove a structure theorem for connected cobordisms of connected manifolds with connected non-empty boundary: such a cobordism splits as a union of left and right product cobordisms. This is a topological counterpart to the algebraic splitting of cobordisms obtained in [BNR]: an algebraic splitting of the chain complex cobordism of a geometric cobordism can be realized topologically by a geometric splitting.

The structure of the paper is the following. After preliminaries in Section 1.1 we study in Section 2 the changes in the topology of the level sets when crossing a boundary critical point. Theorem 2.24 is the main result: passing a boundary stable (unstable) critical point produces a left (right) half-handle attachment. In Section 3 we prove Theorem 3.1.
which moves interior critical points to the boundary. Then we pass to some more standard results, namely rearrangements of critical points in Section 4. We finish the section with our most important — up to now — application, Theorem 4.18, about the splitting of a cobordism into left product and right product cobordisms. Finally, in Section 5 we discuss the possibility of cancelling a pair of critical points.

**Acknowledgements.** The first author wishes to thank Rényi Institute for hospitality, the first two authors are grateful for Edinburgh Mathematical Society for a travel grant to Edinburgh and Glasgow in March 2012. The authors thank András Juhász and András Stipsicz for fruitful discussions and to Rob Kirby for drawing their attention to [KM].

1.1. **Notes on gradient vector fields.** To fix the notation, let us recall what a cobordism of manifolds with boundary is.

**Definition 1.1.** Let $\Sigma_0$ and $\Sigma_1$ be compact oriented, $n$-dimensional manifolds with non-empty boundary $M_0$ and $M_1$. We shall say that $(\Omega, Y)$ is a cobordism between $(\Sigma_0, M_0)$ and $(\Sigma_1, M_1)$, if $\Omega$ is a compact, oriented $(n+1)$-dimensional manifold with boundary $\partial\Omega = Y \cup \Sigma_0 \cup \Sigma_1$, where $Y$ is nonempty, $\Sigma_0 \cap \Sigma_1 = \emptyset$, and $Y \cap \Sigma_0 = M_0$, $Y \cap \Sigma_1 = M_1$.

**Remark 1.2.** Strictly speaking, $\Omega$ is a manifold with corners, so around a point $x \in M_0 \cup M_1$ it is locally modelled by $\mathbb{R}^{n-1} \times \mathbb{R}_{\geq 0}$. Accordingly, sometimes we write that $\Sigma_0$, $\Sigma_1$ and $Y$, as manifolds with boundary, have tubular neighbourhoods in $\Omega$ of the form $\Sigma_0 \times [0, 1)$, $\Sigma_1 \times [0, 1)$, or $Y \times [0, 1)$, respectively. Nevertheless, in most cases it is safe (and more convenient) to assume that $\Omega$ is a manifold with boundary, i.e. that the corners are smoothed along $M_0$ and $M_1$. Whenever possible we make this simplification in order to avoid unnecessary technicalities.

**Example 1.3.** Given a manifold with boundary $(\Sigma, M)$, we call $(\Sigma, M) \times [0, 1) \times \mathbb{R}$ a trivial cobordism, with $\Omega = \Sigma \times [0, 1)$, $Y = M \times [0, 1)$, $\Sigma_i = \Sigma \times \{i\}$, $M_i = M \times \{i\}$ for $i = 0, 1$.

We recall the notion of a Morse function. For this it is convenient to fix a Riemannian metric $g$ on $\Omega$.

**Definition 1.4.** A function $F: \Omega \to [0, 1]$ is called a Morse function on the cobordism $(\Omega, Y)$ if $F(\Sigma_0) = 0$, $F(\Sigma_1) = 1$, $F$ has only Morse critical points, the critical points are not situated on $\Sigma_0 \cup \Sigma_1$, and $\nabla F$ is everywhere tangent to $Y$.

There are two ways of doing Morse theory on manifolds. One can either consider the gradient flow of $\nabla F$ associated with $F$ and the Riemannian metric (in the Floer theory, one often uses $\nabla F$), or, the so-called gradient-like vector field.

**Definition 1.5.** (See [Mi2] Definition 3.1.) Let $F$ be a Morse function on a cobordism $(\Omega, Y)$. Let $\xi$ be a vector field on $\Omega$. We shall say that $\xi$ is gradient-like with respect to $F$, if the following conditions are satisfied:

(a) $\xi \cdot F > 0$ away from the set of critical points of $F$;

(b) if $p$ is a critical point of $F$ of index $k$, then there exist local coordinates $x_1, \ldots, x_{n+1}$ in a neighbourhood of $p$, such that

$$F(x_1, \ldots, x_{n+1}) = F(p) - (x_1^2 + \cdots + x_k^2) + (x_{k+1}^2 + \cdots + x_{n+1}^2)$$

and

$$\xi = (-x_1, \ldots, -x_k, x_{k+1}, \ldots, x_{n+1}) \text{ in } U;$$

(b') furthermore, if $p$ is a boundary critical point, then the above coordinate system can be chosen so that $Y = \{x_j = 0\}$ and $U = \{x_j \geq 0\}$ for some $j \in \{1, \ldots, n+1\}$.

(c) $\xi$ is everywhere tangent to $Y$;
The conditions (a) and (b) are the same as in the classical case. Condition (b\textsuperscript{'}) is a specification of the condition (b) in the boundary case, compare Lemma 2.6.

Smale in [Sm1] noticed that for any gradient-like vector field \( \xi \) for a function \( F \) there exists a choice of the Riemannian metric such that \( \xi = \nabla F \) in that metric. The situation is identical in the boundary case. This is stated explicitly in the following lemma, whose proof is straightforward and will be omitted.

**Lemma 1.6.** Let \( U \) be a paracompact \( k \)-dimensional manifold and \( F: U \to \mathbb{R} \) a Morse function without critical points. Assume that \( \xi \) is a gradient–like vector field on \( U \). Then there exists a metric \( g \) on \( U \) such that \( \xi = \nabla F \) in that metric.

Similar statement holds if \( U \) has boundary and \( \xi \) is everywhere tangent to the boundary.

Hence the two approaches – gradients and gradient-like vector fields – are equivalent. However, we shall need both approaches. In Section 3 we use gradients of functions and a specific choice of a metric, because the argument becomes slightly simpler. In Section 5 we follow [Mi2] very closely; as he uses gradient-like vector fields, we use them as well.

The next result shows that the condition from Definition 1.4 that \( \nabla F \) is everywhere tangent to \( Y \) can be relaxed. We shall use this result in Proposition 4.1.

**Lemma 1.7.** Let \( \Omega \) be a compact, Riemannian manifold of dimension \( (n + 1) \) and \( Y \subset \partial \Omega \) be compact as well. Let \( g \) denote the metric. Suppose that there exists a function \( F: \Omega \to \mathbb{R} \), and a relative open subset \( U \subset Y \) such that \( \nabla F \) is tangent to \( Y \) at each point \( y \in U \). Suppose furthermore, that for any \( y \in Y \setminus U \) we have

\[
\ker dF \not\subset T_y Y.
\]

Then, for any open neighbourhood \( W \subset \Omega \) of \( Y \setminus U \), there exist a metric \( h \) on \( \Omega \), agreeing with \( g \) away from \( W \), such that \( \nabla h F \) (the gradient in the new metric) is everywhere tangent to \( Y \).

**Proof.** Let us fix a point \( y \in Y \setminus U \) and consider a small open neighbourhood \( U_y \) of \( y \) in \( W \), in which we choose local coordinates \( x_1, \ldots, x_{n+1} \) such that \( Y \cap U_y = \{ x_{n+1} = 0 \} \) and \( U_y \subset \{ x_{n+1} \geq 0 \} \). In these coordinates we have \( dF = \sum_{i=1}^{n+1} f_i(x) \, dx_i \) for some smooth functions \( f_1, \ldots, f_{n+1} \). By the assumption, for each \( x \in U_y \), there exists \( i \leq n \) such that \( f_i(x) \neq 0 \). Shrinking \( U_y \) if needed, we may assume that for each \( x \in U_y \) we have \( f_i(x) \neq 0 \) for some \( i \). We may suppose that \( i = 1 \), hence \( \pm f_1(x) > 0 \). Let us choose a symmetric positive definite matrix \( A_y = (a_{ij}(x))_{i,j=1}^{n+1} \) so that \( a_{11} = \pm f_1(x) \) and for \( i > 1 \), \( a_{ii} = a_{i1} = f_i(x) \). \( A_y \) defines a metric \( h_y \) on \( U_y \) such that \( \nabla h_y F = (\pm 1, 0, \ldots, 0) \subset TY \) in that metric.

Now let us choose an open subset \( V \) of \( \Omega \setminus (Y \setminus U) \) such that \( V \cup \bigcup_{y \in Y \setminus U} U_y \) is a covering of \( \Omega \). Let \( \{ \phi_V \} \cup \{ \phi_y \}_{y \in Y \setminus U} \) be a partition of unity subordinate to this covering. Define

\[
h = \phi_V \cdot g + \sum_{y \in Y \setminus U} \phi_y h_y.
\]

Then \( h \) is a metric, which agrees with \( g \) away from \( W \). Moreover, as for each metric \( h_y \), and \( x \in U_y \cap Y \) we have \( \nabla h_y F(x) \in T_x Y \) by construction, the same holds for a convex linear combination of metrics. \( \square \)

2. **Boundary stable and unstable critical points**

2.1. **Morse function for manifolds with boundary.** The whole discussion of Morse functions on manifolds with boundary would be pointless if we did not have the following.

**Lemma 2.1.** Morse functions exist. In fact, for any Morse function \( f: Y \to [0,1] \) with \( f(M_0) = 0, f(M_1) = 1 \) there exists a Morse function \( F: \Omega \to [0,1] \) whose restriction to \( Y \) is \( f \).
Proof. Let \( f : Y \to [0, 1] \) be a Morse function on the boundary, such that \( f(M_0) = 0 \) and \( f(M_1) = 1 \). We want to extend \( f \) to a Morse function on \( \Omega \).

First, let us choose a small tubular neighbourhood \( U \) of \( Y \) and a diffeomorphism \( U \cong Y \times [0, \varepsilon) \) for some \( \varepsilon > 0 \). Let \( \tilde{F} : U \to [0, 1] \) be given by the formula
\[
(2.2) \quad U \cong Y \times [0, \varepsilon) \ni (x, t) \to \tilde{F}(x, t) = f(x) - f(x)(1 - f(x))t^2.
\]
The factor \( f(x)(1 - f(x)) \) ensures that \( \tilde{F} \) attains values in the interval \([0, 1]\) and \( \tilde{F}^{-1}(i) \subset \Sigma_i \) for \( i \in \{0, 1\} \). It is obvious that there exists a smooth function \( F : \Omega \to [0, 1] \), which agrees on \( Y \times [0, \varepsilon/2) \) with \( \tilde{F} \), and it satisfies the Morse condition on the whole \( \Omega \). The gradient \( \nabla F \) is everywhere tangent to \( Y \).

Remark 2.3. The above construction yields a function with the property that all its boundary critical points are boundary stable (see Definition 2.4 below). This is due to the choice of sign \(-1\) in front of \( f(x)(1 - f(x))t^2 \) in (2.2). If we change the sign to \(+1\), we obtain a function with all boundary critical points boundary unstable.

We fix a Morse function \( F : \Omega \to [0, 1] \) and we start to analyze its critical points. Let \( z \) be such a point. If \( z \in \Omega \setminus Y \), we shall call it an interior critical point. If \( z \in Y \), it will be called a boundary critical point. There are two types of boundary critical points.

Definition 2.4. Let \( z \) be a boundary critical point. We shall call it boundary stable, if the tangent space to the unstable manifold of \( z \) lies entirely in \( T_z Y \), otherwise it is called boundary unstable.

The index of the boundary critical point \( z \) is defined as the dimension of the stable manifold \( W^s_z \). If \( z \) is boundary unstable, this is the same as the index of \( z \) regarded as a critical point of the restriction \( f \) of \( F \) on \( Y \). If \( z \) is boundary stable, we have \( \text{ind}_F z = \text{ind}_f z + 1 \). In particular, there are no boundary stable critical point with index 0, nor boundary unstable critical points of index \( n + 1 \).

Remark 2.5. We point out that we use the flow of \( \nabla F \) and not of \(-\nabla F\) as Kronheimer and Mrowka [KM] do, hence our definitions and formulae are slightly different from theirs.

We finish this subsection with three standard results.

Lemma 2.6 (Boundary Morse Lemma). Assume that \( F \) has a critical point \( z \in Y \) such that the Hessian \( D^2F(z) \) at \( z \) is non-degenerate, and \( \nabla F \) is everywhere tangent to \( Y \).
Then there are local coordinates \((x_1, \ldots, x_{n+1})\) in an open neighbourhood \(U \ni z\) such that
\[
U = \{x_1^2 + \cdots + x_{n+1}^2 \leq \varepsilon^2\} \cap \{x_1 \geq 0\}
\]
and \(U \cap Y = \{x_1 = 0\}\) for some \(\varepsilon > 0\), and \(F\) in these coordinates has the form 
\[
\pm u_1^2 \pm x_2^2 \pm \cdots \pm x_{n+1}^2 + F(z).
\]

**Proof.** We choose a coordinate system \(y_1, \ldots, y_{n+1}\) in a neighbourhood \(U \subset \Omega\) of \(z\) such that 
\(z = (0, \ldots, 0)\), \(Y = \{y_1 = 0\}\), \(U = \{y_1 \geq 0\}\), and the vector field \(\frac{\partial \phi}{\partial y_1}\) is orthogonal to \(Y\). We might assume \(F(z) = 0\). The tangency of \(\nabla F\) to \(Y\) implies that at each point of \(Y\)
\[
\frac{\partial F}{\partial y_1}(0, y_2, \ldots, y_{n+1}) = 0.
\]

By Hadamard’s lemma one writes
\[
F = y_1 K_1(y_1, \ldots, y_{n+1}) + \sum_{j=2}^{n+1} y_j K_j(y_1, y_2, \ldots, y_{n+1}),
\]

for some functions \(K_1, \ldots, K_{n+1}\). We can assume that for \(j > 1\), \(K_j\) does not depend
on \(y_1\). Indeed, if it does depend, we might write 
\(K_j(y_1, \ldots, y_{n+1}) = K_j(0, y_2, \ldots, y_{n+1}) + y_1 L_{jj}(y_1, \ldots, y_{n+1})\), and then replace \(K_1\) by 
\(K_j(0, y_2, \ldots, y_{n+1})\) and \(K_1\) by \(K_1 + \sum y_j L_{jj}\).

The condition (2.7) implies now that \(K_1(0, y_2, \ldots, y_{n+1}) = 0\), hence \(K_1 = y_1 H_{11}(y_1, \ldots, y_{n+1})\).

By Hadamard’s lemma applied to \(K_2, \ldots, K_{n+1}\) we get
\[
F = y_1^2 H_{11}(y_1, \ldots, y_{n+1}) + \sum_{j,k=2}^{n} y_j y_k H_{jk}(y_2, \ldots, y_{n+1}).
\]

The non-degeneracy of \(D^2 F(z)\) means that \(H_{11}(z) \neq 0\). Then, after replacing 
\(y_1 \sqrt{\pm H_{11}}\) by \(x_1\), we can assume that \(H_{11} = \pm 1\). Finally, the sum in (2.9) can be written as \(\sum_{j \geq 2} \varepsilon_j x_j^2\) 
\((\varepsilon_j = \pm 1)\) by the classical Morse lemma [Mi1, Lemma 2.2].

The next result is completely standard by now.

**Lemma 2.10.** Assume that \(F\) is a Morse function on a cobordism \((\Omega, Y)\) between \((\Sigma_0, M_0)\)
and \((\Sigma_1, M_1)\). If \(F\) has no critical points then \((\Omega, Y) \cong (\Sigma_0, M_0) \times [0, 1]\). Furthermore, we
can choose the diffeomorphism to map the level set \(F^{-1}(t)\) to the set \(\Sigma_0 \times \{t\}\).

**Proof.** The proof is identical to the classical case, see e.g. [Mi2, Theorem 3.4].

\[
\text{2.2. Half-handles.}\] For any \(k\) we consider the \(k\)-dimensional disc \(D^k = \{x_1^2 + \cdots + x_k^2 \leq 1\}\).

In the classical theory, an \(n\)-dimensional handle of index \(k\) is the \(n\)-dimensional manifold 
\(H = D^k \times D^{n-k}\) with boundary 
\[
\partial H = \left(\partial D^k \times D^{n-k}\right) \cup \left(D^k \times \partial D^{n-k}\right) = B_0 \cup B'_0.
\]

Given an \(n\)-manifold with boundary \((\Sigma, \partial \Sigma)\) and a distinguished embedding \(\phi: B_0 \to \partial \Sigma\),
the effect of a classical handle attachment is the \(n\)-dimensional manifold with boundary 
\[
(\Sigma', \partial \Sigma') = (\Sigma \cup H, (\partial \Sigma \setminus B_0) \cup B'_0),
\]

where we glue along \(\phi(B_0)\) identified with \(B_0\). The boundary \(\partial \Sigma'\) is the effect of surgery
on \(\phi(B_0) \subset \partial \Sigma\). We now extend this construction to relative cobordisms of manifolds with
boundary, using ‘half-handles’. Since our ambient space \(\Omega\) is \((n+1)\)-dimensional, \((n+1)\)
is the dimension of the handles, and they induce \(n\)-dimensional handle attachments on \(Y\).

In order to do this, for any \(k \geq 1\) we distinguish the following subsets of \(D^k\): the
‘half-disc’ \(D^k_+ := D^k \cap \{x_1 \geq 0\}\), and its boundary subsets \(S^{k-1}_+ := \partial D^k \cap \{x_1 \geq 0\}\),
\(S^{k-2}_0 := \partial D^k \cap \{x_1 = 0\}\) and \(D^{k-1}_0 := D^k \cap \{x_1 = 0\}\). Clearly, \(S^{k-2}_0\) is a boundary the two
\((k-1)\)-discs \(S^{k-1}_+\) and \(D^{k-1}_0\). We will call \(x_1\) the *cutting coordinate*. 
Definition 2.11. Let \(0 \leq k \leq n\). An \((n+1)\)-dimensional right half-handle of index \(k\) is the \((n+1)\)-dimensional manifold \(H_{\text{right}} = D^k \times D^{n+1-k}_+\), with boundary subdivided into three pieces \(\partial H_{\text{right}} = B \cup C \cup N\), where

\[
B := \partial D^k \times D^{n+1-k}_+, \quad C := D^k \times D_0^{n-k}, \quad N := D^k \times S_{n-k}^1.
\]

One has the following intersections too

\[
B_0 := C \cap B = \partial D^k \times D_0^{n-k}, \quad N_0 := C \cap N = D^k \times S_{n-k-1}^0.
\]

Hence the handle \(H\) is cut along \(C\) into two pieces, one of them is the half-handle \(H_{\text{right}}\). Note that \((C, B_0)\) is a \(n\)-dimensional handle of index \(k\).

Symmetrically, we define the left half-handles by cutting the handle \(H\) along the left–component disc \(D^k\).

Definition 2.12. Fix \(k\) with \(1 \leq k \leq n+1\). An \((n+1)\)-dimensional left half-handle of index \(k\) is the \((n+1)\)-dimensional disk \(H_{\text{left}} := D^k_+ \times D^{n+1-k}_0\) with boundary subdivided into three pieces \(\partial H_{\text{left}} = B \cup C \cup N\), where

\[
B := S^{k-1}_+ \times D^{n+1-k}_+, \quad C := D^{k-1}_0 \times D^{n+1-k}_0, \quad N := D^k \times \partial D^{n+1-k}.
\]

Furthermore, we specify \(B_0 := C \cap B = S^{k-2}_0 \times D^{n+1-k}_0\) and \(N_0 := N \cap C = D_0^{k-1} \times \partial D^{n+1-k}\).

A half-handle will from now on refer to either a right half-handle or left half-handle. We pass to half-handle attachments. We will attach a half-handle along \(B\). The definitions of the right half-handle attachment and the left half-handle attachment are formally very similar, but there are significant differences in the properties of the two operations.

Definition 2.13. Let \((\Omega, Y; \Sigma_0, M_0, \Sigma_1, M_1)\) be an \((n+1)\)-dimensional relative cobordism. Given an embedding \(\Phi\): \((B, B_0) \hookrightarrow (\Sigma_1, M_1)\) define the relative cobordism \((\Omega', Y'; \Sigma_0, M_0, \Sigma_1', M_1')\) obtained from \((\Omega, Y; \Sigma_0, M_0, \Sigma_1, M_1)\) by attaching a (right or left) half-handle of index \(k\) by

\[
\begin{align*}
\Omega' &= \Omega \cup_B H, & Y' &= Y \cup_{B_0} C, \\
\Sigma_1' &= (\Sigma_1 \setminus B) \cup N, & M_1' &= (M_1 \setminus B_0) \cup N_0.
\end{align*}
\]

See Figure 4 and Figure 5 for right, respectively left half-handle attachments.

We point out that in the case of the right half-handle attachment, any embedding of \(B_0\) into \(M_1\) determines (up to an isotopy) an embedding of pairs \((B, B_0) \hookrightarrow (\Sigma_1, M_1)\). Indeed, as \((B, B_0) = \partial D^k \times (D^{n+1-k}_+ \setminus D_0^{n-k})\), a map \(\phi\): \(B_0 \hookrightarrow M_1\) extends to a map \(\Phi\): \(B \hookrightarrow \Sigma_1\) in a collar neighbourhood of \(M_1\) in \(\Sigma_1\). (This is not the case in the left half-handle attachment.)

In particular, in the case of right attachments, we specify only the embedding \(B_0 \hookrightarrow M_1\).
Example 2.14. (a) We exemplify the left half-handle attachment for $k = 1$. In this case $B_0$ is empty. If we are given an embedding of $B \cong \{1\} \times D^n$ into $\Sigma_1 \setminus M_1$, we glue $[0, 1] \times D^n$ to $\Omega$ along $B$. Then we set $Y' = Y \cup \{0\} \times D^n$, $\Sigma'_1 = (\Sigma_1 \setminus B) \cup [0, 1] \times \partial B$ and $M' = M \cup \{0\} \times \partial B$.

(b) The right half-handle attachment of index 0 is the disconnected sum $\Omega \sqcup D^{n+1}$ with boundary $\partial D^{n+1}_+ = S^n_+ \cup D^n_0$. We think of first disc $S^n_+$ as a part of $\Sigma'_1$, while the second disc as a part of $Y'$, and $M'_1 = M_1 \cup S_0^{n-1}$.

Remark 2.15. In the next subsection we shall see that crossing a boundary stable critical points corresponds to left half-handle attachment, while a boundary unstable critical point corresponds to a right half-handle attachment. Theorem 3.1 can be interpreted informally, as splitting a handle into a right half-handle and left half-handle. This also motivates the name ‘half-handle’.

2.3. Elementary properties of half-handle attachments. The following results are trivial consequences of the definitions.

Lemma 2.16. Let $\Omega'$ be the result of a right half-handle attachment to $\Omega$ along $(B, B_0) \hookrightarrow (\Sigma_1, M_1)$. Let $B'$ be $B$ pushed slightly off $M_1$ into the interior of $\Sigma_1$. Let $\Omega$ be a result of attaching a (standard) handle of index $k$ to $\Omega$ along $B'$. Then $\Omega'$ and $\Omega$ are diffeomorphic.

Proof. This is obvious, since when we forget about $C$ and $B_0$, the pair $(H_{\text{right}}, B)$ is a standard $(n + 1)$-dimensional handle of index $k$. \qed

In particular, the effect of a right half-handle attachment on $\Omega$ is the same as the effect of a standard handle attachment of the same index.

The situation is completely different in the case of left half-handle attachments.

Lemma 2.17. If $\Omega'$ is a result of a left half-handle attachment, then $\Omega'$ is diffeomorphic to $\Omega$.

Proof. By definition the pair $(H_{\text{left}}, B)$ is diffeomorphic to the pair $(D^n \times [0, 1], D^n \times \{0\})$. Attaching $H_{\text{left}}$ along $B$ to $\Omega$ does not change the diffeomorphism type of $\Omega$. \qed

The effect on $Y$ of a right/left half-handle attachments are ‘almost’ the same.

Lemma 2.18. If $(\Omega', Y'; \Sigma_0, M_0, \Sigma'_1, M'_1)$ is the result of left (respectively, right) half-handle attachment to $(\Omega, Y; \Sigma_0, M_0, \Sigma_1, M_1)$ along $(B, B_0) \hookrightarrow (\Sigma_1, M_1)$, then $Y'$ is a result of a classical handle attachment of index $k - 1$ (respectively, $k$) along $B_0$.

Proof. This follows immediately from Definition 2.13. \qed

The effects of half handle attachment on $\Sigma$ are also easily described. The next lemma is a direct consequence of the definitions; its proof is omitted. We refer to Figures 4 and 5.
Lemma 2.19. (a) If \((\Omega', Y'; \Sigma_0, M_0, \Sigma'_1, M'_1)\) is the result of left half-handle attachment to \((\Omega, Y; \Sigma_0, M_0, \Sigma_1, M_1)\) along \((B, B_0) \hookrightarrow (\Sigma_1, M_1)\), then
\[\Sigma'_1 \cong \Sigma_1 \setminus B.\]

(b) If \((\Omega', Y'; \Sigma_0, M_0, \Sigma'_1, M'_1)\) is the result of index \(k\) right half-handle attachment to \((\Omega, Y; \Sigma_0, M_0, \Sigma_1, M_1)\) along \(B_0 \hookrightarrow M_1\), then
\[\Sigma'_1 \cong \Sigma_1 \cup B_0 N,\]
where here \(N\) is an \(n\)-dimensional disk \(D^k \times D^{n-k}\) and \(B_0 = S^{k-1} \times D^{n-k}\).

These facts also emphasize that right half-handle attachments and left half-handle attachments are somehow dual operations on \(\Sigma\). This can be seen also at the Morse function level: changing a Morse function \(F\) to \(-F\) changes all right half-handles to left-half handles and conversely, see Section 2.4 and 2.5 below. But the above lemma shows another aspect as well: a right half handle attachment consists of gluing a disk, a left half-handle attachment consists of removing a disk. Indeed, in the case of right attachment, \((\Sigma'_1, M'_1) = (\Sigma_1 \cup D^k \times D^{n-k}, \partial \Sigma_1)\) associated with an embedding \(\Phi : \partial D^k \times D^{n-k} \to M_1\). On the other hand, by definition, for an embedding \(\Phi' : (D^{k-1} \times D^{n+1-k}, \partial D^{k-1} \times D^{n+1-k}) \to (\Sigma_1, M_1)\) the pair
\[(\Sigma'_1, M'_1) = (\text{closure of } (\Sigma_1 \setminus D^{k-1} \times D^{n+1-k}), \partial \Sigma'_1)\]
is obtained from \((\Sigma_1, M_1)\) by a handle detachment of index \(k - 1\). Hence we obtain:

Corollary 2.20. The effect on \((\Sigma_1, M_1)\) of a right half–handle attachment of index \(k\) (at the level of cobordism) is a handle attachment of index \(k\) at the level of \((\Sigma_1, M_1)\). Conversely, the effect on \((\Sigma_1, M_1)\) of a left half–handle attachment of index \(k\) (at the level of cobordism) is a handle detachment of index \(k - 1\) at the level of \((\Sigma_1, M_1)\). In particular, \(M'_1\) is obtained from \(M_1\) as the result of a \(k\) surgery in the first case, and \((k - 1)\) surgery in the second.

The duality can be seen also as follows: we can cancel any handle attachment by a suitably defined handle detachment, and conversely.

The following definition introduces a terminology which is rather self-explanatory. We include it for completeness of the exposition.

Definition 2.21. We shall say that a cobordism \((\Omega', Y')\) between \((\Sigma, M)\) and \((\Sigma', M')\) is a right (respectively left) half-handle attachment of index \(k\), if \((\Omega', Y', \Sigma', M')\) is a result of right (respectively left) half-handle attachments of index \(k\) (in the sense of Definition 2.13) to \((\Sigma \times [0, 1], M \times [0, 1], \Sigma \times \{0\}, M \times \{0\}, \Sigma \times \{1\}, M \times \{1\})\).
We conclude this section by studying homological properties of the handle attachment. These properties will be used in [BNR]. The proofs are standard and are left to the reader.

Let \((H_+, C, B, N)\) be a half-handle of index \(k\).

**Lemma 2.22.** If \((H_{\text{right}}, C, B, N)\) is a right half-handle, then the pair \((C, B_0)\) is a strong deformation retract of \((H_+^*, B)\), while \((D^k, \partial D^k)\) is a strong deformation retract of \((C, B_0)\). In particular, \(H_j(H_+^*, B) \cong H_j(C, B_0) = \mathbb{Z}\) for \(j = k\), and it is zero otherwise.

The situation is completely different for left half-handles.

**Lemma 2.23.** If \((H_{\text{left}}, C, B, N)\) is a left half-handle, then the pair \((H_{\text{left}}, B)\) retracts onto the trivial pair \((\text{point}, \text{point})\). In particular, all the relative homologies \(H_*(H_{\text{left}}, B)\) vanish. On the other hand, \((D_0^{k-1}, S_0^{k-2})\) is a strong deformation retract of \((C, B_0)\), hence \(H_j(C, B_0) = \mathbb{Z}\) for \(j = k - 1\), and it is zero otherwise. Therefore, the inclusion \((C, B_0) \hookrightarrow (H_{\text{left}}, B)\) induces a surjection on homologies.

### 2.4. Boundary critical points and half-handles.

Consider a Morse function \(F\) on a cobordism \((\Omega, Y)\) and assume that it has a single boundary critical point \(z\) of index \(k\) with critical value \(c\).

**Theorem 2.24.** If \(z\) is boundary stable (unstable), then the cobordism is a left (right) half-handle attachment of index \(k\) respectively.

**Proof.** We can assume that \(c = F(z) = 0\). Let \(\rho\) be a positive number small enough and let \(U\) be a ‘half’ ball around \(z\) of radius \(2\rho\) such that in \(U\) we can choose Morse coordinates \(x_1, \ldots, x_{n+1}\) (cf. Lemma 2.6) with

\[
U = \{x_1^2 + \cdots + x_{n+1}^2 \leq 4\rho^2\} \cap \{x_1 \geq 0\},
\]

and \(Y \cap U\) defined by \(\{x_1 = 0\}\), and

\[
F(x_1, \ldots, x_{n+1}) = -a^2 + b^2,
\]

where if \(z\) is boundary stable we set

\[
a^2 = x_1^2 + x_2^2 + \cdots + x_k^2, \quad b^2 = x_{k+1}^2 + \cdots + x_{n+1}^2 \quad (k \geq 1),
\]

and if \(z\) is boundary unstable

\[
a^2 = x_1^2 + \cdots + x_{k+1}^2, \quad b^2 = x_1^2 + x_{k+2}^2 + \cdots + x_{n+1}^2 \quad (k \geq 0).
\]

We also assume that \(x_1, \ldots, x_{n+1}\) is an Euclidean orthonormal coordinate system.

Next, we consider \(\varepsilon > 0\) such that \(\varepsilon \ll \rho\), and we define the space \(\tilde{H}\) bounded by the following conditions (see Figure [5])

\[
\tilde{H} := \{-a^2 + b^2 \in [-\varepsilon^2, \varepsilon^2], \quad a^2b^2 \leq (\rho^4 - \varepsilon^4)/4, \quad x_1 \geq 0\}.
\]
Lemma 2.28. The flow of $\nabla F$ is tangent to $\overline{K}$.

Proof. Assume the critical point is boundary stable. Then the time $t$ flow acts by

$$(x_1, \ldots, x_{n+1}) \rightarrow (e^{-2t}x_1, \ldots, e^{-2t}x_k, e^{2t}x_{k+1}, \ldots, e^{2t}x_{n+1}),$$

hence $a^2 \rightarrow e^{-4t}a^2$ and $b^2 \rightarrow e^{4t}b^2$, and the hypersurface $a^2b^2 = \text{const}$ is preserved. \hfill $\square$

Lemma 2.29. The inclusion of pair of spaces

$$\left( F^{-1}(-\varepsilon^2) \cup_{\overline{B}} \overline{H}, Y \cap \left( F^{-1}(-\varepsilon^2) \cup_{\overline{B}} \overline{H} \right) \right) \subset (\Omega, Y)$$

admits a strong deformation retract (inducing an isomorphism of these pairs).

Proof. By Lemma 2.10 we can assume that $(\Omega, Y)$ is $(F^{-1}([-\varepsilon^2, \varepsilon^2]), Y \cap F^{-1}([-\varepsilon^2, \varepsilon^2]))$.

First we assume that $\overline{B}$ is not empty, and it is given by the equation (2.27) in $U$. Set $\Sigma_- = \text{closure of } (F^{-1}(-\varepsilon^2) \setminus \overline{B})$ and let $T_- = \text{the part of the boundary of } \Sigma_-$ given by

$$T_- = \text{closure of } (\partial \Sigma_- \setminus \partial F^{-1}(-\varepsilon^2)).$$

Obviously $T_- \subset \overline{B}$, see Figure 7. Let us choose a collar of $T_-$ in $\Sigma_-$, that is a subspace $U_- \subset \Sigma_-$ diffeomorphic to $T_- \times [0, 1]$, $T_-$ identified with $T_- \times \{0\}$ and $\partial T_- \times [0, 1] \subset \partial \Sigma_- \setminus \partial F^{-1}(-\varepsilon^2)$. Let $T'_-$ be the space identified with $T_- \times \{1\}$ by this diffeomorphism.

Similarly, let $\Sigma_+ = \text{closure of } (F^{-1}(\varepsilon^2) \setminus \overline{P})$, and $T_+ = \text{closure of } (\partial \Sigma_+ \setminus \partial F^{-1}(\varepsilon^2))$. We also define $\Omega_0$ as the closure of $\Omega \setminus \overline{H}$. Clearly $F$ has no critical points in $\Omega_0$ and $\nabla F$ is
everywhere tangent to $\partial \Omega \setminus (\Sigma_- \cup \Sigma_+) = (Y \cap \Omega_0) \cup \tilde{K}$ by Lemma 2.28. In particular, by Lemma 2.10, the flow of $\nabla F$ on $\Omega_0$ yields a diffeomorphism between $\Sigma_-$ and $\Sigma_+$, mapping $T_-$ to $T_+$. We define $V \subset \Omega$ as the closure of the set of points $v$ such that a trajectory going through $v$ hits $U_-$. Lemma 2.10 implies that there is a diffeomorphism $V \cong T_- \times [0,1] \times [-\epsilon^2,\epsilon^2]$ such that for $(x,t,s) \in V$ we have $F(x,t,s) = s$. Finally, we also define $V^* := \{(x,t,s) \in V : s \leq \epsilon^2(1-2t)\}$.

We define the contraction in two steps: vertical and horizontal. The vertical contraction is defined as follows. For $v \in \tilde{H} \cup V^*$ we define $\Pi_V(v) = v$. For a point $v \in \Omega_0 \setminus V$ we take for $\Pi_V(v)$ the unique point $s \in \Sigma_-$ such that a trajectory of $\nabla F$ goes from $s$ to $v$. Finally, if $v = (x,t,s) \in V \setminus V^*$ we define $\Pi_V(v) = (x,t,\epsilon^2(1-2t))$.

By construction, the image of $\Pi_V$ is $\tilde{H} \cup V^* \cup F^{-1}(-\epsilon^2)$. Next, we define $\Pi_H$.

It is an identity on $\tilde{H} \cup F^{-1}(-\epsilon^2)$, and maps $(x,t,s) \in V^*$ to $(x,t-(\epsilon^2+s)/(2\epsilon^2),-\epsilon^2)$ if $s \leq \epsilon^2(2t-1)$, and to $(x,0,-\epsilon^2-2\epsilon^2t)$ otherwise. Note that the expressions agree for any $(x,t,s)$ with $s = \epsilon^2(2t-1)$ and these points are sent to $(x,0,\epsilon^2)$. Both $\Pi_H$ and $\Pi_V$ are continuous retractions, by smoothing corners we can modify them into smooth retractions; also they can be extended in a natural way to strong deformation retracts. By construction, the retracts preserve $Y$ too. See also Figure 8.

If $\tilde{B}$ is empty, then $\tilde{H}$ is necessarily a unstable (right) half-handle of index 0, $F^{-1}([-\epsilon^2,\epsilon^2])$ is a disconnected sum of $\tilde{H}$ and the manifold $F^{-1}(-\epsilon^2) \times [-\epsilon^2,\epsilon^2]$. □

**Continuation of the proof of Theorem 2.24.** We want to show that $\tilde{H}$ is a half-handle.

By subsection 2.2 we have the following description in local coordinates of the left half-handle (2.30) and right half-handle (2.31) with cutting coordinate $x_1$:

(2.30) $H_{left} = \{x_1^2 + \cdots + x_k^2 \leq 1\} \cap \{x_{k+1}^2 + \cdots + x_{n+1}^2 \leq 1\} \cap \{x_1 \geq 0\}$
We consider the subsets $R$ and $S$ of $\mathbb{R}^2$ and a diffeomorphism $\psi: R \to S$ continuous on the boundary, where

\[ R = \{(u, v) \in \mathbb{R}^2 : u \geq 0, \ v \geq 0, \ uv \leq (\rho^4 - \varepsilon^4)/4, \ -u + v \in [-\varepsilon^2, \varepsilon^2]\}, \]

\[ S = \{(u, v) \in \mathbb{R}^2 : u \in [0, \varepsilon], \ v \in [0, \varepsilon]\}. \]

Assume that $\psi$ maps the edge of $R$ given by $\{-u + v = -\varepsilon^2\}$ to the edge $\{u = \varepsilon\}$ of $S$ and the images of coordinate axes are the corresponding coordinate axes. (Note that $R$ can be seen on Figure 6 if we replace $a^2$ by $u$ and $b^2$ by $v$.) We lift $\psi$ to a diffeomorphism $\Psi$ between $\tilde{H}$ and $H_{\text{right}}$ (respectively, $H_{\text{left}}$) as follows. First let us write $\psi(u, v) = (\psi_1(u, v), \psi_2(u, v))$. As $\psi$ maps axes to axes, we have $\psi_1(0, v) = 0$ and $\psi_2(u, 0) = 0$. Furthermore $\psi_1, \psi_2 \geq 0$. By Hadamard’s lemma there exists smooth functions $\xi$ and $\eta$ such that

\[ \psi(u, v) = (u\xi(u, v)^2, v\eta(u, v)^2). \]

We define now

\[ \Psi(x_1, \ldots, x_{n+1}) = (\xi(a, b)x_1, \ldots, \xi(a, b)x_k, \eta(a, b)x_{k+1}, \ldots, \eta(a, b)x_{n+1}) \]

if $z$ is boundary stable, and

\[ \Psi(x_1, \ldots, x_{n+1}) = (\eta(a, b)x_1, \xi(a, b)x_2, \ldots, \xi(a, b)x_k, \eta(a, b)x_{k+1}, \ldots, \eta(a, b)x_{n+1}) \]

if $z$ is boundary unstable. Here $a$ and $b$ are given by (2.25) or (2.26). By the very construction, $\Psi$ maps $(\tilde{H}, \tilde{B}, \tilde{C})$ diffeomorphically to the triple $(H, B, C)$, where

\[ H = \{a^2 \in [0, \varepsilon^2], \ b^2 \in [0, \varepsilon^2], \ x_1 \geq 0\} \]

\[ B = \{a^2 = -\varepsilon^2, \ b^2 \in [0, \varepsilon^2], \ x_1 \geq 0\} \]

\[ C = \{a^2 \in [0, \varepsilon^2], \ b^2 \in [0, \varepsilon^2], \ x_1 = 0\}. \]

After substituting for $a$ and $b$ the values from (2.25) or (2.26) (depending on whether $z$ is boundary stable or unstable), we recover the model (2.31) of a right half-handle if $z$ is boundary unstable; or the model (2.30) of a left half-handle (both of index $k$).

The fact that each half-handle can be presented in a left or right model will be now used to show the following converse to Theorem 2.24.

**Proposition 2.32.** Let $(\Omega, Y) = (\Sigma_0 \times [0, 1], M_0 \times [0, 1])$ be a product cobordism between $(\Sigma_0, M_0)$ and $(\Sigma_1, M_1) \cong (\Sigma_0, M_0)$. Let us be given a half-handle $(H, C, B)$ of index $k$ and an embedding of $B_0 = C \cap B$ into $M_1$ (respectively an embedding of $(B, B_0)$ into $(\Sigma_1, M_1)$), and let $(\Omega', Y')$ be the result of a right half-handle attachment along $B_0$ (respectively, a left half-handle attachment along $(B, B_0)$) of index $k$. Then, there exists a Morse function $F: (\Omega', Y') \to \mathbb{R}$, which has a single boundary unstable critical point (respectively, a single boundary stable critical point) of index $k$ on $H$ and no other critical points. In particular, $F$ is a Morse function on a cobordism $(\Omega', Y')$.

**Proof.** We shall prove the result for right half-handle attachment, the other case is completely analogous. In this case we have $B_0$ embedded into $M_1$ and we extend this embedding to an embedding of $B$ into $\Sigma_1$ (see Definition 2.13). We shall be using the notation from the proof of Theorem 2.24 with $\varepsilon^2 = 1$ and $\rho^2 = 4$. We have $\Omega' = \Omega \cup_B H$, where we identify $(H, B, C)$ with $(\tilde{H}, \tilde{B}, \tilde{C})$ using the diffeomorphism $\Psi$. Let now $\Sigma_\omega = \text{closure of } (\Sigma_1 \setminus B)$ and $T_\omega = \text{closure of } (\partial \Sigma_\omega \setminus \partial \Sigma_1)$.

By Lemma 2.29 used ‘back to front’ we deduce that $\Omega''$ is diffeomorphic to $\Omega'$, where

\[ \Omega'' = \Omega' \cup \Sigma_\omega \times [-1, 1]. \]
The gluing in (2.33) is as follows: we glue \( \Sigma_\times \times \{-1\} \) to \( \Sigma_1 \) and \( T_\times \times [-1,1] \) to \( \tilde{K} \) (in the proof of Lemma 2.29 we proved that \( \tilde{K} \) is \( T_\times \times [-1,1] \)). Now let us define a function

\[
F(x) = \begin{cases} 
  t & \text{if } x = (v,t) \in \Sigma_0 \times \{t\} \subset \Sigma_0 \times [0,1] = \Omega \\
  2 + t & \text{if } x = (v,t) \in \Sigma_- \times \{t\} \subset \Sigma_- \times [-1,1] \\
  2 - \sum_{j=1}^{k} x_j^2 + \sum_{j=k+1}^{n+1} x_j^2 & \text{if } x = (x_1, \ldots, x_{n+1}) \in H.
\end{cases}
\]

Checking continuity of \( F \) is straightforward. Moreover \( F \) is piecewise smooth and smooth away from \( B \). Let us choose a vector field \( \bar{e} \) on \( B \), normal to \( B \) and pointing towards \( H \). Then \( \langle \nabla F, \bar{e} \rangle \) is positive on both sides of \( B \). By a standard analytic argument \( F \) can be approximated by a smooth function equal to \( F \) away from some neighbourhood of \( B \), which has no new critical point. \( \square \)

2.5. **Left and right product cobordisms and traces of handle attachments.** In this subsection we relate half-handle attachments to handle attachments and detachments in the sense subsection 2.3. This also creates a dictionary between surgery theoretical notions (traces of handle attachments and detachments) and Morse theoretical (additions of half-handles). To begin with, let \((\Omega, Y)\) be a cobordism between \((\Sigma_0, M_0)\) and \((\Sigma_1, M_1)\).

**Definition 2.34.** We shall say that \( \Omega \) is a **left product cobordism** if \( \Omega \cong \Sigma_0 \times [0,1] \). Similarly, if \( \Omega \cong \Sigma_1 \times [0,1] \), then we shall say that \( \Omega \) is a **right product cobordism**.

**Proposition 2.35.** (a) If \((\Omega, Y)\) is a cobordism between \((\Sigma_0, M_0)\) and \((\Sigma_1, M_1)\) consisting only of left half-handle attachments, then it is a left-product cobordism. Conversely, if it consists only on right half-handle attachments, then it is a right product cobordism.

(b) Let \( F: \Omega \to [0,1] \) be a Morse function in the sense of Definition 1.4. Assume that \( F \) has no critical points in the interior of \( \Omega \). If all critical points on the boundary are boundary stable, then \( F \) is a left-product cobordism. If all critical points are boundary unstable, then \( F \) is a right product cobordism.

**Proof.** The two statements (a) and (b) are equivalent via Theorem 2.24 and Proposition 2.32. The stable-unstable (right-left) statements are also equivalent by replacing the Morse function \( F \) by \(-F\). The stable case follows from Lemma 2.17. \( \square \)

The next results of this subsection will be not used in this paper, but we insert them because they bridge surgery techniques and applications, e.g. with [Ra] or [BNR].

In order to clarify what we wish, let us recall that by Theorem 2.24 if a Morse function \( F \) defined on a cobordism \((\Omega, Y)\) has only one critical point of boundary type then \((\Omega, Y)\) is a half-handle attachment. Proposition 2.32 is the converse of this, the (total) space of a half-handle attachment can be thought as a cobordism with a Morse function on it with only one critical point.

We wish to establish the analogues of these statements ‘at the level of \( \Sigma \)’. In Subsection 2.3 we proved that the output of a right/left half-handle attachment at the level of \( \Sigma \) induces a handle attachment/detachment. The next lemma is the converse of this statement. (In fact, the output cobordism provided by it can be identified with the cobordism constructed in Proposition 2.32.)

**Lemma 2.36.** Assume that \((\Sigma_1, M_1)\) is a result of a handle attachment (respectively detachment) to \((\Sigma_0, M_0)\). Then, there exists a cobordism \((\Omega, Y; \Sigma_0, M_0, \Sigma_1, M_1)\) such that \( \Omega \cong \Sigma_1 \times [0,1] \) (respectively \( \Omega \cong \Sigma_0 \times [0,1] \)).
Lemma 2.36. On the left a 1-handle is attached to $\Sigma_0$. On the right there is a cobordism between $\Sigma_0$ and $\Sigma_1$, which is a right product cobordism.

Proof. Assume that $(\Sigma_1, M_1)$ arises from a handle attachment to $(\Sigma_0, M_0)$, i.e. $\Sigma_1 = \Sigma_0 \cup D^k \times D^{n-k}$. Let us define $\Omega = \Sigma_1 \times [0, 1]$. The boundary $\partial \Omega$ can be split as

$$\partial \Omega = \left( (\Sigma_0 \cup D^k \times D^{n-k}) \times \{0\} \cup (M_1 \times [0, 1]) \cup (\Sigma_1 \times \{1\}) \right)$$

where $Y = D^k \times D^{n-k} \cup (M_1 \times [0, 1])$. Its $D^k \times D^{n-k}$ part can be ‘pushed inside’ $\Omega$ transforming (diffeomorphically) $\Omega$ into a cobordism, see Figure 9.

Analogous construction can be used in the case of a handle detachment. If $(\Sigma'_1, M'_1)$ is a result of a handle detachment from $(\Sigma_0, M_0)$, then the trace of the handle detachment is the cobordism between $(\Sigma_0, M_0)$ and $(\Sigma'_1, M'_1)$ such that

$$(\Omega', Y') = (\Sigma_0 \times [0, 1], M_0 \times [0, 1] \cup D^k \times D^{n-k}).$$

□

Definition 2.37. The cobordism $(\Omega, Y; \Sigma_0, M_0, \Sigma_1, M_1)$ determined by the Lemma 2.36 is called the trace of a handle attachment of $(\Sigma_0, M_0)$ (respectively the trace of a handle detachment).

3. Splitting interior handles

We prove here the theorem about moving critical points to the boundary.

Theorem 3.1. Assume that on a cobordism $(\Omega, Y)$ between $(\Sigma_0, M_0)$ and $(\Sigma_1, M_1)$ we have a Morse function $F$ with a single critical point $z$ of index $k \in \{1, \ldots, n\}$ in the interior of $\Omega$ situated on the level set $\Sigma_{1/2} = F^{-1}(F(z))$. If

$$(3.2) \text{ the connected component of } \Sigma_{1/2} \text{ containing } z \text{ has non-empty intersection with } Y,$$

then there exists a function $G : \Omega \to [0, 1]$, such that

- $G$ agrees with $F$ in a neighbourhood of $\Sigma_0 \cup \Sigma_1$;
- $\nabla G$ is everywhere tangent to $Y$;
- $G$ has exactly two critical points $z^s$ and $z^u$, which are both on the boundary and of index $k$. The point $z^s$ is boundary stable and $z^u$ is boundary unstable.
- There exists a Riemannian metric such that there is a single trajectory of $\nabla G$ from $z^s$ to $z^u$ inside $Y$.

Remark 3.3. A careful reading of the proof shows that we can in fact construct a smooth homotopy $G_t$ such that $F = G_0$, $G = G_1$ and there exists $t_0 \in (0, 1)$ such that $G_t$ has
3.1. **About the proof.** The argument is based on the following two-dimensional picture. Consider the set
\[
\mathcal{Z} = \{(x, y) \in \mathbb{R}^2 : x \geq 0\}
\]
and the function \(D : \mathcal{Z} \to \mathbb{R}\) given by
\[
D(x, y) = y^3 - yx^2 + ay,
\]
where \(a \in \mathbb{R}\) is a parameter. Observe that the boundary of \(\mathcal{Z}\) given by \(\{x = 0\}\) is invariant under the gradient flow of \(D\) (see Figure 10). The proof of the following lemma is completely straightforward and will be omitted.

**Lemma 3.4.** For \(a > 0\), \(D\) has a single Morse critical point in the interior of \(\mathcal{Z}\). For \(a < 0\), \(D\) has two critical points on the boundary of \(\mathcal{Z}\).

For \(a = 0\), \(D\) acquires a \(D_4^-\) singularity at the origin (see e.g. [AGV]).

The proof of Theorem 3.1 starts with introducing 'local/global' coordinates \((x, y, u_1, \ldots, u_{n-1})\) at \(z\), in which \(F\) has the form \(D(x, y) \pm u_1^2 \pm \cdots \pm u_{n-1}^2\), hence it also parametrizes a neighbourhood of a path connecting \(z\) with a point of \(Y\). Then we change the parameter \(a\) (which we originally assume to be equal to 1) to \(-\delta\), where \(\delta\) is very small positive number (which corresponds to the move of the critical point to the boundary along the chosen path).

3.2. **Proof of Theorem 3.1 under additional assumption.** We first give the proof assuming the existence of such coordinate system as in 3.1 described explicitly in the next proposition (which is proved in Section 3.3).

**Proposition 3.5.** There exists \(\eta > 0, \eta \ll 1\) and an open 'half-disc' \(U \subset \Omega\), intersecting \(Y\) along a disk, and coordinates \(x, y, u_1, \ldots, u_{n-1}\) such that in these coordinates \(U\) is given by
\[
0 \leq x < 3 + \eta, \ |y| < \eta, \ \sum_{j=1}^{n-1} u_j^2 < \eta^2,
\]
$U \cap Y$ is given by $\{x = 0\}$, and in these coordinates $F$ is given by

$$y^3 - yx^2 + y + \frac{1}{2} + \sum_{j=1}^{n-1} \epsilon_j u_j^2,$$

where $\epsilon_1, \ldots, \epsilon_{n-1} \in \{\pm 1\}$ are choices of signs. In particular $\# \{j : \epsilon_j = -1\} = k - 1$, where $k = \text{ind}_z F$.

Assuming the proposition, we prove Theorem 3.1. Let us introduce some abbreviations.

(3.6) $\vec{u} = (u_1, \ldots, u_{n-1}), \quad \vec{u}^2 = \sum_{j=1}^{n-1} \epsilon_j u_j^2, \quad ||\vec{u}||^2 = \sum_{j=1}^{n-1} u_j^2.$

We fix a small real number $\varepsilon > 0$ such that $\varepsilon \ll \eta$ and two subsets $U_1 \subset U_2$ of $U$ by

$U_1 = \{|y| \leq \varepsilon, \ x \leq 3\} \cup \{(x-3)^2 + y^2 \leq \varepsilon^2\}$

$U_2 = \{|y| \leq 2\varepsilon, \ x \leq 3\} \cup \{(x-3)^2 + y^2 \leq 4\varepsilon^2\}$.

The difference $U_{21} := U_2 \setminus U_1$ splits into two subsets $S_1 \cup S_2$ (see Figure 11), where

$S_1 = U_{21} \cap \{x \leq 3\}$, $S_2 = U_{21} \cap \{x \geq 3\}$.

For a point $v = (x, y, u_1, \ldots, u_{n-1}) \in U$, let us define:

$$\bar{s}(v) = \begin{cases} 
1 & \text{if } v \in U_1, \\
0 & \text{if } v \in U \setminus U_2, \\
2 - \frac{|y|}{\varepsilon} & \text{if } v \in S_1, \\
2 - \frac{\sqrt{(x-3)^2 + y^2}}{\varepsilon} & \text{if } v \in S_2.
\end{cases}$$

The above formula defines a continuous function $\bar{s} : U_2 \to [0, 1]$. It is smooth away of $\partial S_1 \cup \partial S_2$. We can perturb it to a $C^\infty$ function $s$, with the following properties:

(S1) $s^{-1}(1) = U_1, \ s^{-1}(0) = \{|y| \geq 2\varepsilon - \varepsilon^2\} \cup \{(x-3)^2 + y^2 \geq 4\varepsilon^2 - \varepsilon^3\}$;

(S2) $\frac{\partial s}{\partial u_j} = 0$ for any $j = 1, \ldots, n - 1$;

(S3) $\frac{\partial s}{\partial x} = 0$, and $|\frac{\partial s}{\partial y}| < \frac{2}{\varepsilon}$ at all points of $S_1$. Furthermore $y \frac{\partial s}{\partial y} < 0$ at all points of $S_1$;

(S4) if $v \in S_2$ and we choose radial coordinates $x = 3 + r \cos \phi, \ y = r \sin \phi$ (where $r \in [\varepsilon, 2\varepsilon]$ and $\phi \in [-\frac{\pi}{2}, \frac{\pi}{2}]$), then $|\frac{\partial s}{\partial r}| < \frac{2}{\varepsilon}$ and $|\frac{\partial s}{\partial \phi}| < \varepsilon$.

Figure 11. Sets $U_1, U_{21}, S_1$ and $S_2$ in two dimensions (coordinates $x$ and $y$).
Observe that $\tilde{s}$ satisfies (S1)–(S4) at every point, where it is smooth, the only issue is that on $S_1 \cap S_2$, $\tilde{s}$ fails to be $C^2$.

Now let us choose smooth decreasing function $\phi: [0, \eta^2] \to [0, 1]$, which is equal to 0 on $[\frac{\eta^2}{4}, \eta^2]$ and $\phi(0) = 1$. We define now a new function $b: U_2 \to [0, 1]$ by the formula

$$(3.7) \quad b(x, y, \tilde{u}) = s(x, y, \tilde{u}) \cdot \phi(||\tilde{u}||^2).$$

Let us finally define the function $G: \Omega \to [0, 1]$ by

$$(3.8) \quad G(w) = \begin{cases} F(w) & \text{if } w \notin U_2 \\ y^3 - yx^2 + y - (\delta + 1)b(x, y, \tilde{u})y + \frac{1}{2} + \tilde{u}^2 & \text{if } w = (x, y, \tilde{u}) \in U_2, \end{cases}$$

where $\delta > 0$ is a very small number. Later we shall show that it is enough to take $\delta < \varepsilon^2/2$. In the following lemmas we shall prove that $G$ satisfies the conditions of Theorem 3.1.

**Lemma 3.9.** The function $G$ is smooth.

*Proof.* It is a routine checking and we leave it for the reader. $\square$

In the next two lemmas we show that $G$ has no critical points in $U_{21}$.

**Lemma 3.10.** $G$ has no critical points on $U_{21} \cap \{y = 0\}$.

*Proof.* If $(x, 0, u_1, \ldots, u_{n-1}) \in U_{21}$ then $x > 3$. Consider the derivative over $y$ of $G$:

$$\begin{equation} \label{eq:partialGbyy} \frac{\partial G}{\partial y} = 3y^2 - x^2 + 1 - (\delta + 1)b - (\delta + 1)\phi(u_1^2 + \cdots + u_{n-1}^2)\frac{\partial s}{\partial y}. \end{equation}$$

Taking $y = 0$ we get $-x^2 + 1 - (\delta + 1)b$. Since $b \in [0, 1]$ and $x > 3$, one gets $\frac{\partial G}{\partial y} < 0$. $\square$

**Lemma 3.12.** If $\delta < 3\varepsilon^2$, then $G$ has no critical points on $U_{21} \cap \{y \neq 0\}$.

*Proof.* Assume that $\frac{\partial G}{\partial x} = 0$ for some $(x, y, \tilde{u})$. Then

$$y \left(-2x - (\delta + 1)\frac{\partial s}{\partial x} \phi\right) = 0.$$

As $y \neq 0$, the expression in parentheses should be zero. If $0 < x \leq 3$, then by (S3) we have $\frac{\partial s}{\partial x} = 0$. Hence the above equality can not hold. Assume that $x = 0$. In the derivative over $y$ (see equation (3.11)), the expression $- (\delta + 1)\frac{\partial s}{\partial y} \cdot \phi$ is non-negative by (S3). Furthermore $b < 1$, hence

$$\frac{\partial G}{\partial y} \geq 3y^2 - \delta.$$

Now if $\delta < 3\varepsilon^2$ then there is no critical points with $x = 0$. It remains to deal with the case $(x, y, u_1, \ldots, u_{n-1}) \in S_2$. Consider the derivative $\frac{\partial G}{\partial y}$. By (S4) we have

$$\left| \frac{\partial s}{\partial y} \right| = \left| \frac{y^2}{r^3} \cdot \frac{\partial s}{\partial r} \right| + \left| \frac{(x - 3)y}{r^2} \cdot \frac{\partial s}{\partial \phi} \right| < \frac{2}{\varepsilon} + \varepsilon < 5.$$

Furthermore $|1 - (\delta + 1)b| \leq 1$, and $|3y^2| < 1$ because $\varepsilon$ is small. As $x \geq 3$, we have $\frac{\partial G}{\partial y} < 0$ on $S_2$. $\square$

On $U_1$ the function $G$ is given by

$$(3.13) \quad G(x, y, \tilde{u}) = y^3 - yx^2 - \delta y + \tilde{u}^2 + \frac{1}{2}.$$

As in Section 3.1 we study the critical points in $U_1$. 

**Proof.** It is a routine checking and we leave it for the reader. $\square$
Lemma 3.14. \( G \) has two critical points on \( U_1 \) at
\[
\begin{align*}
z^s &:= (0, \sqrt{\delta/3}, 0, \ldots, 0) \\
z^u &:= (0, -\sqrt{\delta/3}, 0, \ldots, 0).
\end{align*}
\]
Both critical points are boundary, both of Morse indices \( k \), \( z^s \) is stable, while \( z^u \) is unstable.

Proof. The derivative of \( G \) vanishes only at \( z^s \) and \( z^u \). Indices are immediately computed from (3.13). The point \( z^s \) is boundary stable, because for \( z^s \) the expression \(-yx^2\) is negative and the boundary is given by \( x = 0 \), hence it is attracting in the normal direction. Similarly we prove for \( z^u \). See also Figure 10 for the two-dimensional picture.

Remark 3.15. If we define \( G_t = y^3 - yx^2 + y - t(\delta + 1)b \cdot y + \frac{1}{2} + \bar{u}^2 \) for \( t \in [0, 1] \), then the same argument as in Lemmas 3.10 and 3.12 shows that \( G_t \) has no critical points in \( U_2 \setminus U_1 \). As for critical points in \( U_1 \), observe that on \( U \) can be changed near \( \gamma \) from (3.13). The point \( z \) is boundary stable, because \( z \) is in the unstable space. Hence \( z \) is attracting in the normal direction. Similarly we eliminate the others, we need the following lemma.

Let us now choose a Riemannian metric \( g' \) on
\[
U'_1 := U_1 \cap \{ ||\bar{u}|| < \varepsilon \}
\]
by the condition that \( (x, y, u_1, \ldots, u_{n-1}) \) be orthonormal coordinates. Clearly, any metric \( g \) on \( \Omega \) can be changed near \( U_1 \) so as to agree with \( g' \) on \( U'_1 \). In this metric the gradient of \( G \) is
\[
(-2xy, 3y^2 - x^2 - \delta, 2\varepsilon_1 u_1, \ldots, 2\varepsilon_{n-1} u_{n-1}).
\]
We want to show that there is a single trajectory starting from \( z^s \) and terminating at \( z^u \). Clearly, there is one trajectory from \( z^s \) to \( z^u \) which stays in \( U'_1 \) (having \( y = 0 \) and \( \bar{u} = 0 \)). In order to eliminate the others, we need the following lemma.

Lemma 3.16. Let \( \gamma \) be a trajectory of \( \nabla G \) starting from \( z^s \). Let \( w \) be the point, where \( \gamma \) hits \( \partial U'_1 \) for the first time. If \( \delta \) is sufficiently small, then \( G(w) > G(z^u) \).

Proof. Assume that \( \gamma(t) \) is such trajectory. Assume that among numbers \( \varepsilon_i \), we have \( \varepsilon_i = -1 \) for \( i \leq k - 1 \) and \( \varepsilon_i = 1 \) otherwise. As \( z^s \) is a critical point of the vector field \( \nabla G \) with a non-degenerate linear part, we conclude that the limit
\[
\lim_{t \to -\infty} \frac{\gamma'(t)}{||\gamma'(t)||} =: v = (x_0, y_0, u_{01}, \ldots, u_{0n-1})
\]
extists. The vector \( v \) is the tangent vector to the curve \( \gamma \) at the point \( z^s \), and it lies in the unstable space. Hence \( x_0 = 0 \) as \((1,0,\ldots,0)\) is a stable direction; similarly \( u_{01} = \cdots = u_{0,k-1} = 0 \). Therefore, until \( \gamma \) hits the boundary of \( U'_1 \) for the first time, we have
\[
x = u_1 = \cdots = u_{k-1} = 0.
\]
Set also \( g(y) = y^3 - \delta y \). One has the following cases, depending the position of \( w \), where \( \gamma \) hits \( \partial U'_1 \) for the first time: (a) \( y = -\varepsilon \), (b) \( y = \varepsilon \), or (c) \( ||\bar{u}||^2 = \eta^2 \). The case (a) cannot happen since \( G \) is increasing along the trajectory, hence \( G(w) > G(z^s) \), a fact which contradicts \( g(-\varepsilon) < g(\sqrt{\delta/3}) \) valid for \( 2\delta < \varepsilon^2 \). In case (b), \( G(w) > G(z^u) \) follows from \( g(\varepsilon) > g(-\sqrt{\delta/3}) \). Finally, assume the case (c). Then, as \( u_{01} = \cdots = u_{0,k-1} = 0 \), we obtain \( \bar{u}^2 = ||\bar{u}||^2 = \eta^2 \). Then \( G(w) - G(z^s) > \eta^2 \), because the contribution to \( G \) from \( y^3 - \delta y \) increases along \( \gamma \). Hence \( G(w) > G(z^u) \) follows again since \( \varepsilon \ll \eta \).
Figure 12. Sets $Z_1$ and $Z_2$ from Section 3.3. There are drawn also the stable and unstable manifolds of $\nabla A$.

Given the above lemma it is clear that if a trajectory $\gamma$ leaves $U_1'$, then $G$ becomes bigger than $G(z^n)$. As $G$ increases along any trajectory, it is impossible that such trajectory limits in $z^n$. The proof of Theorem 3.11 up to Proposition 3.5 is accomplished.

3.3. An auxiliary construction. We provide now a construction, a crucial ingredient in the proof of Proposition 3.5, see next Section. Set

$$Z = \{(x, y) \in \mathbb{R}^2 : x \geq 0\},$$

and define the two functions

$$A(x, y) = \frac{x^3}{3\sqrt{3}} - \sqrt{3}xy^2 - \frac{x}{\sqrt{3}} + \frac{2}{3\sqrt{3}}, \quad B(x, y) = y^3 - yx^2 + y.$$  

(3.17) They are the real/imaginary parts of a complex function; indeed (with $i = \sqrt{-1}$)

$$A + iB = \left(\frac{x}{\sqrt{3}} - iy\right)^3 - \left(\frac{x}{\sqrt{3}} - iy\right) + \frac{2}{3\sqrt{3}}.$$

It follows from the Cauchy–Riemann equations (or straightforward computations), that the level sets of $A$ and $B$ are orthogonal, away from the point $(1, 0)$. If we choose a Riemannian metric such that $(x, y)$ are orthonormal coordinates, then $A$ is constant on trajectories of $\nabla B$ (i.e. $A$ is a first integral of $\nabla B$). Let us fix $\delta > 0$ smaller than $\frac{2}{3\sqrt{3}}$. Consider two sets

$$Z_1 = \{(x, y) \in Z, x < 1, A(x, y) \geq \delta\}, \quad Z_2 = \{(x, y) \in Z, x > 1, A(x, y) \geq \delta\}.$$  

We have the following result.

Lemma 3.19. The map $\psi(x, y) = (A(x, y), B(x, y))$ maps $Z_1$ and $Z_2$ diffeomorphically onto $E_1$ and $E_2$ respectively, where

$$E_1 = \{ (a, b) \in \mathbb{R}^2 : a \in \left[ \delta, \frac{2}{3\sqrt{3}} \right] \}, \quad E_2 = \{ (a, b) \in \mathbb{R}^2 : a \geq \delta \}.$$  

Moreover, $\psi$ maps trajectories of $\nabla B$ onto vertical lines.

Proof. This follows from the covering theory of one-variable complex functions. Alternatively, one readily checks that $\psi: Z_1 \to V_1$ and $\psi: Z_2 \to V_2$ are bijections. As $D\psi$ is non-degenerate on $Z_1 \cup Z_2$, $\psi$ is a diffeomorphism between the two pairs of sets. Now $A$ is a first integral of the flow $\nabla B$, so each trajectory of $\nabla B$ lies in the set $A^{-1}(c)$.  

3.4. Proof of Proposition 3.5. First, as \( z \) is a critical point of index \( k \in \{1, \ldots, n\} \), by Morse lemma we can find a neighbourhood \( \tilde{V} \) of \( z \) and a chart \( h_1 : \tilde{V} \to \mathbb{R}^{n+1} \), with coordinates \((x', y, \vec{u})\) such that
\[
F \circ h_1^{-1}(x', y, \vec{u}) = x'y + \vec{u}^2 + \frac{1}{2}.
\]
Let us define a map \( h_2(x, y, \vec{u}) = (x', y, \vec{u}) \), where \( x' = y^2 + 1 - x^2 \). By the inverse function theorem, \( h_2 \) is a local diffeomorphism near \((1, 0, \ldots, 0)\). Shrinking \( \tilde{V} \) if needed, and considering \( h_3 = h_2^{-1} \circ h_1 \), we obtain \( h_3(z) = (1, 0, \ldots, 0) \) and
\[
F \circ h_3^{-1}(x, y, \vec{u}) = y^3 - yx^2 + y + \vec{u}^2 + \frac{1}{2} = B(x, y) + \vec{w}^2 + \frac{1}{2}.
\]
Let us pick now \( \xi > 0 \) such that the cylinder
\[
V = \{ |x - 1| < \xi, \ |y| < \xi, \ ||\vec{u}|| < \xi \} \subset \mathbb{R}^{n+1}
\]
lies entirely in \( h_3(\tilde{V}) \). By shrinking \( \tilde{V} \) we may in fact assume that \( h_3(\tilde{V}) = V \). If \( 0 < \delta \ll 1 \) is sufficiently small then \( A(x, 0) < \delta \) implies \( |x - 1| < \xi \). Choose such a \( \delta \), and set
\[
V_1 := V \cap \{ x < 1, \ A(x, y) \geq \delta \}.
\]
(compare (3.18)). By Lemma 3.19 the map

\[(3.20)\quad \Psi_1(x, y, \bar{u}) = (A(x, y), B(x, y) + \bar{u}^2, \bar{u}),\]

is a diffeomorphism (being the composition of \(\psi\) and a ‘triangular’ map). Set \(C_1 := \Psi_1(V_1)\) and \(\bar{V}_1 := h_3^{-1}(V_1)\). Finally, let

\[h = \Psi_1 \circ h_3.\]

Using Lemma 3.19 again we obtain that

\[F \circ h^{-1}(a, b, \bar{u}) = b + \frac{1}{2}.\]

Let \(\theta > \delta\) be sufficiently closed to \(\delta\) satisfying the inclusion

\[D_1 := [\delta, \theta] \times (-\theta, \theta) \times (-\theta, \theta)^{n-1} \subset C_1.\]

Let \(\tilde{D}_1 = h^{-1}(D_1) \subset \bar{V}_1\), see Figure 13.

**Lemma 3.21.** If \(\theta\) and \(\delta\) are small enough, there is an open/closed ball \(\tilde{W}\) in \(\Omega\), containing \(\tilde{D}_1\), such that \(h\) extends to a diffeomorphism between \(\tilde{W}\) and \([\delta, \frac{2}{3\sqrt{3}}] \times (-\theta, \theta) \times (-\theta, \theta)^{n-1}\)

with \(F \circ h^{-1}(a, b, \bar{u}) = b + \frac{1}{2},\) sending points with \(a = \frac{2}{3\sqrt{3}}\) to \(Y\).

In the proof we shall use the following result.

**Lemma 3.22.** There exists a smooth curve \(\gamma: [\delta, \frac{2}{3\sqrt{3}}] \rightarrow \Omega,\) such that \(\gamma(\frac{2}{3\sqrt{3}}) \in Y,\) \(\gamma(t) \in \Sigma_{1/2},\) \(\gamma(t) \in \tilde{D}_1\) if and only if \(t \in [\delta, \theta]\) and \(h(\gamma(t)) = (t, 0, \ldots, 0)\) and \(\gamma\) omits \(\bar{V} \setminus \bar{V}_1\).

Furthermore, \(\gamma\) is transverse to \(Y\).

**Proof of Lemma 3.22.** Let \(p = h^{-1}(\theta, 0, \ldots, 0) \in \Sigma_{1/2}\). Let \(B \subset \Sigma_{1/2}\) be an open ball with centre \(z\) and \(p \in \partial B\). Let \(\Sigma'\) be the connected component of \(\Sigma_{1/2}\) containing \(p\). We consider two cases.

If \(\Sigma' \setminus B\) is connected, it is also path connected. By (3.2), there exists a path \(\bar{\gamma} \subset \Sigma' \setminus B\) joining \(p\) with a point on the boundary. We can assume that \(\bar{\gamma}\) is transverse to \(\gamma\). We choose \(\gamma = h^{-1}([\delta, \theta] \times \{0, \ldots, 0\}) \cup \bar{\gamma}\) (and we smooth a possible corner at \(p\)). It is clear that \(\gamma\) omits \(\bar{V} \setminus \bar{V}_1\) and that we can find a parametrization of \(\gamma\) by the interval \([\delta, \frac{2}{3\sqrt{3}}]\).

If \(\Sigma' \setminus B\) is not connected, then as \(\Sigma'\) is connected, by a homological argument we have \(n = 1\) and \(k = 1\). Since \(\Sigma'\) is connected and has boundary, then \(\Sigma'\) is an interval and \(B\) is an interval too. Then \(\Sigma' \setminus B\) consists of two intervals, each intersecting \(Y\). One of these intervals contains \(p\). So \(p\) is connected with \(Y\) by an interval, which omits \(B\). We conclude the proof by the same argument as in the above case, when \(\Sigma' \setminus B\) was connected. \(\square\)

**Proof of Lemma 3.21.** Given Lemma 3.22 let us choose a tubular neighbourhood \(X\) of \(\gamma\) in \(F^{-1}(1/2) \setminus (\bar{V} \setminus \bar{V}_1)\). Shrinking \(X\) if needed we can assume that it is a disk and \(X_1 := X \cap \bar{V} = \tilde{D}_1 \cap F^{-1}(1/2)\). Now let \(\tau\) be a vector field on \(\tilde{D}_1\) given by \((Dh)^{-1}(1, 0, \ldots, 0)\), where \(Dh\) denotes the derivative of \(h\). This vector field is everywhere tangent to \(X_1\) and

\[(3.23)\quad \tau|_{\gamma \cap \tilde{D}_1} = \frac{d}{dt}\gamma(t)\]

by the very definition of \(\gamma\). We extend \(\tau\) to a smooth vector field on the whole \(X\), such that (3.23) holds on the whole \(\gamma\). For any point \(z \in \gamma\), the trajectory of \(\tau\) (which is \(\gamma\)) hits eventually \(Y\) and, on the other end, it hits the ‘right wall’

\[\bar{R} = h^{-1}([\delta] \times \{0\} \times (-\theta, \theta)^{n-1}).\]

(compare Figure 14 note that the horizontal coordinate there increases from right to left for consistency with Figure 13). Since \(\gamma\) is transverse to \(\bar{R}\) and to \(Y\), by implicit function
Figure 14. Proof of Lemma 3.21. Construction of the vector field $\tau$. Picture on $F^{-1}(1/2)$. The parallel vector field from the region on the right is extended to the whole $X$ so that it is tangent to $\gamma$.

Theorem trajectories close to $\tau$ also start at $\tilde{R}$ and end up at $Y$. Shrinking $X$ if necessary we may assume that each point of $X$ lies on the trajectory of $\tau$ which connects a point of $\tilde{R}$ to some point of $Y$, and all the trajectories are transverse to both $Y$ and $\tilde{R}$.

We can now rescale $\tau$ (that is multiply by a suitable smooth function constant on trajectories) so that all the trajectories go from $\tilde{R}$ to $Y$ in time $\frac{2}{3\sqrt{3}} - \delta$, i.e. the same time as $\gamma$ does. The rescaled vector field allows us to introduce coordinates on $X$ in the following way. For any point $z \in X$, let $\gamma_z$ be the trajectory of $\tau$, going through $z$. We can assume that $\gamma_z(\delta) \in \tilde{R}$. Let $t_z = \gamma_z^{-1}(z)$, i.e. the moment when $\gamma_z$ passes through $z$. Since we normalized $\gamma_z$, we know that $t_z \in [\delta, \frac{2}{3\sqrt{3}}]$ and $t_z = \frac{2}{3\sqrt{3}}$ if and only if $z \in Y \cap X$.

Let $\vec{u}_z$ be such that $h(\gamma_z(\delta)) = (\delta, 0, \vec{u}_z)$. The vector $\vec{u}_z$ might be thought off as a coordinate on $\tilde{R}$. We define now $h(z) = (t_z, 0, \vec{u}_z)$.

This maps clearly extends $h$ to the whole $X$.

Now let $\tilde{W}$ be a tubular neighbourhood of $X$ in $\Omega \setminus (\tilde{V} \setminus \tilde{V}_1)$. We use the flow of $\nabla F$ to extend coordinates from $X$ to $\tilde{W}$. More precisely, shrinking $\tilde{W}$ if needed we may assume that for each $w \in \tilde{W}$ the trajectory of $\nabla F$ intersects $X$. This intersection is necessarily transverse and it is in one point, which we denote by $z_w \in X$. We define now $h(w) = (t_{z_w}, F(w) - F(z_w), \vec{u}_{z_w})$.

As $h$ is a local diffeomorphism on $X$ (because $\nabla F$ is transverse to $X$), it is also a local diffeomorphism near $X$. We put $W = h(\tilde{W})$. Clearly both definitions of $h$ on $\tilde{V}$ and $\tilde{W}$ agree. We may now decrease $\theta$ and shrink $W$ so that

$$W = [\delta, \frac{2}{3\sqrt{3}}] \times (-\theta, \theta) \times (-\theta, \theta)^{n-1}.$$ 

We have $F \circ h^{-1}(a, b, \vec{u}) = b + \frac{1}{2}$. We now extend $h_3$ over $\tilde{W}$ by the formula $h_3 = \Psi^{-1} \circ h$. □

Consider now

$$V_2 := V \cap \{x > 1, \ A(x, y) \geq \delta\}.$$ 

Let $\Psi_2 : V \to \mathbb{R}^{n+1}$ be given by $\Psi_2(x, y, \vec{u}) = (a, b, \vec{u}) = (A(x, y), B(x, y) + \vec{u}^2, \vec{u})$, provided by the same formula as $\Psi_1$ in (3.20) but the image now satisfies $a \geq \delta$, cf. Lemma 3.19.
Let \( C_2 = \Psi_2(V_2) \), and let us choose \( \theta' \) sufficiently small such that
\[
D_2 := [\delta, \theta'] \times (-\theta', \theta') \times (-\theta', \theta')^{n-1} \subset C_2.
\]
We shall denote \( h = \Psi_2 \circ h_3 \) and \( \tilde{D}_2 = h^{-1}(D_2) \).

Let us now fix \( M > 0 \) large enough and consider a map \( h_4: \mathbb{R}^{n+1} \to \mathbb{R}^{n+1} \) of the form
\[
h_4(a, b, \vec{u}) = (\phi(a), b, \vec{u}),
\]
where \( \phi: [\delta, \theta'] \cong [\delta, M] \) is a strictly increasing smooth function, which is an identity near \( \delta \).

Consider the map \( h'_3: \Psi_2^{-1} \circ h_4 \circ h_3: \tilde{D}_2 \to \mathbb{R}^{n+1} \). Since \( h \) is an identity for \( a \) close to \( \delta \), this map agrees with \( h_3 \) for \( a \) close to \( \delta \). Furthermore \( F \circ h^{-1}(a, b, \vec{u}) = F \circ h^{-1} \circ h_3^{-1}(a, b, \vec{u}) = b + \frac{1}{2} \) by a straightforward computation. On the other hand, the point \( h^{-1}(\theta', 0, \ldots, 0) \in \tilde{D}_2 \) is mapped by \( h'_3 \) to \((M, 0, \ldots, 0) \in \mathbb{R}^{n+1} \), where \( M \) can be arbitrary large, e.g. \( M > 3 \).

Having gathered all the necessary maps, we now conclude the proof. Let
\[
\tilde{U} = \tilde{W} \cup (\tilde{V} \setminus h_3^{-1}(V_1 \cup V_2)) \cup \tilde{D}_2.
\]
The map \( h_3: \tilde{U} \to [0, \infty) \times \mathbb{R}^n \) is given by \( h_3 \) on \( \tilde{W} \) and on \( \tilde{V} \setminus h_3^{-1}(V_2) \), and by \( h'_3 \) on \( \tilde{D}_2 \). This map is clearly a diffeomorphism onto its image, so it is a chart near \( z \).

By construction \( F \circ h_3^{-1} \) is equal to \( y^3 - yx^2 + y + u^2 + 1/2 \) and \( h_3(\tilde{W}) \) contains the segment with endpoints \((0, 0, \ldots, 0) \) and \((3, 0, \ldots, 0) \). Since it is an open subset, it contains \([0, 3 + \eta) \times (-\eta, \eta) \times (-\eta, \eta)^{n-1} \) for \( \eta > 0 \) small enough. The inverse image of this cube gives the required chart.

This ends the proof of Theorem 3.1 which moves a single interior critical point to the boundary. Section 4 generalizes this fact for more critical points; one of the needed tools will be the rearrangements of the critical values/points.

3.5. Condition (3.2) revisited. We will provide two characterizations of Condition (3.2). One is valid for arbitrary \( n \geq 1 \), the other one holds only in the case \( n = 1 \). We shall keep the notation from previous subsections, in particular \((\Omega, Y)\) is a cobordism between \((\Sigma_0, M_0)\) and \((\Sigma_1, M_1)\), \( F: \Omega \to [0, 1] \) is a Morse function with a single critical point \( z \in \Omega \) in the interior of \( \Omega \), and \( F(z) = 1/2 \). Let \( \Sigma_{1/2} = F^{-1}(1/2) \) and \( \Sigma' \) be the connected component of \( \Sigma_{1/2} \) such that \( z \in \Sigma' \).

**Proposition 3.24.** If \( \Sigma_0, \Sigma_1 \) and \( \Omega \) have no closed connected components, then \( \Sigma' \cap Y \neq \emptyset \).

In particular, in Theorem 3.7 we can assume that \( \Sigma_0, \Sigma_1 \) and \( \Omega \) have no closed connected components instead of (3.2).

**Proof.** Let \( p = h^{-1}(\theta, 0, \ldots, 0) \in \tilde{D}_1 \subset \Omega \) and let \( B \) be an open ball in \( \Sigma' \) near \( z \), such that \( p \in \partial B \). It is enough to show that \( p \) can be connected to \( Y \) by a path in \( \Sigma_{1/2} \), which omits \( B \) (compare Lemma 3.22).

Let us choose a Riemannian metric on \( \Omega \). Let \( W^s_z \) be the stable manifold of \( z \) and let \( T \) be intersection of \( W^s_z \) and \( \Sigma_0 \). This is a \((k-1)\)-dimensional sphere. The flow of \( \nabla F \) induces a diffeomorphism \( \Phi: \Sigma_{1/2} \setminus B \cong \Sigma_0 \setminus B_0 \), where \( B_0 \) is a tubular neighbourhood of \( T \) in \( \Sigma_0 \) (here we tacitly use the fact that \( \delta \) and \( \theta \) are small enough), see Figure 15. Let \( p_0 = \Phi(p) \). Let \( \Sigma'_0 \) be the connected component of \( \Sigma_0 \), which contains \( B_0 \).

Now we will analyze several cases. Recall that \( k = \ind_z F \in \{1, \ldots, n\} \). First we assume that \( k < n \). Then \( \Sigma'_0 \setminus T \) is connected, so \( p_0 \) can be connected to the boundary of \( \Sigma'_0 \) — which is non-empty by assumptions of the proposition — by a path \( \gamma_0 \). Now the inverse image \( \Phi^{-1}(\gamma_0) \) is the required path.

If \( k = n > 1 \) then we reverse the cobordism and look at \(-F\), hence this case is covered by the previous one (since \( k = n \) will be replaced by \( k = 1 < n \)).
Finally, it remains to deal with the situation $k = n = 1$. Then $\dim \Sigma_0 = 1$. $T$ consists of two points. Assume first that they lie in a single connected component $\Sigma_0'$ of $\Sigma_0$. We shall show that this is impossible. As $\Sigma_0'$ is connected with non-trivial boundary, it is an interval. The situation is like on Figure 16. Now as $F$ has precisely one Morse critical point of index 1, $\Sigma_1$ is the result of a surgery on $\Sigma_0$. This surgery consists of removing two inner segments from $\Sigma_0$ and gluing back two other segments, which on Figure 16 are drawn as dashed arc. But then $\Sigma_1$ has a closed connected component, which contradicts assumptions of Theorem 3.1.

Therefore, $T$ lies in two connected components of $\Sigma_0$. The situation is drawn of Figure 17, and it is straightforward to see that $p_0$ can be connected to $M_0$ by a segment omitting $B_0$. \hfill $\square$

The proof of Proposition 3.24 suggests that the case $n = 1$ is different than case $n > 1$. We shall provide now a full characterization of a failure to (3.2).
Proposition 3.25. Assume that $k = n = 1$ and $\Omega$ is connected. If (3.2) does not hold, then $\Omega$ is a pair of pants, $\Sigma_0$ is a circle and $\Sigma_1$ is a disjoint union of two circles; or converse: $\Sigma_1 = S^1$ and $\Sigma_0$ is a disjoint union of two circles. In particular, $Y = \emptyset$.

Proof. A one-handle attached to a surface changes the number of boundary components by ±1. Let us assume that $\Sigma_1$ has less components than $\Sigma_0$, if not we can reverse the cobordism. As $\Omega$ is connected, $\Sigma_0$ has two components and $\Sigma_1$ only one. Let $A_0 \subset \Sigma_0$ be the attaching region, i.e. the union of two closed intervals to which the one-handle is attached. With the notation of Section 3.5, we have $Y = (\Sigma_0/A_0, A_0/A_0)$, where the quotient denotes collapsing a space to a point. In particular $z$ can not be joined to $Y$ by a path in $\Sigma'$ if and only if $\Sigma_0$ is disjoint from $Y$. Hence $\Sigma_0$ is closed, that is, it is a union of two circles. □

4. Rearrangements of boundary handles

4.1. Preliminaries. Let $(\Omega, Y)$ be a cobordism between two $n$-dimensional manifolds with boundary $(\Sigma_0, M_0)$ and $(\Sigma_1, M_1)$. Let $F$ be a Morse function, with critical points $w_1, \ldots, w_k \in \text{Int} \Omega$ and $y_1, \ldots, y_l \in Y$. In the classical theory (that is, when $Y = \emptyset$), the Thom–Milnor–Smale theorem (see [Mi2, Section 4]) says that we can alter $F$ without introducing new critical points such that if $\text{ind } w_i < \text{ind } w_j$, then $F(w_i) < F(w_j)$ as well. We want to prove similar results in our more general case.

In this section we rely very strongly on [Mi2, Section 4].

4.2. Elementary rearrangement theorems. We shall begin with the case $k + l = 2$, i.e. $F$ has two critical points. For a critical point $p$ we shall denote by $K_p$ the union $W_p^u \cup \{p\} \cup W_p^s$, i.e. the set of all points $x \in \Omega$, such that the trajectory $\phi_t(x)$ ($t \in \mathbb{R}$), of the gradient vector field $\nabla F$ has $p$ as its limit set. Elementary rearrangement theorems deal with the case when the two sets $K_{p_1}$ and $K_{p_2}$ for the two critical points are disjoint.

Proposition 4.1 (Rearrangement of critical points). Let $p_1$ and $p_2$ be two critical points, and assume that $K_1 := K_{p_1}$ and $K_2 := K_{p_2}$ are disjoint. Let us choose $a_1, a_2 \in (0, 1)$. Then, there exist a Morse function $G: \Omega \to [0, 1]$, with critical points exactly at $p_1$ and $p_2$, such that $G(p_i) = a_i$, $i = 1, 2$; furthermore, near $p_1$ and $p_2$, the difference $F - G$ is a constant function.

Remark 4.2. If both critical points are on the boundary, in order to guarantee the above existence, we need even to change the Riemannian metric away from $K_1$ and $K_2$.

Proof. Similarly as in [Mi2, Section 4] we first prove an auxiliary result.

Lemma 4.3. There exists a smooth function $\mu: \Omega \to [0, 1]$ with the following properties:

(M1) $\mu \equiv 0$ in a neighbourhood of $K_1$;
(M2) $\mu \equiv 1$ in a neighbourhood of $K_2$;
(M3) $\mu$ is constant on trajectories of $\nabla F$.

Furthermore, if at least one of the critical points is interior, we have

(M4) $\mu$ is constant on $Y$.

Given the lemma, we choose also a smooth function $\Psi: [0, 1] \times [0, 1] \to [0, 1]$ with

(PS1) $\frac{\partial \Psi}{\partial y}(x, y) > 0$ for all $(x, y) \in [0, 1] \times [0, 1]$;
(PS2) there exists $\delta > 0$, such that $\Psi(x, y) = x$ for all $x \in [0, \delta] \cup [1 - \delta, 1]$ and $y \in [0, 1]$;
(PS3) for any $s \in (-\delta, \delta)$ we have $\Psi(F(p_1) + s, 0) = a_1 + s$ and $\Psi(F(p_2) + s, 1) = a_2 + s$. 


For any $\eta \in \Omega$ we define $G(\eta) = \Psi(F(\eta), \mu(\eta))$. From the properties (PS3), (M1) and (M2) we see that near $p_i$, $G$ differs from $F$ by a constant. The property (PS2) ensures that $G$ agrees with $F$ in a neighbourhood of $\Sigma_0$ and $\Sigma_1$. Let us show that $\nabla G$ does not vanish away from $p_i$. By a chain rule we have

\begin{equation}
\nabla G = \frac{\partial \Psi}{\partial x} \nabla F + \frac{\partial \Psi}{\partial y} \nabla \mu.
\end{equation}

Since $\mu$ is constant on all trajectories of $\nabla F$, the scalar product $\langle \nabla F, \nabla \mu \rangle = 0$. Then the property (PS1) guarantees that $\langle \nabla G, \nabla F \rangle > 0$ away from $p_1$ and $p_2$.

We need to show that $\nabla G$ is everywhere tangent to $Y$. If one of the points is interior, by (M4) $\nabla \mu$ vanishes on $Y$, hence $\nabla G$ is parallel on $Y$ to $\nabla F$ and we are done. Next assume that both critical points are on the boundary. Let us choose an open subset $U$ of $Y$ such that $\nabla \mu|_U = 0$ and $K_1 \cup K_2 \subset U$. This is possible, because of the properties (M1) and (M2). Then let us choose a neighbourhood $W$ in $Y$ of $Y \cup U$, disjoint from $K_1$ and $K_2$. Observe that $dG(\nabla F) = \langle \nabla G, \nabla F \rangle > 0$. As $\nabla F \in TY$ one has $\ker dG \not\subset TY$, so by Lemma 1.7 we can change the metric in $W$ so that $\nabla G$ is everywhere tangent to $Y$.

Proof of Lemma 4.5. Let us define $T_1 = K_1 \cap \Sigma_0$ and $T_2 = K_2 \cap \Sigma_0$. Assume that $T_1$ and $T_2$ are not empty. For each $\eta \in \Omega \setminus (K_1 \cup K_2)$, let $\pi(\eta)$ be the intersection of the trajectory of $\eta$ under $\nabla F$ with $\Sigma_0$. This gives a map $\pi: \Omega \setminus (K_1 \cup K_2) \to \Sigma_0 \setminus (T_1 \cup T_2)$.

Let us define $\mu$ first on $\Sigma_0$ by the following conditions: $\mu \equiv 1$ in a neighbourhood of $T_2$, $\mu \equiv 0$ in a neighbourhood of $T_1$. Furthermore, if either $T_1$ or $T_2$ is disjoint from the boundary $M_0$ we extend $\mu$ to a constant function on $M_0$. Finally, we extend $\mu$ to the whole $\Omega$ by picking $\mu(\eta) = \mu(\pi(\eta))$ if $\eta \not\in K_1 \cup K_2$, and $\mu(\pi(\eta)) = i - 1$, $i = 1, 2$.

If $T_1 = \emptyset$, then $\ind_F p_1 = 0$ and the proof of the rearrangement theorem is completely straightforward.

4.3. Morse–Smale condition on manifolds with boundary. In the classical theory, the Morse–Smale condition imposed on a Morse function $F: M \to \mathbb{R}$ means that for each pair of two critical points $p_1, p_2$ of $M$ the intersection of stable manifold $W^s_{p_1}$ with the unstable manifold of $W^u_{p_2}$ is transverse. (Note that this Morse–Smale condition also depends on the choice of Riemannian metric on $M$.) Following [KM, Definition 2.4.2], we reformulate the Morse–Smale condition in the following way

Definition 4.5. The function $F$ is called Morse–Smale if for any two critical points $p_1$ and $p_2$, the intersection of $\text{Int} \, \Omega \cap W^s_{p_1}$ with $\text{Int} \, \Omega \cap W^u_{p_2}$ is transverse (as the intersection in the $(n+1)$-dimensional manifold $\Omega$) and the intersection of $Y \cap W^s_{p_1}$ with $Y \cap W^u_{p_2}$ is transverse (as an intersection in the $n$-dimensional manifold $Y$).

It is clear, that regular functions form an open-dense subset of all $C^2$ functions, which satisfy the Morse condition from Definition 1.4.

Assume now that $F$ is Morse–Smale. Given two critical points of $F$, $p_1$ and $p_2$, we want to check whether $W^s_{p_1} \cap W^u_{p_2} = \emptyset$. This depends not only on the index, but also on whether any of the two points is boundary stable. We show this in a tabulated form in Table 1, where $\ind p_1 = k$ and $\ind p_2 = l$. In studying the intersection, we remark that $W^s_{p_1} \cap W^u_{p_2}$ is formed from trajectories, so if for dimensional reasons we have $\dim W^s_{p_1} \cap W^u_{p_2} < 1$, it immediately follows that this intersection is empty.

4.4. Global rearrangement theorem. Let us combine the rearrangement theorems from Section 4.2 with the computations in Table 1.

Proposition 4.6. Let $F$ be a Morse function on a cobordism $(\Omega, Y)$ between $(\Sigma_0, M_0)$ and $(\Sigma_1, M_1)$. Let $w_1, \ldots, w_m$ be the interior critical points of $F$ and $y_1, \ldots, y_{k+l}$ be the boundary
critical points, where \(y_1, \ldots, y_k\) are boundary stable and \(y_{k+1}, \ldots, y_{k+l}\) boundary unstable. Let us choose real numbers \(0 < c_0 < c_1 < \cdots < c_{n+1} < 1, 0 < c_1^s < \cdots < c_{n+1}^s < 1, 0 < c_0^u < \cdots < c_n^u < 1\) satisfying
\[
\begin{align*}
c_i^s - 1 < c_i < c_{i+1}^s \\
c_i^u - 1 < c_i < c_{i+1}^u \\
c_i^u - 1 < c_i^s < c_i^u
\end{align*}
\]
for all \(i \in \{0, \ldots, n+1\}\) (we can assume that \(c_{-1} = c_0^s = c_{-1}^u = 0, c_{n+2} = c_{n+2}^s = c_{n+1}^u = 1\) so that (4.7) makes sense for all \(i\)).

Then, there exists another Morse function \(G\) on the cobordism \((\Omega, Y)\) with critical points \(w_1, \ldots, w_m\) in the interior, \(y_1, \ldots, y_{k+l}\) on the boundary, such that if \(\text{ind } w_j = l\), then \(G(w_j) = c_l\), if \(\text{ind } y_i = l\) and \(y_i\) is boundary stable then \(G(y_i) = c_i^s\), and if \(\text{ind } y_i = l\) and \(y_i\) is boundary unstable, then \(G(y_i) = c_i^u\).

Proof. Given the elementary rearrangement result (Proposition 4.1), the proof is completely standard (see the proof of Theorem 4.8 in \([Mi2]\)). Note only that we need to have \(c_i^s < c_i^u\) in the statement, because there might be a trajectory from a boundary stable critical point to a boundary unstable of the same index. However, we are free to choose \(c_i < c_i^s\) or \(c_i \in (c_i^s, c_i^u)\) or \(c_i > c_i^u\). \(\square\)

4.5. Moving more handles to the boundary at once. Before we formulate Theorem 4.10, let us introduce the following technical notion.

**Definition 4.8.** The Morse function \(F\) on the cobordism \((\Omega, Y)\) is called technically good if it has the following properties.

(TG1) If \(p_1, p_2\) are (interior or boundary) critical points of \(F\) then \(\text{ind } p_1 < \text{ind } p_2\) implies \(F(p_1) < F(p_2)\);

(TG2) There exist regular values of \(F\), say \(c, d \in [0, 1]\), with \(c < d\) such that \(F^{-1}[0, c]\) contains those and only those critical points, which have index 0 or which are boundary stable critical points of index 1, and \(F^{-1}[d, 1]\) contains those and only those critical points, which have index \(n+1\) and boundary unstable critical points of index \(n\).

(TG3) There are no pairs of 0 and 1 (interior) handles of \(F\) that can be cancelled (in the sense of Section 5);

(TG4) There are no pairs of \(n\) and \(n+1\) (interior) handles of \(F\) that can be cancelled.
Lemma 4.9. Each function $F$ can be made technically good without introducing new critical points.

Proof. By Proposition 4.6 we can rearrange the critical points of $F$, proving (TG1) and (TG2). The properties (TG3) and (TG4) can be guaranteed, using the handle cancellation theorem (e.g. [Mi2, Theorem 5.4]. We refer to the beginning of Section 5 for an explanation, that one can use the handle cancellation theorem if the manifold in question has boundary. □

Theorem 4.10. Let $(\Omega, Y)$ be a cobordism between $(\Sigma_0, M_0)$ and $(\Sigma_1, M_1)$. Let $F$ be a technically good Morse function on that cobordism, which has critical points $y_1, \ldots, y_k$ on the boundary $Y$, $z_1, \ldots, z_{l+m}$ in the interior $\text{Int} \ \Omega$, of which $z_{m+1}, \ldots, z_{l+m}$ have index 0 or $n+1$ and the indices of $z_1, \ldots, z_m$ are in $\{1, \ldots, n\}$. Suppose furthermore the following properties are satisfied:

(I1) $\Sigma_0$ and $\Sigma_1$ have no closed connected components;

(I2) $\Omega$ has no closed connected component.

Then, there exist a Morse function $G: \Omega \to [0, 1]$, on the cobordism $(\Omega, Y)$, with critical points $y_1, \ldots, y_k \in Y$, $z_{m+1}, \ldots, z_{l+m}$ and $z_1^s, z_2^s, \ldots, z_m^s, z_m^u$ such that:

- $\text{ind}_G y_i = \text{ind}_F y_i$ for $i = 1, \ldots, k$ and for $j = m + 1, \ldots, m + l$ we have $\text{ind}_G z_j = \text{ind}_F z_j$;
- for $j = 1, \ldots, m$, $\text{ind}_G z_j^s = \text{ind}_G z_j^u = \text{ind}_F z_j$;
- for $j = 1, \ldots, m$, $z_j^s$ and $z_j^u$ are on the boundary $Y$, furthermore $z_j^s$ is boundary stable, $z_j^u$ is boundary unstable and $G(z_j^s) < G(z_j^u)$.

In other words, we can move all critical points to the boundary at once. To prove Theorem 4.10 we use Theorem 3.1 independently for each critical point $z_1, \ldots, z_m$. We need to ensure that Condition 3.2 holds. This is done in Proposition 4.11 stated below. Given these two ingredients the proof is straightforward.

4.6. Topological ingredients needed in the proof of Theorem 4.10.

Proposition 4.11. Let $F$ be a technically good Morse function on the cobordism $(\Omega, Y)$. Assume that $\Sigma_0$, $\Sigma_1$ and $\Omega$ have no closed connected components. Let $c, d$ be as in Definition 4.8. Then

(a) If $n > 1$, then for any $y \in [c, d]$, the inverse image $F^{-1}(y)$ has no closed connected component.

(b) If $n = 1$, then after possibly rearranging the critical values of the interior critical points of index 1, for any interior critical point $z \in \Omega$ of index 1, $z$ can be connected with $Y$ by a curve lying entirely in $F^{-1}(F(z))$; and furthermore all the critical points are on different levels.

Remark 4.12. The distinction between cases $n > 1$ and $n = 1$ is necessary. Point (a) of Proposition 4.11 is not necessarily valid if $n = 1$, see Figure 18 for a simple counterexample.

First let us prove several lemmas, which are simple consequences of the assumptions of Proposition 4.11. We use assumptions and notation of Proposition 4.11.

Lemma 4.13. Let $x, y \in [0, 1]$ with $x < y$. If $\Omega'$ is a connected component of $F^{-1}[x, y]$ then either $\Omega' \cap Y = \emptyset$, or for any $u \in [x, y] \cap [c, d]$ we have $F^{-1}(u) \cap \Omega' \cap Y \neq \emptyset$.

Proof. Assume that for some $u \in [x, y] \cap [c, d]$ the intersection $F^{-1}(u) \cap \Omega' \cap Y = \emptyset$ and $\Omega' \cap Y \neq \emptyset$. Then either $\Omega' \cap Y \cap F^{-1}[0, u]$ or $\Omega' \cap Y \cap F^{-1}[u, 1]$ is not empty. Assume the first possibility (the other one is symmetric) and let $Y' = \Omega' \cap Y \cap F^{-1}[0, u]$. Let $f = F|_{Y'}$.
Figure 18. The statement of Proposition 4.11(a) does not hold if \( n = 1 \). Here, \( F \) is the height function. The level set \( F^{-1}(1/2) \), drawn on the picture, has two connected components, one of which is closed.

be the restriction. Then \( Y' \) is compact and \( f \) has a local maximum on \( Y' \). This maximum corresponds to a critical point of \( f \) of index \( n \), so either a boundary stable critical point of \( F \) of index \( n + 1 \), or a boundary unstable critical point of index \( n \). But the corresponding critical value is smaller than \( u \), so smaller than \( d \), which contradicts property (TG2).

Lemma 4.14. For any \( x \in [c, 1] \) the set \( F^{-1}[0, x] \) cannot have a connected component disjoint from \( Y \).

Proof. Assume the contrary, and let \( \Omega' \) be a connected component of \( F^{-1}[0, x] \) disjoint from \( Y \). Let \( \Omega_1 \) be the connected component of \( \Omega \) containing \( \Omega' \). Suppose that \( \Omega_1 \cap Y = \emptyset \). Then either \( \Omega_1 \cap (\Sigma_0 \cup \Sigma_1) = \emptyset \) or \( \Omega_1 \cap (\Sigma_0 \cup \Sigma_1) \neq \emptyset \). In the first case \( \Omega_1 \) is a closed connected component of \( \Omega \), in the second either \( \Omega_1 \cap \Sigma_0 \), or \( \Omega_1 \cap \Sigma_1 \) is not empty, so either \( \Sigma_0 \) or \( \Sigma_1 \) has a closed connected component. The contradiction implies that \( \Omega_1 \cap Y \neq \emptyset \).

By Lemma 4.13 we have then \( F^{-1}(x) \cap \Omega_1 \cap Y \neq \emptyset \), hence \( \Omega'' := (F^{-1}[0, x] \cap \Omega_1) \setminus \Omega' \) is not empty and is disjoint from \( \Omega' \). As \( \Omega' \) and \( \Omega'' \) both belong to \( \Omega_1 \) which is connected, there must be a critical point \( z \in \Omega_1 \) of index 1, which joins \( \Omega' \) to \( \Omega'' \). We have \( F(z) > x \) and \( \Omega', \Omega'' \) belong to two different connected components of \( F^{-1}[0, F(z)] \) and to a single connected component of \( F^{-1}[0, F(z)] \). The connected component of \( F^{-1}[0, F(z)] \) containing \( \Omega' \) has empty intersection with \( Y \) (by Lemma 4.13) hence \( z \) must be an interior critical point of index 1. We also remind that all critical points of \( F \) on \( \Omega' \) are interior critical points, because \( \Omega' \cap Y = \emptyset \).

Let \( W^s \) be the stable manifold of \( z \) of the vector field \( \nabla F \). Then \( W^s \cap \Omega_1 \) must be a connected curve, with non-empty intersection with \( \Omega' \). One of its boundaries is either on \( \Sigma_0 \cap \Omega' \) or it is a critical point of \( F \) in \( \Omega' \), necessarily interior and by the Morse-Smale condition, its index is 0. In the first case, \( \Sigma_0 \cap \Omega' \) is not empty and since it is disjoint from \( Y \), \( \Sigma_0 \) has a closed connected component. In the other case, we have in \( \Omega_1 \) a single trajectory between a critical point of index 0 and a critical point of index 1. This contradicts (TG3).

Lemma 4.15. Assume that for some \( y \in [c, d] \) \( \Sigma_1 \) and \( \Sigma_2 \) are two disjoint connected components of \( F^{-1}(y) \). If \( F \) has no interior or boundary unstable critical points of index \( n \)
with critical value in \([c, y)\), then \(\Sigma_1\) and \(\Sigma_2\) belong to two different connected components of \(F^{-1}[0, y]\).

**Proof.** For \(x < y\) and close to \(y\) the sets \(\Sigma_1\) and \(\Sigma_2\) lie in two different connected components of \(F^{-1}(x, y)\). Let \(x_0 > 0\) be the smallest \(x\) with that property. Assume \(x_0 > 0\). Then \(x_0\) is a critical value of \(F\). The number of connected components of \(F^{-1}(x)\) increases as \(x\) crosses \(x_0\). Thus the corresponding critical point is either an interior critical point of index \(n\), or a boundary unstable critical point of index \(n\). But then \(x_0 > c\) because of (TG2), so we have \(x_0 \in [c, y]\) which contradicts the assumptions of the lemma.

It follows that \(x_0 = 0\). As \(F\) has no critical points on \(\Sigma_0\), it follows that \(\Sigma_1\) and \(\Sigma_2\) belong to different components of \(F^{-1}[0, y]\).

**Lemma 4.16.** Let \(y \in [c, d]\) be chosen so that there are no interior or boundary unstable critical points of index \(n\) with critical values in \([c, y)\). Then \(F^{-1}(y)\) has no closed connected components.

**Proof.** Assume that there is a closed connected component of \(F^{-1}(y)\). Let \(\Sigma'' = F^{-1}(y) \setminus \Sigma'\), it is not empty by Lemma 4.13 (applied for \(x = 0, y = 1\)), for otherwise \(\Omega \cap Y = \emptyset\). Let \(\Omega'\) be the connected component of \(F^{-1}[0, y]\) containing \(\Sigma'\). By Lemma 4.15, \(\Omega'\) and \(\Sigma''\) are disjoint, in particular \(\Omega' \cap F^{-1}(y) \cap Y = \Sigma' \cap Y = \emptyset\). By Lemma 4.13, \(\Omega' \cap Y = \emptyset\). But this contradicts Lemma 4.14.

**Remark 4.17.** There exists a symmetric formulation of the last three lemmas, which can be obtained by considering the function \(1 - F\) instead of \(F\). For instance, in Lemma 4.16 the symmetric assumption is that there are no interior or boundary stable critical points of index 1 in \((y, d]\). The statement is the same.

**Proof of Proposition 4.11.** Case \(n > 1\). Let \(x \in [c, d]\) be a non-critical value such that all the critical points of \(F\) with index \(n\) have critical value greater than \(x\), and all critical points with index smaller than \(n\) have critical values smaller than \(x\). Such \(x\) exists because of (TG1). If \(y \leq x\), then Lemma 4.16 guarantees that \(F^{-1}(y)\) has no closed connected components. If \(y > x\), then \(F^{-1}[y, d]\) has no critical points of index 1 (as \(n > 1\)), so we apply the symmetric counterpart of Lemma 4.16.

**Proof of Proposition 4.11.** Case \(n = 1\). The property (TG2) implies that the only critical points of \(F|_{[c, d]}\) are the interior critical points of index 1. Let us call them \(z_1, \ldots, z_m\). Since they are all of the same index, by Proposition 4.16 we are able to rearrange the values \(F(z_1), \ldots, F(z_m)\) at will. Let us fix \(c_1, \ldots, c_m\) with the property that \(c < c_1 < \cdots < c_m < d\). Let us first rearrange the points \(z_1, \ldots, z_m\) so that \(F(z_1) = \cdots = F(z_m) = c_1\).

The singular level set \(F^{-1}(c_1)\) is a manifold with \(m\) singular points \(z_1, \ldots, z_m\), which are double points. By Lemma 4.16, \(F^{-1}(c_1)\) has no closed connected components. In particular each of the points \(z_1, \ldots, z_m\), can be connected to \(Y\) by a curve lying in \(F^{-1}(c_1)\). At least one of those points can be connected to \(Y\) by a curve \(\gamma\), which omits all the other critical points. We relabel the critical points so that this point is \(z_m\). We rearrange the critical points so that \(F(z_m) = c_m\) and the value \(F(z_1) = \cdots = F(z_{m-1}) = c_1\). By construction, \(z_m\) can be connected with \(Y\) by a curve lying in \(F^{-1}(c_m)\).

The procedure now is repeated, i.e. assume that we have already moved \(z_{k+1}, \ldots, z_m\) to levels \(c_{k+1}, \ldots, c_m\) respectively. Then \(F^{-1}(c_1)\) still has no closed connected components by Lemma 4.16. We assume that \(z_k\) can be connected to \(Y\) by a curve in \(F^{-1}(c_1)\) omitting all the other critical points. Then we rearrange the critical values so that \(F(z_k) = c_k\). The proof is accomplished by an inductive argument.
4.7. Splitting of cobordisms. We have now all the ingredients needed to prove our main theorem about splitting cobordisms. We slightly change the notation in this subsection, the cobordism will be between $(\Sigma, M)$ and $(\Sigma', M')$.

**Theorem 4.18.** Let $(\Omega, Y)$ be a cobordism between $(\Sigma, M)$ and $(\Sigma', M')$. If the following conditions are satisfied

- $\Sigma$ and $\Sigma'$ have no closed connected components;
- $\Omega$ has no closed connected component;

Then the relative cobordism can be expressed as a union

$$\Omega = \Omega_0 \cup \Omega_{1/2} \cup \Omega_1 \cup \Omega_{3/2} \cup \cdots \cup \Omega_{n+1/2} \cup \Omega_{n+1}$$

such that $\partial \Omega_s = \Sigma_s \cup \Sigma_{s+1/2} \cup Y_s$ with $\Sigma_0 = \Sigma$, $\Sigma_{n+3/2} = \Sigma'$, $Y = Y_0 \cup \cdots \cup Y_{n+1}$. In other words $(\Omega_s, Y_s)$ is a cobordism between $(\Sigma_s, M_s)$ and $(\Sigma_{s+1/2}, M_{s+1/2})$, where $M_s = \partial \Sigma_s = \Sigma_s \cap Y_s$. Furthermore

- $(\Omega_0, Y_0)$ is a cobordism given by a sequence of index 0 handle attachments;
- for $k = 1, \ldots, n+1$, $(\Omega_{k-1/2}, Y_{k-1/2})$ is a left product cobordism, given by a sequence of index $k$ left half-handle attachments;
- for $k = 1, \ldots, n$, $(\Omega_k, Y_k)$ is a right product cobordism, given by a sequence of index $k$ right half-handle attachments;
- $(\Omega_{n+1}, Y_{n+1})$ is a cobordism provided by a sequence of index $(n+1)$ handle attachments.

**Proof.** Let us begin with a Morse function $F$ on the cobordism which has only boundary stable critical points (see Remark 2.33). Assume that $w_1, \ldots, w_m$ are the interior critical points and $y_1, \ldots, y_k$ are the boundary critical points. After a rearrangement of critical points and the cancellation of pairs of critical points as in Lemma 4.9 we can make $F$ technically good. After applying Theorem 4.10 we get that $F$ can have only 0 handles and $n + 1$ handles as interior handles. Let us write $\theta = 1/(4n + 6)$ and choose $c_0 = \theta$, $c_1 = 3\theta$, $c_1 = 5\theta$, $\ldots$, $c_k = (4k - 1)\theta$, $c_k = (4k + 1)\theta$, $\ldots$, $c_{n+1} = 1 - 3\theta$, $c_{n+1} = 1 - \theta$. We rearrange the function $F$ according to Proposition 4.6. Then we define for $k = 0, 1/2, 1, \ldots, n + 1$ the manifold $\Omega_k = F^{-1}[4k\theta, (4k + 2)\theta]$, $Y_k = \Omega_k \cap Y$ and $\Sigma_k = F^{-1}(4k\theta)$.

By construction, each part $(\Omega_k, Y_k)$ contains critical points only of one type: for $k = 0$ and $n + 1$ they are interior critical points, for $k = 1, \ldots, n$ they are boundary unstable of index $k$ and for $k = 1/2, \ldots, n + 1/2$, they are boundary stable of index $k + 1/2$ and we conclude the proof by Proposition 2.35. \qed

**Remark 4.19.** If the cobordism is a product on the boundary, i.e. $Y = M \times [0, 1]$, we can choose the initial Morse function to have no critical points on the boundary. Then all the critical points of $F$ come in pairs, $z^i_j$ and $z^u_j$ with $z^i_j$ boundary stable, $z^u_j$ boundary unstable and there is a single trajectory of $\nabla F$ going from $z^i_j$ to $z^u_j$.

The strength of Theorem 4.18 is that it is much easier to study the difference between the intersection forms on $(\Sigma_k, M_k)$ and on $(\Sigma_{k+1/2}, M_{k+1/2})$. We refer to [BNR] for an application of this fact.

5. The cancellation of boundary handles

In this section we assume that $F$ is a Morse function on the cobordism $(\Omega, Y)$ satisfying the Kronheimer–Mrowka–Morse–Smale regularity condition (Definition 1.5). We assume that $F$ has precisely two critical points $z$ and $w$, with $\text{ind } z = k$ and $\text{ind } w = k + 1$ and that there exists a single trajectory $\gamma$ of $\nabla F$ going from $z$ to $w$. If $z$ and $w$ are both interior critical points, then [Mi2, Theorem 5.4] implies that $(\Omega, Y)$ is a product cobordism. In fact,
Milnor’s proof modifies $F$ only in a small neighbourhood of $\gamma$, which avoids the boundary $Y$. Hence, it does not matter, that in our case the cobordism has a boundary.

We want to extend this result to the case of boundary critical points. In some cases an analogue of the Milnor’s theorem holds, in other cases we can show that it cannot hold.

5.1. Elementary cancellation theorems.

**Theorem 5.1.** Let $\gamma$ be a boundary critical points of index $k$ and $k + 1$, respectively. Assume that $\gamma$ is a single trajectory joining $z$ and $w$. Furthermore, assume that both $z$ and $w$ are boundary stable, or both boundary unstable. Then $(\Omega, Y)$ is a product cobordism.

As usual, it is enough to prove the result for boundary unstable critical points, the other case is covered if we change $F$ to $1 - F$. Note also, cf. Section 5.2 the assumption that both critical points are boundary stable, or both boundary unstable is essential.

A careful reading of Milnor [Mi2, pages 46–66] shows that the proof there applies to this situation with only small modifications. Below we present only three steps of that proof, adjusted to our situation. We refer to [Mi2] for all the missing details.

Let $\xi$ be the gradient vector field of $F$. The proof relies on the following proposition (see the Preliminary Hypothesis 5.5 in [Mi2], proved on pages 55–66).

**Proposition 5.2.** There exist an open neighbourhood $U$ of $\gamma$ and a coordinate map $g: U \to \mathbb{R}_{\geq 0} \times \mathbb{R}^n$ and a gradient-like vector field $\xi'$ agreeing with $\xi$ away from $U$ such that

- $g(Y) \subset \{x_1 = 0\}$, and $g(U) \subset \{x_1 \geq 0\}$;
- $g(z) = (0, 0, \ldots, 0)$;
- $g(w) = (0, 1, \ldots, 0)$;
- $g_*\xi' = \eta = (x_1, v(x_2), -x_3, \ldots, -x_k, x_{k+1}, \ldots, x_{n+1})$, where $v$ is a smooth function positive in $(0, 1)$, zero at $0$ and $1$ and negative elsewhere. Moreover $\left| \frac{dv}{dx_2} \right| = 1$ near $0$ and $1$.

Furthermore, $U$ can be made arbitrary small (around $\gamma$).

Given the proposition, we argue in the same way, as in the classical case, cf. [Mi2, pages 50–55]: we improve the vector field $\xi'$ in $U$ so that it becomes a gradient like vector field of a function $F'$, which has no critical points at all. Then the cobordism is a product cobordism.

The proof of Proposition 5.2 is a natural modification of the Milnor’s proof. After applying arguments as in [Mi2, pages 55–58] the proof boils down to the following result.

**Proposition 5.3** (compare [Mi2, Theorem 5.6]). Let $a + b = n$, $a \geq 1$ and $b \geq 0$ and write a point $x \in \mathbb{R}_{\geq 0} \times \mathbb{R}^{a-1} \times \mathbb{R}^b$ as $(x_a, x_b)$ with $x_a \in \mathbb{R}_{\geq 0} \times \mathbb{R}^{a-1}$ and $x_b \in \mathbb{R}^b$. Assume that $h: (\mathbb{R}_{\geq 0} \times \mathbb{R}^{a-1}, \{0\} \times \mathbb{R}^{a-1}) \to (\mathbb{R}_{\geq 0} \times \mathbb{R}^{n-1}, \{0\} \times \mathbb{R}^{n-1})$ is an orientation preserving diffeomorphism such that $h(0) = 0$. Suppose that $h(\mathbb{R}_{\geq 0} \times \mathbb{R}^{a-1} \times \{0\})$ intersects $\{0\} \times \{0\} \times \mathbb{R}^b$ only at the origin and the intersection is transverse and the intersection index is $-1$. Then, given any neighbourhood $N$ of $0 \in \mathbb{R}_{\geq 0} \times \mathbb{R}^{a-1}$, there exists a smooth isotopy $h'_t$ for $t \in [0, 1]$ of diffeomorphisms from $(\mathbb{R}_{\geq 0} \times \mathbb{R}^{n-1}, \{0\} \times \mathbb{R}^{n-1})$ to itself with $h'_0 = h$ such that

(I) $h'_t(x) = h(x)$ away from $N$;

(II) $h'_t(x) = x$ in some small neighbourhood $N_1$ of $0$ such that $\overline{N_1} \subset N$;

(III) $h'_1(\mathbb{R}_{\geq 0} \times \mathbb{R}^{a-1} \times \{0\}) \cap \{0\} \times \{0\} \times \mathbb{R}^b = \{0\} \in \mathbb{R}_{\geq 0} \times \mathbb{R}^n$.

**Remark 5.4.** The transversality assumption from the assumption of Proposition 5.3 is equivalent to the flow of $\xi$ being Morse–Smale.

The proof of Proposition 5.3 in Milnor’s book is given on pages 59–66. We prove here only the analogue of [Mi2, Lemma 5.7]. For all other results we refer to Milnor’s book.
Lemma 5.5. Let h be as in hypothesis of Proposition 5.5. Then there exists a smooth isotopy \( h_t : \mathbb{R}_{\geq 0} \times \mathbb{R}^{n-1} \rightarrow \mathbb{R}_{\geq 0} \times \mathbb{R}^{n-1} \), with \( h_0 \) the identity map and \( h_t = h \), such that for each t we have \( h_t (\mathbb{R}_{\geq 0} \times \mathbb{R}^{a-1}) \cap \mathbb{R}^b = 0 \).

Proof. We follow the proof of [Mi2] Lemma 5.7. We shall construct the required isotopy in two steps. First we isotope h by \( h_t (x) = \frac{1}{t} h(tx) \). Then \( h_1 = h \) and \( h_0 \) is a linear map. If this is an identity, we are done. Otherwise \( h_0 \) is just a nondegenerate linear map and clearly it maps \( \mathbb{R}_{\geq 0} \times \mathbb{R}^{a-1} \times \mathbb{R}^b \) diffeomorphically onto itself. It means that under the decomposition \( \mathbb{R}^n = \mathbb{R} \oplus \mathbb{R}^{a-1} \oplus \mathbb{R}^b \), \( h_0 \) has the following block structure

\[
h_0 = \begin{pmatrix} a_{11} & 0 & 0 \\ * & A & B \\ * & C & D \end{pmatrix},
\]

where \( a_{11} > 0 \), and stars denote unimportant terms. As \( h_0 \) is orientation-preserving, \( \det h_0 > 0 \). We can apply a homotopy of linear maps which changes the first column of \( h_0 \) to \((a_{11}, 0, \ldots, 0) \) and preserves all the other entries of \( h_0 \). We do not change the determinant and the condition \( h_0 (\mathbb{R}_{\geq 0} \oplus \mathbb{R}^{a-1}) \cap \mathbb{R}^b = 0 \) is preserved (it means that \( a_{11} \det A > 0 \)). Let

\[
h_{00} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}.
\]

Obviously \( \det h_{00} = a_{11} \det h_{00} \), so \( \det h_{00} > 0 \). We use the same reasoning as in the Milnor’s proof to find a homotopy of \( h_{00} \) to the identity matrix, finishing the proof. \( \square \)

5.2. Non-cancellation results. The two results below have completely obvious proofs, we state them to contrast with Theorem 5.1.

Lemma 5.6. Assume that a Morse function F on the cobordism \((\Omega, Y)\) between \((\Sigma_0, M_0)\) and \((\Sigma_1, M_1)\) has two critical points z and w. Suppose z is an interior critical point and w is a boundary critical point. Then \((\Omega, Y)\) is not a product cobordism.

Proof. F restricted to Y has a single critical point, so the cobordism between \( M_0 \) and \( M_1 \) cannot be trivial. \( \square \)

Lemma 5.7. Suppose that F has two critical points z and w. Assume that z is boundary stable and w is boundary unstable. Then \((\Omega, Y)\) is not a product.

Proof. If it were a product, we would have \( H_\ast (\Omega, \Sigma_0) = 0 \). We shall show that this is not the case.

If \( F(z) = F(w) \), then there are no trajectories between z and w, so by Proposition 4.11 we can ensure that \( F(z) < F(w) \). So we can always assume that \( F(z) \neq F(w) \). For simplicity assume that \( F(z) < F(w) \). Let c be a regular value such that \( F(z) < c < F(w) \).

By Lemma 2.17 \( F^{-1}[0, c] \sim \Sigma_0 \times [0, c] \). Then \( H_\ast (\Omega, \Sigma_0) \cong H_\ast (\Omega, F^{-1}[0, c]) \). Now \( \Omega \) arises from \( F^{-1}[0, c] \) by a right half-handle addition, hence \( H_\ast (\Omega, F^{-1}[0, c]) \cong H_\ast (H, B) \), where \((H, B)\) is the corresponding right half-handle. But \( H_\ast (H, B) \) is not trivial by Lemma 2.22 (or Lemma 2.16).

\( \square \)

References


Institute of Mathematics, University of Warsaw, ul. Banacha 2, 02-097 Warsaw, Poland.

E-mail address: mcboro@mimuw.edu.pl


E-mail address: nemethi@renyi.hu

School of Mathematics, University of Edinburgh, Edinburgh EH9 3JZ, Scotland UK.

E-mail address: a.ranicki@ed.ac.uk