Lecture III: h-cobordism Theorem

September 20, 2020

1 Introduction

Theorem 1.1. *Let W be a compact, simply connected manifold of dimension $n \geq 6$. Suppose $\partial W = \partial_- W \coprod \partial_+ W$ with each $\partial_\pm W$ being a union of connected components of ∂W. Suppose that the inclusion of each $\partial_\pm W \to W$ is a homotopy equivalence. Then there is a diffeomorphism of triples*

$$(\partial_- W \times I, \partial_- W \times \{0\}, \partial_- W \times \{1\}) \to (W, \partial_- W, \partial_+ W).$$

In particular, $\partial_- W$ and $\partial_+ W$ are diffeomorphic.

Remark 1.2. 1. The h-cobordism theorem holds for $n \leq 3$. The cases $n = 1, 2$ are trivial. The case $n = 3$ is equivalent to the celebrated Poincaré Conjecture, proved by Perelman. By work of Donaldson, the result does not hold for $n = 5$. Whether it holds for $n = 4$ is unknown and is equivalent to whether every smooth, homotopy 4-sphere is diffeomorphic to S^4. By Freedman's work a homotopy 4-sphere is homeomorphic to S^4, so another equivalent question is whether S^4 has a unique differentiable structure up to diffeomorphism.

2. The h-cobordism theorem as stated holds for PL manifolds and topological manifolds as well as smooth manifolds. The proof in the PL case is a fairly straight-forward modification of the smooth proof.

3. We will discuss the non-simply connected case in the next lecture.

Corollary 1.3. *(The Generalized Poincaré Conjecture) Let Σ^n be a smooth manifold homotopy equivalent to S^n. If $n \geq 6$, then Σ^n is the union of two smooth n-balls glued together by a diffeomorphism of their boundaries. In particular, Σ^n is homeomorphic to S^n.*

Proof. Let B_-, B_+ be disjoint closed n-balls in Σ^n and let $W = \Sigma^n \setminus (\mathrm{int}(B_- \coprod B_+)$. with $\partial_\pm W$ being $\partial B_\pm$. Since Σ^n is homotopy equivalent to S^n, W is homotopy equivalent to $S^{n-1} \times I$ and the inclusion of each boundary component is a homotopy equivalence. By the h-cobordism theorem W is diffeomorphic to $S^{n-1} \times I$. Consequently, the union of $B_- \cup W$ is diffeomorphic to a ball and the first statement follows.

For the second it is enough to prove that every homeomorphism of the sphere extends over the ball. Given a homeomorphism of the unit sphere in $\mathbb{R}^n$, $h\colon S^{n-1} \to S^{n-1}$, the map that sends $x \neq 0$ in the unit ball to $|x| \cdot h(x/|x|)$ extends continuously to a map fixing the origin. The result is a continuous bijective map of the unit ball to itself which is automartcally a homeomorphism since the unit ball is a compact Hausdorff space. $\square$

Remark 1.4. **(a)** Special arguments in dimensions ≤ 5 show that the purely topological version of Poincaré's conjecture is true in all dimensions.
(b) It is not true in general that Σ^n as in the corollary is diffeomorphic to S^n. There are examples starting in dimension 7. The reason is that not every diffeomorphism of the S^{n-1} extends to diffeomorphism of the unit ball B^n. The group of diffeomorphsims of S^{n-1} modulo the normal subgroup of diffeomorphisms that extend over the ball is denoted Γ_n. The result is that for $n \geq 5$ the set of homotopy n-spheres up to diffeomorphism is identified with Γ_n. Indeed, the set of homotopy n-spheres is a group under connected sum and this group is identified with Γ_n.
(c) Every PL manifold of dimension ≥ 5 that is homotopic equivalent to a sphere is PL isomorphic to the sphere. The proof is the PL h-cobordism theorem plus the fact that the same coning argument shows that any PL isomorphism of the sphere. extends to a PL homeomorphism of the ball.

In the rest of this lecture we first state the two main cancellation results that will allow us to remove all the critical points of a Morse functioin on an h-cobordism, and then we give the proof assuming these two results. Those will be established in the next lecture.

2 Basic Cancellation Results

Fix a manifold W as in the statement of the h-cobordism theorem except that we make no assumption on its dimension n. We shall work with Morse functions $f\colon W \to \mathbb{R}$ consistent with the decomposition of $\partial W = \partial_- W \coprod \partial_+ W$ in the sense that Df is positive, resp. negative, on outward pointing normal vectors along $\partial_+ W$, resp. $\partial_- W$.

There are two main theorems allowing us to cancel pairs of critical points under certain hypotheses. We postpone their proofs until the next lecture.

Theorem 2.1. *(First Cancellation Result)*
Suppose:

1. *M is compact (possibly with boundary) and $f \colon M \to \mathbb{R}$ is a Morse function.*
2. *p and q are critical points of indices r and $r+1$ respectively with $f(p) < f(q)$; there are constants $a < f(p) < f(q) < b$ such that p and q are the only critical points with critical values in $[a,b]$.*
3. *χ is a gradient-like vector field for f.*
4. *Setting $N = f^{-1}(\ell)$ for some $\ell \in (f(p), f(q))$, the intersection of the descending disk for q, the ascending disk for p and N is a single point x of transverse intersection.*

Then there is are: (i) a Morse function f' that agrees with f outside $f^{-1}([a,b])$ and has no critical points in $f^{-1}([a,b]) = (f')^{-1}([a,b])$ and (ii) a gradient-like vector field χ' for f' that agrees with χ outside $f^{-1}(a,b])$.

Indeed, the proof shows that we choose χ' to agree with χ outside any neighborhood of the closure of the unique flowline T for χ from p to q.

Theorem 2.2. *(Second Cancellation Result) Suppose that W is a compact manifold of dimension $n \geq 6$ and that $f \colon W \to \mathbb{R}$ is a self-indexing Morse function without critical points of index $0, 1, n-1, n$. Suppose that $\partial_- W$ and $\partial_+ W$ are simply connected and let k be the smallest index of a critical point with p a critical point of index k and q a critical point of index $k+1$. Suppose that χ is a gradient-like vector field for f and that the ascending sphere of p at level $\ell = ((2k+1)/2)$ meets the descending sphere of q at level ℓ transversely. Fix orientations for the descending manifolds for p and q giving a sign to each intersection point of the descending sphere for q and the ascending sphere for p at level ℓ. If the sum of the local algebraic intersections is ± 1, then there is a new gradient-like vector field χ' for f, agreeing with χ outside $[k+(1/4), k+(3/4)]$ so that there is only one χ' flow connection between p to q, and this is a transverse flow line. Hence, we can cancel p and q using the First Cancellation Theorem applied to f and χ'.*

3 First steps in the proof of the h-cobordism theorem assuming the cancellation results

First we establish two additional results that we shall need in the proof.

Proposition 3.1. *(An Algebraic Lemma) Let* L *and* L' *be integral lattices of rank* k *with given ordered bases* $\{e_1, \ldots, e_k\}$ *and* $\{f_1, \ldots, f_k\}$*, respectively. let* $A\colon L \to L'$ *be a linear isomorphism of lattices. Then there is a sequence of ordered bases* $\beta_1, \ldots, \beta_T$ *for* L *such that*

- $\beta_1 = \{e_1, \ldots, e_k\}$.
- $\beta_T = \{A^{-1}(f_1), \ldots, A^{-1}(f_k\}\}$,
- *For each* $1 \leq i \leq T-1$ *the basis* β_{i+1} *is obtained from the basis* β_i *by one of the following three* elementary operations*:*

  (i) reordering of the basis,

  (ii) multiplying one of the basis elements by -1 *and leaving the others unchanged, or*

  (iii) replacing a basis element by sum of it with an integral multiple of another basis element and leaving the other basis elements unchanged.

Proof. Translated to matrices this is the statement that given a square integer $k \times k$ matrix of determinant ± 1 there are elementary column operations that reduce it to the identity matrix, where the elementary column operations are: (i) interchanging two columns; (ii) multiplying a single column by -1; and (iii) replacing a column by its sum with an integral multiple of another column and leaving all others unchanged.

To prove this consider the first row. The gcd of its entries is 1. Thus, by the Euclidean algorithm, by replacing one entry by the sum or difference of it with another entry and leaving all other entires unchanged we can reduce the minimum absolute value of the entries in this row unless this minimum absolute value is one. This addition or subtraction is achieved by an elementary column operation of the third type. We continue in this way until the minimum absolute value is 1. Multiplying the relevant column by -1 if necessary and switching this column with the first column, we arrive at a matrix with a 1 in the upper left corner. Once we have this, subtracting multiples of the first column from the other columns allows us to make all the entries in the first row except the first equal to 0.

The determinant of the minor made out of columns and rows 2 through k is one. By induction we can assume that column operations using only columns 2 through n allow us to make this minor the identity. At this point the diagonal entries are 1 and all off-diagonal entries are 0 except those off-diagonal entries in the first column. Now adding multiples of columns 2 through n to the first column allows us to arrive at the identity matrix. $\square$

Corollary 3.2. *Suppose that L, L' are lattices with ordered bases $\{e_1, \ldots, e_s\}$ and $\{f_1, \ldots, f_t\}$ respectively and suppose that $A\colon L \to L'$ is a surjective linear map. Then there is a sequence of elementary operations on the first basis producing a new basis $\{e_1', \ldots, e_s'\}$ for L such that:*

$$A(e_i') = \begin{cases} f_i & \text{for } 1 \leq i \leq t \\ 0 & \text{for } t+1 \leq i \leq s \end{cases}.$$

Proof. Let $K \subset L$ be the kernel of A. We have a short exact sequence

$$0 \to K \longrightarrow L \xrightarrow{A} L' \longrightarrow 0.$$

Since L' is free, there is a splitting $B\colon L \to K$ of the inclusion. Then $A \oplus B\colon L \to L' \oplus K$ is an isomorphism of lattices. Choose an ordered basis $\{k_1, \ldots, k_{s-t}\}$ for K and apply the previous theorem to the isomorphism $A \oplus B$ with $\{e_1, \ldots, e_s\}$ as basis for L and $\{f_1, \ldots, f_t, k_1, \ldots, k_{s-t}\}$ as basis for $L' \oplus K$. □

Proposition 3.3. *(Birth Result) Let B be the unit ball in $\mathbb{R}^n$ (with coordinates $(x^1, \ldots, x^n)$), for some $n \geq 1$. Fix $0 \leq k < n$. Then there is a smooth one-parameter family $f_t\colon B \to \mathbb{R}$ parameterized by $t \in (-\epsilon, \epsilon)$ for some $\epsilon > 0$ such that:*

1. *near ∂B the function f_t agrees with x^1 for every t,*
2. *for $t < 0$ f_t has no critical points,*
3. *for $t > 0$ f_t has two critical points p_t and q_t with $f_t(p_t) < f_t(q_t)$ of index k and $k+1$ respectively, and*
4. *for any ℓ with $f_t(p_t) < \ell < f_t(q_t)$ the descending sphere from q_t w/r/t ∇f_t at level ℓ meets the ascending sphere from p_t at level ℓ transversally in exactly one point.*

Proof. (Following Milnor) The models for a birth are as follows. Consider $\mathbb{R} \times \mathbb{R}^r \times \mathbb{R}^{n-r-1}$ with coordinates (x, y, z) (so that y and z are vector coordinates) and the family

$$g_t(x, y, z) = (\frac{1}{3}x^3 - tx) - |y|^2 + |z|^2.$$

For $t < 0$ there are no critical points; $t = 0$ is a degenerate critical point at the origin; and $t > 0$ has a pair of non-degenerate critical points at

$p = (-\sqrt{t}, 0, 0)$ and $q = (\sqrt{t}, 0, 0)$ of indices $r+1$ and r, respectively. For every $t > 0$ the ascending disks for p are the lie in the (x, z)-hyperplane and the descending disks for q lie in the (x, y)-hyperplane. Hence, there is a unique flow line for ∇f_t between the critical points, namely the interval on the x-axis from $\sqrt{t}$ to $-\sqrt{t}$ and this is flow line the the transverse intersection of the ascending disk for p and the descending disk for q.

We only consider the parameter t only in the range $(-\epsilon, \epsilon)$ for some sufficiently small $\epsilon > 0$. We shall damp out this family producing a family $f_t(x, y, z)$ that agrees with $g_t(x, y, z)$ on a neighborhood of 0 (independent of $t \in (-\epsilon, \epsilon)$ and is equal to x outside a fixed compact set (independent of t.). We start by letting $a_t(x)$ be function equal to $(\frac{1}{3}x^3 - tx)$ for $|x| \leq 1/2$ and equal to x for $|x| \geq 1$. In doing this damping we introduce no new critical points for any $t \in (-\epsilon, \epsilon)$. We write $s_t(x) = a_t(x) - x$, so that s_t is supported in $[-1, 1]$ for all $t \in (-\epsilon, \epsilon)$. Now choose $\alpha(r)$ so that it is equal to 1 on $[0, 1]$ and is zero outside some compact set and satisfies $|\alpha'(r)| < 1/\max_{x,t}(|s_t(x)|)$ for all r. Choose $\beta(r)$ such that $\beta(r) = 1$ if $\alpha(r) \neq 0$ and $\beta = 0$ outside a compact set. Lastly, choose $\gamma(x)$ supported in a compact set with $\gamma(x) = 1$ for $|x| \leq 1$ and $|\gamma'(r)| < 1/\max_t(t\beta(t))$. We define

$$f_f(x, y, z) = x + s_t(x)\alpha(|y|^2 + |z|^2) + \beta(|y|^2 + |z|^2)(-|y|^2 + |z|^2))\gamma(x).$$

Clearly, $f_t(x, y, z) = g_t(x, y, z)$ in some neighborhood of 0 which depends only on the choice of α, β, γ. and $f_t(x, y, z) = x$ outside some ball centered at the origin (depending on α, β, γ). Direct computation (see p.102 in Milnor's "Lectures on the H-cobordism Theorem") shows that for every $t \in (-\epsilon, \epsilon)$ all critical points of f_t are contained in the neighborhood of the origin where $f_t = g_t$.

We can rescale the ball to be of any given positive size and multiply each member of the family f_t by the same parameter to make it equal to x near the boundary of this rescaled ball. $\square$

With these preliminary results in place, the rest of this lecture is devoted to proving the h-cobordism theorem assuming the two cancellation theorems.

4 Proof of the h-cobordism theorem, assuming the two cancellation results stated above

Let $(W, \partial_- W, \partial_+ W)$ be a cobordism as in the statement of the h-cobordism theorem with W connected. Let $n \geq 6$ be the dimension of W. We begin with a self-indexing Morse function $f\colon W \to \mathbb{R}$ and a gradient-like vector field χ. The method of proof is to cancel all the critical points by varying f and χ using the result stated above.

Cancellation of critical points of index 0 and n. Since W is connected and by Corollary 1.4 of Lecture 2, W is homotopy equivalent to a space obtained from $f^{-1}(-\infty, 3/2])$ by attaching cells of dimension ≥ 2, it follows that $f^{-1}(-\infty, 3/2])$ is connected. Applying the same result again, the union of $f^{-1}(-\infty, 1/2])$ and the descending one-disks from the critical points of index 1 is connected. Of course, $f^{-1}(-\infty, 1/2])$ is diffeomorphic to the disjoint union of $\partial_- W \times [-1, 1/2]$ with ascending n disk up to level $1/2$ for each critical point of index 0. If there is a critical point p of index 0, then there is a critical point q of index 1 whose descending 1-disk connects the ascending n disk of some critical point p' of index 0 to $\partial_- W$. In level $1/2$ the descending manifold of q meets the ascending manifold of p' in a single point, which is a point or transverse intersection. Now shift f slightly near the points p', q allows us to assume that the critical value of p' is $+\epsilon$ and the critical value of q is $1 - \epsilon$ and at all other critical points f is self-indexing. Let $W_0 = f^{-1}([\epsilon/2, 1 - \epsilon/2]$. The cancellation Theorem applies allowing us to change $f|_{W_0}$ and $\chi|_{W_0}$ to a Morse function f' and gradient-like vector field χ' in W_0 that agree with f and χ near ∂W_0 and so that f' has no critical points. Then $f_1 = f' \cup f_{W \setminus W_0}$ is a Morse function and $\chi_1 = \chi' \cup \chi|_{W \setminus W_0}$ is a gradient-like vector field for this Morse function. The set of critical points for the Morse function f_1 is the set of all critical points of f except p', q and near any critical point for f_1, the functions f and f_1 agree. In particular, f_1 is a self-indexing Morse function with one fewer critical point of index 0 and one fewer critical point of index 1.

Repeating this argument as long as there are critical points of index 0, we remove al critical points of index 0 and an equal number of critical points of index 1.

To remove the critical points of index n, we apply the same argument to $-f$ and $-\chi$ and use the observation that these two Morse functions f and $-f$ have the same critical points but the indices are complementary, so that the critical points of index 0 for $-f$ are those of index n for f.

Removal of Critical Points of Index 1 and $(n-1)$.

Choose a small ball B centered at a point $x \in f^{-1}(21/8)$ and choose coordinates of B such that $f = x^1$ and $\chi = \lambda \partial / \partial x^1$. Using the Birth Theorem there is a smooth function $g \colon B \to \mathbb{R}$ that agrees with f near ∂B and a gradient-like vector field for g that agrees with χ near the boundary such that g has two critical points p_2, p_3 of index 2 and 3, respectively both at levels above $5/2$ with a single transverse flow line between these critical points. We replace $f|_B$ by g_B to produce two new critical points p_2 and

p_3, of indices 2 and 3 respectively, for the Morse function, By the Isotopy Theorem (Theorem 3.3 from Lecture 2), varying gradient-like vector field slightly allows us to assume that the descending manifold for p_2 is disjoint from all the all the other critical points of index 2. Thus, the descending disk for p_2 extends down to level $5/2$. Let C be its boundary circle.

Claim 4.1. *For any critical point q of index 1 there is a smoothly embedded loop γ_q in level $3/2$ meeting the ascending disk for q transversally in a single point.*

Proof. There is a parallel copy γ^0 of the descending arc from q in level $3/2$ that crosses the ascending manifold transversely in one point and whose endpoints are in the boundary of the canonical neighborhood. The function f and the gradient-like vector field χ determine a product structure $(\partial_- W \setminus \coprod S^0 \times D^{n-1}) \times I$ for the complement of this neighborhood in $f^{-1}([-1, 3/2])$. Under this product structure, the endpoints of γ^0 lie in $X = (\partial_- W \setminus \coprod S^0 \times D^{n-1}) \times (3/2)$. Since $\partial_- W$ is connected and of dimension at least 5, it follows that X is also connected. Thus, we can connect the endpoints of γ^0 by an arc γ^1 in X. The union $\gamma_q = \gamma^0 \cup \gamma^1$ is as required. $\square$

Fix γ_q as in the previous claim. Since the dimension of $f^{-1}(3/2)$ is $n - 1 \geq 5$, by general position and the Isotopy Theorem, we can assume that γ_q is disjoint from all the descending disks of critical points of index 2. Thus, γ_q flows under the gradient-like vector field to a soothly embedded loop γ'_q at level $5/2$. Since W is simply connected and homotopy equivalent to the union of $f^{-1}([-1, 5/2]$ and handles of dimension ≥ 3, it follows that $f^{-1}([-1, 5/2])$ is also simply connected. Replacing $f|_{[-1,5/2]}$ by its negative we see that $f^{-1}([-1, 5/2])$ is obtained from $f^{-1}(5/2)$ by adding handles of degrees $n-2$ and $n-1$. Since $n \geq 6$, it follows that $f^{-1}(5/2)$ is also simply connected. Thus, γ'_q and C are homotopic in $f^{-1}(5/2)$. Since the manifold $f^{-1}(5/2)$ is of dimension at least 5, it follows that C and γ'_q are isotopic in $f^{-1}(5/2)$. Applying the Isotopy Theorem we can change the gradient-like vector field in the region $f^{-1}([5/2, 5/2 + \epsilon])$ keeping it the same near the boundary of this region so that the descending disk for p_2 has boundary sphere γ'_q at level $5/2$. Hence, the descending disk from p extends to level $3/2$ where its boundary is γ_q. Now apply the Cancellation Theorem to cancel the critical point q of index 1 and the critical point p_2 of index 2, leaving an addition critical point p_3 of index 3. Then use the VCV result, Proposition 2.4 from Lecture 2 to move the the critical value of this p_3 to 3 without affecting f near the other critical points. The effect of this is to 'trade' a critical point of index 1 for a critical point of index 3. (Though the argument uses $n \geq 6$ explicitly, It actually only requires $n \geq 5$.)

Repeat this argument for each critical point of index 1 to trade all of these for critical points of index 3 ending up with no critical points of index 1.

The critical points of index $n-1$ are critical points for $-f$ of index 1 and are traded in the same way for critical points of index $n-3 \geq 3$.

At this point we assume that f has critical points only if indices k for $2 \leq k \leq n-2$.

Cancellation of critical points of index k for $2 \leq k \leq n-3$.

Suppose that $f\colon W \to \mathbb{R}$ is a self-indexing Morse function with no critical points of index $0, 1, n-1, n$, Suppose that it has at least one critical point.

Let $2 \leq k \leq n-2$ be the smallest index of a critical point. In fact, $k \leq n-3$, since if the smallest index of a critical point were $n-2$, then all critical points would be of index $n-2$ and the relative homology of $(W, \partial_- W)$ would be non-trivial. Let $\{q_1 \ldots, q_s\}$ be the critical points of index k and give orientations to their descending disks so as to determine an ordered basis $\{f_1, \ldots, f_t\}$ for $C_k(f, \chi)$. Similar choices for the critical points $\{p_1, \ldots, p_s\}$ of index $k+1$ determine an ordered basis $\{e_1, \ldots, e_s\}$ for $C_{k+1}(W, \partial_- W)$.

By the Isotopy Theorem and general position, we arrange that the intersections at level $(2k+1)/2$ between the descending spheres for critical points p_i and ascending spheres of critical points q_j are transverse. As described in Lecture 2, the orientations of the descending manifolds determine local signs (± 1) at each point and the sum over all intersection points between the i^{th} descending sphere and the j^{th} ascending sphere determines a matrix (a_{ij}) which is the boundary map from degree $k+1$ to degree k in the Morse chain complex:

$$\partial e_i = \sum a_{ij} f_j.$$

The Morse complex computes the relative homology of $(W, \partial_- W)$. In our case, the relative homology is trivial. The Morse chain group in degree $k-1$ is zero (since there are no critical points of index $k-1$). Thus, the matrix (a_{ij}) represents a surjective map $C_{k+1}(W, \partial_- W) \to C_k(W, \partial_- W)$. By Corollary 3.2 there is a sequence of elementary transformations that carry the basis $\{e_1, \ldots, e_s\}$ to $\{e'_1, \ldots, e'_s\}$, where $\partial(e'_i) = f_i$ for $i \leq t$ and $\partial(e'_i) = 0$ for $i > t$. The elementary transformations are: (i) reordering the basis, (ii) changing the sign of a basis element, (iii) replacing e_1 by $e_1 + e_2$ and leaving the others fixed. Our goal is to change the geometric data in a way that reflects these changes so that the descending disks for the new

Morse function and gradient-like vector field give the basis $\{e'_1, \ldots, e'_t\}$. The first two are easy to implement: reordering the basis of $C_{k+1}(f, \chi)$ is implemented by renumbering the critical points of index $k + 1$. Changing the sign of a basis element is implemented by reversing the orientation of the corresponding descending disk from the critical point of index $k + 1$.

Lemma 4.2. *For any $\epsilon > 0$ sufficiently small, we can achieve the third transformation by changing f and χ in $f^{-1}[(k + 1 - \epsilon), (k + 1 + \epsilon)]$.*

.

Proof. It suffices to show that we can implement the following elementary basis change:

$$\{e_1, e_2, \ldots, e_t\} \mapsto \{e_1 + e_2, e_2, \ldots, e_t\}.$$

Using the VCV lemma we find a new Morse function f' equal to $f + \epsilon/2$ near p_1 and equal to f outside a canonical neighborhood of p_1 and outside $f^{-1}[(k + 1 - \epsilon), (k + 1 + \epsilon)]$, with the property that χ is gradient-like for f'. Consider the descending sphere S^k_1 from p_1 at level $k + 1 + \epsilon/4$. It is disjoint from all the ascending spheres from the other critical points of index $k + 1$ (which have critical value $k + 1$). There is an embedded arc $[-1, 2] \subset f^{-1}(k+1+\epsilon/4)$ that meets S^k_1 at $\{0\}$ and mets the ascending disk S^{n-k-2}_2 from p_2 at $\{1\}$. We can extend this to an embedding $(-1, 2] \times \mathbb{R}^k \times \mathbb{R}^{n-k-2} \to f^{-1}(k+1+\epsilon/4)$ so that its intersection of S^k_1 is $\{0\} \times \mathbb{R}^k \times \{0\}$ and its intersection with the ascending sphere for p_2 is $\{1\} \times \{0\} \times \mathbb{R}^{n-k-2}$ with the orientation of $\mathbb{R}^k$ agreeing with that of S^k_1 along the first intersection and agreeing with the normal orientation of S^{n-k-2}_2 at the second intersection.

Now we use this embedded copy of $[-1, 2] \times \mathbb{R}^k \times \mathbb{R}^{n-k-2}$ to define an isotopy of S^{k-1}_1 supported in its intersection with $\{0\} \times \mathbb{R}^{k-1} \times \{0\}$. To define the isotopy let $\rho\colon \mathbb{R}^k \to [-1, 2]$ be a smooth function supported in the ball of radius 2 about the origin and equal to 2 on the ball of radius 1 and damped out by a positive function ≤ 1 in the annular region. Then define $\varphi_t = t\rho$ in $\mathbb{R}^k \subset S^k_1$ and $\varphi_t = \mathrm{Id}$ outside $\mathbb{R}^k$. The result is an isotopy of S^k_1 which is the identity at time 0, that crosses S^{n-k-2}_2 transversely at time $t = 1/2$ in a single point and is otherwise disjoint from S^{n-k-2}_2. We extend this isotopy to an ambient isotopy of $f^{-1}(k + 1 + \epsilon/4)$ and reparameterize this isotopy by a weak reparameterization function $[k + 1 + \epsilon/4, k + 1] \to [0, 1]$. The resulting ambient isotopy is equal to the identity near the upper boundary and a constant isotopy near the lower boundary. According to the Isotopy Theorem there is a gradient-like vector field for f', equal to χ outside $f^{-1}([k + 1, k + 1 + \epsilon/4]$ so that the descending disk from p_1 agrees with the original descending disk on $f^{-1}([k+1+\epsilon/4, k+1+\epsilon/2])$ and is the image under the isotopy of the descending disk on $f^{-1}(k+1,[k+1+\epsilon/4]$. Thus, the descending disk from p_1 meets each of the ascending disks from p_1 and p_2 transversely in a single point with intersection number $+1$ and is disjoint from all other ascending disks from critical points of index $k+1$. According to Lemma 4.1 of Lecture 2, this implies that the descending disk from p_1 for χ' represents e_1+e_2. Since $\chi = \chi'$ on $f^{-1}([(2k+1)/2, k+1])$ the descending manifolds from the critical points $p_2, \ldots, p_t$ are unchanged and hence their oriented descending disks represent $e_2, \ldots, e_t$, respectively. Lastly, use the VCV lemma to move the critical value of p_1 back to $k+1$ keeping the Morse function fixed outside $f^{-1}([k+1-\epsilon, k+1+\epsilon)]$, so that the new Morse function is self-indexing. □

Repeating these three geometric operations a finite number of times we arrive at a Morse function f' that agrees with f outside $f^{-1}([(2k+1)/2, k+1+\epsilon])$ has the same critical points and critical values as f and a gradient-like vector field χ' such that for appropriate ordering of the critical points of index $k+1$ and orientations of their descending manifolds the matrix for $\partial\colon C_{k+1}(f',\chi') \to C_k(f',\chi')$ is given by the matrix (a_{ij}) where $a_{ij} = \delta_{ij}$. We rename these objects f and χ.

Now we can use the Second Cancellation theorem to remove critical points in pairs. Use the VCV result to lower the critical value at p_1 and raise the critical value at q_1. Let f' be the resulting Morse function. The vector field χ is gradient-like for f'. Since the algebraic intersection of the descending sphere from p_1 with the ascending from q_1 at level $(2k+1)/2$ is $+1$ and since there are no other critical points with critical values in the interval $[f(q), f(p)]$, the Second Cancellation Theorem tells us that we can cancel p_1 and q_1. That is to say there is a Morse function f'' with no critical points in the open interval $(k, k+1)$ and agreeing with f outside $f^{-1}([k+\epsilon, k+1-\epsilon])$. Thus, f' is self-indexing, has no critical points of index $< k$ and has one fewer critical point of index k than f does.

We repeat this entire argument killing the all the critical points of index k. (Actually, the argument shows that the deformation of f and χ to remove the critical points q_1, p_1 can be done without affecting the descending manifolds from $p_2, \ldots, p_t$, so that once we have found a basis in which the boundary map is diagonal basis, as we remove pairs of critical points the remaining basis elements continue to give a diagonal matrix for the boundary map.)

Now by induction on k we can arrange that f has no critical points of index $0, 1, n-1, n$ and no critical points of index k with $2 \leq k \leq n-3$.

As we showed at the beginning of the proof, this implies that f has no critical points at all. By Lemma 1.1 of Lecture 2, this implies that W is diffeomorphic to $\partial_- W \times I$.