Pergamon

Topology Vol. 38, No. 4, pp. 861–888, 1999

Printed in Great Britain
0040-9383/99 $19.00 + 0.00

PII: S0040-9383(98)00044-5

CIRCLE-VALUED MORSE THEORY, REIDEMEISTER TORSION, AND SEIBERG–WITTEN INVARIANTS OF 3-MANIFOLDS

MICHAEL HUTCHINGS and YI-JEN LEE

(*Received* 25 *September* 1996; *in revised form* 16 *July* 1998)

Let X be a closed oriented Riemannian manifold with $\chi(X) = 0$ and $b_1(X) > 0$, and let $\phi : X \to S^1$ be a circle-valued Morse function. Under some mild assumptions on ϕ, we prove a formula relating

1. the number of closed orbits of the gradient flow of ϕ in different homology classes;
2. the torsion of the Novikov complex, which counts gradient flow lines between critical points of ϕ; and
3. a kind of Reidemeister torsion of X determined by the homotopy class of ϕ.

When $\dim(X) = 3$, we state a conjecture related to Taubes's "SW = Gromov" theorem, and we use it to deduce (for closed manifolds, modulo signs) the Meng–Taubes relation between the Seiberg-Witten invariants and the "Milnor torsion" of X.

1. INTRODUCTION

1.1. Background on circle-valued Morse theory and torsion

Let X be an n-dimensional closed Riemannian manifold and $\phi : X \to \mathbb{R}$ a Morse function.

Definition 1.1. Recall that the *Morse complex* (C^{Morse}, d) is defined as follows. The chain module C_i^{Morse} is the free $\mathbb{Z}$-module generated by Crit_i, the set of critical points of index i. The differential $d : C_i^{\text{Morse}} \to C_{i-1}^{\text{Morse}}$ sends a critical point $x \in \text{Crit}_i$ to

$$d(x) := \sum_{y \in \text{Crit}_{i-1}} \langle dx, y \rangle y$$

where $\langle dx, y \rangle$ denotes the signed number of downward gradient flow lines (i.e. flow lines of $-\nabla\phi$) from x to y. (If the metric on X is generic, then there are a finite number of such flow lines, and we can attach a sign to each one after certain orientation data are chosen, see Section 2.2.)

One of the fundamental theorems of Morse theory (which evolved in [36, 32, 21, 41]) is:

THEOREM 1.2 (cf. [1, 5, 31]). *$d^2 = 0$ and $H_i(C_*^{\text{Morse}}, d) = H_i(X)$.*

Novikov [25] generalized this to multiple-valued functions, i.e. closed 1-forms. In the simplest version of Novikov's construction, we consider a circle-valued Morse function $\phi : X \to S^1$. There is an analogue of the Morse complex which counts gradient flow lines of ϕ. In this circle-valued case, to obtain a finite count we need to classify flow lines using some information about their homotopy classes. A minimal way to do this is as follows.

Definition 1.3. Define the *Novikov complex* $(C_*^{\text{nov}}, \mathbf{d})$ as follows. Let C_i^{nov} denote the free $\mathbb{Z}((t))$-module generated by Crit_i. (Notation: if R is a ring, then $R((t))$ denotes the ring of

Laurent series with coefficients in R, i.e. formal sums $\sum_{k=k_0}^{\infty} a_k t^k$ with $k_0 \in \mathbb{Z}$ and $a_k \in R$.) Assume that 0 is a regular value of ϕ and consider the level set

$$\Sigma := \phi^{-1}(0) \subset X.$$

If $x \in \mathrm{Crit}_i$, we define

$$\mathbf{d}(x) := \sum_{y \in \mathrm{Crit}_{i-1}} \sum_{k=0}^{\infty} \langle \mathbf{d}x, y \rangle^k t^k y$$

where $\langle \mathbf{d}x, y \rangle^k$ denotes the signed number of flow lines of $-\nabla\phi$ from x to y that cross Σ a total of k times. (For a generic metric, $\mathbf{d}$ is well defined, and $\mathbf{d}^2 = 0$.)

The Novikov complex has a topological counterpart.

Notation 1.4. (a) Let $\tilde{X}$ denote the infinite cyclic cover of X induced by ϕ, i.e. the fiber product of X and $\mathbb{R}$ over $S^1 = \mathbb{R}/\mathbb{Z}$:

$$\begin{tikzcd} \tilde{X} \arrow[r] \arrow[d] & \mathbb{R} \arrow[d] \\ X \arrow[r, "\phi"] & \mathbb{R}/\mathbb{Z} \end{tikzcd}$$

We sometimes regard $\tilde{X}$ as a subset of $X \times \mathbb{R}$.

(b) Choose a cell decomposition of X, and lift the cells to obtain an equivariant cell decomposition of $\tilde{X}$. Let $C_*^{\mathrm{cell}}(\tilde{X})$ denote the cellular chain complex. This is a module over $\mathbb{Z}[t, t^{-1}]$, where t acts by the "downward" deck transformation sending $(x, \lambda) \mapsto (x, \lambda - 1)$.

THEOREM 1.5. (Novikov [25], cf. [26, 29, 19]).

$$H_i(C_*^{\mathrm{nov}}) \simeq H_i(C_*^{\mathrm{cell}}(\tilde{X}) \otimes_{\mathbb{Z}[t,t^{-1}]} \mathbb{Z}((t))).$$

One can think of $C_*^{\mathrm{cell}}(\tilde{X}) \otimes_{\mathbb{Z}[t,t^{-1}]} \mathbb{Z}((t))$ as the complex of "half-infinite" chains in $\tilde{X}$. Note that the Novikov homology $H_i(C_*^{\mathrm{nov}})$ is usually quite different from $H_i(X) \otimes \mathbb{Z}((t))$; for example, if $d\phi$ is not exact, then $H_0(C_*^{\mathrm{nov}}) = H_n(C_*^{\mathrm{nov}}) = 0$.

When $\chi(X) = 0$ and $b_1(X) > 0$, it may happen that all of the Novikov homology vanishes, at least after tensoring with the field $\mathbb{Q}((t))$. In this case it is interesting to consider the Reidemeister torsion of the Novikov complex.

Definition 1.6. Suppose $C_m \xrightarrow{\partial} C_{M-1} \xrightarrow{\partial} \cdots \xrightarrow{\partial} C_0$ is an acyclic complex of finite-dimensional vector spaces over a field F, and suppose that each vector space C_i has a volume form chosen on it. Choose $\omega_i \in \wedge^* C_i$ for each i so that $0 \neq \partial\omega_{i+1} \wedge \omega_i \in \wedge^{\mathrm{top}} C_i$. Then the *Reidemeister torsion* $\tau(C_*, \partial) \in F$ is defined to be

$$\tau(C_*) := \prod_{i=0}^{m} \mathrm{vol}(\partial\omega_{i+1} \wedge \omega_i)^{(-1)^{i+1}} \tag{1.1}$$

(where we interpret $\partial\omega_{m+1} = 1$). One can check, using the fact that $\partial^2 = 0$, that $\tau(C_*)$ does not depend on the choice of ω_i's.

The Novikov complex has a natural basis consisting of the critical points, and this gives a volume form on $C_i^{\mathrm{nov}} \otimes \mathbb{Q}((t))$, up to sign.

Definition 1.7. If $C_*^{\mathrm{nov}} \otimes \mathbb{Q}((t))$ is acyclic, we define the *Morse-theoretic Reidemeister torsion*

$$T_{\mathrm{Morse}} := \tau(C_*^{\mathrm{nov}} \otimes \mathbb{Q}((t))) \in \frac{\mathbb{Q}((t))}{\pm 1}.$$

(Note that choosing a different level set Σ will multiply T_{Morse} by a power of t. More intrinsically, one should define the Novikov complex using the critical points in $\tilde{X}$; a basis is then determined by choosing lifts of the critical points in X to $\tilde{X}$.)

If $C_*^{\mathrm{nov}} \otimes \mathbb{Q}((t))$ is acyclic, then so is the cell complex $C_*^{\mathrm{cell}}(\tilde{X}) \otimes \mathbb{Q}(t)$ (where $\mathbb{Q}(t)$ is the field of rational functions), by Novikov's theorem. The latter complex has a basis consisting of a lift of each cell in X to $\tilde{X}$, and this gives a volume form, defined up to sign and powers of t.

Definition 1.8. If $C_*^{\mathrm{cell}}(\tilde{X}) \otimes \mathbb{Q}(t)$ is acyclic, we define the *topological Reidemeister torsion*,

$$T_{\mathrm{top}} := \tau(C_*^{\mathrm{cell}}(\tilde{X}) \otimes \mathbb{Q}(t)) \in \frac{\mathbb{Q}(t)}{\pm t^k}.$$

If $C_*^{\mathrm{cell}}(\tilde{X}) \otimes \mathbb{Q}(t)$ is not acyclic, we define $T_{\mathrm{top}} := 0$.

The invariant T_{top} depends only on the homotopy class of ϕ, namely the cohomology class $[d\phi] \in H^1(X; \mathbb{Z})$. (This follows for instance from Lemma 2.6.) We sometimes denote T_{top} by $T_{\mathrm{top}}(X, [d\phi])$ to indicate this dependence.

Example 1.9. Let $K \subset S^3$ be a knot, and let X be the 3-manifold obtained by 0-surgery on K. Milnor observed (cf. [37] or Lemma 2.6) that if α is a generator of $H^1(X; \mathbb{Z}) \simeq \mathbb{Z}$ then

$$T_{\mathrm{top}}(X, \alpha) = \frac{\mathrm{Alex}(K)}{(1-t)^2}$$

where $\mathrm{Alex}(K) \in \mathbb{Z}[t,t^{-1}]/\pm t^k$ is the Alexander polynomial of K.

We can compare the Morse-theoretic and topological Reidemeister torsion using the natural inclusion $\imath : \mathbb{Q}(t) \hookrightarrow \mathbb{Q}((t))$. It turns out that usually

$$T_{\mathrm{Morse}} \neq \imath(T_{\mathrm{top}}).$$

In fact, the Morse-theoretic torsion is not a topological invariant when $d\phi$ is not exact.

Example 1.10. Let $X = S^1$, and suppose $\phi : X \to S^1$ has degree $k \neq 0$. If ϕ has $2c > 0$ critical points then for a certain choice of orientation data, the differential $\mathbf{d} : C_1^{\mathrm{nov}} \to C_0^{\mathrm{nov}}$ has the form

$$\mathbf{d} = \begin{pmatrix} t^{a_1} & -t^{b_1} & 0 & \cdots & 0 \\ 0 & t^{a_2} & -t^{b_2} & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ -t^{b_c} & 0 & \cdots & 0 & t^{a_c} \end{pmatrix}.$$

Since ϕ has degree k, the exponents satisfy $\sum_i a_i - \sum_i b_i = k$, so up to signs and powers of t,

$$T_{\mathrm{Morse}} = (1 - t^k)^{-1}.$$

(We also have $T_{\mathrm{top}} = (1 - t^k)^{-1}$, as one can see by choosing a cell decomposition of X with one 1-cell and one 0-cell.) However, if we choose a different function ϕ with no critical points, then $T_{\mathrm{Morse}} = 1$.

1.2. The main theorem

To obtain a topologically invariant Reidemeister torsion from the Morse-theoretic data, we must also consider the closed orbits of the gradient flow. This is the novelty of the present paper, from the standpoint of Morse theory. Of course, closed orbits of flows have been extensively studied in the dynamical systems literature, and they are typically counted by various "zeta functions" going back to [40]. We will use the following basic "zeta function".

Definition 1.11. (a) For $k = 1, 2, \ldots$ we define partially defined *return maps*

$$f^k : \Sigma \to \Sigma$$

as follows. If $p \in \Sigma$, then $f^k(p)$ is the kth intersection with Σ of the trajectory of the flow $-\nabla\phi$ starting at p, if it exists. If the flow from p does not cross Σ k times (due to the intervention of a critical point), then $f^k(p)$ is not defined.

(b) We define the *zeta function*

$$\zeta := \exp\left(\sum_{k=1}^{\infty} \operatorname{Fix}(f^k)\frac{t^k}{k}\right).$$

Here $\operatorname{Fix}(f^k)$ counts the fixed points of f^k with signs. If $p \in \Sigma$ is a fixed point of f^k, we define its sign to be the sign of $\det(1 - df_p^k)$. (For a generic metric on X, these determinants are always nonzero, and there are only finitely many fixed points for each k.)

THEOREM 1.12 (Main result). *Let X be a closed smooth manifold and $\phi : X \to S^1$ a circle-valued Morse function. Choose a generic metric on X so that $(C_*^{\mathrm{nov}}, \mathbf{d})$ and ζ are defined. Suppose $C_*^{\mathrm{cell}}(\tilde{X}) \otimes \mathbb{Q}(t)$ is acyclic.* (*In particular, $\chi(X) = 0$ and $0 \neq [d\phi] \in H^1(X; \mathbb{Z})$.*) *Then*

$$T_{\mathrm{Morse}} \cdot \zeta = \imath(T_{\mathrm{top}}) \tag{1.2}$$

in $\mathbb{Q}((t))$, up to sign and multiplication by powers of t.

We can re-express equation (1.2) as a kind of Lefschetz formula for the partially defined maps f^k, in which the Reidemeister torsion of the Novikov complex appears as a correction term.

Notation 1.13. Let $Q_i : H_i(\tilde{X}; \mathbb{Q}) \to H_i(\tilde{X}; \mathbb{Q})$, or $Q : H_*(\tilde{X}; \mathbb{Q}) \to H_*(\tilde{X}; \mathbb{Q})$, denote the map in rational homology induced by the "upward" deck transformation of $\tilde{X}$ which sends $(x, \lambda) \mapsto (x, \lambda + 1)$.

When $C_*^{\mathrm{nov}} \otimes \mathbb{Q}((t))$ is acyclic, the rational homology $H_i(\tilde{X}; \mathbb{Q})$ is a finite dimensional $\mathbb{Q}$-vector space (see Lemma 2.5), and we will see in Section 2.6 that

$$T_{\mathrm{top}} = c \prod_{i=0}^{n-1} \det(1 - tQ_i)^{(-1)^{i+1}} \tag{1.3}$$

(modulo $\pm t^k$) where $c \in \mathbb{Q}$. Hence, taking the logarithmic derivative of (1.2) gives:

THEOREM 1.14 (Main result, alternate statement). *Under the assumptions of Theorem* 1.12, *we have*

$$\sum_{k=1}^{\infty} t^k Fix(f^k) + t\frac{d}{dt}\log T_{\mathrm{Morse}} = \operatorname{Str}((1 - tQ)^{-1}) + m \tag{1.4}$$

in $\mathbb{Z}((t))$, where $m \in \mathbb{Z}$.

(Here Str denotes graded trace, using the homology grading, and $(d/dt)\log T_{\mathrm{Morse}}$ denotes $(T_{\mathrm{Morse}})^{-1}(d/dt)T_{\mathrm{Morse}}$, which is well defined even though T_{Morse} has a sign ambiguity. The integer m absorbs the ambiguity of multiplication by powers of t in Theorem 1.12.)

Example 1.15. If ϕ has no critical points, then f is a diffeomorphism of Σ, we can identify $H_i(\tilde{X}) \simeq H_i(\Sigma)$, and under this identification $Q_i = H_i(f)$. (This is consistent with the fact that Q_i goes "up" and f goes "down", cf. Section 3.1.) Since $T_{\mathrm{Morse}} = 1$, Eq. (1.4) reduces to the Lefschetz fixed point theorem for f and its iterates. (In this case, the relation between zeta functions and torsion goes back to Milnor [22], and has been generalized by Fried [6, 7] to count closed orbits of certain nonsingular hyperbolic flows.)

Remark 1.16. Earlier papers such as [16, 27, 9] considered a different kind of torsion (Whitehead torsion, modulo units in the Novikov ring with leading coefficient ± 1) for a more refined version of the Novikov complex (using the universal cover), when the latter is acyclic. The version of torsion studied in these papers does not detect the zeta function, but it does give an obstruction to finding non-singular closed 1-forms in a given cohomology class, which is more or less sharp when $\dim(X) \geqslant 6$.

In Section 2.6 we will show that, in fact, Theorem 1.14 implies Theorem 1.12. We will then prove Theorem 1.14 in Section 3, by modifying a classical proof of the Lefschetz fixed point theorem.

1.3. Seiberg–Witten invariants of 3-manifolds

Our original motivation for this work was to compute the Seiberg–Witten invariants of a closed oriented 3-manifold X with $b_1(X) > 0$, following a suggestion of Taubes. The Seiberg–Witten invariant (see e.g. [15, 24, 2, 20], or [42, 23] for the four-dimensional case) is a map

$$\mathrm{SW} : \mathrm{Spin}^c(X) \to \mathbb{Z}.$$

Here $\mathrm{Spin}^c(X)$ denotes the set of spin-c structures on X; a spin-c structure is equivalent to a $U(2)$-bundle $W \to X$ with some extra structure, see Section 4.1. (The sign of SW depends on a choice of homology orientation of X. When $b_1 = 1$, the invariant SW also depends on a choice of "chamber".)

By a sort of dimensional reduction of Taubes' "SW = Gr" theorem relating Seiberg–Witten invariants of symplectic 4-manifolds to counting pseudoholomorphic curves, we conjecture (Conjecture 4.8) that the Seiberg–Witten invariant of our 3-manifold X equals an invariant I, defined in Section 4.1, which counts certain unions of closed orbits and flow lines between critical points of the vector field dual to a closed 1-form. We explain the motivation for this conjecture in more detail in Section 4.2.

The invariant I is similar to the left-hand side of Theorem 1.12, but it is sharper, because it keeps track of the relative homology classes of closed orbits and flow lines (and not just intersection numbers with Σ). Theorem 1.12 thus allows us to compute some of the information in I, leading (in Section 4.3) to:

THEOREM 1.17. *Let X be a closed oriented* 3*-manifold with $b^1 > 0$ and $0 \neq \alpha \in H^1(X; \mathbb{Z})$. Then Conjecture* 4.8 *implies that*

$$\sum_{W \in \mathrm{Spin}^c(X)} \mathrm{SW}(W) t^{\alpha(c_1(\det W))/2} = T_{\mathrm{top}}(X, [\alpha]) \tag{1.5}$$

modulo multiplication by $\pm t^k$. (*If* $b_1 = 1$, *then we mean* SW *to be computed in the chamber defined by* α, *see Conjecture* 4.8).

Remark 1.18. (a) When $b^1(X) > 1$, the t^k ambiguity can be resolved by applying the "charge conjugation invariance" of the Seiberg–Witten equations, which asserts that SW is invariant under conjugation of spin-c structures and implies that the left-hand side of (1.5) is invariant under $t \mapsto t^{-1}$. See [20], or [42, 23] for the four-dimensional case. (The invariant I of Conjecture 4.8 has the same duality property, as a result of the involution that replaces a circle-valued Morse function with its additive inverse.)

(b) Even if we apply this theorem to every $\alpha \in H^1(X; \mathbb{Z})$, we cannot recover all of the Seiberg–Witten invariants of X, because the theorem does not distinguish between spin-c structures that differ by the action of a torsion element of $H^2(X; \mathbb{Z})$ (i.e. by tensoring with a line bundle for which c_1 is torsion). However, we can recover (see Section 4.4), in the closed case modulo signs, the theorem of Meng and Taubes [20] which relates SW to "Milnor torsion".

1.4. Other approaches and further developments

In a sequel to this paper [12], we define a refined version of the invariant $T_{\mathrm{Morse}} \cdot \zeta$ for a circle-valued Morse function (or closed 1-form) and show that it equals a form of topological torsion defined by Turaev [38]. The refined Morse theory invariant in [12] is a natural generalization of the three-dimensional invariant I to n dimensions using ideas from [38]. The method of proof in [12] (following [18]) is different: we compute the topological torsion directly using a cell decomposition adapted to the Morse function. (The homological methods used to compute topological torsion in the present paper, cf. Section 2.5, are not sufficient to compute more refined versions of torsion.) Roughly speaking, the cell decomposition splits X into the union of a Morse-theoretic component, whose torsion equals that of the Novikov complex, and something similar to a mapping torus, whose torsion equals the zeta function. A preprint by Pajitnov [38], proving a similar result, appeared at about the same time as [12].

A third approach to Theorem 1.12 (and the sharper result of [12]) is to first prove that $T_{\mathrm{Morse}} \cdot \zeta$ is a topological invariant, and then use this fact to compute it (using tricks of Latour [16] or Pozniak [29] to reduce to the easier case when ϕ lifts to a real-valued function). To prove invariance, we study what happens to T_{Morse} and ζ as one deforms the Morse function ϕ and the metric through a generic one-parameter family.

If ϕ lifts to a real-valued function, then T_{Morse} is invariant (cf. [17]). For example, if a critical point $p \in \mathrm{Crit}_i$ "slides over" $q \in \mathrm{Crit}_i$ (i.e. if at some time in the deformation there is a flow line from p to q), then the effect on the Novikov complex is a change of basis in which p is replaced by $p \pm q$. This does not change the torsion since the change of basis matrix has determinant one. However, in the circle-valued case, it is possible for p to slide over $t^k p$. This multiplies T_{Morse} by $(1 \pm t^k)^{\pm 1}$. At the same time, the boundary of the graph of f^k (see equation Eq. (2.2)) crosses the diagonal in $\Sigma \times \Sigma$, causing a closed orbit to be created or destroyed. Due to a product formula for the zeta function (see Remark 4.6), the result is that the zeta function is multiplied by $(1 \pm t^k)^{\pm 1}$. Further analysis shows that the signs work out so that $T_{\mathrm{Morse}} \cdot \zeta$ is unchanged.

In the circle-valued case, T_{Morse} and ζ also change when two critical points of index difference one are created or destroyed; when two critical points die, it turns out that every k-times broken closed orbit or flow line at the time or bifurcation leads to an unbroken closed orbit or flow line after the bifurcation. By a little algebra, $T_{\text{Morse}} \cdot \zeta$ stays invariant when this happens. See [13] for details.

Update. In [12] we also show that Conjecture 4.8 implies that the full Seiberg–Witten invariant of a 3-manifold with $b_1 > 0$, up to sign, equals a refined form of Reidemeister torsion defined by Turaev in [38]. Turaev conjectured this conclusion in [38] and later showed in [39] how it may be proved by refining the work of Meng and Taubes [20]. Thus, the combination of [12, 39] confirms, indirectly, that Conjecture 4.8 is true.

2. PRELIMINARIES

2.1. More notation

The symbol $\cdot$ indicates (sometimes) intersection number of oriented submanifolds. We take $\alpha \cdot \beta$ to be zero if α and β do not have complementary dimension.

The Novikov complex possesses a natural inner product

$$\langle\ ,\ \rangle : C_*^{\text{nov}} \otimes C_*^{\text{nov}} \to \mathbb{Z}((t))$$

in which the critical points are orthonormal. We denote the adjoint (with respect to $\langle\ ,\ \rangle$) of the differential $\mathbf{d} : C_*^{\text{nov}} \to C_{*-1}^{\text{nov}}$ by $\mathbf{d}^* : C_*^{\text{nov}} \to C_{*+1}^{\text{nov}}$. We define

$$\Delta := \mathbf{d}\mathbf{d}^* + \mathbf{d}^*\mathbf{d}.$$

2.2. Sign conventions

If Y is an oriented manifold, $\phi : Y \to \mathbb{R}$, and $\lambda \in \mathbb{R}$ is a regular value of ϕ, we orient the level set $W = \phi^{-1}(\lambda)$ by declaring

$$TY|_W = \mathbb{R} \cdot \operatorname{grad}(\phi) \oplus TW$$

to be an isomorphism of oriented vector bundles.

If Y has a metric and p is a critical point of ϕ, the *ascending manifold* $\mathscr{A}(p)$ (resp. *descending manifold* $\mathscr{D}(p)$) is the set of $y \in Y$ such that downward (resp. upward) gradient flow starting at y converges to p. If ϕ is a Morse function, then $\mathscr{D}(p)$ and $\mathscr{A}(p)$ are embedded manifolds intersecting transversely at p. We choose orientations of $\mathscr{D}(p)$ and $\mathscr{A}(p)$ such that the oriented intersection number

$$\mathscr{D}(p) \cdot \mathscr{A}(p) = (-1)^{\text{index}}(p).$$

If $p \in \text{Crit}_i$ and $q \in \text{Crit}_{i-1}$, and if $\mathscr{D}(p)$ and $\mathscr{A}(q)$ intersect transversely along a flow line from p to q, then we define the sign of the flow line to be minus the local intersection number (at the point corresponding to the flow line) of $\mathscr{D}(p) \cap W$ with $\mathscr{A}(q) \cap W$ in W, where W is a level set of ϕ.

If Y is an oriented manifold with boundary, we orient the boundary via the convention

$$TY|_{\partial Y} = \mathbb{R} \cdot v \oplus T(\partial Y)$$

where v points outwards.

2.3. Morse-theoretic preliminaries

Our proof of Theorem 1.14 will consist in manipulation of a few simple formulas (stated in Lemma 2.3 below) concerning various maps between the Morse complex and chains in $\tilde{X}$ and Σ defined using the gradient flow. These formulas also lead to proofs of the fundamental isomorphisms between Morse/Novikov homology and ordinary homology.

Consider a "Morse cobordism", i.e. let Y be a compact n-dimensional manifold with boundary $\partial Y = Y_0 \cup Y_1$, and let $\phi : Y \to [0, 1]$ be a Morse function with $Y_0 = \phi^{-1}(0)$ and $Y_1 = \phi^{-1}(1)$. (In the case of interest later, Y will be a subset of $\tilde{X}$ sandwiched between two lifts of Σ.)

Choose a metric on Y so that the ascending and descending manifolds of the critical points intersect transversely. Let $(C_*^{\mathrm{Morse}}, d)$ denote the Morse complex as in Section 1.1. For $p \in \mathrm{Crit}$, we define the ascending and descending "slices" $A(p) := \mathscr{A}(p) \cap Y_1$ and $D(p) := \mathscr{D}(p) \cap Y_0$. The downward gradient flow defines a diffeomorphism

$$f : Y_1 \setminus \bigcup_{p \in \mathrm{Crit}} A(p) \to Y_0 \setminus \bigcup_{p \in \mathrm{Crit}} D(p).$$

We let $G(f) \subset Y_1 \times Y_0$ denote its graph. If $\alpha \subset Y_1$ is a submanifold (possibly with boundary), we let $\mathscr{F}(\alpha)$ denote the submanifold of all $y \in Y$ such that upward gradient flow from y converges to a point in α.

For any manifold Z, let $(C_*^{\mathrm{sing}}(Z), \partial)$ denote the complex of singular chains in Z. We call a singular chain in $C_*^{\mathrm{sing}}(Y_1)$ *generic* if every face of every simplex is smooth and transverse to the ascending manifolds of the critical points. Generic chains form a subcomplex $(C_*^{\mathrm{gen}}(Y_1), \partial)$, and a standard argument shows that its homology equals that of $C_*^{\mathrm{sing}}(Y_1)$. (See [8] for a similar use of generic singular chains.)

We will now define the following chains and maps:

$$\begin{aligned}
\overline{G(f)} &\in C_{n-1}^{\mathrm{sing}}(Y_1 \times Y_0) \\
\bar{\mathscr{D}} &: C_*^{\mathrm{Morse}} \to C_*^{\mathrm{sing}}(Y) \\
\bar{D} &: C_*^{\mathrm{Morse}} \to C_{*-1}^{\mathrm{sing}}(Y_0) \\
\bar{A} &: C_*^{\mathrm{Morse}} \to C_{n-*-1}^{\mathrm{sing}}(Y_1) \\
\bar{\mathscr{F}} &: C_*^{\mathrm{gen}}(Y_1) \to C_{*+1}^{\mathrm{sing}}(Y) \\
\overline{f_*} &: C_*^{\mathrm{gen}}(Y_1) \to C_*^{\mathrm{sing}}(Y_0).
\end{aligned} \tag{2.1}$$

The bars indicate that these are "compactifications" – more precisely, pushforwards of triangulations of certain abstract compactifications. We construct the latter using "broken flow lines", a standard technique in Floer theory.

Definition 2.1. For $k \geqslant 0$, a (k-times) *broken flow line* from $y \in Y$ to $y' \in Y$ is a sequence $(p_0, \ldots, p_{k+1})$ such that $p_0 = y$, $p_{k+1} = y'$, and $p_i \in \mathrm{Crit}$ for $1 \leqslant i \leqslant k$, together with downward flow lines from p_i to p_{i+1} for $0 \leqslant i \leqslant k$.

The abstract compactified graph, $\overline{G(f)}_{\mathrm{abs}}$, consists of broken flow lines from points in Y_1 to points in Y_0. The space $\overline{G(f)}_{\mathrm{abs}}$ naturally has the structure of a smooth manifold with corners, and the codimension k stratum consists of k-times broken flow lines. (For the analytic details of very similar constructions see e.g. [3, 8, 31].) There is a natural endpoint map

$$e : \overline{G(f)}_{\mathrm{abs}} \to Y_1 \times Y_0$$

which is smooth on each stratum. (The map e sends a broken flow line with itinerary $(p_0, \ldots, p_{k+1})$ to the pair (p_0, p_{k+1}).) We can choose a smooth triangulation of $\overline{G(f)}_{\mathrm{abs}}$, and we define $\overline{G(f)} \in C_{n-1}^{\mathrm{sing}}(Y_1 \times Y_0)$ to be the pushforward of this triangulation under e.

The map $\bar{\mathscr{D}}$ is defined analogously. If $p \in \mathrm{Crit}$, then $\bar{\mathscr{D}}(p)_{\mathrm{abs}}$ consists of broken flow lines starting at p and ending anywhere in Y. Again this is a manifold with corners whose codimension k stratum consists of k-times broken flow lines. To obtain the chain $\bar{\mathscr{D}}(p)$, we push forward a smooth triangulation of $\bar{\mathscr{D}}(p)_{\mathrm{abs}}$ to Y via the lower endpoint map $\bar{\mathscr{D}}(p)_{\mathrm{abs}} \to Y$ (which sends a broken flow line with itinerary $(p_0 = p, \ldots, p_{k+1})$ to p_{k+1}). We then extend $\bar{\mathscr{D}}$ linearly from the set Crit to the module C_*^{Morse}.

It should now be clear how we define the remaining objects in (2.1). We define one more map

$$A^\dagger : C_*^{\mathrm{gen}}(Y_1) \to C_*^{\mathrm{Morse}}$$

by requiring that for $\alpha \in C_*^{\mathrm{gen}}(Y_1)$ and $p \in \mathrm{Crit}$,

$$\alpha \cdot A(p) = \langle A^\dagger \alpha, p \rangle.$$

Before stating our key formulas, we need to discuss certain harmless error terms which arise in them.

Definition 2.2. A singular chain $\alpha \in C_k^{\mathrm{sing}}(Z)$ is *degenerate* if it is a linear combination of chains satisfying (a) or (b) below:

(a) There is a compact k-dimensional manifold with corners M, a map $e: M \to Z$, and two chains $\beta_1, \beta_2 \in C_k(M)$ obtained from different triangulations of M, such that

$$\alpha = e_*(\beta_1 - \beta_2).$$

(b) There are compact manifolds with corners M and N of dimensions $j < k$ and $k - j$, a map $e: M \to Z$, and a chain $\beta \in C_k(M \times N)$ coming from a triangulation, such that $\alpha = (e \circ \pi_M)_*(\beta)$, where π_M is the projection $M \times N \to M$.

Observe that degenerate chains form a subcomplex, and modding out by this subcomplex does not affect homology. Moreover, smooth generic chains have intersection number zero with smooth chains that intersect them transversely. Thus we can and *will mod out by degenerate chains* in the intersection-theoretic and homological calculations in the rest of the paper.

LEMMA 2.3. *The following equations hold, modulo degenerate chains*:

$$\partial \overline{G(f)} = \sum_{i=1}^{n-1} \sum_{p \in \mathrm{Crit}_i} (-1)^{(n-1)(i-1)} \bar{A}(p) \times \bar{D}(p) \tag{2.2}$$

$$\partial \bar{\mathscr{F}} = -\bar{\mathscr{F}} \partial + 1 - \bar{f}_* - \bar{\mathscr{D}} A^\dagger \tag{2.3}$$

$$\partial \bar{\mathscr{D}} = \bar{\mathscr{D}} d - \bar{D} \tag{2.4}$$

$$\partial \bar{D} = -\bar{D} d \tag{2.5}$$

$$\partial \bar{A} = -\bar{A} d^* (-1)^{\mathrm{index}} \tag{2.6}$$

$$d A^\dagger = A^\dagger \partial \tag{2.7}$$

$$\partial \bar{f}_* = \bar{f}_* \partial + \bar{D} A^\dagger. \tag{2.8}$$

Proof. Except for the orientations, these formulas are fairly clear from the construction of the compactifications, and express the fact that certain submanifolds of different abstract compactifications map to the same submanifold of Y. For example, in (2.2), the boundary (i.e. the codimension one stratum) of $\overline{G(f)}_{\text{abs}}$ comes from flow lines broken by a single critical point p, and the image under the endpoint map of the set of such flow lines is by definition $A(p) \times D(p)$; moreover the closure of the codimemsion one stratum in $\overline{G(f)}_{\text{abs}}$ maps to $\bar{A}(p) \times \bar{D}(p)$.

Two types of error terms may arise in these formulas. First, there might be disagreement between the chosen triangulations of submanifolds of different abstract compmactifications that map to the same submanifold of Y. Second, in the formulas after (2.2), there are error terms consisting of chains supported in submanifolds of lower dimension than the degree of the chain. For example, in (2.4), if $p \in \mathrm{Crit}_i$, then the codimension 1 stratum of $\bar{\mathscr{D}}(p)_{\text{abs}}$ includes components corresponding to flow lines broken by critical points q of index less than $i-1$ (which are then mapped to $\mathscr{D}(q)$, which has dimension less than $i-1$, in Y). All of these errors are degenerate in the sense of Definition 2.2.

One can check the orientations in the formulas explicitly for a special metric in which the gradient flow near each critical point has the nice form $(-x^1\partial_1, \dots, -x^i\partial_i, x^{i+1}\partial_{i+1}, \dots, x^n\partial_n)$ for some local coordinates $(x^1, \dots, x^n)$. The orientations agree in the general case by continuity. □

Remark 2.4. (a) Instead of dealing with degenerate chains, it might be possible to use currents here. An apparently related use of currents in Morse theory appears in [10].

(b) One can see the standard isomorphism

$$H_i(C_*^{\text{Morse}}) = H_i(Y, Y_0) \tag{2.9}$$

using similar formalism. Namely, one can introduce a complex $C_*^{\text{gen}}(Y, Y_0)$ of "generic" chains in Y relative to Y_0 which intersect all ascending manifolds transversely. One can extend $\bar{\mathscr{F}}$ to such chains and define a map $\mathscr{A}^\dagger$ on them by analogy with $A^\dagger$. Then we observe that:

(i) Equation (2.4) shows that we have a chain map

$$\bar{\mathscr{D}} : C_*^{\text{Morse}} \to C_*^{\text{gen}}(Y, Y_0).$$

(ii) As in (2.7) we have $d\mathscr{A}^\dagger = \mathscr{A}^\dagger\partial$, so

$$\mathscr{A}^\dagger : C_*^{\text{gen}}(Y, Y_0) \to C_*^{\text{Morse}}$$

is a chain map.

(iii) By definition, $\mathscr{A}^\dagger\bar{\mathscr{D}}$ equals the identity at the chain level.

(iv) As in (2.3) we have (modulo degenerate chains as usual)

$$\partial\bar{\mathscr{F}} = -\bar{\mathscr{F}}\partial + 1 - \bar{f}_* - \bar{\mathscr{D}}\mathscr{A}^\dagger$$

on $C_*^{\text{gen}}(Y, Y_0)$, so $\bar{\mathscr{F}}$ is a chain homotopy between $\bar{\mathscr{D}}\mathscr{A}^\dagger$ and the identity.

(c) Note also that under the isomorphism (2.9), the connecting homomorphism $\delta : H_*(Y, Y_0) \to H_{*-1}(Y_0)$ is induced by $\bar{D}$.

(d) The Novikov isomorphism of Theorem 1.5 can be proved similarly to (b).

2.4. Homological assumptions

We will now clarify the homological assumption in our main theorem. Versions (a), (b), and (d) below are required for the statements of the main theorem to make sense, while version (c) will be needed in the proof in Section 3.

LEMMA 2.5. *The following are equivalent* (*the notation is defined in Section* 1.1):

(a) $C_*^{\mathrm{nov}} \otimes \mathbb{Q}((t))$ *is acyclic.*
(b) $C_*^{\mathrm{cell}}(\tilde{X}) \otimes \mathbb{Q}(t)$ *is acyclic.*
(c) *The map* $H_*(\Sigma; \mathbb{Q}) \to H_*(\tilde{X}; \mathbb{Q})$ *induced by the inclusion* $x \mapsto (x, 0)$ *is surjective.*
(d) $H_*(\tilde{X}; \mathbb{Q})$ *is a finite-dimensional* $\mathbb{Q}$*-vector space.*

Proof (a)$\Leftrightarrow$(b): This follows from the Novikov isomorphism (Theorem 1.5).

(b) $\Rightarrow$ (c): Assume (b), and let $\beta \in C_*^{\mathrm{cell}}(\tilde{X}; \mathbb{Q})$ be a cycle. Assumption (b) implies that the complexes $C_*^{\mathrm{cell}}(\tilde{X}) \otimes \mathbb{Q}((t))$ and $C_*^{\mathrm{cell}}(\tilde{X}) \otimes \mathbb{Q}((t^{-1}))$ are acyclic. So we can write $\beta = \partial v_- = \partial v_+$ where v_- and v_+ are locally finite chains in $\tilde{X}$ supported in $\tilde{f}^{-1}(-\infty, R)$ and $\tilde{f}^{-1}(-R, +\infty)$ respectively for some $R \in \mathbb{R}$. Without loss of generality, we can choose our cell decomposition of $\tilde{X}$ so that Σ is a subcomplex. Now let $\alpha := v_+ \cap \tilde{f}^{-1}(-R, 0] + v_- \cap \tilde{f}^{-1}[0, R)$. Then $\beta - \partial\alpha$ is a cycle supported in Σ. Since α is compactly supported, this means that β is equal in $H_*(\tilde{X}; \mathbb{Q})$ to a cycle in Σ.

(c) $\Rightarrow$ (d): This is clear since Σ is compact.

(d) $\Rightarrow$ (b): Assume (d). If $\alpha \in H_*(\tilde{X}; \mathbb{Q})$, then for large k the vectors $\alpha, t\alpha, \ldots, t^k\alpha$ must be linearly dependent, so α is annihilated by a nonzero polynomial in $P \in \mathbb{Z}[t]$. Since P is invertible in $\mathbb{Q}(t)$, the homology class α vanishes in $H_*(C_*^{\mathrm{cell}}(\tilde{X}) \otimes \mathbb{Q}(t))$ (cf. [22]). □

2.5. Algebraic facts about torsion

We will use the following lemma in Section 2.6 to compute T_{top} in terms of the homology of $\tilde{X}$. (It should be noted that more refined versions of topological torsion are not homotopy invariant and thus cannot be computed this way.)

If R is a ring, let $R^\times$ denote its group of units. If E is a finitely generated module over a ring R, the first *Fitting ideal* of E is generated by the determinants of the $n \times n$ minors of the matrix of relations for a presentation of E with n generators. This does not depend on the presentation. If greatest common divisors exist in R, let $\mathrm{ord}(E) \in R/R^\times$ denote the greatest common divisor of the elements in the first Fitting ideal. If R is an integral domain, let $Q(R)$ denote its field of fractions.

LEMMA 2.6 (Turaev [37, Section 2.1]). *Let R be a Noetherian UFD* (*e.g.* $\mathbb{Z}[t, t^{-1}]$ *or* $\mathbb{Z}((t))$). *Let* (C_*, ∂) *be a finite complex of finitely generated free R-modules. If* $C_* \otimes_R Q(R)$ *is acyclic, then its Reidemeister torsion is given by*

$$\tau(C_* \otimes_R Q(R)) = \prod_{i=0}^{m} (\mathrm{ord}\, H_i(C))^{(-1)^{i+1}} \in \frac{Q(R)}{R^\times}.$$

(*Here* $\tau(C_* \otimes_R Q(R))$ *is defined using volume forms induced by arbitrary bases for* C_* *over* R.)

Note also that if $\iota: Q(R) \hookrightarrow F$ is an inclusion into a field, then $C_* \otimes_R F$ is acyclic if and only if $C_* \otimes_R Q(R)$ is, and

$$\tau(C_* \otimes_R F) = \iota(\tau(C \otimes_R Q(R))) \tag{2.10}$$

modulo units in R.

2.6. Equivalence of Theorems 1.12 and 1.14

This follows from:

LEMMA 2.7. (a) *The leading coefficients of the left and right sides of Theorem* 1.12 *are equal, up to sign.*

(b) *Theorem* 1.14 *is the logarithmic derivative of Theorem* 1.12.

Proof. (a) The leading coefficient of $\zeta(f)$ is 1 by definition, so we need to check that the leading coefficients of T_{Morse} and T_{top} agree up to sign. By Eq. (2.10), if $\imath : \mathbb{Q}(t) \hookrightarrow \mathbb{Q}((t))$ is the natural inclusion, then

$$\imath(T_{\mathrm{top}}) = \tau(C_*^{\mathrm{cell}}(\tilde{X}) \otimes_{\mathbb{Z}[t,t^{-1}]} \mathbb{Q}((t))). \tag{2.11}$$

Applying Lemma 2.6 to the complexes C_*^{nov} and $C_*^{\mathrm{cell}}(\tilde{X}) \otimes_{\mathbb{Z}[t,t^{-1}]} \mathbb{Z}((t))$ over $\mathbb{Z}((t))$, and using the Novikov isomorphism (Theorem 1.5), we obtain

$$\tau(C_*^{\mathrm{cell}}(\tilde{X}) \otimes_{\mathbb{Z}[t,t^{-1}]} \mathbb{Q}((t))) = T_{\mathrm{Morse}} \tag{2.12}$$

up to multiplication by units in $\mathbb{Z}((t))$. By (2.11) and (2.12) we are done, since every unit in $\mathbb{Z}((t))$ has leading coefficient ± 1.

(b) We just need to prove Eq. (1.3) (cf. Notation 1.13). For this purpose, it will suffice to show that the order of the $\mathbb{Z}[t, t^{-1}]$-module $H_i(\tilde{X})$ is given by

$$\mathrm{ord}(H_i(\tilde{X})) = c_i \det(1 - tQ_i) \tag{2.13}$$

for some $c_i \in \mathbb{Z}$. We will then be done by Lemma 2.6. (Note that $\mathbb{Z}[t, t^{-1}]^\times = \{\pm t^k\}$, so the ambiguity in Lemma 2.6 is no worse than the ambiguity in the definition of T_{top}.)

To prove (2.13), let V denote the $\mathbb{Z}$-torsion part of $H_i(\tilde{X}; \mathbb{Z})$, i.e.

$$V := \{\beta \in H_i(\tilde{X}) \mid (\exists k \in \mathbb{Z}^+)\ k\beta = 0\}.$$

Then $H_i(\tilde{X})/V$ is free and by Lemma 2.5(d) we can choose a finite set $S \subset H_i(\tilde{X})$ which projects to a basis for $H_i(\tilde{X})/V$ over $\mathbb{Z}$. Since $\mathbb{Z}[t, t^{-1}]$ is Noetherian, we can choose a finite set T which generates V over $\mathbb{Z}[t, t^{-1}]$. So S and T generate $H_i(\tilde{X})$, and the matrix of relations is

$$\begin{array}{c@{}c} & \begin{array}{ccc} S & T & T \end{array} \\ \begin{array}{c} S \\ T \end{array} & \begin{pmatrix} 1 - tQ_i & 0 & 0 \\ ? & D & ? \end{pmatrix}. \end{array}$$

Here the columns represent relations. The only relations on $H_i(\tilde{X}; \mathbb{Q})$ are $1 - tQ_i$, and when we lift these to $H_i(\tilde{X})$ via our choice of S, there may be an additional component in T, which is the lower left block of the matrix. D is a diagonal matrix of integers asserting that the elements of T are torsion. The lower right block of the matrix expresses whatever additional relations the elements of T may satisfy amongst themselves.

Every minor of this matrix is divisible by $\det(1 - tQ_i)$, so $\det(1 - tQ_i)$ divides $\mathrm{ord}(H_i(\tilde{X}))$. On the other hand one of the minors is $\det(D)\det(1 - tQ_i)$, so $\mathrm{ord}(H_i(\tilde{X}))$ divides $\det(D)\det(1 - tQ_i)$. This proves (2.13). □

3. PROOF OF THEOREM 1.14

3.1. Outline of the proof

One of the classical proofs of the Lefschetz fixed point formula $\mathrm{Fix}(f) = \mathrm{Str}(H_*(f))$ for a diffeomorphism f on a manifold Σ proceeds as follows. We wish to calculate the intersection number of the graph $G(f)$ with the diagonal Diag in $\Sigma\times\Sigma$. We can replace the diagonal with a homologous cycle in $C_*^{\mathrm{sing}}(\Sigma)^{\otimes 2}$, and then we are reduced to intersection theory in Σ. Let $\{e_k\}$ be representatives of a basis for $H_*(\Sigma;\mathbb{Q})$, and let $\{e_k^*\}$ be the (Poincaré) dual basis. Then Diag is homologous to $\sum_k e_k\times e_k^*$, so the intersection number

$$G(f)\cdot\left(\mathrm{Diag}-\sum_k e_k\times e_k^*\right)=0. \tag{3.1}$$

This gives the Lefschetz formula, since $G(f)\cdot\mathrm{Diag}=\mathrm{Fix}(f)$ and

$$G(f)\cdot\sum_k e_k\times e_k^*=\mathrm{Str}(H_*(f)). \tag{3.2}$$

We wish to extend this reasoning to the partially defined maps f^k in Section 1.2. There are two difficulties. First, Eq. (3.1) is no longer true, because the graph $G(f^k)$ is not a cycle. Second, the r.h.s of Eq. (3.2) does not make sense for f^k, because f^k is not a chain map and does not induce a map in homology.

We can correct (3.1) by completing the graph to a cycle as follows. To start, it is useful to combine the graphs of f^k for different k into a single "generating function". Namely, define a chain $\mathbf{G}\in C_{n-1}^{\mathrm{sing}}(\Sigma\times\Sigma;\mathbb{Z}((t)))$ by

$$\mathbf{G}:=\sum_{k=1}^{\infty}t^k\overline{G(f^k)}$$

(where the graph $G(f^k)$ is compactified as in Section 2.3, using $Y=\{(x,\lambda)\in\tilde{X}\,|\,0\leqslant\lambda\leqslant k\}$).

We wish to find $Z\in C_*^{\mathrm{sing}}(\Sigma)^{\otimes 2}$ such that $\partial(\mathbf{G}-Z)=0$. Fortuitously, as we will see in Section 3.3, if $C_*^{\mathrm{nov}}\otimes\mathbb{Q}((t))$ is acyclic, then a canonical such Z exists, constructed by Morse theory. Equation (3.1) is now replaced by

$$(\mathbf{G}-Z)\cdot\left(\mathrm{Diag}-\sum_k e_k\times e_k^*\right)=0. \tag{3.3}$$

Theorem 1.14 then results from the following three calculations:

LEMMA 3.1. (a) $\mathbf{G}\cdot\mathrm{Diag}=\sum_{k=1}^{\infty}t^k\mathrm{Fix}(f^k)$.
(b) $Z\cdot\mathrm{Diag}=-t(d/dt)\log T_{\mathrm{Morse}}$.
(c) $(\mathbf{G}-Z)\cdot\sum_k e_k\times e_k^*-\mathrm{Str}((1-tQ)^{-1})\in\mathbb{Z}$.

Part (a) is clear, and part (b) will be proved by direct calculation in Section 3.4.

Part (c) is the appropriate analogue of (3.2) in our situation. To prove it, we first compute in Section 3.4 that

LEMMA 3.2. $(\mathbf{G}-Z)\cdot\sum_i e_i\times e_i^*=\mathrm{Str}(B\colon H_*(\Sigma;\mathbb{Q}((t)))\to H_*(\Sigma;\mathbb{Q}((t))))$.

Here B is a chain map, defined on generic chains in $C_*^{\mathrm{sing}}(\Sigma;\mathbb{Q}((t)))$, which is obtained (in Section 3.2) roughly by adding a Morse-theoretic correction term to $\sum_{k=1}^{\infty}t^k f_*^k$, under the assumption that $C_*^{\mathrm{nov}}\otimes\mathbb{Q}((t))$ is acyclic. The affinity of B with f^k, or more precisely with Q^k, is clarified by the following lemma:

LEMMA 3.3. *The diagram*

$$\begin{tikzcd} H_*(\Sigma;\ \mathbb{Q}((t))) \arrow[r, "B"] \arrow[d, "i_*"] & H_*(\Sigma;\ \mathbb{Q}((t))) \arrow[d, "i_*"] \\ H_*(\tilde{X};\ \mathbb{Q}((t))) \arrow[r, "tQ\,(1-tQ)^{-1}"] & H_*(\tilde{X};\ \mathbb{Q}((t))) \end{tikzcd}$$

commutes. (Here $\iota : \Sigma \hookrightarrow \tilde{X}$ is the inclusion sending $p \mapsto (p, 0)$.)

(The idea of the proof is as follows. Let $\gamma \in C_*(\Sigma; \mathbb{Q})$ be a cycle, and suppose the downward gradient flow takes γ around X k times without hitting any critical points. Then $\gamma \times \{k\}$ and $f^k(\gamma) \times \{0\}$ are homologous in $\tilde{X}$, because their difference is the boundary of the entire gradient flow between them. This means that $Q^k(\iota_*\gamma) = \iota_* f^k(\gamma)$. More generally, if the downward gradient flow of γ hits some critical points, then the gradient flow no longer gives a homology between $\gamma \times \{k\}$ and $f^k(\gamma) \times \{0\}$, because it has additional boundary components arising from the critical points. But the Morse-theoretic correction term in B is exactly what is needed to cancel these.)

Lemma 3.3 implies (with the help of Lemma 2.5(c), asserting that ι_* is surjective) that

$$\mathrm{Str}(B) = \mathrm{Str}(tQ(1 - tQ)^{-1}) + \mathrm{Str}(B \,|\, \mathrm{Ker}(\iota_*)).$$

To complete the proof of Lemma 3.1(c) and hence of Theorem 1.14, we need to show that $\mathrm{Str}(B|\mathrm{Ker}(\iota_*)) \in \mathbb{Z}$. We do this in Section 3.6 by describing $\mathrm{Ker}(\iota_*)$ in terms of Morse theory and directly evaluating B on it.

3.2. Some useful chain maps

We begin by introducing four chain maps $\mathbf{D}$, $\mathbf{A}$, $\mathbf{A}^\dagger$, and B, which will be needed for the calculations outlined above.

Let $C_*^{\mathrm{gen}+}(\Sigma)$ (resp. $C_*^{\mathrm{gen}-}(\Sigma)$) denote the subcomplex of (smooth) singular chains intersecting all ascending (resp. descending) manifolds in X transversely.

If $p \in \mathrm{Crit}$, let $\tilde{p} \in \tilde{X}$ denote the lift (p, λ) of p with $-1 < \lambda < 0$. Define

$$\mathbf{D}(p) := \sum_{k=0}^{\infty} t^k (\mathscr{D}(\tilde{p}) \cap (\Sigma \times \{-1-k\})) \in C_*^{\mathrm{gen}+}(\Sigma; \mathbb{Z}((t))).$$

Here we are identifying $\Sigma \times \{-k\}$ with Σ. Also we are implictly making the noncompact manifold $\mathscr{D}(\tilde{p}) \cap (\Sigma \times \{-k\})$ into a chain using the machinery of Section 2.3 (namely using the operator $\bar{D}$ on $Y = \{(x, \lambda) \in \tilde{X} \,|\, -k \leqslant \lambda \leqslant 0\}$). Intuitively, $\mathbf{D}(p)$ is the intersection in X of $\mathscr{D}(p)$ and Σ, with the components weighted by powers of t. Now extend linearly over $\mathbb{Z}((t))$ to obtain a map

$$\mathbf{D} : C_*^{\mathrm{nov}} \to C_{*-1}^{\mathrm{gen}+}(\Sigma; \mathbb{Z}((t))).$$

We define a $\mathbb{Z}((t))$-linear map

$$\mathbf{A} : C_*^{\mathrm{nov}} \to C_{n-1-*}^{\mathrm{gen}-}(\Sigma; \mathbb{Z}((t)))$$

similarly, using ascending manifolds; if $p \in \mathrm{Crit}$, then

$$\mathbf{A}(p) := \sum_{k=0}^{\infty} t^k (\mathscr{A}(\tilde{p}) \cap (\Sigma \times \{k\})).$$

The maps $\mathbf{D}$ and $\mathbf{A}$ arise naturally as follows:

LEMMA 3.4. *The boundary of the graph is given by*

$$\partial \mathbf{G} = t \sum_{i=1}^{n-1} \sum_{p \in \mathrm{Crit}_i} (-1)^{(n-1)(i-1)} \mathbf{A}(p) \times \mathbf{D}(p).$$

Proof. Applying (2.2) to $Y = \{(x,\lambda) \in \tilde{X} \mid -k \leqslant \lambda \leqslant 0\}$, we obtain

$$\begin{aligned}\partial \overline{G(f^k)} &= \sum_{i=1}^{n-1} \sum_{p \in \mathrm{Crit}_i} (-1)^{(n-1)(i-1)} \sum_{j=0}^{k-1} \bar{A}(t^j \tilde{p}) \times \bar{D}(t^j \tilde{p}) \\ &= \sum_{i=1}^{n-1} \sum_{p \in \mathrm{Crit}_i} (-1)^{(n-1)(i-1)} \sum_{j=0}^{k-1} (\mathbf{A}(p))^j \times (\mathbf{D}(p))^{k-j-1}\end{aligned}$$

where $(\mathbf{A}(p))^j$ denotes the t^j term of $\mathbf{A}(p)$, etc. This proves the t^k term of the lemma. □

In the rest of the paper, we will omit the details when applying formulas from Lemma 2.3 in a straightforward way as above.

We next define a map

$$\mathbf{A}^\dagger : C_*^{\mathrm{gen}+}(\Sigma; \mathbb{Z}((t))) \to C_*^{\mathrm{nov}}$$

by requiring that for $\alpha \in C_*^{\mathrm{gen}+}(\Sigma; \mathbb{Z}((t)))$ and $x \in C_*^{\mathrm{nov}}$,

$$\alpha \cdot \mathbf{A}(x) = \langle \mathbf{A}^\dagger \alpha, x \rangle.$$

The maps $\mathbf{D}$, $\mathbf{A}$, and $\mathbf{A}^\dagger$ are all chain maps:

LEMMA 3.5. (a) $\partial \mathbf{D} = -\mathbf{Dd}$,
(b) $\partial \mathbf{A} = (-1)^{i-1} \mathbf{Ad}^*$ *on* C_i^{nov},
(c) $\mathbf{dA}^\dagger = \mathbf{A}^\dagger \partial$.

Proof. These equations follow from (2.5), (2.6) and (2.7), respectively. □

By the machinery of Section 2.3 (applied to $Y = \{(x, \lambda) \in \tilde{X} \mid 0 \leqslant \lambda \leqslant k\}$), we have a map $\overline{f_*^k} : C_*^{\mathrm{gen}+}(\Sigma) \to C_*^{\mathrm{gen}+}(\Sigma)$. We now define

$$\mathbf{f} := \sum_{k=1}^{\infty} t^k \overline{f_*^k} : C_*^{\mathrm{gen}+}(\Sigma; \mathbb{Z}((t))) \to C_*^{\mathrm{gen}+}(\Sigma; \mathbb{Z}((t))).$$

Because f^k is only partially defined, $\mathbf{f}$ is not a chain map; by (2.8) we have

$$\partial \mathbf{f} = \mathbf{f} \partial + t \mathbf{DA}^\dagger. \tag{3.4}$$

However we can add a correction term to $\mathbf{f}$ to obtain a chain map. We are assuming that $C_*^{\mathrm{nov}} \otimes \mathbb{Q}((t))$ is acyclic, so Δ is invertible over $\mathbb{Q}((t))$. We then define

$$B := \mathbf{f} + t \mathbf{Dd}^* \Delta^{-1} \mathbf{A}^\dagger. \tag{3.5}$$

(This will arise naturally in the calculations in Section 3.4.)

LEMMA 3.6. $\partial B = B \partial$.

Proof. By (3.4) we have

$$\partial B = \mathbf{f} \partial + t \mathbf{DA}^\dagger + t \partial \mathbf{Dd}^* \Delta^{-1} \mathbf{A}^\dagger. \tag{3.6}$$

Using the identity $\mathbf{dd}^*\Delta^{-1} + \mathbf{d}^*\Delta^{-1}\mathbf{d} = 1$ and Lemma 3.5(a), (c) gives

$$\begin{aligned} t\mathbf{DA}^\dagger &= t\mathbf{D}(\mathbf{dd}^*\Delta^{-1} + \mathbf{d}^*\Delta^{-1}\mathbf{d})\mathbf{A}^\dagger \\ &= -t\partial\mathbf{Dd}^*\Delta^{-1}\mathbf{A}^\dagger + t\mathbf{Dd}^*\Delta^{-1}\mathbf{A}^\dagger\partial. \end{aligned}$$

Substituting this into (3.6) gives $\partial B = B\partial$. □

3.3. Closing off the boundary of the graph

We will now find $Z \in C_*^{\mathrm{sing}}(\Sigma; \mathbb{Q}((t)))^{\otimes 2}$ with $\partial Z = \partial\mathbf{G}$. Let P be the composition

$$\mathrm{Hom}(C_i^{\mathrm{nov}}, C_j^{\mathrm{nov}}) \xrightarrow{\rho} C_j^{\mathrm{nov}} \otimes C_i^{\mathrm{nov}} \xrightarrow{\mathbf{A}\otimes\mathbf{D}} C_{n-1-j}^{\mathrm{gen}-}(\Sigma; \mathbb{Z}((t))) \otimes C_{i-1}^{\mathrm{gen}+}(\Sigma; \mathbb{Z}((t))).$$

Here ρ is the canonical isomorphism given by the inner product $\langle\ ,\ \rangle$. Our ansatz will be

$$Z = P(W)$$

for some $W \in \mathrm{Hom}(C_*^{\mathrm{nov}}, C_{*-1}^{\mathrm{nov}})$.

LEMMA 3.7. *Let* $W = \sum_{i=1}^n W_i$ *with* $W_i \in \mathrm{Hom}(C_i^{\mathrm{nov}}, C_{i-1}^{\mathrm{nov}})$. *Then*

$$\partial P(W) = \sum_{i=1}^{n-1} P((-1)^i\mathbf{d}^*W_i + (-1)^{n-i}W_{i+1}\mathbf{d}^*).$$

Proof. By Lemma 3.5(a) and (b) we have

$$\partial P(W) = \sum_{i=1}^{n} (\mathbf{A}\otimes\mathbf{D})((-1)^i\mathbf{d}^*\otimes 1 + (-1)^{n-i+1}1\otimes\mathbf{d})\,\rho(W_i).$$

(A factor of $(-1)^{n-i}$ arises because $\partial(a\times b) = \partial a\times b + (-1)^{\dim(a)}a\times\partial b$.) Now use the facts $(\mathbf{d}^*\otimes 1)\rho(W_i) = \rho(\mathbf{d}^*W_i)$ and $(1\otimes\mathbf{d})\rho(W_i) = \rho(W_i\mathbf{d}^*)$. □

In this notation, Lemma 3.4 says that

$$\partial\mathbf{G} = t\sum_{i=1}^{n-1} (-1)^{(n-1)(i-1)}P(1 : C_i^{\mathrm{nov}} \to C_i^{\mathrm{nov}}).$$

So by Lemma 3.7, $\partial Z = \partial\mathbf{G}$ if and only if

$$(-1)^i\mathbf{d}^*W_i + (-1)^{n-i}W_{i+1}\mathbf{d}^* = (-1)^{(n-1)(i-1)}t$$

on C_i^{nov} for $0 < i < n$. Thanks to our standing assumption that $C_*^{\mathrm{nov}} \otimes \mathbb{Q}((t))$ is acyclic, such a W exists, and we will make the natural choice

$$W_i := (-1)^{ni+n+1}t\Delta^{-1}\mathbf{d}.$$

3.4. Calculating intersection numbers

We will now prove Lemmas 3.1(b) and 3.2. For these calculations, it is convenient to choose a basis $\{x_{ij}\}$ for $\mathrm{Ker}(\mathbf{d}^*|C_i^{\mathrm{nov}})$ with $\|x_{ij}\| = 1$ and $\Delta x_{ij} = \lambda_{ij}x_{ij}$. (To find such a basis we may have to extend to coefficients in the algebraically closed field $\mathbb{C}((t))$, which causes no problems.) In this notation we have

$$Z = \sum_{i=1}^{n} (-1)^{ni+n+1}\sum_j \lambda_{ij}^{-1}\mathbf{A}(\mathbf{d}x_{ij}) \times \mathbf{D}(x_{ij}) \tag{3.7}$$

(because the x_{ij}'s, together with an orthonormal basis of Ker($\mathbf{d}$), constitute an orthonormal basis of C_*^{nov}). We also have

$$T_{\mathrm{Morse}} = \prod_{i=1}^{n} \left(\prod_j \sqrt{\lambda_{ij}} \right)^{(-1)^i} \tag{3.8}$$

as one can see by choosing $\omega_i = \bigwedge_j x_{ij}$ in (1.1).

LEMMA 3.8. *Let* $x \in C_*^{\mathrm{nov}}$ *and* $y \in C_{*-1}^{\mathrm{nov}}$. *Then*

$$\mathbf{D}x \cdot \mathbf{A}y = -\frac{d}{dt}\langle \mathbf{d}x, y \rangle + \left\langle \mathbf{d}\left(\frac{d}{dt}x\right), y \right\rangle + \left\langle \mathbf{d}x, \frac{d}{dt}y \right\rangle.$$

Proof. If $x, y \in \mathrm{Crit}$, then the two rightmost terms are zero, and the formula follows directly from the definitions. The general case follows by expanding x and y in powers of t. □

Proof of Lemma 3.1(b). If α, β are two chains of complementary dimension a,b then $(\alpha \times \beta) \cdot \mathrm{Diag} = (-1)^{(a+1)b} \beta \cdot \alpha$. So by (3.7) we have

$$Z \cdot \mathrm{Diag} = t \sum_{i,j} (-1)^i \lambda_{ij}^{-1} \mathbf{D}(x_{ij}) \cdot \mathbf{A}(dx_{ij}). \tag{3.9}$$

Lemma 3.8 gives

$$\mathbf{D}x \cdot \mathbf{A}dx = -\frac{d}{dt}\|\mathbf{d}x\|^2 + \left\langle \mathbf{d}\left(\frac{d}{dt}x\right), dx \right\rangle + \left\langle dx, \frac{d}{dt}\mathbf{d}x \right\rangle.$$

If $x = x_{ij}$, then the middle term on the right vanishes:

$$\begin{aligned} \left\langle \mathbf{d}\left(\frac{d}{dt}x_{ij}\right), \mathbf{d}x_{ij} \right\rangle &= \left\langle \frac{d}{dt}x_{ij}, \mathbf{d}^*\mathbf{d}x_{ij} \right\rangle \\ &= \lambda_{ij}\left\langle \frac{d}{dt}x_{ij}, x_{ij} \right\rangle \\ &= \frac{\lambda_{ij}}{2}\frac{d}{dt}\|x_{ij}\|^2 \\ &= 0. \end{aligned}$$

Thus

$$\mathbf{D}x_{ij} \cdot \mathbf{A}\mathbf{d}x_{ij} = -\frac{1}{2}\frac{d}{dt}\|\mathbf{d}x_{ij}\|^2 = -\frac{1}{2}\frac{d}{dt}\lambda_{ij}.$$

Substituting this into (3.9) and comparing with (3.8) proves Lemma 3.1(b). □

Proof of Lemma 3.2. By definition we have

$$\mathbf{G} \cdot \sum_k e_k \times e_k^* = \sum_k (-1)^{\dim(e_k)} \mathbf{f}(e_k) \cdot e_k^*. \tag{3.10}$$

By (3.7),

$$Z \cdot \sum_k e_k \times e_k^* = t \sum_{i,j,k} (-1)^i \lambda_{ij}^{-1} (e_k \cdot \mathbf{A}\mathbf{d}x_{ij})(\mathbf{D}x_{ij} \cdot e_k^*). \tag{3.11}$$

We rewrite

$$\begin{aligned} e_k \cdot \mathbf{A}\mathbf{d}x_{ij} &= \langle \mathbf{A}^\dagger e_k, \mathbf{d}x_{ij} \rangle \\ &= \langle \mathbf{d}\mathbf{d}^* \Delta^{-1} \mathbf{A}^\dagger e_k, \mathbf{d}x_{ij} \rangle \\ &= \lambda_{ij} \langle \mathbf{d}^* \Delta^{-1} \mathbf{A}^\dagger e_k, x_{ij} \rangle. \end{aligned}$$

(We have $\mathbf{A}^\dagger e_k = \mathbf{d}\mathbf{d}^* \Delta^{-1} \mathbf{A}^\dagger e_k$ above because $\mathbf{d}\mathbf{A}^\dagger e_k = 0$ by Lemma 3.5(c).) Putting this into (3.11) gives

$$Z \cdot \sum_k e_k \times e_k^* = t \sum_k (-1)^{\dim(e_k)+1} \mathbf{D}\mathbf{d}^* \Delta^{-1} \mathbf{A}^\dagger e_k \cdot e_k^*.$$

Subtracting this from (3.10) gives Lemma 3.2. □

3.5. The action of B on homology

We will now prove Lemma 3.3. We begin by defining $\mathbb{Q}$-linear maps

$$\begin{aligned} \mathscr{F}_k &: C_*^{\mathrm{gen}+}(\Sigma; \mathbb{Q}) \to C_{*+1}^{\mathrm{sing}}(\tilde{X}; \mathbb{Q}) \\ \mathscr{D}_k &: C_*^{\mathrm{nov}} \otimes \mathbb{Q}((t)) \to C_*^{\mathrm{sing}}(\tilde{X}; \mathbb{Q}) \end{aligned}$$

for $k = 1, 2, \ldots$ as follows. If $\gamma \in C_*^{\mathrm{gen}+}(\Sigma; \mathbb{Q})$ then we take $Y = \{(x, \lambda) \in \tilde{X} \mid -k \leqslant \lambda \leqslant 0\}$ in Section 2.3 and define $\mathscr{F}_k(\gamma) := \bar{\mathscr{F}}(\gamma \times \{0\})$. If $p \in \mathrm{Crit}$ and $l \in \mathbb{Z}$, then we take $Y = \{(x, \lambda) \mid -k \leqslant \lambda \leqslant -l\}$ and define $\mathscr{D}_k(t^l p) := \bar{\mathscr{D}}(t^l \tilde{p})$. (Recall that $t^l \tilde{p} \in \tilde{X}$ is the lift (p, λ) with $\lambda \in (-l-1, -l)$.) If $l \geqslant k$ then $\mathscr{D}_k(t^l p) = 0$. We extend $\mathscr{D}_k$ linearly over $\mathbb{Q}$ (it is not $\mathbb{Q}((t))$-linear).

LEMMA 3.9. (a) *If* $\gamma \in C_*^{\mathrm{gen}+}(\Sigma; \mathbb{Q})$ *is a cycle and* $k > 0$, *then*

$$\partial \mathscr{F}_k(\gamma) = \gamma \times \{0\} - f^k(\gamma) \times \{-k\} - \mathscr{D}_k \mathbf{A}^\dagger \gamma. \tag{3.12}$$

(b) *If* $x \in C_*^{\mathrm{nov}}$ *then*

$$\partial \mathscr{D}_k(x) = \mathscr{D}_k \mathbf{d}x - (\mathbf{D}x)^{k-1} \times \{-k\}. \tag{3.13}$$

(Here $(\mathbf{D}x)^{k-1}$ denotes the coefficient of t^{k-1} in $\mathbf{D}x$.)

Proof. (a) and (b) follow from equations (2.3) and (2.4), respectively. □

Proof of Lemma 3.3. Let $\gamma \in C_*^{\mathrm{gen}+}(\Sigma; \mathbb{Q})$ be a cycle; we need to show that

$$tQ(1 - tQ)^{-1} \imath_* \gamma = \imath_* B\gamma.$$

Equating coefficients of t^k, this is equivalent to:

(a) If $k > 0$, then $(B\gamma)^k \times \{-k\}$ is homologous to $\gamma \times \{0\}$ in $C_*^{\mathrm{sing}}(\tilde{X}; \mathbb{Q})$ (where $(B\gamma)^k$ denotes the t^k coefficient of $B\gamma$).

(b) If $k \leqslant 0$, then $(B\gamma)^k \times \{-k\}$ is nullhomologous in $C_*^{\mathrm{sing}}(\tilde{X}; \mathbb{Q})$.

Suppose $k > 0$. By Lemma 3.5(c), $\mathbf{d}\mathbf{A}^\dagger \gamma = 0$, so we can write $\mathbf{A}^\dagger \gamma = \mathbf{d}(\mathbf{d}^* \Delta^{-1} \mathbf{A}^\dagger \gamma)$. Thus, putting $x = \mathbf{d}^* \Delta^{-1} \mathbf{A}^\dagger \gamma$ into (3.13) gives

$$\partial \mathscr{D}_k \mathbf{d}^* \Delta^{-1} \mathbf{A}^\dagger \gamma = \mathscr{D}_k \mathbf{A}^\dagger \gamma - (\mathbf{D}\mathbf{d}^* \Delta^{-1} \mathbf{A}^\dagger \gamma)^{k-1} \times \{-k\}. \tag{3.14}$$

Adding this to (3.12) gives

$$\begin{aligned}\partial(\text{something}) &= \gamma \times \{0\} - f^k(\gamma) \times \{-k\} - (\mathbf{D}\mathbf{d}^*\Delta^{-1}\mathbf{A}^\dagger\gamma)^{k-1} \times \{-k\} \\ &= \gamma \times \{0\} - (B\gamma)^k \times \{-k\}.\end{aligned}$$

This proves (a).

If $k \leqslant 0$, then $\mathscr{D}_k\mathbf{A}^\dagger\gamma = 0$, because $\mathbf{A}^\dagger\gamma$ contains no negative powers of t. Then (3.14) gives

$$\partial\mathscr{D}_k\mathbf{d}^*\Delta^{-1}\mathbf{A}^\dagger\gamma = -(B\gamma)^k \times \{-k\}.$$

This implies (b). □

3.6. Completing the proof of the main theorem

We will now prove that $\mathrm{Str}(B|\mathrm{Ker}(\imath_*)) \in \mathbb{Z}$ (Lemma 3.12), which will complete the proof of Theorem 1.14, as explained in Section 3.1.

We first need to describe the kernel of $\imath_* : H_*(\Sigma; \mathbb{Q}) \to H_*(\tilde{X}; \mathbb{Q})$ in terms of Morse theory. Let $V_- \subset H_*(\Sigma; \mathbb{Q})$ denote the subspace consisting of cycles of the form $(\mathbf{D}x)^0$, where $x \in C_*^{\mathrm{nov}}$ and $(\mathbf{d}x)^{\leqslant 0} = 0$. (Here $(\mathbf{D}x)^0$ denotes the constant coefficient of $\mathbf{D}x$, and $(\mathbf{d}x)^{\leqslant 0}$ the portion of $\mathbf{d}x$ containing nonpositive powers of t.) Similarly, let

$$V_+ := \{(\mathbf{A}y)^0 | (\mathbf{d}^*y)^{\leqslant 0} = 0\}.$$

Lemma 3.10 (a) $\mathrm{Ker}(\imath_* : H_*(\Sigma; \mathbb{Q}) \to H_*(\tilde{X}; \mathbb{Q})) = \mathrm{span}(V_+, V_-)$.
(b) *If* $H_*(C_*^{\mathrm{nov}} \otimes \mathbb{Q}((t))) = 0$, *then* $V_+ \cap V_- = \{0\}$.

Proof. Define

$$\begin{aligned}\tilde{X}^+ &:= \{(x, \lambda) \in \tilde{X} | \lambda \geqslant 0\}\\ \tilde{X}^- &:= \{(x, \lambda) \in \tilde{X} | \lambda \leqslant 0\}.\end{aligned}$$

The relative homology exact sequence

$$H_{*+1}(\tilde{X}^-, \Sigma) \xrightarrow{\delta} H_*(\Sigma) \to H_*(\tilde{X}^-)$$

and Remark 2.4(c) imply that

$$\mathrm{Ker}(H_*(\Sigma) \to H_*(\tilde{X}^-)) = V_+.$$

(Here we are identifying Σ with $\Sigma \times \{0\} \subset \tilde{X}$, and all homology is with rational coefficients.) Furthermore, the kernel of $H_*(\tilde{X}^-) \to H_*(\tilde{X})$ is given by the image of the connecting homomorphism δ in the exact sequence

$$H_{*+1}(\tilde{X}, \tilde{X}^-) \xrightarrow{\delta} H_*(\tilde{X}^-) \to H_*(\tilde{X}).$$

By excision, $H_{*+1}(\tilde{X}, \tilde{X}^-) = H_*(\tilde{X}^+, \Sigma)$, and δ sends this to V_-. This proves (a).

To prove (b), suppose $u \in V_+ \cap V_-$. Write $u = (\mathbf{A}x)^0 = (\mathbf{D}y)^0$. Let $v \in H_{*+1}(\tilde{X})$ be the cycle obtained by gluing together the upward gradient flow of x (up to Σ) and the downward gradient flow of y (down to Σ), compactified into chains as in Section 2.3. Note that u is the image of v under the connecting homomorphism δ in the Mayer–Vietoris sequence

$$H_{k+1}(\tilde{X}^-) \oplus H_{k+1}(\tilde{X}^+) \to H_{k+1}(\tilde{X}) \xrightarrow{\delta} H_k(\Sigma).$$

By Lemma 2.5(c), v is in the image of $\imath_* : H_{k+1}(\Sigma) \to H_{k+1}(\tilde{X})$. But $\delta\imath_* = 0$, so $u = 0$. □

We now compute $B|\mathrm{Ker}(\imath_*)$. Let R be the $\mathbb{Q}$-linear operator that sends t^k to t^{-k}.

LEMMA 3.11. (a) *If* $(\mathbf{d}x)^{\leqslant 0} = 0$ *then* $B((\mathbf{D}x)^0) = -(\mathbf{D}x)^{\leqslant 0}$ *in* $H_*(\Sigma; \mathbb{Q})$.
(b) *If* $(\mathbf{d}^*y)^{\leqslant 0} = 0$ *then* $B((\mathbf{A}y)^0) = R((\mathbf{A}y)^{<0})$ *in* $H_*(\Sigma; \mathbb{Q})$.

Proof. Without loss of generality, $x^{>0} = 0$. Then

$$\mathbf{f}((\mathbf{D}x)^0) = (\mathbf{D}x)^{>0}. \tag{3.15}$$

From the definitions of $\mathbf{A}^\dagger$ and $\mathbf{d}$, we have

$$\mathbf{A}^\dagger((\mathbf{D}x)^0) = -t^{-1}(\mathbf{d}x)^{>0} = -t^{-1}\mathbf{d}x.$$

Then

$$\begin{aligned} t\mathbf{D}\mathbf{d}^*\Delta^{-1}\mathbf{A}^\dagger((\mathbf{D}x)^0) &= -\mathbf{D}\mathbf{d}^*\Delta^{-1}\mathbf{d}x = -\mathbf{D}x + \mathbf{D}\mathbf{d}\mathbf{d}^*\Delta^{-1}x \\ &= -\mathbf{D}x + (\text{nullhomologous cycle}) \end{aligned} \tag{3.16}$$

by Lemma 3.5(a). Putting (3.15) and (3.16) into the definition of B (Eq. (3.5)) proves (a).

To prove (b), let $\gamma \in C^{\text{gen}+}(\Sigma)$ be a perturbation of $(\mathbf{D}y)^0$. (We need to perturb because $(\mathbf{D}y)^0 \notin C^{\text{gen}+}(\Sigma)$, so we cannot apply B to $(\mathbf{D}y)^0$ at the chain level.) We create γ by replacing $(\mathbf{A}y)^0$ with the intersection of $\Sigma \times \{0\}$ and the ascending slices of small spheres linking the descending manifolds of the critical points in y at generic points. For any positive integer k, we can choose these spheres to be small enough that

$$\mathbf{f}(\gamma) = R((\mathbf{A}y)^{<0}) + (\text{nullhomologous cycle}) + O(t^k).$$

Similarly

$$\mathbf{A}^\dagger(\gamma) = \pm R((\mathbf{d}^*y)^{\leqslant 0}) + O(t^k) = O(t^k).$$

Now apply the definition of B. □

LEMMA 3.12. $\operatorname{Str}(B|\operatorname{Ker}(\imath_*)) \in \mathbb{Z}$.

Proof. By Lemma 3.10,

$$\operatorname{Str}(B|\operatorname{Ker}(\imath_*)) = \operatorname{Str}(B|V_+) + \operatorname{Str}(B|V_-). \tag{3.17}$$

(We are assuming that $H_*(C_*^{\text{nov}} \otimes \mathbb{Q}((t))) = 0$, which is necessary for B to be defined.) Now we use a trick. Equation (3.3) and the combination of Lemmas 3.1(b), 3.2, 3.3 and 2.5(c) imply that

$$\operatorname{Str}(B|\operatorname{Ker}(\imath_*)) = \mathbf{G}\cdot\operatorname{Diag} + t\frac{d}{dt}\log T_{\text{Morse}} - \operatorname{Str}(tQ(1-tQ)^{-1}). \tag{3.18}$$

It follows from (3.18) that $\operatorname{Str}(B|\operatorname{Ker}(\imath_*))$ contains no negative powers of t, and moreover the constant coefficient of $\operatorname{Str}(B|\operatorname{Ker}(\imath_*))$ is an integer (namely the smallest exponent in T_{Morse}). It then follows from (3.17) and Lemma 3.11 that $\operatorname{Str}(B|V_-) \in \mathbb{Z}$, since all the negative degree terms must vanish. We also see from Lemma 3.11 that the coefficients of $\operatorname{Str}(B|V_+)$ are exactly minus what the nonconstant coefficients of $\operatorname{Str}(B|V_-)$ would be if we inverted the Morse function (with appropriate new orientations on the ascending and descending manifolds). So $\operatorname{Str}(B|V_+) = 0$. Thus $\operatorname{Str}(B|\operatorname{Ker}(\imath_*)) \in \mathbb{Z}$ by (3.17). □

4. SEIBERG–WITTEN INVARIANTS OF 3-MANIFOLDS

From now on we assume that X is a closed oriented 3-manifold with $b^1 > 0$. In Section 4.1, we define the invariant I which is an analogue of the Gromov invariant for X, and in Section 4.2 we explain the motivation for our conjecture that I equals the Seiberg–Witten invariant. In Section 4.3 we apply Theorem 1.12 to compute an "averaged" version of the invariant I, and in Section 4.4 we deduce that our conjecture implies the Meng–Taubes formula (for closed manifolds, modulo signs) relating the Seiberg–Witten invariant to Milnor torsion.

4.1. An analogue of the Gromov invariant

Let $\phi : X \to S^1$ be a generic Morse function, and let $\Sigma = \phi^{-1}(0)$ as usual. Let $\eta = -d\phi$. Assume that ϕ has no index 0 or 3 critical points. (This assumption is not really necessary to define our invariant I but will simplify the algebra, and is justified in the Seiberg–Witten context of Section 4.2, where we will take ϕ to be harmonic.) In particular this implies that the homology class of η is nontrivial.

4.1.1. Parametrization of the invariant

Definition 4.1. Within the relative homology $H_1(X, \eta^{-1}(0))$, we consider the subset

$$H_1(X, \eta) := \{\gamma \in H_1(X, \eta^{-1}(0)) | \partial\gamma = [\eta^{-1}(0)]\}.$$

(Here $\eta^{-1}(0)$ is oriented in the usual way, which means via minus the parity of the index of critical points of ϕ.) The set $H_1(X, \eta)$ is naturally an affine space modelled on $H_1(X)$.

Our analogue of the Gromov invariant will initially be defined on the set $H_1(X, \eta)$. To state a conjectural analogue of the "SW = Gr" theorem, we need a way of identifying $H_1(X, \eta)$ with the set of spin-c structures, on which the Seiberg–Witten invariant is defined.

Definition 4.2. A *spin-c structure* on our 3-manifold X is a $U(2)$ vector bundle $W \to X$ together with a *Clifford multiplication* map $\mathrm{cl} : T^*X \to \mathrm{End}(W)$, satisfying the axioms

$$\mathrm{cl}(v)^2 = -|v|^2$$
$$\mathrm{cl}(e_1)\,\mathrm{cl}(e_2)\,\mathrm{cl}(e_3) = -1 : W_p \to W_p$$

where $\{e_1, e_2, e_3\}$ is an oriented orthonormal basis for T^*_pX. The set $\mathrm{Spin}^c(X)$ of spin-c structures on X is an affine space modelled on $H^2(X; \mathbb{Z}) = H_1(X)$; a cohomology class $\alpha \in H^2(X; \mathbb{Z})$ acts by $\alpha \cdot W = W \otimes L$, where L is the complex line bundle with $c_1(L) = \alpha$.

Given a spin-c structure W, the endomorphism $\mathrm{cl}(\eta/|\eta|)$, defined on $X \backslash \mathrm{Crit}$, has square -1 and splits W into $\pm i$ eigenspaces, which we denote by $E_\pm$.

LEMMA 4.3. *There is a well-defined $H_1(X)$-equivariant isomorphism*

$$j_\eta : \mathrm{Spin}^c(X) \to H_1(X, \eta)$$

which sends $W \in \mathrm{Spin}^c(X)$ to the Poincaré-Lefschetz dual of

$$c_1(E_-) \in H^2(X \backslash \eta^{-1}(0)).$$

Proof. To show that j_η is well defined, i.e. that $\mathrm{PD}(c_1(E_-)) \in H_1(X, \eta)$, we must check that if p is a zero of η with sign ± 1 and S is a small 2-sphere around p, then the line bundle $E_-|_S$ has degree ∓ 1. Choose local coordinates x_1, x_2, x_3 near p. The isomorphism class of $E_-|_S$ depends only on the homotopy class of the map $\eta/|\eta| : S \to S^2$, and not on the metric. So we can assume the metric is Euclidean in these coordinates and consider the special case

$$\eta = x_1 dx_1 + x_2 dx_2 \pm x_3 dx_3.$$

We can trivialize the spin bundle W in this neighborhood so that

$$\mathrm{cl}(dx_1) = \begin{pmatrix} -i & \\ & i \end{pmatrix}, \quad \mathrm{cl}(dx_2) = \begin{pmatrix} & -1 \\ 1 & \end{pmatrix}, \quad \mathrm{cl}(dx_3) = \begin{pmatrix} & i \\ i & \end{pmatrix}.$$

We can then define a section s of E_- by

$$s = \begin{pmatrix} -x_2 \pm \mathrm{i}x_3 \\ \mathrm{i}x_1 - \mathrm{i}|x| \end{pmatrix}.$$

We see that $s^{-1}(0)$ is the positive x_1-axis, and the sign of the zero on S is ∓ 1.

This shows that j_η is well defined. The map j_η is clearly $H_1(X)$-equivariant, and it follows that it is an isomorphism. $\square$

4.1.2. The Novikov ring. The definition of our invariant involves some algebraic counting which takes place in the following ring.

Definition 4.4. If G is an abelian group and $N: G \to \mathbb{R}$ is a homomorphism, then the *Novikov ring* $\mathrm{Nov}(G; N)$ is defined to be the ring of formal sums $\sum_{g \in G} a_g \cdot g$ with $a_g \in \mathbb{Z}$ (equivalently, functions $G \to \mathbb{Z}$ sending $g \mapsto a_g$) such that for each $R \in \mathbb{R}$, there are only finitely many $g \in G$ such that $a_g \neq 0$ and $N(g) < R$. Multiplication in $\mathrm{Nov}(G; N)$ is given by the convolution product

$$\left(\sum_{g \in G} a_g \cdot g\right)\left(\sum_{g \in G} b_g \cdot G\right) := \sum_{g \in G}\left(\sum_{h \in G} a_h b_{g-h}\right) \cdot g.$$

The Novikov ring is an enlargement of the group ring $\mathbb{Z}[G]$. (See e.g. [11] for more about Novikov rings.)

We are interested in the case when G is the relative homology $H_1(X, \eta^{-1}(0))$ and N is intersection number with Σ, and we denote the corresponding Novikov ring by Λ.

4.1.3. Refined zeta function

Definition 4.5. Let $\mathcal{O}$ denote the set of (nonconstant) closed orbits of the flow dual to η. We define (cf. [6, 28])

$$\hat{\zeta} := \exp\left(\sum_{\gamma \in \mathcal{O}} \frac{(-1)^{\varepsilon(\gamma)}}{p(\gamma)}[\gamma]\right) \in \Lambda.$$

Here $[\gamma] \in H_1(X)$ denotes the homology class of γ and $p(\gamma)$ the *period* of γ (i.e. the largest integer k such that the closed orbit $\gamma: S^1 \to X$ factors through a k-fold covering $S^1 \to S^1$). The sign $(-1)^{\varepsilon(\gamma)}$ is defined by the same local information as the Lefschetz sign in Section 1.2; it is the sign of $\det(1 - df_x)$, where $x \in X$ is a point on the closed orbit and f is the (pth) return map on the hyperplane in T_xX normal to γ. A compactness argument shows that $\hat{\zeta}$ is a well defined element of the Novikov ring Λ.

Remark 4.6. This definition makes sense in n dimensions, and for motivation elsewhere, we note that there is a product formula (cf. [7, 14])

$$\hat{\zeta} = \prod_{\gamma \in \mathcal{I}} (1 - (-1)^{i_-}[\gamma])^{-(-1)^{i_0}}. \tag{4.1}$$

Here $\mathcal{I}$ denotes the set of irreducible (i.e. period 1) closed orbits; $i_-(\gamma)$ and $i_0(\gamma)$ denote the numbers of eigenvalues of the return map that are real and in the intervals $(-\infty, -1)$ and

$(-1, 1)$ respectively. This formula is easily verified by taking the formal logarithm of both sides.

4.1.4. Definition of the invariant I. Let $\Lambda\langle \mathrm{Crit}_i \rangle$ denote the free Λ-module generated by the set Crit_i. Define a map

$$\hat{\mathbf{d}} : \Lambda\langle \mathrm{Crit}_2 \rangle \to \Lambda\langle \mathrm{Crit}_1 \rangle$$

as follows. If $x \in \mathrm{Crit}_2$ and $y \in \mathrm{Crit}_1$, let $\mathscr{P}(x, y)$ denote the set of flow lines from x to y (of the flow dual to η), with the orientation induced by η. If $\gamma \in \mathscr{P}(x,y)$, let $(-1)^{\varepsilon(\gamma)}$ denote its sign, as in Section 2.2. (For this to be defined, we need to choose orientations of the ascending and descending manifolds of the critical points, as in Section 2.2.) For $x \in \mathrm{Crit}_2$, define

$$\hat{\mathbf{d}}(x) := \sum_{y \in \mathrm{Crit}_1} \left(\sum_{\gamma \in \mathscr{P}(x,y)} \varepsilon(\gamma)[\gamma] \right) y.$$

Next define

$$I_\eta := \det(\hat{\mathbf{d}}) \cdot \hat{\zeta} \in \Lambda. \tag{4.2}$$

(To fix the sign of the determinant, we have to choose orderings of the sets Crit_1 and Crit_2. Note that $\hat{\mathbf{d}}$ is a square matrix because $\chi(X) = 0$.) The formal sum I_η can be regarded as a function $H_1(X, \eta^{-1}(0)) \to \mathbb{Z}$, and because of the determinant term, this function is zero on any element of $H_1(X, \eta^{-1}(0))$ not contained in the subset $H_1(X, \eta)$. We now finally define

$$I := I_\eta \circ j_\eta : \mathrm{Spin}^c(X) \to \mathbb{Z}.$$

We will not use the following theorem in this paper, but it provides some assurance that I is a reasonable object to define.

THEOREM 4.7. *The map I depends only on X, with the following exceptions*:

(a) *Changing the orientation choices above will multiply I by ± 1.*
(b) *If $b_1(X) = 1$ then I depends on the sign of the cohomology class $[\eta]$.*

See Section 1.4 for some of the ideas in the proof. An n-dimensional generalization is proved in [13] and also follows *a posteriori* from the result of [12].

4.2. *Conjectural analogue of Taubes' "SW = Gr" theorem*

The idea of the following conjecture was suggested to us by Taubes. Recall that $\mathrm{SW} : \mathrm{Spin}^c(X) \to \mathbb{Z}$ denotes the Seiberg–Witten invariant of X (see Section 1.3).

CONJECTURE 4.8 *Let X be a closed oriented 3-manifold with $b^1(X) > 0$. Then*

$$\mathrm{SW} = \pm I.$$

(*If $b_1(X) = 1$, then we mean* SW *to be computed in the chamber determined by η, i.e. counting solutions to the SW equations perturbed by $ir * \eta$ for $r \gg 0$.*)

4.2.1. Motivation. The idea is that this is a dimensional reduction of Taubes's "SW = Gr" theorem [33], extended to certain singular symplectic forms.

The "SW = Gr" theorem asserts roughly that on a symplectic four-manifold, the Seiberg–Witten invariant of a given spin-c structure equals a certain count of pseudoholomorphic curves in a homology class determined by the spin-c structure. The Seiberg–Witten invariants of a 3-manifold X equal the Seiberg–Witten invariants of product spin-c structures on the 4-manifold $X \times S^1$. (A standard integration by parts which Taubes showed us proves that for the product metric, any Seiberg-Witten solution on $X \times S^1$ comes from a solution on X.) If η is a *harmonic* 1-form on X, then on $X \times S^1$ the 2-form

$$\omega := *_3\eta + \eta \wedge d\theta$$

is symplectic, except on the circles $\eta^{-1}(0) \times S^1$ where it vanishes. Moreover, if a 1-manifold $\gamma \subset X$ is parallel to η, then $\gamma \times S^1 \subset X \times S^1$ is pseudoholomorphic (with respect to the almost complex structure on $X \times S^1$ determined by ω and the product metric). This suggests that the Seiberg–Witten invariant of X should somehow count unions of closed orbits and flow lines between critical points.

More specifically, we can try to imitate on X the proof of Taubes's theorem. The strategy is to perturb the curvature equation in the Seiberg-Witten equations by $ir * \eta$, where $r \gg \mathbf{0}$. If r is large, then given a solution of the perturbed equations, the zero set of the E_- component of the spinor should be approximately parallel to η. This suggests that for a spin-c structure W, the Seiberg–Witten invariant $\mathrm{SW}(W)$ counts unions of closed orbits and flow lines between critical points of η whose total homology class equals $j_\eta(W)$. Now our invariant $I(W)$ gives just such a count. The term $\det(\hat{\mathbf{d}})$ is a generating function counting unions of flow lines between critical points whose boundary contains each critical point once. Moreover, by the product formula (4.1), the refined zeta function $\hat{\zeta}$ is a generating function counting unions of closed orbits, where some orbits may be multiply covered (i.e. those for which i_0 is even—which, because η is harmonic, means those orbits which are not hyperbolic). The signs (especially those arising from the determinant) are subtle but apparently necessary to obtain a topological invariant. The treatment of multiply covered orbits is a special case of Taubes' treatment of multiply covered tori in four dimensions [34].

Indeed, if η has no zeroes (and is harmonic, which we can achieve either by deforming η or by choosing an appropriate metric), then the conjecture is a consequence of Taubes's theorem applied to the 4-manifold $X \times S^1$ with the symplectic form ω defined above. (One can check that every pseudoholomorphic curve on $X \times S^1$ in a homology class in $H_1(X) \times H_1(S^1)$ is a union of closed orbits of η crossed with S^1.) Moreover, Taubes [35] has extended part of the "SW $=$ Gr" story to "singular symplectic forms", including forms such as ω above when η has zeroes, and this analysis may eventually lead to a proof of our conjecture.

In the case when η has no zeroes, an alternate proof of the conjecture is provided by a result of Salamon [30] which turns out to be equivalent and asserts that the Seiberg–Witten invariants of the mapping torus of a surface diffeomorphism are given by Lefschetz numbers of the induced maps on the symmetric products of the surface. (Different spin-c structures correspond to different homological fixed point classes.)

4.3. Computing Seiberg–Witten invariants

We now prove Theorem 1.17 (using Theorem 1.12). Let $\phi : X \to S^1$ be a generic Morse function with no index 0 or 3 critical points and with $[d\phi] = \alpha$. (One way to find such a ϕ is to start with the harmonic representative of α and perturb it slightly.) Let 0 be a regular value of ϕ, and let $\Sigma = \phi^{-1}(0)$.

Let $W \in \mathrm{Spin}^c(X)$. We have

$$\alpha(c_1(\det W)) = \int_\Sigma c_1(\det W).$$

The tangent bundle $T\Sigma$ has a complex structure induced by the orientation and metric on X, and Clifford multiplication (identifying $T\Sigma \subset T^*X$ using the metric) gives an isomorphism of complex line bundles

$$T\Sigma \otimes E_-|_\Sigma \simeq E_+|_\Sigma.$$

Therefore $W|_\Sigma \simeq E_-|_\Sigma \oplus (T\Sigma \otimes E_-|_\Sigma)$, so

$$\alpha(c_1(\det W)) = \chi(\Sigma) + 2\Sigma \cdot j_\eta(W).$$

By Conjecture 4.8,

$$\sum_{W \in \mathrm{Spin}^c(X)} \mathrm{SW}(W) t^{\Sigma \cdot j_\eta(W)} = \pm \rho(I_\eta)$$

where $\rho : \Lambda \to \mathbb{Z}((t))$ sends $\gamma \mapsto t^{\Sigma \cdot \gamma}$. Thus

$$\sum_{W \in \mathrm{Spin}^c(X)} \mathrm{SW}(W) t^{\alpha(c_1(\det W))/2} = \pm t^{\chi(\Sigma)/2} \rho(I_\eta).$$

To compute $\rho(I_\eta)$, observe that

$$\rho(\det(\hat{\mathbf{d}})) = \det(\mathbf{d} : C_2^{\mathrm{nov}} \to C_1^{\mathrm{nov}})$$

$$\rho(\hat{\zeta}) = \zeta.$$

(The first equation is clear, and the second is proved by a short combinatorial calculation.) Therefore

$$\sum_{W \in \mathrm{Spin}^c(X)} \mathrm{SW}(W)\, t^{\alpha(c_1(\det W))/2} = \det(\mathbf{d} : C_2^{\mathrm{nov}} \to C_1^{\mathrm{nov}}) \cdot \zeta \tag{4.3}$$

modulo multiplication by $\pm t^k$.

Since C_2^{nov} and C_1^{nov} are the only nontrivial terms in the Novikov complex, $\det(\mathbf{d} : C_2^{\mathrm{nov}} \to C_1^{\mathrm{nov}})$ equals T_{Morse} if $C_*^{\mathrm{nov}} \otimes \mathbb{Q}((t))$ is acyclic, and zero otherwise. In conclusion, if $C_*^{\mathrm{nov}} \otimes \mathbb{Q}((t))$ and $C_*^{\mathrm{cell}}(\tilde{X} \otimes \mathbb{Q}((t)))$ are acyclic, then equation (1.5) which we need to prove follows from equation (4.3) and Theorem 1.12. If $C_*^{\mathrm{nov}} \otimes \mathbb{Q}((t))$ and $C_*^{\mathrm{cell}}(\tilde{X}) \otimes \mathbb{Q}((t))$ are not acyclic, then the left-hand side of (1.5) equals zero because $\det(\mathbf{d})$ vanishes, and the right side is zero by the definition of T_{top}. □

4.4. The Meng–Taubes formula

Theorem 1.17 turns out to be equivalent to a formula for the Seiberg–Witten invariants (in the boundaryless case, modulo signs) obtained by Meng and Taubes [20], as we will now explain.

Let $H = H_1(X)/\mathrm{Torsion} = H^2(X;\mathbb{Z})/\mathrm{Torsion}$, and let $\hat{X}$ be the universal free abelian covering of X, whose monodromy is the projection $\pi_1(X) \to H$. The *Milnor torsion*

$$\mathrm{MT}(X) \in Q(\mathbb{Z}[H])/\pm H$$

is defined as follows. Let $C_*^{\mathrm{cell}}(\hat{X})$ be the chain complex (over $\mathbb{Z}[H]$) coming from an H-equivariant triangulation. This has a basis consisting of a lift of each cell in X to $\hat{X}$, giving a volume form (defined modulo $\pm H$). If $C_*^{\mathrm{cell}}(\hat{X}) \otimes Q(\mathbb{Z}[H])$ is acyclic, define

$$\mathrm{MT} := \tau(C_*^{\mathrm{cell}}(\hat{X}) \otimes Q(\mathbb{Z}[H])) \in \frac{Q(\mathbb{Z}[H])}{\pm H}.$$

If $C_*^{\mathrm{cell}}(\hat{X}) \otimes \mathbb{Q}(\mathbb{Z}[H])$ is not acyclic, define $\mathrm{MT} := 0$.

When $b^1(X) > 1$, it turns out that $\mathrm{MT} \in \mathbb{Z}[H]/\pm H$. (See Turaev [37, Theorem 1.1.2].) Furthermore [37, Section 1.11.5] there is, up to sign, a unique element in this equivalence class invariant under the map that sends $h \to h^{-1}$ for $h \in H$. When $b^1(X) > 1$ we will identify MT with this distinguished lift in $\mathbb{Z}[H]/\pm 1$.

Next, following Meng–Taubes [20], define

$$\underline{\mathrm{SW}} := \sum_{W \in \mathrm{Spin}^c(X)} \mathrm{SW}(W) \frac{c_1(\det W)}{2} \in \mathbb{Z}[[H]].$$

Here $\mathbb{Z}[[H]]$ denotes the set of arbitrary functions $H \to \mathbb{Z}$.

Theorem 4.9. *Let X be a closed oriented 3-manifold with $b^1(X) > 0$. Then Conjecture* 4.8 *implies that*

$$\underline{\mathrm{SW}} = \pm \mathrm{MT}. \tag{4.4}$$

(Equation (4.4), with the sign ambiguity resolved, and for some three-manifolds with boundary as well, was proved by Meng and Taubes, see the announcement in [20].)

The idea of the proof of Theorem 4.9 is to apply "tomography" using Theorem 1.17 and the following simple lemma.

Lemma 4.10. *Let G be a finitely generated free abelian group and let $f, g \in \mathbb{Z}[G]$. Suppose that for every homomorphism $\alpha : G \to \mathbb{Z}$, we have $\alpha_*(f) = \alpha_*(g)$ in $\mathbb{Z}[\mathbb{Z}]$, up to sign. Then $f = \pm g$.*

Proof. Let $\{e_i\}$ be a basis for G. Choose an integer N such that f and g are supported in the set $S := \{\sum a_i e_i \mid |a_i| < N\}$. Let α send e_i to $(2N)^i$. Then for any integer k, the hyperplane $\{x \in G \mid \alpha(x) = k\}$ contains at most one point of S. Apply the hypothesis to this α. □

Proof of Theorem 4.9. If $b^1(X) = 1$ then this is just Theorem 1.17. Assume $b^1(X) > 1$. We have already remarked that $MT \in \mathbb{Z}[H]$. We also have $\underline{\mathrm{SW}} \in \mathbb{Z}[H]$, by the well-known *a priori* bounds for the Seiberg–Witten equations. (See [42] for the four-dimensional case. Note that we do not necessarily have $\underline{\mathrm{SW}} \in \mathbb{Z}[H]$ when $b^1 = 1$, because here we are making a large perturbation to the equations which destroys the *a priori* bounds on their solutions, and the Seiberg–Witten invariants are not invariant under perturbation when $b^1(X) = 1$.) A cohomology class $\alpha \in H^1(X; \mathbb{Z})$ gives rise to a function $\alpha_* : \mathbb{Z}[H] \to \mathbb{Z}[\mathbb{Z}]$, and if we identify $\mathbb{Z}[\mathbb{Z}] = \mathbb{Z}[t, t^{-1}]$, then the left hand side of Theorem 1.17 equals $\alpha_*(\underline{\mathrm{SW}})$, modulo signs. On the other hand, the right side of Theorem 1.17 equals $\alpha_*(MT)$. (This is easy when both the complexes involved are acyclic, and the general case follows from [37, Theorem 1.1.3].) So Theorem 1.17 asserts that $\alpha_*(\underline{\mathrm{SW}}) = \alpha_*(MT)$, modulo signs and powers of t. Since both $\underline{\mathrm{SW}}$ and MT are symmetric, it follows that $\alpha_*(\underline{\mathrm{SW}}) = \alpha_*(MT)$ modulo signs. We are done by Lemma 4.10. □

Acknowledgements—We are grateful to Cliff Taubes for the excellent suggestion which got this project started, for helpful conversations, and for his advice and support in general. The first author thanks R. Bott, K. Fan, D.

Salamon, P. Seidel, and R. Vakil for helpful conversations, and J. Bryan for teaching him the importance of Morse theory. The second author thanks P. Kronheimer for additional helpful conversations.

REFERENCES

1. Bott, R., Morse Theory Indomitable. *I.H.E.S. Publ. Math.* **68**, 1989, pp 99–114.
2. Carey, A. L., McCarthy, J., Wang, B. L. and Zhang, R. B., Topological quantum field theory and Seiberg–Witten monopole. Preprint, 1995.
3. Cohen, R., Jones, J. and Segal, G., Morse theory and classifying spaces. Warwick University, Preprint, 1994.
4. Fintushel, R. and Stern, R., Knots, links, and 4-manifolds. Preprint, 1996.
5. Floer, A., Witten's complex and infinite dimensional Morse theory. *Journal of Differential Geometry*, 1989, **30**(1), 207–222.
6. Fried, D Homological identities for closed orbits. *Inventiones Mathematicae*, 1983, **71**, 419–442.
7. Fried, D Lefschetz Formulas for Flows. Contemporary Mathematics Vol. **58**, Part III, AMS, Providence, RI, 1987, pp. 19–67.
8. Fukaya, K Floer homology of connected sum of homology 3-spheres. *Topology*, 1996, **35**(1), 89–136.
9. Fukaya K., The symplectic *s*-cobordism conjecture: a summary. *Geometry and Physics*, Aarhus, 1995, Lecture Notes in Pure and Applied Mathematics Vol. 184, Dekker, New York, 1997, pp. 209–219.
10. Harvey, R. and Lawson, H. B., Morse theory and Stokes' theorem. Preprint, (1996).
11. Hofer, H. and Salamon, D. A., Floer homology and Novikov rings, The Floer Memorial Volume, Birkhäuser, Basel, 1995, pp. 483–524.
12. Hutchings, M. and Lee, Y.-J., Circle-valued Morse theory and Reidemeister torsion. Preprint, dg-ga/9706012.
13. Hutchings, M., Reidemeister torsion in generalized Morse theory. Thesis, Harvard University, 1998.
14. Ionel, E. and Parker, T., Gromov invariants and symplectic maps. Preprint, dg-ga/9703013.
15. Kronheimer, P. and Mrowka, T., The genus of embedded surfaces in the projective plane. *Mathematical Research Letters*, 1994, **1**, 797–808.
16. Latour, F., Existence de 1-formes fermées non-singulières dans une classe de cohomologie de de Rham. *Publ. Math. IHES*, **80**, 1994.
17. Laudenbach, F., On the Thom–Smale complex, appendix to J-M. Bismut and W. Zhang, An extension of a theorem by Cheeger and Müller, *Asterisque*, 1992, **205**.
18. Lee, Y.-J., Morse theory and Seiberg–Witten monopoles on 3-manifolds. Thesis, Harvard University, 1997.
19. Le Hong Van and Ono, K., Symplectic fixed points, the Calabi invariant and Novikov homology. *Topology*, 1995, **34**(1), 155–176.
20. Meng, G. and Taubes, C., $\underline{\mathrm{SW}}$ = Milnor torsion. *Mathematical Research Letters*, 1996, **3**(5), 661–674.
21. Milnor, J., Lectures on the *h*-cobordism theorem. Notes by L. Siebenmann and J. Sondow. Princeton University Press, Princeton, NJ, 1965.
22. Milnor, J., Infinite cyclic coverings. In *Conference on the Topology of Manifolds*, ed. Hocking, Prindle, Weber, Schmidt, Boston, MA, 1968.
23. Morgan, J. W., The Seiberg–Witten Equations and Applications to the Topology of Smooth Four-Manifolds. Princeton University Press, Princeton, NJ, 1996.
24. Morgan, J. W., Szabo, Z. and Taubes, C. H., A product formula for the Seiberg–Witten invariants and the generalized Thom conjecture. *Journal of Differential Geometry*, 1996, **44**(4), 706–788.
25. Novikov, S. P., Multivalued functions and functionals. An analogue of the Morse theory. *Soviet Mathematics Doklady*, 1981, **24**(2), 222–226.
26. Pazhitnov, A., On the Novikov complex for rational Morse forms, Toulouse. *Faculte des Sciences Annales Mathematiques Series* (6), 1995, **4**(2), 297–338.
27. Pajitnov, A., Surgery on the Novikov complex. *K-theory*, 1996, **10**, 323–412.
28. Pajitnov, A., Simple homotopy type of Novikov complex for closed 1-forms and Lefschetz zeta functions of the gradient flow. Preprint dg-ga/9706014.
29. Pozniak, M., Floer homology, Novikov rings, and clean intersections. Thesis, University of Warwick, 1994.
30. Salamon, D., Seiberg–Witten invariants and symplectic fixed points, (in preparation).
31. Schwarz, M., *Morse Homology*. Birkhäuser, Basel, 1993.
32. Smale, S., Morse inequalities for dynamical systems. *Bulletin American Mathematical Society*, **66**, 43–49 (1960).
33. Taubes, C. H., The Seiberg–Witten and the Gromov invariants. *Mathematical Research Letters*, 1995, **2**, 221–238.
34. Taubes, C. H., Counting pseudoholomorphic submanifolds in dimension four. *Journal of Differential Geometry*, 1996, **44**(4), 818–893.
35. Taubes, C. H., Seiberg–Witten invariants and pseudoholomorphic subvarieties for self-dual, harmonic 2-forms. Preprint, 1998.
36. Thom, R., Sur une partition en cellules associée à une fonction sur une variété. *Comptes Rendus Academic Science, Paris*, 1949, **228**, 973–975.
37. Turaev, V., Reidemeister torsion in knot theory. *Russian Mathematical Surveys*, 1986, **41**(1), 119–182.

38. Turaev, V., Torsion invariants of spin-c structures on 3-manifolds. *Mathematical Research Letters*, 1997, **4**(5), 679–695.
39. Turaev, V., A combinatorial formulation for the Seiberg–Witten invariants of 3-manifolds. Preprint, 1998.
40. Weil, A., Numbers of solutions of equations in finite field. *Bulletin of the American Mathematical Society*, 1949, **55**, 497–508.
41. Witten, E., Supersymmetry and Morse theory. *Journal of the Differential Geometry*, 1982, **17**(4), 661–692.
42. Witten, E., Monopoles and four-manifolds. *Mathematical Research Letters*, 1994, **1**, 769–796.

Harvard University
Department of Mathematics
Cambridge, MA 02138
U.S.A.

Princeton University
Mathematics Department
Princeton, NJ 08544
U.S.A.