RAOUL BOTT

# Morse theory indomitable

*Publications mathématiques de l'I.H.É.S.*, tome 68 (1988), p. 99-114
<http://www.numdam.org/item?id=PMIHES_1988__68__99_0>

# MORSE THEORY INDOMITABLE

*by* Raoul BOTT[1]

*Dedicated to René Thom.*

In the early fifties there were many occasions when Thom's name and work impinged on my life, but maybe none quite as dramatic as when some time in 1952-53 I picked Norman Steenrod up at the Princeton railway station upon his return from Europe. Steenrod was a great hero of mine and my first real topology teacher. I had taken his course on Fiber Bundles in '49 and ever since had pursued the poor man with a vengeance. My strategy was to loiter in parts of Princeton which Norman was known to frequent, and then greet him with surprise when he duly showed up. For Steenrod was wonderful to talk to. He really *listened*! Also he never feigned understanding, and the dreaded words " of course " never crossed his lips.

In any case on this occasion my rendez-vous was on the up and up, because I had a new version of my paper on the Steenrod Squares to deliver to him in his editorial capacity for the Annals. After our greeting he accepted my paper with a quizzical look and showed me a white, relatively slim, envelope in his jacket pocket. " This is a paper of Thom and Wu dealing with the same subject—but I don't want to show it to you until yours is all finished and found to be correct ", he said. And so it came to be that the names " Thom and Wu " haunted my dreams for several months to come, until my paper was accepted and I was finally allowed to see their much more professional one.

In one way or another, both Thom and Wu, and I, came up with intrinsic definitions of the Steenrod operations in terms of the Smith theory of cyclic products. These were then gropings towards Steenrod's ultimate definition in terms of equivariant theory. In any case, I could have slept soundly for Thom and Wu's paper was never officially published, but of course I didn't know that at the time, and so it was that I was greatly relieved when I met Thom in the flesh several years later. For I of course found him the most delightful and unthreatening of people to meet. This was in the year 1955-56 at the IAS where we overlapped for a term.

---

1. Lecture delivered at the Conference in honour of René Thom, Paris, September 1988. Research supported in part by NSF Grant # DMS-86-05482.

As all of us here know, Thom does his mathematics with his fingers and hands, and I still recall the motions of his hands as he taught me that for manifolds with boundary, only half of the critical points on the boundary really " counted ". And the term " critical point " of course brings me to my topic proper of this morning: " Morse Theory Indomitable ". I think Morse would have approved the title for when I first met him, he preached the gospel of " Critical point theory " first, last and forever, to such an extent, that we youngsters would wink at each other whenever he got started. Actually in '49 he was not directly involved in critical point theory. He was completely immersed in delicate questions of lower and upper semi-continuity of quadratic forms on function spaces. Still the motivation behind this excursion was critical point theory—but in infinite dimensions—and it would not surprise me if his papers of that period still have something to teach us. For of course the infinite-dimensional case is very much at the center of our interest at present.

In any case from 1949-51 when I learned the Morse theory—essentially by fiddling with it, and trying to read his and Seifert and Threlfall's books—Morse was off in Hilbert Space and more involved in analysis than topology. In fact the only person who seemed interested in the Morse Theory in Princeton at that time was—of all people—again Steenrod. By 1950 I had come to understand how the " Thom isomorphism " fitted into the Morse Theory and so saw how to " count " higher dimensional critical sets. From this point of view the cohomology of the symmetric product of a sphere, for instance, was very transparent, and I remember that this method pleased Norman.

But it is time to get down to business, and what I would like to do here today is to try and sort out the various stages of " Morse Theory " that I have now lived through, and in particular to comment on the new points of view initiated by Thom (40's), Smale (60's) and Witten (70's) and Floer (80's).

Let me first of all remind you what I am talking about, at least in the most elementary setting, that is: The Nondegenerate Morse Theory on a smooth compact manifold M. In this case we assume as given a smooth function $f: \mathrm{M} \to \mathbf{R}$, whose extrema, i.e., the points $p$ where the differential $df$ vanishes:

$$df_p = 0,$$

are non degenerate in the sense that near such a point the determinant of the Hessian of $f$ at $p$

$$\mathrm{H}_p f = \left\| \frac{\partial^2 f}{\partial x_i \, \partial x_j} \right\|, \quad x_i \text{ local order at } p,$$

does not vanish. These data present us with a second order invariant at each point $p$ which Morse called the " index " of $f$ at $p$, notated by $\lambda_p$, and which he defined as the number of negative eigenvalues of $\mathrm{H}_p f$ at $p$. Thus:

$$\lambda_p = \text{number of negative eigenvalue of } \mathrm{H}_p f.$$

This granted, he formed the polynomial

$$\mathcal{M}_t(f) = \sum_{\{p\}} t^{\lambda_p}, \quad df_p = 0,$$

and observed that this polynomial is always coefficient by coefficient larger than the Poincaré polynomial $P_t(M; K)$ of M, taken relative to any field K. Actually, he showed more, namely that if we write

$$(*) \qquad \mathcal{M}_t(f) - P_t(M; K) = (1 + t)\, Q_t(f; K),$$

and think of $Q_t(f, K)$ as the K-error of $f$, then this error is always *non-negative* in the sense that all the coefficients of Q are nonnegative:

$$Q_t(f, K) = \Sigma a_i t^i, \quad \text{with } a_i \geqslant 0.$$

The $(1 + t)$ term on the right gives this inequality much more power than it would have without it. The $(1 + t)$ feeds back information from the critical points of $f$ to the topology of M.

There is no question that in Morse's mind these inequalities expressed the change in the topology of the half-spaces

$$M_a = \{ m \in M \mid f(m) < a \}$$

determined by $f$, and in the terminology of the late forties I liked to summarize his arguments of the twenties by the following two statements:

*Theorem A. — If there is no critical value of f in the range* $[a, b]$ *then* $M_a$ *and* $M_b$ *are diffeomorphic*

$$M_a \cong M_b.$$

*Theorem B. — If there is precisely one nondegenerate critical point of f of index* $\lambda_p$ *in the range* $[a, b]$ *then the homotopy type of* $M_b$ *is obtained from* $M_a$ *by " attaching " a* $\lambda_p$*-cell:*

$$M_b \sim M_a \cup_\alpha e_{\lambda_p},$$

*by a map* $\alpha : \partial e_{\lambda_p} \to M_a$.

Combining these two theorems with the by then well-known exact sequences associated to an attaching of cells easily yields the Morse inequalities for any cohomology theory satisfying the Eilenberg-Steenrod axioms. Everett Pitcher also made a similar translation of the Morse theory, at about the same time, and of course Thom used it as well. As I remarked earlier, it was he who a few years later explained to me how these principles has to be altered when M is a manifold with nontrivial boundary. I also recall Thom proving the Lefschetz theorem via Morse Theory—but the details of his argument eluded me. At infinity his hands and fingers seemed to take over. Still, these lectures—also never published—were the impetus for the subsequent Andreotti-Frankel Proof, as well as of my version of the theorem, published roughly at the same time.

The Theorems A and B, or " halfspace " approach is in my mind still the simplest ans most straightforward path to the Morse inequalities, and it suffices for purely homotopy theoretic applications. These of course were my main motivation, and my insight was, simply, that what Morse had done for the spheres could be done for the compact Lie groups. Eventually Samelson and I discovered that it works also for all symmetric spaces. Indeed the great symmetry of these spaces is to a large extent reflected in the beautiful properties of the " Energy function "—or free particle " Lagrangian "—

$$\mathscr{E}(\mu) = \int \frac{1}{2} \dot{\mu}(t)^2 \, dt,$$

on the space of paths L(M) on M which we are *still* in the process of uncovering. For Samelson and me the crucial property of $\mathscr{E}$ was simply that on every fiber, $\Omega_{p,q}$, of the mapping $\pi$:

$$\Omega_{p,q} \to \mathrm{L} \xrightarrow{\pi} \mathrm{M} \times \mathrm{M},$$

which sends a path $\mu$ to its endpoints $p$ and $q$, $\mathscr{E}$ restricts to a *nondegenerate* and $\mathbf{Z}_2$-*perfect* Morse Function. Although this situation is *a priori* infinite-dimensional, it can be treated purely by finite-dimensional techniques and in that context $\mathbf{Z}_2$-perfect simply means that (1) the critical sets of $\mathscr{E}$ occur along manifolds (N) along which H is nondegenerate in the normal direction, and (2) that the error term Q in (*) vanishes for the field $\mathbf{Z}_2$. By the way, the appropriate generalization of Theorem B in this situation is simply that

$$\mathrm{M}_b \cong \mathrm{M}_a \cup_\alpha \mathrm{E}^-(\mathrm{N})$$

with $\mathrm{E}^-(\mathrm{N})$ the negative bundle of the normal bundle to the critical set N between $a$ and $b$. (Thus $\mathrm{E}^-(\mathrm{N})$ is a maximal subbundle on which $\mathrm{H}f \leqslant 0$). In short one now attaches a *disc bundle* to $\mathrm{M}_\alpha$ along its " boundary *sphere bundle* ".)

Morse's Technique for proving his beautiful theorem of the twenties, that two points $p$ and $q$ on the $n$ sphere in any Riemann structure are joined by an infinite number of geodesics, was as follows. He first used the " round " metric and its known geodesics to compute the critical points of $\mathscr{E}$ on suitable fibres of $\mathrm{LS}^n$. He then found $\mathscr{E}$ to be perfect—by *purely combinatorial means*, that is, the Morse inequalities force actual equality! In this way he computed the Poincaré Series of the space of paths $\Omega_{p,q}$ from $p$ to $q$, $\Omega_{p,q}$, on $\mathrm{S}^n$ to be

$$\mathrm{P}_t(\Omega_{p,q}\,\mathrm{S}^n) = \frac{1}{1 - t^{(n-1)}}$$

for any field K.

From this he then argues—again—using the Morse inequalities, that for any other metric, the energy function $\mathscr{E}$ would still have to have an infinite number of critical points, and as these correspond to geodesics joining $p$ to $q$, one finds an infinite number of them also in the new structure. Q.E.D.

In my work I used a similar technique. One again uses the "normal" metric on the symmetric space M to actually compute the critical sets, but my strategy was to find a pair of points $p$, $q$ on a given symmetric space M, with the property that the space of "minimal geodesics" from $p$ to $q$, say M′, was as *large* as *possible* and then to estimate the indexes of all higher critical points. This led to formulae of the sort

$$\Omega_{p,q}\,M = M' \cup e_k \cup \ldots, \qquad k > \ell,$$

from which we see that $\pi_r(M) \cong \pi_{r-1}(\Omega M) \cong \pi_{r-1}(M')$ for $r \ll \ell$, and applied appropriately, this procedure yields all the periodicity theorems as well as homotopy equivalences of the form:

$$\Omega O = O/U,$$

$$\Omega(O/U) = U/SP, \text{ etc.}$$

But enough said here about those old '40's, '50's homotopy aspects of Morse Theory.

It was Smale who refined this "half-space" approach I have been discussing ten years later when he subtly recast Theorem B in a greatly strengthened form.

If $\dim M = n$, then the formulation reads:

$$M_b \cong M_a \cup e_{\lambda_p} \times e_{n-\lambda_p}$$

that is, we have attached a "thickened disc", $e_{\lambda_p} \times e_{n-\lambda_p}$, as we cross the critical point, *but now the equivalence is in the* "$C^\infty$ category", and so allows for a stepwise inductive procedure for determining diffeomorphism type. This then led him to his handlebody theory, the generalized Poincaré Conjecture and the $h$-cobordism theorem! Not bad.

Smale enhanced our understanding of Morse Theory in two more ways. First of all he and Palais formulated abstract criteria under which Theorems A and B could be expected to hold. These are normally referred to as the "condition C" of Palais-Smale. These conditions hold, for instance, for the Lagrangian $\mathscr{E}$ discussed earlier when the space LM is given in the appropriate topology, but unfortunately they fail for many of the other geometrically induced Lagrangians, such as those of minimal surface theory or the Yang-Mills theory.

Secondly Smale saw how to fit Morse Theory into the scheme of dynamical systems and so to complete a program that really started in a 1949 *Comptes Rendus* note of R. Thom's. It is in this framework that we now also make contact with physics, with terms like "instantons" and with what I would like to call the "Thom-Smale" and the "Witten" complexes.

The starting point of these developments is to pass from $f$ on M, to the gradient of $f$, say $\nabla f$, relative to a smooth Riemann structure $g$ on M. Let's stay in the truly *non-degenerate* case so that the vector field $\nabla f$ then vanishes only at the finite number of critical points of $f$. In this way every point $q$ which is not a critical point of $f$ lies on a unique 1-dimensional integral manifold $X_q$ of $\nabla f$ which will "start" at some critical point $p$ and end at some other critical point $r$ (I am assuming that M is compact, as throughout).

The physicists call such an integral manifold an " instanton " for the following reason: if we parametrize the $q$ " trajectory " $X_q$, by solving the differential equation

$$\frac{du(t)}{dt} = -\nabla f(u(t)), \qquad \mu(0) = q,$$

then $u$ will be defined on all of $\mathbf{R}$ with $\lim_{t \to -\infty} u(t) = p$ the initial point of the trajectory, and $\lim_{t \to +\infty} u(t) = r$ its final point. The path $t \to u(t)$ will furthermore hover near its initial point for " most of the time " $t < t_0$ and near its " final point " for most of the time $t > t_1$. In short, just as in the great scheme of things our lives take only an instant to live, these " instantons " stay put at $p$ and $q$ most of the time and then whip from $p$ to $q$ in an instant.

In any case, terminology aside, what Thom pointed out already in 1949 was that if we gather together the instantons having a given critical point $p$ as their *initial* point, then this set—denoted by $W_p$—is a cell of *dimension* $\lambda_p$. This is the " descending cell " through $p$.

It is clear enough, then, that the cells $\{ W_p \}$, as $p$ ranges over the critical points of $f$, decompose M into disjoint sets. Unfortunately, however, this is, in general, not a " good cell decomposition " in the usual sense of the word. The closures of the cells can be complicated and to use this construction *per se* to derive the Morse inequalities requires some additional pushing and pulling. I was vaguely aware of Thom's paper, but only saw Samelson's review which only seemed to give the weaker Morse inequality, that is, without the feedback term. Also, I am afraid that *Comptes Rendus* notes—especially by inspired dreamers like Thom—did not carry a very high credit rating. All in all then, I forgot this paper, until in 1960, I believe, when at a conference in Zürich, Smale and I went to have a swim at a beautiful pool in the environs of Zürich. There Steve explained to me his approach to dynamical systems, which appeared in an *A.M.S. Bulletin* note of 1959.

By the way, it is interesting to meditate on the relative credit rating of *Bulletin* notes by Smale and *Comptes Rendus* notes by Thom. Michael Atiyah and I tend to call communications of this type " morally correct ", and I leave it to you to ponder the implications of this expression on our morals and the validity of the results in question. In any case, Smale there introduces the concept of transversality into Thom's cell-decomposition and at the same time extends it to integral manifolds of more general vector fields subject to certain axioms.

In the Morse theoretic context Smale's idea is this:

As we saw, the gradient field $\nabla f$ decomposes M into the descending cells

$$M = \bigoplus_p W_p.$$

If we change $f$ to $-f$, the new " descending " cells are called the " ascending " cells of $f$:

$$M = \bigoplus_p W'_p$$

and Smale would call $\nabla f$ *transversal*, if these two types of cells always meet in as " generic a way " as they can. Precisely, this means that for any point $q$ in the intersection $W_p \cap W'_r$ the tangent spaces $T_q W$ and $T_q W'$ should span $T_q M$.

Thus for each $q \in W_p \cap W'_r$ one then has the exact sequence

$$0 \to T_q X_q \to T_q W_p \oplus T_q W'_r \to T_q M \to 0,$$

for, of course, the tangent to the trajectory of $\nabla f$ through $q$ is contained in both $T_q W_p$ and $T_q W'_r$.

The transversality condition is best understood by an example in which it fails. And perversely enough this is the case for the " prime " example all of us use when we explain the Morse theory to the uninitiated! That is, of course, the height function $z$, restricted to the torus standing on the $x, y$ plane. The picture is as indicated below,

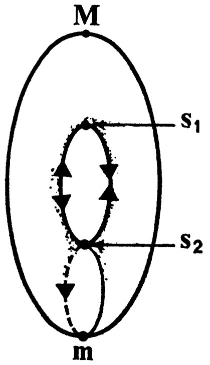


FIG. 1

and the height $z$ takes on its maximum at M, minimum at $m$, and clearly has two saddle points $s_1$ and $s_2$. The gradient of $-z$, then starts a 2-cell at M, and 1-cells at $s_1$ and $s_2$ which I have also indicated. This decomposition violates Smale's axiom, for the ascending cell of $s_2$ at $q$ agrees with the descending cell at $s_1$, so that their tangent spaces do *not* span the whole tangent space.

If one perturbs—$\nabla f$ a little, this phenomenon will disappear, and, in fact, $W_{s_1}$ and $W'_{s_2}$ will then fail to intersect at all!

The cells of a perturbed version of $-\nabla f$, therefore, will look roughly like:

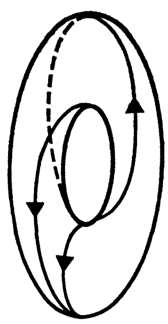

FIG. 2

Notice by the way that the transversality condition implies that

$$\dim(W_p \cap W'_q) = \lambda_p - \lambda_q + 1$$

from which it follows that the number of " instantons " joining two critical points whose indexes differ by 1 is finite!

I wish Steve had written a more extended account of his note, for then surely he would have pointed out that in fact the disposition of these finite numbers of " proper instantons ", as I will call them, are precisely what one needs to compute the homology of M! Instead he derives the Morse inequalities and their generalization to more general flows and hurries on to other things. But certainly implicit in his note is the following algorithm, which I consciously became aware of only in the early '80's when Witten came to my office one day and asked me whether it was well-known that a procedure of the sort he was describing yielded the cohomology of M.

The algorithm in question is the following one.

Given a truly nondegenerate $f$ with transversal gradient $\nabla f$, orient the descending cells arbitrarily and consider the free group over $\mathbf{Z}$ generated by them:

$$C^f(M) = \{[W_p]\}.$$

We grade $C^f(M)$ by the dimension of $W_p$, and define a boundary operator

$$\partial : C_r \to C_{r-1}$$

by simply counting each fundamental instanton $\gamma$ with a $\pm 1$; thus

$$\partial[W_p] = \Sigma e(\gamma)\,[W_q]$$

with $\gamma$ a proper instanton from the critical point $r$ of index $k-1$, and $e(\gamma) = \pm 1$ according to whether for some $q \in \gamma$ the exact sequence

$$0 \to T_q X_q \to T_q W_p \oplus T_q W'_r \to T_q M \to 0$$

preserves orientation or not.

(For simplicity sake, assume here that M is oriented. Then the orientations on $W_p$ induce orientations on $W'_p$ while $T_q X_q$ is oriented by $-\nabla f_x$ so $e(\gamma)$ is well-defined).

With this understood one has the following easy consequence of Smale's work:

*Theorem. — The complex* $C^f(M)$ *defined above, has the property that* $\partial^2 \simeq 0$, *and*

$$H^*(M; \mathbf{Z}) \cong H^*\{C^f(M)\}. \tag{**}$$

This relation of course implies the Morse inequalities because of the purely algebraic fact that the counting series,

$$C_t = \Sigma t^q \dim C_q,$$

of any finite-dimensional chain complex satisfy the Morse inequalities relative to the Poincaré series of its cohomology,

$$P_t = \Sigma t^q \dim H^q(C; K),$$

relative to any field K.

Whether the complex $C^f(M)$ should be called the " Smale complex of $f$ " or the " Smale-Witten " complex or the " Thom-Smale-Witten " complex I leave to you. In any case, I think of it as the most beautiful formulation of the nondegenerate Morse theory, with the analysis not only prescribing the dimensions of the cells but even the attaching maps. This is also the formulation Marston Morse was groping for in his many later papers with Cairns, for instance. On the other hand, one cannot really call this procedure practical. In general, it is difficult to find a metric whose gradient flow for $f$ is transversal (Smale proves that they are dense), and it is then a difficult matter to actually compute the instantons. On these grounds, Morse might well have found this formulation wanting. Still, in my view it is a beautiful and simple a statement as one might wish for. And what I like even more is the road that brought Witten to ask h s question.

His approach is, as you will see, along *quite* different lines, and really lines which are *characteristic* of the modern physicists' world view. For the grand lesson of quantum theory is that the Hilbert space of functions on M is in some sense more " real " than the points of M, and correspondingly, in Witten's view the deformations of M—that is, the pushing of M along the gradient of $f$, which underlies both approaches to the Morse theory outlined so far—are replaced by a quite different deformation which takes place in the function-space attached to M by the Hodge theory.

But first a little history. In August, 1979, I gave some lectures at Cargèse on equivariant Morse theory, and its pertinence to the Yang-Mills theory on Riemann surfaces. I was reporting on joint work with Atiyah to a group of very bright physicists, young and old, most of whom took a rather detached view of the lectures. " Beautiful, but oh so far from Physics " was Wilson's reaction, I remember. On the other hand, Witten followed the lectures like a hawk, asked questions, and was clearly very interested.

I therefore thought I had done a good job indoctrinating him in the rudiments of the half-space approach, etc., so that I was rather nonplussed to receive a letter from him some eight months later, starting with the comment, " Now I finally understand Morse theory! "

(This comment was actually very reminiscent of Smale's comment to me in 1960! For Smale was of course also a student of mine, whom I believed to have taught the Morse theory in the " proper way ". It is quite a hardship to have such bright students!)

In any case, let me now explain to you the gist of Witten's approach, at least the way I understand it.

We start with M, $f$ and a metric $g$ on M just as before. But we now consider the de Rham complex

$$\Omega^* : \Omega^0 \to \Omega^1 \to \dots \Omega^n$$

of M, and its " Hodge theory " relative to $g$. Thus $g$ induces an adjoint $d^*$ to $d$ going the other way. The resulting Laplacian

$$\Delta = dd^* + d^* d$$

can then serve to decompose $\Omega^q$ into a direct sum of finite-dimensional eigenspaces:

$$\Omega^q = \bigoplus_{\lambda} \Omega^q_\lambda, \qquad \lambda \in \mathbf{R},$$

with $\Omega^q_\lambda = \{ \varphi \in \Omega^q \mid \Delta\varphi = \lambda\varphi \}$.

The Hodge theory then implies that:

1) $\Omega^q_0 \cong H^q(M)$, and

2) $0 \to \Omega^q_\lambda \xrightarrow{d_\lambda} \Omega^{q+1}_\lambda \to \ldots \Omega^n_\lambda \to 0$ is exact for $\lambda > 0$.

Here, $d_\lambda$ is the restriction of $d$ to $\Omega^*_\lambda = \bigoplus_q \Omega^q_\lambda$. These two conditions trivially imply that *all* the finite-dimensional complexes

$$\Omega^*_a = \bigoplus_{\lambda \leqslant a} \Omega^*_\lambda, \qquad a > 0,$$

have $H^*(M)$ as their cohomology.

In particular, then all the counting series

$$\Omega^*_{a,t} = \Sigma \dim(\Omega^q_a)\, t^q$$

satisfy the Morse inequalities relative to $P_t(M)$. As we move the metric $g$ on M the spaces $\Omega^q_a$ of course jump about, but they must always respect these inequalities, because all the complexes $\Omega^*_a$ compute $H^*(M)$.

Now comes Witten's idea. Introduce the operator

$$d_s = e^{-sf} \circ d \circ e^{sf}$$

with $s \in \mathbf{R}$ a real parameter. In short, conjugate $d$ by multiplication with $e^{sf}$. We clearly have $d_s^2 = 0$ and hence cohomology groups

$$H_s(M) = \operatorname{Ker} d_s/\operatorname{Im} d_s.$$

However, it is an easy matter to see that $\dim H^*_s(M)$ is independent of $s$.

$$H^*_s(M) \cong H^*(M),$$

because we are just conjugating.

On the other hand, we can also compute $H_s(M)$ via the Hodge theory: this leads to the operator

$$\Delta_s = d_s d^*_s + d_s d^*_s,$$

and the corresponding decompositions

$$\Omega^*(s) = \bigoplus_{\lambda} \Omega^*_\lambda(s)$$

into the eigenspace of $\Delta_s$. Just as before, we also have for every $a > 0$ the finite-dimensional complex of differential forms $\Omega^*_a(s)$ spanned by all eigenforms of $\Delta_s$ with eigenvalues $\lambda \leqslant a$ and the counting series of all these $\Omega^*_a(s)$ must for the same reason as before again satisfy the Morse inequalities relative to $P_t(M)$.

It is this curve of finite-dimensional chain complexes $\Omega^*_a(s)$ which Witten uses in his version of the Morse theory!

Namely, he argues that for $s$ very large the dimensions of this complex become independent of $s$, so that we can denote them by $\Omega_a^*(\infty)$, say, and that furthermore:

1. $\dim \Omega_a^k(\infty) =$ number of critical points of *index k*, and
2. the differential operator induced by $d$ on $\Omega_a^k(\infty)$ is carried by the proper instantons from the critical points of index $k$ to those of index $k + 1$.

But let me spell this all out in greater detail in a simple example.

Consider a function with four critical points on $S^1$, say as depicted below:

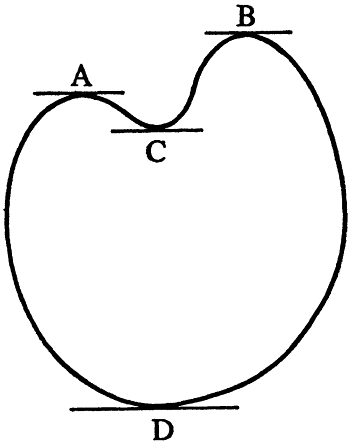


FIG. 3

and let us plot the spectrum of the laplacian $\Delta$ on $H^0$ and $H^1$, respectively:

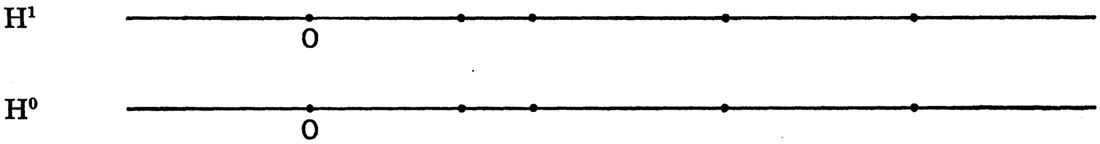


FIG. 4

By Hodge theory we know that the distribution of eigenspaces must have the disposition indicated in Fig. 4 with the crosses bearing equal multiplicity in the vertical direction.

Now the idea behind the Witten construction is that for $s$ very large the spectrum of $\Delta_s$ will have moved—in a possibly very complicated way—to a terminal position of the type described in the next figure.

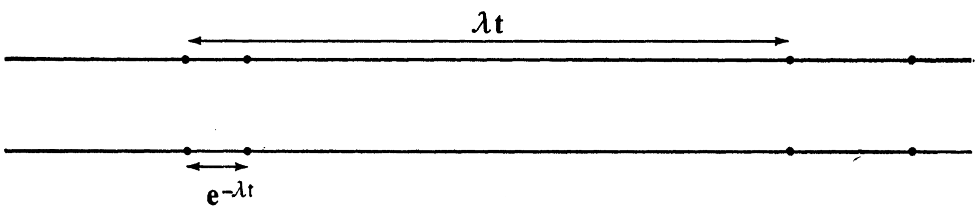


FIG. 5

Thus $\Omega_a^*(+\infty)$ will be a 2-step chain complex with 2-dimensional components. Furthermore, a basis for $\Omega_a^1(\infty)$ will consist of 1-forms which are essentially concentrated at the maxima A and B, while a basis for $\Omega^0(\infty)$ will consist of functions essentially concentrated at the minimal C and D. Furthermore, the operator $d_\infty : \Omega_\infty^0 - \Omega_\infty^1$ now describes the " tunnelling effect " between the minima and the maxima, and in the physics literature this effect is computed precisely by estimating the contribution of the proper instantons in question. Actually, these effects were first proposed in the much grander and infinite-dimensional setting of Yang-Mills theory, so that Witten was " coming down " to the much more mundane mathematical scene when he, as he put it, finally " understood " what was going on.

I do not have much more time left, so let me close by trying to give you a little bit of the flavor of how Witten's assertion comes about at least in this simplest case. He did not *prove* it by mathematical standards in his wonderful paper [8]. For that we had to wait for the quite difficult papers of Helder and Sjöstrand [5], where of course Smale's transversality condition enters. (In fact, it is nowadays possible to write long *Comptes Rendus* notes and publish them in Journals like *Differential Geometry*, thanks to the Pioneering work in this direction of Dennis Sullivan and Bill Thurston, say, and so you should really think of Witten's paper [8] as being in a direct line with the notes of Thom and Smale.)

The clue to the whole phenomenon is to study the nature of $\Delta_s$ in detail near a critical point of $f$. Let us assume then that near a maximum, say A, of $f$, we have

$$f(x) = c - \frac{x^2}{2},$$

with $x$ a local coordinate centered at A on $S^1$. Assume also that we have given $\mathbf{R}$ its flat Riemann structure. Then in the basis 1, $dx$ for $\Omega^0$ and $\Omega^1$ respectively, $d_s$ is simply represented by

$$d_s = \partial_x - sx$$

and hence $d_s^*$ by

$$d_s^* = -\partial_x - sx.$$

It follows that

$$\begin{aligned} \Delta_s^0 = d_s^* \circ d_s &= (\partial_x + sx)\,(\partial_x - sx) \\ &= -\partial_x^2 + s^2 x^2 + s.1, \\ \Delta_s^1 = d_s \circ d_s^* &= -\partial_x^2 + s^2 x^2 - s.1. \end{aligned}$$

Thus if $H_s$ denotes the quantum mechanical harmonic oscillator $-\partial_x^2 + sx^2$, then

$$\Delta_s^0 = H_s + s.1,$$

while

$$\Delta_s^1 = H_s - s.1.$$

Now the spectrum of $H_s$ is given by

$$\operatorname{Spec}(H_s) = |s|, 3|s|, \ldots$$

from which it follows that, for $s > 0$,

$$\operatorname{Spec}(\Delta_s^0) = 2s, 4s \ldots$$

while $$\operatorname{Spec}(\Delta_s^1) = 0, s, \ldots.$$

Similarly, near a minimum these two spectra are reversed:

$$\operatorname{Spec}(\Delta_s^0) = 0, s, \ldots$$

$$\operatorname{Spec}(\Delta_s^1) = 2s, 4s \ldots.$$

The global consequence of these estimates is that if $f$ is an eigenfunction of $\Delta_s^0$ with small eigenvalue (as $s \to +\infty$) then it must be concentrated near the minima, and vice-versa for 1-forms. Inversely, if we start with the " vacuum state " of $\Delta_0$, $\Phi$ of a minimum in $L^2(x)$ and smooth it out to give a global function $\widetilde{\Phi}$, say by cutting its support to a compact region, then, for large $s$, $\Delta_s \widetilde{\Phi}$ will involve only small eigenvalues for arbitrary large $s$.

These heuristic are of course second nature to physicists and led Witten to the conclusion stated above.

To recapitulate: the curve of elliptic differential operators $\Delta_s = d_s d_s^* + d_s^* d_s$ has, for large $s$, the property of dividing the spectrum into a *low-lying sector*, and a *large sector*. For very large $s$ the low-lying sector has eigenvalues $\geqslant 0$ and arbitrarily close to 0 and this low-lying sector $\Omega_\infty^*$ constitutes the " Witten complex ". Because it computes $H^*(M)$ and on the other hand has a basis given in terms of the critical points of $f$ this complex in particular proves the Morse inequalities. Q.E.D.

Witten's paper on Morse theory and super-symmetry is a gold mine also in other respects, and teaches us how the physicist's " thinking in terms of harmonic oscillators " enters profitably into many questions. For instance, it teaches us that a similar technique gives beautiful and transparent proofs of the fixed point theorems Atiyah and I proved in the sixties, and in general points the way to a more " hands on " approach of all the classical index and equivariant index problems. The work of Bismuth, Getzler, Vergne and others all fall under this heading: pointwise formulae for indexes and pushforwards, which we previously only understood in cohomology. In this paper we also first learn about the pertinence of $S^1$-equivariant theory to these questions, and the free loopspace as the crucial instance of an $S^1$-space. In short, Witten's approach not only taught us to " recall "—in the platonic sense—the most satisfying version of the nondegenerate Morse theory, but also taught us to relate the quantum mechanical concept of " tunnelling " to topology.

But let me turn to a different development of the late seventies when for me the Morse theory suddenly illuminated a completely new field. I am speaking of the relation of Morse theory to symplectic geometry. However, as my time is up I must now really

" change scale " altogether if I am to give you at least a glimpse of Morse theory on the more contemporary scene.

In fact, it is quite depressing to see how long it is taking us collectively to truly sort out symplectic geometry. I became aware of this especially when one fine afternoon in 1980, Michael Atiyah and I were trying to work in my office at Harvard. I say trying, because the noise in the neighboring office made by Sternberg and Guillemin made it difficult. So we went next door to arrange a truce and in the process discovered that we were *grosso modo* doing the same thing. Later Mumford joined us, and before the afternoon was over we saw how Mumford's " stability theory " fitted with the Morse theory.

The important link here is the concept of a moment map, which in turn is the mathematical expression of the relation between symmetries of Lagrangians and conserved quantities; in short, what the physicists call " Noether's theorem " and which is one of their great paradigms.

Precisely, let X be a vector field preserving the symplectic form $\omega$ on the symplectic manifold M.

Thus, by my favorite formula, we have

$$\mathscr{L}_X \omega = di_X \omega + i_X d\omega = 0,$$

whence $i_X \omega$ is a closed 1-form. Now, if we can find a function $f_X$ on M such that

$$df_X = i_X \omega,$$

we call $f_X$ a moment function for X, and these functions are endowed with especially miraculous properties—provided X generates a *compact* group of symplectic diffeomorphisms. Indeed, under these assumptions,

(1) $f_X$ will be a perfect Morse function on M. Furthermore

(2) the pushforward $f_{X*}$ of the measure $\omega^n$ on M, is piecewise polynomial on **R**, or, quite equivalently,

$$\int e^{itf_X} \frac{\omega^n}{n!} = \frac{1}{(it)^u} \sum_p \frac{e^{itf(p)}}{e_p},$$

where the sum is taken over the fixed points of X.

The first property was noted long ago by Frankel in the " Kaehler case ", the second is in a theory of Duistermaat-Heckman [3] of the seventies, and the proof of their theorem via the formulation above was noted by Atiyah and myself [2].

This recipe for constructing perfect functions gives all my earlier examples on homogeneous spaces such as K/T (T a maximal torus of the compact Lie group K) and the Bruhat cells of these spaces agree with the Thom-Smale cells relative to the invariant metric of certain $f_X$'s.

In the infinite-dimensional context the recipe recreates the energy function $\mathscr{E}$, discussed early on the loop space on K, and its " cells " give the Bruhat decomposition on loop groups!

To fit these concepts into the Mumford theory, recall that the problem there is to define a suitable algebrogeometric notion of quotient " M/G " in the algebraic category. If $K \subset G$ is the maximal compact subgroup of G and $\omega \in \Omega^2(M)$ a Kaehler form on M preserved by K, then the moment map for this situation becomes an equivariant map

$$M \xrightarrow{f} k^*$$

and its norm $||f||^2$ is the pertinent function for Mumford's stability theory. *Grosso modo*—the " stable points " of Mumford's theory correspond to the big open cell *ascending* from the minimum of $||f||^2$, and the algebraic quotient " M/G " is to be taken to be $||f^2||^{-1}$ (minimum)/K. Thus

$$\text{“ } M/G \text{ ”} \simeq ||f^2||^{-1}(\min)/K.$$

In the infinite-dimensional case Atiyah and I were discussing at the time, $||f||^2$ could be interpreted as the Yang-Mills functional on the space of connections $\mathscr{A}$ of a principal bundle P over a Riemann surface M,

$$YM : A \to \int_M || F_A ||^2,$$

and this YM turned out to be *equivariantly perfect*, thereby enabling us to compute the cohomological properties of the minimum of YM—that is, the space of flat bundles over M [1].

These then were the directions in which the interaction between symplectic structure and the Morse theory arose in our work. They all spring from the duality $X \sim f_X$, which of course identifies the fixed points of X with critical points of $f_X$.

The attempt to " integrate " this duality from symplectic vector fields to symplectic diffeomorphisms of course brings us into the mainstream of the symplectic school going back to Poincaré and Birkhoff and it is in this area that the work of Arnold, Gromov, Zehnder, Conley, Chaperon, and many others has recently been beautifully extended by Floer [4]. I have just one minute left to comment on it. Note first that the symplectic structure $\omega$ on M induces a new function on the space of paths LM we encountered earlier. Indeed, the endpoint projection $LM \xrightarrow{\pi} M \times M$ induces an evaluation map

$$LM \times I \xrightarrow{e} M \times M$$

defined by $e(\mu, t) = \mu(t)$, so that the pullback of the symplectic form $\omega' = 1 \otimes \omega - \omega \otimes 1$ under $e$ gives rise to a 2-form on $LM \times I$, and hence, by integration over the integral I, to a 1-form

$$\theta = \int_I e^* \omega$$

on LM. Now let $V \subset M \times M$ be a Lagrangian submanifold of M (e.g., the graph of a symplectic diffeomorphism $\varphi$), and let $L_V M \subset LM$ be the subspace $\pi^{-1}(V)$ in LM. Then it is easy to see that *restricted to* $L_V M$, $\theta$ becomes closed and hence at least locally $df$.

The " Morse theory " of this closed 1-form now leads Floer to an infinite-dimensional analogue of the Thom-Smale-Witten complex C, whose cohomology is of interest in several contexts [4]. The crucial remark here is that in infinite-dimensional situations, where the Hessian of a function $f$ at a critical point $p$ might well have an infinite number of negative and positive eigenvalues so that it is meaningless to speak of an index $\lambda_p$, it is still possible to make sense of the " relative index " of two critical points $p$ and $q$, by measuring the " spectral flow of $\mathrm{H}f$ " along a curve joining $p$ and $q$. Thus $\mathrm{C}^f$ can be given a relative grading, proper instantons can be defined, and so the boundary operator $\partial$ carried by these is again given meaning.

But my time is up and I must stop even though I have in no way exhausted my topic. There is, for instance, no mention of the " beyond condition C " work of Uhlenbeck and Taubes and others and not a word about the achievements of surgery theory. Still, I hope to have convinced you a little that Morse theory is indeed indomitable, and hopefully this rehearsal will also underscore for you Thom's dictum that the simple ideas are the ones that yield the greatest power.

## BIBLIOGRAPHY

[1] M. F. Atiyah and R. Bott, The Yang-Mills equations over Riemann surfaces, *Phil. Trans. Roy. Soc.*, London, A308 (1982), 523-615.

[2] M. F. Atiyah and R. Bott, The moment map and equivariant cohomology, *Topology*, **21** (1) (1984), 1-28.

[3] J. J. Duistermaat and G. J. Heckman, On the variation in the cohomology in the symplectic form of the reduced phase space, *Invent. Math.*, **69** (1982), 259-268.

[4] A. Floer, Morse theory for Lagrangian intersections, *J. Diff. Geom.*, **28** (1988), 513-547.

[5] B. Helffer, J. Sjöstrand, Points multiples en mécanique semiclassique IV, étude du complexe de Witten, *Comm. Par. Diff. Equ.*, **10** (1985), 245-340.

[6] S. Smale, Differentiable dynamical systems, *Bull. Am. Math. Soc.*, **73** (1967), 747.

[7] René Thom, Sur une partition en cellules associée à une fonction sur une variété, *C.R. Acad. Sci. Paris*, **228** (1949), 661-692.

[8] E. Witten, Supersymmetry and Morse theory, *J. Diff. Geom.*, **17** (1982), 661-692.

Harvard University
Department of Mathematics
One Oxford Street
Cambridge, Massachusetts 02138

*Manuscrit reçu le 28 novembre 1988.*