Phil. Trans. R. Soc. Lond. A **308**, 523–615 (1982)

Printed in Great Britain

THE YANG–MILLS EQUATIONS OVER RIEMANN SURFACES

By M. F. ATIYAH, F.R.S.,† and R. BOTT‡

† *Mathematical Institute, University of Oxford, 24–29 St Giles, Oxford OX1 3LB, U.K.*

‡ *Department of Mathematics, Harvard University, Cambridge, Massachusetts* 02138, *U.S.A.*

(*Received* 4 *March* 1982)

CONTENTS

The Yang–Mills functional over a Riemann surface is studied from the point of view of Morse theory. The main result is that this is a 'perfect' functional provided due account is taken of its gauge symmetry. This enables topological conclusions to be drawn about the critical sets and leads eventually to information about the moduli space of algebraic bundles over the Riemann surface. This in turn depends on the interplay between the holomorphic and unitary structures, which is analysed in detail.

Introduction

This paper is a greatly expanded account of the preliminary material that appeared in Atiyah & Bott (1980). Part of the reason for the long delay between that paper and this present version was that new view-points emerged that added further interest to the topic and enabled us to provide proofs for what had earlier been conjectures. The length of this paper is due to our desire to present the several different aspects of the problem. We feel that this is justified because the main interest lies not so much in the actual applications as in the methods employed and the interaction between different approaches.

Because of its long gestation period and its consequent size, we feel that we owe the reader a substantial and leisurely introduction that puts the paper into historical perspective. In fact, on a personal level, one of the attractions of this subject to us is that it brings together algebraic vector bundles and Morse theory, topics to which we separately made contributions a quarter of a century ago (Atiyah 1955, 1957; Bott 1958, Bott &Samelson 1958). Even better, the catalyst that produced this interaction came from a quite unexpected quarter, namely that of theoretical physics in the form of the Yang–Mills equations.

At this stage we should perhaps explain that our main idea is to apply Morse theory methods to the Yang–Mills functional over a compact Riemann surface M (or algebraic curve) and deduce results about the cohomology of the moduli spaces of stable algebraic vector bundles over M. To explain the background we shall now digress to give brief historical accounts of both Morse theory and algebraic bundle theory.

Morse theory is concerned with the relation between the homology of a manifold M and the critical points of a real-valued function f on M. When M is finite-dimensional these ideas go back at least to Poincaré but they have been applied in more refined form in recent times to derive deep results concerning the geometry of manifolds. Morse's great contribution was to deal with the infinite-dimensional case arising from variational problems for functions of one variable. The most noteworthy geometrical application was to the 'Energy' function on the loop space, which yielded significant results concerning closed godesics. In Bott (1958), Bott & Samelson (1958), Morse theory was applied to nice spaces arising from Lie groups, such as (some) homogeneous spaces G/H and the loop space ΩG, where explicit knowledge of natural functions could be exploited to derive information about the cohomology of the spaces concerned.

As a very simple example consider, on the complex projective n-space $P_n(C)$, the function f defined by

$$f(z_0, \ldots, z_n) = |z_0|^2/\Sigma|z_i|^2,$$

where we use standard homogeneous coordinates. Clearly f has a unique maximum at the point $(1, 0, 0, \ldots, 0)$ and a minimum along the hyperplane $z_0 = 0$. Morse theory then allows one to conclude that the cohomology of $P_n(C)$ differs from that of $P_{n-1}(C)$ by a single free generator in dimension $2n$. This is an easy consequence of the fact that the complement of $z_0 = 0$ is a copy of C^n. Of course in this example, and in many other explicit cases, one does not need the function f to produce the decomposition into such pieces (or strata). In fact for complex homogeneous spaces one can always produce such a stratification from orbits of suitable groups. The same applies to ΩG. Thus in these cases derived from groups, Morse theory, which uses *real* functions, can be replaced by *complex* analytic methods. A much more sophisticated, though computationally simple method of computing Betti numbers is to use the Weil conjectures as established

by Deligne, which involves counting the number of points over finite fields. This works for algebraic manifolds and for $P_n(F_q)$ one finds

$$N(n,q) = (q^{n+1}-1)/(q-1) = 1+q+q^2+\ldots+q^n.$$

Note that the equality

$$N(n,q) = N(n-1,q)+q^n$$

corresponds to the stratification of P_n discussed above, which indicates the common role this plays in all three methods of computing Betti numbers. This number-theory approach is frequently very effective but it does not give as much information as the more direct geometrical methods. In particular the fundamental group and torsion cannot be computed by number theory. In fact the Morse theory proof in Bott & Samelson (1958) that ΩG was torsion-free was a significant triumph of the method, particularly since G itself *can* have torsion.

We turn next to the topic of algebraic vector bundles over an algebraic curve. Over the complex numbers these are equivalent to *holomorphic* vector bundles over the associated compact Riemann surface. For vector bundles with fibre of dimension one, i.e. line-bundles, the classical divisor theory of Abel–Jacobi expresses the fact that the isomorphism classes of line-bundles form an abelian group isomorphic to $Z \times J$, where J is the Jacobian of the curve and the integers Z correspond to the Chern class of the line-bundle (or the degree of the divisor). Weil (1938) began the generalization of divisor theory to that of *matrix* divisors, which correspond to the modern notion of vector bundle. The classification problem for bundles of rank $n > 1$ is much harder than for line-bundles partly because there is no group structure. Grothendieck (1957)showed that for genus 0 the classification is trivial, in the sense that every bundle is a sum of line-bundles. Atiyah (1957) extended the classification to genus 1 and (Atiyah 1955) treated the case of rank 2 bundles for genus 2. In general in order to get a good *moduli space* one has to restrict to the class of *stable* bundles as introduced by Mumford; otherwise one gets non-Hausdorff phenomena. A major breakthrough came with the discovery by Narasimhan & Seshadri (1965) that bundles are stable if and only if they arise from irreducible (projective) unitary representations of the fundamental group. This connection between holomorphic and unitary structures was already apparent in Weil's paper, and in the classical case of line-bundles it is essentially equivalent to the identification between holomorphic and harmonic 1-forms, which in turn was the starting point for Hodge's general theory of harmonic forms.

The unitary view-point enabled Newstead to examine the topological properties of the moduli space for rank 2, obtaining in particular formulae for the Betti numbers. A direct generalization of this method to higher rank appeared intractible. A quite different approach, initiated by Harder (1970) for rank 2 and successfully generalized by Harder & Narasimhan (1975) for higher rank, was number-theoretical based on the Weil conjectures and counting points over finite fields. This method, pursued further by Desale & Ramanan (1975) led to an explicit inductive formula for the Betti numbers of the moduli space for arbitrary rank n. At this point we should comment that when the Chern class k is prime to n the moduli space $N(n,k)$ is compact and non-singular, and this is the case for which Betti numbers are computed. If $(n,k) \neq 1$ then the moduli space needs to be compactified and the geometry is more complicated.

The success of the Harder method depends on the fact that the moduli space $N(n,k)$ for a curve over a finite field F_q has another description, showing that it is the function field analogue of the classical moduli space for elliptic curves (i.e. the upper half plane divided by the modular group). In modern terminology $N(n,k)$ is a double coset space of an adèle group and counting points in $N(n,k)$ can be reduced to computing adèlic measures.

We come now to the third and most recent ingredient in the story, namely the Yang–Mills equations. These have become prominent in elementary particle physics and they have been studied both in Minkowski space, where they are of hyperbolic type, and in Euclidean 4-space where they are of elliptic type. In this latter case they have deep connections with three-dimensional algebraic geometry for which the interested reader may consult Atiyah (1979). The Yang–Mills equations can be formulated on any Riemannian manifold M and they depend on a fixed compact Lie group G. In dimension 2, i.e. when M is a surface, the equations are practically trivial and all solutions can easily be described. Despite this apparent triviality our first surprising observation was that, for a 2-sphere, the Yang–Mills equations for G essentially reproduced the Morse theory picture of ΩG. The Yang–Mills functional plays the role of the Energy and the explicit solutions correspond to the explicit geodesic structure of G. The space ΩG is replaced by the space of G-connections modulo (based) equivalence. Unlike the four-dimensional case studied in Atiyah (1979) where the Yang–Mills functional for $SU(2)$ appears only to have minima, in two dimensions there are critical points of arbitrarily high Morse index.

With this encouraging start it seemed natural to take the next step and investigate the Yang–Mills equations over a Riemann surface of arbitrary genus. The Narasimhan–Seshadri unitary approach fits naturally into this picture since the bundles arising from representations of $\pi_1(M)$ are easily seen to give the critical points, and the irreducible representations give the Yang–Mills minimum.

It seemed reasonable to hope that, as for the genus 0 case, we would have a perfect Morse function, i.e. that the critical point structure would correspond precisely to the homology. Comparison with the results of Newstead showed that this was not true in the naïve sense, but it eventually became apparent that if we used the full symmetry of the situation we should again have a perfect Morse theory. Technically this meant that we needed to use *all* bundle automorphisms not just *based* automorphisms. The lesson learnt from this example is of wider validity and in § 1 we begin with a general discussion of equivariant Morse theory, illustrated with some very simple examples. For an interesting application of these ideas see Kirwan (1982).

In the application of Morse theory to ΩG by Bott & Samelson (1958) the conclusions drawn related to the cohomology of the whole space, since the cohomology of the various critical manifolds was all known. In the Yang–Mills case the situation is different, in that the critical manifolds are complicated and we would like to reverse the procedure, using information about the whole space to deduce results on the critical manifolds. This procedure works for two reasons. In the first place the cohomology of the whole space can be easily computed by relating it in fact to $\Omega U(n)$ (or equivalently to the Yang–Mills situation for genus 0). Secondly the critical manifolds *other than the minimum* can all be expressed in terms of the minima for $U(m)$ with $m < n$, so that we can apply an inductive argument.

At this stage we reach the position that, provided the basic analysis works as expected, we have a perfect Morse theory and can inductively deduce information about the space of Yang–Mills minima, which by the Narasimhan–Seshadri theorem can be identified with the moduli space of stable bundles (in the coprime case $(n, k) = 1$). What has to be shown analytically is that the Yang–Mills paths of steepest descent always converge in a suitably strong sense to a critical point. We understand that Uhlenbeck (1982) has preliminary results in this direction that may do what is required. However, we have found an alternative presentation that is more direct and by-passes this question.

This alternative is a purely complex-analytic approach developed in § 7 and it begins with the

observation that the space $\mathscr{A}$ of *unitary connections* on a given C^∞ vector bundle E over a Riemann surface can also be viewed as the space $\mathscr{C}$ of all *holomorphic structures* on E. One can then define a stratification of $\mathscr{C}$ in which the one open stratum corresponds to semi-stable bundles and the other strata are described in terms of the canonical flags or filtrations introduced by Harder & Narasimhan (1975). Looked at equivariantly, relative to the group Aut (E) of automorphisms of E, this turns out to be a 'perfect' stratification and enables us to deduce information about the equivariant cohomology of the semi-stable stratum, and hence in the coprime case about the cohomology of the moduli space of stable bundles.

This complex approach is analogous to the use of complex cell decompositions to compute the cohomology of $P_n(C)$ and other homogeneous spaces. However, the stratification of $\mathscr{C}$ is not given by orbits of a group, except in the case of genus 0. Although technically independent of the Morse theory approach based on the Yang–Mills functional our complex approach was motivated by Morse theory and, as explained in § 8, it is essentially equivalent to it. By this we mean that, if the basic analytic facts of the Morse theory about convergence of trajectories are assumed, then our complex strata must coincide with the Morse strata, i.e. the stable manifolds of the critical sets.

The fact that stability in Mumford's sense and stability in Morse theory coincide in this situation is not accidental. As has been pointed out to us very recently by D. Mumford and S. Sternberg, this phenomenon occurs quite generally in the context studied in Mumford (1965) of reductive groups acting on Kähler manifolds. The novelty in our situation is that we have an infinite-dimensional example of this type, although the resulting moduli spaces are finite-dimensional. The key observation in all cases is that one should introduce the 'moment map' familiar in symplectic geometry. This point of view will be explained at the end of § 9.

The detailed results that our methods yield on the cohomology of the moduli space $N(n, k)$, in the coprime case, are described in § 9. First of all we obtain inductive formulae to calculate the Poincaré polynomials $P_t(N(n, k))$. These formulae are essentially the same as those obtained by the Harder–Narasimhan method and we shall comment on the comparison shortly. In addition, however, we prove that $N(n, k)$ has *no torsion* in its cohomology. We also prove the same thing for the moduli space $N_0(n, k)$ for stable bundles *with fixed determinant*, and we show that $N_0(n, k)$ is *simply connected*. Finally our methods give a natural and explicit set of multiplicative generators for the cohomology ring (theorem 9.11).

Although the number-theory approach of Harder–Narasimhan appears totally different from our geometric method there are close analogies, which are very intriguing. We discuss these analogies in detail in § 11.

We now review rapidly the contents of the sections not explicitly mentioned above. In § 2 we study the topology of the gauge group which from the Morse theory view-point determines the homotopy of the space on which the Yang–Mills function is naturally defined. Sections 3 and 4 develop basic general facts about the Yang–Mills equations while § 5 deals with the special case of Riemann surfaces. In § 6 we pursue the Yang–Mills solutions globally and show how they correspond essentially to (projective) unitary representations of the fundamental group. Up to this point we treat the general case of a compact Lie group G but in §§ 7 and 8 we concentrate on the unitary group $U(n)$ in order to make the connection with the theory of holomorphic vector bundles. However, we return to the general case in § 10, showing rather briefly how the whole theory extends to any G. The only notable difference is that we do not now get results about torsion: in fact the presence of torsion in G almost certainly implies torsion in the corresponding moduli spaces.

Sections 12 and 13 are both in the nature of technical appendixes. Thus in § 12 we review some elementary, though not widely known, facts about convexity and Lie groups. These play an important role in the partial ordering of the strata in our stratification of $\mathscr{C}$. An important notion that emerges in our analysis is that of a convex invariant function ϕ on the Lie algebra of a compact group. As we show in § 8 we get essentially the same theory if, in the definition of the Yang–Mills functional, we replace the norm-square $\| \ \|^2$ by ϕ. Finally § 13 summarizes facts about equivariant cohomology and in particular we formulate a result (proposition 13.4) that is used in § 1 to give a criterion (proposition 1.9) for a stratification to be 'equivariantly perfect'. This criterion is closely related to an argument due to Frankel (1959), which asserts that Morse functions arising from circle actions on Kähler manifolds are perfect.

It remains for us to make some comments about infinite-dimensional manifolds. The function-space manifolds that we shall meet such as the space of unitary connections or the space of maps of M into $U(n)$ can be given various topologies, depending on the class of functions we take. As long as our functions are at least continuous the homotopy type of the function spaces will be essentially the same. Technically it is usually convenient to work with Banach manifolds (so as to have the implicit function theorem) and one introduces Sobolev norms for this purpose. We explain in § 14 how this is done, much of it being fairly standard. In the main body of the paper we have ignored these technicalities and worked rather heuristically with smooth functions in order to concentrate on the geometrical ideas. Section 14 redresses the balance and provides the justification. Essentially this is a matter of establishing local regularity properties. For the global properties we need an additional argument and for this we fall back on algebro-geometric methods to which we devote § 15.

From this summary of the various sections it will be clear that not all sections are strictly necessary for the proof of our main results on the cohomology of moduli spaces of vector bundles. The proofs are essentially contained in §§ 1, 2, 7, 9, 13, 14 and 15.

We should perhaps point out that the theory of stable bundles over Riemann surfaces is only of real interest for genus $g \geqslant 2$. However, most of our discussion goes through for all values of the genus and is interesting even for $g = 0, 1$, from the Morse theory point of view. There are a few minor differences in the rational and elliptic case and we comment on these in the appropriate places.

Finally we should warn the reader that the level of exposition and sophistication is not uniform throughout the paper. Thus the first few sections are written at a more leisurely pace and make fewer demands on the reader. The technical requirements increase substantially in the later sections.

1. Equivariant Morse theory

We start with a brief review of the Morse theory of a non-degenerate smooth function f on a compact C^∞ manifold M.

Recall, first of all, that a critical point of f is a point p at which $\mathrm{d}f$ vanishes, and that at such a point the *Hessian*, $H_p f$, is a well defined quadratic form on $T_p M$, the tangent space to M at p. In local coordinates $\{x^i\}$ centred at p, the matrix of $H_p f$ relative to the base $\partial/\partial x^i$ at p is then given by

$$H_p f = \| \partial^2 f / \partial x^i \partial x^j \|$$

and p is called a non-degenerate critical point of f, if $\det H_p f \neq 0$.

At such a point the number of negative eigenvalues in a diagonalization of $H_p f$ is called the *index* of p (as a critical point of f) and is denoted by $\lambda_p(f)$.

Now with any function f all of whose critical points are non-degenerate we associate the *Morse counting-series*

$$M_t(f) = \sum_p t^{\lambda_p(f)}, \quad \mathrm{d}f_p = 0,$$

where the sum ranges over the necessarily finite number of critical points of f.

The Morse theory in its most elementary manifestation sets topological bounds for $M_t(f)$. Precisely, suppose that

$$P_t(M;K) = \sum t^i \dim H^i(M;K)$$

is the Poincaré series for M relative to a coefficient field K. Then if f is any non-degenerate function on M, its Morse series satisfies the following *Morse inequalities: there exists a polynomial $R(t)$ with non-negative coefficients*, such that

$$M_t(f) - P_t(M,K) = (1+t)\,R(t).$$

Thus in particular, the coefficients of $M_t(f)$ dominate those of $P_t(M)$. On the other hand setting $t = -1$ we see that $M_{-1}(f)$ always yields the *Euler number* $P_{-1}(M)$ of M.

We shall call a function f, a *K-perfect Morse-function* on M if

$$M_t(f) = P_t(M;K),$$

and call f perfect if this equality holds for all fields K.

Hence a perfect Morse function can exist only on a torsion-free manifold. In general it is of course difficult to decide whether a given f is perfect. However, there are two criteria for establishing 'perfection'.

First of all, *if the set $\{\lambda_p(f)\}$ of all indices of f contains no consecutive integers, then f is perfect.* This is the lacunary principle of Morse. For instance, if it can be shown that f has only even indices at its critical points, then this principle immediately yields the perfection of f, and this is the method that can be used to show that the Energy function on the space of loops of a Lie group is perfect (Bott & Samelson 1958).

Failing such a fortuitous disposition of the indices $\{\lambda_p(f)\}$, one has the 'completion principle' also used by Morse and already foreshadowed by Birkhoff's minimax principle.

Suppose then that p is a non-degenerate critical point p of f at level c, and of index λ_p. The 'Morse lemma' then asserts that in a suitable coordinate system $x_1, \ldots, x_n$ centred at p, the function f has near p the form

$$f = c - x_1^2 - x_2^2 - \ldots - x_k^2 + x_{k+1}^2 + \ldots + x_n^2,$$

where $k = \lambda_p(f)$. The set

$$\nu_p^- = \{x \mid x_1^2 + \ldots + x_k^2 \leqslant \epsilon, \quad x_{k+1} = \ldots = x_n = 0\}$$

is then a disc near p, whose boundary $\partial\nu_p^-$ is a $(k-1)$-sphere in the space

$$M_{c-\epsilon} = \{m \in M \mid f(m) \leqslant c - \epsilon\}.$$

We now call p '*completable*' if this sphere $\partial\nu_p^-$ bounds a singular chain in $M_{c-\epsilon}$ for small enough $\epsilon > 0$. With this understood one has the following:

Completion principle. *If f is non-degenerate and all its critical points are completable, then f is a perfect Morse function.*

Both these principles are easy consequences of the main structure theorem of the non-degenerate Morse theory. This theorem asserts that the sets $M_a = \{p \in M | f(p) \leqslant a\}$ change their homotopy type only at critical values of f and then only by the attaching of a cell of dimension $\lambda_p(f)$. Thus we have:

$M_a \sim M_b$ if there are no critical values between a and b, while $M_b \sim M_a \cup e_\lambda$ if there is a single critical point p of index λ in $M_b - M_a$.

From the standard exact sequences relating the cohomology of M_b and M_a under these circumstances one may then easily deduce the Morse inequalities as well as the completion principle, which we have just described.

So much then for a quick review of the Morse theory in its most elementary form. For our purposes we must now extend the concept of non-degeneracy of f in the following manner.

If $N \subset M$ is a connected submanifold of M, it will be called a *non-degenerate critical manifold* for f if and only if

$$\mathrm{d}f \equiv 0 \quad \text{along} \quad N \tag{1.1}$$

$$H_N f \text{ is non-degenerate on the normal bundle } \nu(N) \quad \text{of} \quad N. \tag{1.2}$$

Note that because of (1.1) the Hessian $H_N f$ of f is a well defined quadratic form on $\nu(N)$, so that (1.2) is the natural extension of the non-degeneracy hypothesis for critical points.

In the following a function on M will be called *non-degenerate* if its critical set is a union of *non-degenerate critical manifolds*. A prime example, which in a sense explains the virtue of this extension of the non-degeneracy concept, is the following.

Suppose $E \xrightarrow{\pi} M$ is a fibring and f a non-degenerate function on M in our new sense. Then it is easy to see that $\pi^* f$ on E is again non-degenerate in our new sense. On the other hand $\pi^* f$ will never have isolated critical sets unless E is a covering.

We next formulate the proper way to 'count' a non-degenerate critical manifold N of f. The recipe is as follows. We first endow $\nu(N)$ with a Riemannian metric. Then of course our Hessian $H_N f$ defines a canonical self-adjoint endomorphism

$$A_N : \nu(N) \to \nu(N)$$

by the formula

$$(A_N x, y) = H_N f(x, y), \quad x, y \in \nu(N).$$

The non-degeneracy of $H_N f$ now implies that the eigenvalues of A_N are all non-zero, and hence that A_N decomposes $\nu(N)$ into an orthogonal direct sum

$$\nu(N) = \nu^+(N) \oplus \nu^-(N)$$

spanned by the positive and negative eigenvalues of A_N respectively. We call the *fibre dimension of $\nu^-(N)$ the index of N* – as a critical manifold of f – and say that we are in the *orientable case* if this 'negative' bundle $\nu^-(N)$ is orientable. With this understood, and having chosen a coefficient field, K, we 'count' a non-degenerate critical manifold N of f with the polynomial

$$M_t(f, N) = \sum t^i \dim H^i_c\{\nu^-(N)\}$$

where now H^i_c denotes the compactly supported cohomology. In particular, by the Thom isomorphism, this polynomial reduces to

$$t^{\lambda_N} P_t(N)$$

in the orientable case, whereas in the non-orientable case $P_t(N)$ has to be computed relative to a twisted system of coefficients. This procedure turns out to be the proper one for 'counting' in the sense that if f is non-degenerate and $M_t(f)$ is defined by

$$M_t(f) = \sum_N M_t(f, N)$$

the summation extending over the critical manifolds of f, *then the Morse inequalities persist*, provided of course that the same coefficients are used on both sides of the equation. One may therefore speak of *K-perfect Morse functions also in this extended sense*. They are non-degenerate functions f, with

$$M_t(f) = P_t(M).$$

We have already remarked that the main advantage of this extended notion of non-degeneracy is its functorial nature under pull-back. Precisely, this amounts to the following.

PROPOSITION 1.3. *Let $E \xrightarrow{\pi} M$ be a smooth fibring. Then f is non-degenerate on M if and only if $\pi^* f$ is non-degenerate on E. Further the index of N as a non-degenerate critical manifold of M equals the index of $\pi^{-1}N$ as a critical manifold of E.*

The proof is self-evident, as $\pi^{-1}N$ is clearly a manifold if N is one, and its normal bundle in E is $\pi^{-1}\nu(N)$.

It remains to formulate the completion process in this extended context. The pertinent diagram is the following one:

$$\begin{array}{ccccc} H_*\{\nu_\epsilon^-(N)\} & \longrightarrow & \tilde{H}_*(\nu_\epsilon^-(N), \partial\nu_\epsilon^-(N)) & \xrightarrow{\partial} & \tilde{H}(\partial\nu_\epsilon^-(N)) \\ & & \pi^{-1}\uparrow & & \downarrow \\ & & H_{*-\lambda_N}(N) & \dashrightarrow & \tilde{H}(M_{f(c)-\epsilon}), \end{array} \tag{1.4}$$

where we have used the following notation.

We assume that $f(N) = c$, and write $\nu_\epsilon^-(N)$ for the set in the exponential image of $\nu^-(N)$ in M, where $f \geqslant c - \epsilon$. This will be an λ_N-disc-bundle over N, if $\epsilon > 0$ is small enough. We write π for the projection of this disc-bundle, so that π^{-1} corresponds to the Thom isomorphism and H for homology with coefficients in K. The $\tilde{H}$ denotes reduced homology. With this understood we say that N is *K-completable if the dashed arrow in* (1.4) *is zero*.

It is easy to check that this condition reduces to the previous one for a non-degenerate critical point of f, and again a standard argument implies the following:

Completion principle. *If all the critical manifolds of f are K-completable then f is a K-perfect Morse function on M.*

Remarks. Note that, as opposed to a critical point, a critical manifold can essentially be 'self-completing' in the following sense. By commutativity and the exactness of the horizontal sequence in (1.4), it is clear that a class $\alpha \in H_{*-\lambda_N}(N)$ certainly goes to zero under the dashed arrow if $\pi^{-1}\alpha$ is in the image of $H_*(\nu_\epsilon^-(N))$. Hence we call these classes N-completed. This phenomenon, of course, occurs only if the bundle $\nu^-(N)$ is non-trivial over N, and in a compact finite-dimensional setting it will not occur for all $\alpha \in H_*(N)$. However, in the infinite-dimensional or equivariant case, which we shall encounter in a moment, this will happen, and then one is in the fortuitous circumstance that we refer to as 'self-completing'.

We are now finally ready to discuss the question that is central for our considerations.

Suppose that f is a smooth function on M that is invariant under the smooth action of a Lie group G on M. When is such a function to be considered a perfect G-invariant function?

If G acts freely on M, it is clear enough that such an f should be considered perfect – as a G-invariant function – if the induced function on M/G is perfect. On the other hand if the action of G is not free, this procedure is certainly not correct, and one has to bring the different stability groups of the critical sets into play in some fashion or another. The manner of doing this that we advocate is the following one.

Consider any smooth principal G-bundle E over a base-manifold B, and the corresponding mixing diagram

$$\begin{array}{ccccc} E & \xleftarrow{\tilde{\alpha}} & E\times M & \xrightarrow{\tilde{\beta}} & M \\ \downarrow{\scriptstyle \pi'} & & \downarrow{\scriptstyle \pi} & & \downarrow{\scriptstyle \pi''} \\ E/G & \xleftarrow[\alpha]{} & E\times_G M & \xrightarrow[\beta]{} & M/G \end{array}$$

of the G-actions on M and E. Here of course the middle action is diagonal, so that under π, $(e\cdot g, m)$ is identified with $(e, g\cdot m)$.

Now because the action on E is free this diagonal action is also free. On the other hand a G-invariant f on M clearly lifts to a G-invariant f on $E\times M$, and hence *descends* to a smooth function f_E on $E\times_G M$. Now the space $E\times_G M$ is itself a fibre space over the base $B = E/G$ of E, and is of course the *bundle associated to E with fibre M*. In short then *every G-invariant function f naturally defines a function f_E on any smooth fibre* bundle with M as fibre and structure group G. Furthermore we have the following.

PROPOSITION 1.5. *If f is a non-degenerate function on M, then for every smooth principal G-bundle E, f_E is non-degenerate on $E\times_G M$. Furthermore, if N is a non-degenerate critical manifold of f on M, then f_E will have as corresponding critical manifold the space $E\times_G N$. Finally, the indices of N rel f and $E\times_G N$ rel f_E are equal.*

The proof is again self-evident in view of the functoriality of our concept. Indeed it is clear that

$$\pi^{-1}(E\times_G N) = \tilde{\beta}^{-1}N,$$

and now proposition 1.3 implies the rest.

Now there are very many different G-bundles but they are all induced from a *universal G-bundle* that is unique up to homotopy. Such a universal G-bundle is characterized by having its total space E *contractible*. It is then reasonable to say that our function f is perfect in the domain of G-invariant functions, or *G-equivariantly perfect* if the induced function f_E is perfect for the universal G-bundle E. In this universal case we shall simply write M_G for the space $E\times_G M$, BG for E/G and f_G for f_E.

To summarize then, this construction converts f into f_G, which is a function on the space M_G constructed functorially out of *the group G and its action on M*. In homotopy theory this is of course a well known procedure and in fact M_G is called the *homotopy quotient* of M by G. It has the following properties.

PROPOSITION 1.6. *If G acts freely on M (i.e. defines a fibration) then the natural map*

$$M_G \xrightarrow{\beta} M/G$$

is a homotopy equivalence. On the other hand M_G is always a fibring over BG with M as fibre, and its homotopy type depends only on the homotopy type of G and its action on M.

There is just one difficulty with this construction, and that is that in general EG and BG will not be realizable as *finite-dimensional* manifolds. Hence M_G is not usually a finite-dimensional

manifold. However, this is not a serious problem and can be overcome in several ways. When G is a compact Lie group, which is the only case of essential interest, BG can be realized as an infinite-dimensional manifold or as a suitable limit of finite-dimensional ones. In the former situation all sub-manifolds occurring will have finite codimension and cause no problems. Alternatively, and this is the point of view that we shall adopt, we can stick to our original manifold M and function f but introduce *equivariant cohomology* as the appropriate functor. By definition for every G-space X its equivariant cohomology $H_G(X)$ is *defined* by

$$H_G(X) = H(X_G).$$

In the category of G-spaces it has the usual properties of cohomology. In § 13 for the convenience of the reader we recall some of the basic facts about H_G and prove some particular results that we shall be needing.

To illustrate these ideas let us consider an example in which M is the 2-sphere

$$S^2 = \{(x, y, z) | x^2 + y^2 + z^2 = 1\}$$

in R^3 and let $f(x, y, z) = z$, be the *height function* on S^2. Also let $G = S^1$ be the group of rotations about the z-axis in R^3. Then f is clearly G-invariant, and also intuitively looks about as perfect as one could hope for. To construct an approximation to BS^1, let C^{l+1} be complex $(l+1)$-space and consider the action of S^1 on C^{l+1} given by

$$(z_0, \ldots, z_l) \to (e^{i\theta} z_0, \ldots, e^{i\theta} z_l).$$

Restricted to the unit sphere S^{2l+1} this action is free and gives rise to the Hopf fibring

$$\begin{array}{c} S^{2l+1} \\ {\scriptstyle S^1}\big\downarrow \\ P_l(C), \end{array}$$

with base space the complex projective l-space. Now

$$\pi_k(S^{2l+1}) = 0 \quad \text{for} \quad k < 2l+1.$$

Hence this sequence of finite-dimensional fibrings approximates the universal one, which for S^1 may be taken to be the fibring of the unit sphere $S(H)$ in a Hilbert space H over the space $P(H)$ of rays in H.

Let us now consider the spaces

$$M_l = S^{2l+1} \times_{S^1} S^2.$$

They are the finite-dimensional approximations to M_{S^1}, and are naturally 2-sphere bundles over $P_l(C)$. We have schematically indicated this below:

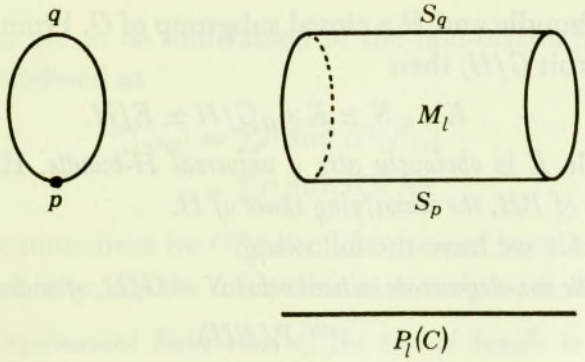

FIGURE 1.

Note that for the fixed point p (or q)

$$p \times_{S^1} S^{2l+1} \cong P_l(C).$$

Thus *the critical manifolds corresponding to p and q, are the sections S_p and S_q in M_l indicated above.* Hence, since

$$P_t(P_l(C)) = 1 + t^2 + \ldots + t^{2l}$$

and the index of f on S^2 is 0 at p and 2 at q, we obtain the formula

$$M_t(f_l) = (1+t^2)(1+t^2+\ldots+t^{2l})$$

for the Morse series of the function f_l induced by f on S^{2l+1}.

When $l \to \infty$, this polynomial therefore becomes the *formal power series*

$$M_t(f_{S^1}) = (1+t^2)/(1-t^2).$$

Now cohomologically, the fibring M_l over $P_l(C)$ is trivial (2-sphere bundles always are). Hence by the Künneth formula

$$P_t(M_l) = (1+t^2)(1+t^2+\ldots+t^{2l})$$

and so we see by inspection that our f_l was indeed perfect, in fact not only on M_{S^1} but also *in each approximation M_l.*

Let us next modify f so that it has a maximum at p and at q, and a minimum along the equator, but still keeping it $G = S^1$-invariant. Then, p and q both contribute

$$t^2(1+t^2+\ldots+t^{2l})$$

to $M_t(f_l)$. On the other hand the critical set in M_l corresponding to the equator on S^2, is given by

$$S^1 \times_{S^1} S^{2l+1} \simeq S^{2l+1}.$$

Hence it contributes $1+t^{2l+1}$. Thus

$$M_t(f_l) = (2t^2)(1+t^2+\ldots+t^{2l}) + (1+t^{2l+1}).$$

These functions are therefore not perfect for any particular l. On the other hand, letting $l \to \infty$, we obtain

$$\begin{aligned} M_t(f_{S^1}) &= 1 + 2t^2 + 2t^4 + \ldots \\ &= (1+t^2)/(1-t^2), \end{aligned}$$

so that this new f is again perfect according to our definition, i.e. on M_{S^1}.

This example illustrates two phenomena. First of all that a perfect f_G on M_G need not come from a perfect f on M. It also shows that in some sense the larger the orbit of a critical set of f on M, the 'smaller' its contribution is in M_G. The precise formulation of this principle is as follows.

First of all recall the identity

$$E/H \simeq E \times_G G/H,$$

when E is a principal G-bundle and H a closed subgroup of G. From this it follows immediately that if $N \subset M$ is the G-orbit G/H, then

$$E \times_G N \simeq E \times_G G/H \simeq E/H.$$

But a universal G-bundle E is *obviously also a universal H-bundle.* Hence in the universal case *E/H has the homotopy type of BH, the classifying space of H.*

Hence by proposition 1.5 we have the following

Counting principle. *The non-degenerate critical orbit $N = G/H$, of index $\lambda(N)$ for f on M, contributes*

$$t^{\lambda(N)} P_t(BH) \tag{1.7}$$

to $M_t(f_G)$, the 'counting series' of f_G on M_G.

Note also that if H is connected then the $P_t(BH)$ in (1.7) can be taken literally, whereas if H is not then a local coefficient system might still be needed.

More precisely, the correspondence (1.7) can still be refined in one very important respect. Namely if N is the orbit of G through q, then H the stability group of q acts on the normal space to N at q and, using an H-invariant metric, also on the *negative normal space* $\nu_q^-(N)$. It follows that $\nu^-(N)$ is associated with the principal bundle $G/H = N$, via this representation, and correspondingly that $\nu^-(BH)$ is associated with the universal H-bundle EH over BH, by that same representation.

Thus we arrive at the following refinement of (1.7).

Equivariant correspondence principle. *Under the correspondence* $N_q \mapsto BH$ *of* (1.7), *the negative bundle of* f_G *along* BH *becomes the bundle associated with the universal* H*-bundle via the 'negative representation':*

$$\lambda_N \colon H \to \operatorname{Aut} \nu_q^-(N). \tag{1.8}$$

Remark. We have here used the same notation λ_N for the index of N, and the negative representation for obvious reasons; and in the future the context will make it clear which is meant.

The importance of (1.8) is that standard methods allow one to compute the characteristic classes in $H^*(BH)$ for bundles associated with representations, and one may therefore use (1.8) to compute to what extent the critical set BH for f_G is 'self-completing'. For instance, in the 2-sphere example, for the critical point q, which is the maximum of f, we find that $H = S^1$, and that λ_p is the standard representation of S^1 on $\mathbb{R}^2$. The Euler class $e(\lambda_N)$ of $\nu^-(BH)$ is therefore a generator of $H^2(BS^1)$ and hence *generates* $H^*(BS^1)$ *multiplicatively*. It follows that *multiplication by* $e(\lambda_N)$ *induces an injection of* $H^*(BS^1)$ *into* $H^*(BS^1)$ *for any coefficient system*. Dually, this implies precisely that $H_*(\nu_\epsilon^-(BH))$ maps *onto* $H_*(\nu_\epsilon^-, \partial\nu_\epsilon^-)$ in the diagram (1.4), i.e. it implies that BH is *self-completing, as a critical set of* f_G. Now as the minimum is always self-completing (the condition of (1.4) becomes vacuous), it follows that we have in this instance established the 'perfection' of f_G by *purely local considerations* as opposed to our earlier global proof of the same fact. This state of affairs turns out to be the one we shall encounter for the Yang–Mills functional. For future reference we therefore formalize this principle in the following.

COROLLARY. *If in* (1.8) *the Euler class of* λ_N *induces an injection of* $H^*(BH)$ *into itself for a coefficient system* K *then, as a critical set of* f_G, BH *is self-completing relative to* K.

There now remains only one more appropriate extension of these concepts. In the domain of G-invariant functions, the formula (1.7) corresponds to a non-degenerate critical *point*. For a non-degenerate critical manifold N the contribution to the equivariant Morse series $M_t(f_G)$ is

$$t^{\lambda(N)} P_t(N_G)$$

and again local coefficients are to be understood in the non-orientable case. The equivariant Poincaré series is of course defined as

$$\begin{aligned} P_t(N_G) &= \Sigma t^i \dim H^i(N_G) \\ &= \Sigma t^i \dim H_G^i(N) \end{aligned}$$

and we shall also denote it sometimes by $GP_t(N)$. The normal bundle to N has an equivariant Euler class and as before we have, for the orientable case and any field K,

PROPOSITION 1.9. *If the equivariant Euler class of the normal bundle to* N *is not a zero-divisor in* $H_G^*(N, K)$, *then* N *is equivariantly self-completing for* K.

If all critical manifolds satisfy the hypothesis of (1.9) then f will be equivariantly perfect over K, so that the equivariant Morse and Poincaré series coincide.

In §13 we establish a useful sufficient criterion for (1.9) to hold. This criterion (see proposition 13.4) involves simply the isotropy group structure of the action of G and is easy to verify. It provides the key to the perfect nature of the Yang–Mills functional, which we shall be explaining in subsequent sections.

So far we have concentrated exclusively on the homological aspects of Morse theory. There is however more detailed geometrical information about the structure of our manifold M that is provided by a function f. If we introduce a Riemannian metric on M we can define the vector field grad f, dual to the differential df. The 'gradient flow' of f is then given by the paths of steepest descent, i.e. the trajectories of $-\text{grad}f$. If f has only non-degenerate critical points p then every trajectory converges to some p, and the set of all points on trajectories converging to a given p form a cell $M^+(p)$. This cell is called the stable manifold of p since f, restricted to p, has an absolute minimum at p. Similarly, replacing f by $-f$ we get another cell $M^-(p)$ called the unstable manifold of p. The dimension of $M^-(p)$ (or the codimension of $M^+(p)$) is equal to the Morse index of p. Thus f defines a cell decomposition

$$M = \bigcup_p M^+(p) \tag{1.10}$$

and the Morse inequalities follow at once by using these cells to compute the homology of M.

More generally if there are non-degenerate critical manifolds N we have stable manifolds $M^+(N)$ that are cell-bundles over N and we get a stratification

$$M = \bigcup_N M^+(N), \tag{1.11}$$

which we shall call the *Morse stratification*.

One easy consequence of this stratification, which goes beyond homology, is the following:

PROPOSITION 1.12. *Let N_0 be the manifold giving the absolute minimum of f and assume that, for all other critical manifolds, the Morse index is $\geqslant 3$. Then, if M is connected, N_0 is also connected and we have an isomorphism of fundamental groups*

$$\pi_1(N_0) \cong \pi_1(M).$$

For the equivariant case if f is G-invariant, where G is a compact Lie group, we can always pick a G-invariant metric. Then the gradient flow is G-invariant so that the stratification (1.11) is G-invariant. The equivariant analogue of (1.12) holds but is in fact equivalent to it because the fibration $M \to M_G \to BG$ gives an exact sequence

$$\to \pi_1(M) \to \pi_1(M_G) \to \pi_1(BG) \to$$

and there is a similar one with N_0 replacing M.

The critical manifolds N of our function f have a natural partial ordering. We first define a pre-ordering $\prec$ by

$$N_1 \prec N_2 \Leftrightarrow \text{the boundary of } M^+(N_1) \text{ intersects } M^+(N_2).$$

By following the trajectories of grad$f$ it is then easy to show that

$$N_1 \prec N_2 \Rightarrow \text{there is a trajectory of grad}f \text{ starting on } N_1 \text{ and passing within } \epsilon \text{ of } N_2.$$

Here ϵ is any positive constant. In particular taking ϵ to be less than $f(N_2) - f(N_1)$ it follows that

$$N_1 \prec N_2 \Rightarrow f(N_1) < f(N_2).$$

Hence the transitive relation $<$ generated by $\prec$ is a partial ordering and has the property that

$$\text{closure of } M^+(N) \subset \bigcup_{N' \geqslant N} M^+(N'). \tag{1.13}$$

Sometimes one may be given an explicit finite stratification of M

$$M = \bigcup_{\lambda} M_\lambda, \tag{1.14}$$

where each M_λ is a locally closed submanifold of M, and the index set of λ is partially ordered so that

$$\bar{M}_\lambda \subset \bigcup_{\mu \geqslant \lambda} M_\mu \tag{1.15}$$

holds for all λ (we assume the partial ordering is strict, i.e. $\lambda \leqslant \mu$ and $\mu \leqslant \lambda$ implies $\lambda = \mu$) One can then use the stratification to get Morse-type information about the homology of M. We start with the open strata, given by minimal λ, and inductively add other strata. At each stage we can write down the exact cohomology sequence for a pair $(U, U-V)$ where V is a closed submanifold of U. More formally this can be described as follows. Define a subset I of indices to be

$$\begin{aligned} &\textit{open} && \text{if } \lambda \in I \text{ and } \mu \leqslant \lambda && \Rightarrow \mu \in I \\ &\textit{closed} && \text{if } \lambda \in I \text{ and } \mu \geqslant \lambda && \Rightarrow \mu \in I. \end{aligned}$$

It is easy to check that I is closed if and only if its complement I' is open. Moreover the subspace of M defined by

$$M_I = \bigcup_{\lambda \in I} M_\lambda$$

is open or (or closed) if I is open (or closed): this follows from (1.15). If I is open and $\lambda \in I'$ is minimal then $J = I \cup \lambda$ is open, and our inductive step is from M_I to M_J. From (1.15) it follows that $M_\lambda = M_J - M_I$ is a closed submanifold of M_J. Assuming for simplicity that the normal bundles to all strata in M are orientable we have the exact sequence

$$\to H^{q-k}(M_\lambda) \to H^q(M_J) \to H^q(M_I) \to, \tag{1.16}$$

where we have used the Thom isomorphism

$$H^{q-k}(M_\lambda) \cong H^q(M_J, M_I)$$

with $k = k_\lambda = \operatorname{codim} M_\lambda$.

If, for a given field K of coefficients, (1.16) breaks up into short exact sequences for all q and all λ it follows that

$$P_t(M) = \Sigma t^{k_\lambda} P_t(M_\lambda).$$

In such a case we shall say that the stratification is *perfect* over K. If this holds for $K = Z_p$, for all primes p, we shall simply call it perfect. Thus a perfect Morse function defines a perfect stratification.

If the stratification is G-invariant and the corresponding equivariant cohomology sequences break up we shall call the stratification equivariantly or G-perfect. Proposition 1.9 has an obvious analogue in this context with the normal bundle in question being the normal to a stratum.

Examples of manifolds with naturally arising stratifications are the flag manifolds G/T, where $T \subset G$ is a maximal torus. Using the complexification G^c of G one also has a complex description,

namely as G^c/B where B is a Borel subgroup. The left action of B on G^c/B then has finitely many orbits. These are the Bruhat cells and they give a (complex) stratification of the flag manifold. The loop space ΩG has also such a stratification (Pressley 1980) and we shall meet other examples in dealing with the space of G-connections on a Riemann surface. The last two examples are both infinite-dimensional, but the strata will have *finite codimension*. The indexing sets will be countably infinite but will have the following finiteness property.

For every finite subset I there are a finite number of minimal elements of the complement I' (1.17)

so that our inductive procedure still applies. Although the induction never terminates, only finitely many steps will be needed to compute $H^q(M)$ for any given q provided the stratification satisfies the following further finiteness condition.

For each integer q there are only finitely many indices $\lambda \in I$ for which codim $M_\lambda < q$. (1.18)

Thus when (1.17) and (1.18) are satisfied we may proceed to compute the cohomology of M as in the finite-dimensional case.

Sometimes we may be given a stratification of M and a function f and we might like to know if the stratification is the Morse stratification (by stable manifolds) arising from f (for some metric on M). Thus for the flag manifold one has natural functions arising from considering G/T as an orbit in the Lie algebra of G and restricting a linear function. It is not hard to axiomatize the Morse stratification, and one can then test any given stratification to see whether the axioms are satisfied. We shall prove the following.

PROPOSITION 1.19. *Let $f: M \to R$ have only non-degenerate critical manifolds N_λ and let $M = \bigcup_\lambda M_\lambda$ be a stratification by disjoint locally closed submanifolds M_λ, such that, for some partial ordering on the set of λ, the following properties hold*:

(1) $\lambda \leqslant \mu \Rightarrow f(\lambda) \leqslant f(\mu)$,
(2) $\bar{M}_\lambda \subset \bigcup_{\mu \geqslant \lambda} M_\mu$,
(3) grad f *at any $x \in M$ is tangential to the M_λ containing x.*
(4) $N_\lambda \subset M_\lambda$.
(5) index N_λ = codim M_λ.

Then M_λ is the stable manifold S_λ of N_λ so that we have the Morse stratification.

Proof. We have only to show that the trajectory $x(t)$ of $-\text{grad} f$ through any point x of M_λ converges to N_λ as $t \to \infty$. Now (3) guarantees that $x(t)$ remains in M_λ for all finite t and (2) implies that $x(\infty) \in N_\mu$ for some $\mu \geqslant \lambda$. Now if x is sufficiently close to N_λ the trajectory $x(t)$, as $t \to \infty$, either converges to N_λ or 'falls below N_λ' (this basic fact is needed to establish the existence and properties of the S_λ and is formally a consequence of (1.13)). Since $x(\infty) \in N_\mu$ for some $\mu \geqslant \lambda$ property (1) shows that $x(\infty)$ cannot be below N_λ and so $x(\infty) \in N_\lambda$. Thus locally near N_λ we have $M_\lambda \subset S_\lambda$. By (5) we see that dim M_λ = dim S_λ and so (near N_λ) M_λ is an open set of S_λ. By (4) we see that M_λ and S_λ must coincide near N_λ. Now return to a general point $x \in M_\lambda$ with $x(\infty) \in N_\mu$. Then for large t, $x(t)$ is close to N_μ and in S_μ. Hence by what we have just proved (with μ for λ) $x(t)$ lies in M_μ for large t. On the other hand $x(t) \in M_\lambda$ for all finite t. Since different M_λ are disjoint this implies $\mu = \lambda$ and completes the proof.

This proposition can be applied for example to the flag manifolds to show that the Bruhat cells coincide with the Morse cells of an appropriate function. For a detailed discussion of this

and related questions refer to Atiyah (1982). Proposition 1.19 also applies on the infinite-dimensional manifold ΩG to show that the Birkhoff cells (see Pressley 1980) are the Morse cells of the Energy function (see Bott & Samelson (1958) for a discussion of this case).

In the Yang–Mills situation, which we shall be treating in this paper, we shall exhibit a stratification satisfying properties (1)–(5) of (1.19). If the Morse strata exist, i.e. if one can prove good properties about the trajectories $x(t)$ as $t \to \infty$, then (1.19) will identify the Morse strata with our strata. However, there are analytical difficulties involved here because the manifold M is infinite-dimensional and the critical sets N have singularities. We shall therefore by-pass these difficulties by simply using our stratification directly to compute cohomology. The connection with Morse theory is then left at a slightly conjectural level, but this is of no consequence for the topological applications.

2. The topology of the gauge group

Throughout this section all maps, bundles, sections and other objects will be taken as smooth, i.e. of class C^∞. From the point of view of homotopy theory this gives essentially the same result as the continuous maps and we shall on occasion blur the distinction when we are dealing with homotopy computations.

If P is a principal G-bundle over X, Ad P shall denote the bundle associated to P with fibre G, the action of G on itself however being the adjoint one. Thus

$$\mathrm{Ad}\, P = P \times_G G$$

is *not* a principal bundle any more, but rather a bundle of groups over X, whose sections can be identified with maps $f: P \to G$ satisfying

$$f(pg) = g^{-1} f(p)\, g. \tag{2.1}$$

The space of such sections $\Gamma \mathrm{Ad}\, P$ forms a group under pointwise multiplication and this is by definition, the 'gauge group' $\mathcal{G}(P)$ of P:

$$\mathcal{G}(P) = \Gamma \mathrm{Ad}\, P. \tag{2.2}$$

This group acts naturally as a group of G-equivariant maps of P, which cover the identity map of X. It can in fact be identified with the group of such automorphisms:

$$\mathcal{G}(P) \simeq \mathrm{Aut}\,(P). \tag{2.3}$$

To see this let $f: P \to G$ represent a section of Ad P. Then define

$$f_*: P \to P$$

by

$$f_*(p) = p \cdot f(p).$$

The relation (2.1) then shows that f_* is equivariant and covers the identity. Conversely given a map $f_*: P \to P$ covering the identity, f_* defines a unique map $f: P \to G$ such that

$$f_*(p) = p \cdot f(p),$$

and now G-equivariance forces the relation (2.1) on f. This establishes (2.3).

The purpose of this section is to describe the topology of the classifying space $B\mathcal{G}(P)$ of the gauge group when M, the base space of P, is a compact Riemann surface, and G is the unitary group.

PROPOSITION 2.4. *Let BG be the classifying space for G. Then in homotopy theory*

$$B\mathcal{G}(P) = \mathrm{Map}_P(M, BG).$$

Here the subscript P denotes the component of a map of M into BG which induces P.

Proof. Let

$$G \to EG \to BG$$

by a universal bundle for G, and consider the space $\mathrm{Map}_G(P, E)$ of G-equivariant maps of P to E.

The group $\mathcal{G}(P)$ now acts naturally on this space by composition, to yield the principal fibring

$$\mathcal{G}(P) \longrightarrow \mathrm{Map}_G(P, E) \xrightarrow{\pi} \mathrm{Map}_P(M, BG).$$

If BG is paracompact and locally contractible, which is easily arranged, π will be a locally trivial principal fibring, as follows easily from the homotopy properties of fibrings. The total space $\mathrm{Map}_G(P, E)$ is contractible so that this is a universal bundle for $\mathcal{G}(P)$, and

$$B\mathcal{G}(P) = \mathrm{Map}_P(M, BG)$$

as was asserted.

Using (2.4) we now compute $B\mathcal{G}(P)$ for the cases we have in mind.

Case I. *The unitary group* $U(1)$

The group $U(1)$ is the circle S^1 of complex numbers of norm 1. These act naturally on the unit sphere $S(H)$ of a Hilbert space H over C, and the quotient space

$$P(H) = S(H)/S^1$$

is the projective space of rays in H. When $\dim H = \infty$, $S(H)$ becomes contractible, and hence

$$S^1 \longrightarrow S(H) \xrightarrow{\pi} P(H) \tag{2.5}$$

is a universal S^1-bundle. From the corresponding exact homotopy sequence it now follows that

$$\pi_k\{P(H)\} = 0 \quad \text{for} \quad k \neq 2,$$

while

$$\pi_2\{P(H)\} = Z.$$

Thus $P(H)$, which is the BG in this case, is an Eilenberg–Maclane space $K(Z; 2)$. Now it is a theorem of René Thom that, if Y is such a space, and X any finite complex, then $\mathrm{Map}\,(X; Y)$ is again a product of such spaces. Precisely,

THEOREM (*Thom*). *Let $\pi_q(Y) = 0$ for $q \neq n$ and let $\pi_n(Y) = \pi$. Then*

$$\mathrm{Map}\,(X, Y) = \prod_q K\{H^q(X, \pi); n - q\}. \tag{2.6}$$

For a Riemann surface M of genus g this yields the corollary:

$$\mathrm{Map}\,(M; BS^1) = Z \times \underbrace{S^1 \times \ldots \times S^1}_{2g} \times P(H)$$

corresponding to the fact that

$$H^1(M; Z) \cong Z \oplus \ldots \oplus Z \quad (2g \text{ factors}),$$
$$H^0(M; Z) \cong H^2(X; Z) \cong Z.$$

In particular then, we see that in this case $B\mathcal{G} = B\mathcal{G}(P)$ has *no torsion* and its Poincaré series

$$P_t(B\mathcal{G}) = \sum_{i \geqslant 0} \dim H^i(B\mathcal{G}; Q)\, t^i$$

is given by

$$P_t(B\mathcal{G}) = (1+t)^{2g}/(1-t^2). \tag{2.7}$$

When P is a $U(1)$-bundle over S^4, on the other hand, the above recipe only yields

$$B\mathcal{G} = P(H). \tag{2.8}$$

Remark. The gauge group $\mathcal{G}(P)$ as we have defined it here really corresponds to the 'local gauge group' as this term is used in physics. The global gauge group corresponds to the 'constant sections' of Ad P, which are given by the *centre* Z of G. Indeed every $z \in Z$ is invariant under the adjoint action of G on itself and hence induces a canonical section f_z of Ad P, given by $f_z(p) = z$.

We next consider the more general

Case II. $G = U(n)$, $n > 1$.

Now it is no longer true that BG is an Eilenberg–Maclane space. However, over the *rationals* Q, BG is simply a product of Eilenberg–Maclane spaces:

$$BU(n) \underset{Q}{\simeq} K(Z;2) \times K(Z;4) \times \ldots \times K(Z;2n).$$

Indeed each Chern class $c_i \in H^{2i}(BU(n), Z)$ induces a map.

$$BU(n) \xrightarrow{c_i^*} K(Z;2i)$$

and, since $H^*(BU(n))$ is the polynomial ring in $c_1, \ldots, c_n$, the product of these maps induces a Q-equivalence of these spaces. Hence over Q we may apply Thom's theorem as before, at least to compute the Poincaré series of a component of Map (M, BG). Further as these behave multiplicatively under products, it is enough to take each $K(Z; 2k)$ at a time. Now

$$P_t\{K(Z;2k)\} = 1/(1-t^{2k}),$$

while

$$P_t(K(Z;2k-1)) = 1+t^{2k-1}.$$

Hence

$$P_t \operatorname{Map}(M, K(Z;2k)) = (1+t^{2k-1})^{2g}/(1-t^{2k-2})(1-t^{2k}) \quad \text{for} \quad k \geqslant 2.$$

Together with (2.7) for the case $k = 1$ this yields

$$P_t\{\operatorname{Map}_P(M; BU(n))\} = \prod_{k=1}^{n} (1+t^{2k-1})^{2g} \Big/ \left\{\prod_{k=1}^{n-1} (1-t^{2k})^2\right\} (1-t^{2n}) \tag{2.9}$$

for any component, i.e. any P over M.

Actually more is true.

PROPOSITION 2.10. *The space* Map $_P(M, BU(n))$ *under consideration, is free of torsion.*

Proof. To see this we have to come to grips with the fibrings that lie behind Thom's theorem. First recall that a compact Riemann surface M, can be obtained from a wedge of $2g$ circles by attaching a 2-cell. This implies that there is a cofibration

$$\bigvee_{2g} S^1 \to M \to S^2, \tag{2.11}$$

which, by the exactness of the mapping functor, gives rise to the fibring

$$\begin{array}{ccc} \operatorname{Map}^*(S^2, BU(n)) & \longrightarrow & \operatorname{Map}^*(M; BU(n)) \\ & & \downarrow \\ & & \operatorname{Map}^*(\bigvee S^1, BU(n)). \end{array} \tag{2.12}$$

Here the $*$ denotes base-point-preserving maps, which, by the same principle, are linked to the unrestricted maps via the fibring

$$\begin{array}{ccc} \mathrm{Map}^*(M; BU(n)) & \longrightarrow & \mathrm{Map}(M, BU(n)) \\ & & \downarrow \\ & & BU(n). \end{array} \tag{2.13}$$

We wish to show that both these fibrations are homologically trivial, and that all spaces involved are torsion-free.

Now recall that $BU(n)$ and its loopspace $\Omega BU(n) = U(n)$ and its second loopspace $\Omega U(n)$ are all torsion-free. Hence after (2.12) is rewritten with the standard identifications, one obtains

$$\begin{array}{ccc} \Omega U(n) & \longrightarrow & \mathrm{Map}^*(M, BU(n)) \\ & & \downarrow \\ & & U(n) \times \ldots \times U(n) \quad (2g \text{ factors}) \end{array} \tag{2.14}$$

with both fibre and base torsion-free. Hence any non-trivial homology-twisting, i.e. a non-zero differential in the spectral sequence, or a non-trivial coefficient system, would be detectable over Q and produce a Poincaré polynomial for the middle term that would be smaller than the product of the Poincaré polynomials of the factors. On the other hand by Thom's theorem, applied to pointed maps, it must be the product. This completes the proof.

The same argument now applies to (2.13) and we are done.

To recapitulate, we have established

THEOREM 2.15. *Let P be a $U(n)$-bundle over the compact Riemann surface. M Then if $\mathcal{G} = \mathcal{G}(P)$ is the gauge group of P, $B\mathcal{G}$ is torsion-free and has Poincaré series*

$$P_t(B\mathcal{G}) = \prod_{k=1}^{n} (1+t^{2k-1})^{2g} \Big/ \left\{ \prod_{k=1}^{n-1} (1-t^{2k})^2 \right\} (1-t^{2n}).$$

In the course of our proof we have also shown that, in the fibration (2.14), the fundamental group of the base, namely

$$\Gamma = \pi_1(U(n)^{2g}) \cong Z^{2g} \cong H_1(M, Z),$$

acts trivially on the cohomology of the fibre $\Omega U(n)$. This implies that the cohomology is unaltered on lifting to a finite covering corresponding to a subgroup Γ' of finite index in Γ. Moreover from (2.13) and (2.14) we see that

$$\pi_1(B\mathcal{G}) \cong \pi_1(U(n)^{2g}) \cong \Gamma.$$

But $\pi_1(B\mathcal{G}) \cong \pi_0(\mathcal{G})$ is the group of components of $\mathcal{G}$. Hence a subgroup Γ' of Γ of finite index corresponds to a subgroup $\mathcal{G}'$ of $\mathcal{G}$ of finite index and so we have

PROPOSITION 2.16. *In the situation of theorem (2.15), for any subgroup of $\mathcal{G}'$ of $\mathcal{G}$ of finite index, $B\mathcal{G}'$ is torsion-free and has the same Poincaré series as $B\mathcal{G}$.*

We shall now describe a way of producing explicit generators for the integral cohomology of $B\mathcal{G}$. This will eventually enable us to describe corresponding generators for the cohomology of the moduli space of stable bundles. It also provides an independent proof of the cohomological triviality of the fibrations (2.12) and (2.13) without appealing to Thom's theorem.

We begin by considering the natural evaluation map

$$e : \mathrm{Map}(M, BU(n)) \times M \to BU(n).$$

Pulling back the universal vector bundle over $BU(n)$ we then get a vector bundle V over $B\mathcal{G} \times M$. Since M has no torsion the Künneth formula gives, for integer cohomology,

$$H^{2r}(B\mathcal{G} \times M) \cong H^{2r}(B\mathcal{G}) \oplus H^{2r-1}(B\mathcal{G}) \otimes H^1(M) \oplus H^{2r-2}(B\mathcal{G}) \otimes H^2(M). \tag{2.17}$$

Taking the Chern class $c_r(V)$ and decomposing it, we get from this Künneth decomposition, elements

$$a_r \in H^{2r}(B\mathcal{G}),$$
$$b_r^j \in H^{2r-1}(B\mathcal{G}), \quad j = 1, \ldots, 2g,$$
$$f_r \in H^{2r-2}(B\mathcal{G}),$$

relative to a basis α_j of $H^1(M)$. These will give rational generators, but to get integral generators we need to replace the f_r by elements d_r constructed as follows. We introduce K-theory instead of cohomology and analogous to (2.17) we have a Künneth formula

$$K^0(B\mathcal{G} \times M) \cong K^1(B\mathcal{G}) \otimes K^1(M) \oplus K^0(B\mathcal{G}) \otimes K^0(M). \tag{2.18}$$

Now we have

$$K^0(M) \cong Z \oplus Z$$

with two canonical generators, the first given by the trivial line-bundle and the second by the reduced line-bundle of Chern class 1 (i.e. $H-1$, where H is the line-bundle and 1 the trivial line-bundle). Starting with the class of V in $K^0(B\mathcal{G} \times M)$ and projecting onto the second component then gives an element $W \in K^0(B\mathcal{G})$. An alternative description of W is to say that

$$W = f_!(\tilde{V})$$

where $f: B\mathcal{G} \times M \to B\mathcal{G}$ is the projection, $f_!$ is the direct image map in K-theory and $\tilde{V} = V - 1 \otimes V_0$ with $V_0 = V|B\mathcal{G} \times \text{point}$. Since $f_!(1) = 1-g$ it follows that

$$W = f_!(\tilde{V}) = f_!(V) + (g-1)\,V_0. \tag{2.19}$$

Finally, taking the Chern classes of W we get an infinite sequence of elements

$$e_r \in H^{2r}(B\mathcal{G}), \quad r = 1, 2, \ldots.$$

We shall now prove

PROPOSITION 2.20. *The elements a_r, b_r^j, e_r constructed above are multiplicative generators of the integral cohomology ring of $B\mathcal{G}$. The elements a_r, b_r^j, f_r are multiplicative generators of the rational cohomology ring.*

Remark. The a_r are the Chern classes of V_0 so that by (2.19) we get another set of generators by replacing the e_r by the Chern classes d_r of $f_!(V)$. These will occur more naturally in algebraic geometry.

The three types of element will in fact provide generators for the three factors in the fibration decomposition (2.13) and (2.14). Clearly the a_r give the Chern classes of $BU(n)$ and so generate its cohomology. The classes b_r^j (for fixed j) are easily seen to give the generators for the cohomology of the jth factor $U(n)$ in $U(n)^{2g}$. It remains to show that the elements e_r give generators for the cohomology of $\Omega U(n)$. Now we have a natural stabilization map

$$i: \Omega U(n) \to \Omega U,$$

where $U = \lim_{m\to\infty} U(n)$ is the stable unitary group. The periodicity theorem gives a homotopy equivalence

$$\Omega U \sim Z \times BU$$

so that $H(\Omega U)$ is a polynomial ring on all the universal Chern classes $c_1, c_2, \ldots$. Pulling these back by i we get classes in $H(\Omega U(n))$. These coincide with the classes e_r introduced above in view of the relation between K-theory and the periodicity theorem. To prove (2.20) it remains therefore to show that i^* is surjective in cohomology, or in steps that the inclusions

$$j: \Omega U(n) \to \Omega U(n+1)$$

have this property for all n. This can easily be deduced from the explicit description of $H^*(\Omega U(n))$ given by Bott (1958). This completes the proof for the integral cohomology. Over the rationals the proof is similar but easier.

As mentioned earlier our proof of (2.20) produced cohomology classes in the total spaces of the fibrations (2.13) and (2.14), which generated the cohomology of the fibres in each case. This gives an independent proof of their triviality.

Over the rationals we can use the Chern character to compare (2.17) and (2.18). This enables us to express the infinite sequence of e_r in terms of n elements f_r. The fact that the e_r are *integral* then leads to an infinite sequence of integrality relations involving polynomials (with rational coefficients) in the f_r.

Finally we shall derive a result that concerns the role of the constant $U(1)$-subgroup of $\mathcal{G}$ representing the central automorphisms of P. This will play an important role in §9 when we study the cohomology of the moduli spaces. We shall prove the following.

PROPOSITION 2.21. *Assume that the Chern class k of P and the rank n are coprime. Then the inclusion of the constant central $U(1)$ in $\mathcal{G}$ induces a surjection*

$$H^2(B\mathcal{G}, Z) \to H^2(BU(1), Z).$$

Using the cohomological triviality of the fibrations (2.12) and (2.13) it will be enough to check the surjectivity when M is the 2-sphere S^2. In this case $\mathcal{G}$ is connected and

$$H^2(B\mathcal{G}, Z) \cong H^1(\mathcal{G}, Z)$$

so we are reduced to checking surjectivity of

$$H^1(\mathcal{G}, Z) \to H^2(BU(1), Z) \tag{2.22}$$

or equivalently that

$$\pi_1(U(1)) \to \pi_1(\mathcal{G}) \tag{2.23}$$

gives a direct summand of $\pi_1(\mathcal{G})$. Now $\pi_1(\mathcal{G}) \cong \pi_2(B\mathcal{G})$ and, since $M = S^2$, this can be calculated from the fibration (2.13), which gives the short exact sequence

$$0 \longrightarrow \pi_3(U(n)) \longrightarrow \pi_1(\mathcal{G}) \xrightarrow{e} \pi_1(U(n)) \longrightarrow 0, \tag{2.24}$$

Thus $\pi_1(\mathcal{G})$ is free abelian on two generators. Note that the projection e is given by evaluation at a base point of $M = S^2$.

A more convenient description of (2.24) is given in terms of K-theory. Let E be the vector bundle defined by P and write $\mathcal{G}(E)$ for $\mathcal{G}(P)$. Then to every map

$$f: S^1 \to \mathcal{G}(E)$$

we form the bundle E_f over $M \times S^2$ by using f as clutching data and consider the element

$$[E_f] - [E_1] \in K(M \times S^2, M \times \text{point}).$$

The assignment $(f) \to [E_f] - [E_1]$ gives an isomorphism

$$\pi_1(\mathscr{G}) \to K(M \times S^2, M \times \text{point}).$$

and (2.24) corresponds to the exact sequence obtained by restricting to a point in M. All calculations can then be made in $K(M \times S^2)$, which is generated by the Hopf bundles H and L of the two factors. Using this description of $\pi_1(\mathscr{G})$ we shall now prove the following lemma, true for all pairs (n, k) with $0 < k < n$.

LEMMA 2.25. *Let E be the direct sum of k copies of H and $n-k$ trivial factors and let $U(k) \times U(n-k)$ be the corresponding constant automorphisms of E. Then the induced map*

$$\pi_1(U(k) \times U(n-k)) \to \pi_1(\mathscr{G}(E))$$

is an isomorphism.

Proof. Let $f, g: S^1 \to \mathscr{G}(E)$ come from the standard generators of $\pi_1(U(k))$ and $\pi_1(U(n-k))$ respectively. Then in $K(M \times S^2)$ we have

$$[E_f] - [E_1] = H \otimes (L-1),$$
$$[E_g] - [E_1] = 1 \otimes (L-1),$$

and these generate the kernel of

$$K(M \times S^2) \to K(M \times \text{point}).$$

Since by tensoring with line-bundles we can always reduce k modulo n we deduce immediately what we want, namely

COROLLARY 2.26. *If $(n, k) = 1$ the homomorphism*

$$\pi_1(U(1)) \to \pi_1(\mathscr{G}(E)),$$

coming from the constant central automorphisms, is a direct summand.

This completes the proof of proposition 2.21. Our use of K-theory in this proof becomes very natural if we consider briefly the situation for manifolds M of arbitrary dimension. Let $\mathscr{G}(E)$ denote the automorphism group for a vector bundle E over M and let $\mathscr{G}_0(E)$ denote its identity component. Then we have a homomorphism $U(1) \to \mathscr{G}_0(E)$ given by the constant scalars and hence a homomorphism

$$\pi_1(U(1)) \to \pi_1(\mathscr{G}_0(E)). \tag{2.27}$$

Now if we are in the stable range $n > \frac{1}{2} \dim M$ then we can show, using the construction in our proof of (2.21), that

$$\pi_1(\mathscr{G}_0(E)) \cong K(M).$$

Moreover, the image of the generator of $\pi_1(U(1)) \cong Z$ under (2.27) is just the class $[E]$ in $K(M)$. Hence (2.27) defines a direct summand if and only if $[E]$ is a primitive element of the abelian group $K(M)$. When $\dim M = 2$, $K(M) \cong Z \oplus Z$ and $[E]$ is represented by the pair of integers (n, k). Thus the coprime condition $(n, k) = 1$ generalizes naturally to the primitivity of $[E]$ in $K(M)$.

3. THE YANG–MILLS FUNCTIONAL

In this section G denotes a fixed compact connected Lie group, and P a fixed principal G-bundle over the compact manifold M.

The identity element e of G then defines a canonical section s_e of $\operatorname{Ad} P$ over M, and we use this section to pull back to M the *tangent bundle* along the fibres $(T_F \operatorname{Ad} P)$, of $\operatorname{Ad} P$.

The resulting bundle on M, is denoted by $\operatorname{ad}(P)$, this being an abbreviation for

$$\operatorname{ad}(P) = s_e^{-1}(T_F \operatorname{Ad} P).$$

Alternatively, ad (P) may be thought of as the bundle associated with P via the adjoint action of G on its tangent space T_eG at e. This space is in turn identified with the Lie algebra $\mathfrak{g}$ of left invariant vector fields on G. Thus we also have

$$\text{ad}\,(P) = P \times_G \mathfrak{g},$$

and both these descriptions make it clear that the space of sections

$$\Gamma \text{ad}\,(P)$$

has a natural Lie algebra structure induced by the structure on each fibre. Hence this space plays the role of the Lie algebra of the gauge group. Correspondingly we sometimes write

$$\mathfrak{g}(P) \quad \text{for} \quad \Gamma \text{ad}\,(P).$$

Actually it is useful to extend this Lie algebra to the following *graded Lie algebra*:

$$\Omega^*(M; \text{ad}\,(P))$$

consisting of the forms on M with values in ad (P). Precisely then,

$$\omega^q \in \Omega^*(M; \text{ad}\,(P))$$

is a smooth section of $\Lambda^q T^*M \otimes \text{ad}\,(P)$ and the bracket operation in $\mathfrak{g}$ together with the usual exterior multiplication, combines to define a pairing

$$\Omega^p \otimes \Omega^q \to \Omega^{p+q},$$

which we write $[\omega^p, \omega^q]$. This operation clearly satisfies the formula

$$[\omega^p, \omega^q] = (-1)^{pq+1}[\omega^q, \omega^p]$$

and the corresponding Jacobi identity

$$[\omega^p, [\omega^q, \omega^r]] = [[\omega^p, \omega^q], \omega^r] \pm [\omega^q, [\omega^p, \omega^r]].$$

Now a compact group G always admits a *positive definite* inner product $\langle , \rangle$ on its Lie algebra $\mathfrak{g}$, which is invariant under the adjoint action. Hence a choice of such an inner product $\langle , \rangle$ on $\mathfrak{g}$ induces a Riemannian metric on ad (P), and then naturally extends to induce a pairing

$$\Omega^p(M; \text{ad}\,(P)) \otimes \Omega^q(M; \text{ad}\,(P)) \to \Omega^{p+q}(M),$$

which *we simply write* $\omega^p \wedge \omega^q$. The invariance of $\langle , \rangle$ on $\mathfrak{g}$ implies that

$$\langle [x, y], z \rangle = \langle x, [y, z] \rangle$$

and, as in this identity all terms retain their natural order, it persists for any three elements u, v, w in our complex $\Omega^*(M; \text{ad}\,(P))$:

$$[u, v] \wedge w = u \wedge [v, w].$$

Suppose finally that a fixed Riemannian metric and a fixed orientation is chosen on M, and that $*$ is the corresponding duality operator, in $\Omega^*(M)$. Thus $*$ is characterized by

$$\theta \wedge *\theta = \langle \theta, \theta \rangle_M \,\text{vol}\,(M) \quad \text{for} \quad \theta \in \Omega,$$

where $\langle , \rangle_M$ denotes the natural Riemannian structure on $\Omega^q(M)$, and vol (M) is the unique form of length 1 in the orientation of M. Then the inner product on $\mathfrak{g}$, and the Riemannian

metric of M combine to give $\Omega^*(M; \mathrm{ad}\,(P))$ a natural inner product structure, which we write $(,)$.

Precisely then

$$(\theta, \varphi) = \int_M \theta \wedge *\varphi.$$

With these conventions out of the way we only need two more concepts to write down the Yang–Mills equations. These are the notions of a *connection* A, for P, and of its curvature

$$F(A) \in \Omega^2(M; \mathrm{ad}\,(P)).$$

By definition, a connection A for P is a *G-invariant splitting of the natural exact sequence*

$$0 \to T_F P \to TP \to \pi^{-1} TM \to 0 \tag{3.1}$$

of vector bundles over P. Here T_F denotes the tangent bundle along the fibres in P, and TM the tangent bundle of M. The group G acts on all terms of this sequence and so a G-invariant splitting of (3.1) is a well defined concept. There are two complementary ways of describing a splitting, and we correspondingly introduce the following notation: ω_A is the projection on the 'vertical bundle' $T_F P$, defined by A, and $T_A P$ is the complement to $T_F P$, defined by A; $T_A P$ is also called the 'horizontal bundle of A', and is the kernel of ω_A.

The splitting A is therefore also equivalent to a G-invariant direct sum decomposition

$$TP \cong T_F P \oplus T_A P.$$

Connections clearly exist. For instance $T_A P$ can be taken to be the orthocomplement to $T_F P$ relative to a G-invariant metric on P. Furthermore, the space of connections $\mathscr{A}(P)$ naturally has an *affine structure*, with associated vector space $\Omega^1(M; \mathrm{ad}\,(P))$.

To see this most clearly it is best to use the description of (3.1) as a pull-back under π of an exact sequence of vector bundles on M. For this purpose let $E = E(P)$ be the vector bundle over M whose fibre at $q \in M$ is equal to the G-invariant sections of TP, along the fibre $\pi^{-1}(q) \in P$:

$$E(P)_q = \Gamma\{TP | \pi^{-1}(q)\}^G.$$

Then $E(P)$ is easily seen to define a vector bundle over M, with a natural projection to TM. There results an exact sequence *on* M.

$$0 \longrightarrow \mathrm{ad}\,(P) \xrightarrow{i} E(P) \xrightarrow{\mathrm{d}\pi} TM \longrightarrow 0, \tag{3.2}$$

whose kernel is the earlier bundle ad (P), which under the pull-back to P goes over into (3.1). Finally because G-invariance is clearly built into this sequence, a connection A can also be *defined* simply as *a splitting* of (3.2). Thus in this picture, ω_A is an arrow

$$\mathrm{ad}\,(P) \xleftarrow{\omega_A} E(P)$$

with $\omega_A \cdot i = 1$; and the difference $\omega_A - \omega_{A'}$ therefore factors to an arrow

$$\mathrm{ad}\,(P) \leftarrow T(M),$$

i.e. gives rise to a 1-form $\eta \in \Omega^1(M; \mathrm{ad}\,(P))$:

$$\omega_A - \omega_{A'} = \eta. \tag{3.3}$$

This shows that $\mathscr{A}(P)$ is an affine space as asserted.

We next recall how the curvature $F(A)$ of a connection A arises. This curvature has many interpretations, but from the point of view of (3.2) it precisely measures to what extent ω_A fails to preserve the Lie structures in (3.2).

Indeed, it is clear from the definition of $E(P)$ that

$$\Gamma E(P) \cong \textit{G-invariant vector fields on } P,$$

while

$$\Gamma \mathrm{ad}\,(P) \cong \textit{G-invariant vertical vector fields on } P.$$

Hence both of these spaces have natural Lie algebra structures, and so interpreted

$$0 \to \Gamma \mathrm{ad}\,(P) \to \Gamma E(P) \to \Gamma(TM) \to 0 \tag{3.4}$$

defines an *extension* of the *Lie algebra* $\Gamma(TM)$ by the Lie algebra of the gauge group. Here $\Gamma E(P)$ is the Lie algebra of automorphisms of P, which do not necessarily cover the identity on M.

Now a connection A assigns to every $X \in \Gamma(TM)$ a unique *horizontal* vector field $\tilde{X} \in \Gamma E(P)$ projecting onto X. Hence the element

$$F_A(X, Y) = \omega_A[\tilde{X}, \tilde{Y}]$$

is a natural measure of the extent to which A fails to split $\Gamma E(P)$ as a Lie algebra. It is now easy to verify that F_A is linear over the C^∞ functions on M and hence defines a unique 2-form

$$F(A) \in \Omega^2(M; \mathrm{ad}\,(P)).$$

With all this understood one now has the following

DEFINITION. *The Yang–Mills functional L on the space of connections $\mathscr{A}(P)$ is the function*

$$L(A) = \|F(A)\|^2,$$

where $F(A)$ is the curvature of A, and $\| \|$ denotes the L^2 norm in $\Omega^*(M; \mathrm{ad}\,(P))$.

Remarks and examples. To get a feeling for this function we start by considering the case of a circle bundle P over M. In this case the choice of an invariant form on $\mathfrak{g}$ reduces to choosing a generator v of the invariant 1-forms on S^1, and in the sequel we shall assume that this v is normalized to have

$$\int_{S^1} v = 1.$$

We next write Z for the dual generator of $\mathfrak{g} = \mathbb{R}^1$, so that $v(Z) = 1$. We also write Z for the unique vertical G-invariant vector field on P, which is the infinitesimal generator corresponding to Z under the action of S^1 on P.

With these conventions, a connection A for P is completely described by a 1-form θ_A on P, which has the properties

$$\theta_A(Z) = 1, \quad \text{that is} \quad i(Z)\,\theta_A = 1, \tag{3.5}$$

$$\mathscr{L}(Z)\,\theta_A = 0, \tag{3.6}$$

where $\mathscr{L}(Z)$ denotes the Lie derivative in the direction Z.

Note further that in this instance $\mathrm{ad}\,(P)$ is a trivial one-dimensional bundle and hence $F(A) \in \Omega^2(M)$ is an ordinary 2-form on M. It is characterized by the equation

$$\mathrm{d}\theta_A = -\pi^* F(A). \tag{3.7}$$

Indeed, let $\tilde{X}, \tilde{Y}$ be A-horizontal lifts of X, Y on M. Then

$$\mathrm{d}\theta_A[\tilde{X}, \tilde{Y}] = -\theta_A[\tilde{X}, \tilde{Y}] \tag{3.8}$$

as the other terms such as $\tilde{X}\theta_A(\tilde{Y})$ disappear. On the other hand $\mathrm{d}\theta_A$ vanishes in the direction Z:

$$i(Z)\,\mathrm{d}\theta_A = 0, \quad \mathscr{L}(Z)\,\mathrm{d}\theta_A = 0$$

as follows from (3.5) and (3.6) via the basic identity

$$\mathscr{L}(Z) = i(Z)\,\mathrm{d} + \mathrm{d}i(Z).$$

Hence (3.7) follows from (3.8).

Equation (3.7) implies that

$$\mathrm{d}F(A) = 0.$$

Furthermore, the formula (3.3) now yields

$$F(A) - F(A') = \mathrm{d}\eta, \quad \eta \in \Omega^1(M). \tag{3.9}$$

Thus the map $A \to F(A)$ *sends* $\mathscr{A}(P)$ *precisely onto a certain cohomology class* $k(P) \in H^2(M)$:

$$\mathscr{A}(P) \xrightarrow{F} k(P) \longrightarrow 0.$$

We next turn to the fibre of this map F. Again from (3.9) it follows that if $F(A) = F(A')$, then $\eta = A - A'$ is closed, and conversely. Thus the fibre of F consists precisely of the space of closed 1-forms $Z^1(M) \in \Omega^1(M)$. Hence we have the 'exact sequence'

$$0 \to Z^1(M) \to \mathscr{A}(P) \to k(P) \to 0,$$

which is unorthodox in that $\mathscr{A}(P)$ is *only an affine space*, and $k(P)$ denotes the whole coset in $Z^2(M)$ representing a class usually denoted by $-2\pi i c_1(P)$ in $H^2(M)$.

We next describe the action of $\mathscr{G}(P)$ on this sequence. In the present instance Ad P is clearly trivial, and hence

$$\mathscr{G}(P) \cong \mathrm{Map}\,(M, S^1). \tag{3.10}$$

We now have

LEMMA 3.11. *If* $f: M \to S^1$ *is a smooth map, then its effect, via* (3.10), *on* θ_A *is given by*

$$f^*\theta_A = \theta_A + \pi^* f^* v$$

where v *is the form on* S^1 *discussed earlier.*

As a first consequence we see immediately that, because v is closed, $F(f^*A) = \mathrm{d}\theta_A = F(A)$. Thus F *is invariant under the gauge group.*

Next we see that $\mathscr{G}(P)$ acts on $Z^1(M)$ by translations

$$\xi \to \xi + f^* v.$$

Consider now the action of the identity component $\mathscr{G}_0(P)$ on $Z^1(M)$. Clearly such a map lifts to a map $\tilde{f}$ in the diagram

$$\begin{array}{ccc} M & \xrightarrow{\tilde{f}} & R \\ & \searrow^{f} & \downarrow \exp \\ & & S^1 \end{array}$$

and if x is the linear function on R with

$$\exp^* v = \mathrm{d}x$$

then

$$f^* v = \mathrm{d}\tilde{f}.$$

We have the converse also, so that under $\mathscr{G}_0(P)$, $\xi \in Z^1(M)$ is moved in its entire cohomology class: thus

$$Z^1(M)/\mathscr{G}_0(P) \cong H^1(M; R).$$

Finally we turn to the components of $\mathscr{G}(P)$. Because S^1 is an Eilenberg–Maclane space, these components are isomorphic to $H^1(M; Z)$, the correspondence being

$$f \to \text{class of } f^*v.$$

Putting all this together we obtain the formula

$$Z^1(M)/\mathscr{G}(M) \cong H^1(M; R)/H^1(M; Z).$$

Thus in this case $\mathscr{A}(P)/\mathscr{G}(P)$ is a torus $T(M)$ of dimension equal to the dimension of $H^1(M; R)$, as it should be, because the action of $\mathscr{G}(P)$ on $\mathscr{A}(P)$ only has the global gauge group S^1 in its kernel, whence

$$B\mathscr{G}(P) \simeq T(M) \times BS^1.$$

4. The Yang–Mills equations

We turn next to the equations of variation of $L(A) = \|F(A)\|^2$, and because $\mathscr{A}(P)$ is naturally an affine space, it suffices to vary A along lines

$$A_t = A + t\eta, \quad \eta \in \Omega^1(M; \text{ad}(P)).$$

The first problem is therefore to compute $F(A_t)$ for such a variation.

To do this recall that a connection A on P induces a natural covariant derivative ∇_A^X on *all* associated vector bundles of P. Thus if

$$V(P) = P \times_G V$$

is associated with a representation

$$\rho : G \to \text{Aut } V,$$

then A induces a way of differentiating sections s, of $V(P)$:

$$s \to \nabla_X^A s, \tag{4.1}$$

in any given direction X on M. This 'covariant derivative' then dually corresponds to a differential operator

$$\mathrm{d}_A : \Omega^0(M; V(P)) \to \Omega^1(M; V(P)),$$

which finally *extends uniquely* to a differential operator

$$\mathrm{d}_A : \Omega^*(M; V(P)) \to \Omega^{*+1}(M; V(P)),$$

compatible with the natural pairing

$$\Omega^*(M; R) \otimes \Omega^*(M; V(P)) \to \Omega^*(M; V(P))$$

given by multiplication. Compatibility simply means

$$\mathrm{d}_A(\theta \wedge \omega) = \mathrm{d}\theta \wedge \omega \pm \theta \wedge \mathrm{d}_A \omega.$$

Recall here that (4.1) is defined by the following construction. We have

$$\Gamma(V(P)) = \text{Map}_G(P; V).$$

Hence $s \in \Gamma(V(P))$ corresponds to a G-equivariant map

$$\tilde{s}\colon P \to V, \quad \tilde{s}(pg) = p(g^{-1})\tilde{s}(p).$$

Now given X, a vector field on M, its A-lift $\tilde{X}$ to P is a well defined G-invariant vector field. Hence

$$\tilde{X} \cdot \tilde{s}\colon P \to V$$

is again G-equivariant, and corresponds to the section $\nabla_X^A s \in \Gamma(V(P))$.

In our situation, the bundle ad (P) is associated to P via the adjoint representation, so that $\Omega^*(M; \mathrm{ad}\,(P))$ inherits a natural differential operator from the connection A. Explicitly, we have for $s \in \Gamma(\mathrm{ad}\,(P))$, $X \in \Gamma(T)$ and $\tilde{X} \in \Gamma(E)$ the A-lift of X to $E(P)$ in (3.2)

$$\nabla_X^A s = \mathrm{d}_A s(X) = [\tilde{X}, s],$$

and, for instance, if $\theta \in \Omega^1(M; \mathrm{ad}\,(P))$ then

$$\mathrm{d}_A\theta(X, Y) = \nabla_X \theta(Y) - \nabla_Y \theta(X) - \theta[X, Y]. \tag{4.2}$$

The associated connections to A, on the various associated bundles are all of course compatible, in the sense that, if

$$V \to W$$

is a G-equivariant map, then d_A commutes with the induced maps

$$\Omega^*(M; V(P)) \to \Omega^*(M; W(P)).$$

Furthermore, these covariant derivatives behave like derivatives relative to tensor-products.

It follows in particular that d_A *acting on* $\Omega^*(M; \mathrm{ad}\,(P))$ *behaves as a derivation under both the bracket* [ ] *and the* $\wedge$ *operation*:

$$\mathrm{d}_A[\alpha, \beta] = [\mathrm{d}_A\alpha, \beta] \pm [\alpha, \mathrm{d}_A\beta], \tag{4.3}$$

$$\mathrm{d}(\alpha \wedge \beta) = \mathrm{d}_A\alpha \wedge \beta \pm \alpha \wedge \mathrm{d}_A\beta. \tag{4.4}$$

With all these functorial remarks out of the way, we have the following:

LEMMA 4.5. *Let* A_t *be the line of connections*

$$A_t = A + t\eta, \quad \eta \in \Omega^1(M; \mathrm{ad}\,(P)).$$

Then the curvature of A_t *is given by*

$$F(A_t) = F(A) + t\mathrm{d}_A\eta + \tfrac{1}{2}t^2[\eta, \eta].$$

Proof. By definition, the horizontal lifts $\tilde{X}_t$ of a vector field X on M relative to A_t are related by:

$$\tilde{X}_t = X_t + t\eta(X)$$

and, correspondingly, the vertical projections relative to A_t are related by

$$\omega_t = \omega - t\eta \circ \pi.$$

Expanding $\omega_t[\tilde{X}_t, \tilde{Y}_t]$ now yields

$$\begin{aligned} F_t(X, Y) &= \omega_t\{[\tilde{X}, \tilde{Y}] + [\tilde{X}, \eta(Y)] + [\eta(X), \tilde{Y}] + t^2[\eta(X, \eta(Y)]\} \\ &= \omega_0[\tilde{X}, \tilde{Y}] - t\eta[X, Y] + t[\tilde{X}, \eta(Y)] - t[\tilde{Y}, \eta(X)] + t^2[\eta(X), \eta(Y)], \end{aligned}$$

which is the desired formula by (4.2), since $[\alpha, \beta]$ for 1-forms α, β is defined as the two form

$$[\alpha, \beta](X, Y) = [\alpha(X), \beta(Y)] - [\alpha(Y), \beta(X)].$$

With the help of our lemma, it is now an easy matter to compute the first and second variations of L.

PROPOSITION (4.6). *The connection A is stationary for $L(A) = \|F(A)\|^2$, if and only if*

$$d_A * F(A) = 0.$$

Proof. Expanding $F_t = F(A_t)$ according to (4.5) gives

$$\|F_t\|^2 = \|F\|^2 + 2t(d_A\eta, F) + t^2\{\|d_A\eta\|^2 + (F, [\eta, \eta])\} + \text{higher terms.} \tag{4.7}$$

Hence at an extremum $(d_A\eta, F) = 0$, or equivalently

$$(\eta, d_A^* F) = 0,$$

for all $\eta \in \Omega^1(M; \mathrm{ad}\,(P))$. Hence at an extremum

$$d_A^* F(A) = 0. \tag{4.8}$$

Here d_A^* is the adjoint of d_A relative to our norm on $\Omega^*(M; \mathrm{ad}\,(P))$ and, just as in the usual Hodge theory, it is given by $\pm * d_A *$. *Precisely*, if $m = \dim M$, then

$$d_A^* = (-1)^{m+mp+1} * d_A * \quad \text{on} \quad \Omega^p. \tag{4.9}$$

Hence (4.9) implies (4.6).

For completeness we quickly review the proof of (4.9). Suppose then that $\varphi \in \Omega^{p-1}$, $\psi \in \Omega^p$, one of them having compact support. Then by (4.4)

$$d_A\{\varphi \wedge * \psi\} = d_A\varphi \wedge * \psi + (-1)^{p-1}\varphi \wedge d_A * \psi.$$

Hence integration yields

$$0 = (d_A\varphi, \psi) + (-1)^{p-1}(\varphi, *^{-1} d_A * \psi).$$

But

$$*^{-1} = (-1)^{(p-1)(m-p+1)} *$$

on $(p-1)$-forms, yielding (4.9).

Remark. The Bianchi identities assert that for every A, we have $d_A F(A) = 0$. Hence at a stationary point we have both

$$d_A F(A) = 0 \quad \text{and} \quad d_A^* F(A) = 0.$$

Forms satisfying these two equations are clearly (*nonlinear*!) analogues of *harmonic forms* in the usual Hodge theory. In short the condition for A to be extremal is precisely that its curvature $F(A)$ be harmonic in $\Omega^2(M; \mathrm{ad}\,(P))$.

The expansion (4.7) of course also yields the Hessian of L at an extremal connection A. This Hessian is a quadratic form on the tangent space to $\mathscr{A}(P)$ at A, which in our identification is precisely $\Omega^1(M; \mathrm{ad}\,(P))$. With this understood we have the formula for the second variation.

PROPOSITION (4.10). *The quadratic form $Q(\eta, \eta)$ defined by the Hessian of L at an extremal connection A is given by*

$$Q(\eta, \eta) = (d_A^* d_A \eta + *[*F, \eta], \eta).$$

Proof. From (4.7) we have

$$Q(\eta, \eta) = \frac{1}{2}\frac{d^2}{dt^2}\|F_t\|^2\Big|_{t=0} = \|d_A\eta\|^2 + (F, [\eta, \eta]).$$

To bring this expression into the required form observe that $\|d_A\eta\|^2 = (d_A^* d_A\eta, \eta)$, while

$$(F, [\eta, \eta]) = \int_M [\eta, \eta] \wedge * F = \int_M \eta \wedge [\eta, * F] = (-1)^{m+1}\int_M \eta \wedge * *^{-1}[*F, \eta]. \tag{4.11}$$

Using the formula for $*^{-1}$ this reduces to $(\eta, *[*F, \eta])$.

We write $\hat{F}$ for the endomorphism

$$\hat{F}\colon \eta \to *[*F, \eta]$$

of $\Omega^1(M; \mathrm{ad}\,(P))$. This is clearly a degree-zero operator, which by the argument of (4.11) is also characterized by

$$(\hat{F}\eta, \xi) = (F, [\eta, \xi])$$

and is therefore self-adjoint.

Remarks. The operator

$$\mathrm{L}_A \equiv \mathrm{d}_A^* \mathrm{d}_A + *[*F, \ ],$$

which appears in (4.10), can also be interpreted as the Jacobi operator associated with an extremal of $L(A)$. That is, if $A_t = A + t\eta + t^2\eta' \ldots$ describes a curve of connections for which $L(A)$ is extremal, then

$$\mathrm{L}_A \eta = 0.$$

To see this we differentiate the equations

$$\mathrm{d}_{A_t}^* F(A_t) = 0$$

with respect to t and set $t = 0$. If a dot denotes such a differentiation then we clearly have

$$\dot{\mathrm{d}}_A = \eta, \quad F(\dot{A}) = \mathrm{d}_A \eta,$$

whence the derivative of

$$*\,\mathrm{d}_A * F(A) = 0$$

is given by

$$*[\eta, *F(A)] + *\,\mathrm{d}_A * \mathrm{d}_A \eta = 0,$$

which, once the signs are taken care of, is equivalent to

$$\mathrm{L}_A \eta = 0, \quad \eta \in \Omega^1(M; \mathrm{ad}\,(P)). \tag{4.12}$$

The solutions of (4.12) are therefore the 'Jacobi-fields' of L, and describe the *tangent space* to the space of solutions of the nonlinear extremal equations for L.

On the other hand our functional L is clearly *invariant under the action of the group of gauge transformations*. Hence the proper measure of the tangent space to the space of solutions is given by the quotient of the solutions of (4.12) by the directions along the orbits of the action of $\mathcal{G}(P)$. Now this space is, as we shall show in a moment, precisely the image of $\Omega^0(M; \mathrm{ad}\,(P))$ in $\Omega^1(M; \mathrm{ad}\,(P))$ under d_A. Thus the corrected tangent space to the space of solutions is the quotient N_A of the space of Jacobi fields J_A by the image of d_A and therefore fits into the exact sequence

$$\Omega^0(M; \mathrm{ad}\,(P)) \xrightarrow{\mathrm{d}_A} J_A \longrightarrow N_A \longrightarrow 0. \tag{4.13}$$

We shall call N_A the null space of Q_A, and its dimension the nullity of A. This nullity is always *finite* because of the following argument.

In our norm on $\Omega^1(M; \mathrm{ad}\,(P))$ the orthocomplement of the image of the d_A in (4.13) is precisely the kernel of d_A^*. Thus N_A may be identified with the space of $\eta \in \Omega^1$ satisfying the equations

$$\mathrm{L}_A \eta = 0, \quad \mathrm{d}_A^* \eta = 0$$

or equivalently

$$\mathrm{d}_A^* \mathrm{d}_A + \mathrm{d}_A \mathrm{d}_A^* + *[*F, \ ] = 0, \quad \mathrm{d}_A^* = 0. \tag{4.14}$$

The first operator on the left is the *Laplacian* Δ_A of d_A and hence elliptic. Hence the solutions of (4.14) are *finite-dimensional*, and therefore the nullity is also.

This argument immediately extends to the Morse index of A, which is defined as the dimension of a maximal negative subspace of Q, or, equivalently, as the dimension of a maximal subspace in the kernel of d_A^* on which the form

$$\hat{Q}(\eta) = (\Delta_A \eta + *[*F, \eta], \eta), \quad F = F(A),$$

is negative definite.

Thus, to sum up, we have the following finiteness result.

PROPOSITION 4.15. *The index and nullity of an extremal A are finite and equal to the index and nullity of the quadratic form*

$$\hat{Q}(\eta) = (\Delta_A \eta + \hat{F}\eta, \eta), \quad \hat{F} = *[*F, \eta],$$

on the kernel of d_A^* in $\Omega^1(M; \mathrm{ad}\, P)$.

Finally, if $X \in \Gamma(\mathrm{Ad}\, P)$ is a left invariant vertical vector field it acts on $T(P)$ via the Lie bracket. Hence, if $\tilde{Y}$ is an A-horizontal vector field, $-[X, Y]$ measures the first-order effect on the A-horizontal spaces. It follows from (4.5) that in our identification

$$T_A \mathscr{A} = \Omega^1(M; \mathrm{ad}\,(P))$$

the tangent space to the orbit of $\mathscr{G}(P)$ at A is given by the image of

$$d_A: \Omega^0(M; \mathrm{ad}\,(P)) \to \Omega^1(M; \mathrm{ad}\,(P)),$$

as was asserted earlier.

5. YANG–MILLS OVER A RIEMANN SURFACE

When the base-manifold of P is two-dimensional the Yang–Mills equations naturally relate to holomorphic structures and can therefore be understood best in a holomorphic context.

To see this recall first that when $\dim M = 2$, the $*$ operator of a Riemannian structure on M maps Ω^1 to Ω^1, with $*^2 = -1$. Hence we have a natural decomposition

$$\Omega^1_C(M) = \Omega^{1,0}(M) \oplus \Omega^{0,1}(M),$$

with Ω_C denoting $\Omega \otimes C$, and

$$* = -\mathrm{i} \quad \text{on} \quad \Omega^{1,0}, \quad * = \mathrm{i} \quad \text{on} \quad \Omega^{0,1},$$

of the complexified de Rham complex. This decomposition splits $d: \Omega^0 \to \Omega^1$, into $d': \Omega^0 \to \Omega^{1,0}$ and $d'': \Omega^0 \to \Omega^{0,1}$, and so induces a holomorphic structure on M, whose holomorphic functions f correspond locally to solutions of

$$d''f = 0.$$

Suppose now that P is a principal G-bundle over M (M and G both compact), and that A is a connection for P. Then the above argument can be applied to the complex $\Omega^*(M; \mathrm{ad}\,(P))$, and d_A, giving a decomposition

$$\Omega^1_C(M; \mathrm{ad}\,(P)) = \Omega^{1,0}(M; \mathrm{ad}\,(P)) \oplus \Omega^{0,1}(M; \mathrm{ad}\,(P))$$

according to the eigenspaces of $*$:

$$* = -\mathrm{i} \quad \text{on} \quad \Omega^{1,0}, \quad * = \mathrm{i} \quad \text{on} \quad \Omega^{0,1}.$$

There is a corresponding decomposition of d_A, so that we have the diagram:

$$\begin{array}{ccc} \Omega^{1,0}(M; \mathrm{ad}\,(P)) & \xrightarrow{d_A''} & \Omega^2_C(M; \mathrm{ad}\,(P)) \\ \uparrow{\scriptstyle d_A'} & & \uparrow{\scriptstyle d_A'} \\ \Omega^0_C(M; \mathrm{ad}\,(P)) & \xrightarrow{d_A''} & \Omega^{0,1}(M; \mathrm{ad}\,(P)) \end{array},$$

which is of course compatible with the corresponding decomposition of $\Omega_C(M)$, and now the operator d''_A defines a *holomorphic structure on the vector bundle* ad (P) *over* M. This can be proved as in Atiyah *et al.* (1978, theorem 5.1) by applying the Newlander–Nirenberg integrability theorem for complex structures. Alternatively we can give a more elementary proof as follows.

Clearly all that has to be done is to construct local frames s, for ad (P), with $\mathrm{d}''_A s = 0$. Now if s_U is any frame over $U \subset M$, we have

$$\mathrm{d}''_A s_U = \theta s_U,$$

where θ is some matrix of 1-forms of type $(0, 1)$ on M. Further if we change s_U to fs_U where f is an appropriate matrix-valued function then

$$\mathrm{d}''_A f s_u = (\mathrm{d}'' f + f\theta)\, s_u.$$

Hence we need to solve the equation

$$f^{-1}\mathrm{d}'' f + \theta = 0. \tag{5.1}$$

First consider (5.1) globally over the 2-sphere S^2 (for the trivial bundle). Working with Sobolev spaces H^k and using their basic properties, explained in § 14, we see that the map

$$P : H^2 \to H^1$$

given by $P(f) = f^{-1}\mathrm{d}'' f$ is smooth. Moreover its derivative at $f = 1$ (the identity matrix) is the linear elliptic operator d'', which on S^2 is surjective and has the constant matrices as kernel. The implicit function theorem for Banach spaces then ensures that the equation $P(f) = -\theta$ has (near $f = 1$) a unique solution $f \in H^2$ orthogonal to the constants, provided θ is close to zero in H^1. If θ is in C^∞ then so is f.

To deduce the local solvability of (5.1) around $z = 0$ we introduce a cut-off function $\rho(|z|)$ with graph of the form

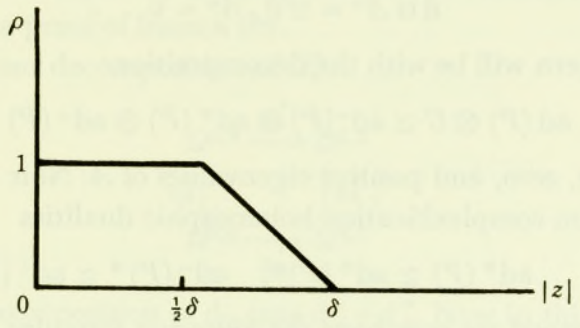

FIGURE 2.

This function is in H^1 and we have a universal bound (independent of δ) for its H^1-norm:

$$\|\rho\|_1 \leqslant 2\pi^{\frac{1}{2}}.$$

Putting $\phi = \rho\theta$ we can then estimate the H^1-norm of ϕ:

$$\|\phi\|_1^2 = \|\rho\theta\|^2 + \|\rho'\theta + \rho\theta'\|^2 \leqslant 2\{\|\rho\theta\|^2 + \|\rho'\theta\|^2 + \|\rho\theta'\|^2\} \leqslant 8\pi \sup |\theta|^2 + 2\|\theta\|_1^2.$$

Here $\|\ \|$ stands for the usual L^2-norm, ρ' stands for $\mathrm{d}\rho$ and we restrict throughout to the disc $|z| \leqslant \delta$, which is the support of ρ. Now we can always assume that our frame was so chosen that $\theta(0) = 0$ and so, for δ sufficiently small, both $\sup |\theta|^2$ and $\|\theta\|_1^2$ can be made as small as we please. Thus, for small δ, $\|\phi\|_1^2$ is small and applying the global solvability of (5.1) to ϕ we find an f

satisfying $f^{-1}\mathrm{d}''f+\phi=0$. Restricting to $|z|<\frac{1}{2}\delta$ we have $\phi=\theta$ and so f is the required local solution.

Using the holomorphic structure defined by d_A'' we shall continue our analysis of Yang–Mills in the two-dimensional case and discuss the implication of having a connection A that is extremal for $L(A)$ and therefore satisfies the equation

$$\mathrm{d}_A * F(A) = 0. \tag{5.2}$$

Now $F(A)\in\Omega^2(M,\mathrm{ad}\,(P))$ and hence $*F(A)\in\Omega^0(M;\ \mathrm{ad}\,(P))$, and is therefore a section of $\mathrm{ad}\,(P)$, on which both d_A' and $\mathrm{d}_A''=0$. Thus (5.2) implies that $*F(A)$ is a *holomorphic* section of $\mathrm{ad}\,(P)$, which is covariant constant, and we claim that we can therefore decompose $\mathrm{ad}\,(P)$ *completely according to the eigenvalues of the endomorphism*

$$\Lambda = \mathrm{i}\,\mathrm{ad} * F(A),$$

that is

$$\Lambda\alpha = \mathrm{i}[*F(A),\alpha],\quad \alpha\in\Omega^*(M;\mathrm{ad}\,(P)).$$

More precisely, we claim that the eigenvalues of Λ on $\mathrm{ad}\,(P)$ *are locally constant*, and, as Λ is self-adjoint, there is therefore a natural decomposition

$$\mathrm{ad}\,(P)\otimes C = \bigoplus_{\lambda}\mathrm{ad}_\lambda\,(P),\quad \lambda\in R,$$

of $\mathrm{ad}\,(P)\otimes C$ into orthogonal sub-bundles where Λ reduces to the scalar λ.

Furthermore, (5.2) now implies that the induced decomposition

$$\Omega^*_C(M;\mathrm{ad}\,(P)) = \bigoplus_{\lambda}\Omega^*(M;\mathrm{ad}_\lambda\,(P))$$

is stable under d_A. The local constancy of the eigenvalues of Λ follows by considering the trace functions $\mathrm{tr}\,\Lambda^n$, $n=1,2,\ldots$. We have

$$\mathrm{d}\,\mathrm{tr}\,\Lambda^n = \mathrm{tr}\,\mathrm{d}_A\Lambda^n = 0.$$

Actually our main concern will be with the decomposition

$$\mathrm{ad}\,(P)\otimes C \cong \mathrm{ad}^-(P)\oplus\mathrm{ad}^*\,(P)\oplus\mathrm{ad}^+\,(P)$$

corresponding to negative, zero, and positive eigenvalues of Λ. Note also that the Reimannian metric on $\mathrm{ad}\,(P)$ induces on complexification holomorphic dualities

$$\mathrm{ad}^*\,(P)\cong\mathrm{ad}^*\,(P)^*,\quad \mathrm{ad}^-\,(P)^*\cong\mathrm{ad}^+\,(P). \tag{5.3}$$

With these matters understood we now have the following formulae for the index and nullity.

PROPOSITION 5.4. *Let A be an extremal connection for P. Then*

$$\mathrm{nullity}\,(A) = 2\dim_C H^1(M;\mathrm{ad}^*\,(P)),$$
$$\mathrm{index}\,(A) = 2\dim_C H^1(M;\mathrm{ad}^-\,(P)),$$

where H^i denotes the cohomology of the sheaf of holomorphic sections of the bundle indicated.

Proof. From the discussion in §4 we have to compute the nullity and index of the quadratic form

$$\hat{Q}(\eta) = (\Delta_A\eta+\hat{F}\eta,\eta),\quad \eta\in\mathrm{Ker}\,\mathrm{d}_A^*,$$

in $\Omega^1(M;\mathrm{ad}\,(P))$, where as before we have written $\hat{F}$ for the transformation

$$\hat{F}\colon\eta\to *[*F(A),\eta].$$

We can furthermore clearly decompose this equation according to the eigenvalues of Λ, and so are naturally reduced to three cases with $\Lambda = 0$, or a positive or negative scalar. Consider first $\Lambda = 0$, that is the case corresponding to $\mathrm{ad}^*(P)$. In this situation $\hat{Q}(\eta) = (\Delta_A \eta, \eta)$ is semi-definite. Hence the index is zero, and the nullity equals the dimension of the harmonic forms

$$\Delta_A \eta = 0 \tag{5.5}$$

in $\ker \mathrm{d}_A^*$. But (5.5) implies that $\mathrm{d}_A \eta = 0$ and $\mathrm{d}_A^* \eta = 0$. Hence nullity $\hat{Q}$ = dim harmonic forms in $\Omega^1(M; \mathrm{ad}^*(P))$.

We next turn to the case when Λ is a *positive* scalar, that is corresponding to $\mathrm{ad}_\lambda(P)$ with $\lambda > 0$ in the spectrum of Λ. In this situation we shall need the following estimate on the first positive eigenvalue of the Laplacian.

LEMMA 5.6. *Consider Δ_A acting on Ω^1_C for the bundle $\mathrm{ad}_\lambda(P)$ with $\lambda > 0$. Then Δ_A preserves the spaces $\Omega^{1,0}$ and $\Omega^{0,1}$, and the first positive eigenvalue of $\Delta_A | \Omega^{1,0}$ is $\geqslant 2\lambda$.*

From this lemma and the self-evident formula

$$\hat{F} | \Omega^{1,0} = -\lambda, \quad \hat{F} | \Omega^{0,1} = \lambda,$$

it follows immediately that our operator

$$\Delta_A + \hat{F}$$

is positive on $\Omega^{0,1}$, and has on $\Omega^{1,0}$ the single negative eigenvalue $-\lambda$ with multiplicity the dimension of the harmonic forms in $\Omega^{1,0}$. Hence (5.6) leads to

COROLLARY (5.7). *The quadratic form $\hat{Q}$ of (4.15) has nullity zero and, index $\hat{Q} | Ker\, \mathrm{d}_A^*$ = Index $\hat{Q}$ = dim harmonic forms in $\Omega^{1,0}$.*

In short then, this corollary computes the contribution of $\mathrm{ad}_\lambda(P)$ to the index of A, in terms of the dimension of the harmonic forms in $\Omega^{1,0}$, and it is then quite standard Hodge theory to translate this answer into the statement in proposition 5.4. We shall review these matters in a moment, but turn first to a proof of lemma 5.6.

For this purpose recall our decomposition of Ω^*_C:

$$\begin{array}{ccc} \Omega^{1,0} & \xrightarrow{\mathrm{d}_A''} & \Omega^{1,1} \\ {\scriptstyle \mathrm{d}_A'}\uparrow & & \uparrow{\scriptstyle \mathrm{d}_A'} \\ \Omega^{0,0} & \xrightarrow[\mathrm{d}_A'']{} & \Omega^{0,1} \end{array} , \tag{5.8}$$

and the corresponding decomposition of d_A into $\mathrm{d}_A' + \mathrm{d}_A''$. Now in this diagram each arrow has a natural adjoint and we can therefore associate a *Laplacian with each arrow*. Each such Laplacian gives a self-adjoint operator on the spaces at both ends of the arrow. Thus we have a *lower* and an *upper* $\Box_A''$, defined by

$$\Box_A'' = \mathrm{d}_A''(\mathrm{d}_A'')^* + (\mathrm{d}_A'')^* \mathrm{d}_A''$$

as well as *left* and *right* $\Box_A'$ defined by

$$\Box_A' = \mathrm{d}_A'(\mathrm{d}_A')^* + (\mathrm{d}_A')^* \mathrm{d}_A'.$$

Now the basic relations between these operators are given by the following.

LEMMA 5.9. *The Laplacians $\Box_A'$ and $\Box_A''$ induce the same operator on $\Omega^{1,0}$ and $\Omega^{0,1}$. Further Δ_A preserves these spaces and*

(i) $$\Delta_A = 2\Box_A' = 2\Box_A'' \quad \text{on} \quad \Omega^{1,0} \quad \text{and} \quad \Omega^{0,1},$$

while

(ii) $$\Delta_A = \Box'_A + \Box''_A \quad \text{on} \quad \Omega^{0,0} \quad \text{and} \quad \Omega^{1,1},$$

and, finally, on $\Omega^{0,0}$ *these two Laplacians differ by* λ:

(iii) $$\Box'_A - \Box''_A = \lambda \quad \text{on} \quad \Omega^{0,0}.$$

Proof. Both (i) and (ii) are formal consequences of the equation $\mathrm{d}^* = -*\mathrm{d}*$ and the fact that $* = -\mathrm{i}$ on $\Omega^{1,0}$ and $* = \mathrm{i}$ on $\Omega^{0,1}$.

For instance, for $\alpha \in \Omega^{0,0}$ we obtain

$$\begin{aligned}\Delta_A \alpha &= -*\mathrm{d}_A * \mathrm{d}_A \alpha \\ &= +*\mathrm{d}(\mathrm{i}\mathrm{d}'_A\alpha - \mathrm{i}\mathrm{d}''_A\alpha) \\ &= \mathrm{i}*(\mathrm{d}''_A\mathrm{d}'_A - \mathrm{d}'_A\mathrm{d}''_A)\alpha \\ &= \Box'_A\alpha + \Box''_A\alpha.\end{aligned}$$

The last relation (iii) now follows from

$$\mathrm{i}\,\mathrm{d}^2\alpha = (\mathrm{d}'_A\mathrm{d}''_A + \mathrm{d}''_A\mathrm{d}'_A)\alpha = *[\Lambda, \alpha] \quad \text{for} \quad \alpha \in \Omega^{0,0}.$$

Now this lemma serves to estimate the spectrum of Δ_A by means of the standard theory of elliptic complexes. Indeed each arrow of diagram (5.8) is an *elliptic* operator. Hence by the Hodge theory *the positive spectra of the two associated Laplacians are isomorphic*. Thus the positive spectra of $\Box'_A$ on $\Omega^{0,0}$ and $\Omega^{1,1}$ are equal. On the other hand by (iii) this spectrum is bounded below by λ, because $\Box''_A$ is semi-definite. But on $\Omega^{1,0}$ we have $\Delta_A = 2\Box'_A$. This completes the proof of lemma 5.6.

It remains to translate the harmonic forms into sheaf cohomology terms in the standard way. This translation is based on the fact that each of the operators in our diagram can be interpreted as a resolution for the kernel sheaf of the operator.

Thus the lower d''_A, together with Hodge theory, yields

$$H^i(M; \mathrm{ad}_\lambda(P)) \cong \mathrm{Ker}\{\Box''_A | \Omega^{0,i}\},$$

while the upper d''_A gives

$$H^i(M; \mathrm{ad}_\lambda(P) \otimes \Omega^1) \cong \mathrm{Ker}\,(\Box''_A | \Omega^{1,i}).$$

Finally Serre duality gives:

$$H^i(M; \mathrm{ad}_\lambda(P) \otimes \Omega^1) \cong H^{1-i}(M; \mathrm{ad}_\lambda(P)\ *).$$

Thus our index formula (5.7) for $\lambda > 0$, becomes

$$\mathrm{index}\, Q = \dim H^1(M; \mathrm{ad}_{-\lambda}(P))$$

once we recall the duality (5.3).

Finally if we take $\lambda < 0$ a completely analogous argument yields

$$\mathrm{index}\, Q = \dim H^1(M; \mathrm{ad}_\lambda(P)).$$

Summing over λ then completes the proof of proposition 5.4.

Remarks. Although the formulae for the index and nullity seem similar, there is an essential difference between them. The nullity is essentially unstable (under changes of A) while the index is not. This stability of the index follows from the Kodaira vanishing theorem and the

Riemann–Roch formula. Indeed on $\mathrm{ad}_\mu(P)$, with $\mu < 0$, $H^0(M; \mathrm{ad}_\mu(P)) = 0$, as also follows from the formula

$$\Box' = \Box'' - \mu, \quad \mu > 0,$$

on $\Omega^{0,0}$, and the diagram (5.8).

Hence the index is also minus the Euler characteristic of $\mathrm{ad}_\mu(P)$. Now, by Riemann–Roch,

$$\dim H^0(M; E) - \dim H^1(M; E) = c_1(E) + (g-1)\dim E.$$

Applied in our situation, this then leads to the following *rigid and computable* formula for the index:

$$\text{index}\, A = 2\{c_1(\mathrm{ad}^+(P)) + (g-1)\dim_C(\mathrm{ad}^+ P)\}. \tag{5.10}$$

Here c_1 denotes the first Chern class and we have switched to $\mathrm{ad}^+(P)$ via the formula

$$c_1(E^*) = -c_1(E).$$

6. Representations of the fundamental group

In the previous section we saw that, over a compact connected Riemann surface M, a connection A is extremal for the Yang–Mills functional if and only if $*F(A)$ is covariant constant (relative to A), that is

$$\mathrm{d}_A * F(A) = 0. \tag{6.1}$$

In particular, if our G-bundle is topologically trivial, a flat connection, that is with $F = 0$, necessarily satisfies (6.1) and corresponds to the absolute minimum of the Yang–Mills functional. Now a flat G-connection is locally trivial and globally corresponds to a homomorphism

$$\pi_1(M) \to G \tag{6.2}$$

describing the holonomy. Solutions of (6.1) that are non-zero can be described in a similar manner by using a suitable central extension of $\pi_1(M)$.

We recall that $\pi_1(M)$ is a group with $2g$ generators $A_1, B_1, \ldots, A_g, B_g$ satisfying the single relation

$$\prod_{i=1}^{g} [A_i, B_i] = 1, \tag{6.3}$$

where $[A, B]$ is the commutator $ABA^{-1}B^{-1}$. It follows that, for $g \geqslant 1$, $\pi_1(M)$ has a universal central extension

$$1 \to Z \to \Gamma \to \pi_1(M) \to 1,$$

where Γ is generated by $A_1, B_1, \ldots, A_g, B_g$ and a central element J satisfying the single relation

$$\prod_{i=1}^{g} [A_i, B_i] = J. \tag{6.4}$$

Let Γ_R denote the group obtained from Γ by extending the centre to R, so that we have a central extension:

$$1 \to R \to \Gamma_R \to \pi_1(M) \to 1. \tag{6.5}$$

On dividing by Z the group Γ_R clearly becomes a direct product

$$1 \to Z \to \Gamma_R \to U(1) \times \pi_1(M) \to 1. \tag{6.6}$$

Now let $Q \to M$ be a $U(1)$-bundle with Chern class 1 endowed with a fixed harmonic or Yang–Mills connection. If we normalize the metric on M so that it has total volume 1 the curvature of this harmonic connection on Q is $-2\pi i\omega$, where ω is the volume form on M. The universal covering $\tilde{M} \to M$ is of course a flat $\pi_1(M)$-bundle, so that the fibre product $Q \times_M \tilde{M}$ is a $U(1) \times \pi_1(M)$-bundle with connection still having curvature $-2\pi i\omega$. Lifting to Γ_R we then get a Γ_R-bundle over M with connection and curvature $-2\pi i\omega$. In particular this connection A is a Yang–Mills connection, a notion which makes sense even though Γ_R is not compact.

Given any homomorphism

$$\rho : \Gamma_R \to G$$

we then get an induced G-connection A_ρ also satisfying the Yang–Mills equations, since (6.1) is clearly functorial for homomorphisms. Our observation is that all Yang–Mills connections are obtained in this way, namely we have

THEOREM 6.7. *The mapping $\rho \to A_\rho$ induces a bijective correspondence between conjugacy classes of homomorphisms $\rho : \Gamma_R \to G$ and equivalence classes of Yang–Mills connections over M.*

To prove this theorem we have to understand the significance of the Yang–Mills equation (6.1). First we shall show that, as a consequence of (6.1), the *conjugacy class* of $*F = *F(A)$ is constant. To see this recall that $*F$ can be considered as a $\mathfrak{g}$-valued function on P,

$$*F : P \to \mathfrak{g},$$

which is equivariant under G, i.e.

$$*F(pg) = \operatorname{Ad} g^{-1} *F(p).$$

Hence the values of $*F$ certainly lie in a fixed conjugacy class of $\mathfrak{g}$ (i.e. orbit of $\operatorname{Ad} G$) on each fibre of P. On the other hand the condition

$$d_A *F = 0$$

asserts that for any vector field X on M, its horizontal lift $\tilde{X}$ (relative to A) annihilates $*F$. It follows that $*F$ is constant along horizontally lifted curves.

To proceed further let us now choose a point X in the orbit of $*F(A)$, and set P_X equal to the inverse image of X under $*F$:

$$P_X = *F^{-1}\{X\}.$$

Because $*F$ maps *onto* the orbit of X, this set will be a submanifold of P, and in fact it defines a *reduction of the structure group of G to G_X*, the centralizer of X in G, that is P_X is stable under the action of G_X and

$$P_X/G_X = M. \tag{6.8}$$

Furthermore the horizontal subspaces of A are tangent to P_X (again because $\tilde{X} *F \equiv 0$) so that *A restricts to a connection of P_X* over M, with curvature $F(A)$, where now $*F(A)$ is the constant map

$$P_X \to X. \tag{6.9}$$

It follows that *we may think of $F(A)$ as the Lie algebra valued* 2-form,

$$F(A) = X \otimes \omega \in \Omega^2(M; \mathfrak{g}_X), \tag{6.10}$$

with ω the volume form on M.

The Yang–Mills connection A_ρ defined by a homomorphism $\rho : \Gamma_R \to G$ has curvature $X_\rho \otimes \omega$

where X_ρ is the element of the Lie algebra $\mathfrak{g}$ of G defined by $\mathrm{d}\rho: R \to \mathfrak{g}$. Since R is central in Γ_R it follows that $\rho(\Gamma_R)$ centralizes X_ρ and so ρ is actually a homomorphism

$$\rho: \Gamma_R \to G_X \quad \text{with} \quad X = X_\rho.$$

Thus in proving theorem 6.7 we can restrict ourselves to the case when X is central, i.e. when $G_X = G$.

Next let us consider the case when G is a torus, so that we are dealing with a direct sum of line-bundles. Now line-bundles with harmonic connection form an abelian group under $\otimes$ and so can be uniquely expressed as $Q^k \otimes L_0$ where k is the Chern class, Q is our fixed line-bundle of degree 1 and L_0 is flat. Taking direct sums then shows that theorem 6.7 is true in this abelian case. As we have already remarked it is also true in the flat case, Γ_R now factoring through $\pi_1(M)$.

The general case is essentially a combination of these extreme cases but to proceed further we need to recall the basic facts about the structure of compact connected Lie groups G. First of all the commutator subgroup $S = [G, G]$ is the *maximal connected semi-simple subgroup* of G. The connected component H of the centre of G is a torus, which together with S generates G. The intersection $D = S \cap H$ is a finite subgroup of the centre of S, and so

$$H \times S \to G$$

is a finite covering with group D (acting diagonally). Thus we may write $G = H \times_D S$, and factoring out further by D we can put

$$\bar{G} = G/D, \quad \bar{H} = H/D, \quad \bar{S} = S/D$$

so that we have

$$\bar{G} = \bar{H} \times \bar{S}. \tag{6.11}$$

Any G-bundle P with connection induces a $\bar{G}$-bundle $\bar{P}$ with connection. Conversely if $\bar{P}$ lifts to P then P is unique and inherits a connection from that of $\bar{P}$. Similarly a homomorphism $\rho: \Gamma_R \to G$ induces $\bar{\rho}: \Gamma_R \to \bar{G}$. Moreover if $\rho(R)$ is central in G then $\rho(Z) \subset D$, since $Z \subset [\Gamma, \Gamma]$, and so $\bar{\rho}$ factors through

$$\Gamma_R/Z = U(1) \times \pi_1(M).$$

In view of (6.11) we see that $\bar{\rho}$ is determined by a pair of homomorphisms

$$\begin{cases} \alpha: U(1) \times \pi_1(M) \to \bar{H} \\ \beta: \pi_1(M) \to \bar{S}. \end{cases} \tag{6.12}$$

A central Yang–Mills connection for $\bar{G}$ is equivalent to a Yang–Mills connection for $\bar{H}$ and a flat connection for $\bar{S}$. We have already seen that (6.7) holds in these two separate cases so that we end up precisely with the pair of homomorphisms α and β.

This completes the proof of theorem 6.7. We should note, however, that in this theorem we have simultaneously considered all topological types. It remains therefore to describe the topology of the bundle associated with a given representation $\rho: \Gamma_R \to G$. Now G-bundles over M are trivial over the 1-skeleton of M (since G is connected) and are classified by a class in

$$H^2(M, \pi_1(G)) \cong \pi_1(G).$$

For the group $\bar{G}$ of (6.11) we have

$$\pi_1(\bar{G}) \cong \pi_1(\bar{H}) \oplus \pi_1(\bar{S}).$$

The homomorphisms α, β of (6.12) determine classes

$$[\alpha] \in \pi_1(\bar{H}), \quad [\beta] \in \pi_1(\bar{S}).$$

The definition of $[\alpha]$ is clear, we simply regard the restriction to $U(1)$ as a loop. For $[\beta]$ we note that β extends canonically to a homomorphism

$$\hat{\beta}\colon \Gamma \to \tilde{S},$$

where $\tilde{S}$ is the (finite) universal covering of $\bar{S}$: this is because Γ is the universal central extension of $\pi_1(M)$. The class $[\beta]$ is then just given by

$$[\beta] = \hat{\beta}(J),$$

where as before J generates the kernel of $\Gamma \to \pi_1(M)$.

To any homomorphism $\rho\colon \Gamma_R \to G$ with $\rho(R)$ central we then pass to $\bar{\rho}\colon \Gamma_R \to \bar{G}$, which is described by the pair α, β. The class $[\alpha] \oplus [\beta] \in \pi_1(\bar{G})$ then determines the topology of the associated $\bar{G}$-bundle, and so that of the G-bundle. Since $G \to \bar{G}$ is a finite covering with group D it follows that we have an exact sequence

$$0 \to \pi_1(G) \to \pi_1(\bar{G}) \to D \to 0$$

and the pair $[\alpha] \oplus [\beta]$ in $\pi_1(\bar{G})$ belongs to the subgroup $\pi_1(G)$ if and only if $[\alpha]$ and $[\beta]$ have opposite images in D, using the exact sequences

$$0 \to \pi_1(H) \to \pi_1(\bar{H}) \to D \to 0,$$
$$0 \to \pi_1(S) \to \pi_1(\bar{S}) \to D \to 0.$$

Since for the semi-simple group S we have $[S, S] = S$ it follows that, for $g \geqslant 1$, the equations

$$\prod_{i=1}^{g} [A_i, B_i] = \eta \tag{6.13}$$

have solutions with $A_i, B_i \in S$ for any given $\eta \in S$. In particular on replacing S by its universal covering and taking η to be any element of the centre it follows that $\beta \to [\beta]$ defines a *surjection*

$$\mathrm{Hom}\,(\pi_1(M), \bar{S}) \to \pi_1(\bar{S}). \tag{6.14}$$

More trivially $\alpha \to [\alpha]$ gives a surjection since restriction to $U(1)$ defines an *isomorphism*

$$\mathrm{Hom}\,(U(1), \bar{H}) \cong \pi_1(\bar{H}). \tag{6.15}$$

Thus we have proved

Proposition 6.16. *Every topological G-bundle P over M possesses a central Yang–Mills connection. The space of (equivalence classes of) such connections is given by all (conjugacy classes of) solutions of* (6.13) *with given η, multiplied by the torus* $\mathrm{Hom}\,(\pi_1(M), \bar{H})$.

As explained above the element η in (6.16) is determined by the topological type of P. The curvature also is determined by the topology of P. More precisely it depends on the characteristic classes via (6.15). Hence the value of the Yang–Mills functional is also determined by the topology of P. It is not hard to check and will be proved in § 12 that this value is the *absolute minimum* for P.

The general Yang–Mills G-connection for P then arises from a Yang–Mills minimum for the group G_X. If S_X is the maximal connected semi-simple subgroup of G_X then $G_X/S_X = \bar{H}_X$ is a torus and

$$L_X = \pi_1(G_X/S_X)$$

is a lattice. This contains $\pi_1(H_X)$ as a lattice of finite index, so that we may view L_X as sitting inside the Lie algebra of H_X, which in turn is in the Lie algebra of G. In this way

$$X \in L_X \subset \mathfrak{g}. \tag{6.17}$$

Note that X can now be identified with the class $[\alpha]$ of the homomorphism $\alpha\colon U(1) \to G_X/S_X$.

It remains to examine the topological relation between the bundles P_X and P. The inclusion $G_X \subset G$ sends S_X into S (since S is maximal semi-simple) and hence $G_X/S_X \to G/S$. Also for the universal coverings $\tilde{S}_X$ is a factor of $\tilde{S}$ and so the centre of $\tilde{S}_X$ is a subgroup of the centre of S. Elements of $\pi_1(G)$ can be viewed as pairs (a, b) with

$$a \in \pi_1(G/S), \quad b \in \text{centre}\, S.$$

Similarly elements of $\pi_1(G_X)$ are given by pairs a_X, b_X and the homomorphism

$$\pi_1(G_X) \to \pi_1(G)$$

then assigns (a_X, b_X) to (a, b) in the obvious way. In particular b *determines* b_X *uniquely*. The element a_X is then constrained by two relations

$$\begin{cases} a_X \to a \\ a_X \equiv b_X \bmod D_X. \end{cases} \tag{6.18}$$

Here the congruence is to be understood in the sense that we use the two natural homomorphisms

$$\text{centre}\, \tilde{S}_X \to \text{centre}\, S_X$$
$$L_X \to D_X \subset \text{centre}\, S_X.$$

For the G_X-bundle defined by a pair of homomorphisms (α_X, β_X) as in (6.12), we have $a_X = [\alpha_X]$ and $b_X = [\beta_X]$, and $[\alpha_X]$ can also be identified with the point X in the lattice L_X.

To sum up we see therefore that, for a given C^∞ G-bundle P the Yang–Mills connections fall into a countable number of families or *types* determined by conjugacy classes of elements X in $\mathfrak{g}$. These X are constrained by the condition (6.17) and (6.18) rewritten in the form

$$\begin{cases} X \to a \\ X \equiv b \bmod D_X, \end{cases} \tag{6.19}$$

where (a, b) are the pair determining the topology of P with $a \in \pi_1(G/S)$, $b \in \text{centre}\, \tilde{S}$.

In theorem 6.7 we formulated the results for full equivalence classes. If instead we pick a base point $x_0 \in M$ and work with the subgroup $\mathcal{G}_0$ of gauge transformations that are the identity at x_0, then (6.7) becomes the statement that we have a bijection

$$\text{Hom}\,(\Gamma_R, G) \to \mathcal{M}/\mathcal{G}_0$$

where $\mathcal{M}$ is the space of all Yang–Mills connections. The group $G = \mathcal{G}/\mathcal{G}_0$ acts on both sides and induces the bijection on quotients expressed in (6.7).

We shall now spell this out in more detail for the case of $G = U(n)$. The Lie algebra is then the space of skew-hermitian matrices. We write such a matrix as $-2\pi i\Lambda$ so that Λ is hermitian. Its conjugacy class is determined by the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ of Λ, which we may arrange in descending order:

$$\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n. \tag{6.20}$$

The maximal semi-simple subgroup is $SU(n)$ while the diagonal $U(1)$ is the centre. The group

$$D = SU(n) \cap U(1) \cong Z_n$$

is the group of nth roots of unity. The homomorphism $U(n) \to U(n)/SU(n) \cong U(1)$ is of course given by the determinant. The lattice

$$L = \pi_1(U(n)/SU(n)),$$

therefore corresponds to diagonal hermitian matrices with integral trace, i.e. the diagonal entries λ are such that $n\lambda$ is integral.

If $X = -2\pi i\Lambda$ with Λ having the eigenvalues in (6.20), the centralizer G_X depends on how many coincidences there are amongst the λ_j. Thus if the first n_1 are equal, the next n_2 are equal and so on, we have

$$G_X = U(n_1) \times \ldots \times U(n_r). \tag{6.21}$$

The lattice L_X then has dimension r and the condition (6.17) becomes

$$n_j\lambda_j \quad \text{is integral for all } j. \tag{6.22}$$

Thus when the λ_j are all distinct they must all be integers, while at the opposite extreme, when they are all equal, they are rational with denominator n.

Since $\pi_1(U(n)) \cong Z$ is free abelian, a $U(n)$-bundle over M is determined topologically by a single integer, its Chern class. Thus the general description we have used for $\pi_1(G)$ contains redundant information in this case. More precisely we considered the finite Z_n-covering

$$U(n) \to U(1) \times PU(n),$$

where $PU(n)$ is the projective unitary group, and identified $\pi_1(U(n))$ with the appropriate subgroup of

$$\pi_1(U(1)) \oplus \pi_1(PU(n)) \cong Z \oplus Z_n.$$

It is easy to see that our subgroup is generated by the element $1 \oplus 1$. Thus for a pair $(a, b) \in Z \oplus Z_n$ to represent an element of $\pi_1(U(n))$ we must have $a \equiv b \bmod n$, and our element is then given by the integer a. Condition (6.19) now reduces to the obvious requirement

$$\text{trace } X = a \tag{6.23}$$

where a is the Chern class of P.

In terms of vector bundles, a reduction from $U(n)$ to a G_X of the form (6.21) corresponds to a direct sum decomposition

$$E = E_1 \oplus \ldots \oplus E_r.$$

The condition (6.22) merely asserts that the Chern classes of the E_j must be integers while (6.23) asserts that the sum of these Chern classes must coincide with the Chern class of E.

For $U(n)$, a homomorphism $\Gamma_R \to U(n)$ is just a unitary representation of Γ_R. If a representation is *irreducible* then X is necessarily central so that all its eigenvalues λ_j are equal and given by k/n where k is the Chern class. The converse is true when $(n, k) = 1$, since a reducible representation can only produce eigenvalues with smaller denominators in view of (6.22). Narasimhan & Seshadri (1965) have shown that, provided $g \geqslant 2$, irreducible representations exist for all (n, k). The proof is a simple matter of exhibiting irreducible sets of matrices satisfying (6.4) with J any given nth root of unity. Naturally for $g = 1$, $\pi_1(M)$ is abelian and so has no irreducible unitary representations for $n > 1$. Thus for $k = 0$ and $n > 1$ irreducible representations do not exist, while for $(n, k) = 1$ they do exist. This is consistent with the results of Atiyah (1957) on the classification of holomorphic bundles over elliptic curves.

Yang–Mills $U(n)$-connections for which X is diagonal (with entries $-2\pi ik/n$) give rise as we have mentioned to the absolute minimum $4\pi^2k^2/n$ for the Yang–Mills functional. We have shown therefore that the most general Yang–Mills connection for a vector bundle E is simply a direct sum of Yang–Mills minima for sub-vector bundles.

7. Holomorphic bundles

In this section we shall consider holomorphic vector bundles over our compact Riemann surface M and discuss the general nature of the classification problem. In particular we shall explain how to compute the cohomology of the moduli space of stable bundles. In principle the approach we shall give is entirely 'non-unitary' and does not involve Morse theory. However, in §8 we shall explain the relation between this holomorphic approach and the unitary approach based on the Yang–Mills functional and Morse theory ideas.

To demonstrate clearly the structure of the argument we shall not enter here into any technicalities. The relevant analytical details are treated in §§14 and 15.

We consider therefore a fixed C^∞ complex vector bundle E over M of rank n and Chern class k and we denote by $\mathscr{C}(E)$ or $\mathscr{C}(n,k)$, or simply $\mathscr{C}$, the space of all holomorphic structures on E. In concrete terms a holomorphic structure may be defined by its d''-operator, so that the local holomorphic sections are the solutions of $\mathrm{d}''u = 0$. Relative to a C^∞ local basis of E we can write

$$\mathrm{d}'' = \mathrm{d}_0'' + B$$

where d_0'' is the usual Cauchy–Riemann operator and B is a matrix-valued $(0,1)$-form on M. Since $\dim M = 1$ there are no integrability constraints on B, so that B can be chosen arbitrarily (see §5). From a global point of view it follows that $\mathscr{C}$ is a complex affine space whose vector space of translations is $\Omega^{0,1}(\mathrm{End}\, E)$, where $\mathrm{End}\, E$ denotes the C^∞ vector bundle of complex endomorphisms of E.

Let $\mathrm{Aut}\,(E)$ denote the group of automorphisms of E so that an element of this group is locally a C^∞ map of M into $GL(n,C)$. Then $\mathrm{Aut}\,(E)$ acts on $\mathscr{C}(E)$ and the orbits are, by definition, the isomorphism classes of complex analytic bundles on M with rank n and Chern class k. Our aim is to describe this orbit structure as fully as possible and in particular to discuss the 'moduli space'.

As usual with classification problems in algebraic geometry, in order to get a good 'moduli space', we have to consider a restricted class of holomorphic structures, those that are *stable* in the sense of Mumford (1965). The set of stable points in $\mathscr{C}(E)$ forms an open set $\mathscr{C}_s(E)$ and the corresponding orbits are then closed in $\mathscr{C}_s(E)$ so that the quotient space $\mathscr{C}_s(E)/\mathrm{Aut}\,(E)$ is a Hausdorff space. In fact it turns out to be a complex manifold and is compact if $(n,k) = 1$. This is the moduli space we are primarily interested in studying and whose cohomology we want to compute.

We recall now the precise definition of stability. It will be convenient first to introduce the *normalized Chern class* or 'slope' (in the terminology of Shatz (1977)) μ = Chern class/rank. Then a holomorphic bundle E is *stable* if, for every proper holomorphic sub-bundle D of E, we have $\mu(D) < \mu(E)$. *Semi-stable* is defined similarly but we allow now the weak inequality $\mu(D) \leqslant \mu(E)$. Elementary arguments as in Harder & Narasimhan (1975) then show that every holomorphic bundle E has a *canonical filtration*.

$$0 = E_0 \subset E_1 \subset E_2 \subset \ldots \subset E_r = E \tag{7.1}$$

with $D_i = E_i/E_{i-1}$ semi-stable and

$$\mu(D_1) > \mu(D_2) > \ldots > \mu(D_r).$$

Of course if E is semi-stable then $r = 1$.

If D_i has rank n_i and Chern class k_i so that $n = \sum n_i$, $k = \sum k_i$ we shall call the sequence of pairs (n_i, k_i) $i = 1, \ldots, r$ the *type* of E. It is sometimes convenient to describe the type equivalently by the single n-vector μ whose components are the ratios k_i/n_i each represented n_i times and arranged in decreasing order. Thus

$$\mu = (\mu_1, \ldots, \mu_n)$$

with $\mu_1 \geqslant \mu_2 \geqslant \ldots \geqslant \mu_n$, where the first n_i are equal to k_1/n_1, the next n_2 are equal to k_2/n_2 and so on.

All the holomorphic bundles of a given type μ define a subspace $\mathscr{C}_\mu$ of $\mathscr{C}$. In particular if all components of μ are equal (hence are all k/n) then $\mathscr{C}_\mu = \mathscr{C}_{ss}$ is the semi-stable part of $\mathscr{C}$.

Since the filtration (7.1) is canonical the subspaces $\mathscr{C}_\mu$ are preserved by the action of the automorphism group, so that each $\mathscr{C}_\mu$ is a union of orbits.

It is well known that the infinitesimal variations of a holomorphic bundle E are classified by the elements of the sheaf cohomology group $H^1(M, \operatorname{End} E)$. In our picture this gets interpreted as follows. The orbit in $\mathscr{C}$ corresponding to a given holomorphic bundle E is, locally, a manifold of finite codimension in $\mathscr{C}$ and its normal can be identified with $H^1(M, \operatorname{End} E)$. This is because an infinitesimal gauge transformation, namely a global C^∞ endomorphism ϕ of E, alters d'' by the addition of $\mathrm{d}''\phi$ and the cokernel of

$$\Omega^0(\operatorname{End} E) \xrightarrow{\mathrm{d}''} \Omega^{0,1}(\operatorname{End} E)$$

is just $H^1(M, \operatorname{End} E)$.

In the same way we can identify the conormal to $\mathscr{C}_\mu$. Since $\mathscr{C}_\mu$ is a union of orbits its conormal should be a quotient of $H^1(M, \operatorname{End} E)$. Now let $\operatorname{End}' E$ denote the bundle of holomorphic endomorphisms of E that preserve its canonical filtration and define $\operatorname{End}'' E$ by the exact sequence

$$0 \to \operatorname{End}' E \to \operatorname{End} E \to \operatorname{End}'' E \to 0. \tag{7.2}$$

From the exact cohomology sequence of (7.2) we see that $H^1(M, \operatorname{End}'' E)$ is indeed a quotient of $H^1(M, \operatorname{End} E)$ and this is clearly the right candidate for the conormal to $\mathscr{C}_\mu$, since $H^1(M, \operatorname{End}' E)$ describes variation inside $\mathscr{C}_\mu$. The important point to notice at this stage is

$$\dim H^1(M, \operatorname{End}'' E) \quad \textit{depends only on } \mu. \tag{7.3}$$

This follows from Riemann–Roch together with the key fact that

$$H^0(M, \operatorname{End}'' E) = 0. \tag{7.4}$$

This in turn is an easy corollary, by induction, of

(7.5) *If E, D are both semi-stable and $\mu(E) > \mu(D)$ then every homomorphism $E \to D$ is zero.*

The proof of (7.5) is a simple consequence of the definitions and can be deduced from Narasimhan & Seshadri (1965, proposition 4.4): it is in any case an essential step in the proof of the uniqueness of the canonical filtration.

From (7.3) we can deduce that $\mathscr{C}_\mu$ is locally a submanifold of finite codimension in $\mathscr{C}$. Thus the picture that is emerging of $\mathscr{C}$ is that it has a stratification by submanifolds $\mathscr{C}_\mu$, giving a sort of generalized cell-structure. To understand the mutual positions of the $\mathscr{C}_\mu$ we need to know something about the *closure* of $\mathscr{C}_\mu$. In algebraic terms we want to know what happens to the canonical filtration (7.1) under 'specialization'. This problem has been studied, in the framework of

algebraic geometry, by Shatz (1977) who describes how the *type* changes under specialization. To explain this result we have to introduce a partial ordering on the vectors μ that parametrize our types. This partial ordering can be described in several equivalent ways. First we follow Shatz and associate with the type μ the convex polygon P_μ with vertices $(0,0), (n_1,k_1), (n_1+n_2,k_1+k_2), \ldots$:

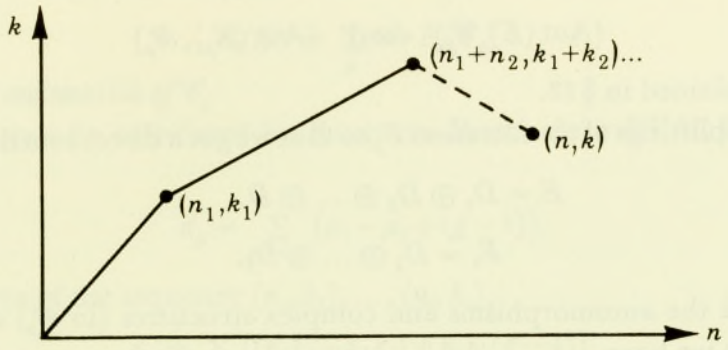

FIGURE 3.

Note that the convexity of P_μ is equivalent to monotonicity of the quotients k_i/n_i. Shatz now defines the partial ordering by:

(7.6) $$\lambda \geqslant \mu \quad \textit{if} \quad P_\lambda \quad \textit{is above} \; P_\mu.$$

If we consider P_μ as the *graph* of a concave function p_μ, then p_μ is defined at the integers by

$$p_\mu(i) = \sum_{j \leqslant i} \mu_j$$

and interpolates linearly between integers. Here the μ_j are the components of our n-vector μ. Hence, for our vector notation, (7.6) translates into the following partial ordering:

(7.7) $$\lambda \geqslant \mu \quad \text{if} \quad \sum_{j \leqslant i} \lambda_j \geqslant \sum_{j \leqslant i} \mu_j, \quad j = 1, \ldots, n-1.$$

Note that $\Sigma\lambda_j = \Sigma\mu_i = k$ is fixed. This partial ordering on vectors in R^n is well known in various contexts (see Marshall & Olkin 1979) and we shall discuss its Lie group significance in §12. For the present we return to the result of Shatz, which now takes the form

(7.8) $$\mathscr{C}_\mu \subset \bigcup_{\lambda \geqslant \mu} \mathscr{C}_\lambda.$$

In the next section we shall give a differential–geometric proof of (7.8) that is more in the spirit of this paper.

It is clear that this partial ordering on types satisfies condition (1.17). We shall check condition (1.18) later (see (7.16)). We can thus use the stratification of $\mathscr{C}$ by the $\mathscr{C}_\mu$ to describe the equivariant cohomology of $\mathscr{C}$ in terms of that of the $\mathscr{C}_\mu$. It remains to show that this stratification is 'perfect' in the sense of §1.

Let $\mathscr{F}_\mu$ denote the space of all C^∞ filtrations of E of type μ. Thus points $f_\mu \in \mathscr{F}_\mu$ are sections of the fibre bundle over M with fibre the manifold $F_\mu = GL(n,C)/B_\mu$ where B_μ is the parabolic subgroup preserving a fixed (partial) flag of subspaces of C^n of dimensions $n_1, n_1+n_2, \ldots$. The sequence of Chern classes k_i corresponds to picking a definite component of the space of all sections. Since the filtration (7.1) is canonical we have a map (the continuity of which will be established in §§14 and 15)

(7.9) $$\mathscr{C}_\mu \to \mathscr{F}_\mu.$$

If we fix a base-point of $\mathscr{F}_\mu$ corresponding to a definite C^∞ filtration E_μ of E, the fibre of (7.9) over this point is the subspace $\mathscr{B}_\mu \subset \mathscr{C}_\mu$ of complex structures compatible with the given filtration. If $\mathrm{Aut}\,(E_\mu)$ is the group of C^∞ automorphisms of E preserving the filtration then $\mathrm{Aut}\,(E_\mu)$ acts on $\mathscr{B}_\mu$, $\mathscr{F}_\mu$ is the homogeneous space $\mathrm{Aut}\,(E)/\mathrm{Aut}\,(E_\mu)$ and $\mathscr{C}_\mu$ can be identified with the associated bundle. Hence for the purposes of equivariant cohomology the pairs

$$(\mathrm{Aut}\,(E), \mathscr{C}_\mu) \quad \text{and} \quad (\mathrm{Aut}\,(E_\mu), \mathscr{B}_\mu)$$

are equivalent as explained in §13.

Next let us choose splittings of the filtration E_μ so that we get a direct sum decomposition E_μ^0 of E

$$E = D_1 \oplus D_2 \oplus \ldots \oplus D_r \tag{7.10}$$

with

$$E_i = D_1 \oplus \ldots \oplus D_i.$$

and let $\mathrm{Aut}\,E_\mu^0, \mathscr{B}_\mu^0$ be the automorphisms and complex structures (in $\mathscr{B}_\mu$) compatible with this decomposition. Then we have

$$\mathrm{Aut}\,(E_\mu^0) \cong \prod_{l=1}^{r} \mathrm{Aut}\,(D_i), \tag{7.11}$$

$$\mathscr{B}_\mu^0 \cong \prod_{l=1}^{r} \mathscr{C}_{\mathrm{ss}}(D_i),$$

On the other hand the natural homomorphism

$$\mathrm{Aut}\,(E_\mu) \to \mathrm{Aut}\,(E_\mu^0)$$

is a homotopy equivalence, and the fibration

$$\mathscr{B}_\mu \to \mathscr{B}_\mu^0$$

has a vector space as fibre and is compatible with the group actions. It follows that, for purposes of equivariant cohomology, the pairs

$$(\mathrm{Aut}\,(E_\mu), \mathscr{B}_\mu) \quad \text{and} \quad (\mathrm{Aut}\,(E_\mu^0), \mathscr{B}_\mu^0)$$

are equivalent. Together with (7.11) this finally yields for rational cohomology

Proposition 7.12. *The equivariant cohomology of the stratum $\mathscr{C}_\mu(E)$ is isomorphic to the tensor product of the equivariant cohomology of the semi-stable strata for the quotients D_i.*

Here of course the equivariant cohomology is always taken with respect to the automorphism group of the appropriate bundle.

We also need to look at the equivariant cohomology of the conormal bundle N_μ to $\mathscr{C}_\mu$ in $\mathscr{C}$. By this we mean of course the appropriate relative cohomology or the cohomology of the Thom space of N_μ. Exactly the same reduction as above shows that we can replace the triple $(\mathrm{Aut}\,(E), \mathscr{C}_\mu, N_\mu)$ by the triple $(\mathrm{Aut}\,(E_\mu^0), \mathscr{B}_\mu^0, N_\mu^0)$ where N_μ^0 is the restriction of N_μ to $\mathscr{B}_\mu^0$. Now from (7.11) we see that $\mathrm{Aut}\,(E_\mu^0)$ contains the r-dimensional torus T^r, which acts trivially on $\mathscr{B}_\mu^0$. To show that our stratification is perfect it remains to show, using (1.9) and (13.4), that the representation of T^r on the fibre of N_μ is primitive. Now at a point of $\mathscr{B}_\mu^0$ our bundle E is a holomorphic direct sum of the D_i and so the bundle $\mathrm{End}'\,E$ of endomorphisms preserving the filtration is the direct sum of $\mathrm{Hom}\,(D_i, D_j)$ for $i \geqslant j$. Hence

$$\mathrm{End}''\,E \cong \bigoplus_{i<j} \mathrm{Hom}\,(D_i, D_j). \tag{7.13}$$

Now on Hom (D_i, D_j) the element $(t_1, \ldots, t_r) \in T^r$ acts by $t_i^{-1} t_j$ and so it acts by the same character on $H^1(M, \mathrm{Hom}\,(D_i, D_j))$. Since the fibre of N is $H^1(M, \mathrm{End}'' E)$ it follows from (7.13) that the representation of T^r on N is indeed primitive. Thus we have proved

THEOREM 7.14. *The stratification of $\mathscr{C}$ by the $\mathscr{C}_\mu$ is equivariantly perfect so that for the equivariant Poincaré series we have*

$$P_t(\mathscr{C}) = \sum_\mu t^{2d_\mu} P_t(\mathscr{C}_\mu)$$

where d_μ is the complex codimension of $\mathscr{C}_\mu$.

The dimension d_μ can be calculated by Riemann–Roch, since $H^0(M, \mathrm{End}'' E) = 0$, and we find (as in (5.10))

$$d_\mu = \sum_{\mu_i > \mu_j} \{\mu_i - \mu_j + (g-1)\}. \tag{7.15}$$

Alternatively, in terms of the sequence $(n_1, k_1), \ldots, (n_r, k_r)$,

$$d_\mu = \sum_{i>j} \{(n_i k_j - n_j k_i) + n_i n_j (g-1)\}. \tag{7.16}$$

In particular this shows that our stratification does indeed satisfy the finiteness condition (1.18).

The first term in the series of (7.14) arises from the semi-stable bundles. All the remainder can be calculated inductively by (7.12). Hence knowledge about $P_t(\mathscr{C})$ (from §1) leads to inductive formulae for $P_t(\mathscr{C}_{ss})$.

Since the stratification of $\mathscr{C}$ is perfect over the integers we can also deduce results about torsion. First we should note that the equivariant cohomology of $\mathscr{C}$, namely the cohomology of $B\,\mathrm{Aut}\,(E)$, has no torsion. This follows from the identification with $B\mathscr{G}$, to be explained in §8, together with the results of §2. It follows therefore that all strata $\mathscr{C}_\mu$ have no torsion (equivariantly). In particular therefore

(7.17) *the semi-stable stratum $\mathscr{C}_{ss}$ has no torsion in its equivariant cohomology.*

In the coprime case $(n, k) = 1$ we have $\mathscr{C}_{ss} = \mathscr{C}_s$ and $\mathrm{Aut}\,(E)$ acts on $\mathscr{C}_s$ with the constant scalars as the only isotropy group (Narasimhan & Seshadri 1965). From this we can derive results for the ordinary cohomology of the moduli space $N(n, k) = \mathscr{C}_s/\mathrm{Aut}\,(E)$. Thus we get a formula for its Poincaré polynomial and we shall also see that it has no torsion. This will be treated in detail in §9.

8. RELATION WITH YANG–MILLS

In the previous section we have given a purely complex analytic approach to the moduli space of stable bundles. We want now to look at the same problem from the unitary or differential–geometric point of view. The connecting link is the key result of Narasimhan & Seshadri (1965) identifying stable bundles as those that arise from irreducible unitary representations. Translated into the notation we have introduced in §6 their result can be formulated as follows.

(8.1) *A holomorphic vector bundle of rank n is stable if and only if it arises from an irreducible representation $\rho : \Gamma_R \to U(n)$. Moreover isomorphic bundles correspond to equivalent representations.*

Remarks. 1. Actually the description given by Narasimhan & Seshadri (1965) is slightly different from (8.1) though equivalent to it. They puncture M at one point p and consider coverings of M with ramification of order n at p. This leads to a purely holomorphic description whereas our version, with connections, is a differential–geometric version.

2. Donaldson (1983) has recently given a new proof of (8.1) in the spirit of this paper.

To understand the geometric significance of (8.1) we recall first, as explained in §5, that a unitary connection A for a vector bundle E over our Riemann surface M defines a holomorphic structure simply by taking the $(0, 1)$-component d_A'' of the covariant derivative d_A. This gives a map $\mathscr{A} \to \mathscr{C}$, which is in fact an affine-linear isomorphism. Locally this corresponds to the isomorphism

$$\Omega^1(\mathfrak{u}(n)) \cong \Omega^{0,1}(\mathfrak{gl}(n, C))$$

for the Lie algebra valued 1-forms. Note that $\mathscr{C}$ is defined independently of any metric on the bundle, whereas $\mathscr{A}$ and hence the isomorphism $\mathscr{A} \to \mathscr{C}$ depend on such a choice. The connection A associated in this way to the holomorphic structure will be called the metric connection. Since any two metrics differ by a complex gauge transformation, i.e. an element of Aut (E), it will be immaterial which metric we pick. The group Aut (E) may now be viewed as the complexification $\mathscr{C}^c$ of the group of unitary gauge transformations $\mathscr{G}$ of E.

Now let $\mathscr{N} \subset \mathscr{A}$ denote the set of connections giving the *minimum* for the Yang–Mills functional. As we have shown in §6 these are precisely $\mathscr{G}$-equivalent to those given by representations $\rho : \Gamma_R \to U(n)$ with $\rho(R)$ central. Let $\mathscr{N}_s \subset \mathscr{N}$ be those given by irreducible representations. Then (8.1) can be reformulated as follows:

(8.2) *Under the identification of $\mathscr{A}$ with $\mathscr{C}$, we have $\mathscr{N}_s \subset \mathscr{C}_s$ and the induced map of quotient spaces*

$$\mathscr{N}_s/\mathscr{G} \to \mathscr{C}_s/\mathscr{G}^c$$

is a homeomorphism.

The proof of Narasimhan & Seshadri (1965) is essentially of (8.2). It is easy to prove the inclusion $\mathscr{N}_s \subset \mathscr{C}_s$ and infinitesimal arguments show that the map of quotients (which are manifolds) is a local diffeomorphism. The hardest part of the proof is the surjectivity and this requires compactifying both sides and a consequent induction on n. The real explanation for (8.2) is probably to be found in the moment map ideas indicated at the end of §10 (see also remark 2 above).

Note that the quotient space $\mathscr{G}^c/\mathscr{G}$ may be identified with the space of hermitian metrics on E. Since this is a convex set in a linear space it is contractible and so $\mathscr{G}$ and $\mathscr{G}^c$ have the same homotopy type. Hence equivariant cohomology is the same for the two groups.

Since direct sums of stable bundles with the same 'slope' are semi-stable it follows that $\mathscr{N} \subset \mathscr{C}_{ss}$: note also that $\mathscr{N} = \mathscr{N}_s$ in the coprime case. More generally now let us transport the stratification of $\mathscr{C}$ by the $\mathscr{C}_\mu$ defined in §7, to give a stratification of $\mathscr{A}$ by strata $\mathscr{A}_\mu$. Let $\mathscr{N}_\mu$ denote the Yang–Mills connections whose curvature is of type μ. Such connections are direct sums of connections of the form $\mathscr{N}_s$ for smaller ranks. This shows that $\mathscr{N}_\mu \subset \mathscr{A}_\mu$.

Our first aim in this section is to show how to characterize the $\mathscr{A}_\mu$ strata by properties of the curvature. In particular we shall eventually show that the $\mathscr{A}_\mu$ are the 'Morse strata' of the Yang–Mills functional. In fact we shall prove that this holds for a much wider class of functionals than Yang–Mills. These functionals are obtained as follows. Let ϕ be any smooth function on the Lie algebra $\mathfrak{g}$ of G that is *invariant* under the adjoint action and is *convex*. For example when $G = U(n)$, so that $x \in \mathfrak{g}$ is a skew-hermitian matrix with eigenvalues $i\lambda_1, \ldots, i\lambda_n$, we can take for ϕ any of the following

(i) $\Sigma\lambda_j^{2k}$,
(ii) $\exp(k\lambda_j)$,

$k = 1, 2, \ldots$.

These can be written directly in terms of x, without resort to the eigenvalues, and are clearly smooth invariant functions. Why they are convex will be explained in § 12 when we discuss this notion more systematically. Given such a ϕ we now define a function Φ on the space of G-connections over M in the obvious way:

$$\Phi(A) = \int_M \phi(*F(A)). \tag{8.3}$$

This reduces to the Yang–Mills functional when $\phi(x) = \text{trace } x^*x$ (corresponding to $k = 1$ in (i) above). Since ϕ is invariant under the adjoint action, Φ is gauge-invariant. Taking $A_t = A + t\eta$ and computing as in § 4 we see that

$$\Phi(A_t) = \Phi(A) + t\int \langle \phi'(*F(A)), *\mathrm{d}_A \eta\rangle \bmod t^2 \tag{8.4}$$

where $\phi' : \mathfrak{g} \to \mathfrak{g}$ is the derivative of ϕ, i.e.

$$\phi(x+ty) = \phi(x) + t\langle \phi'(x), y\rangle \bmod t^2.$$

From (8.4) we see that the gradient of Φ, relative to the metric on the space of connections, is

$$\operatorname{grad}\Phi = -*\mathrm{d}_A\phi'(*F(A)). \tag{8.5}$$

This reduces to the formulae of § 4 for the Yang–Mills functional when $\phi(x) = \text{trace } x^*x$ so that $\frac{1}{2}\phi'$ is the identity map $\mathfrak{g} \to \mathfrak{g}$. In general ϕ' is an equivariant map (relative to the adjoint action of G) and so for any section s of ad (P) the covariant derivative of $\phi'(s)$ can be obtained from that of s by the composition rule

$$\mathrm{d}_A\phi'(s) = \phi''(s) \circ \mathrm{d}_A s. \tag{8.6}$$

This is most easily understood by viewing s as an equivariant function $P \to \mathfrak{g}$ and noting that $\mathrm{d}_A s$ is just the horizontal part (relative to the connection A) of the ordinary differential $\mathrm{d}s$.

From (8.5) and (8.6) we deduce at once

(8.7) *a critical connection for the Yang–Mills functional L is also critical for Φ, and the converse holds if ϕ is strictly convex.*

By a *strictly* convex function we mean as usual a function for which the second derivative ϕ'' is everywhere positive *definite*, so that the linear transformation $\phi''(s)$ in (8.6) is invertible. The quadratic function trace x^*x used for the Yang–Mills function is clearly strictly convex and any (weakly) convex ϕ becomes strongly convex if we add a positive multiple of trace x^*x. We see therefore that there are many functionals Φ that have precisely the same set of critical points as the Yang–Mills functional L.

So far we have not really used the convexity of ϕ, only the non-degeneracy of the second derivative ϕ''. Thus $-\phi$ would have had the same properties. The significance of convexity is that the Hessian of Φ always has a *finite Morse index*. To see this we compute as in § 4 and find that the Hessian H corresponds to the self-adjoint differential operator

$$\Phi_A = \phi''(*F(A))\,\mathrm{d}_A^*\mathrm{d}_A + \text{lower order}. \tag{8.8}$$

Since Φ is gauge invariant we can, as in § 4, restrict ourselves to ad (P)-valued 1-forms η with $\mathrm{d}_A^*\eta = 0$, so that in (8.8) $\mathrm{d}_A^*\mathrm{d}_A$ can be replaced by the covariant Laplacian

$$\Delta_A = \mathrm{d}_A^*\mathrm{d}_A + \mathrm{d}_A\mathrm{d}_A^*.$$

Then Φ_A becomes a second-order elliptic differential operator and the (strong) convexity of ϕ asserts that the leading-order terms are positive definite. This is enough to make the spectrum discrete and bounded below so that there are only finitely many negative eigenvalues, showing that H_Φ has finite Morse index.

Equations (8.5) and (8.6), together with the strong convexity of ϕ, imply that

$$(\text{grad}\,\Phi, \text{grad}\,L) \geqslant 0 \tag{8.9}$$

with equality only if A is a critical connection. This means that Φ is strictly decreasing along the paths of steepest descent for L, i.e. the trajectories of $-\text{grad}\,L$. In finite dimensions this would imply that at the common critical points the Morse indices of L and Φ coincide. In our situation this can be seen directly as follows. Expanding (8.9) at a critical connection A, and discarding higher-order terms, we deduce

$$(H_\Phi \eta, H_L \eta) \geqslant 0 \tag{8.10}$$

with equality only if η is in the null-space of H_L (which coincides with that of H_Φ). By restricting η to the negative space V of H_L we reduce (8.10) to a finite-dimensional inequality, which easily implies that H_Φ is negative definite on V (e.g. diagonalize H_L on V). Thus the Morse index of Φ is at least equal to that of L. Reversing the roles of L and Φ we get therefore

(8.11) *the Morse indices of Φ and L all coincide.*

To sum up we see that any one of our functionals Φ, defined by a strongly convex invariant function ϕ on $\mathfrak{g}$, has exactly the same critical point structure as the Yang–Mills functional L. Our next aim is to show that all such functionals lead in fact to the same Morse strata and that these strata coincide with the complex strata introduced in §7.

We now return to the identification of the space $\mathscr{A}$ of unitary connections with the space $\mathscr{C}$ of complex structures, explained at the beginning of this section, together with the actions of the groups $\mathscr{G}$ and $\mathscr{G}^c$ of unitary and complex automorphisms. The tangent space to the $\mathscr{G}$-orbit through A consists of vectors $\mathrm{d}_A\alpha$ with $\alpha \in \Omega^0(M, \text{ad}\,(P))$ while for the $\mathscr{G}^c$-orbit it consists of $\mathrm{d}''_A\beta$ with $\beta \in \Omega^0(M, \text{ad}\,(P^c))$. Since we are identifying $\Omega^1(M, \text{ad}\,(P))$ with $\Omega^{0,1}(M, \text{ad}\,(P^c))$ on which $* = \mathrm{i}$ we can say that the tangent space to the $\mathscr{G}^c$-orbit at A consists of vectors

$$\mathrm{d}_A\alpha_1 + *\,\mathrm{d}_A\alpha_2 \quad \text{with} \quad \alpha_1, \alpha_2 \in \Omega^0(M, \text{ad}\,(P)).$$

In particular then formula (8.5) for grad Φ shows that

(8.12) *grad Φ is tangential to the $\mathscr{G}^c$-orbits.*

In other words the 'gradient flow' of Φ preserves the $\mathscr{G}^c$-orbits and hence also the strata $\mathscr{C}_\mu$ of §7 since these are unions of orbits. Since the stratum $\mathscr{A}_\mu$ contains a unique component $\mathscr{N}_\mu$ of the critical set of Φ it is then reasonable to expect $\mathscr{A}_\mu$ to be just the Morse stratum or stable manifold of Φ associated with $\mathscr{N}_\mu$. For this to be true it is of course necessary that Φ on $\mathscr{A}_\mu$ should take its minimum on $\mathscr{N}_\mu$. Now for any $A \in \mathscr{N}_\mu$ the conjugacy class of $*F(A)$ is constant and is represented by the skew-hermitian diagonal matrix Λ_μ with entries $-2\pi\mathrm{i}\mu_j\,(j = 1, \ldots, n)$. Since the volume of M is normalized to be unity it follows that $\Phi(A)$ takes the constant value $\phi(\Lambda_\mu)$ and we shall write this simply as $\phi(\mu)$. For example, for the Yang–Mills functional, we have

$$\phi(\mu) = \|\mu\|^2 = \Sigma\mu_i^2.$$

Thus we might expect the following to hold.

PROPOSITION 8.13. *For any* $A \in \mathscr{A}_\mu$, *and for any convex invariant function* ϕ *on* $\mathfrak{u}(n)$, *we have* $\Phi(A) \geqslant \phi(\mu)$.

We shall begin by proving (8.13) in the simple case when

$$\mu_1 = \mu_2 = \ldots = \mu_r > \mu_{r+1} = \ldots = \mu_n$$

so that the canonical filtration of the bundle E has just two steps. This means that, for the holomorphic structure defined by $A \in \mathscr{A}_\mu$, we have an exact sequence of vector bundles

$$0 \to D_1 \to E \to D_2 \to 0,$$

where D_j has rank m_j, Chern class k_j $(j = 1, 2)$ so that $\mu_1 = k_1/m_1$ and $\mu_n = k_2/m_2$. For convenience we shall use the notation $\mu^j = k_j/m_j$ $(j = 1, 2)$. The curvature $F(A)$ can then be written in the form

$$F(A) = \begin{bmatrix} F_1 - \eta \wedge \eta^* & \mathrm{d}\eta \\ -\mathrm{d}\eta^* & F_2 - \eta^* \wedge \eta \end{bmatrix}, \tag{8.14}$$

where F_j is the curvature of the metric connection of D_j, $\eta \in \Omega^{0,1}(M, \mathrm{Hom}\,(D_2, D_1))$, η^* is its transposed conjugate and $\mathrm{d}\eta$ is the covariant differential. Now let f_j, α_j be scalar $m_j \times m_j$ matrices such that

$$\begin{gathered} \mathrm{trace}\, f_j = \mathrm{trace} * F_j, \\ \mathrm{trace}\, \alpha_1 = \mathrm{trace} * (\eta \wedge \eta^*) = -\mathrm{trace} * (\eta^* \wedge \eta) = -\mathrm{trace}\, \alpha_2. \end{gathered} \tag{8.15}$$

Then some elementary inequalities concerning convex invariant functions ϕ (which will be treated in §12) show that

$$\phi(*F(A)) \geqslant \phi \begin{bmatrix} f_1 - \alpha_1 & 0 \\ 0 & f_2 - \alpha_2 \end{bmatrix}.$$

The convexity of ϕ, together with the fact that M has normalized volume, then implies that

$$\Phi(A) = \int_M \phi(*F(A)) \geqslant \phi \int_M \begin{bmatrix} f_1 - \alpha_1 & 0 \\ 0 & f_2 - \alpha_2 \end{bmatrix}. \tag{8.16}$$

But the Chern class k_j of D_j is given by

$$k_j = \frac{\mathrm{i}}{2\pi} \int_M \mathrm{trace}\, f_j.$$

Since f_j is a scalar matrix this means that $\int_M f_j$ is a scalar matrix whose diagonal entries are $-2\pi \mathrm{i} k_j/m_j = -2\pi \mathrm{i} \mu^j$. Also from (8.15) (since $\eta \in \Omega^{0,1}$) it follows that $-\mathrm{i}$ trace α_1 is *non-negative* and so

$$\int_M \alpha_1 = 2\pi \mathrm{i} a_1,$$

where a_1 is a non-negative scalar $m_1 \times m_1$ matrix. Then

$$\int_M \alpha_2 = 2\pi \mathrm{i} a_2,$$

where a_2 is the non-positive scalar $m_2 \times m_2$ matrix such that trace $a_2 = -$ trace a_1. Hence we have

$$\int_M \begin{bmatrix} f_1 - \alpha_1 & 0 \\ 0 & f_2 - \alpha_2 \end{bmatrix} = -2\pi \mathrm{i} [\mu + a], \tag{8.17}$$

where [ ] denotes the diagonal matrix defined by a vector, so that $[a]$ denotes the matrix $\begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix}$.

From (8.16) and the convention we have adopted for defining ϕ on n-vectors (absorbing the factor $-2\pi i$) we then obtain

$$\Phi(A) \geqslant \phi(\mu + a). \tag{8.18}$$

But since $a_1 \geqslant 0$, $a_2 \leqslant 0$ with trace $a_1 = -$ trace a_2 it follows easily that $\mu + a \geqslant \mu$ with respect to the partial ordering defined in (7.7). As will be explained in §12 this implies, and in fact is equivalent to,

$$\phi(\mu + a) \geqslant \phi(\mu)$$

for all convex invariant ϕ. This completes the proof of proposition 8.13 for the two-step case. As can be seen the essence of the proof is the basic principle that 'curvature decreases in holomorphic sub-bundles and increases in quotients' (Griffiths & Harris 1978, p. 79). The general case of (8.13) proceeds in the same manner and we simply have to keep track of the notation. The details are as follows.

We start with a holomorphic bundle E with its canonical filtration of type μ:

$$0 = E_0 \subset E_1 \subset \ldots \subset E_r = E,$$

where the quotients $D_j = E_j/E_{j-1}$ have normalized Chern classes μ^j with

$$\mu^1 > \mu^2 > \ldots > \mu^r.$$

The curvature $F(A)$ can then be expressed in a block form generalizing (8.14). For every $j < k$ we have an element

$$\eta_{jk} \in \Omega^{0,1}(M, \mathrm{Hom}\,(D_k, D_j))$$

so that $\mathrm{d}\eta_{jk}$ appears in the (j,k)-block. The η_{jk} are the components of the element

$$\eta_k \in \Omega^{0,1}(M, \mathrm{Hom}\,(D_k, E_{k-1}))$$

related to the exact sequence

$$0 \to E_{k-1} \to E_k \to D_k \to 0.$$

Now define scalar non-negative $m_j \times m_j$ matrices a_{jk} for $j < k$ by

$$\mathrm{trace}\, a_{jk} = \frac{1}{2\pi i}\int_M \mathrm{trace}\,(\eta_{jk} \wedge \eta_{jk}^*) \geqslant 0,$$

and define a_{kk} by

$$\mathrm{trace}\, a_{kk} = \frac{1}{2\pi i}\int \mathrm{trace}\,\eta_k^* \wedge \eta_k \leqslant 0,$$

so that $\sum_{j \leqslant k} \mathrm{trace}\, a_{jk} = 0$. Then the convexity of ϕ leads to the inequality

$$\Phi(A) \geqslant \phi[\mu + a],$$

where a stands for the vector (or diagonal matrix) whose jth block a^j is the scalar (matrix)

$$a^j = \sum_{k \geqslant j} a_{jk}.$$

Equivalently the vector a can be written as a sum

$$a = \Sigma b_k$$

where b_k is the vector corresponding to the diagonal matrix whose jth block is a_{jk} for $j \leqslant k$ (and zero for $j > k$). The fact that

$$\operatorname{trace} a_{jk} \geqslant 0 \quad \text{for} \quad j < k \quad \text{and} \quad \sum_{j \leqslant k} \operatorname{trace} a_{jk} = 0$$

implies that $b_k \geqslant 0$ relative to the partial ordering (7.7). Hence $a = \sum b_k \geqslant 0$ and so $\mu + a \geqslant \mu$. As before this then implies that $\phi(\mu + a) \geqslant \phi(\mu)$ and so completes the general proof of proposition 8.13.

It will be noted that we have nowhere used the maximal nature of the canonical filtration, i.e. the fact that the quotients D_j are semi-stable. Once we use this we shall be able to strengthen (8.13). For this we shall need to use the Narasimhan–Seshadri result (8.1).

If, for any holomorphic vector bundle E over M, and any convex invariant ϕ we define

$$\Phi(E) = \inf \Phi(A),$$

where A runs over all metric connections on E, then (8.1), together with proposition 8.13, implies that for *stable bundles* E we have $\Phi(E) = \phi(\mu)$. We shall now extend this to all bundles. First suppose we have a holomorphic exact sequence

$$0 \to D_1 \to E \to D_2 \to 0.$$

Then a metric on E gives rise to a connection whose curvature is given by (8.14). The element $\eta \in \Omega^{0,1}(M, \operatorname{Hom}(D_2, D_1))$ defines a cohomology class $[\eta] \in H^1(M, \operatorname{Hom}(D_2, D_1))$, which classifies the extension. Replacing η by $t\eta$ with t a non-zero constant alters the extension class but does not alter the isomorphism class of E, since the new extension is isomorphic to the original by a diagram of the form

$$\begin{array}{ccccccccc} 0 & \longrightarrow & D_1 & \longrightarrow & E & \longrightarrow & D_2 & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle t} & & \downarrow & & \downarrow{\scriptstyle 1} & & \\ 0 & \longrightarrow & D_1 & \longrightarrow & E & \longrightarrow & D_2 & \longrightarrow & 0. \end{array}$$

Hence replacing η by $t\eta$ and then letting $t \to 0$ shows that

$$\Phi(E) \leqslant \Phi(D_1 \oplus D_2).$$

Similarly if E has a filtration of arbitrary length with quotients D_j we have

$$\Phi(E) \leqslant \Phi(\textstyle\bigoplus D_j). \tag{8.19}$$

Now an elementary argument (see Seshadri 1967) shows that any semi-stable bundle has a filtration with stable quotients all of which have the same normalized Chern class. From this, together with (8.13) and the result for stable bundles, it follows that we have the equality $\Phi(E) = \phi(\mu)$ for all semi-stable bundles. Finally using the canonical filtration of any E, (8.19) and (8.13) yield the equality in general. Thus we have now established

PROPOSITION (8.20). *If a holomorphic bundle E is of type μ then for any convex invariant ϕ we have $\Phi(E) = \phi(\mu)$, where*

$$\Phi(E) = \inf_A \int_M \phi((*F(A))$$

and A runs over all metric connections on E.

Now in § 12 we shall see that if μ, ν are any two n-vectors (with $\mu_1 \geqslant \ldots \geqslant \mu_n$ and $\nu_1 \geqslant \ldots \geqslant \nu_n$) then

$$\phi(\mu) = \phi(\nu) \quad \textit{for all convex invariant } \phi \quad \Rightarrow \mu = \nu. \tag{8.21}$$

In view of this we see that (8.20) amounts to a differential–geometric characterization of the type, i.e. we have

(8.22) *E is of type μ if and only if $\Phi(E) = \phi(\mu)$ for all convex invariant ϕ.*

Since $\Phi(E)$ is defined by an infimum it follows that

$$\mathscr{C}_\lambda \text{ in the closure of } \mathscr{C}_\mu \Rightarrow \Phi(\mathscr{C}_\lambda) \leqslant \Phi(\mathscr{C}_\mu)$$
$$\Rightarrow \phi(\lambda) \leqslant \phi(\mu) \quad \text{by (8.20).}$$

As will be shown in § 12

(8.23) $$\phi(\lambda) \leqslant \phi(\mu) \quad \text{for all } \phi \quad \Leftrightarrow \lambda \leqslant \mu,$$

where $\lambda \leqslant \mu$ refers to the partial ordering (7.7). Hence we have established by differential–geometric means the result (7.8) of Shatz.

In § 5 we computed the index of the Yang–Mills functional at any critical point, and we obtained in proposition 5.4 the formula

$$\text{index } A = 2 \dim_C H^1(M, \text{ad}^-(P)).$$

If the critical point A is of type μ, so that the curvature is $-2\pi i$ times the diagonal matrix given by μ, then the holomorphic bundle E determined by the connection is a direct sum

$$E = \bigoplus D_j,$$

where the D_j are semi-stable and have as normalized Chern classes the distinct components of μ. We see therefore that the bundle $\text{ad}^-(P)$ of § 5 coincides with the bundle $\text{End}'' E$ of § 7, both being isomorphic (see (7.13)) to $\bigoplus_{j<k} \text{Hom}(D_j, D_k)$. Hence the index of A is equal to the codimension of the stratum $\mathscr{A}_\mu$ containing A. In fact the normal to $\mathscr{A}_\mu$ at A actually coincides with the negative eigenspace of the Hessian of L at A, both being given by the appropriate space of harmonic forms. In view of (8.11) it follows that the codimension of $\mathscr{A}_\mu$ is equal to the Morse index of any of our functionals Φ.

To sum up we see that the $\mathscr{A}_\mu$ play the role of the Morse strata not only for the Yang–Mills functional L but more generally for any functional Φ defined by a strongly convex invariant function ϕ on the Lie algebra of $U(n)$. This statement is to be understood in the sense that our strata $\mathscr{A}_\mu$ satisfy all the properties of (1.19) relative to Φ, which in good cases, as explained in § 1, characterize the Morse strata. This suggests that each critical set $\mathscr{N}_\mu$ should be an equivariant deformation retract of the corresponding stratum $\mathscr{A}_\mu$, the retraction being given by following the trajectories of $-\text{grad}\,\Phi$. To prove this it would be enough to check it for the minimal stratum (for all $U(n)$). In the coprime case $(n, k) = 1$ this is a consequence of (8.2), but in general the singularities of $\mathscr{N}_\mu$ give rise to difficulties and we shall not pursue this question further. Thus although we have shown that the stratification of $\mathscr{A}$ by the $\mathscr{A}_\mu$ is equivariantly perfect we have not actually proved that the Yang–Mills functional is an equivariantly perfect Morse function, although this seems very likely and would follow from sufficiently good properties about the Yang–Mills flow.

9. Cohomology of the moduli spaces

We have now shown how to compute inductively the equivariant cohomology of the space $\mathscr{C}_{ss}$ of semi-stable bundles. In this section we shall show how to derive the integral cohomology of the moduli space $N(n, k)$ in the coprime case $(n, k) = 1$, and also that of the moduli space

$N_0(n,k)$ for bundles with fixed determinant. First, however, we shall need to extend theorem 7.14 by replacing the group $\mathcal{G}$ with a subgroup of finite index. As explained in §2, the group Γ of components of $\mathcal{G}$ is $H^1(M,Z) \cong Z^{2g}$, and so a subgroup $\mathcal{G}'$ of $\mathcal{G}$ of finite index is specified by giving a sublattice of maximal rank $\Gamma' \subset \Gamma$. As shown in §2 the classifying space $B\mathcal{G}'$, which is a finite covering of $B\mathcal{G}$, has no torsion, it has the same Poincaré series as $B\mathcal{G}$ and Γ/Γ' acts trivially on its cohomology. We now consider our stratification of $\mathcal{A}$ relative to $\mathcal{G}'$. We proceed exactly as with $\mathcal{G}$. The only point to comment on is that the space $\mathcal{F}_\mu$ occurring in §7 (namely the space of all C^∞ filtrations of E of type μ) is *connected*. In more concrete terms this means that any two filtrations of E of type μ are homotopic. To see this we note first that, over the 1-skeleton of M, all bundles are trivial and all filtrations homotopic (since the partial flag manifolds of $U(n)$ are all simply connected). We can therefore collapse the 1-skeleton to a point and reduce to the case $M = S^2$, but now $\mathcal{G}$ becomes connected and so two filtrations of the same type, being isomorphic, are necessarily homotopic. Hence $\mathcal{F}_\mu$ is equally a homogeneous space of $\mathcal{G}'$ and if $\mathcal{F}_\mu = \mathcal{G}/\mathcal{H} = \mathcal{G}'/\mathcal{H}'$ then $\mathcal{H}' \subset \mathcal{H}$ is of finite index and corresponds to the same sublattice Γ' of Γ.

Thus our stratification of $\mathcal{C}$ is also perfect relative to $\mathcal{G}'$. In particular the $\mathcal{G}'$-equivariant cohomology of $\mathcal{C}_{ss}$ has no torsion and it is acted on trivially by Γ/Γ', so that the $\mathcal{G}'$ and $\mathcal{G}$ Poincaré series of $\mathcal{C}_{ss}$ coincide.

We move on now to consider the coprime case $(n,k) = 1$. Then stable and semi-stable coincide, so that $\mathcal{C}_s = \mathcal{C}_{ss}$ and Aut E acts on $\mathcal{C}_s(E)$ with only the constant central scalars as isotropy group (Narasimhan & Seshadri 1965). The *moduli space* $N(n,k)$ is then the quotient of $\mathcal{C}_s(E)$ by this action. It is a compact complex manifold: it even inherits a natural Kähler structure as we shall see later. We want now to deduce what we can about the cohomology of $N(n,k)$ from our general results about equivariant cohomology.

Let us denote by $\bar{\mathcal{G}}$ the quotient of $\mathcal{G}$ by its constant central $U(1)$-subgroup, and similarly $\bar{\mathcal{G}}^c$ will be the quotient of $\mathcal{G}^c = \text{Aut}\,(E)$ by C^*. Thus $\bar{\mathcal{G}}^c$ acts freely on $\mathcal{C}_s$ with $N(n,k)$ as quotient. Hence (for any coefficients)

$$H^*(N(n,k)) \cong H^*_{\bar{\mathcal{G}}}(\mathcal{C}_s) \tag{9.1}$$

where on the right we have replaced $\bar{\mathcal{G}}^c$ by $\bar{\mathcal{G}}$ since they give the same cohomology. It remains to investigate the relation between $\mathcal{G}$-cohomology and $\bar{\mathcal{G}}$-cohomology. This depends on the fibration

$$BU(1) \to B\mathcal{G} \to B\bar{\mathcal{G}}, \tag{9.2}$$

which is always trivial in rational cohomology. This is because restriction to a point of M followed by taking determinants defines a homomorphism $\mathcal{G} \to U(1)$ and the composition $U(1) \to \mathcal{G} \to U(1)$ is of degree n. This implies that

$$H^*(B\mathcal{G}, Q) \to H^*(BU(1), Q)$$

is surjective, which gives the triviality of the fibration over Q. Hence for any $\bar{\mathcal{G}}$-space X the $\mathcal{G}$-Poincaré series of X is the product of the $\bar{\mathcal{G}}$-Poincaré series of X and $(1-t^2)^{-1}$. Together with (9.1) this then gives the formula for the Poincaré series of $N(n,k)$:

$$P_t(N(n,k)) = (1-t^2)\,\mathcal{G}P_t(\mathcal{C}_s) \tag{9.3}$$

where $\mathcal{G}P_t(\mathcal{C}_s)$ is given inductively by theorem 7.14. As noted in the Introduction, and will be elaborated in §11, this formula coincides with that of Desale & Ramanan (1975), which rests on the Harder–Narasimhan approach.

Taking determinants gives a natural map

$$\det: N(n,k) \to J_k, \tag{9.4}$$

where J_k is the Jacobian of M, parametrizing line-bundles of degree k. Clearly J_0 acts on $N(n,k)$ by tensor product and the determinant becomes a J_0-equivariant map if we make J_0 act on J_k via the nth power map

$$\sigma_n: J_0 \to J_0.$$

This shows that after lifting to a finite covering, with group $\operatorname{Ker}\sigma_n \cong H^1(M, Z_n) \cong \Gamma_n = \Gamma/n\Gamma$, (9.4) becomes a product. Thus if we denote by $N_0(n,k)$ the fibre of (9.4) then

$$N(n,k) = (N_0(n,k) \times J_k)/\Gamma_n. \tag{9.5}$$

The manifold $N_0(n,k)$ is the moduli space of stable bundles *with fixed determinant*. If we now take $\mathscr{G}' \subset \mathscr{G}$ corresponding to the lattice $n\Gamma \subset \Gamma$, so that $\mathscr{G}/\mathscr{G}' = \Gamma/n\Gamma = \Gamma_n$ then the analogue of (9.1) becomes

$$H^*(N_0(n,k) \times J_k) \cong H^*_{\bar{\mathscr{G}}}(\mathscr{C}_s). \tag{9.6}$$

Since $\mathscr{G}$ and $\mathscr{G}'$ give the same equivariant cohomology of $\mathscr{C}_s$ (over Q) the same holds for $\bar{\mathscr{G}}$ and $\bar{\mathscr{G}}'$, and so comparing (9.1) and (9.6) we get

PROPOSITION 9.7. *For rational cohomology we have*

$$H^*(N(n,k)) \cong H^*(N_0(n,k)) \otimes H^*(J)$$

or in terms of Poincaré polynomials

$$P_t(N(n,k)) = P_t(N_0(n,k))\,(1+t)^{2g}.$$

This proposition, which is equivalent to saying that Γ_n acts trivially on the rational cohomology of $N_0(n,k)$, was the main result of Harder & Narasimhan (1975) where it was proved by number-theoretic methods comparing $GL(n)$ with $SL(n)$. For us the triviality of the action of Γ_n is a consequence of its triviality on the cohomology of $B\mathscr{G}'$.

We turn next to the integral cohomology of the moduli space $N(n,k)$. We want to prove that it has no torsion. We already know by (7.17) that $H_{\mathscr{G}}(\mathscr{C}_s)$ has no torsion and, in view of (9.1), we want to deduce the same result for $H_{\bar{\mathscr{G}}}(\mathscr{C}_s)$. It will be sufficient to prove that the fibration (9.2) is in fact a product so that

$$H_{\mathscr{G}}(\mathscr{C}_s) \cong H_{\bar{\mathscr{G}}}(\mathscr{C}_s) \otimes H(BU(1)).$$

Now $BU(1)$ is an Eilenberg–Maclane space $K(Z,2)$ and so (9.2) has a characteristic class in $H^3(B\bar{\mathscr{G}}, Z)$ whose vanishing will imply the triviality of the fibration. Equivalently we need to show that

$$H^2(B\mathscr{G}, Z) \to H^2(BU(1), Z) \tag{9.8}$$

is surjective, but this was the content of proposition 2.21. Thus we have now proved

THEOREM 9.9. *If $(n,k) = 1$ the moduli space $N(n,k)$ of stable bundles has no torsion.*

For the space $N_0(n,k)$ we use the commutative diagram of fibrations

$$\begin{array}{ccccc} BU(1) & \longrightarrow & (\mathscr{C}_s)_{\mathscr{G}'} & \longrightarrow & N_0 \times J \\ \| & & \downarrow & & \downarrow \\ BU(1) & \longrightarrow & (\mathscr{C}_s)_{\mathscr{G}} & \longrightarrow & N \end{array}$$

where $N_0 \times J \to N$ is the finite Γ_n-covering. Since the bottom row has now been shown to be a product the same is true for the top row. Since we showed earlier that $\mathscr{C}_s$ has no torsion for its $\mathscr{G}'$-cohomology it follows that $N_0 \times J$ has no torsion, and hence also N_0 has no torsion. Thus we have

THEOREM 9.10. *If $(n, k) = 1$ the moduli space $N_0(n, k)$ of stable bundles with fixed determinant has no torsion.*

The triviality of the fibration (9.2) when $(n, k) = 1$ is essentially equivalent to the existence of a (topological) universal or Poincaré bundle over $M \times N$ as we shall now explain. By definition a universal bundle is a holomorphic vector bundle V over $M \times N$ so that for all $n \in N$ the restriction V_n of V to $M \times \{n\}$ is in the isomorphism class represented by the point n. The projective bundle $P(V)$ exists naturally. To see this we recall that we have an obvious holomorphic bundle W over $M \times \mathscr{C}_{ss}$ and $\mathscr{G}^c$ acts holomorphically on W with the constant scalars C^* acting trivially on the base and as scalars in the fibre of W. Thus $\bar{\mathscr{G}}^c = \mathscr{G}^c/C^*$ acts freely on $P(W)$ and the quotient gives $P(V)$ over $M \times N$. A universal vector bundle is therefore a 'lift' back from this natural projective bundle over $M \times N$. If V is holomorphic on each M_n but only *continuous* in n we refer to it as a *topological* universal bundle.

Quite generally there is an obstruction to the extension of such a lift called the 'Brauer class'. It arises from the sequence

$$1 \to C^* \to GL(n) \to PGL(n) \to 1$$

and lies in $H^2(M \times N, \mathcal{O}^*)$ where $\mathcal{O}^*$ is the sheaf of multiplicative holomorphic functions. Taking the coboundary of the exponential sequence

$$0 \longrightarrow Z \xrightarrow{e^{2\pi i}} \mathcal{O} \longrightarrow \mathcal{O}^* \longrightarrow 0$$

leads to the topological Brauer class, which is an n-torsion class in $H^3(M \times N, Z)$. Explicitly in terms of transition matrices p_{ij} for a $PGL(n)$-bundle we lift these locally to g_{ij} in $GL(n)$ and define the scalar λ_{ijk} by the formula

$$g_{ij}g_{jk} = \lambda_{ijk}g_{ik}.$$

This is a 2-cocycle for $\mathcal{O}^*$ and taking determinants shows that λ_{ijk}^n is a coboundary.

In our case since M and N (by (9.9)) are torsion-free it follows that the topological Brauer class must be zero and from this one can deduce that a topological universal vector bundle V does indeed exist. In fact our proof of (9.9) depended essentially on the triviality of (9.2) and this in slightly disguised form is equivalent to the vanishing of the topological Brauer class, as one might suspect from the fact that the characteristic class of (9.2) is an element of $H^3(B\bar{\mathscr{G}}, Z)$. To explain this we note that the bundle W on $M \times \mathscr{C}_{ss}$ gives rise naturally to a bundle $W_{\mathscr{G}}$ on $M \times (\mathscr{C}_{ss})_{\mathscr{G}}$: this bundle is holomorphic only in the M-directions, since $(\mathscr{C}_{ss})$ is only a topological space. Passing to the projective bundles we see that $P(W)_{\bar{\mathscr{G}}}$ lifts to $P(W_{\mathscr{G}})$ under the natural map

$$\pi : M \times (\mathscr{C}_{ss})_{\mathscr{G}} \to M \times (\mathscr{C}_{ss})_{\bar{\mathscr{G}}}.$$

Thus the Brauer class of $P(W)_{\bar{\mathscr{G}}}$ lifts under π^* to the Brauer class of $P(W_{\mathscr{G}})$, which is zero since this bundle comes from the vector bundle $W_{\mathscr{G}}$. On the other hand the fibration π is trivial, since (9.2) is trivial, and so the Brauer class of $P(W)_{\bar{\mathscr{G}}}$ must itself vanish. More explicitly any section σ of π induced by a section of (9.2) defines a vector bundle σ^*W that has $P(W)_{\bar{\mathscr{G}}}$ as projective bundle. Finally we have only to observe that, homotopically,

$$(\mathscr{C}_{ss})_{\bar{\mathscr{G}}} \sim N$$

and that $P(W)_{\bar{\mathcal{G}}}$ corresponds (topologically) to the projective bundle $P(W)$, while σ^*W corresponds to a topological universal bundle V. This shows the tie-up between the different points of view.

In fact in this coprime case a more refined argument shows that the analytic Brauer class in $H^2(M\times N, \mathcal{O}^*)$ is zero, so that a *holomorphic* universal bundle exists. Let us recall briefly the essential point of the proof, which is to construct a holomorphic line-bundle L over $M\times\mathcal{C}_{ss}$ on which $\mathcal{G}^c$ acts such that $C^*\subset\mathcal{G}^c$ acts via scalar multiplication in the fibres of L. Then C^* acts trivially on $W\otimes L^{-1}$ and so this is acted on by the quotient group $\bar{\mathcal{G}}^c=\mathcal{G}^c/C^*$. We can now descend the bundle $W\otimes L^{-1}$ to the quotient space

$$(M\times\mathcal{C}_{ss})/\bar{\mathcal{G}}^c = M\times N$$

to obtain the universal bundle. The bundle L is constructed on $\mathcal{C}_{ss}$ (and then lifted to $M\times\mathcal{C}_{ss}$), by using the vector bundles given by the various H^0. More precisely if k is sufficiently large (the precise values will be given later) then for any semi-stable bundle of rank n and Chern class k we have

$$H^1(M,E) = 0$$

$$\dim H^0(M,E) = k-n(g-1).$$

This gives a holomorphic bundle of dimension $k-n(g-1)$ over $\mathcal{C}_{ss}$. Taking determinants (i.e. the highest exterior power) gives a line bundle A_k on $\mathcal{C}_{ss}$. The group C^* of scalar automorphisms of E acts on this with weight $m=k-n(g-1)$, i.e. $\lambda\in C^*$ acts by multiplication with λ^m. Tensoring E with a fixed line-bundle of Chern class 1 replaces k by $k+n$ so giving a line bundle A_{k+n} over $\mathcal{C}_{ss}$ on which C^* acts with weight $m+n$. Since $(k,n)=1$ we have $(m,m+n)=1$ and we can find integers a,b such that

$$am+b(m+n)=1.$$

Hence $L=A_k^a\otimes A_{k+n}^b$ is acted on with weight 1 and leads to the universal bundle.

We note finally that the universal bundle is not unique and can be altered by tensoring with any holomorphic line-bundle L on $M\times N$. On M the universal property implies that L must have degree zero (and must moreover satisfy $L^n=1$) but the Chern class on N is arbitrary, as is the component in $H^1(M)\otimes H^1(N)$.

We recall that in §2 we proved that the integral cohomology ring $H^*(B\mathcal{G})$ was generated multiplicatively by certain explicit classes constructed from the canonical bundle on $M\times B\mathcal{G}$. This canonical bundle restricts to W when we embed $(\mathcal{C}_{ss})_{\mathcal{G}}$ in $\mathcal{C}_{\mathcal{G}}=B\mathcal{G}$, and since our stratification of $\mathcal{C}$ is equivariantly perfect (theorem 7.14) it follows that our generators for $H^*(B\mathcal{G})$ restrict to give generators for $H^*((\mathcal{C}_{ss})_{\mathcal{G}})$ and hence generators for $H^*(N)$ (after pulling back by a section σ of the fibration π). Since $\sigma^*W\cong V$ (topologically), where V is the universal bundle on $M\times N$ we see finally that the integral cohomology ring $H^*(N)$ is multiplicatively generated by explicit elements constructed (as in §1) from the universal bundle V on $M\times N$. These classes are of three types.

(i) The Chern classes a_r of V restricted to N.

(ii) The odd-dimensional classes b_r^j $(j=1,\ldots,2g)$, which occur in the $(1,2r-1)$ Künneth component of the rth Chern class of V on $M\times N$.

(iii) The Chern classes d_r of $f_!(V)\in K(N)$.

Note that in (i) and (ii) r runs from 1 to n while in (iii) it is unrestricted. Now since V is holomorphic its K-theory direct image $f_!$ may be computed directly. If $H^1(M,V_y)=0$ for all $y\in N$

then $f_!(V)$ is simply the vector bundle H on N whose fibre at y is $H^0(M, V_y)$. Now by Serre duality $H^1(M, V_y) = 0$ provided every homomorphism $V_y \to K$ (where K is the canonical line-bundle) is zero. By (7.5) this will hold provided $k/n > 2g-2$ in which case Riemann–Roch gives the dimension of $H^0(M, V_y)$ as $k-n(g-1)$. Now by tensoring with line-bundles we can always arrange that k is of the form

$$k = (2g-2)n + k' \quad \text{with} \quad 0 < k' < n$$

(recall that $(k, n) = 1$), so that we then have

$$\dim H = n(g-1) + k'.$$

Finally then we have proved the following theorem (cf. Newstead 1972 for $n = 2$).

THEOREM 9.11. *Let $k = (2g-2)n + k'$ with $0 < k' < n$ and $(k, n) = 1$, and let V be a universal bundle over $M \times N(n, k)$. Define integral cohomology classes a_r, b_r^j, d_r on N by*

$$a_r = c_r(V|N), \quad 1 \leqslant r \leqslant n,$$

$$\sum_{j=1}^{2g} \alpha_j \otimes b_r^j = c_r(V), \quad 1 \leqslant r \leqslant n \quad \text{and} \quad \alpha_j \text{ a basis for } H^1(M),$$

$$d_r = c_r(H(V)), \quad 1 \leqslant r \leqslant n(g-1) + k',$$

where $H(V)$ is the bundle over N whose fibre at y is $H^0(M, V_y)$. Then the integral cohomology ring of N is generated by these classes.

The moduli spaces N and N_0 are torsion-free when $(n, k) = 1$, and theorem 9.11 provides us with a system of integral generators $\{a_r, b_r^j, d_r\}$ while our Poincaré series formulae determine the dimensions of their span. Hence these rings are in principle determined once a complete set of relations for their generators is written down over the rationals. Ideally one should be able to derive these from the Thom classes of the various strata of $\mathscr{A}$, but we have been unable to make much headway in that direction, except for the computation of the fundamental group.

Note that the complex codimension d_μ of any stratum $\mathscr{C}_\mu$ other than the semi-stable stratum $\mathscr{C}_{ss}$, as given by (7.16), satisfies

$$d_\mu \geqslant 1 + (g-1) \geqslant 2 \quad \text{if} \quad g \geqslant 2.$$

This implies as in (1.12) that

$$\pi_1(\mathscr{C}_{ss}) \cong \pi_1(B\mathscr{G}) \cong H_1(M, Z).$$

On the other hand the triviality of (9.2) shows that

$$\pi_1((\mathscr{C}_{ss})_{\mathscr{G}}) \cong \pi_1((\mathscr{C}_{ss})_{\bar{\mathscr{G}}}) \cong \pi_1(N).$$

Hence

$$\pi_1(N) \cong H_1(M, Z),$$

and this isomorphism is naturally induced by the determinant map to the Jacobian

$$N \xrightarrow{\det} J.$$

Thus the fibre N_0 of this map is simply connected. We recall that N_0 is the moduli space for stable bundles with fixed determinant. Thus we have proved

THEOREM 9.12. *The moduli space of stable bundles of fixed determinant, and with $(n, k) = 1$, is simply connected.*

Remark. This result can also be deduced from the fact that N_0 is at least uni-rational. Its rationality is conjectured but not yet proved.

Returning to the general problem of computing the relations in the cohomology ring, a careful analysis of the implications of the Riemann–Roch theorem applied to $f_!$ has a good chance of succeeding, as was already noted independently by D. Mumford long before our involvement. The hope is to derive all the necessary relations from the vanishing of the Chern classes of $f_!V_k$ beyond their dimension.

To provide evidence for this conjecture we shall discuss the rank two case in some detail below and roughly compare our relations with those obtained by Ramanan (1973) for genus three. Mumford is investigating this question more generally with the aid of a computer and has verified it up to genus five. But first it is expedient to make some general remarks on the normalization of the V_k and their relation to the tangent bundle T of N. This material can also be found in Ramanan's paper but is considerably simpler in our context because theorem 9.11 furnishes us with integral generators that are *a priori* Chern classes of holomorphic line bundles. For simplicity we shall only treat the case $k = 2n(g-1)+1$, and write $\bar{g}$ for $g-1$.

Recall now that V_k is not unique, though its projective class is. It follows that we may twist any V_k by the pullback $f^{-1}L$ of any holomorphic line bundle on N relative to the projection

$$M \times N \xrightarrow{f} N.$$

Under such a twist our generators $a_1 = c_1(V)$ and $d_1 = c_1(f_!V)$ change by $nc_1(L)$ and $(n\bar{g}+1)c_1(L)$ respectively. Hence $\bar{g}a_1 - d_1$ changes by $c_1(L)$. But $\bar{g}a_1 - d_1$ is the Chern class of the holomorphic line-bundle $L = A_1^q D_1^{-1}$ where $A_1 = \det(V_k|N)$ and $D_1 = \det f_!(V)$. It follows that the bundle

$$V = V_k \otimes f^{-1}L^{-1}$$

is now determined up to isomorphism and is called the 'normalized universal bundle over N'.

In what follows all our generators will be associated to this normalized V, so that in particular

$$\bar{g}a_1 = d_1.$$

We next relate T to V in the K-theory of N. For each $y \in N$, the tangent space to N at y is canonically given by $H^1(M; \operatorname{End} V_y)$. Furthermore, as stable bundles admit only trivial automorphisms, $H^0(M; \operatorname{End} V_y) = C$. Thus in the K-theory of N

$$1 - T = f_!(\operatorname{End} V).$$

Next observe that if, as before, Ω^1 denotes the line bundle of holomorphic 1-forms along the fibre of f, then by Serre duality

$$H^1(M; \operatorname{End} V_y)^* \cong H^0(M; \operatorname{End} V_y \otimes \Omega^1),$$

whence

$$T^* - 1 = f_!(\operatorname{End} V \otimes \Omega^1)$$

so that subtracting these two expressions yields the relation

$$T + T^* - 2 = f_!\{\operatorname{End} V \otimes (\Omega^1 - 1)\}.$$

As a first corollary of these relations we prove the following proposition.

PROPOSITION 9.13 (Ramanan). *The cohomology group $H^2(N_0; Z)$ is infinite cyclic and is generated by half the first Chern class of N_0.*

Proof. Recall that the Riemann–Roch theorem in our present simple context is given by the formula

$$\operatorname{ch}(f_!W) = f_*\{\operatorname{ch}(W)\}\{1 - \bar{g}\omega\},$$

where $\omega \in H^2(M)$ is the orientation class on M, W is any holomorphic bundle on $M \times N$, and f_* denotes integration over the fibre M.

Applied to End V this leads to the relation

$$-\mathrm{ch}_1(T) = \mathrm{ch}_1(f_!\,\mathrm{End}\,V) = f_*\,\mathrm{ch}_2(\mathrm{End}\,V),$$

while with $W = V$ it yields the formula:

$$d_1 = \mathrm{ch}_1(f_!V) = -\bar{g}f_*\{\mathrm{ch}_1(V)\cdot\omega\} + f_*(\mathrm{ch}_2\,V).$$

Here of course ch_i denotes the part of the character of dimension $2i$; so that in terms of the Chern-classes

$$\mathrm{ch}_1 = c_1 \quad \mathrm{ch}_2 = \tfrac{1}{2}c_1^2 - c_2.$$

Now recall our definitions of the a_r and b_r as Künneth components relative to a fixed base $\alpha_j \in H^1(M)$. They imply the formulae

$$c_r = a_r + \sum_j \alpha_j b_r^j + f_r\omega,$$

where $c_r = c_r(V)$ and a_r, f_r, b_r^j are classes on N identified with their pullback to $M \times N$. We also write

$$c_r = a_r + \xi_r + f_r\omega$$

for these relations, so that

$$\xi_r = \sum \alpha_j b_r^j.$$

As a consequence note that the ξ_r are nilpotent of order three: $\xi_r\xi_s\xi_t = 0$, and that $\xi_r\omega = 0$, while $\xi_r\xi_s$ is a multiple of ω. We therefore set

$$\xi_r^2 = 2\Lambda_{rr}\omega$$

and

$$\xi_r\xi_s = \Lambda_{rs}\omega, \quad r \neq s.$$

In terms of the skew form μ_{ij} given by the intersection pairing in $H^1(M)$, that is

$$\alpha_i\alpha_j = \mu_{ij}\omega,$$

we have

$$\Lambda_{rr} = -\tfrac{1}{2}\sum \mu_{ij}b_r^i b_r^j, \quad \Lambda_{rs} = -\sum \mu_{ij}b_r^i b_s^j$$

so that these are integral non-degenerate forms in the b_r. They are pertinent for our purposes because of the following easily proved result.

The push-forward f_*c_α of any monomial in the c_i, is given by a universal polynomial $P_\alpha(a, \Lambda, f)$ in terms of the variables a_i, f_i and Λ_{ij}, $i \leqslant j = 1, \ldots, n$.

For example

$$f_*c_1^2 = f_*(a_1^2 + 2a\xi + 2f_1\omega + 2\Lambda_{11}\omega)$$

$$= 2(a_1 + \Lambda_{11}),$$

so that by our formula for d_1

$$d_1 = -\bar{g}a_1 + a_1f_1 + \Lambda_{11} - f_2.$$

Now $f_1 = 2n\bar{g} + 1$, as we are dealing with V, whence

$$d_1 = \{(2n-1)\bar{g} + 1\}a_1 + \Lambda_{11} - f_2.$$

Together with the normalization $d_1 = \bar{g}a_1$ this yields the formula

$$f_2 = \{(2n-1)\bar{g} + 1\}a_1 + \Lambda_{11},$$

and at this stage we are ready to compute $c_1(T)$. Recall first that for any n-dimensional bundle V

$$c_2(\mathrm{End}\, V) = (n-1)\, c_1^2(V) - 2nc_2(V).$$

Hence
$$\begin{aligned} c_1(f_!\,\mathrm{End}\, V) &= \mathrm{ch}_1(f_!\,\mathrm{End}\, V) \\ &= -f_*\{(n-1)\, c_1^2 - 2nc_2\} \\ &= -\{2(n-1)\, a_1 f_1 + \Lambda_{11}) - 2nf_2\} \\ &= -2(a_1 + \Lambda_{11}), \end{aligned}$$

so that

$$c_1(T) = 2(a_1 + \Lambda_{11}).$$

The first part of our proposition now follows by restriction to N_0. The second part also follows because a_1 must generate $H^2(N_0, Z)$ by theorem 9.11. Note also that $\dim H^2(N_0) = 1$ from the Poincaré series for N_0.

Before proceeding to a more detailed account of the case $n = 2$, observe that if we define the total Pontryagin class $p(T)$ of T as the product of the total Chern classes of T and T^*,

$$p(T) = c(T)\, c(T^*),$$

then our relation for $T + T^*$ in $K(N)$ implies the formula

$$p(T) = c(\mathrm{End}|N_0)^{2\bar{g}}.$$

This follows from the Riemann–Roch theorem, or also from the fact that the support of $\Omega^1 - 1$ is at a point of M and $c_1(\Omega^1 - 1) = 2\bar{g}$. The formula is especially simple in the rank 2 case where (cf. Newstead 1972)

$$c(\mathrm{End}\, V|N) = 1 + (a_1^2 - 4a_2).$$

Thus the ring Pont (T) generated by all the Pontryagin classes of T is actually generated by the single element $(a_1^2 + 4a_2) = -p_1$.

Newstead made two conjectures about the Pontryagin and Chern classes of T: first of all that $c_i(T) = 0$ for $i > 2\bar{g}$, and secondly that Pont $(T) = 0$ in dim $> 4\bar{g}$. The first of these conjectures has been recently proved by Gieseker (1982). The second conjecture remains open but, in view of our formula for $p(T)$, is equivalent to the assertion that $p_1^g = 0$.

We turn now finally to a more detailed examination of the relations we are after in the rank 2 case. Here we have to deal with only two Chern classes for V and they are given by

$$\begin{aligned} c_1 &= a_1 + \xi_1 + k\omega, \quad k = 2n\bar{g} + 1, \\ c_2 &= a_2 + \xi_2 + \{(2\bar{g} + 1)\, a_1 + \Lambda_{11}\}\, \omega, \end{aligned}$$

where we have substituted for f_2 the expression already found earlier. Applying the Riemann–Roch procedure one therefore obtains universal polynomials $Q_q(a; \Lambda)$ in $(a_1, a_2, \Lambda_{11}, \Lambda_{12}, \Lambda_{22})$ such that

$$c_q(f_!\, V) = Q_q(a; \Lambda).$$

Hence the relations take the form

$$Q_q(a; \Lambda) = 0 \quad \text{for} \quad q \geqslant 2g,$$

so that the 'first' of these asserts that

$$Q_{2g}(a; \Lambda) = 0.$$

To analyse the implications of this relation further, recall the diagram

$$\begin{array}{ccc} N_0 & \xrightarrow{i} & N, \\ & {\scriptstyle\eta}\searrow \quad \swarrow{\scriptstyle\pi} & \\ & N/J_0 & \end{array}$$

which gave rise to the decomposition over Q,

$$H^*(N) \simeq H^*(N_0) \otimes H^*(J),$$

of proposition 9.7. Note in particular that η^* is an isomorphism over Q. It follows that if we introduce the new rational classes

$$\begin{aligned} \boldsymbol{b}_2^j &= \pi^* (\eta^*)^{-1} i^* b_2^j, \quad j = 1, \ldots, 2g, \\ \boldsymbol{a}_k &= \pi^* (\eta^*)^{-1} i^* a_k, \quad k = 1, 2, \end{aligned}$$

then these will generate $\pi^* H^*(N/J_0)$ so that as a $\Lambda^*(b_1^j)$-module, $H^*(N)$ is *freely* generated by their span.

In terms of these variables, and the corresponding $\boldsymbol{\Lambda}$, our polynomial $Q_g(a, \Lambda)$ is now transformed into a polynomial $R_g(\boldsymbol{a}; \boldsymbol{\Lambda})$ and if this expression is expanded in terms of the basis

$$b_1^I = b_1^{i_1} \ldots b_1^{i_k}, \quad i_1 < i_2 < \ldots < i_k$$

for $\Lambda^*(b_1^j)$, then $R_g(\boldsymbol{a}, \boldsymbol{\Lambda}) = 0$ implies that *each coefficient* in this expansion must vanish. Equivalently one can multiply R_g by b_1^I and integrate over the fibre of π, to obtain

$$\pi_* b_1^I R_g(\boldsymbol{a}; \boldsymbol{\Lambda}) = 0,$$

yielding a large number of relations in a_1, a_2 and b_2^i.

To carry out this process one first of all has to determine the old generators in terms of the new bold-faced ones, and this is done quite easily by observing that End V descends to N/J_0 so that the characteristic classes of End V certainly are in the image of π^*. Thus $a_1 + \Lambda_{11}$ is in the image of π^* and restricts to $i^* a_1$ on N_0 whence

$$\boldsymbol{a}_1 = a_1 + \Lambda_1 \quad \text{or} \quad a_1 = \boldsymbol{a}_1 - \Lambda_{11}.$$

Similarly $p_1(T)$ is in this image. Hence

$$\pi^* (\eta^*)^{-1} i^* (a_1^2 - 4a_2) = a_1^2 - 4a_2.$$

On the other hand as i^*, η^* and π^* are ring-homomorphisms the expression on the left is also equal to $\boldsymbol{a}_1^2 - 4\boldsymbol{a}_2$. Eliminating one obtains:

$$a_2 = \boldsymbol{a}_2 - \tfrac{1}{2}\Lambda_{11}\boldsymbol{a}_1 + \tfrac{1}{4}\Lambda_{11}^2.$$

Finally to determine $\boldsymbol{b}_2^i$ consider $c_2(\text{End } V) = c_1^2 - 4c_2$, whose Künneth components must again all be in the image of π^*. Applied to the first mixed component this yields the result that

$$(2a, b_1^i - 4b_2^i)$$

is in the image of π^*, whence

$$b_2^i = \boldsymbol{b}_2^i + \tfrac{1}{2}(\boldsymbol{a}_1 - \Lambda_{11})\, b_1^i.$$

Let us now expand our first relation $R_g(\boldsymbol{a}, \boldsymbol{\Lambda})$ in terms of Λ_{11} and Λ_{12}, when $g = 3$. Also, in part to avoid subscripts and in part to come closer to Ramanan's notation, let us set

$$h = \boldsymbol{a}_1, \quad \nu = \boldsymbol{a}_2, \quad \theta = \Lambda_{22}$$

and

$$\omega = \Lambda_{11}, \quad \Delta = \Lambda_{12}.$$

Then for dimensional reasons the possible monomials in ω and Δ occurring in R_3 are given by the following table:

6	–	–	–	ω^3	–	–	$\Delta^2\omega^2$
5	–	–	–	–	$\Delta\omega^2$	–	–
4	–	–	ω^2	–	–	$\Delta^2\omega$	–
3	–	–	–	$\Delta\omega$	–	-	Δ^3
2	–	ω	–	–	Δ^2	–	–
1	–	–	Δ	–	–	–	–
0	–	–	–	–	–	–	–
	0	2	4	6	8	10	12

Here the dimension of a term relative to the b_1^j is indicated on the left and the total dimension below. The total dimension of R_3 is 12 so that the coefficient of one of these monomials (which is a polynomial in h, v and θ) must make up for the deficiency in the dimension. Note further that every element in the top row must be a multiple of $b_1^1 \wedge \ldots \wedge b_1^6 = \frac{1}{6}\omega^6$. Hence $\Delta^2\omega^2$ is also such a multiple and, as is easily checked, is in fact given by

$$\Delta^2\omega^2 = \text{const.} \times \theta^2\omega^3.$$

In short then, the expression R_3 takes the form

$$R_3 = A_6\omega^3 + B_4\Delta\omega^2 + C_8\omega^2 + D_2\Delta^2\omega + E_0\Delta^3 + F_6\Delta\omega + \ldots.$$

The first two relations now follow immediately. We must have

$$A_6 = 0 \quad \text{and} \quad B_4 b_2^i = 0, \quad i = 1, \ldots, n.$$

At the next level more care has to be taken as the two terms interfere. One procedure is to write down the implications $\pi_*\omega R_g = 0$, and $(\pi_* u)R_g = 0$ with $u \in \Lambda^2(b_1^1, \ldots, b_1^n)$, subject to $u\omega^2 = 0$. The first of these leads to a relation of the form

$$C_8 + \text{const.} \times \theta_2 D_2 = 0$$

while the second one implies that

$$D_2 x = 0 \quad \text{for all} \quad x \in \Lambda^2(b_2^1, \ldots, b_2) \quad \text{of the form} \quad \pi_* u\Delta^2, \text{ with } u\omega^2 = 0.$$

In short then, in these and subsequent relations the decomposition of $\Lambda^*(b_1^i)$ into primitive classes relative to ω makes its appearance, and as is really not too surprising this decomposition corresponds under π_* to the corresponding decomposition under θ in $\Lambda^*(b_2^1, \ldots, b_2^n)$. Thus the last relation is equivalent to

$$D_2 x = 0 \quad \text{for} \quad x\theta^2 = 0.$$

Similarly the $E_0\Delta^3$ term produces the relation

$$E_0 x = 0 \quad \text{for} \quad x \in \Lambda^3(b_2^i) \quad \text{with} \quad x\theta = 0.$$

We have traced the nature of these five relations so carefully because they correspond precisely to the complete set of relations actually found by Ramanan in this case using a quite different method. Here are his relations:

(1) $3h^3 - 10hv - 4\theta = 0$,

(2) $(h^2 - 2v)\,V = 0 \quad (V = \text{span of } b_2^j)$,

(3) $(h^2 - 3\nu)\,\nu = 0$,

(4) $hx = 0$, for $x \in \Lambda^2 V$, with $x\theta^2 = 0$,

(5) $y = 0$, for $y \in \Lambda^3 V$, with $y\theta = 0$.

This concludes our remarks on the relations among the generators of theorem 9.11. Clearly the computations involved in carrying out the programme laid out here are quite astronomical and therefore appropriately best left to a computer. Note that in the present case of genus 3 the single equation $c_{2g}(f_! V) = 0$ implied all the necessary relations. On the other hand it is disappointing, and shows how deep Newstead's conjectures lie, that even with all the relations before one, the formulae $p_1(T)^g = 0$ and $c_i(T) = 0$ for $i \geqslant 2\bar{g}$ are by no means obvious.

Finally we revert to the geometry of the moduli space $N(n, k)$ and show that it inherits a natural Kähler structure. The essential observation is that the space $\mathscr{A}$ of unitary connections has a natural *symplectic* structure: if α, β are two ad (P)-valued 1-forms on M they have a skew product $\int_M \alpha \wedge \beta$ (we recall that this uses the inner product in the Lie algebra). This symplectic structure is preserved by the action of $\mathscr{G}$. Moreover the curvature

$$F: \mathscr{A} \to \Omega^2(M; \mathrm{ad}\,(P))$$

can be identified with the corresponding *moment map*. To see this we first note that $\Omega^2(M, \mathrm{ad}\,(P))$ is canonically dual to $\Omega^0(M, \mathrm{ad}\,(P))$, which is the Lie algebra of $\mathscr{G}$. Hence for any $\phi \in \Omega^0(M, \mathrm{ad}\,(P))$ we have a real-valued function F_ϕ on $\mathscr{A}$, defined by $F_\phi(A) = (F_A, \phi)$. To say that F is the moment map for the $\mathscr{G}$-action on $\mathscr{A}$ means that the Hamiltonian vector field on $\mathscr{A}$ defined by F_ϕ coincides with the vector field given by the Lie algebra action of ϕ. Equivalently we have to show that, for any $\psi \in \Omega^1(M, \mathrm{ad}\,(P))$,

$$(\mathrm{d}F_\phi, \psi) = \int (\mathrm{d}_A \phi) \wedge \psi. \tag{9.14}$$

But, as we saw in §3, the left-hand side is equal to $\int (\mathrm{d}_A \psi) \wedge \phi$. Since d_A is a derivation and $\int \mathrm{d}(\psi \wedge \phi) = 0$ we have the usual formula for integration by parts

$$\int (\mathrm{d}_A \psi) \wedge \phi = -\int \psi \wedge (\mathrm{d}_A \phi)$$

which verifies (9.14).

The constant central $U(1)$ subgroup of $\mathscr{G}$ acts trivially on $\mathscr{A}$ corresponding to the fact that the function $\int \mathrm{trace}\, F_A$ is constant and equal to $-2\pi i k$ (where k is the Chern class).

The moment map is $\mathscr{G}$-equivariant and so to every orbit $C \subset \Omega^2(M, \mathrm{ad}\,(P))$ the inverse image $F^{-1}(C) \subset \mathscr{A}$ is $\mathscr{G}$-invariant. The quotient $F^{-1}(C)/\mathscr{G}$ is sometimes called the Marsden–Weinstein quotient. Under appropriate non-degeneracy conditions, it is a manifold and it inherits a natural symplectic structure from that of $\mathscr{A}$. In particular taking C to be the orbit given by the Yang–Mills minimum (i.e. the constant conjugacy class with all eigenvalues $-2\pi i k/n$) we obtain the moduli space $N(n, k)$. Thus $N(n, k)$ inherits a natural symplectic structure.

The symplectic structure on $\mathscr{A}$ together with its natural metric defines the complex structure of $\mathscr{C}$. Similarly the induced symplectic structure and metric on N define its complex structure. Thus N is a Kähler manifold.

Note that the tangent space to N at E is $H^1(M, \mathrm{End}\, E)$ and it is easy to define the metric, complex structure and symplectic structure on this tangent space. What is not immediately clear is the global integrability condition of the complex and symplectic structures so defined on N. The complex structure becomes clear by expressing N as the quotient $\mathscr{C}_s/\mathrm{Aut}\,(E)$ while the symplectic structure is similarly transparent as the 'Marsden–Weinstein quotient'.

10. The stratification for general G

In this section we shall indicate briefly how to extend the results of the previous sections from $U(n)$ to general compact Lie groups G. We shall content ourselves now with the basic results about *rational* cohomology, since the presence of torsion in G makes it difficult to say much in general about *integral* cohomology.

On the algebro-geometric side the work of Narasimhan–Seshadri has been extended to general reductive groups by Ramanathan (1975). We shall, however, adopt a slightly different, although equivalent, approach to stability and the canonical filtration, reducing everything to the vector bundle case by a systematic use of the adjoint representation.

The general set-up is much the same as before and we shall use the same notation. Thus we start with a given C^∞ principal G-bundle P over M and we denote by $\mathscr{A}$ the space of connections and $\mathscr{G}$ the group of automorphisms. It is again true that a connection on P defines a holomorphic structure on P^c, the associated bundle with group G^c the complexification of G. Conversely a holomorphic G^c-bundle together with a reduction of structure group to G determines a canonical G-connection (Singer 1959) so that we may identify $\mathscr{A}$ with the space of holomorphic structures on P^c.

To proceed further we need to introduce the appropriate stratification of $\mathscr{C}$ by strata $\mathscr{C}_\mu$ analogous to the Harder–Narasimhan stratification for the case of $GL(n)$. We shall in fact define such a stratification by using the canonical filtration of the vector bundle ad (P^c) in an appropriate way. First of all, however, we need a few lemmas concerning vector bundles.

We have already noted in §8 that a semi-stable vector bundle of slope (or normalized Chern class) μ has a filtration with stable quotients of slope μ. The converse is also true in view of lemma 7.5. This enables us to extend results for stable bundles to semi-stable bundles by induction. In this way we shall prove

Lemma 10.1. *If E, F are semi-stable of slopes μ, ν then $E \otimes F$ is semi-stable of slope $\mu + \nu$.*

Proof. Consider the first case when E, F are both stable. According to the Narasimhan–Seshadri theorem 8.1 they then arise from unitary representations of the extended fundamental group Γ_R (as in §6) with slopes μ, ν. The tensor product $E \otimes F$ then arises from the tensor product of the two unitary representations. This tensor product is not necessarily irreducible but, being unitary, it is a direct sum of irreducible pieces. Moreover the slope of a representation is given by the character of the centre of Γ_R and this therefore takes the same value $\mu + \nu$ on all the irreducible pieces. Hence $E \otimes F$ is a direct sum of stable bundles of slope $\mu + \nu$ and so is semi-stable and of the same slope. Now we move on to the general case and use filtrations of E and F with stable quotients D_j, G_k respectively. The tensor product then inherits a filtration with quotients $D_j \otimes G_k$, which as we have just proved are stable and of slope $\mu + \nu$. Hence $E \otimes F$ is semi-stable and of slope $\mu + \nu$.

For our next lemmas, which concern general vector bundles E, it will be convenient to introduce some additional notation. Let

$$0 = E_0 \subset E_1 \subset E_2 \subset \ldots \subset E_r$$

be the canonical filtration of E with semi-stable quotients

$$D_j = E_j/E_{j-1}, \quad \text{slope}\,(D_j) = \mu_j, \quad \mu_1 > \mu_2 > \ldots > \mu_r.$$

We shall write

$$\inf E = \mu_r, \quad \sup E = \mu_1.$$

Thus E is semi-stable if and only if $\inf E = \sup E$.

The next lemma is then a straightforward consequence of lemma 7.5, with double induction on the steps of the canonical filtrations.

LEMMA 10.2. *$\inf E \geqslant \mu$ if and only if, for every F with $\sup F < \mu$, every homomorphism $F \to E$ is zero.*

Using (10.1) and (10.2) and again using double induction one obtains

LEMMA 10.3. $\inf(E \otimes F) = \inf E + \inf F$.

With these vector bundle lemmas out of the way we return to consider a holomorphic G^c-bundle ξ over M. Let $E = \mathrm{ad}(\xi)$ be the vector bundle associated with the adjoint representation, so that E is actually a holomorphic *bundle of Lie algebras*. Since the Lie algebra of G^c has a non-degenerate invariant quadratic form so does the bundle E. In particular E is self-dual so that its canonical filtration must be of the form

$$0 \subset E_{-r} \subset E_{-r+1} \subset \ldots \subset E_{-1} \subset E_0 \subset E_1 \subset \ldots \subset E_{r-1} \subset E,$$

where E_{-j} is the polar space (relative to the quadratic form) of E_{j-1}. We have indexed things in such a way that $D_0 = E_0/E_{-1}$ has slope zero. Since E_{-1} is the polar space of E_0 we have an induced non-degenerate quadratic form on D_0.

Consider now the Lie bracket

$$\phi : E_0 \otimes E_0 \to E/E_0.$$

Since $\inf(E_0 \otimes E_0) = 0$, by (10.3), and $\sup E/E_0 < 0$, lemma 10.2 implies that $\phi = 0$. Hence *E_0 is a Lie sub-algebra bundle of E*. For similar reasons

$$[E_{-j}, E_{-1}] \subset E_{-j-1} \quad \text{for} \quad j \geqslant 0$$

so that E_{-1} is a nilpotent ideal: it is the nilpotent radical and D_0 the reductive quotient of E_0. It now follows (see lemma below) that E_0 is a *parabolic* sub-algebra bundle, i.e. it contains (over every point of M) a maximal solvable (Borel) sub-algebra. Now a parabolic sub-algebra generates a parabolic subgroup and this is its own normalizer. Hence the sub-algebra bundle $E_0 \subset \mathrm{ad}(\xi)$ determines a reduction of the structure group of ξ to this parabolic subgroup Q. We denote this new principal bundle by ξ_Q and call it the canonical parabolic reduction of ξ.

For $G^c = GL(n, C)$ the parabolic subgroups are the stabilizers of partial flags and a parabolic reduction of the principal bundle is equivalent to giving a filtration of the associated vector bundle. We shall now show that the canonical parabolic reduction defined above does indeed coincide with the canonical filtration of Harder–Narasimhan. So let V be a holomorphic vector bundle and let

$$0 = V_0 \subset V_1 \subset V_2 \subset \ldots \subset V_r = V$$

be its canonical filtration so that the associated quotients

$$W_j = V_j/V_{j-1}$$

are semi-stable and have slopes μ_j strictly decreasing with j. The adjoint bundle E is now End V. The filtration of V induces a filtration

$$0 = E^{-r} \subset \ldots \subset E^{-1} \subset E^0 \subset E^1 \subset \ldots \subset E^{r-1},$$

where ϕ is in E^j provided $\phi(V_i) \subset V_{i+j}$ for all i. Thus E^0 consists precisely of endomorphisms preserving the filtration of V so that

$$E^0/E^{-1} \cong \bigoplus_i \operatorname{End} V_i.$$

More generally

$$E^j/E^{j-1} \cong \bigoplus_i \operatorname{Hom}(V_i, V_{i+j}).$$

In view of (10.1) it follows that E^0/E^{-1} is semi-stable of slope 0 while

$$\inf E^{-1} > 0 \quad \sup E/E^0 < 0.$$

Comparing the filtration E^j of E with the canonical filtration and using (10.2) it easily follows that E^0 coincides with E_0 and E^{-1} with E_{-1}. Thus the two parabolic reductions do in fact agree.

Remark. Note that the filtration E^j does not totally coincide with the canonical filtration E_j: it has to be further refined depending on the particular sequence of slopes before it does so.

Next we shall prove

PROPOSITION 10.4. *The canonical parabolic reduction is functorial with respect to group homomorphisms.*

Thus let $\phi: G^c \to H^c$ be a homomorphism, ξ a principal G^c-bundle, and $\eta = \phi(\xi)$ the associated H^c-bundle. Then we have a homomorphism of Lie-algebra bundles

$$\phi: \operatorname{ad}(\xi) \to \operatorname{ad}(\eta).$$

Since G^c is reductive the homomorphism of Lie algebras induced by ϕ has G^c-invariant complements to the kernel and image. Hence putting $E = \operatorname{ad}(\xi)$, $F = \operatorname{ad}(\eta)$ we can decompose $\phi: E \to F$ into *split* exact sequences

$$0 \to K \to E \to I \to 0, \quad 0 \to I \to F \to J \to 0.$$

Now for any direct sum $A \oplus B$ of vector bundles it is easy to see, using 10.3, that the parts of the canonical filtration with slope $\geqslant 0$ are additive: $(A \oplus B)_0 = A_0 \oplus B_0$. Applying this to our situation we see that

$$E_0 \cong K_0 \oplus I_0, \quad F_0 \cong I_0 \oplus J_0$$

so that ϕ sends E_0 into F_0. This proves that the canonical parabolic reduction of η is induced by that of ξ.

For a vector bundle we defined its *type* μ in terms of the Chern classes of the semi-stable quotients of its canonical filtration. We shall introduce the corresponding notion for a general group. Thus let ξ be a principal G^c-bundle, ξ_Q its canonical parabolic reduction. To every character χ of Q (i.e. a homomorphism $\chi: Q \to C^*$) we have a line-bundle $\chi(\xi_Q)$ over M and so an integer Chern class $c_1\chi(\xi_Q)$. In this way we obtain a homomorphism

$$\hat{Q} \to Z, \tag{10.5}$$

where $\hat{Q}$ is the abelian group of characters of Q. This will essentially be our *type*. To see more clearly what it involves let us pass to the reductive quotient S of Q, i.e. the quotient by its unipotent radical R (maximal connected normal unipotent subgroup). For $GL(n)$ we have

$$S = GL(n_1) \times \ldots \times GL(n_r)$$

where the n_j are the dimensions of the quotients in the canonical filtration. Clearly $\hat{Q} = \hat{S}$ so that for $GL(n)$ the homomorphism $\hat{Q} \to Z$ consists precisely of assigning the Chern classes $k_1, \ldots, k_r$ to appropriate semi-stable factors. The general case is similar in that $\hat{S}$ is a lattice of rank equal to

the dimension of the centre of S and the type will then be a vector in the dual lattice. Now for $GL(n)$ we found it convenient to replace the sequence of (n_j, k_j) by a single n-vector μ and we shall reinterpret our type in a similar way for the general case.

The group Q is the semi-direct product RS. In fact S can be identified with the complexification K^c of the maximal compact subgroup of K. If T_0 is the connected component of the centre then a character of S defines a (unitary) character of T_0 and the map $\hat{S} \to \hat{T}_0$ is injective with finite cokernel. Now we may assume $K \subset G$ and that the maximal torus T_1 of K is contained in the maximal torus T of G. Passing to characters gives surjective maps

$$\hat{T} \to \hat{T}_1 \to \hat{T}_0$$

while taking Hom (, Z) gives an inclusion of the corresponding dual lattices

$$L_0 \subset L_1 \subset L.$$

Each lattice here can be identified with the integral points in the Lie algebra of the corresponding torus (i.e. the kernel of $\exp 2\pi i$). The lattice Hom $(\hat{S}, Z)$ then contains L_0 as a sublattice of finite index. In particular we may view Hom $(\hat{S}, Z)$ as a subgroup of the Lie algebra of T. In this way the *type* of our G^c-bundle ξ becomes an element μ of the Lie algebra of T.

For $GL(n, C)$ our vector μ satisfied the inequalities

$$\mu_1 \geqslant \mu_2 \geqslant \ldots \geqslant \mu_n,$$

which describe a fundamental chamber for the action of the Weyl group W. The choice of this chamber derived from the parabolic subgroup determined by the complex structure. From the unitary point of view μ, or rather its W-orbit, corresponds naturally to a conjugacy class in the Lie algebra of $U(n)$ and this (up to $2\pi i$) is the curvature of the Yang–Mills connection associated with μ. In this way we saw that each stratum $\mathscr{C}_\mu$ contains a unique component of the Yang–Mills connection. For general groups the situation is now exactly the same: μ determines a conjugacy class in the Lie algebra of G and hence a component of the Yang–Mills connection. The group K is the local holonomy group and this determines the integrality conditions on μ.

The stratum $\mathscr{C}_\mu$, consisting of all ξ of given type μ, has a conormal bundle whose fibre at ξ is $H^1(M, \mathrm{ad}\,(\xi)/E_0)$ where E_0 is as above the canonical parabolic subalgebra of ad (ξ).

If $\rho : G \to H$ is a homomorphism and ξ is a G^c-bundle of type μ then proposition 10.4 implies that $\rho(\xi)$ is an H^c-bundle of type $\rho(\mu)$. Here μ and $\rho(\mu)$ are best considered as conjugacy classes of the appropriate Lie algebras. Even if ρ is an embedding $\rho(\mu)$ does not necessarily determine μ, but the Peter–Weyl theorem implies that if $\rho(\mu) = \rho(\nu)$ for all unitary representations ρ of G then $\mu = \nu$. Thus we have

PROPOSITION 10.6. *A G^c-bundle ξ is of type μ if and only if $\rho(\xi)$ is of type $\rho(\mu)$ for all unitary representations ρ of G.*

This proposition, together with (7.8), which describes the closure properties of the strata for $U(n)$, enables us to derive similar results for every G. Thus let $\mathscr{C}_\lambda$ be a stratum for G lying in the closure of $\mathscr{C}_\mu$. Then (10.6) implies that $\rho(\mathscr{C}_\lambda) = \mathscr{C}_{\rho(\lambda)}$ lies in the closure of $\mathscr{C}_{\rho(\mu)}$ for all ρ. Hence by (7.8) we have $\rho(\lambda) \geqslant \rho(\mu)$ for the partial ordering on conjugacy classes of $\mathfrak{u}(n)$. But by (12.18) this is equivalent to $\lambda \geqslant \mu$ where this is the partial ordering on conjugacy classes of $\mathfrak{g}$ defined in §12 (and corresponds to inclusion of convex hulls). Thus (7.8) holds for all G.

As with the unitary case the conormal to the stratum $\mathscr{C}_\mu$ at ξ can be identified with

$$H^1(M, \mathrm{ad}\,(\xi)/E_0)$$

where E_0 gives the canonical parabolic reduction of ξ. Since $\inf(\mathrm{ad}\,(\xi)/E_0) < 0$ it follows that $H^0(M, \mathrm{ad}\,(\xi)/E_0) = 0$ and by Riemann–Roch we can compute the complex codimension d_μ of $\mathscr{C}_\mu$. One finds the following generalization of (7.15):

$$d_\mu = \sum_{\alpha(\mu)>0} \{\alpha(\mu) + g - 1\}, \tag{10.7}$$

where α runs over the positive roots of G, and μ is the representative in the positive Weyl chamber (so that $\alpha(\mu) \geqslant 0$). In particular we see that

$$d_\mu = 0 \Leftrightarrow \mu \text{ is in the centre.} \tag{10.8}$$

In this case μ is uniquely determined by the topology of ξ (there will be one Chern class for each circle factor in the centre). We define this stratum to be the *semi-stable* stratum: it is necessarily open and non-empty. Moreover μ is central if and only if $\mathrm{ad}\,(\mu) = 0$ so that we have in this case the following strengthening of (10.6).

PROPOSITION 10.9. *A G^c-bundle ξ is semi-stable if and only if* $\mathrm{ad}\,(\xi)$ *is a semi-stable vector bundle.*

As with the unitary case a general stratum $\mathscr{C}_\mu$ is equivariantly equivalent to a semi-stable stratum for the group K. Moreover the connected centre T_0 of K acts on $H^1(M, \mathrm{ad}\,(\xi)/E_0)$ with no trivial character: in fact the connected centralizer of T_0 in G^c is just K^c.

We now have all the ingredients to deduce as in (7.14)

THEOREM 10.10. *For any G the stratification of $\mathscr{C}$ by the $\mathscr{C}_\mu$ is equivariantly perfect over the rationals so that for Poincaré series we have*

$$P_t(\mathscr{C}) = \sum_\mu t^{2d_\mu} P_t(\mathscr{C}_\mu),$$

where d_μ is given by (10.8).

In principle this enables us to calculate the equivariant cohomology of the semi-stable stratum by induction on the dimension of G. The point is that, for any other stratum $\mathscr{C}_\mu$, the equivariant cohomology is equal to that of a semi-stable stratum for a proper subgroup K of G, namely the maximal compact subgroup of the parabolic subgroup of G^c determined by μ. When $G = U(n)$ the group K is always of the form $U(n_1) \times \ldots \times U(n_r)$ and so our induction in the unitary case did not use other groups. However, for general G the groups K that occur are centralizers of tori and can be of many types.

To relate this to the Morse theory for the Yang–Mills functional L we note first that, after suitable normalization L is functorial for homomorphisms of Lie groups. Hence (10.6) together with (8.13) enables us to deduce, for any G,

$$A \text{ of type } \mu \Rightarrow L(A) \geqslant L(\mu) = |\mu|^2. \tag{10.11}$$

On the other hand our description of Yang–Mills connections shows that every stratum $\mathscr{A}_\mu$ does in fact contain a critical set $\mathscr{N}_\mu$, so that, on $\mathscr{A}_\mu$, $L(A)$ achieves its minimum. To go further and establish the generalization of (8.20) we need the following lemma, in which Γ_R denotes the central extension of $\pi_1(M)$ by R defined in § 6.

LEMMA 10.12. *A holomorphic G^c-bundle ξ arises from a homomorphism $\rho : \Gamma_R \to G$ if and only if* $\mathrm{ad}\,(\xi)$ *arises from a unitary representation of Γ_R.*

Proof. In one direction this is trivial. For the converse let $\mathrm{ad}\,(\xi)$ arise from a unitary representation of Γ_R, and consider the Lie bracket homomorphism of vector bundles:

$$\mathrm{ad}\,(\xi) \otimes \mathrm{ad}\,(\xi) \to \mathrm{ad}\,(\xi).$$

Both sides are vector bundles arising from unitary representations of Γ_R and, as proved by Narasimhan & Seshadri (1965), this implies the homomorphism is *covariant constant*. This means that ad (ξ) as an ad (G)-bundle comes from a homomorphism $\pi_1(M) \to \text{ad}\,(G)$. It is then easy, using the theory of line-bundles, to lift this to a homomorphism $\rho : \Gamma_R \to G$, which will, on extension to G^c, define ξ.

Arguing along the lines of §8 and using (10.12) one can then prove

PROPOSITION 10.13. *If a holomorphic G^c-bundle ξ is of type μ then*

$$\inf_A L(A) = L(\mu),$$

where A runs over all compact connections on ξ.

Quite likely (10.11) and (10.13) hold for all convex invariant functions on the Lie algebra of G: they certainly do for any ϕ induced from a representation.

To sum up therefore we see that the picture for general G is in practically all respects similar to the unitary case, with the notable difference that we have had to switch from integral to rational cohomology.

11. COMPARISON WITH HARDER–NARASIMHAN APPROACH

As mentioned in the Introduction the Poincaré polynomials of the moduli spaces of stable bundles have been computed by number-theory methods in Harder & Narasimhan (1975) and Desale & Ramanan (1975). In this section we shall compare those methods with ours.

We begin with an example by spelling out in detail our results for the simplest interesting case, namely for $n = 2$ and $k = 1$. Our basic theorem 7.14 becomes

$$\mathcal{G}P_t(\mathcal{C}_s) + \sum_{r=0}^{\infty} t^{2(2r+g)}\, \mathcal{G}P_t(\mathcal{C}_r) = \mathcal{G}P_t(\mathcal{C}), \tag{11.1}$$

where $\mathcal{G}P_t$ stands for $\mathcal{G}$-equivariant Poincaré series and $\mathcal{C}_r$ is the stratum corresponding to unstable bundles of type $(r+1, -r)$ (i.e. of the form (11.10)). As shown in §9 (see (9.3) and (9.7)), for the stable bundles we have

$$\mathcal{G}P_t(\mathcal{C}_s) = \frac{P_t(N(2,1))}{1-t^2} = \frac{(1+t)^{2g}\, P_t(N_0(2,1))}{1-t^2}.$$

For the unstable stratum $\mathcal{C}_r$ we apply (7.12) to see that

$$\mathcal{G}P_t(\mathcal{C}_r) = \left\{\frac{(1+t)^{2g}}{1-t^2}\right\}^2.$$

Finally for the whole space we apply theorem 2.15, which, for $n = 2$, gives

$$\mathcal{G}P_t(\mathcal{C}) = P_t(B\mathcal{G}) = \frac{\{(1+t)\,(1+t^3)\}^{2g}}{(1-t^2)^2\,(1-t^4)}.$$

Substituting these into (11.1) and cancelling a common factor $(1+t)^2/(1-t^2)$, we get

$$P_t(N_0(2,1)) + \frac{(1+t)^{2g}}{1-t^2} \sum_{r=0}^{\infty} t^{2(2r+g)} = \frac{(1+t^3)^{2g}}{(1-t^2)\,(1-t^4)}. \tag{11.2}$$

Summing the geometric series we see that this gives the formula

$$P_t(N_0(2,1)) = \frac{(1+t^3)^{2g} - t^{2g}(1+t)^{2g}}{(1-t^2)\,(1-t^4)}. \tag{11.3}$$

It is true, though not entirely transparent, that this rational function is in fact a *polynomial* in t with non-negative integer coefficients (giving the Betti numbers). Moreover

$$\dim N_0(2,1) = 6g-6$$

and so Poincaré duality requires that

$$P_t(N_0) = t^{6g-6} P_{1/t}(N_0). \tag{11.4}$$

We turn now to summarize the methods of Harder & Narasimhan.

We begin by taking a curve M of genus g defined over a finite field F_q. The ζ-function of M then has the form

$$\zeta_M(s) = \left\{\prod_{i=1}^{2g} (1-\omega_i q^{-s})\right\} \Big/ (1-q^{-s})\,(1-q.q^{-s}), \tag{11.5}$$

where the ω_i are algebraic integers (depending on M) with $|\omega_i| = q^{\frac{1}{2}}$. We now consider vector bundles E over M that are defined over F_q and have given rank n and *fixed determinant* of degree k. This means that we fix the isomorphism class of the line-bundle $L = \det E$. Then the Siegel formula is the following:

$$\sum_E \frac{1}{|\mathrm{Aut}\,(E)|} = \frac{1}{q-1}\, q^{(n^2-1)(g-1)} \zeta_M(2) \ldots \zeta_M(n), \tag{11.6}$$

where the sum is over all isomorphism classes (with $\det E$ fixed), and $|\mathrm{Aut}\,(E)|$ is the number of automorphisms of E.

Thus (11.6) *counts* the number of isomorphism classes of E, each being weighted inversely by its number of automorphisms. In particular stable bundles that admit only scalar automorphisms occur with weight $(q-1)^{-1}$ and so contribute

$$|N_L(n,k)|/(q-1) \tag{11.7}$$

to the sum in (11.6), where the numerator denotes the number of classes of stable bundles of rank n and determinant L (of degree k) defined over F_q. Now when $(n,k) = 1$ the moduli space $N_L(n,k)$ of stable bundles of rank n and determinant L is a projective non-singular variety and we can suppose that it is also defined over F_q (if not replace F_q by a finite extension). Then, as the notation suggests, the numerator in (11.7) is just the number of points of the moduli space that are defined over F_q. By the Weil conjectures, as established by Grothendieck and Deligne, the numbers of rational points over F_{q^n}, for all n, determine the Betti numbers of the 'corresponding variety' over C. In our case this means the moduli space for stable bundles of fixed determinant over a Riemann surface of genus g: the variety denoted in §9 by $N_0(n,k)$.

In this way (11.6) will lead to a formula for the Poincaré polynomial $P_t(N_0(n,k))$ provided we can deal with all the terms arising from unstable bundles. This can be done inductively, but for this purpose we need to consider also the non-coprime case and so we introduce

$$\beta(n,k) = \sum \frac{1}{|\mathrm{Aut}\,(E)|} \quad \text{summed over } \textit{semi-stable } E. \tag{11.8}$$

Using the canonical filtration of Harder–Narasimhan explained in §7 we can collect together in (11.6) all terms of the same *type*. These can then be summed explicitly in terms of $\beta(n_i,k_i)$ and the number of rational points J_q on the Jacobian of M. Now in terms of the ω_i occurring in (11.5) this is given by

$$J_q = \prod_{i=1}^{2g} (1-\omega_i). \tag{11.9}$$

Finally therefore (11.6) gives an explicit inductive formula for $\beta(n,k)$ in terms of $\beta(n_i,k_i)$ with $n_i < n$. The formula is given rationally in q and the ω_i and is independent of the line-bundle L (Desale & Ramanan 1975; proposition 1.7). To get the Poincaré polynomial $P_t(N_0(n,k))$ one now makes the substitution

$$\omega_i \to -t,\ q \to t^2.$$

For the purposes of comparison with our method let us now examine in detail the case $n = 2$ and $k = 1$. In (11.6) we then have stable bundles, which contribute (11.7), and unstable bundles, which have a canonical filtration

$$0 \to L_1 \to E \to L_2 \to 0 \tag{11.10}$$

where $\deg L_1 = r+1$, $\deg L_2 = -r$ for $r = 0, 1, \ldots$ and $L_1 \otimes L_2 \cong L$. To compute $|\mathrm{Aut}\,(E)|$ consider separately the trivial and non-trivial extensions. For $E = L_1 \oplus L_2$ the automorphisms consist of $F_q^* \times F_q^*$ together with the unipotents of the form $1+\phi$ with

$$\phi \in \mathrm{Hom}\,(L_2, L_1) = H^0(M, L_2^* \otimes L_1).$$

Hence

$$|\mathrm{Aut}\,(E)| = (q-1)^2 h_0,$$

where $h_0 = |H^0(M, L_2^* \otimes L_1)|$. On the other hand for non-trivial extensions we have only one copy of F_q^* and so

$$|\mathrm{Aut}\,(E)| = (q-1)\, h_0.$$

The non-trivial extensions correspond to non-zero elements of $H^1(M, L_2^* \otimes L_1)$ and proportional vectors give isomorphic bundles. Hence the number of isomorphism classes of bundles E for which (11.10) is non-trivial is

$$(h_1 - 1)/(q-1),$$

where

$$h_1 = |H^1(M, L_2^* \otimes L_1)|.$$

Hence the contribution to the sum in (11.6) arising from a given L_1 (L_2 being then determined as $L \otimes L_1^*$) is

$$\frac{1}{(q-1)^2 h_0} + \frac{h_1 - 1}{(q-1)^2 h_0} = \frac{h_1}{(q-1)^2 h_0}.$$

Now by Riemann–Roch we have

$$\dim H^0(M, L_2^* \otimes L_1) - \dim H^1(M, L_2^* \otimes L_1) = 2r + 2 - g,$$

and so

$$h_0/h_1 = q^{2r+2-g}.$$

Thus (11.6) becomes

$$\frac{|N_L(2,1)|}{q-1} + \frac{J_q}{(q-1)^2} \sum_{r=0}^{\infty} \frac{1}{q^{2r+2-g}} = \frac{1}{q-1} q^{3g-3} \zeta_M(2). \tag{11.11}$$

Substituting for $\zeta_M(2)$ from (11.5) and for J_q from (11.9) we get

$$|N_L(2,1)| + \frac{q^{g-2} \prod_{i=1}^{2g} (1-\omega_i)}{(q-1)(1-q^{-2})} = \frac{q^{3g-3} \prod_{i=1}^{2g} (1-\omega_i q^{-2})}{(1-q^{-2})(1-q^{-1})}. \tag{11.12}$$

If we now make the substitution $\omega_i \to -t, q \to t^2$ then the expression for $|N_L(2,1)|$ given by (11.12) converts into the formula for the Poincaré polynomial given in (11.3).

Comparing these two derivations of the formula for $P_t(N_0(2,1))$, we see that they are formally

very similar with (11.11) playing the role of (11.2). We note, however, that (11.11) involves a convergent power series in q^{-1} while (11.2) involves a formal power series in t. This makes it surprising that we should have made the substitution $q \to t^2$. In fact it is better to make the substitution

$$q \to t^{-2}, \quad \omega_i \to -t^{-1}. \tag{11.13}$$

In view of the Poincaré duality formula (11.4) we must now get

$$|N_L(2,1)| \to t^{6-6g} P_t(N_0(2,1)). \tag{11.14}$$

Making the substitution (11.13) in (11.11) we see that, after removing the factor $(q-1)^{-1}$ and multiplying by q^{6g-6}, we get precisely (11.2), with a correspondence term by term.

If we were to compute for $n = 2$ and $k = 0$, we could still compare (11.6) with (7.14) but we would not be dealing with the moduli space. Thus the leading term in (11.6) is the quantity $\beta(2,0)$ defined in (11.8), while in (7.14) it would be the equivariant Poincaré series of the semi-stable stratum. Instead of (11.11) and (11.1) we then get

$$\beta(2,0) + \frac{J_q}{(q-1)^2} \sum_{r=1}^{\infty} \frac{1}{q^{2r+1-g}} = \frac{q^{3g-3}}{q-1} \zeta_M(2), \tag{11.15}$$

$$\mathcal{G}P_t(\mathcal{C}_{ss}) + \sum_{r=1}^{\infty} t^{2(2r+g-1)} \mathcal{G}P_t(\mathcal{C}_r) = \mathcal{G}P_t(\mathcal{C}), \tag{11.16}$$

where the stratum $\mathcal{C}_r$ corresponds now to extensions (11.10) with $\deg L_1 = r$, $\deg L_2 = -r$. Comparing these two formulae we see that the substitution (11.13) now leads to

$$\beta(2,0) \to \frac{t^{6-6g}}{(1+t)^{2g}} \mathcal{G}P_t(\mathcal{C}_{ss}), \tag{11.17}$$

the denominator $(1+t)^{2g}$ arising only because on the right we did not fix the determinant.

We see therefore that, by making $\beta(n,k)$ in general correspond to the equivariant Poincaré series of the semi-stable stratum, (11.6) and (7.14) lead to identical inductive procedures. It remains now to explain the origin of (11.6) and its relation to (7.14).

Just as the Jacobian arises classically as the group of divisor classes so moduli spaces of vector bundles can be viewed in terms of 'matrix divisor classes' as originally described by Weil (1938). In modern terminology this is best formulated in the language of adèles. Thus let K be the function field of M over F_q and for any affine algebraic group G let G_A be the adèle group of K, i.e. the restricted product of G_{K_p} where p runs over all valuations of K, and K_p is the corresponding local field. Then G_A is a locally compact topological group and G_K is a discrete subgroup. For $G = GL_n$, the isomorphism classes of vector bundles of rank n over M (defined over F_q) are in bijective correspondence with the double coset space

$$\mathfrak{K} \backslash G_A / G_K,$$

where $\mathfrak{K}$ is a maximal compact subgroup of G_A. To understand this correspondence one should think of $\mathfrak{K}\backslash G_A$ as a (multiplicative) matrix divisor and dividing by G_K as rational equivalence. In terms of bundles it corresponds to describing a bundle by a basis of meromorphic sections. If we take $G = SL_n$ there are different maximal compact subgroups $\mathfrak{K}$ for different choices of the line bundle $L = \det E$, and the corresponding double cosets are in bijective correspondence with classes of *L-oriented* bundles, i.e. bundles E together with a chosen isomorphism $\det E \cong L$. Now on G_A one introduces a special choice of Haar measure, the Tamagawa measure τ. The total

measure of G_A/G_K is finite and is called the Tamagawa number. For SL_n it turns out to be 1. Decomposing G_A/G_K into $\mathfrak{K}$-orbits then leads to the formula

$$\sum_\alpha \tau(\mathfrak{K}/\mathfrak{K}_\alpha) = \tau(G_A/G_K) = 1,$$

where α runs over the orbits and $\mathfrak{K}_\alpha$ is the (finite) isotropy group of the orbit. Dividing by $\tau(\mathfrak{K})$ then gives

$$\sum_\alpha \frac{1}{|\mathfrak{K}_\alpha|} = \frac{1}{\tau(\mathfrak{K})}. \tag{11.18}$$

If we denote by $|E|$ the number of inequivalent orientations on E, then for each orientation α on E one has

$$\frac{|\mathrm{Aut}\,(E)|}{|K_\alpha|} = \frac{|F_q^*|}{|E|} = \frac{q-1}{|E|}.$$

Hence if we rewrite (11.18) as a sum over isomorphism classes of vector bundles E with fixed determinant, by ignoring the orientation, we get

$$(q-1)\sum_E \frac{1}{|\mathrm{Aut}\,(E)|} = \frac{1}{\tau(\mathfrak{K})}.$$

This is the same as (11.6) in view of the formula

$$\tau(\mathfrak{K})^{-1} = q^{(n^2-1)(g-1)}\zeta_M(2)\ldots\zeta_M(n). \tag{11.19}$$

The factor $q-1 = |F_q^*|$ has arisen because of the passage from GL_n to SL_n.

In comparing the derivation of (11.6) and (7.14) we see that in both cases we start from an *infinite-dimensional space* that describes all bundles, but in a redundant fashion. In one case this space is G_A/G_K while in the other it is the space $\mathscr{C}$. As already noted the first description of algebraic bundles relies on the fact that every bundle is trivial over K, i.e. that it has a basis of rational or meromorphic sections. In the Riemann surface case we used instead the fact that all holomorphic bundles with the same degree (or Chern class) are *differentiably* equivalent.

In both cases we now stratify this infinite-dimensional space according to the *type* of the bundle, so that we have a unique open stratum given by semi-stable bundles. Moreover we have a group acting, preserving the strata, so that the equivalence classes represent isomorphism classes of bundles. In one case the group is $\mathfrak{K}$, the maximal compact subgroup of G_A, while in the other it is the group of C^∞ complex automorphisms. These equivariant stratifications can now be used to compute appropriate invariants. In the number-theory situation we compute Tamagawa measures to get (11.6) while in the geometric situation we compute equivariant cohomology to get (7.14). The parallel between these two procedures should be viewed as similar to that involved in the elementary computation with $P_n(C)$ in the Introduction. There are two notable differences here. In the first place the spaces concerned are infinite-dimensional and in the second place we work with equivariant notions relative to the appropriate group.

In these parallel treatments we see that in both cases the infinite-dimensional space itself is, in the appropriate sense, trivial. Thus the space $\mathscr{C}$ is *contractible* so that its ordinary Poincaré series is identically 1, while the Tamagawa number of G_A/G_K is also equal to 1 (notably it is independent of q). The next step is to 'divide' in the appropriate sense by the group action and to compute the result globally and locally and equate. On the global level we see therefore that the equivariant Poincaré series of $\mathscr{C}$, which is the same as the ordinary Poincaré series of $B\mathscr{G}$ and was computed in (2.15), corresponds to the measure $\tau(\mathfrak{K})^{-1}$ given by (11.19). Using the explicit

formula (11.5) for $\zeta_M(s)$ and applying the substitution $\omega_i \to -t^{-1}$, $q \to t^{-2}$ we see that except for trivial factors corresponding to the difference between SL_n and GL_n and a power of t (related to the dimension of the moduli space)

$$\tau(\mathfrak{R})^{-1} \to P_t(B\mathcal{G}). \tag{11.20}$$

Thus the 'global' terms in (11.6) and (7.14) correspond. On the other hand each stratum or type μ produces a 'local' term in both cases and when due account is taken of the isotropy groups the resulting formulae in the two cases again correspond. Thus the 'weighted counting' process corresponds to the use of equivariant cohomology.

When we compare these two basic methods of computing Betti numbers, i.e. the number-theory method and the Morse-theory method, we see that in each case we need to be fortunate to get an explicit answer. Thus when counting up points with a stratification the answer is clearly additive but in general we may not know how to compute the number of points in each stratum. In the Morse theory method each stratum retracts onto its critical set but we have no guarantee that the exact sequences split, i.e. that we have a perfect Morse stratification. In our present case the reason why we can count points effectively is that each stratum is made up of affine spaces corresponding to extensions as illustrated above. On the topological side the perfect nature of the stratification arises from the isotropy group behaviour. This is presumably linked in some way with the affine space decomposition of the strata.

Another reason that sometimes simplifies the process of counting points is if all homology is represented by algebraic cycles. In that case Frobenius acts on H^{2r} by q^r and so there are no mysterious eigenvalues. In our case this is nearly true in the sense that all rational cohomology of the moduli space N or N_0 is generated, as shown in §9, by the Künneth components of the Chern classes of the universal bundle on $M \times N$. Thus the only eigenvalues other than powers of q arise from $H^1(M)$ and these are the ω_i that appeared above. This explains why the simple substitution $\omega_i \to -t^{-1}$, $q \to t^{-2}$ is all that is required to convert the number-theory formulae into Poicaré series formulae.

Now that we have described the detailed correspondence between our method and that of Harder–Narasimhan many questions arise. In the first place why is the Tamagawa number of SL_n equal to 1? This is not very well understood but analogy with our method suggests that it might have some cohomological significance. Why moreover do we have the remarkable correspondence (11.20) and the analogy exhibited in (11.6) and (2.9), between the separate factors of both sides, namely $\zeta_M(k)$ and $P_t(\mathrm{Map}\,(M, K(Z, 2k))$? This and other aspects of the comparison suggest that the basic relation between numbers of points and Betti numbers for algebraic varieties may have some extension to infinite dimensions in which counting of points is replaced by a suitable measure.

Speculating in another direction we recall that the Yang–Mills equations arise in physics and that to quantize them involves, at least heuristically, some process of integration over function spaces. Comparison with the number theory suggests that there might be a natural measure, depending perhaps on some real parameter t, so that what we have been computing as Poincaré series actually turn out to be measures.

12. Convexity and Lie groups

This section is essentially an appendix concerned with the partial ordering that we have encountered in our stratification of the space $\mathscr{C}$. We shall take this opportunity of giving a brief but essentially self-contained account, which emphasizes the role of convexity in Lie groups. The results are not essentially new, and can mainly be found in Horn (1954) for the unitary case and in Kostant (1973) for the general groups, but our presentation brings out those aspects that are of particular relevance to the theory of bundles and connections. In particular we stress the role of convex invariant functions on the Lie algebra. For an extensive account of some aspects see also Marshall & Olkin (1979).

For simplicity we shall begin with the partial ordering (7.7) for sequences $(\lambda_1, \ldots, \lambda_n)$ of real numbers. Thus one defines $\mu \leqslant \lambda$ if, after arranging each sequence in decreasing order, we have

$$\text{(12.1)} \qquad \begin{cases} \sum_{j=1}^{i} \mu_j \leqslant \sum_{j=1}^{i} \lambda_j \quad \text{for} \quad i = 1, \ldots, n-1, \\ \sum_{j=1}^{n} \mu_j = \sum_{j=1}^{n} \lambda_j. \end{cases}$$

This partial ordering occurs in Horn (1954) where it is shown to be equivalent to either of the following properties:

$$\text{(12.2)} \qquad \sum_j f(\mu_j) \leqslant \sum_j f(\lambda_j) \quad \textit{for every convex function } f: R \to R;$$

$$\text{(12.3)} \qquad \mu = P\lambda \textit{ where } \lambda, \mu \in R^n \textit{ and } P \textit{ is a doubly stochastic matrix.}$$

We recall that a real square matrix $P = (p_{ij})$ is *stochastic* if $p_{ij} \geqslant 0$ and $\sum_j p_{ij} = 1$ for all i. If in addition the transposed matrix is also stochastic then P is called *doubly stochastic*. A theorem of G. D. Birkhoff identifies doubly stochastic matrices in terms of permutation matrices, namely

The doubly stochastic $n \times n$ matrices are the convex hull of the permutation matrices.

In view of this (12.3) can be replaced by

$$\text{(12.4)} \qquad \hat{\Sigma_n}\mu \subset \hat{\Sigma_n}\lambda$$

where $\Sigma_n x$ denotes the orbit of any $x \in R^n$ under the permutation group Σ_n, and $\hat{C}$ denotes the convex hull of the set $C \subset R^n$.

Geometric notions of convexity can be dualized into statements about convex functions by virtue of the fact that, for $C \subset R^n$,

$$x \in \hat{C} \Leftrightarrow \phi(x) \leqslant \sup_C \phi \quad \text{for all convex} \quad \phi: R^n \to R.$$

Thus taking ϕ to be a convex *symmetric* function on R^n (i.e. invariant under Σ_n) we can see that (12.4) implies

$$\text{(12.5)} \qquad \phi(\mu) \leqslant \phi(\lambda) \textit{ for all convex symmetric functions on } R^n.$$

Since (12.2) is the special case of (12.5) for functions $\phi(x_1, \ldots, x_n)$ of the form $\sum_{i=1}^n f(x_i)$, it follows that (12.5) implies (12.2) and so is equivalent to all the other properties.

Schur showed that if μ_j $(j = 1, \ldots, n)$ are the diagonal elements of a hermitian matrix whose eigenvalues are λ_j, then $\mu \leqslant \lambda$ in the sense of (12.1). Horn (1954) proved the converse so that another equivalent of (12.1) is

$$\text{(12.6)} \qquad \textit{the } \lambda_j \textit{ are the eigenvalues of a hermitian matrix with diagonal elements } \mu_j.$$

A hermitian matrix A is determined, up to conjugacy by $U(n)$, by the unordered set of its eigenvalues, or equivalently by the orbit $\Sigma_n \lambda \in R^n$. If $\Sigma_n \mu$ corresponds to the conjugacy class $C(B)$ of a hermitian matrix B then (12.4) clearly implies that $C(B)$ lies in the convex hull of $C(A)$. Conversely if a diagonal matrix B, with eigenvalues μ_j, lies in the convex hull of $C(A)$ it must lie in the convex hull of the diagonal parts of the matrices in $C(A)$. But by Schur's result this means that $\mu \in R^n$ is in the convex hull of $\Sigma_n \lambda$, where the λ_j are the eigenvalues of A. Hence (12.1) is also equivalent to

$$\widehat{C(\mu)} \subset \widehat{C(\lambda)}, \tag{12.7}$$

where $C(\lambda)$ denotes the conjugacy class of hermitian matrices with the given eigenvalues λ_j.

As before (12.7) implies

(12.8) $\psi(B) \leqslant \psi(A)$ *for all convex invariant functions* ψ *on the space of hermitian matrices*,

where $B \in C(\mu)$ and $A \in C(\lambda)$. Clearly such a convex invariant ψ defines a convex symmetric function ϕ on R^n by putting $\psi(A) = \phi(\lambda)$. Thus (12.8) is also directly implied by (12.5). The converse is not quite so clear because it is by no means obvious that convexity of ϕ on R^n implies convexity of ψ on the space of hermitian matrices. We shall, however, prove that this is in fact true, so that (12.8) is equivalent to all the earlier properties. This proof is just as easily given in the more general context of a general compact Lie group, so we move on now to consider how one generalizes all the preceding ideas.

For a general compact Lie group G, the role of the hermitian (or rather skew-hermitian) matrices is played now by the Lie algebra $\mathfrak{g}$ of G. The diagonal matrices are replaced by the Lie algebra $\mathfrak{t}$ of a maximal torus T of G and Σ_n becomes the Weyl group W. Writing a set of λ_j in decreasing order corresponds to picking a (closed) positive Weyl chamber C in $\mathfrak{t}$: this is a fundamental domain for the action of W. If we fix once and for all a bi-invariant inner product on $\mathfrak{g}$ we get a W-invariant inner product on $\mathfrak{t}$ and we define C^* to be the dual cone of C, namely

$$x \in C^* \Leftrightarrow \quad \langle x, y \rangle \geqslant 0 \text{ for all } y \in C. \tag{12.9}$$

In the semi-simple case C and C^* are both of maximal dimension but if $\mathfrak{g}$ has a non-trivial centre then C^* lies in the subspace orthogonal to the centre, i.e. in the semi-simple part. The following lemma relating W, C and C^* is then standard (Bourbaki 1968, ch. VI, prop. 18).

LEMMA 12.10. $x \in C \Leftrightarrow (1-\omega)x \in C^*$ *for all* $\omega \in W$.

The cone C^* defines a natural partial ordering on $\mathfrak{t}$ by

$$x \geqslant y \Leftrightarrow x - y \in C^*. \tag{12.11}$$

With this notation (12.10) can be rewritten as

$$x \in C \Leftrightarrow x \geqslant \omega x \text{ for all } \omega \in W. \tag{12.12}$$

For $U(n)$ the cone C as already mentioned is given by the conditions $x_1 \geqslant x_2 \geqslant \ldots \geqslant x_n$, the standard inner product $\langle x, y \rangle = \sum x_i y_i$ can be rewritten as

$$\langle x, y \rangle = (x_1 - x_2) y_1 + (x_2 - x_3)(y_1 + y_2) + \ldots + (x_{n-1} - x_n)(y_1 + \ldots + y_{n-1}) + x_n (y_1 + \ldots + y_n) \tag{12.13}$$

showing that the cone C^* is given by

$$\sum_{j=1}^{i} y_j \geqslant 0 \quad \text{for} \quad 1 \leqslant i \leqslant n-1$$

and

$$\sum_1^n y_j = 0.$$

Thus the partial ordering (12.11) reduces in this case to that given by (12.1).

The following lemma, which is an easy corollary of (12.10), generalizes the equivalence of (12.1) and (12.5).

LEMMA 12.14. *If* $x, y \in C$ *then*

$$y \in \widehat{W}x \Leftrightarrow y \leqslant x.$$

Proof. Let us first illustrate the geometric meaning of this for $SU(3)$, in which case the diagram is as below.

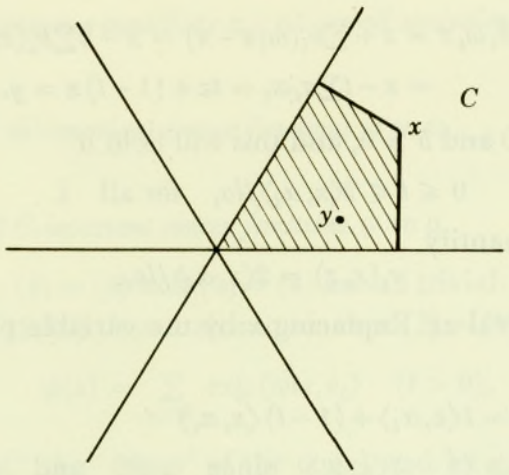

FIGURE 4.

The shaded region is the intersection C with the 'backward' C^*-cone centred at x. It therefore describes the set y such that

$$y \in C \quad \text{and} \quad y \leqslant x.$$

The lemma asserts that this set is also the intersection of C and the hexagon $\widehat{W}x$. In one direction this is easy because for $y \in \widehat{W}x$ we have

$$y = \sum_{\omega \in W} a_\omega \omega x, \quad a_\omega \geqslant 0, \quad \sum a_\omega = 1$$

so that

$$x = (\sum a_\omega)\, x \geqslant \sum a_\omega \omega x = y$$

by (12.12). For the converse it will be enough by continuity to assume that x is an interior point of C and that $x - y$ is an interior point of C^*. The directed line xy then meets the boundary of C in a point z and we must show that the whole finite interval of xz lies in $\widehat{W}x$. Since the relation $y \in \widehat{W}x$ is transitive, it will be enough to show that there is a constant $c(x, z)$ so that, if $y = tz + (1-t)x$, with $0 \leqslant t \leqslant 1$, is any point in the interval zx, then

(i) $c(y, z) \geqslant c(x, z)$,

(ii) $y \in \widehat{W}x$ if $t \leqslant c(x, z)$.

A finite number N of repetitions, where $N^{-1} < c(x, z)$, will then prove that the whole interval $zx \in \widehat{W}x$.

Now let α_i be the simple roots normalized to have length one, so that the α_i are the unit normals to the faces of C and form the basis for C^*. Since x is assumed interior to C we have

$$\langle x, \alpha_i \rangle > 0 \quad \text{for all} \quad i.$$

Since $x - z$ is assumed interior to C^* we have

$$x - z = \sum_{i=1}^{l} a_i \alpha_i, \quad a_i > 0.$$

Now let $\omega_i \in W$ be the reflexion in the face $\langle x, \alpha_i \rangle = 0$ so that

$$\omega_i x = x - 2\langle x, \alpha_i \rangle \alpha_i,$$

and define for $0 \leqslant t \leqslant 1$ constants

$$b_i = t a_i / 2\langle x, \alpha_i \rangle, \quad b = 1 - \Sigma b_i.$$

Then

$$\begin{aligned} bx + \Sigma b_i \omega_i x &= x + \Sigma b_i(\omega_i x - x) = x - 2\Sigma b_i \langle x, \alpha_i \rangle \alpha_i \\ &= x - t\Sigma a_i \alpha_i = tz + (1-t)x = y. \end{aligned}$$

Hence $y \in \widehat{W}x$ provided $b_i \geqslant 0$ and $b \geqslant 0$, and this will hold if

$$0 \leqslant t \leqslant 2\langle x, \alpha_i \rangle / l a_i \quad \text{for all} \quad i.$$

It remains to examine the quantity

$$c_i(x, z) = 2\langle x, \alpha_i \rangle / l a_i,$$

when we vary x on the interval zx. Replacing x by the variable point $y = tz + (1-t)x$, a_i gets replaced by $(1-t)a_i$ and

$$\begin{aligned} \langle y, \alpha_i \rangle &= t\langle z, \alpha_i \rangle + (1-t)\langle x, \alpha_i \rangle \\ &\geqslant (1-t)\langle x, \alpha_i \rangle \quad \text{since} \quad z \in C \quad \text{and} \quad t \geqslant 0. \end{aligned}$$

Hence $c_i(y, z) \geqslant c_i(x, z)$ and the proof is completed by taking $c(x, z) = \min_i c_i(x, z)$.

Remark. The partial ordering $y \leqslant x$ for $x \in C$ is the usual ordering for dominant weights of representations, when we consider not the Lie algebra of T but its dual. The reinterpretation in terms of convex hulls of W-orbits is given in Adams (1969). In our case we are interested not in representations but in conjugacy classes but the partial ordering is essentially the same.

Kostant (1973) proved the following generalization of the Schur–Horn theorem:

$$\pi(Gy) = \widehat{W}y \tag{12.15}$$

where $y \in \mathfrak{t}$, $\pi : \mathfrak{g} \to \mathfrak{t}$ is orthogonal projection and Gy denotes the G-orbit of y under the adjoint action. See also Atiyah (1982) for a different proof in a more general context. Using (12.15), or rather the easier half that gives the inclusion $\pi(Gy) \subset Wy$, we shall now prove the promised result about convex invariant functions:

PROPOSITION 12.16. *Let ϕ be a W-invariant convex function on $\mathfrak{t}$ and ψ the corresponding G-invariant function on $\mathfrak{g}$. Then ψ is also convex.*

Proof. For any function on R^n it will be convenient to define $\Gamma(f)$ to be the region 'above its graph', i.e. all points (x, y) with $x \in R^n$, $y \in R$ such that $y > f(x)$. Convexity of the function f is then equivalent to convexity of $\Gamma(f)$. Recall also that Γ is convex if every boundary point a has a supporting hyperplane H_a (i.e. Γ is contained in one of the two half-spaces complementary to H_a). Now consider the functions ϕ, ψ and the corresponding regions

$$\Gamma(\phi) \subset \mathfrak{t} \oplus R, \quad \Gamma(\psi) \subset \mathfrak{g} \oplus R.$$

Because ψ is G-invariant so is $\Gamma(\psi)$ and it is therefore sufficient to prove the existence of a supporting hyperplane to $\Gamma(\psi)$ at boundary points (λ, y) of $\Gamma(\phi)$. By hypothesis $\Gamma(\phi)$ is convex so we have a supporting hyperplane $H \subset \mathfrak{t} \oplus R$. Let $H' = \pi^{-1}(H) \subset \mathfrak{g} \oplus R$ where π is orthogonal projection. We shall show that H' is the required supporting hyperplane for $\Gamma(\psi)$. Any point $(x, y) \in \Gamma(\psi)$ satisfies $y > \psi(x)$. From (12.15) and the convexity of ϕ we see that

$$\psi(x) \geqslant \phi(\pi x).$$

Hence $y > \phi(\pi x)$ so that $(\pi x, y) \in \Gamma(\phi)$ and hence is on one side of H. This means precisely that (x, y) is on the corresponding side of H', which completes the proof.

As with $U(n)$ we can give several further equivalent definitions of the partial ordering and we summarize this in

PROPOSITION 12.17. *The following conditions $x, y \in \mathfrak{t}$ are all equivalent*:

(1) $\widehat{Wy} \subset \widehat{Wx}$;

(2) $\phi(y) \leqslant \phi(x)$ *for all W-invariant convex functions ϕ on* $\mathfrak{t}$;

(3) $\widehat{Gy} \subset \widehat{Gx}$;

(4) $\psi(y) \leqslant \psi(x)$ *for all G-invariant convex functions ψ on* $\mathfrak{g}$.

Proof. (1) $\Rightarrow$ (2), (2) $\Rightarrow$ (4), (1) $\Rightarrow$ (3) and (3) $\Rightarrow$ (4) are all trivial. (3) $\Rightarrow$ (1) follows from (12.15) and (4) $\Rightarrow$ (2) follows from (12.16). It remains to see that (2) $\Rightarrow$ (1). For this we take

$$\phi(x) = \sum_{\omega \in W} \exp\langle \omega t x, e_i \rangle \quad (t > 0),$$

where the e_i are a basis of C (the 'edges' of the cone) and let $x, y \in C$. In view of (12.11) and (12.12) we have

$$\langle x, e_i \rangle \geqslant \langle \omega x, e_i \rangle \quad \text{for} \quad \omega \neq 1 \quad \text{and all} \quad i,$$
$$\langle y, e_i \rangle \geqslant \langle \omega y, e_i \rangle.$$

It will be sufficient by continuity to suppose that both x and y are interior to C; then the above inequalities are strict so that for large t the first term (for $\omega = 1$) in the sum defining ϕ is dominant. Hence

$$\begin{aligned} \phi(y) \leqslant \phi(x) &\Rightarrow \langle x, e_i \rangle \geqslant \langle y, e_i \rangle \quad \text{for all} \quad i \\ &\Rightarrow x - y \in C^* \\ &\Rightarrow y \leqslant x \quad \text{by} \quad (12.11). \end{aligned}$$

Remark. Proposition 12.17 remains true when 'convex' is interpreted as 'smooth convex' (or even analytic). This is clear from the proof because, for the essential implication (2) $\Rightarrow$ (1), we use only exponential functions.

If we take any irreducible representation $\rho: G \to U(n)$ it has weights $\lambda_1, \ldots, \lambda_n$, which we may view as elements of $\mathfrak{t}$, so that any $x \in \mathfrak{t}$ gives rise to the hermitian matrix with eigenvalues $x_j = \langle x, \lambda_j \rangle$. If λ_1 is the maximal weight then $\lambda_1 \in C$ and for all $j \neq 1$, $\lambda_1 > \lambda_j$, i.e. $\lambda_1 - \lambda_j \in C^*$. This means that, if $x \in C$, then $x_1 \geqslant x_j$ for $j \neq 1$, so that x_1 is the largest eigenvalue. Hence if $x, y \in C$ and we assume $\rho(y) \leqslant \rho(x)$ then in particular $y_1 \leqslant x_1$, i.e. $\langle y, \lambda_1 \rangle \leqslant \langle x, \lambda_1 \rangle$. If we let ρ run over all irreducible representations then λ_1 runs over all integral dominant weights and these span C: in fact there are l basic integral weights e_i that lie in the edges of C and generate it. This proves (cf. Kostant 1973)

PROPOSITION 12.18. *For $x, y \in \mathfrak{g}$ we have $y \leqslant x \Leftrightarrow \rho(y) \leqslant \rho(x)$ for all unitary representations ρ.*

This completes our survey of convexity and essentially contains all the results we have used earlier. Thus in §8 we used the inequality

$$\phi\begin{bmatrix} A & B \\ C & D \end{bmatrix} \geqslant \phi\begin{bmatrix} \alpha & 0 \\ 0 & \delta \end{bmatrix}$$

for every convex invariant function ϕ on $\mathfrak{u}(n)$, where $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$ is a skew-hermitian matrix in block form and α, δ are the central components of A, D. This follows in fact from Horn's theorem and the observation that $\begin{bmatrix} \alpha & 0 \\ 0 & \delta \end{bmatrix}$ is in the convex hull of the Σ_n-orbit of the diagonal part of $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$. The same proof holds for the more general block decomposition also used in §8.

The result (8.21) (and its generalization to any G) is an easy consequence of (12.17). In fact if $\phi(y) = \phi(x)$ for all W-invariant convex functions ϕ then $\widehat{Wy} = \widehat{Wx}$ and so the extreme points of these two convex polyhedra must coincide. But the extreme points of $\widehat{Wx}$ are certainly among the finite set Wx. Hence Wx and Wy intersect and so coincide.

13. Equivariant cohomology

In this section we shall review some of the general facts about equivariant cohomology and establish some of the more particular results that we have had to use.

We recall first that for any topological group G the classifying space BG is defined as the base of a principal G-bundle whose total space is contractible. It is unique up to homotopy so that in particular $H^*(BG)$ depends only on G. More generally if G acts on a space X we define X_G as the associated bundle over BG with fibre X and define the equivariant cohomology by

$$H_G(X) = H(X_G),$$

so that $H(BG)$ is now the equivariant cohomology of a point. If G acts freely on X so that $X \to X/G$ is a principal G-bundle then the map $X_G \to X/G$ has contractible fibres and so is a homotopy equivalence. Thus in this case

$$H_G(X) \cong H(X/G).$$

Suppose now that K is a closed normal subgroup of G, and that X is a G-space on which K acts freely (with $X \to X/K$ a principal K-bundle). Then the quotient group $S = G/K$ acts on $Y = X/K$, and we have

$$H_G(X) \cong H_S(Y). \tag{13.1}$$

To see this let E_1, E_2 be the total spaces of universal bundles of G, S respectively. Note that G acts on E_2 via S so that $E_1 \times E_2$ is also a free contractible G-space; we can thus take

$$X_G = X \times_G (E_1 \times E_2).$$

Projecting onto $X \times_G E_2$ with fibre E_1 is a homotopy equivalence and

$$X \times_G E_2 = Y \times_S E_2 = Y_S$$

so that $X_G \sim Y_S$ proving (13.1).

So far we have been rather imprecise about the class of topological groups to be considered and the reader might feel uneasy about the use of these ideas for the large infinite-dimensional groups $\mathscr{G}$ of gauge transformations. In view of (13.1) we can effectively reduce all our application

of equivariant cohomology to the case when G is a compact Lie group. In fact if $\mathcal{G}$ is the group of gauge transformations of a principal G-bundle over a manifold M, it has a normal subgroup $\mathcal{G}_0$ consisting of transformations that are the identity at some fixed base point of M and the quotient group is isomorphic to G. Moreover $\mathcal{G}_0$ acts freely on the space of connections so that by (13.1)

$$H_{\mathcal{G}}(\mathcal{A}) \cong H_G(\mathcal{A}/\mathcal{G}_0)$$

with a similar result for any $\mathcal{G}$-stable subspace of $\mathcal{A}$. We could therefore always work in $\mathcal{A}/\mathcal{G}_0$ and use G-equivariant cohomology if we wished.

If $K \subset G$ and Y is a K-space we define its 'extension' to a G-space X by putting $X = G \times_K Y$. Note that X is just the bundle with fibre Y and base G/K associated with the principal K-bundle $G \to G/K$. If E is a free contractible G-space then

$$X_G = E \times_G X = E \times_G G \times_K Y = E \times_K Y = Y_K$$

so that $H_G(X) \cong H_K(Y)$.

We come now to some more specific results, which concern compact connected Lie groups G *without torsion* in their cohomology. The examples we need are just $U(n)$ and more generally products of the form $U(n_1) \times \ldots \times U(n_r)$. If T is a maximal torus of G then it is well known that the fibration

$$G/T \to B_T \to B_G$$

behaves like a product for integral cohomology and all the spaces involved have no torsion. It follows that, for any G-space X, the induced fibration

$$G/T \to X_T \to X_G$$

is multiplicative for integral cohomology

$$H(X_T) \cong H(X_G) \otimes H(G/T),$$

so that

(13.2) $$H_G(X) \quad \textit{is a direct summand of} \quad H_T(X).$$

or equivalently for all primes p

(13.3) $$H_G(X, Z_p) \to H_T(X, Z_p) \quad \textit{is injective.}$$

Next let $T = T_0 \times T_1$ be the product of two subtori with T_0 acting trivially on the connected T-space X. Then

$$X_T = BT_0 \times X_{T_1}$$

so that for Z_p coefficients (and any prime p)

$$H_T(X) \cong H(BT_0) \otimes H_{T_1}(X).$$

Now $H(BT_0)$ is a polynomial ring and so any $\alpha_0 \in H(B_{T0})$ with $\alpha_0 \neq 0$ is not a zero-divisor in $H_T(X)$. More generally if $\alpha \in H_T(X)$ restricts to such an α_0, i.e. if

$$\alpha = \alpha_0 \otimes 1 + \text{terms of positive degree in } H_{T_1}(X),$$

the same holds. This follows on filtering by the degree in $H_{T_1}(X)$ and noticing that α acts via α_0 on the associated graded module.

In our application the element α_0 above will occur as the Chern class of a vector bundle N_{T_0} over BT_0 arising from a representation N of T_0. For dim $N = 1$ the assignment

$$N \to c_1(N_{T_0})$$

gives rise to an isomorphism

$$\hat{T}_0 \cong H^2(BT_0, Z),$$

(where $\hat{T}_0$ is the character group of T_0) which we shall consider as an identification. The whole cohomology ring $H^*(BT_0, Z)$ can then be viewed as the symmetric algebra of the lattice $\hat{T}_0$. For an n-dimensional representation N therefore we decompose

$$N = \sum_{j=1}^{n} L_j$$

into one-dimensional representations, and

$$c_n(N_{T_0}) = \prod_{j=1}^{n} L_j.$$

If each L_j is primitive, i.e. is not divisible in $\hat{T}_0$ by any prime p, we shall say that N is *primitive*. In this case $c_n(N_{T_0})$ is clearly non-zero when reduced mod p for any p.

We shall now put all these remarks together into the following.

PROPOSITION 13.4. *Let X be a connected G-space on which some subtorus T_0 acts trivially and let N be a G-vector bundle on X. Assume that the representation of T_0 on the fibre of N is primitive and that $H(G)$ has no torsion. Then multiplication by the top Chern class $\alpha = c_n(N_G)$ on $H_G(X, Z_p)$ is injective for all primes p.*

The proof follows from 13.3, which allows us to restrict from G to a maximal torus $T \supset T_0$, so that we are in the situation just discussed.

14. SOBOLEV SPACES

In this section we shall show, by introducing appropriate Sobolev spaces of functions, how to justify our heuristic use of infinite-dimensional manifolds. Much of this is standard and can be found in Narasimhan & Ramadas (1979), Uhlenbeck (1981) or Mitter & Viallet (1981) but some of the more detailed results related to the complex structure depend of course on the dimensionality of the base manifold being 2. For this reason we shall give a self-contained account tailored to our purposes.

For the convenience of the reader we shall now recall some of the basic facts about Sobolev spaces. For fuller details we refer to Palais (1965, ch. 9). On a compact smooth n-dimensional manifold M the space L^p_k (for $1 \leqslant p < \infty$) denotes those functions f all of whose derivatives up to and including order k are in the Lebesgue space L^p. The definition can be extended to non-integral k and to sections f of any smooth complex vector bundle over M. Each L^p_k is a Banach space and for $p = 2$ is a Hilbert space also denoted by H^k. The Sobolev embedding theorems assert that

(14.1) $L^p_k \subset L^q_l$ if $k \geqslant l$ and $k - n/p \geqslant l - n/q$ and the inclusion is *compact* if we have *strict* inequalities,

(14.2) $L^p_k \subset C^l$ if $k - n/p > l$, and the inclusion is compact.

Here C^l (for integral $l \geqslant 0$) denotes as usual sections whose partial derivatives of order $\leqslant l$ are continuous. In particular

(14.3) $$\bigcap_{k=1}^{\infty} H^k = C^\infty.$$

Recall now the Hölder inequality

$$\|fg\|_r \leqslant \|f\|_p \, \|g\|_q,$$

where $\| \|_p$ denotes the L^p-norm and $r^{-1} = p^{-1} + q^{-1}$. This implies the continuity of the multiplication map

$$L_0^p \times L_0^q \to L_0^r. \tag{14.4}$$

Applying this together with (14.1) one deduces

(14.5) L_k^p is a Banach algebra for $k > n/p$ and L_j^p is a topological L_k^p-module for $0 \leqslant j \leqslant k$.

In the good range $k > n/p$ one can also define nonlinear generalizations of the spaces L_k^p. Thus one can define the space $L_k^p(M, N)$ of maps $f: M \to N$ where N is another C^∞ manifold. More generally one can define the L_k^p-sections of a C^∞ fibre bundle over M with N as fibre. These spaces are dense in the space of continuous sections. In particular on taking N to be a Lie group and using (14.5) the L_k^p automorphisms of a vector bundle (or a unitary bundle) form Lie groups.

We come now to the special case that interests us, namely $n = \dim M = 2$. For a complex C^∞ vector bundle E over M with hermitian metric we then have for $k \geqslant 2$ the real Lie group of unitary automorphisms of class $H^k = L_k^2$, which we shall denote by $\mathcal{G}^k$. Its 'complexification' $(\mathcal{G}^c)^k$ is the (complex) Lie group of all automorphisms of E of class H^k. Since automorphisms act on connections by affine transformations it follows from (14.5) that we can define the space of unitary connections $\mathcal{A}^{k-1}$ of class H^{k-1} and that $\mathcal{G}^k$ acts smoothly on $\mathcal{A}^{k-1}$. For the same reasons, when we view $\mathcal{A}$ as the space $\mathcal{C}$ of (almost) complex structures (or d″-operators) we see that the complex Lie group $(\mathcal{G}^c)^k$ also acts smoothly on $\mathcal{A}^{k-1}$.

For $k > 2$ the space $\mathcal{A}^{k-1}$ consists of *continuous* connections. However, the most natural space for our purposes is in fact $\mathcal{A}^1$, so the reader should remember that this includes discontinuous connections. A little more care will be necessary in various places but there is no fundamental difficulty. As an indication of this we shall establish the following regularity results.

LEMMA 14.6. *For $k \geqslant 2$ and any $A \in \mathcal{A}^{k-1}$ let $F: (\mathcal{G}^c)^k \to \mathcal{A}^{k-1}$ be the map given by the action on A i.e. $F(g) = g(A)$. Then the differential* dF *at the identity is a Fredholm operator.*

Proof. The differential dF at the identity is just the operator d_A'' acting from H^k-sections of End E to H^{k-1}-sections of $\Omega^{0,1}(\mathrm{End}\, E)$. If we fix a standard C^∞ connection A_0 on E then $A = A_0 + B$ and

$$\mathrm{d}_A'' \phi = \mathrm{d}_0'' \phi + [B, \phi].$$

Since $B \in H^{k-1}$ and $\phi \in H^k$ the mapping $\phi \to [B, \phi]$ can by (14.5) be factored through the compact inclusion $H^k \to H^{k-\frac{1}{2}}$ and so is compact. Since, for the smooth connection A_0 the operator d_0'' is elliptic of order 1 and so Fredholm, it follows that d_A'' is also a Fredholm operator.

Applying the smooth group action it follows that dF is a Fredholm operator at all points of the orbit of A. The implicit function theorem for Banach manifolds then implies (for $k \geqslant 2$)

(14.7) *for neighbourhoods U of the identity in $(\mathcal{G}^c)^k$ and V of A in $\mathcal{A}^{k-1}$, the image $U(A)$ is a closed Banach submanifold of V of finite codimension.*

From this we shall deduce

LEMMA 14.8. *For $k \geqslant 2$, every $(\mathcal{G}^c)^k$-orbit in $\mathcal{A}^{k-1}$ contains a C^∞-connection.*

Proof. Let N be a finite-dimensional subspace of $\mathcal{A}^{k-1}$ transversal to the orbit at A, i.e. N is a complement to the image of dF. Then (14.7) implies that for a suitably small neighbourhood V of $A \in \mathcal{A}^{k-1}$ we have a continuous map $\pi: V \to N$ with $\pi^{-1}(A) = U(A)$. Now for any $r+1$ points $B_1, \ldots, B_{r+1} \in V$ (where $r = \dim N$) let

$$f_B: \sigma^r \to V$$

be the affine linear map sending the vertices of σ^r to the B_i. Composing with π we then get a continuous map

$$\pi f_B : \sigma^r \to N,$$

which depends continuously on the B_i. Start now with any set of $B_i \in N$ that span a maximal-dimensional simplex having A as barycentre. For this choice of B, πf_B restricts to give a map

$$\partial\sigma^r \to N - A,$$

which generates $H_{r-1}(N-A) \cong Z$. By continuity it follows that this will be true for πf_C with $|C_i - B_i| < \epsilon$. Hence πf_C must take the value A on some point of σ^r, i.e. $f_C(\sigma^r)$ intersects $U(A)$. Finally since $\mathscr{A} = \mathscr{A}^\infty$ is dense in $\mathscr{A}^{k-1}$ we can find such $C_i \in \mathscr{A}$ for any ϵ. Then every point of the linear span of $C_1, \ldots, C_{r+1}$ is in $\mathscr{A}$ and so the intersection $f_C(\sigma^r) \cap U(A)$ is in $\mathscr{A}$, proving that the orbit of A contains a C^∞-connection.

Conversely we shall prove

LEMMA 14.9. *For $k \geqslant 2$ let $A, B \in \mathscr{A}$ and $g \in (\mathscr{G}^c)^k$ with $B = g(A)$. Then $g \in \mathscr{G}^c$, i.e. g is C^∞.*

Proof. The two connections B, A differ by a C^∞ 1-form $B - A$. The condition $B = g(A)$ is more explicitly written

$$g^{-1} \mathrm{d}''_A g = (B-A)'',$$

where ω'' is the $(0,1)$-part of the 1-form ω. Hence

$$\mathrm{d}''_A g = g(B-A)''.$$

Since $(B-A)'' \in C^\infty$ and $g \in H^k$ the product lies also in H^k (by (14.5)). The standard regularity theorem for the smooth elliptic operator d''_A then implies that $g \in H^{k+1}$. By iteration this proves that g is C^∞.

We have now established all the local regularity properties that we need concerning the action of the group $\mathscr{G}^c$ on the space $\mathscr{A}$. In particular the orbit through any point A has, as local transversal, the harmonic space $H^{0,1}(\mathrm{End}\, E)$, which is isomorphic to the sheaf cohomology group $H^1(M, \mathrm{End}\, E)$. The structure of nearby orbits is then entirely determined by their intersection with this (or any other) transversal slice N. More precisely the union of all nearby orbits in $\mathscr{A}^k$ is $(\mathscr{G}^c)^k$-equivariantly homeomorphic to the fibre bundle over the orbit of A with fibre N and group the stabilizer of A (which is finite-dimensional and consists of the automorphisms of the holomorphic bundle $E(A)$ defined by A).

In the next section we shall use standard algebro-geometric methods to establish the *global* properties of our stratification. For the present we note simply that the stratification of $\mathscr{A}$, which has been defined so far only for smooth connections, extends naturally to $\mathscr{A}^k$ for any $k \geqslant 1$ by our regularity results. The discussion in §7 can then all be made rigorous in terms of Sobolev spaces and Banach Lie groups. Thus the groups $\mathrm{Aut}\,(E)$, $\mathrm{Aut}\,(E_\mu)$ will be replaced by the Banach Lie groups $\mathrm{Aut}^k(E)$, $\mathrm{Aut}^k(E_\mu)$ and the space $\mathscr{F}_\mu$ by $\mathscr{F}^k_\mu$, which can be identified with the homogeneous space

$$\mathrm{Aut}^k(E)/\mathrm{Aut}^k(E_\mu).$$

Similarly replacing $\mathscr{B}_\mu, \mathscr{C}_\mu$ by $\mathscr{B}^{k-1}_\mu, \mathscr{C}^{k-1}_\mu$ we have a continuous map

$$\mathrm{Aut}^k(E) \times \mathscr{B}^{k-1}_\mu \to \mathscr{C}^{k-1}_\mu.$$

This map is constant on the orbits of $\mathrm{Aut}^k(E_\mu)$ and so induces a map

$$\mathscr{H}^{k-1}_\mu \to \mathscr{C}^{k-1}_\mu,$$

where $\mathscr{H}_\mu^{k-1}$ denotes the total space of the homogeneous fibre bundle over $\mathscr{F}_\mu^k$ with fibre $\mathscr{B}_\mu^{k-1}$. Our regularity results tell us that this map is continuous and bijective. To establish that it is actually a homeomorphism we need finally to prove that the map

$$f_\mu^k : \mathscr{C}_\mu^{k-1} \to \mathscr{F}_\mu^k \tag{14.10}$$

refining (7.9) is continuous. In other words we have to show that the canonical filtration varies continuously along $\mathscr{C}_\mu$ with a 'gain of one derivative'.

Since the group $\mathrm{Aut}^k(E)$ acts continuously on both $\mathscr{C}_\mu^{k-1}$ and $\mathscr{F}_\mu^k$, and commutes with f_μ^k, it will be sufficient by (14.8) to prove continuity of f_μ^k at C^∞ points A of $\mathscr{C}_\mu$. Moreover by our regularity theorems it will then be sufficient to prove continuity in the harmonic space $H^{0,1}(\mathrm{End}\, E)$ at A (or any other smooth transversal N). On such a finite-dimensional space all the Sobolev norms are now equivalent and the problem can be reduced to one of algebraic geometry, which will be dealt with in the next section.

Once the continuity of f_μ^k is proved it follows that our strata are locally closed submanifolds of finite codimension. Moreover the homotopy properties of the various function spaces are all independent of k by standard approximation theorems (Palais 1965, th. 13.14). This then justifies our heuristic arguments in § 7.

Finally we note the continuity properties of the Yang–Mills functional.

(14.11) *The curvature $F(A)$ extends by continuity to a quadratic function $\mathscr{A}^1 \to H^0 = L^2$, so that the Yang–Mills functional L gives a smooth function $\mathscr{A}^1 \to R$.*

The proof (given under more general conditions by Uhlenbeck (1982)) is a straightforward consequence of the multiplicative properties of Sobolev spaces. In fact, writing $A \in \mathscr{A}^1$ in the form $A = A_0 + B$ with A_0 a fixed C^∞ connection, we see that

$$F(A) = F(A_0) + \mathrm{d}_0 B + \tfrac{1}{2}[B, B].$$

Since $B \in H^1$ we have $\mathrm{d}_0 B \in H^0$ and $[B, B] \in H^0$ (using the inclusion $H^1 \to L_0^4$ from (14.1) and the multiplication (14.4)).

This result explains why $\mathscr{A}^1$ is the most natural Sobolev space for the Yang–Mills functional, although for our purposes any $\mathscr{A}^k$ with $k \geqslant 1$ would do equally well.

As we have seen in earlier sections the strata should be seen as the Morse strata of the Yang–Mills functional. A more careful analysis of the gradient flow or some alternative differential-geometric argument might be able to show this directly and in particular to establish that $\mathscr{A}_\mu^k$ is a locally closed submanifold of $\mathscr{A}^k$ for all μ. We have not found an argument on these lines, which is why we have to resort in the next section to algebraic geometry.

15. The stratification in algebraic geometry

In this section M will denote a complete non-singular algebraic curve defined over a ground field k of characteristic zero. As observed by Harder & Narasimhan (1975) the definition of semi-stability and the canonical filtration of vector bundles over M does not require k to be algebraically closed. In fact the *uniqueness* of the canonical filtration over $\bar{k}$ implies that it is already defined over k. Moreover if E_k is a vector bundle defined over k, E_K its extension to *any* larger (finitely-generated) field K, then

$$E_k \text{ is semi-stable} \Leftrightarrow E_K \text{ is semi-stable.} \tag{15.1}$$

To see this we may first replace k by its algebraic closure in K so that K is purely transcendental over k. Assume now that E_K is not semi-stable so that there exists a sub-vector bundle F_K with $\mu(F_K) > \mu(E_K)$ where μ denotes as usual the normalized Chern class. If $K = k(x_1, \ldots, x_n)$ the bundles E_K, F_K can be represented by vector bundles $\mathscr{E}$, $\mathscr{F}$ over $M \times U$ where U is some Zariski open set of k^n. Moreover $\mathscr{E}$ can be taken to be the pull-back of E_k under the projection $M \times U \to M$. Now restrict $\mathscr{E}, \mathscr{F}$ to any point of U algebraic over k and we find a sub-bundle F_k of E_k with $\mu(F_k) > \mu(E_k)$ so that E_k is not semi-stable. The opposite implication is trivial so that (15.1) is proved.

Now let k be algebraically closed, S an irreducible algebraic variety over k and let $\mathscr{E}$ be a vector bundle over $M \times S$, which we interpret as an algebraic family E_s of vector bundles over M parametrized by $s \in S$. A fairly elementary result proved by Narasimhan & Seshadri (1965) is that the set of points s for which E_s is semi-stable is a *constructible* sub-set of S. We recall that a constructible set is a finite disjoint union of locally closed subsets in the Zariski topology, and X is locally closed if it is open in the closure $\bar{X}$. Constructibility is preserved under finite unions, intersections, complements, direct and inverse images. Since we shall need to refine this result of Narsimhan & Seshadri we recall the essentials of the proof. First one shows that any indecomposable bundle F of smaller rank such that

(i) $\mu(F) > \mu = \mu(E_s)$,

(ii) $\mathrm{Hom}\,(F, E_s) \neq 0$ for some $s \in S$

must belong to one of a finite number of irreducible families. Let T be the parameter space of one of these families. Then the subset $Z \subset T \times S$ consisting of all points (t, s) such that

$$\mathrm{Hom}(F_t, E_s) \neq 0$$

is a *closed* subset. Its projection onto S is not necessarily closed but it is constructible. This shows that the set of $s \in S$ for which E_s is not semi-stable is constructible and so therefore is the complementary set.

We want to prove the following

LEMMA 15.2. *Let $K = k(S)$ be the function field of S, E_K the bundle over M defined over K arising from $\mathscr{E}$. Assume E_K is semi-stable, then there exists an open set $U \subset S$ such that E_s is semi-stable for all $s \in U$.*

Proof. Assume the conclusion false. Then for at least one of the parameter spaces T occurring above the corresponding sub-set $Z \subset T \times S$ must project onto a dense set of S (i.e. containing an open set). Replace Z by an irreducible component with the same property and it follows that $K' = k(Z)$ is an extension of $K = k(S)$. The definition of Z, together with the coherence of direct images, shows that we have a non-zero homomorphism

$$F_{K'} \to E_{K'}.$$

Since $\mu(F_{K'}) > \mu(E_{K'})$ this means $E_{K'}$ is not semi-stable. By (15.1) this means E_K is not semi-stable and gives the required contradiction.

We return now to consider a general family E_s parametrized by S. Passing to the quotient field $K = k(S)$ we consider the canonical filtration of E_K. This filtration can be represented by a filtration for the family $\mathscr{E}$ restricted to some open set $U \subset S$. The associated quotient bundles being semi-stable over K will, by (15.2), remain semi-stable over suitable open subsets of U. Hence there is an open set $V \subset S$ so that our filtration is canonical at all points of V. In particular the *type* of E_s is constant for all $s \in V$. Removing V from S we get a variety (possibly reducible) of smaller dimension. Applying induction therefore we have proved the following

PROPOSITION 15.3. *Let E_s be a family of bundles over M parametrized by S, and stratify S according to the type of E_s. Then each stratum is a constructible set.*

Note. This result is proved by Shatz (1977) in a different way. Our proof, using the approach of Narasimhan & Seshadri (1965), is more in line with the rest of our paper.

When $k = C$, the field of complex numbers, we have shown in §8 that, with respect to the partial ordering studied in §12, the subset $\bigcup_{\lambda \geqslant \mu} \mathscr{A}_\lambda$ is closed in $\mathscr{A}$. This implies in particular that in any algebraic family, as in (15.3), the corresponding set $\bigcup_{\lambda \geqslant \mu} S_\lambda$ is closed in the usual complex numbers topology. Since it is also constructible it follows (Narasimhan & Seshadri 1965, lemma 12.2) that it is Zariski closed. Hence each stratum S_λ must be locally closed in the Zariski topology (cf. Shatz 1977). This is nearly but not quite enough to show that the $\mathscr{A}_\mu$ themselves are locally closed. For this we need to examine further the continuity properties of the canonical filtration.

The proof of 15.3 shows that over a Zariski dense open set V of each stratum S_μ the canonical filtration varies algebraically. If we introduce the appropriate flag-bundle F_μ over S_μ this means we have a regular section of F_μ over V: in particular this section is continuous (for the C-topology). In fact continuity holds everywhere:

PROPOSITION 15.4. *Let E_s be a family of bundles over M parametrized by an irreducible variety S and assume all E_s are of the same type μ. Then there is a continuous filtration of the bundle $\mathscr{E}$ over $M \times S$ that induces the canonical filtration on each E_s.*

Proof. As we have observed above there will be a Zariski open set $V \subset S$ with the required property. Also we can proceed by induction on the length of the filtration so we can restrict essentially to filtrations of length two. Such a filtration is determined by a section ξ of the appropriate Grassmann bundle, and it will be sufficient to show that the Zariski closure of ξ over $M \times V$ coincides everywhere with the canonical section (because $\bar{\xi} \to S$ is then proper and bijective, hence a homeomorphism). Since every point in the Zariski closure can be approached along a curve we can suppose that $\dim S = 1$. Moreover there is no essential change in replacing S by its desingularization so we may suppose S non-singular. Our section ξ over $M \times V$ is then a surface and its Zariski closure intersects the Grassmann bundle G_s over $M \times \{s\}$, for $s \in S - V$, in some algebraic curve $\bar{\xi}_s$. We have to show that $\bar{\xi}_s$ is just the canonical section ξ_s. Consider the irreducible curves that make up $\bar{\xi}_s$. We claim there is just one of these, say C, giving a section of G_s over M and any others, say D_j, must lie entirely in the fibres (over points of M). The reason is purely homological: since ξ is a section generically the intersection number of $\bar{\xi}_s$ with a fibre over $G_s \to M$ must be one. Now we shall use the assumption that E_s is of constant type to deduce that there are no D_j. To do this let F_s be the universal vector bundle over G_s, i.e. the fibre of F_s at a point $\gamma \in G_s$ is the vector space represented by γ. Hence the bundle $\xi_s^*(F_s)$ over M is by definition the canonical sub-bundle of E_s and so has Chern class k_1 say (independent of s). On the other hand it is well known (cf. §8) that on the Grassmannian itself the universal bundle has *negative* Chern class. Hence

$$c_1(F_s|D_j) < 0$$

for any component D_j of $\bar{\xi}$ lying in the fibres. On the other hand the intersection number

$$c_1(F_s) \cdot \bar{\xi}_s$$

must be independent of s and hence is equal to k_1. But since $\bar{\xi}_s = C + \sum D_j$ we have

$$k_1 = c_1(F_s) \cdot C + \sum_j c_1(F_s) \cdot D_j \leqslant c_1(F_s) \cdot C$$

and the inequality is strict unless the D_j do not occur. Now C defines a section of $G_s \to M$ and so a sub-bundle F of E_s with Chern class

$$c_1(F_s) \cdot C > k$$

if the D_j occur. But the assumption about constancy of type of the E_s means that no such F_s can exist. Hence $\bar{\xi}_s = C$ and the proof is complete.

Remark. The continuous filtration ξ in 15.4 defines a continuous section $\xi(z, s)$ of the flag bundle over $M \times S$. Now, for each $s \in S$, ξ is holomorphic in z and hence its z-derivatives can be estimated in terms of its sup norm. This shows that ξ is actually continuous from S to the space $\mathscr{F}_\mu$ (the smooth filtrations of E over M of type μ).

Proposition 15.4 shows that, for an *algebraic* family of bundles over M, the canonical stratification is continuous. To prove the continuity of the map (14.10) it remains now to show that we can always construct 'sufficiently large' algebraic families. More precisely we need to show that for any $A \in \mathscr{E}$ we can find a smooth transversal N to the $\mathscr{G}^c$-orbit through A that represents (locally) an algebraic family. This means we have to prove the following lemma.

LEMMA 15.5. *Let E_0 be an algebraic vector bundle over the algebraic curve M. Then there exists an algebraic family of bundles E_s parametrized by a non-singular variety S such that*

(i) $E_0 \cong E_{s_0}$ for some $s_0 \in S$,

(ii) *the infinitesimal deformation map*

$$\phi : T_{s_0}(X) \to H^1(M, \operatorname{End} E_0)$$

is an isomorphism.

Before giving the proof we make a few comments on (ii). Here $T_{s_0}(S)$ denotes the tangent space to S at s_0. The map ϕ is defined quite generally in such circumstances as follows. Consider the sheaf $\mathscr{O}^1(M) = \mathscr{O}(M \times S)/m^2$ where m is the ideal sheaf of $M \times s_0$ in $M \times S$. We then have an exact sequence of sheaves

$$0 \to \mathscr{O}(M) \otimes T^* \to \mathscr{O}^1(M) \to \mathscr{O}(M) \to 0,$$

where $T = T_{s_0}(S)$. For the bundle $\mathscr{E}$ on $M \times S$ representing the family E_s we have correspondingly an exact sequence

$$0 \to (\operatorname{End} E_0) \otimes T^* \to \mathscr{O}^1(M) \otimes (\operatorname{End} \mathscr{E}) \to \operatorname{End} E_0 \to 0.$$

From the cohomology of this sequence we obtain the coboundary

$$\delta : H^0(M, \operatorname{End} E_0) \to H^1(M, \operatorname{End} E_0) \otimes T^*.$$

The image of the identity endomorphism gives therefore an element of

$$\operatorname{Hom}(T, H^1(M, \operatorname{End} E_0))$$

and this is the infinitesimal deformation map ϕ.

From the Riemann surface point of view this map can also be defined as follows. First we restrict to a small neighbourhood U of s_0 in S over which the family E_s is a product family, so that we can identify all E_s with $E_{s_0} = E_0$ differentiably. Next fix a hermitian metric on E_0 so that we get a family of unitary connections on E_0 parametrized by U. This gives a map

$$\psi : U \to \mathscr{A}$$

with $\psi(s_0) = A$ representing the bundle E_0. The differential of ψ at s_0 is then a map of T into the tangent space to $\mathscr{A}$ at A. Projecting onto the normal to the $\mathscr{G}^c$-orbit then gives the infinitesimal

map. Thus the condition (ii) precisely guarantees that $N = \psi(U)$ will be a smooth transversal to the $\mathscr{G}^{\mathbb{C}}$-orbit of A. As we observed in § 14 continuity of the map (14.10) is equivalent to continuity of the canonical filtration along the corresponding stratum in N. But N is (locally) diffeomorphic to U and so the continuity follows from proposition 15.4 (and the subsequent remark).

We return now to give the proof of lemma 15.5. Observe first that it is sufficient to find an S with ϕ surjective, because we can then always pick a submanifold of S transversal to the kernel of ϕ to get an isomorphism. We now proceed by induction on the rank n of E_0. Let $c_1(E_0) = q$. Then we can always represent E_0 as an extension of the form

$$0 \to I_{n-1}(-m) \to E_0 \to L_0 \to 0,$$

where I_{n-1} is a trivial bundle of rank $n-1$ and m is a suitably large integer (depending on E_0). For the proof see Atiyah (1957) where a stronger result is proved, namely that for indecomposable E_0 the integer m depends only on n, q and the genus of M. Moreover we can assume m chosen so that

$$q + mn > 2g - 2,$$

which will imply that

$$\begin{cases} H^0(M, L^* \otimes I_{n-1}(-m)) = 0, \\ H^1(M, L \otimes I_{n-1}(m)) = 0. \end{cases} \tag{15.6}$$

Now consider bundles E given by extensions of the form

$$0 \to F(-m) \to E \to L \to 0, \tag{15.7}$$

where F is a bundle of rank $n-1$ and $c_1(F) = 0$ while L is a line-bundle with

$$c_1(L) = c_1(L_0) = q + m(n-1).$$

Applying our inductive hypothesis to the trivial bundle I_{n-1} we obtain a family F_r parametrized by $r \in R$ having properties (i) and (ii). We then take for our family E_s all extensions of the form (15.7) with $F = F_r$ parametrized by R, and L parametrized by the Jacobian J of M. Now extensions of this type are classified by elements of

$$H^1(M, L^* \otimes F(-m)). \tag{15.8}$$

By (15.6) the corresponding H^0 vanishes when $F = I_{n-1}$ and hence for all F_r with $r \in R_1$, some Zariski neighbourhood of r_0 in R. Then H^1 will have constant dimension and so our parameter space S is fibred over $R_1 \times J$ with fibre the vector space (15.8).

We must now investigate $H^1(M, \operatorname{End} E)$ for any E in our family. Denote by $\operatorname{End}' E$ the subspace of endomorphisms preserving the exact sequence (15.7), and by $\operatorname{End}'' E$ the quotient:

$$0 \to \operatorname{End}' E \to \operatorname{End} E \to \operatorname{End}'' E \to 0. \tag{15.9}$$

Clearly $\operatorname{End}'' E = \operatorname{Hom}(F(-m), L)$. By (15.6) H^1 of this vanishes when $F = I_{n-1}$ and so it will vanish for all F_r with $r \in R_2 \subset R_1$, some new Zariski neighbourhood of r_0 in R_1. This then implies, from (15.9), that

$$H^1(M, \operatorname{End}' E) \to H^1(M, \operatorname{End} E) \tag{15.10}$$

is surjective.

On the other hand we have the exact sequence

$$0 \to L^* \otimes F(-m) \to \operatorname{End}' E \to \operatorname{End} F \oplus \mathscr{O} \to 0,$$

which gives the exact cohomology sequence

$$\to H^1(M, L^* \otimes F(-m)) \to H^1(M, \mathrm{End}' E) \to H^1(M, \mathrm{End}\, F) \oplus H^1(M, \mathcal{O}) \to 0.$$

If we now compute the infinitesimal deformation map for our family S at the point s_0, and recall that S is fibred over $R_2 \times J$ with fibre $H^1(M, L^* \otimes F(-m))$ we see that the surjectivity for S follows from that of R_2 and J. For R_2 this is our induction assumption and for J it is of course classical. Together with the surjectivity of (15.10) this completes the proof of lemma 15.5.

We have had the benefit of discussions with many colleagues on the topics in this paper and we are in particular grateful to N. Ekedahl, G. Harder, N. J. Hitchin, D. Mumford, M. S. Narasimhan, T. R. Ramadas and J.-P. Serre for helpful observations.

This work was supported by N.S.F. under contract MCS–8006104.

References

Adams, J. F. 1969 *Lectures on Lie groups*. New York: Benjamin.

Atiyah, M. F. 1955 Complex fibre bundles and ruled surfaces. *Proc. Lond. math. Soc.* **5**, 407–434.

Atiyah, M. F. 1957 Vector bundles over an elliptic curve. *Proc. Lond. math. Soc.* **7**, 414–452.

Atiyah, M. F. 1979 Geometry of Yang–Mills fields. *Fermi lectures*, Scuola Normale Pisa.

Atiyah, M. F. 1982 Convexity and commuting Hamiltonians. *Bull. Lond. math. Soc.* **14**, 1–15.

Atiyah, M. F. & Bott, R. 1980 Yang–Mills and bundles over algebraic curves. *Geometry and analysis, Patodi memorial volume of Indian Academy of Sciences.*

Atiyah, M. F., Hitchin, N. J. & Singer, I. M. 1978 Self-duality in four-dimensional Riemannian geometry. *Proc. R. Soc. Lond.* A **362**, 425–461.

Bott, R. 1958 The space of loops on a Lie group. *Mich. Math. J.* **5**, 35–61.

Bott, R. & Samelson, H. 1958 Applications of the theory of Morse to symmetric spaces. *Am. J. Math.* **80**, 964–1029.

Bourbaki, N. 1968 *Elements de mathematique, groupes et algèbres de Lie*, ch. 4–6. Paris: Hermann.

Desale, U. V. & Ramanan, S. 1975 Poincaré polynomials of the variety of stable bundles. *Math. Annln* **216**, 233–244.

Donaldson, S. K. 1983 A new proof of a theorem of Narasimhan and Seshadri. *J. diffl Geom.* (To appear.)

Frankel, T. 1959 Fixed points on Kähler manifolds. *Ann. Math.* **70**, 1–8.

Gieseker, D. 1982 A degeneration of the moduli space of stable bundles. (To appear.)

Griffiths, P. & Harris, J. 1978 *Principles of algebraic geometry*. New York: Wiley.

Grothendieck, A. 1957 Sur la classification des fibrés holomorphes sur la sphère de Riemann. *Am. J. Math.* **79**, 121–138.

Harder, G. 1970 Eine Bemerkung zu einer Arbeit von P. E. Newstead. *J. Math.* **242**, 16–25.

Harder, G. & Narasimhan, M. S. 1975 On the cohomology groups of moduli spaces of vector bundles over curves. *Math. Annln* **212**, 215–248.

Horn, A. 1954 Doubly stochastic matrices and the diagonal of a rotation matrix. *Am. J. Math.* **76**, 620–630.

Kirwan, F. 1982 Cohomology of quotient spaces *C. R. Acad. Sci. Paris* **295–1**, 261–264.

Kostant, B. 1973 On convexity, the Weyl group and the Iwasawa decomposition. *Annls scient Éc. norm. sup., Paris* **6**, 413–455.

Marshall, A. W. & Olkin, I. 1979 *Inequalities: theory of majorization and its applications*. New York: Academic Press.

Mitter, P. K. & Viallet, C. M. 1981 On the bundle of connections and the gauge orbit manifold in Yang–Mills theory. *Communs math. Phys.* **79**, 457–472.

Mumford, D. 1965 *Geometric invariant theory*. Berlin: Springer-Verlag.

Narasimhan, M. S. & Ramadas, T. R. 1979 Geometry of SU(2) gauge fields. *Communs math. Phys.* **67**, 121–136.

Narasimhan, M. S. & Ramanan, S. 1969 Moduli of vector bundles on compact Riemann surface. *Ann. Math.* **89**, 1201–1208.

Narasimhan, M. S. & Seshadri, C. S. 1965 Stable and unitary vector bundles on a compact Riemann surface *Ann. Math.* **82**, 540–567.

Newstead, P. E. 1972 Characteristic classes of stable bundles over an algebraic curve. *Trans. Am. Math. Soc.* **169**, 337–345.

Palais, R. 1965 *Foundations of global non-linear analysis*. New York: Benjamin.

Pressley, A. 1980 Decompositions of the space of loops on a Lie group. *Topology* **19**, 65–80.

Ramanan, S. 1973 The moduli spaces of vector bundles over an algebraic curve. *Math. Annln* **200**, 69–84.

Ramanathan, A. 1975 Stable principal bundles on a compact Riemann surface. *Math. Annln* **213**, 129–152.

Seshadri, C. S. 1967 Space of unitary vector bundles on a compact Riemann surface. *Ann. Math.* **85**, 303–336.
Shatz, S. S. 1977 The decomposition and specialization of algebraic families of vector bundles. *Compositio math.* **35**, 163–187.
Singer, I. M. 1959 The geometric interpretation of a special connection. *Pacif. J. Math.* **9**, 585-590.
Singer, I. M. 1978 Some remarks on the Gribov ambiguity. *Communs math. Phys.* **60**, 7–12.
Uhlenbeck, K. 1982 Connections with L^p bounds on curvature. *Communs math. Phys.* **83**, 31–42.
Weil, A. 1938 Généralisation des fonctions abéliennes. *J. Math. pures appl.* **17**, 47–87.