SU(3) GENERALIZATIONS OF THE CASSON INVARIANT FROM GAUGE THEORY

CHRISTOPHER HERALD

Department of Mathematics, University of Nevada
Reno NV 89557-0045, USA

Abstract. This paper is a survey of some recent joint work of Hans Boden, Paul Kirk and the author, as well as work by Cappell, Lee, and Miller, on generalizing the Casson invariant to the group SU(3). The main challenge here is that in this setting there are nontrivial reducible representations. Because of this, the irreducible stratum is not compact, and as a consequence an algebraic count of points does not provide a topological invariant (independent of perturbation).

1. Introduction

In the late 1980’s, Andrew Casson defined a new invariant of closed, oriented 3-manifolds $X$ which have the same homology as $S^3$. Roughly speaking, the Casson invariant $\lambda_{SU(2)}(X)$ is an algebraic count of the conjugacy classes of irreducible $SU(2)$ representations $\rho: \pi_1 X \to SU(2)$. The reason for restricting to $SU(2)$ and imposing this homology restriction on the 3-manifold $X$ is that reducible representations $\rho: \pi_1 X \to SU(2)$ are necessarily abelian and hence factor through the homology $H_1(X)$. For a homology 3-sphere, this homology group is trivial and so the only reducible representation is the trivial one. For cohomological reasons, the trivial representation is isolated, which implies that the irreducible portion of the set of representations modulo conjugation is compact.

We shall describe Casson’s method for defining the invariant in more detail in the next section. For the moment, suffice it to say that Casson devised a means of perturbing to achieve a 0-dimensional set of points to count, in case the representation variety was not already a 0-dimensional manifold. The
compactness of the irreducible stratum was then the key to showing that the
signed count of points was independent of perturbation.

This paper is a survey of some recent joint work of Hans Boden, Paul Kirk
and the author, as well as work by Cappell, Lee, and Miller, on generalizing
the Casson invariant to the group \( SU(3) \). The main challenge here is that in
this setting there are nontrivial reducible representations. Because of this, the
irreducible stratum is not compact, and as a consequence an algebraic count of
points does not provide a topological invariant (independent of perturbation).

The first work in this direction was a generalization of the \( SU(2) \) invariant
to rational homology spheres (i.e. manifolds \( X \) with \( H_*(X; \mathbb{Q}) = H_*(S^3; \mathbb{Q}) \))
by Walker [10]. On rational homology spheres, there are nontrivial abelian
representations and a count of the irreducible \( SU(2) \) representations depends
on the perturbation. Walker devised a correction term involving the abelian
representations that compensated for this dependence.

The \( SU(3) \) generalizations discusses in this paper also involve correction terms
involving reducible representations. These generalizations follow a different
approach from Casson’s original description of \( \lambda_{SU(2)} \), viewing \( \lambda_{SU(2)} \) in terms
of connections and gauge theory (following Taubes [9]) rather than representa-
tions. In this gauge theory context, we will describe the various issues that
complicate the \( SU(3) \) setting and discuss several alternatives for dealing with
them. Finally, we summarize the properties of the different generalized invari-
ants. Before we do so, though, we briefly explain Casson’s approach to the
transversality and perturbation issues.

2. Transversality in the framework of Casson and Walker

For general homology spheres \( X \), the \( SU(2) \) representation variety

\[
R(X) = \text{Hom}(\pi_1 X, SU(2))/\text{conjugation}
\]

is not necessarily a finite collection of points. Casson took care of this problem
in the following way. Choose a Heegard decomposition of \( X \), i.e., a closed
surface \( F \subset X \) which divides \( X \) into two solid handlebodies \( X = X_1 \cup_F X_2 \).
Then the Seifert van Kampen theorem indicates how to calculate \( \pi_1 X \) from
the fundamental groups of the pieces, and suggests that representations of \( \pi_1 X \)
be viewed as intersections of \( \text{Hom}(\pi_1 X_1, SU(2)) \) and \( \text{Hom}(\pi_1 X_2, SU(2)) \)
inside \( \text{Hom}(\pi_1 F, SU(2)) \). (More precisely, they are viewed as points in
\( \text{Hom}(\pi_1 X_1, SU(2)) \times \text{Hom}(\pi_1 X_2, SU(2)) \) which map to the same point in
\( \text{Hom}(\pi_1 F, SU(2)) \).) Dividing by the conjugation action of \( SU(2) \) on the rep-
resentation spaces, we obtain an intersection interpretation of \( R(X) \).

The quotient spaces \( R(X_1) \) and \( R(F) \) have singularities at the reducible re-
presentations (where the stabilizer is larger than the center \( \mathbb{Z}_2 \subset SU(2) \)), but
and describe three types of correction terms that have been used to define generalized Casson invariants.

3. The Casson Invariant and Gauge Theory

Throughout the remainder of the paper, we assume that $X$ is an oriented homology 3-sphere. We now recall the gauge theoretic description of the Casson invariant provided by Taubes [9]. Let $P_n = X \times SU(n)$ be the trivial principal bundle. Let $\mathcal{A}_n$ denote the space of connections on this bundle. To get a Banach manifold, one must complete the space of smooth connections with respect to an appropriate Sobolev norm. The group $\mathcal{G}_n$ of bundle automorphisms of $P_n$ (also suitably completed) is called the gauge group. The gauge group acts on $\mathcal{A}_n$, with the action defined by $d_{g \cdot A}s = g \circ d_A \circ g^{-1}s$ for any section $s : X \to P_n$.

Since $P_n$ is trivial, we can fix a trivialization and let $\theta$ denote the canonical trivial connection. Any other connection differs from $\theta$ by a $su(n)$ valued 1-form. We will not distinguish in the notation between the connection and its connection 1-form. The group action on 1-forms takes the form

$$g \cdot A = gAg^{-1} - dg^{-1}.$$ 

Fixing a basepoint $x_0 \in X$, any connection then associates to each loop in $X$ based at $x_0$ a holonomy element in $SU(n)$, obtained by parallel translating point in the fiber above the basepoint around the loop. A connection is called flat if its holonomy depends only on the homotopy class of the loop. A flat connection, then, gives a representation of the fundamental group to $SU(n)$, and flat connections modulo gauge transformation correspond bijectively to representations modulo conjugation.

The Chern-Simons function $CS : \mathcal{A}_n \to \mathbb{R}$, given by

$$CS(A) = \frac{1}{8\pi^2} \int_X \text{trace} \left( dA \wedge A + \frac{2}{3} A \wedge A \wedge A \right)$$

is a smooth function whose critical points are exactly the flat connections. Thus, the Casson invariant is roughly an algebraic count of critical points of the Chern-Simons function on connections modulo gauge. This suggests an analogy with the following interpretation of the Euler characteristic in finite dimensions.

Let $M$ be a smooth, finite dimensional manifold, and let $f : M \to \mathbb{R}$ be a function which is Morse. This means that (if we choose a Riemannian metric on $M$) the gradient vector field of $f$ is transverse to the zero vector field. For a critical point $p \in M$ of $f$ (i.e. a zero of the gradient vector field), we define
the Morse index to be the number of negative eigenvalues of the Hessian of $f$ at $p$. Then the Poincaré–Hopf theorem states that the number of critical points, counted with sign determined by the Morse index, equals the Euler characteristic $\chi(M)$.

To make a similar count of critical points of the Chern–Simons function one must deal with several issues. The first issue is that $CS(g\cdot A) = CS(A) + \text{deg } g$, where $\text{deg} : \pi_0 G_n \rightarrow \mathbb{Z}$ is an isomorphism. If we view the Chern–Simons function as an $S^1 = \mathbb{R}/\mathbb{Z}$ valued function on $\mathcal{A}_n/G_n$, this is not a serious problem. The second issue concerns the signs of the points.

Interpreted appropriately, the Hessian of the Chern–Simons function on $\mathcal{A}_n/G_n$ at a critical point $[A]$ gives an essentially self adjoint Fredholm operator. It has infinitely many positive and negative eigenvalues, so the traditional Morse index is infinite.

Taubes showed that one could replace the Morse index with a relative Morse index between the critical point $A$ and a fixed base point. For the base point he chose the trivial connection $\theta$. The relative Morse index, then, is the number of eigenvalues of the Hessian which cross zero, counted with sign depending on which way they cross, along a path connecting $\theta$ to $A$. We call this number the spectral flow and denote it $SF(\theta, A)$. If a different gauge representative for the same orbit $[A] \in \mathcal{A}_n/G_n$ is chosen, this spectral flow changes by a multiple of eight, so the parity of the spectral flow is independent of the choice.

To achieve transversality in this Morse theory framework, we perturb the Chern–Simons function by adding a suitable gauge invariant function $h : \mathcal{A}_n \rightarrow \mathbb{R}$ to it to obtain a function with isolated critical points. The flatness equation (curvature equals zero) on $\mathcal{A}_n/G_n$ which cuts out the critical set of the Chern–Simons function is a Fredholm map with index zero. Taubes and Floer described a class of functions $h$ which did not destroy this property. Now consider the space $\mathcal{A}_2/G_2$ of $SU(2)$ connections modulo gauge. The main result in [9] states that for a generic small perturbation $h$,

$$\lambda_{SU(2)}(X) = \sum_{[A] \in \mathcal{M}_h} (-1)^{SF(\theta, A)}.$$

4. Reducible Connections, Singularities and Bifurcations

As explained in the previous section, Taubes interpreted the Casson invariant as a sort of infinite dimensional Morse theoretic Euler characteristic for the space of $SU(2)$ connections modulo gauge. This quotient space has singularities, but (except for the orbit of the trivial connection, which is isolated from other critical orbits) the critical set of the Morse function avoids the singularities and so they don’t cause great difficulties.
In the setting of $SU(3)$ gauge theory, reducible flat connections need not be abelian (and hence on homology spheres need not be trivial). The non-abelian reducible flat connections have holonomies contained in the subgroup $S(U(2) \times U(1)) \subset SU(3)$ (or a conjugate subgroup) and stabilizer given by the 1-dimensional torus consisting of diagonal matrices with eigenvalues $\lambda, \lambda, \lambda^{-2}$ for $\lambda \in U(1)$. Because this stabilizer is larger than the stabilizer of an irreducible $SU(3)$ connection (the latter is isomorphic to $\mathbb{Z}_3$, the center of $SU(3)$), the orbits of these reducible connections form a singular stratum in the quotient $\mathcal{A}_3/\mathcal{G}_3$.

The existence of flat reducible $SU(3)$ connections means that the singular stratum is not disjoint from the critical set of the Chern–Simons function. Indeed the flat $SU(2)$ connections considered by Taubes may be viewed as $S(U(2) \times U(1))$ flat connections. Before proceeding to describe the methods for extending the Casson invariant to the $SU(3)$ setting, we now illustrate the difficulties involved in defining a Morse theoretic Euler characteristic for a manifold with similar quotient singularities.

Since the singularities involved in $SU(3)$ gauge theory involve orbits with stabilizer given by $U(1)$, we use an example of a finite dimensional manifold $M$ with a semifree $U(1)$ action. Let $L$ denote the set of fixed points of the group action. If $f : M \to \mathbb{R}$ is an $U(1)$ invariant smooth function, then its critical set consists of a collection of fixed points together with a union of circle orbits of critical points in $M - L$. Fix an invariant Riemannian metric on $M$ so that we can identify the critical set of $f$ with the zero set of the gradient vector field $\nabla f$. Because $U(1)$ acts without nonzero fixed vectors on the normal bundle to $L$, if $p \in L$ then $\nabla f(p) \in T_p L$. It follows that the critical fixed points are exactly the critical points of $f|_L$.

A well-known theorem in Morse theory (on manifolds without group actions) states that for a generic path of functions connecting two Morse functions, the only topological changes in the critical set are births and deaths of pairs of critical points of Morse indices differing by one. This makes it clear that the signed count of critical points is independent of the choice, for the algebraic contribution from such a pair will be zero.

We may view $M/U(1)$ as a union of two manifolds, $(M - L)/U(1)$ and $L/U(1) \cong L$. The singular nature of the quotient space has to do with normal bundle structure of the second in the first. A theorem of Wasserman [11] asserts in this setting that a generic invariant function on $M$ will give a Morse function with finite critical set on each of these strata. Since $(M - L)/U(1)$ is not compact, finiteness here is nontrivial; what is being asserted is that at each critical point in $L$ the function is nondegenerate not only in the tangent direction to $L$ but also in the normal direction, and hence critical orbits in
\((M - L)/U(1)\) cannot limit to it.

For a generic path connecting two such functions, three types of bifurcations can occur in the critical set. The first two correspond to standard Morse births and deaths of cancelling pairs of critical points in \(L\), and in the (smooth) quotient space \((M - L)/U(1)\). In other words, a pair of critical points in \(L\) of Morse index difference one can be born or can annihilate one another, and similarly for a pair of circle orbits of critical orbits of index difference one. The third type of bifurcation involves a critical circle orbit popping out of a critical fixed point (or this process in reverse). That is to say, there is a critical point in \(L\), with no other critical points nearby for \(t \leq t_0\), and then for \(t > t_0\) there is in addition a circle orbit of critical points nearby. Figure 1 illustrates the topology of the critical set (in a neighborhood nearby). This phenomenon can be illustrated with the path of functions \(f_t : C \rightarrow \mathbb{R}, f_t(z) = -t|z|^2 + |z|^4\).

![Figure 1. A local picture of the parameterized critical set, drawn as a fibration over a small interval around \(t_0\).

For \(t \leq t_0\), the critical set consists of a fixed point. Beginning at \(t = t_0\), there is a new circle orbit of critical points in addition to the fixed point.]

As a consequence, for a generic path of functions, the union of the critical sets divided by the group \(U(1)\) has the form of a 1-dimensional cobordism of circle orbits together with a 1-dimensional cobordism of fixed point orbits. Both cobordisms are compact except for "T"-intersections where an end of the former limits to a point on the latter (see Fig. 2).

To obtain an analogue of the Euler characteristic, in this \(U(1)\) manifold setting, we must modify the usual signed count of critical points (or, here, critical orbits) to obtain a formula which is invariant under all three types of bifurcations. The crucial observation is that in the third type of bifurcation described above, simultaneous to the birth of a critical circle, there is a change in the normal Morse index of the function at the critical fixed point. For this reason, we refine our notion of Morse index at the critical points in \(L\). The fact that the Hessian of an invariant function at a critical point \(p \in L\) is invariant under the (linearized) group action on the tangent space \(T_pM\) implies that the Hessian decomposes
into a direct sum of operators on the two summands $T_p L \oplus N_p L \cong T_p M$. We define the tangential and normal components of the Morse index to be the number of negative eigenvalues of these summands, and denote them by $\mu_T(p)$ and $\mu_N(p)$. Since the normal summand commutes with the $U(1)$ action, it is a Hermitian bilinear form with respect to the compatible complex structure on the normal bundle fiber $N_p L$. Hence its eigenspaces are all complex subspaces and $\mu_N(p)$ is always even.

![Figure 2. A cobordism with "T"-intersections](image)

The dashed curves represent the varying critical set in $L$, and the solid curves represent the varying critical set (modulo $U(1)$) in $M - L$.

A close examination of the relationship between the Morse indices of the critical points involved in the bifurcations demonstrates that the quantity

$$
\sum_{q \in ((M-L) \cap crit(f))/U(1)} (-1)^{\mu(q)} - \sum_{p \in L \cap crit(f)} (-1)^{\mu_T(p)} \left( \frac{\mu_N(p)}{2} \right)
$$

is invariant under all three bifurcations and hence is independent of the invariant Morse function. In fact, one can see by choosing the function appropriately that this invariant of the $U(1)$ manifold $M$ is the relative Euler characteristic of the quotient space relative to the singular set (i.e. the Euler characteristic of the relative cohomology).

5. **Correction Terms in the $SU(3)$ Setting**

In this section, we consider the $SU(3)$ gauge theory setting (and we’ll drop the 3 subscripts). There is a completely analogous classification of the types of bifurcations that occur the critical set for a generic one parameter family of admissible perturbations of the Chern–Simons function. (Here, admissibility involves some technical constraints, mainly that the perturbation does not alter the Fredholm properties of the equation $\nabla CS(A) = 0$ defining the critical
set). That is to say, the bifurcations involve standard Morse births/deaths in the irreducible stratum of \( \mathcal{A}/\mathcal{G} \), standard Morse births/deaths in the reducible (i.e. \( S(U(2) \times U(1)) \)) stratum of \( \mathcal{A}/\mathcal{G} \), and irreducible orbits of critical points popping out of reducible critical points.

Motivated by the finite dimensional model discussed in the previous section and Taubes’ gauge theoretic description of the \( SU(2) \) Casson invariant, it is natural to try to define an \( SU(3) \) invariant by simply replacing the Morse indices in (1) with the spectral flow analogues. Indeed, the decomposition of Morse index \( \mu(p) = \mu_T(p) + \mu_N(p) \) at critical points with circle stabilizer has a spectral flow analogue, which we explain briefly.

The tangent space of the space \( A \) of \( SU(3) \) connections may be identified with the space of differential 1-forms on \( X \) with values in the Lie algebra \( su(3) \). After a gauge transformation, each connection \( A \in \mathcal{A} \) which is reducible may be viewed as an \( S(U(2) \times U(1)) \) connection. Under the conjugation action of \( S(U(2) \times U(1)) \) on \( su(3) \), \( su(3) \) decomposes into \( s(u(2) \times u(1)) \), the Lie algebra, and a 4-dimensional subspace isomorphic to \( C^2 \). On the latter factor, \( B \oplus B' \in S(U(2) \times U(1)) \) acts on \( C^2 \) by multiplication by the \( U(2) \) matrix \( (\det(B))^{1/2}B \).

Any reducible connection can be connected to the trivial connection by a path of (possibly abelian or trivial) \( S(U(2) \times U(1)) \) connections, and so one can make a similar decomposition of the spectral flow of the Chern–Simons function to \( s(u(2) \times u(1)) \) and \( C^2 \) components, which we denote by \( SF_T \) and \( SF_N \). Furthermore, it is again true that the third type of bifurcation, where a new irreducible critical point pops out of a reducible one, coincides with a (multiplicity two) eigenvalue in the normal Hessian crossing zero.

If we simply replace the Morse indices in formula (1) with the corresponding spectral flows, we obtain

\[
\sum_{[A] \in \text{cent}(cs+h)} (-1)^{SF_A(\theta,A)} - \sum_{[B] \in \text{cent}(cs+h)} (-1)^{SF_B(\theta,B)} \left( \frac{SF_N(\theta, B)}{2} \right)
\]

(2)

where the first sum is over the orbits of irreducible critical points and the second sum is over the orbits of reducible (nontrivial) critical points. But recall that spectral flow is not well-defined on gauge orbits. For the exponents of \(-1\), this is not a problem, since \( SF_{su(3)}(\theta, gA) = SF_{su(3)}(\theta, A) + 12 \deg g \) and \( SF_T(\theta, gA) = SF_T(\theta, A) + 8 \deg g \) where \( \deg g \in \pi_0 \mathcal{G} \cong \mathbb{Z} \). The gauge ambiguity of the normal spectral flow, however, makes formula (2) dependent on the gauge representatives chosen for the reducible perturbed flat orbits.

Several methods have been devised to overcome this problem. In the first method, used in [2], the correction term (i.e. the second sum) was modified (roughly) by subtracting from the normal spectral flow an appropriate multiple of the Chern–Simons value at the reducible connection. Since both the normal
spectral flow and Chern–Simons change by multiples of the degree of a gauge transformation, this difference is independent of gauge representative. The resulting invariant, denoted by \( \lambda_{SU(3)} \), was shown in [3] to satisfy the relatively simple connected sum formula

\[
\lambda_{SU(3)}(X_1 \# X_2) = \lambda_{SU(3)}(X_1) + \lambda_{SU(3)}(X_2) + 4\lambda_{SU(2)}(X_1)\lambda_{SU(2)}(X_2).
\] 

It follows that \( \lambda_{SU(3)} - 2\lambda_{SU(2)}^2 \) is additive.

The behavior of \( \lambda_{SU(3)} \) under Dehn surgery appears to be very complicated. Calculations were made for surgeries on \((2, q)\)-torus knots, and the values of the invariant in these cases are rational functions of the surgery coefficient. In general, however, it is not known whether the Chern–Simons values that enter into the correction term are rational. Hence it is unclear whether \( \lambda_{SU(2)} \) is always rational valued.

More recently, two alternatives to the correction term in [2] have been found, independently and more or less simultaneously. In [6], Cappell, Lee and Miller noted that since both the tangential and normal components of the spectral flow were gauge dependent, an appropriate linear combination of these could be used to produce a gauge invariant quantity with which to replace the normal Morse index. This linear combination causes another problem, though. A correction term of this form fails to be invariant under the simpler bifurcation wherein two new reducible perturbed flat connections are born. Thus in [6] a tertiary correction term was added, involving the rank of the Floer instanton homology complex boundary operator. This was based on another clever idea, essentially that the Floer homology is independent of the perturbation, so if two new reducibles are born they must cancel in Floer homology, and thus the rank of the boundary operator must also have changed. The behavior of the Cappell, Lee, Miller invariant under connected sum is similar to that of \( \lambda_{SU(3)} \) except for this tertiary term, which on its face seems to be much more complicated (since the Floer homology of connected sums is quite subtle). Calculations of this invariant have again been made for the surgeries on the \((2, q)\)-torus knots. The third alternative was described in [5] and is based on the idea of using a gauge dependent choice of base point in place of \( \theta \) for each normal spectral flow. The space \( \mathcal{A} \) is contractible, so \( \pi_1 \mathcal{A}/\mathcal{G} \cong \pi_0 \mathcal{G} \cong \mathbb{Z} \). Let \( \mathcal{G}_0 \) denote the identity component of \( \mathcal{G} \). Then

\[
\mathcal{A}/\mathcal{G}_0 \to \mathcal{A}/\mathcal{G}
\]

is a nontrivial \( \mathbb{Z} \) covering which is classified by the Chern–Simons function. On the other hand, the flat moduli space (the critical set of \( cs \) modulo gauge) is compact, and hence has a finite collection of components, and the Chern–Simons function is constant along each of these components. Therefore, normal
spectral flow is well-defined on paths in $\mathcal{A}/\mathcal{G}$ which stay in a small neighborhood of a component. Let $X$ be a homology sphere and index the components of the reducible flat moduli space $C_1, \ldots, C_n$. Pick a base point $[B_i] \in C_i$ in each component (we’ll come back shortly to the question of how to make this selection). For small perturbations, the perturbed flat moduli space (the critical set of $cs + h$ modulo gauge) is close to the unperturbed flat moduli space, and so there is an unambiguous notion of normal spectral flow from the basepoint in the nearby component.

Now consider the question of making a canonical choice of $[B_i] \in C_i$. Since each component $C_i$ is compact, one can show that there are basepoints $[B_i^{\pm}]$ (not uniquely determined) which give well-defined maximal and minimal normal spectral flow to the other points near the component. From consideration of how the primary term (the count of irreducibles) behaves with respect to change of orientation, it is natural to hope for an $SU(3)$ Casson invariant to be independent of orientation, and in fact with suitable normalizations the other two invariants discussed above have this property. In the present approach, to obtain an invariant which is orientation independent, we must average the results of the maximal and minimal choices. That is, for the correction term we sum the following quantity over orbits $[B]$ of reducible critical points of the perturbed Chern–Simons function

$$(-1)^{SF_T(\theta,B)} \left( \frac{SF_N(B_i^+, B) + SF_N(B_i^-, B)}{4} \right).$$

Although it is not obvious from what has been presented here, this correction term gives an integer valued invariant, which we denote by $\tau_{SU(3)}$, not a half-integer. By construction, $\tau_{SU(3)}(-X) = \tau_{SU(3)}(X)$. Furthermore, $\tau_{SU(3)}$ satisfies the same connected sum formula (3) as $\lambda_{SU(3)}$. See [5] for details.

The invariant $\tau_{SU(3)}$ enjoys one more pleasant property. Suppose $X$ is a homology 3-sphere $X$ and for each reducible representation $\rho: X \to SU(2)$, the cohomology of $X$ with twisted $\mathbb{C}^2$ coefficients in the flat bundle determined by $\rho$ is trivial. Then the correction term vanishes for this manifold. In other words, the correction term is only nonzero in situations where the correction term is truly necessary because the signed count of irreducible orbits of critical points is perturbation dependent.

Finally, we note that from calculations currently in progress, $\tau_{SU(3)}$ on $\frac{1}{n}$ surgeries on $(p,q)$-torus knots is polynomial in the surgery coefficient $n$. This appears to offer the best hope yet of an $SU(3)$ invariant with a reasonable Dehn surgery formula.
Acknowledgements

The author gratefully acknowledges partial support from the National Science Foundation (DMS-0072564), an American Mathematical Society Centennial Research Fellowship, and a Junior Faculty Research Award from the University of Nevada, Reno.

References


