EQUIVARIANT SOBOLEV THEOREMS AND YANG-MILLS-HIGGS FIELDS

*Thomas H. Parker**

1. Introduction

Many of the most important non-linear P.D.E.'s that occur in geometry and physics—the Yamabe equation, the harmonic map equation, the Yang-Mills and Yang-Mills-Higgs equations—can be cast as variational problems. In each case there is a Lagrangian of the general form

$$L(\phi) = \int |\nabla\phi|^2 + |\phi|^p$$

whose critical points are the solutions of the given P.D.E. There is a standard approach to finding such critical points. One first shows that L defines a smooth function on an appropriate Sobolev space and that the downward gradient flow lines exist. The limit of these flow lines should be a critical point of L. The usual proof that these flow lines converge uses a compactness condition—the Palais-Smale Condition (PS)—that depends on the compactness of a certain Sobolev embedding. Unfortunately, for each of the problems mentioned above the required Sobolev embedding is not compact, and the direct minimization method fails.

In this paper we develop a simple way of avoiding this difficulty for manifolds with symmetry. When a compact Lie group H acts isometrically on a compact Riemannian manifold M, one can define Sobolev spaces of invariant functions (and more generally, equivariant sections of a vector bundle). Our main result—the "Equivariant Sobolev Embedding Theorem"—states that if all the orbits of H have dimension $d \geq 1$ then the Sobolev spaces of

* Supported in part by the NSF.

invariant functions satisfy Sobolev inequalities as if they were on a manifold of dimension $n - d$ (even when the orbit space is not a manifold). Thus, restricting to the invariant functions gives improved Sobolev embeddings. It is then easy, for at least several of the problems above, to verify the Palais-Smale Condition and directly minimize the Lagrangian on the space of invariant functions. This yields points that are critical with respect to invariant perturbations. By the "symmetric criticality principle" (see section 3) these are, in fact, critical with respect to all perturbations, and hence are solutions of the variational P.D.E.

This article describes this method in detail. We begin in section 1 by reviewing the geometric approach to global variational problems and describing the difficulties that arise because of borderline Sobolev embeddings. In section 2 we state and prove the Equivariant Sobolev Embedding Theorem, and in section 3 we describe our scheme for approaching variational problems with symmetry. As an initial application we prove in section 4 that the Yamabe equation has infinitely many solutions on a compact Riemannian manifold with symmetry.

In the remaining sections we apply the equivariant Morse theory to find Yang-Mills-Higgs fields on compact Riemannian 4-manifolds M. In variational form, YMH fields are the critical points of the Lagrangian

$$L(\nabla, \phi) = \frac{1}{2} \int_M |F^\nabla|^2 + |\nabla\phi|^2 + \frac{\lambda}{2}(|\phi|^2 - \mu)^2 \, dv$$

where ∇ is a connection on a bundle over M with curvature F^∇, ϕ is a section of an associated vector bundle, and $\lambda \neq 0$ and $\mu > 0$ are "coupling constants". These fields arose in physics: it is the YMH equations—rather than the plain Yang-Mills equations—that are at the core of the theory of weak interactions, and the quantum tunneling amplitudes of the theory are expressed in terms of YMH fields on $\mathbb{R}^4$ (with its *euclidean* metric). On the mathematical side, the analytic properties of YMH fields were studied more than a decade ago ([JT], [Pa1]). However, the existence of YMH fields was entirely unknown on any Riemannian 4-manifold until very recently, when existence was established on S^4 using the Equivariant Morse Theory ([Pa2]). Our aim here is to extend that first existence result.

We begin in section 5 by considering the critical points of the YMH Lagrangian with the connection fixed. There are two cases to consider, depending on the choice of sign of the coupling constant λ. The qualitative behavior of the Lagrangian is very different in the two cases, but both are amenable to the above approach—after imposing symmetry they satisfy the Palais-Smale Condition and a Morse theory. Two concrete examples are given in section 6. The first shows how the Palais-Smale Condition can hold for invariant functions, yet fail for non-invariant functions. The second (Corollary 6.5) describes an interesting bifurcation phenomenon for Higgs fields on the four-sphere: if one fixes the constant $\lambda\mu > 0$ in the Lagrangian and varies the Riemannian metric g to $R^2 g$, then as R increases (that is, as the radius of the sphere increases) there are more and more critical points. Alternatively, one can fix the metric and obtain more and more critical points by increasing $\lambda\mu$.

Finally, in section 7 we turn to the full gauge theory by incorporating the gauge group and considering the coupled YMH equations (see equation (5.2)). We review the Morse Theorem for equivariant connections proved in an earlier paper [Pa2] and briefly describe how it yields a proof of the existence of non-self-dual Yang-Mills fields on S^4. We then combine this with the results of sections 5 and 6 to prove the main results of this paper. These show that, with either choice of sign for λ, there exist YMH fields on S^4. Specifically, section 7 contains the details of the following theorem.

THEOREM. Let P be a principal $SU(2)$ bundle over S^4 and let E be a non-trivial associated bundle. Then

(a) for $\lambda < 0$ there are infinitely many distinct solutions of the YMH equations on (P, E) which are neither uncoupled or reducible,
(b) for $\lambda > 0$ and $\lambda\mu$ sufficiently large, there is a solution of the YMH equations on (P, E) which is neither uncoupled or reducible.

The results of section 6 suggest an improved version of statement (b): the bifurcation phenomenon mentioned above for the uncoupled equations should continue to hold for the coupled equations. Thus one expects that as the constant $\lambda\mu > 0$ increases there are more and more solutions to the coupled YMH equations.

The idea of seeking symmetric solutions to a P.D.E. is of course a very old and standard procedure. Typically, one imposes invariance under a group action whose principal orbits have codimension 1, such as the rotational action of $SO(n)$ on $\mathbb{R}^n$ or S^{n-1}. This reduces the P.D.E. to an O.D.E. One must then solve the O.D.E. with the appropriate boundary conditions—boundary conditions arising from the requirement that the solution extend smoothly over the lower-dimensional orbits. The approach taken here is different. Instead of rewriting the P.D.E. as an equation on orbit space we work directly with invariant functions on the original manifold. The singularities of the orbit space never come into play and the analysis is considerably simplified.

The Equivariant Sobolev Theorem was inspired by a theorem of WeiYue Ding [D]. We are indebted to Richard Palais, who developed the foundations of global non-linear analysis, and who has continually espoused the geometric viewpoint taken here.

1. The Global Variational Method

In this section we briefly outline the global analytic approach to geometric variational problems, and explain why the method does not work for problems involving the "borderline case" of the Sobolev inequalities. This material is standard; details can be found in the books of T. Aubin [A], M. Berger [B], and R. Palais [P].

Let (M, g) be a compact n-dimensional Riemannian manifold, and let E be a vector bundle over M with a fixed fiber metric and compatible connection ∇. The space $\Gamma(E)$ of smooth sections of E can be made into a Banach space as follows.

Definition. For $k \geq 0$, $p \geq 1$ the Sobolev space $L^{k,p}(E)$ is the completion of $\Gamma(E)$ with respect to the norm

$$\|\phi\|_{k,p} = \left(\sum_{l=0}^{k} \int_M |\nabla^l \phi|^p \right)^{1/p} \tag{1.1}$$

where $\nabla^l \phi = \nabla \circ \nabla \circ \cdots \circ \nabla \phi \in \Gamma(\overset{l}{\otimes} T^*M \otimes E)$ is the l^{th} covariant derivative of ϕ. The Holder spaces $C^{k,\alpha}(E)$ are defined similarly using the norm

$$\|\phi\|_{k,\alpha} = \sum_{l=0}^{k} \sup |\nabla^l \phi| + \sup_{x,y} \frac{|\nabla^k \phi(x) - \nabla^k \phi(y)|}{|\operatorname{dist}(x, y)|^\alpha}, \tag{1.2}$$

where the last supremum is over all $y \neq x$ contained in a normal coordinate neighborhood of x, and $\nabla^l \phi(y)$ is taken to mean the tensor at x obtained by parallel transport along the geodesic from x to y. The relationships between these spaces are summarized in the following fundamental theorem.

SOBOLEV EMBEDDING THEOREM. Let $1 \leq p < q$, $k \geq l$, and $0 < \alpha < 1$. Then

(i) for $k - \frac{n}{p} \geq l - \frac{n}{q}$ the identity map induces a continuous inclusion $L^{k,p}(E) \hookrightarrow L^{l,q}(E)$, and this inclusion is compact if $k > l$ and $k - \frac{n}{p} > l - \frac{n}{q}$.

(ii) for $k - \frac{n}{p} \geq l + \alpha$ the inclusion $L^{k,p}(E) \hookrightarrow C^{l,\alpha}(E)$ is continuous, and is compact if $k - \frac{n}{p} > l + \alpha$.

Examples.

1) $L^{1,2}(E) \hookrightarrow L^2(E)$ is compact.
2) For a 1-manifold M, $L^{1,2}(E) \hookrightarrow C^0(E)$ is compact.
3) For a 4-manifold M, $L^{1,2}(E) \hookrightarrow L^4(E)$ is continuous, but not compact. This is a borderline case of the Sobolev theorem.

Notice that the Sobolev embeddings are monotone in the dimension, becoming stronger as the dimension is reduced. Thus, for example, the embedding $L^{1,2} \hookrightarrow L^4$ is borderline in dimension 4, but compact in dimension $n \leq 3$.

Sobolev spaces provide the context for the global approach to variational problems. To illustrate the basic technique—and its shortcomings—we will consider several well-known geometric problems. The first is a basic fact about the Laplacian on a Riemannian manifold.

PROPOSITION 1.2. Let M be a compact Riemannian manifold and fix $g \in C^\infty(M)$ with $\int g = 0$. Then there exists $f \in C^\infty(M)$, unique up to a constant, such that $\Delta f = g$.

Here $\Delta = d^*d$ is the (positive) Laplacian. The idea of the proof is to directly minimize the Lagrangian

$$L(f) = \int_M |df|^2 - 2fg,$$

whose formal variational equation is $\Delta f = g$. Since $L(f)$ closely resembles the definition of the norm in the Sobolev space $L^{1,2}$, it is natural to minimize over $f \in L^{1,2}(M)$. Before proceeding, two remarks are in order.

First, by adding a constant to f we may assume its average value is zero, i.e., that $\int f = 0$. Hence it is more convenient to minimize f over

$$H = \left\{ f \in L^{1,2}(M) \mid \int f = 0 \right\}$$

(we will then get uniqueness). This space H is a closed linear subspace of $L^{1,2}(M)$. Second, each $f \in H$ satisfies the Poincaré inequality

$$\int_M f^2 \leq c \int_M |df|^2$$

for some constant c.

PROOF OF PROPOSITION 1.2. Using $2ab \leq \varepsilon a^2 + \varepsilon^{-1} b^2$ and the Poincaré inequality, we have

$$\begin{aligned} L(f) &\geq \int_M |df|^2 - \frac{1}{2c}\int_M f^2 - \frac{c}{2}\int_M g^2 \\ &\geq \int_M |df|^2 - \frac{c}{2c}\int_M |df|^2 - \frac{c}{2}\int_M g^2 \\ &= \frac{1}{2}\int_M |df|^2 - \frac{c}{2}\int_M g^2. \end{aligned} \tag{1.3}$$

Since g is fixed this shows that $L(f)$ is bounded below. Choose a sequence $\{f_k\} \in H$ with $L(f_k) \to L_0 = \inf\{L(f_k) \mid f \in H\}$. Then (1.3) and the Poincaré inequality imply that $\|f_k\|_{1,2}$ is bounded. Since the unit ball in H is weakly compact we can choose a subsequence—still denoted $\{f_k\}$—converging weakly to $f_0 \in L^{1,2}$. Furthermore, $L^{1,2} \hookrightarrow L^2$ is compact so we may also assume that $\{f_k\} \to f_0$ in L^2. Convergence in L^2 implies that $\int f_0 = \lim \int f_k = 0$, so $f_0 \in H$, and weak convergence implies that $L(f_0) \leq \lim L(f_k) = L_0$, so f_0 minimizes L over H. Hence for any $h \in C^\infty(M)$ with $\int h = 0$ and $t \geq 0$

$$L(f_0) \leq L(f_0 + th) = L(f_0) + 2t \int \langle df_0, dh\rangle - hg + 0(t^2). \tag{1.4}$$

This holds for all t, so, integrating by parts,

$$0 \leq \int_M \langle \Delta f_0 - g, h\rangle.$$

Replacing h by $-h$ shows the right-hand side equals 0. Since $\int g = 0$, we conclude that $\Delta f_0 = g$ weakly. Elliptic regularity then implies that f_0 is smooth and satisfies $\Delta f_0 = g$. Finally, if $f_0, \tilde{f}_0 \in H$ are solutions then $0 = \Delta(f_0 - \tilde{f}_0)$; multiplying by $f_0 - \tilde{f}_0$, integrating by parts, and using the Poincaré inequality then shows $f_0 = \tilde{f}_0$. ❖

In the 1960's a general machinery was developed for treating non-linear problems by the same direct approach used in the above linear problem. In the non-linear case, one considers a C^2 Lagrangian function $F: \mathcal{M} \to \mathbb{R}$ defined in a complete C^2 Riemannian Hilbert manifold $\mathcal{M}$ without boundary. We say F satisfies the *Palais-Smale Condition (PS)* if every sequence $\{x_k\}$ in $\mathcal{M}$ with $\{F(x_k)\}$ bounded and $\|(\text{grad } F)_{x_k}\| \to 0$ has a convergent subsequence. This condition means that the downward gradient flow lines of F converge. Assuming it, R. Palais and S. Smale showed that the basic lemmas of Morse Theory apply in this infinite-dimensional context.

THEOREM 1.3 (Palais [P1], Smale [Sm]). If $F: \mathcal{M} \to \mathbb{R}$ satisfies the Palais-Smale Condition and has only non-degenerate critical points, then Morse theory applies, that is:

(a) The critical values of F are isolated and there are only a finite number of critical points of F at any critical level.
(b) If there are no critical values of F in $[a, b]$ then $\mathcal{M}_b = F^{-1}((-\infty, b])$ is diffeomorphic to $\mathcal{M}_a$.
(c) If $a < c < b$ and c is the only critical value of F in $[a, b]$ and $x_i, \ldots, x_n$ are the critical points at level c, then $\mathcal{M}_b$ deformation retracts to $\mathcal{M}_a \cup H_1 \cup \cdots \cup H_n$, where H_i is a cell ("handle") of dimension equal to the index of F at x_i attached to $\partial\mathcal{M}_b$ by a homeomorphism of the boundary spheres.

Consequently, if the critical points of F are non-degenerate, there are at least as many critical points as predicted by the homotopy type of $\mathcal{M}$.

In practice it is often difficult to verify that the critical points are non-degenerate. In these circumstances one may still deduce the existence of critical points using Lusternik-Schnirelman theory.

Recall that the category $\text{Cat}(X)$ of a path-connected space X is the least integer n such that X can be covered by n closed contractible subsets of X, and for a closed subset $A \subset X$, $\text{Cat}(A, X)$ is the least n such that A can be covered by n closed sets contractible in X. A well-known theorem gives the bound $\text{Cat}(X) > \text{cuplength}(X)$, where the cuplength is the largest n such that there exist elements $x_1, \ldots, x_n \in H^*(X)$ with $x_1 \cup \cdots \cup x_n \neq 0$.

In Lusternik-Schnirelman theory one considers a smooth function F on a finite-dimensional manifold X and defines values c_m by the mini-max procedure

$$c_m(F) = \inf_{\text{Cat}(A) \leq m} \left[\sup_{a \in A} F(a) \right] \qquad m \leq \text{Cat}(X).$$

These satisfy $c_1 \le c_2 \le \cdots$. The main result of Lusternik-Schnirelman theory is that

(a) Each finite $c_m(F)$ is a critical value.
(b) If F is bounded below then F has at least $\mathrm{Cat}(X)$ critical points.
(c) If $c_m(F) = c_{m+1}(F) = \cdots = c_{m+n}(F) = c$ is finite then the critical set $K_c = \{x \in X \mid F(x) = c,\ (\nabla F)_x = 0\}$ satisfies $\mathrm{Cat}(K_c) \ge n + 1$.

Palais and J. Schwartz extended this classical result to infinite dimensions:

THEOREM 1.4 ([P3] Theorem 7.2 or [S] Corollary 16). Let $F: \mathcal{M} \to \mathbb{R}$ be a C^2 function on a complete C^2 Banach manifold $\mathcal{M}$ without boundary. If F satisfies the Palais-Smale Condition then Lusternik-Schnirelman theory applies. In particular, if $\mathcal{M}$ has infinite cuplength and F is bounded below then F has infinitely many critical points.

These theorems give a potentially powerful method of establishing existence results for global non-linear equations. In fact, this approach scored an immediate success when applied to the original problem of Morse: finding geodesics on manifolds. This case, which we describe next, is the model example of Palais' vision of global non-linear analysis.

Fix points x, y in a complete Riemannian manifold (M, g) and consider the energy function

$$E(\gamma) = \frac{1}{2} \int_0^1 |\dot{\gamma}(t)|^2\, dt$$

on the path space $\mathcal{P}$ of all $L^{1,2}$ paths $\gamma: [0, 1] \to M$ with $\gamma(0) = x$ and $\gamma(1) = y$. To topologize $\mathcal{P}$, fix an embedding of M into $\mathbb{R}^N$. This defines an inclusion of $\mathcal{P}$ into the Hilbert space

$$H = L^{1,2}([0, 1], \mathbb{R}^N).$$

The Sobolev embedding $L^{1,2} \subset C^{\frac{1}{2}}$ shows that $\mathcal{P}$ is a closed subspace of H with one component for each homotopy class of paths from x to y, and that each path $\gamma \in \mathcal{P}$ is Hölder continuous and lies in a bounded set in M whose radius depends only on $E(\gamma)$. An application of the implicit function theorem shows that $\mathcal{P}$ is a smooth embedded submanifold of H and E is a smooth function (we will give such an argument in Lemma 4.2 below). The inner product on H then induces an $L^{1,2}$ Riemannian metric on $\mathcal{P}$.

Computing the variation, one finds that the differential of E is

$$(dE)_\gamma(X) = \int_0^1 g(X, \nabla_T T)\, dt$$

where $T = \dot{\gamma}$. Thus the critical points are solutions of the geodesic equation $\nabla_T T = 0$.

LEMMA 1.5 (cf. [P1], [S]). The energy function $E\colon \mathcal{P} \to \mathbb{R}$ satisfies the Palais-Smale condition. Each critical point is a smooth geodesic from x to y.

PROOF. For any Palais-Smale sequence $\{\gamma_k\}$ in $\mathcal{P}$ the image curves γ_k all lie in a bounded set B, so $\{\gamma_k\}$ is a bounded sequence in $L^2([0,1], \mathbb{R}^N)$. In fact, since the E_k are bounded, $\{\gamma_k\}$ is a bounded sequence in H and hence is equicontinuous. By weak compactness and the Arzela-Ascoli theorem there is a subsequence $\{\gamma_k\}$ that converges strongly in $L^2 \cap C^0$ and weakly in H to $\gamma_0 \in H$.

It remains to show that the differences $\phi_{kl} = \gamma_k - \gamma_l$ converge to 0 in $L^{1,2}$. Since we already have L^2 convergence it suffices to show that $\dot{\phi}_{kl} \to 0$ in L^2. Integrating by parts and writing $\Gamma(T,T) = \dot{T} - \nabla_T T$, we have

$$
\begin{aligned}
(1.5) \qquad \|\dot{\phi}_{kl}\|^2 &= -\int_0^1 \langle \phi_{kl}, \partial_t (T_k - T_l)\rangle \\
&= -\int_0^1 \langle \phi_{kl}, \nabla_{T_k} T_k - \nabla_{T_l} T_l + \Gamma(T_k, T_k) - \Gamma(T_l, T_l)\rangle \\
&\le \left| \int_0^1 \langle \phi_{kl}, \nabla_{T_k} T_k \rangle \right| + \sup |\Gamma \phi_{kl}| \int_0^1 |T_k|^2 + (k \leftrightarrow l).
\end{aligned}
$$

where $(k \leftrightarrow l)$ denotes the same terms with k replaced by l. These inner products are with respect to the euclidean metric, but on the set B the euclidean metric is uniformly equivalent to the Riemannian metric g of M and Γ is bounded. Thus

$$
\|\dot{\phi}_{kl}\|^2 \le \left| \int_0^1 g\left(\phi_{kl}, \nabla_{T_k} T_k\right) \right| + cE_k \sup |\phi_{kl}| + (k \leftrightarrow l).
$$

The term $E_k \sup |\phi_{kl}|$ vanishes in the limit since the E_k are bounded and $\phi_{kl} \to 0$ pointwise.

Now $\phi_{kl}(t)$ is a vector field along γ_k that vanishes at the endpoints but is not tangent to $M \subset \mathbb{R}^N$. However, its tangential component $\pi_k \phi_{kl}$ lies in $T_{\gamma_k}\mathcal{P}$, so

$$
\begin{aligned}
\left| \int_0^1 g\left(\pi_k \phi_{kl}, \nabla_{T_k} T_k\right) \right| &= |dE_k(\pi_k \phi_{kl})| \le \|\phi_{kl}\|_{1,2} \cdot \|dE_k\|_{-1,2} \\
&= \|\phi_{kl}\|_{1,2} \cdot \|(\nabla E)_k\|_{1,2} \to 0
\end{aligned}
$$

since the ϕ_{kl} are bounded in $L^{1,2}$ and the sequence is Palais-Smale. This leaves only the normal component, for which we have

$$\left|\int_0^1 g\left((1-\pi_k)\phi_{kl}, \nabla_{T_k}T_k\right)\right| \le \|(1-\pi_k)\phi_{kl}\|_{1,2} \cdot \|\nabla_{T_k}T_k\|_{-1,2}.$$

But multiplication induces a bounded linear map $L^2 \times L^2 \to L^{-1,2}$, so the sequence

$$\|\nabla_{T_k}T_k\|^2_{-1,2} = \|\dot{T}_k - \Gamma(T_k, T_k)\|^2_{-1,2} \le 2\|T_k\|_2^2 + c(\Gamma)\|\,|T_k|^2\,\|^2_{-1,2} \le 2E_k + cE_k^2$$

is bounded. Finally, consider the orthogonal projection $\pi_p \colon \mathbb{R}^N \to T_pM$ at a point $p \in M \subset \mathbb{R}^N$. Then for q near p and $X \in T_q\mathbb{R}^N$ we have

$$|X - \pi_p X| \le \operatorname{dist}(p, q) \cdot |X|$$

with c uniform for $p, q \in B$ with $\operatorname{dist}(p, q) \le \varepsilon$. Applying this to ϕ_{kl} and $\dot{\phi}_{kl}$ and integrating shows that

$$\|(1-\pi_k)\phi_{kl}\|_{1,2} \le c \sup|\phi_{kl}| \cdot \|\phi_{kl}\|_{1,2} \to 0.$$

We conclude that each term in (1.5) vanishes in the limit, so $\{\gamma_k\}$ is Cauchy in $\mathcal{P}$. The limit γ_0 is then a weak solution of $\nabla_T T = 0$ and by elliptic regularity (for O.D.E.s!) is a smooth geodesic. ❖

One way to apply Lemma 1.5 is to invoke the fact that for a generic pair of points $x, y \in M$ the critical points of the energy function are non-degenerate ([M], [U]). For these pairs the Morse Theorem 1.3 applies. Alternatively, without assumptions on x and y, we obtain a Lusternik-Schnirelman theory from Theorem 1.4. Either way, we have linked the critical points of L to the topology of $\mathcal{P}$. But $\mathcal{P}$ has the homotopy type of the loop space ΩM and, when M is compact and simply-connected, the spectral sequence for the path space fibration

$$\begin{array}{ccc} \mathcal{P} = \Omega M & \longrightarrow & PM = * \\ & & \downarrow \\ & & M \end{array} \tag{1.6}$$

shows that $\mathcal{P}$ has infinite cuplength ([Se]). Consequently, the energy function E on $\mathcal{P}$ must have infinitely many critical points.

COROLLARY 1.6. On a compact 1-connected Riemannian manifold (M, g), there are infinitely many geodesics joining any two points $x, y \in M$.

Another application is to finding closed geodesics. For this, consider the energy function on the space ΛM of $L^{1,2}$ maps $S^1 \to M$. As above, ΛM is a manifold and the Palais-Smale Condition holds provided M is compact (compactness is needed to ensure boundedness in the first sentence of the proof of Lemma 1.5). Minimizing over each component of ΛM and noting that $\pi_0(\Lambda M) = \pi_1(M)$ by (1.6) gives the following classical theorem.

COROLLARY 1.7. On a compact Riemannian manifold (M, g), every element of $\pi_1(M)$ is represented by a closed geodesic.

Unfortunately, these applications to geodesics were the *only* notable success of this program in the 1960's. It turns out that the Palais-Smale Condition fails for many important non-linear geometric problems.

To understand the way in which the Palais-Smale Condition typically fails, consider the "non-linear eigenvalue problem"

$$\Delta f + f = \lambda f^3 \tag{1.7}$$

on S^4, where λ is an unknown constant. (This equation arises in connection with the Yamabe problem, and we will study a more general version of it in section 4.) To solve this by variational methods, one should minimize

$$L(f) = \int_{S^4} |df|^2 + f^2 \tag{1.8}$$

subject to the constraint that $\int_{S^4} f^4 = 1$ (this constraint gives a Lagrange multiplier and hence the right-hand side of (1.7)). As above, one can choose a minimizing sequence $\{f_k\}$ in $L^{1,2}$, and find a subsequence converging weakly to $f_0 \in L^{1,2}$. However, the inclusion $L^{1,2} \hookrightarrow L^4$ is not compact, and hence the weak limit f_0 *need not satisfy the constraint* and the variational method breaks down. Equivalently, from the Morse Theory perspective, one finds that L, as a function on the constraint manifold, does not satisfy the Palais-Smale Condition.

The difficulty here is that the inclusion $L^{1,2} \hookrightarrow L^4$ is a borderline Sobolev embedding. If we change the equation, replacing the right-hand side of (1.7) by $\lambda f^{3-\varepsilon}$, $\varepsilon > 0$, then the relevant embedding $L^{1,2} \hookrightarrow L^{4-\varepsilon}$ is compact, the variational method works, and in fact the non-linearities in the problem are of little consequence. However, many natural geometric elliptic equations (e.g., the Yamabe and Yang-Mills equations) involve borderline Sobolev embeddings—indeed it is precisely this analytic property that allows the solutions to display interesting nonlinear behavior. In the next section we see how the imposition of symmetry allows one to circumvent this problem of borderline embeddings.

2. The Equivariant Sobolev Theorem

Suppose that a compact Lie group H acts isometrically on the vector bundle $E \to M$, that is, each $h \in H$ gives a smooth metric-preserving bundle map $\tilde{h}\colon E \to E$ covering an isometry $h\colon M \to M$. A section $\phi \in \Gamma(E)$ is *equivariant* if $\phi(h(x)) = \tilde{h}(\phi(x))$ for all $x \in M$. Let $\Gamma_H(E)$ denote the set of all smooth equivariant sections.

The group H also acts on the connections on E. Specifically, $h \in H$ takes a connection ∇ to the connection $h \cdot \nabla$ defined by

$$(h \cdot \nabla)_X \phi = \tilde{h}^{-1}(\nabla_{h_* X}(\tilde{h}\phi)), \tag{2.1}$$

where X is a vector field on M and $\phi \in \Gamma(E)$. By the device of averaging over the group we obtain an invariant connection ∇^0, which we fix once and for all.

Completing the set $\Gamma_H(E)$ with respect to the norms (1.1) and (1.2)—using the invariant connection ∇^0—we obtain Sobolev spaces $L^{k,p}_H(E)$ and $C^k_H(E)$. These equivariant Sobolev spaces are closed subspaces of $L^{k,p}(E)$ and $C^k(E)$, so the Sobolev embeddings of Theorem 1 restrict to corresponding embeddings for $L^{k,p}_H(E)$ and $C^k_H(E)$. However, when the H-orbits have positive dimension one expects a better result: since equivariant sections correspond, in some sense, to sections over the orbit space M/H, the spaces $L^{k,p}_H(E)$ should satisfy the Sobolev theorem with the dimension n replaced by $\dim(M/H)$. While this statement, as it stands, is not correct, the following theorem shows that equivariance does indeed yield improved Sobolev embeddings.

THEOREM 2.1 (Equivariant Sobolev Embedding Theorem). Suppose H acts isometrically on E and that each H-orbit in M has dimension $\geq d$. Then the spaces $L^{k,p}_H(E)$ and $C^k_H(E)$ satisfy the Sobolev Embedding Theorem with n replaced by $n - d$.

PROOF. We give the proof for the spaces $L^{k,p}_H(E)$; the proof for the $C^k_H(E)$ is similar.

The orbit through each $x \in M$ has a tubular neighborhood $\mathcal{U}_x$ equivariantly diffeomorphic (by the exponential map) to $H \times_{H_x} B_x$, where H_x is the isotropy subgroup at x, and B_x is the ball of radius ε in the normal space to the orbit at x. For small ε the differential of the exponential map is uniformly close to the identity, and hence the metric on $\mathcal{U}_x$ is uniformly close to the product metric on $H \times_{H_x} B_x$. Let $\{\mathcal{U}_i = \mathcal{U}_{x_i}\}_{i=1}^{l}$ be a finite subcover of $\{\mathcal{U}_x \mid x \in M\}$. Write E_i for the restriction of the bundle E to $\mathcal{U}_i$ and set

$$(2.2)\quad L^{k,p}_H(\{E_i\}) = \big\{(\phi_1, \ldots, \phi_l) \in \oplus L^{k,p}_H(E_i) \,\big|\, \phi_i = \phi_j \text{ on } \mathcal{U}_i \cap \mathcal{U}_j,\ \forall i, j\big\}.$$

We then have isomorphisms

$$(2.3)\qquad L^{k,p}_H(E) \simeq L^{k,p}_H(\{E_i\})$$

([P4], Theorem 4.3). Furthermore, each equivariant section ϕ on $\mathcal{U}_x$ pulls back by the exponential map to an equivariant section ϕ^* on $H \times_{H_x} B_x$, and the $L^{k,p}$ norm of ϕ on $\mathcal{U}_x$ is uniformly equivalent to the $L^{k,p}$ norm of ϕ^* on $H \times_{H_x} B_x$. It therefore suffices to prove the theorem for equivariant Sobolev spaces on $H \times_{H_x} B_x$ with its product metric.

Each equivariant section ϕ on $H \times_{H_x} B_x$ satisfies

$$(2.4)\qquad |\phi(h(x))| = |\tilde{h}\phi(x)| = |\phi(x)|,$$

i.e., $|\phi|$ is constant on orbits. The total covariant derivative $\nabla^0\phi$ splits into components $(\nabla^0\phi)^T$ tangent to the orbits and components $(\nabla^0\phi)^S$ tangent to the slice $S = \mathrm{Id} \times B_x$. Thus

$$(2.5)\qquad |\nabla^0\phi|^2 = |(\nabla^0\phi)^T|^2 + |(\nabla^0\phi)^S|^2$$

where, by the invariance of ∇^0, each term is constant along orbits. Now each X in the Lie algebra of H gives a vector field $\tilde{X}$ on E, and the Lie derivative $\mathcal{L}_{\tilde{X}}\phi$ vanishes for equivariant ϕ. Hence the difference $\Psi_X = \nabla^0_X - \mathcal{L}_{\tilde{X}}$ is a smooth endomorphism, depending only on ∇^0 and the H-action. Equation (2.5) can then be written as

$$(2.6)\qquad |\nabla^0\phi|^2 = |\nabla^0\bar{\phi}|^2 + |\Psi(\bar{\phi})|^2$$

where $\bar{\phi}$ is the restriction of ϕ to the slice. From (2.4), (2.6), and the fact that Ψ is bounded we obtain

$$\int_{H\times_{H_x}B_x} |\nabla^0\phi|^p = \mathrm{Vol}(H/H_x)\int_S |\nabla^0\bar{\phi}|^p + C\int_S |\bar{\phi}|^2,$$

so the $L^{1,p}_h$-norm of ϕ on $H\times_{H_x} B_x$ is equivalent to the $L^{1,p}$-norm of $\bar{\phi}$ on S. Similar statements hold for higher derivatives, giving isomorphisms

$$L^{k,p}_H(E) \simeq L^{k,p}_{H_x}(E|_{\text{slice}}).$$

Finally, since each H-orbit has dimension $\geq d$, each slice $S_i = \mathrm{Id}\times B_{x_i}$ has dimension $\leq n-d$, so the $(n-d)$-dimensional Sobolev embeddings hold on each slice. ❖

The improvement in the Sobolev embeddings stated in Theorem 2.1—with the dimension n being effectively reduced by the dimension d of the *smallest* orbit—is the best possible. The following example shows why. It is an example of an action with fixed points (so $d = 0$) where equivariance yields no improvement in the Sobolev embeddings. One can easily construct similar examples with $d > 0$ by taking the product of this example with a homogeneous space.

2.2 Example. Fix $n \geq 3$ and set $p = 2n/n-2$. The functions f_λ on $\mathbb{R}^n$ defined by

$$f_\lambda(x) = \left(\frac{\lambda}{\lambda^2+|x|^2}\right)^{(n-2)/2} \qquad \lambda > 0$$

are smooth and positive. Differentiating in polar coordinates shows that $f_\lambda \Delta f_\lambda = -f_\lambda f_\lambda'' + (n-1)r^{-1} f_\lambda f_\lambda' = n(n-2) f_\lambda^p$. Integrating by parts and substituting $x = \lambda y$ shows that

$$\text{(2.7)} \qquad \frac{1}{n(n-2)}\int_{\mathbb{R}^n} |df_\lambda|^2 = \int_{\mathbb{R}^n} f_\lambda^p = \int_{\mathbb{R}^n} (1+|y|^2)^{-n}\, dy = C_n.$$

is a constant independent of λ.

Now consider the functions ϕ_λ corresponding to f_λ under the standard stereographic projection $\sigma\colon \mathbb{R}^n \to S^n$. As is well-known, this map is conformal: the metric g on S^n pulls back to $(1+|x|^2)^{-1}\delta_{ij} = f_1^{p/n}\delta_{ij}$ on $\mathbb{R}^n$. Pulling back $|d\phi_\lambda|^2 = g^{-1}(d\phi_\lambda, d\phi_\lambda)$ and the volume form dv_g, we have

$$\text{(2.8)} \qquad \int_{S^n} |d\phi_\lambda|^2 + \phi_\lambda^p\, dv_g = \int_{\mathbb{R}^n} \left(|f_1^{-p/n} df_\lambda|^2 + f_\lambda^p\right) f_1^p\, dx^1 \cdots dx^n$$

with $f_1 \leq 1$. It follows from (2.7), (2.8) and Hölder's inequality that the functions $\{\phi_\lambda\}$ on S^n are uniformly bounded in $L^{1,2}$. Moreover, the f_λ are clearly invariant under the usual action of $H = SO(n)$ on $\mathbb{R}^n$, and hence the ϕ_λ are invariant under the corresponding action of H on S^n. Fixing a sequence $\lambda_k \to 0$, we obtain a sequence $\{\phi_k\}$ bounded in $L^{1,2}_H(S^n)$ that become more and more concentrated at the south pole as $k \to \infty$.

With our choice of p the embedding $L^{1,2} \hookrightarrow L^p$ is a borderline Sobolev embedding. If equivariance gave any improvement in the Sobolev inequalities the embedding $L^{1,2}_H \hookrightarrow L^p_H$ would be compact, and $\{\phi_k\}$ would have a subsequence converging in L^p to some $\phi_0 \in L^{1,2}_H$. But as $\lambda \to 0$ the f_λ converge to 0 pointwise on $\mathbb{R}^n - \{0\}$, so we would have $\phi_k \to \phi_0 = 0$ in L^p. However, integrating as in (2.7) and (2.8)

$$\int_{S^n} \phi_\lambda^p \, dv_g = \int_{\mathbb{R}^n} f_\lambda^p f_1^p \, dx \geq 2^{-p} \int_{B(0,1)} f_\lambda^p \, dx = 2^{-p} C_n > 0.$$

Thus $\{\phi_k\}$ does not converge in L^p and the embedding $L^{1,2}_H \hookrightarrow L^p_H$ is not compact.

This example illustrates a well-known general phenomenon that occurs for the borderline Sobolev embedding: a bounded sequence in $L^{1,2}$ either has a subsequence that converges in L^p, or it has a subsequence whose L^p norm becomes concentrated at a finite set of points $\{p_i\} \subset M$. The details above simply verify that this latter possibility can occur equivariantly, with a sequence of invariant functions concentrating at a fixed point of the group action.

3. Morse Theory for Equivariant Functions

The Equivariant Sobolev Theorem gives a direct way to approach variational problems that are borderline for the Sobolev embeddings: look for solutions invariant under an appropriate group action, use the Equivariant Sobolev Theorem to verify the Palais-Smale Condition, and apply the theorems of section 1. In subsequent sections we will apply this method to specific problems. Here we will describe the main steps in this approach. As will be clear, this program owes much to the work of Palais.

Suppose that M is a Riemannian manifold with an action of a compact Lie group G and that all orbits have dimension $d \geq 1$. We are given a Lagrangian L defined on a Hilbert or Banach manifold $\mathcal{M}$ associated to M

(e.g., $\mathcal{M} = L^{1,2}(E)$ for some vector bundle $E \to M$) and an induced action of G on $\mathcal{M}$. We can then seek solutions to the variational equations of L using the following procedure.

(i) Show that the fixed point set M^G is a Hilbert or Banach manifold and that $L: \mathcal{M}^G \to \mathbb{R}$ is smooth.
(ii) Prove the PS Condition using the Equivariant Sobolev Theorem.
(iii) Apply the Morse or Lusternik-Schnirelman theorems to obtain points critical with respect to invariant variations.
(iv) Show that these are critical with respect to all variations.
(v) Use elliptic theory to show that these critical points are smooth solutions of the variational equations.

Only step (iv) needs clarification. Here we encounter an interesting general phenomenon: criticality with respect to invariant perturbations implies criticality with respect to all perturbations. This too was studied by Palais [P5], who dubbed it the "Symmetric Criticality Principle". Specifically,

SYMMETRIC CRITICALITY PRINCIPLE. Let $\mathcal{M}$ be a smooth Banach manifold with smooth G action with fixed set $\mathcal{M}^G$. Suppose that $F: M \to \mathbb{R}$ is a smooth G-invariant function. If $x \in \mathcal{M}^G$ is a critical point of $F: \mathcal{M}^G \to \mathbb{R}$, then x is a critical point of F.

The Symmetric Criticality Principle is often implicitly assumed. In a typical example, a physicist seeking critical points for a Lagrangian L for a field on $\mathbb{R}^3$ will impose the Ansatz of spherical symmetry by seeking solutions $f(r)$ that depend only on the distance from the origin. Inserting $f(r)$ into the Lagrangian and computing the Euler-Lagrange equation, yields an O.D.E. for $f(r)$. A solution to this O.D.E is a critical point for L with respect to *spherically symmetric perturbations*—it is only by the symmetric criticality principle that one knows these solve the original problem.

The Symmetric Criticality Principle is not always true (it is a principle, not a theorem; counterexamples are given in [P5]). Nevertheless, it is valid in considerable generality. The following theorem gives conditions ensuring that it holds; these are sufficient for many applications. The proof makes transparent the key point behind the Symmetric Criticality Principle.

PROPOSITION 3.1. Let G be a Lie group acting smoothly on a Banach manifold $\mathcal{M}$, and let $F: \mathcal{M} \to \mathbb{R}$ be a smooth G-invariant function. Assume either (a) G is compact, or (b) G acts isometrically on M. Then the set

$\mathcal{M}^G \subset X$ of fixed points is a smooth closed invariant submanifold and the Symmetric Criticality Principle holds.

PROOF. At each fixed point $x \in \mathcal{M}^G$ the group acts on the tangent space. When M is a Riemannian Hilbert manifold we have a G-invariant orthogonal decomposition

$$T_x\mathcal{M} = T_x\mathcal{M}^G \oplus \text{Normal space} \tag{3.1}$$

and the exponential map is an equivariant diffeomorphism of a neighborhood of 0 in $T_x\mathcal{M}$ to a neighborhood of x in $\mathcal{M}$. It follows that $\mathcal{M}^G$ is a submanifold and that $T_x\mathcal{M}^G$ is exactly the invariant subspace of the action of G on $T_x\mathcal{M}$. Since F is invariant, ∇F is in this invariant subspace. But the component of ∇F in $T_x\mathcal{M}^G$ is precisely the gradient of the restriction of F to $\mathcal{M}^G$. Hence ∇F vanishes if and only if x is critical for $F: \mathcal{M}^G \to \mathbb{R}$.

When G is compact and $\mathcal{M}$ is a Hilbert manifold, we can always find an invariant metric $\bar{g}$ by choosing an arbitrary metric and averaging over the group. The action is then isometric with respect to $\bar{g}$ and the theorem follows as above.

When $\mathcal{M}$ is a non-Hilbert Banach manifold a slightly different argument is needed. By applying the Implicit Function Theorem and averaging over the group, one can show (see [P5], section 5) that the G action is still linearizable at each $x \in \mathcal{M}^G$, i.e., there is an equivariant diffeomorphism of a neighborhood of x in $\mathcal{M}$ with a neighborhood of 0 in $T_x\mathcal{M}$ with a decomposition (3.1). The proof then proceeds as before. ❖

This Proposition shows that the process of restricting to the invariant submanifold and applying the symmetric criticality principle is "soft"—it can be done generally and abstractly. Of course, the same is true of the Equivariant Sobolev Theorem and the Morse and Lusternik-Schnirelman Theorems. It is only in step (ii) of our program—verifying the Palais-Smale Condition—that we need look at specific equations and do hard estimates.

4. The Yamabe Variational Problem

In this section we illustrate the use of the Morse theory for equivariant functions by applying it to the variational problem that arises in the Yamabe problem.

In the Yamabe problem we are given a compact Riemannian manifold (M, g) of dimension $n \geq 3$ and asked to find a metric conformal to g which has constant scalar curvature. This problem can be cast in variational form as follows. Set $p = 2n/n-2$, $a = n-2/4(n-1)$ and let s_g denote the scalar curvature of the metric g. The critical points $f \in L^{1,2}$ of the Lagrangian

$$L(f) = \frac{1}{2}\int_M |df|^2 + a s_g f^2 \tag{4.1}$$

subject to the constraint $\int_M |f|^p = 1$ are solutions of the non-linear equation

$$\Delta f + a s_g f = \lambda f^{p-1}. \tag{4.2}$$

If f is a *minimal* critical point of L then (4.2) and the maximum principle imply that $|f| > 0$. The metric $g' = |f|^{p-2} g$ then has constant scalar curvature (see [A] or [LP] for details).

Equation (4.2) is a direct generalization of (1.7) and its analysis is plagued by the same problem—direct minimization methods fail because the exponent p is critical for the Sobolev embedding $L^{1,2} \hookrightarrow L^p$ in dimension n. The solution of the Yamabe problem—the cumulative work of H. Yamabe, N. Trudinger, T. Aubin and R. Schoen—requires a very detailed analysis of the geometric origins of the constant of this Sobolev embedding (cf. [LP]). Here we will show that these subtleties can be avoided for Riemannian manifolds with symmetry—direct minimization yields infinitely many distinct solutions of (4.2).

THEOREM 4.1. Let (M, g) be a compact Riemannian manifold of dimension $n \geq 3$. Suppose that a compact Lie group H acts isometrically on M with each orbit having dimension d, $1 \leq d < n$. Then there is a sequence $\{f_k\}$ of smooth, H-invariant solutions of (4.2) with $L(f_k) \to \infty$.

Remarks. (1) Wei Yue Ding [D] proved this for $M = S^n$ and $H = 0(k) \times 0(n-k+1)$.

(2) The solutions $\{f_k\}$ with $L(f_k)$ minimal are positive, and hence give solutions of the Yamabe problem. The higher-energy solutions change sign, and hence are not immediately relevant to the Yamabe problem.

The first step toward constructing a Morse theory for the Lagrangian (4.1) is to show that the constraint surface is a Riemannian Hilbert manifold. For this we write the $L^{1,2}$ norm as

$$\|f\|^2 = \int_M |df|^2 + f^2 = \langle f, \Box f\rangle_{L^2} \tag{4.3}$$

where $\Box = \Delta + 1 = d^*d + 1$. Note that $\Box$ is a bounded invertible operator $L_H^{k+2,2} \to L_H^{k,2}$ for each k; let $\Box^{-1}$ denote its inverse.

LEMMA 4.2. The set $X = \{f \in L_H^{1,2} \mid \|f\|_{L^p} = 1\}$ is a smooth complete codimension 1 Riemannian submanifold of $L_H^{1,2}$. At $f \in X$ its tangent space and unit normal vector are

$$T_f X = \{h \in L_H^{1,2} \mid \langle h, |f|^{p-2} f\rangle_{L^2} = 0\}$$
$$\nu_f = \frac{\Box^{-1}(|f|^{p-2}f)}{\|\Box^{-1}(|f|^{p-2}f)\|}. \tag{4.4}$$

PROOF. X is a level set of the function $F(f) = \int |f|^p$. For $h \in L^{1,2}$ and $f \in X$

$$(dF)_f(h) = \frac{d}{dt}F(f+th)\Big|_{t=0} = p\int |f|^{p-2} fh \tag{4.5}$$
$$\leq \|f\|_p^{p-1}\|h\|_p = \|h\|_p \leq C\|h\|_{1,2}$$

so $(dF)_f$ exists and is bounded. The higher derivatives are computed similarly, and one easily sees that F is C^∞. Since dF never vanishes on X the implicit function theorem then implies that X is a smooth, closed—hence complete—codimension 1 submanifold of $L_H^{1,2}$. The metric (4.3) on $L_H^{1,2}$ restricts to give a Riemannian metric on X. Equation (4.5) also shows that the tangent space is as stated in (4.4). Since the $L^{1,2}$ normal ν_f is characterized by

$$0 = \langle h, \nu\rangle_{L^{1,2}} = \langle h, \Box\nu\rangle_{L^2} \qquad \text{for all } h \in T_f X$$

we have $\Box\nu_f = \lambda|f|^{p-2}f$ for some non-zero $\lambda \in \mathbb{R}$. Inverting $\Box$ and normalizing gives the expression (4.4) for the unit normal. ❖

One can visualize X as a sort of ellipsoid in $L_H^{1,2}$ with tangent spaces and normal vectors given by (4.4)—see Figure 1. The next step is to apply the Lagrange Multiplier method as in ordinary calculus.

The Lagrangian (4.1) is a smooth function on X (it is the restriction of a quadratic function on $L^{1,2}$) with differential

$$(dL)_f(h) = \frac{d}{dt}L(f+th)\Big|_{t=0} = \int h(\Delta f + as_g f) \tag{4.6}$$
$$= \langle h, \Box^{-1}(\Delta f + as_g f)\rangle_{L^{1,2}}$$
$$= \langle h, f + \Box^{-1}((as_g - 1)f)\rangle_{L^{1,2}}.$$

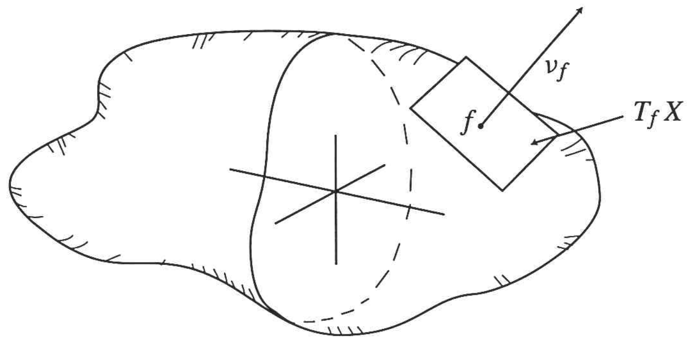

Figure 1. X as a submanifold of $L^{1,2}_H$.

This last expression gives the gradient of L on $L^{1,2}_H$. Consequently, the gradient of L on X is

$$(\nabla L)_f = \left[f + \Box^{-1}(as_g - 1)f\right]^T \in T_f X \tag{4.7}$$

where T denotes the orthogonal projection $\phi^T = \phi - \langle\phi, \nu_f\rangle\nu_f$ onto $T_f X$. With this formula in hand, we can now use the Equivariant Sobolev Theorem to establish the PS Condition.

LEMMA 4.3. $L\colon X \to \mathbb{R}$ satisfies the Palais-Smale Condition.

PROOF. Let $\{f_k\}$ be a sequence in X with $L(f_k) \le B$ and $\|(\nabla L)_{f_k}\| \to 0$. By Hölder's inequality

$$\begin{aligned} \|f_k\|^2 &= L(f_k) + \int (1 - as_g) f^2 \\ &\le B + \|f_k\|^2_{L^p}\|(1 - as_g)\|_{L^{n/2}} \\ &= B + \|(1 - as_g)\|_{L^{n/2}} \end{aligned} \tag{4.8}$$

so the sequence is bounded in $L^{1,2}_H$. Since the unit ball in $L^{1,2}_H$ is weakly compact and the inclusion $L^{1,2}_H \hookrightarrow L^p$ is compact by Theorem 2.1, we can choose a subsequence (also denoted $\{f_k\}$) that converges weakly in $L^{1,2}$ and strongly in L^p to $f \in L^{1,2}_H$. Convergence in L^p implies that $\int f^p = 1$, so $f \in X$. It also implies that $\{f_k^{p-1}\} \to f^{p-1}$ in $L^{p/p-1}$ so, since $\Box^{-1}$ is a bounded map $L^{p/p-1} \to L^{2,p/p-1} \subset L^{1,2}$, the sequence $\{\nu_k = \nu_{f_k}\}$ given by (4.4) converge to ν_f in $L^{1,2}$. Similarly, the Hölder inequality used in (4.8) implies that $\{(1 - as_g)f_k\} \to \{(1 - as_g)f\}$ in $L^{p/p-1}$, so $\{\square^{-1}(1 - as_g)f_k\} \to \square^{-1}(1 - as_g)f$ in $L^{1,2}$.

Now set $\phi_k = f_k + \square^{-1}(as_g - 1)f_k$. Then $\{\phi_k\}$ converges weakly in $L^{1,2}$ and $\{\nu_k\} \to \nu_f$ in $L^{1,2}$. It follows that $\{\langle\phi_k, \nu_k\rangle\nu_k\}$ converges in $L^{1,2}$. But from (4.7) our hypothesis on ∇L is that $(\nabla L)_{f_k} = \phi_k - \langle\phi_k, \nu_k\rangle\nu_k$ converges to zero. We then conclude that $\{\phi_k\}$, and therefore also $\{f_k\}$, converges in $L^{1,2}$. ❖

LEMMA 4.4. $L\colon X \to \mathbb{R}$ has infinitely many critical points.

PROOF. The key point is that the Lagrangian (4.1) is *even*. In fact, the involution $f \to -f$ is a smooth isometry of X leaving L invariant, so we get a 2-fold cover $X \to X/\mathbb{Z}_2$ and Lemma 4.3 implies that $L\colon X/\mathbb{Z}_2 \to \mathbb{R}$ satisfies the Palais-Smale Condition.

Now the well-known fact that the unit sphere in Hilbert space is contractible ([K]) indicates that $X/\mathbb{Z}_2$ is a $K(\mathbb{Z}_2, 1)$. The specific argument requires some care (X is not a CW complex) and goes as follows. First, the map $H(f) = f/\|f\|$ is a homeomorphism from X to the unit sphere S in the Hilbert space $L^{1,2}_H$, which is infinite-dimensional because the H-orbits have dimension $< n$. Thus X is contractible. Next, $L^{1,2}_H$ is filtered by the eigenspaces of $\square$; inductively enumerating the normalized eigenfunctions in the order of their eigenvalues gives a continuous inclusion $\mathbb{R}^\infty \hookrightarrow L^{1,2}_H$ with dense image. Restricting to the unit sphere and composing with the homeomorphism $H^{-1}\colon S \to X$ yields a diagram

$$\begin{array}{ccc} S^\infty & \xhookrightarrow{\;i\;} & X \\ \downarrow & & \downarrow \\ \mathbb{RP}^\infty & \xhookrightarrow{\;j\;} & X/\mathbb{Z}_2 \end{array}$$

where i and j are weak homotopy equivalences. A result of Palais ([P2], Theorem 15) then shows that i, j are in fact homotopy equivalences. Thus $H^*(X/\mathbb{Z}_2; \mathbb{Z}_2) = H^*(\mathbb{RP}^\infty; \mathbb{Z}_2)$ is a polynomial algebra on one generator in dimension 1, so $X/\mathbb{Z}_2$ has infinite cuplength. By the Lusternik-Schnirelman theorem L has infinitely many critical points on $X/\mathbb{Z}_2$, and hence on X. ❖

These critical points are critical with respect to variations tangent to X, that is, variations that preserve both the constraint *and the H-symmetry*. The Symmetric Criticality Principle insures that they are critical with respect to all variations that preserve the constraint. Thus we have infinitely many transforms the Lagrangian (5.1) to

$$\frac{1}{2|\lambda|}\int_M |F^\nabla|^2 + |\nabla\phi|^2 \pm \frac{1}{2}(|\phi|^2 - 1)^2\, dv \tag{5.4}$$

where the $\pm$ is the sign of λ. Hence there is a one-to-one correspondence between the critical points of (5.1) and those of (5.4); equivalently, one can check this directly by substituting (5.3) into (5.2). Moreover, the factor of $|\lambda|^{-1}$ at the front of (5.4) is inconsequential to the variational problem—the critical points are the same for any $\lambda \neq 0$. (This parameter $|\lambda|$ nevertheless has a definite physical meaning: it is the "coupling constant" in the quantum field theory corresponding to these Lagrangians. It determines the strength of the quantum corrections to the classical field theory problem we are considering here.)

Thus after the initial rescaling (5.3) we can assume that $\mu = |\lambda| = 1$. This means that there are really two types of Higgs equations, corresponding to $\lambda > 0$ and $\lambda < 0$. Since we are fixing the connection, we can drop the curvature term from the Lagrangian. We then have two variational problems; they are given by the Lagrangians

$$L^{\pm}(\phi) = \frac{1}{2}\int_M |\nabla\phi|^2 \pm \frac{1}{2}(|\phi|^2 - 1)^2 \tag{5.5}$$

and corresponding variational equations

$$\nabla^*\nabla\phi \pm (|\phi|^2 - 1)\phi = 0. \tag{5.6$^\pm$}$$

The qualitative features of the Lagrangian (5.5) depend very much on the sign of the Higgs potential, and the two problems must be treated separately.

Case 1: $\lambda < 0$. For the case with a minus sign in (5.5) and (5.6) we can apply the method of the previous section to obtain an infinite number of solutions on a manifold with symmetry.

THEOREM 5.1. Let E be a vector bundle over a compact Riemannian 4-manifold M. Suppose that a compact Lie group H acts isometrically on E, with each orbit of the induced action on M having dimension $1 \leq d < 4$, and suppose ∇ is an invariant connection. Then there is a sequence $\{\phi_k\} \in \Gamma(E)$ of smooth, H-equivariant solutions of (5.6$^-$), with $L^-(\phi_k) \to \infty$.

PROOF. We begin by directly minimizing the "energy functional"

$$E(\psi) = \frac{1}{2}\int_M |\nabla\psi|^2 + |\psi|^2 \tag{5.7}$$

on the constraint manifold $X = \{\psi \in L_H^{1,2}(E) \mid \int_M |\psi|^4 = 1\}$. All of the arguments of section 4 apply without change. Thus E has a sequence $\{\psi_k\}$ of smooth, equivariant critical points with $E(\psi_k) \to \infty$. By the symmetric criticality principle these are solutions of the variational equation

$$\nabla^*\nabla\psi_k + \psi_k = \alpha_k|\psi_k|^2\psi_k \tag{5.8}$$

for some constant (Lagrangian multiplier) α_k. Taking the inner product of (5.8) with ψ_k and integrating by parts shows that

$$2E(\psi_k) = \int_M |\nabla\psi_k|^2 + |\psi_k|^2 = \alpha_k \int_M |\psi_k|^4 = \alpha_k.$$

In particular, $\alpha_k > 0$ for all k. But then for each k, $\phi_k = \sqrt{\alpha_k}\psi_k$ satisfies (5.6$^-$) with $\|\phi_k\|_{L^4} = \sqrt{\alpha_k}$, $2E(\phi_k) = \alpha_k^2$, and hence by (5.5)

$$L^-(\phi_k) = \frac{1}{4}[\alpha_k^2 - \mathrm{Vol}(M)] \to \infty. \quad ❖$$

This proof leads to a simple picture of the Lagrangian L^- on $L_H^{1,2}(E)$. First note that $\phi = 0$ satisfies (5.6$^-$) and hence is a critical point with $L^-(0) = -\frac{1}{4}\mathrm{Vol}(M)$. A simple calculation shows that the Hessian of L^- at $\phi = 0$ is

$$H(\eta, \eta) = \int_M |\nabla\eta|^2 + |\eta|^2 = \|\eta\|_{1,2}^2.$$

Hence $\phi = 0$ is an isolated critical point and a local minimum. Next, observe that for $\alpha > 0$ the renormalization $\phi = \sqrt{\alpha}\psi$ used in the above proof gives a one-to-one correspondence between the non-zero critical points ϕ of L^- with $\|\phi\|_{L^4}^2 = \alpha$ and the critical points ψ of (5.7) on the constraint manifold $X = \{\|\psi\|_{L^4} = 1\}$. Thus the solutions $\{\phi_k\}$ obtained in Theorem 5.1 are, together with the solution $\phi = 0$, the complete set of critical points on $L_H^{1,2}(E)$. Finally, note that for each $\phi \in L_H^{1,2}(E)$ the restriction of the Lagrangian to the line through ϕ and the origin has the form $L^-(t\phi) = -at^4 + bt^2 - \frac{1}{4}$ with $a, b > 0$; the graph of this quartic is shaped like an 'M' with a local minimum at $t = 0$ and local maxima at $t = \pm 2a/b$. These facts give a picture of the graph of L^- as a function on $L_H^{1,2}(E)$—see Figure 2. There is one critical point (a local minimum) at $\phi = 0$. Around it are critical points arranged in pairs $\{\pm\phi_k\}$. As k increases the critical points move away from the origin, the values $L^-(\phi_k)$ increase, and the Hessian of L^- at ϕ_k has an increasing (finite) number of negative eigenvalues.

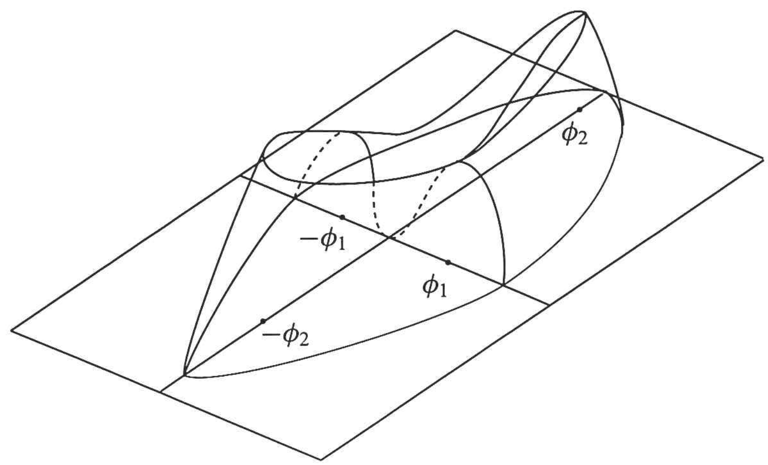

Figure 2. The graph of L^- on $L^{1,2}_H(E)$.

Case 2: $\lambda > 0$. The case of the plus sign in (5.5) and (5.6) is quite different. In contrast to Theorem 5.1, we now expect the Lagrangian (with a generic metric and connection) to have only *finitely many* critical points. To explain this difference we will first describe the qualitative behavior of the Lagrangian L^+ on $L^{1,2}$ and $L^{1,2}_H$, and then give a rigorous result.

The variational equation (5.6$^+$) still has $\phi = 0$ as a solution, so the origin in $L^{1,2}(E)$ is a critical point of L^+. The Hessian of L^+ at $\phi = 0$ is $H(\eta, \eta) = \langle \eta, (\nabla^*\nabla - 1)\eta \rangle$. This already shows that the critical point behavior of L^+ depends on the low eigenvalues of the operator $\nabla^*\nabla$. For example:

PROPOSITION 5.2. If the first eigenvalue of $\nabla^*\nabla$ is ≥ 1 then $\phi = 0$ is the only critical point of L^+ on $L^{1,2}(E)$.

PROOF. Each critical point ϕ satisfies (5.6$^+$). Hence if $\nabla^*\nabla - 1$ is a non-negative operator we have $0 = \int_M \langle \phi, (\nabla^*\nabla - 1)\phi \rangle + |\phi|^4 \geq \int_M |\phi|^4$. ❖

In general, $\nabla^*\nabla$ will have a finite number of eigenvalues ≤ 1 and the Hessian at 0 will have a finite-dimensional negative eigenspace. For each $\eta \in L^{1,2}$ the restriction of the Lagrangian to the line through η and the origin has the form $L^+(t\eta) = at^4 + bt^2 + 1/4$ where $a > 0$ and $b = \frac{1}{2}H(\eta, \eta)$; the graph of this quadric is shaped like a 'W' if $H(\eta, \eta) < 0$ and like a 'U' if $H(\eta, \eta) > 0$. Thus the graph of L^+ on $L^{1,2}(E)$ is as sketched in Figure 3.

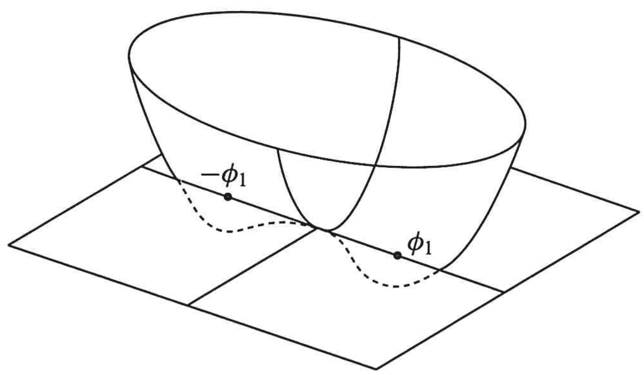

Figure 3. The graph of L^+ on $L^{1,2}_H(E)$.

Since there are only finitely many W-shaped directions we expect only finitely many critical points, with the number of critical points approximately equal to the number of eigenvalues of $\nabla^*\nabla$ that are ≤ 1. Finally, when we restrict to $L^{1,2}_H \subset L^{1,2}$ the eigenvalues shift up and we get a similar picture with probably fewer W-shaped directions and fewer critical points.

It is difficult to make rigorous conclusions from this picture because the Lagrangian L^+ again involves the borderline Sobolev embeddings. But L^+, like L^-, is analytically well-behaved on the subspace of invariant sections.

THEOREM 5.3. With E, M, and H as in Theorem 5.1, the Lagrangian L^+ given by (5.5) satisfies the PS Condition on $L^{1,2}_H(E)$.

PROOF. Suppose $\{\phi_k\}$ is a sequence in $L^{1,2}_H(E)$ with $L^+(\phi_k) \leq B$ and $\|(\nabla L^+)_{\phi_k}\| \to 0$. Using Hölder's inequality we then have

$$4B \geq 4L^+(\phi_k) \geq \int (|\phi_k|^2 - 1)^2 \geq \left(\|\phi_k\|^2_{L^4} - \sqrt{\mathrm{Vol}(M)}\right)^2.$$

It follows that $\{\phi_k\}$ is bounded in L^4 and hence—again by Hölder—in L^2. But we also have $B \geq L^+(\phi_k) \geq \|\nabla\phi_k\|^2_{L^2}$, so $\{\phi_k\}$ is bounded in $L^{1,2}$. The differential $(dL^+)_\phi$ is the left-hand side of (5.6$^+$), and the $L^{1,2}$-gradient is $(\nabla L)_\phi = \Box^{-1}(\nabla^*\nabla\phi + (|\phi|^2 - 1)\phi) = \Box^{-1}(\Box\phi - 2\phi + |\phi|^2\phi)$ (cf. (4.6)). Thus the hypothesis on the gradient is

$$\|\phi_k - 2\Box^{-1}\phi_k + \Box^{-1}(|\phi_k|^2\phi_k)\|_{1,2} \to 0. \tag{5.9}$$

Now recall that the Sobolev space $L^{-k,p}$ is by definition the dual of $L^{k,q}$ and that the Sobolev embedding theorems apply with $k < 0$ (see [P4], section 9). The composition

$$L^4_H \times L^4_H \times L^4_H \longrightarrow L^{4/3}_H \longrightarrow L^{-1,2}_H \xrightarrow{\square^{-1}} L^{1,2}_H \tag{5.10}$$

is therefore bounded (here the first map is induced by multiplication and is continuous by Hölder's inequality, and the second map is induced by the identity and is continuous by the equivariant Sobolev embedding theorem). Since $\{\phi_k\}$ is bounded in $L^{1,2}$ and $L^{1,2}_H \hookrightarrow L^4_H$ is compact, we can replace $\{\phi_k\}$ by a subsequence that converges in L^4, and then (5.10) shows that $\{\square^{-1}(|\phi_k|^2\phi_k)\}$ converges in $L^{1,2}$. Similarly, since $L^{1,2}_H \subset L^2_H$ is compact and $\square^{-1}\colon L^2 \to L^{1,2}$ is bounded, we can take a further subsequence to ensure that $\{\square^{-1}\phi_k\}$ converges in $L^{1,2}$. But then (5.9) shows that $\{\phi_k\}$ converges in $L^{1,2}$. ❖

COROLLARY 5.4. The critical set of L^+ on $L^{1,2}_H(E)$ is compact, and the highest critical value on $L^{1,2}(E)$ occurs when $\phi = 0$.

PROOF. If ϕ is critical it satisfies (5.6$^+$). Multiplying by ϕ and integrating by parts shows that $4L^+(\phi) = \int(1 - |\phi|^4) \leq \mathrm{Vol}(M)$, with equality iff $\phi = 0$. Hence any sequence $\{\phi_k\}$ in the critical set has $\{L^+(\phi_k)\}$ bounded and therefore has a convergent subsequence by the PS Condition. ❖

PROPOSITION 5.5. Let k be the number of eigenvalues of $\nabla^*\nabla$ that are < 1 and assume that no eigenvalue is exactly equal to 1. Then L^+ has at least $2k + 1$ critical points on $L^{1,2}_H(E)$.

PROOF. Replacing ϕ by $t\phi$ shows that for $0 \leq t \leq 1$

$$L^+(t\phi) \leq t^2[L^+(\phi) - a] + a$$

where $a = \mathrm{Vol}(M)/4$ is the value of the critical point $\phi = 0$. This implies that when $b \geq a$ the set $M_b = \{\phi \in L^{1,2}_H(E) \mid L^+(\phi) \leq b\}$ is contractible, since if $\phi \in M_b$, then $L^+(t\phi) \leq t^2(b - a) + a \leq a \leq b$. Corollary 5.4 implies that there is an ε such that a is the only critical value in the interval $[a - \varepsilon, a + \varepsilon]$. Our assumptions mean that the Hessian $H(\eta, \eta) = \langle \eta, (\nabla^*\nabla - 1)\eta \rangle$ of L^+ at $\phi = 0$ is non-degenerate of index k. Then Theorem 1.3(c) gives

$$* = M_{a+\varepsilon} = M_{a-\varepsilon} \cup H_k \qquad \text{and} \qquad M_{a-\varepsilon} \cap H_k = \partial H_k \cong S^{k-1},$$

and hence $H^*(M_{a-\varepsilon}) = H^*(S^{k-1})$ by the Mayer-Vietoris sequence. Now L^+ is invariant under $\phi \mapsto -\phi$ and hence descends to the quotient $\overline{M}_{a-\varepsilon}$ of $M_{a-\varepsilon}$ by $\mathbb{Z}_2$. Using the Gysin sequence with $\mathbb{Z}_2$ coefficients we find that $\overline{M}_{a-\varepsilon}$ is a $\mathbb{Z}_2$-cohomology $\mathbb{RP}^{k-1}$, so has cuplength k. Therefore $\text{Cat}(\overline{M}_{a-\varepsilon}) \geq k$. Furthermore, $L^+ \colon \overline{M}_{a-\varepsilon} \to \mathbb{R}$ satisfies the PS Condition by Theorem 5.3.

We cannot directly apply the L-S Theorem 1.4 because $\overline{M}_{a-\varepsilon}$ has a boundary. However, $\overline{M}_{a-\varepsilon}$ is complete, contains only interior critical points (with our choice of ε) and is invariant under the downward gradient flow. The proofs of the Lusternik-Schnirelman Theorem given in [P3] and [S] then apply without change—the downward gradient flow never encounters the boundary. We therefore conclude L^+ has at least k critical points on $\overline{M}_{a-\varepsilon}$. This gives $2k$ critical points in $M_{a-\varepsilon}$, plus the critical point $\phi = 0$. ❖

The eigenvalues of $\nabla^*\nabla$ vary as we change the rescaling (5.3), and this changes the number of solutions found by Proposition 5.5. We will return to this bifurcation phenomenon at the end of section 6.

The results above give a good understanding of the critical points of the Lagrangians L^+ and L^- on $L^{1,2}_H(E)$. Of course, things are more complicated on the larger space $L^{1,2}(E)$. The invariant critical points will remain critical (by the symmetric criticality principle), but their indices may increase, additional critical points may appear, and there may even be non-compact manifolds of critical points (as occurs in Example 6.1 below). The failure of the PS Condition could also mean that downward gradient flow lines may not converge. It is striking that these analytic difficulties are completely avoided by imposing symmetry.

6. Two examples

The theorems in the previous sections are rather abstract and deceptively easy to prove. To illustrate their implications we will now work out two particular examples. The first is the equation considered by Ding [D]; we include it to show how the Palais-Smale Condition can hold on the space of invariant functions, but fail in general. The second example shows that there are non-trivial Higgs fields on S^4, and shows how the number of solutions changes as we vary the parameter λ.

Example 1. Let $M = (S^4, g)$ be the unit sphere in $\mathbb{R}^5$. The group $H = S^1 \times SO(3)$ acts on $\mathbb{R}^5$ with the S^1 rotating the first two coordinates and the $SO(3)$ rotating the last three coordinates; this gives an isometric action on S^4 all whose orbits have dimension $1 \le d < 4$. Since S^4 has scalar curvature $s_g = 12$, Theorem 4.1 immediately gives the result of Ding [D].

COROLLARY 6.1. There is a sequence $\{f_k\}$ of smooth solutions of the equation

$$\Delta f + 2f = \lambda f^3 \tag{6.1}$$

on S^4 with $\|f_k\|_{L^4} = 1$ and $\lambda_k = \int |\nabla f_k|^2 + 2f_k^2 \to \infty$ as $k \to \infty$.

This equation on S^4 is particularly interesting because it is conformally invariant, as follows. Let $g' = \gamma^2 g$ be a metric conformal to g by a function $\gamma > 0$. Then the scalar curvature of g' is $s' = 6\gamma^{-3}(\Delta_g + 2)\gamma$ (see [LP], §1) and the Laplacians of g and g' are related by $\Delta' = \gamma^{-2}\Delta - 2\gamma^{-3}\langle d\gamma, d\cdot\rangle$. Hence if f satisfies (6.1) then $\gamma^{-1} f$ satisfies

$$\left(\Delta' + \frac{s'}{6}\right)(\gamma^{-1} f) = \lambda(\gamma^{-1} f)^3. \tag{6.2}$$

Now suppose that $h\colon S^4 \to S^4$ is a conformal diffeomorphism. Then $g' = h^*g = \gamma_h^2 g$ for some function $\gamma_h > 0$. By the naturality of curvature and of the Laplacian we then have $s' = h^*s$ and $\Delta' = h^*\Delta$. Hence if f satisfies (6.1) then by (6.2) $\psi = (h^{-1})^*(\gamma_h^{-1} f)$ satisfies

(6.3)
$$h^*[(\Delta + 2)\psi] = \left[\Delta' + h^*\left(\frac{s}{6}\right)\right](h^*\psi) = \left(\Delta' + \frac{s'}{6}\right)(\gamma_h^{-1} f) = h^*[\lambda\psi^3],$$

so ψ also satisfies (6.1). Thus we can construct new solutions by applying conformal transformations to the ones obtained in Corollary 6.1. In fact, the group of conformal diffeomorphisms of S^4 is $SO(5,1)$ and the above procedure yields

LEMMA 6.2. The map $\Phi\colon SO(5,1) \to \mathrm{End}(L^{1,2})$ defined by $\Phi_h(f) = (h^{-1})^*(\gamma_h^{-1} f)$ is an injective homomorphism, and each Φ_h preserves the L^4 norm.

PROOF. The equation $h^*g = \gamma_h^2 g$ implies that $\gamma_{hk} = (k^*\gamma_h)\gamma_k$, and it is then easy to check that Φ is a homomorphism. Injectivity is also easy: if $h \in \ker\Phi$ then $h^*f = \gamma_h^{-1} f$ for all f. Taking $f \equiv 1$ shows that $\gamma_h = 1$ and consequently h is the identity.

For any $h \in SO(5,1)$ the pulled back volume form satisfies $h^*dv = \gamma_h^4\, dv$ and hence by the diffeomorphism invariance of the integral

$$(6.4)\quad \int |\Phi_h(f)|^4\, dv = \int (h^{-1})^*|\gamma_h^{-1} f|^4\, dv = \int \gamma_h^{-4}|f|^4 h^*\, dv = \int |f|^4\, dv$$

and similarly

$$(6.5)\qquad \int |d(\Phi_h(f))|^2\, dv = \int |df|^2\, dv.$$

Thus Φ_h is a bounded linear endomorphism of $L^{1,2}$ that preserves the L^4 norm. ❖

COROLLARY 6.3. Each solution $f \neq 0$ obtained in Corollary 6.1 is part of a non-compact family $\mathcal{F}_f$ of solutions with constant λ.

PROOF. Fix a solution $f \neq 0$. We have seen that the family $\mathcal{F}_f = \{\Phi_h(f) \mid h \in SO(5,1)\}$ consists of solutions with a fixed λ. It remains to show non-compactness.

The stereographic projection map $\sigma : \mathbb{R}^4 \to S^4$ described in Example 2.2 is conformal, as is the multiplication map $M_t : \mathbb{R}^4 \to \mathbb{R}^4$ by $x \mapsto tx$ for each $t > 0$. Hence $h_t = \sigma \circ M_t^{-1} \circ \sigma^{-1}$ is a 1-parameter family of conformal diffeomorphisms of S^4 which fix the north and south poles $\{\pm p\}$ and are invariant under rotations fixing the poles.

Now by (6.4) and (6.5) and Holder's inequality the functions $f_t = \Phi_{h_t}(f)$ are uniformly bounded in $L^{1,2}$. If $\{f_t\}$ were a compact subfamily of $\mathcal{F}_f$ there would be a subsequence $\{f_k \mid k \to \infty\}$ converging in $L^{1,2}$, and hence in L^4. But for any set $A \subset S^4 - \{p\}$ we have, as in (6.4),

$$\lim_{k\to\infty} \int_A |f_k|^4 = \lim_{k\to\infty} \int_{M_k^{-1}\sigma^{-1}(A)} |\sigma^* f|^4 = 0.$$

Thus $f_k \to 0$ in $L^4(S^4)$, contradicting (6.4). ❖

Corollary 6.1 implies that the Palais-Smale Condition does not hold for the Lagrangian (4.1) on $L^{1,2}(S^4)$ (the family $\mathcal{F}_f$ lies in the critical set $L^{-1}(\lambda)$, while the PS Condition implies that each critical set is compact). Thus in this example the PS Condition holds on $L^{1,2}_H$, but fails on $L^{1,2}$. It should by now be clear how this happens. When t is large the function f_t is sharply concentrated around the north pole $p \in S^4$. No such bump function can be H-invariant because the H-action has no fixed points. Thus the non-compact "end" $\{f_t \mid t \gg 0\}$ of $\mathcal{F}_f$ lies outside of $L_H^{1,2}$, while $\mathcal{F}_f \cap L_H^{1,2}$ is compact.

Example 2. Again let M be the unit sphere (S^4, g) and let $H = S^1 \times SO(3)$ act as in Example 1. This induces an isometric action of H on $\Lambda^* T^* M$ which restricts to the subbundle $\Lambda_+^2 T^* M$ of self-dual 2-forms and hence gives, by the metric duality, an isometric action of H on the bundle $\Lambda_+^2 TM$.

Let P be the $k = 1$ principal $SU(2)$-bundle on S^4 ($k = 1$ is the usual notation for the bundle whose associated $\mathbb{C}^2$-vector bundle E has Chern class $c_2(E) = -1$). The adjoint bundle $\mathrm{Ad}\, P = P \times_{\mathrm{Ad}} \mathfrak{su}(2)$ is naturally identified with $\Lambda_+^2 TM$ (see [GP], section 3), and hence has an H-invariant connection ∇ coming from the Riemannian connection on TM; moreover, this is a Yang-Mills connection. We then obtain non-trivial Higgs fields in the presence of a $k = 1$ instanton on S^4.

COROLLARY 6.4. Let $M = (S^4, g)$, H and ∇ be as above. For the Lagrangians $L^{\pm}$ of (5.5)

(a) There is a sequence $\{\phi_k\} \in \Gamma(\mathrm{Ad}\, P)$ of critical points of L^- with $L^-(\phi_k) \to \infty$ as $k \to \infty$.
(b) The only critical point of L^+ on $L^{1,2}(\mathrm{Ad}\, P)$ is $\phi = 0$.

PROOF. The estimates in [TY] imply that the first eigenvalue of $\nabla^*\nabla$ on $\Lambda_+^2 TM$ is ≥ 2. The corollary follows from Theorems 5.1 and 5.2. ❖

Using the rescaling transformations (5.3) we see that Corollary 6.4(a) gives infinitely many solutions of the uncoupled Higgs equations

$$(d^\nabla)^* F^\nabla = 0$$
$$\nabla^*\nabla\phi = -\lambda(|\phi|^2 - \mu)\phi.$$

for any $\lambda < 0$ and $\mu > 0$. When $\lambda > 0$ the rescaling behavior is more complicated: as the scale changes the Higgs fields bifurcate. This can be seen in two ways.

First, one can consider the Lagrangian L^+ of (5.5) (with $\lambda = \mu = 1$) and rescale the metric on S^4. Let $\{\lambda_k\}$ denote the eigenvalues of $\nabla^*\nabla$ on $L_H^{1,2}(\mathrm{Ad}\, P)$ and let m_k be the multiplicity of λ_k. The corresponding eigenvalues on the sphere S_R of radius R are $\{R^{-2}\lambda_k\}$ with multiplicity m_k. Hence the Hessian $\nabla^*\nabla - 1$ of L^+

(i) is positive definite for $R < \sqrt{\lambda_1}$
(ii) has m_1-dimensional negative eigenspace for $\sqrt{\lambda_1} < R < \sqrt{\lambda_2}$

and so forth. Thus as R increases the Hessian has larger negative eigenspaces and by Proposition 5.5 we get more and more critical points (or compact critical manifolds).

Equivalently, one can fix the metric and μ and vary $\lambda > 0$ in the Lagrangian (5.1). For λ near zero the only critical point is $\phi = 0$, but as λ increases we again get more and more critical points or critical manifolds. In general, for the metric $R^2 g$, the Hessian operator of the Lagrangian (5.1) at $\phi = 0$ is $R^{-2}\nabla^*\nabla - \lambda\mu$, and its index grows as $\mu\lambda R^2$ increases.

COROLLARY 6.5. Let S^4_R be the standard sphere of radius R, and let H and ∇ be as above. For any N there is a constant $c(N)$ such that the YMH equations (5.2) with $\lambda > 0$ admit at least N solutions whenever $\mu\lambda R^2 > c(N)$.

Thus far, we have fixed the connection and considered solutions of the second equation in (5.2). We next turn to the full coupled equations, where both the connection and ϕ are allowed to vary.

7. The Full Yang-Mills-Higgs Equations

The Yang-Mills fields on a principal bundle $P \to M$ are the critical points of the Lagrangian

(7.1) $$YM(\nabla) = \frac{1}{2}\int_M |F^\nabla|^2$$

on the space $\mathcal{A}$ of connections. Similarly, solutions of the Yang-Mills-Higgs equations (5.2) are the critical points of the Lagrangian (5.1) on the space $\mathcal{A} \times \Gamma(E)$. These Lagrangians are invariant under the gauge group $\mathcal{G}$, and hence descend to functions on the orbits spaces $\mathcal{B} = \mathcal{A}/\mathcal{G}$ and $\mathcal{E} = \mathcal{A} \times_{\mathcal{G}} \Gamma(E)$ respectively.

Given an action of a group H on M, we can apply the program of section 3 to construct a Morse Theory on the space of invariant connections. This is carried out in [Pa2]. Here we briefly describe the setup and a result on unstable Yang-Mills, and refer the reader to [Pa2] for details. The last two theorems of this section are the main results of the present paper; they show that Yang-Mills-Higgs fields exist in abundance on S^4.

Suppose that a compact connected Lie group H acts isometrically on M. For each $h \in H$ there is a bundle automorphism $\hat{h}: P \to P$ covering h. This $\hat{h}$ is unique up to a gauge transformation, so there is a well-defined action of H on $\mathcal{B}$ given by $[\nabla] \mapsto [\hat{h}^*\nabla]$. Let $\mathcal{B}^H$ denote the fixed set of this action and let $(\mathcal{B}^*)^H$ denote the open dense set of gauge classes of irreducible connections. The sets $\mathcal{E}^H$ and $(\mathcal{E}^*)^H$ are defined similarly.

The set $\mathcal{B}^H$ has components labeled by lifts of the action. This is described in [Pa2] and goes as follows. Let Z be the center of G and let $\hat{H}$ be an extension of H by Z. A *lift* of the H-action is a homomorphism $l: \hat{H} \to \mathrm{Aut}(P)$ covering the action of H on M. Each lift l gives an action of H on $\mathcal{A}$ and hence an action on $\mathcal{B}$ that agrees with the one defined above. The lift also defines a subgroup $\mathcal{G}_l$ of the gauge group consisting of all gauge transformations that commute with the lifted action. The irreducible part of $\mathcal{B}^H$ is then a disjoint union

$$(\mathcal{B}^*)^H = \bigcup_{i\lambda \in L} (\mathcal{A}_l^*/\mathcal{G}_l) = \bigcup_{l\in L} \mathcal{B}_l^*. \tag{7.2}$$

where L labels the isomorphism classes of lifts to an action of an extension $\hat{H}$ of H. There is an analogous decomposition for $\mathcal{E}^H$: if E is associated to P by a representation ρ of G, set $Z' = Z \cap \ker \rho$ (this is the part of Z that acts a trivial gauge transformations of E). Then

$$(\mathcal{E}^*)^H = \bigcup_{l\in L'} (\mathcal{E}_l^*/\mathcal{G}_l) = \bigcup_{l\in L'} \mathcal{E}_l^*. \tag{7.3}$$

where L' labels the isomorphism classes of lifts to an action of an extension $\hat{H}'$ of H by Z'. (This corrects the statement in [Pa2] that the components are labeled by lifts of H itself).

The sets $\mathcal{B}^H$ and $\mathcal{E}^H$ are not manifolds—they have singularities at the reducible connections. Furthermore, the components of the above decomposition may intersect at the reducible connections. The simplest way to avoid these difficulties is to assume that $\mathcal{B}^H$ contains no reducible connections, so $\mathcal{B}^H = (\mathcal{B}^*)^H$. With this assumption we have a Morse theory for equivariant connections.

THEOREM 7.1 [Pa2]. Suppose that a compact connected Lie group H acts isometrically on M with all orbits having dimension $d \geq 1$. Suppose also that $\mathcal{B}^H$ contains no reducible connections. Then for each lift l

(a) $\mathscr{C}_l$ (resp., $\mathscr{B}_l$) is a smooth closed submanifold of $\mathscr{C}$ (resp., $\mathscr{B}$), and is a complete Riemannian manifold with respect to the $L^{1,2}$ metric.
(b) On $\mathscr{C}_l^H$ (resp., $\mathscr{B}_l^H$) the Yang-Mills-Higgs (resp., Yang-Mills) are smooth functions that satisfy the Palais-Smale Condition and hence satisfy Morse and Lusternik-Schnirelman theory.
(c) The critical points are smooth YMH (resp., YM) fields.

This theorem already has interesting implications for Yang-Mills fields. The absolute minima of the Yang-Mills Lagrangian $YM: \mathscr{B} \to \mathbb{R}$ occur at the self-dual (or anti-self-dual) connections. For many years it was not known whether there exist other, non-minimal Yang-Mills fields. In fact, we can use Theorem 7.1 to produce such non-minimal Yang-Mills fields.

For a specific example, we turn to the "quadrapole bundles" $E_{k,l}$ on S^4. These are a collection of $SU(2)$-equivariant bundles, indexed by pairs of integers k, l, that originally arose in quantum mechanics. They are described in detail in [SS] and [Pa2]. For our purposes they have two salient properties. First, the orbits of the underlying action of $SU(2)$ on S^4 all have dimensions 2 or 3, and second, an application of the G-Index Theorem shows that the bundle $E_{k,l}$ has Chern class $c_2 = (l^2 - k^2)/8$ and admits no invariant reducible connections and no self-dual or anti-self-dual connections unless $k = 1$ or $l = 1$. Thus we can realize each Chern class $c_2 \neq \pm 1$ by a quadrapole bundle that admits no SD or ASD connections. On the other hand, Theorem 7.1 ensures that the Yang-Mills function attains its minimum on each $E_{k,l}$. These observations yield the following result of L. Sadun and J. Segert (who prove it using O.D.E. methods).

COROLLARY 7.2 [SS1], [Pa2]. Every principal $SU(2)$ bundle P over S^4 with $c_2(P) \neq \pm 1$ admits an irreducible non-minimal Yang-Mills field.

Remark. Over the last several years the existence of non-minimal Yang-Mills fields has also been proved by a variety of other methods ([SSU], [Pa3], [W]). The solutions produced by Corollary 7.2 have been studied in detail by Sadun and Segert using computer calculations [SS2].

It should be possible to get much more from this argument. The moduli space of invariant self-dual connections is a finite-dimensional subset $\mathcal{M}^H \subset \mathscr{B}^H$, so if one could show that $\mathscr{B}^H$ has infinite cohomological dimension it would force the existence of non-minimal Yang-Mills fields (in fact, dim $\mathcal{M}^H$ is at most the dimension of $\mathcal{M}$ given by the standard index formula, so one only needs a non-zero cohomology class in high dimension). In the YMH case we carry through such an argument in Theorem 7.5 below.

A Yang-Mills field is *a fortiori* a Yang-Mills-Higgs field with $\phi \equiv 0$. In fact there are three cases in which the equations decouple, giving rather trivial solutions (∇, ϕ):

(i) ∇ is a Yang-Mills connection, and $\phi \equiv 0$.
(ii) ∇ is the trivial connection on the trivial bundle, and $\phi \equiv \sqrt{\mu}$ is the constant function.
(iii) ∇ is a Yang-Mills, E is associated to P by a representation ρ that contains a trivial representation (so $E = E' \oplus \tau$ where τ is a trivial line bundle), and $\phi \in \Gamma(\tau)$ satisfies $d^* d\phi = \pm\lambda(|\phi|^2 - \mu)\phi$.

Definition. We will refer to solutions of types (i)–(iii) as "decoupled solutions". We say that E is *nontrivial* if it is associated to P by a representation that contains no invariant lines (so type (iii) solutions cannot occur on E).

Until very recently, *no* coupled solutions were known on compact 4-manifolds. This is in sharp contrast to the Yang-Mills equations, where one has Taubes' existence theorems for self-dual solutions. From the Morse Theory perspective Taubes' method is to construct an approximate solution (by "gluing on instantons") that is near a *stable* critical point of the Yang-Mills function, and then follow the gradient flow downward to a true critical point. This approach cannot work in the Yang-Mills-Higgs case because of the following fact.

PROPOSITION 7.3 [Pa2]. For $G = SU(2)$ a stable Yang-Mills-Higgs field (∇, ϕ) on S^4 is decoupled and ∇ is SD or ASD and ϕ is constant.

This means that the problem of finding coupled YMH fields is analogous to finding *unstable* Yang-Mills fields, which perhaps explains why solutions were not found earlier. Note, however, that the Equivariant Morse Theorem 7.1 applies equally well in this YMH case.

Let P be a principal $SU(2)$ bundle over S^4 and let E be an associated bundle. We can identify P with a quadrapole bundle. The $SU(2)$ action on the quadrapole bundle induces actions on E. Again, there are no invariant reducible connections, so $\mathscr{C}^H = (\mathscr{C}^*)^H$ has a decomposition (7.3). The Morse Theorem immediately gives critical points, but now we must take care that these are not the uncoupled solutions described above. Also, as in section 5 we must treat the two cases ($\lambda > 0$ and $\lambda < 0$) of the Higgs Lagrangian (5.1) separately. Both proofs require interesting arguments beyond simply quoting the Morse Theorem.

First, for $\lambda > 0$ we have the following theorem from [Pa2].

THEOREM 7.4. Let P be a principal $SU(2)$ bundle over S^4 and let E be a non-trivial associated bundle. Then for $\lambda > 0$ and $\lambda\mu$ sufficiently large, there is a solution of the Yang-Mills-Higgs equations (5.2) on (P, E) which is neither uncoupled or reducible.

PROOF. By Theorem 7.1 there exists a minimum of the YMH function on each component of $\mathscr{E}^H$; this is a YMH field on a quadrapole bundle. We will show that it is non-trivial (i.e., $\phi \neq 0$) when $\lambda\mu$ is large.

Choose $A \in \mathscr{B}^H$ and set $C = YM(A)$. Then

$$\mathcal{M}_C^H = \{\, [A] \in \mathscr{B}^H \mid A \text{ is Yang-Mills and } YM(A) \leq C \,\}$$

is non-empty and compact because the Yang-Mills function (7.1) satisfies the PS Condition on $\mathscr{B}^H$. Let Δ_A denote the Laplacian $\nabla^*\nabla$ of the connection A acting on $L_H^{1,2}(E)$. The eigenvalues of Δ_A are continuous functions of the connection, so there is a constant M such that the first eigenvalue satisfies $\lambda_1(\Delta_A) \leq M$ for all $A \in \mathcal{M}_C^H$. Choose $\lambda\mu > M/2$. Computing the second variation of the YMH Lagrangian (5.1), we find that at each $A \in \mathcal{M}_C^H$ the first eigenfunction ψ of Δ_A satisfies

$$\text{Hess } YMH_{(A,0)}(\psi, \psi) = \int_{S^4} |\nabla^A \psi|^2 - \lambda\mu|\psi|^2 = (\lambda_1 - \lambda\mu) \int_{S^4} |\psi|^2 < 0.$$

Thus the fields $(A, \phi) \in \mathscr{E}$ with $\phi \equiv 0$ are never minimal in $\mathscr{E}^H$. ❖

Presumably, the bifurcation phenomenon described at the end of section 6 applies to the coupled equations also, so as $\lambda\mu$ increases more and more solutions appear.

THEOREM 7.5. Let P be a principal $SU(2)$ bundle over S^4 and let E be a non-trivial associated bundle. Then for $\lambda < 0$ there are infinitely many distinct solutions of the Yang-Mills-Higgs equations (5.2) on (P, E) which are neither uncoupled or reducible.

PROOF. As in Theorem 5.1 we first rescale, reducing the Lagrangian to the form

$$L^{-}(\nabla, \phi) = \frac{1}{2}\int_{S^4} |F^{\nabla}|^2 + |\nabla\phi|^2 \pm \frac{1}{2}(|\phi|^2 - 1)^2.$$

We then minimize

$$E(\phi) = \frac{1}{2}\int_{S^4} |F^{\nabla}|^2 + |\nabla\phi|^2 + \phi^2$$

on the constraint manifold

$$X = \left\{[(\nabla, \phi)] \in \mathscr{C}^H \;\middle|\; \textstyle\int_{S^4} |\phi|^4 = 1\right\}.$$

Forgetting ϕ gives a projection $X \to \mathscr{B}^H$ whose fiber S^∞ is contractible (as in Lemma 4.4). Hence X is homotopic to $\mathscr{B}^H$. If $\mathscr{B}^H$ has infinite cuplength we obtain infinitely many critical points by the Lusternik-Schnirelman Theorem 7.1. On the other hand, $E(\phi)$ descends to $\overline{X} = X/\mathbb{Z}_2$, and when $\mathscr{B}^H$ has finite cuplength the spectral sequence for the fibration

$$\begin{array}{ccc} \mathbb{RP}^\infty & \longrightarrow & \overline{X} \\ & & \downarrow \\ & & \mathscr{B}^H \end{array}$$

shows that $\overline{X}$ has infinite cuplength, so we again obtain infinitely many critical points. In either case, we get a sequence $\{\nabla^k, \phi^k\}$ of smooth $SU(2)$-invariant YMH fields with $\phi^k \neq 0$ and $L^{-}(\nabla^k, \phi^k) \to \infty$ as in the proof of Theorem 5.1. It follows that these solutions are neither uncoupled or reducible. ❖

By Proposition 7.3 these fields are unstable critical points of the YMH function: the Lagrangian, although minimal amongst equivariant fields, can be reduced by perturbations in non-equivariant directions.

BIBLIOGRAPHY

[A] T. Aubin, "Nonlinear Analysis on Manifolds. Monge-Ampere Equations," Springer-Verlag, New York, 1982.

[B] M. Berger, "Nonlinearity and Functional Analysis," Academic Press, New York, 1977.

[**D**] W. Y. Ding, *On a conformally invariant elliptic equation on* $\mathbb{R}^n$, Commun. Math. Phys. **107** (1986), 331–335.

[**GP**] D. Groisser and T. H. Parker, *The Riemannian geometry of the Yang-Mills moduli space*, Commun. Math. Phys. **112** (1987), 663–689.

[**JT**] A. Jaffe and C. H. Taubes, "Vortices and Monopoles," Birkhäuser, Boston, 1980.

[**K**] S. Kakutani, *Topological properties of the unit sphere of a Hilbert space*, Proc. Imp. Acad. Tokyo **19** (1943), 269–271.

[**LP**] J. Lee and T. H. Parker, *The Yamabe problem*, A.M.S. Bulletin **17** (1987), 37–91.

[**M**] J. Milnor, "Morse Theory," Princeton University Press, Princeton, N.J., 1963.

[**P1**] R. S. Palais, *Morse Theory on Hilbert manifolds*, Topology **2** (1963), 299–340.

[**P2**] ———, *Homotopy theory of infinite dimensional manifolds*, Topology **5** (1966), 1–16.

[**P3**] ———, *Lusternik-Schnirelman Theory on Banach manifolds*, Topology **5** (1966), 115–132.

[**P4**] ———, "Foundations of Global Nonlinear Analysis," Benjamin, New York, 1968.

[**P5**] ———, *The Principle of Symmetric Criticality*, Commun. Math. Phys. **69** (1979), 19–30.

[**Pa1**] T. H. Parker, *Gauge theories on four-dimensional Riemannian manifolds*, Commun. Math. Phys. **85** (1982), 1–40.

[**Pa2**] ———, *A Morse theory for Equivariant Yang-Mills*, Duke Math. Jour. **66** (1992), 337–356.

[**Pa3**] ———, *Non-minimal Yang-Mills Fields and Dynamics*, Invent. Math. **107** (1992), 397–420.

[**R**] P. H. Rabinowitz, *Variational methods for non-linear eigenvalue problems*, Indiana Univ. Math. J. **23** (1973), 729–754.

[**S**] J. T. Schwartz, *Generalizing the Lusternik-Schnirelman theory of critical points*, Commun. Pure Appl. Math. **17** (1964), 307–315.

[**Se**] J. P. Serre, *Homologie singulière des espaces fibrés*, Annals of Math. **54** (1951), 425–505.

[**Sm**] S. Smale, *Morse Theory and a Non-linear Generalization of the Dirichlet Problem*, Ann. Math. **80** (1964), 382–396.

[**SS1**] L. Sadun and J. Segert, *Non-self-dual Yang-Mills Fields with Quadrapole Symmetry*, Commun. Math. Phys. **145** (1992), 363-392.

[**SS2**] L. Sadun and J. Segert, *Stationary Points of the Yang-Mills Action*, preprint (1992).

[**SSU**] L. Sibner, R. Sibner, and K. Uhlenbeck, *Solutions to the Yang-Mills Equations that are not Self-dual*, Proc. Natl. Acad. Sci. USA **86** (1989), 8608-8613.

[**TY**] S. Tachibana and S. Yamaguchi, *The first proper space for p-forms in a compact Riemannian manifolds of positive curvature operator*, J. Diff. Geo. **15** (1980), 51–60.

[**U**] K. K. Uhlenbeck, *Integrals with Non-degenerate Critical Points*, Bull. Amer. Math. Soc. **76** (1970), 125–128.

[**W**] H. Y. Wang, *The Existence of Non-minimal Solutions to the Yang-Mills Equation with Group* $SU(2)$ *on* $S^2 \times S^2$ *and* $S^1 \times S^3$, J. Diff. Geo. **34** (1991), 701–768.