Compactness results for neck-stretching limits of instantons

David L. Duncan

Abstract
We prove that, under a suitable degeneration of the metric, instantons converge to holomorphic quilts. To prove the main results, we develop estimates for the Yang-Mills heat flow on surfaces and cobordisms.

Contents
1 Introduction ........................................... 1
2 Background, notation and conventions .................. 3
  2.1 Gauge theory ........................................ 4
  2.2 The complexified gauge group ....................... 7
3 Compactness for products $S \times \Sigma$ ................... 11
  3.1 Small curvature connections in dimension 2 ....... 15
  3.2 Proof of Lemma 3.5 ................................ 31
  3.3 Proof of Theorem 3.3 ............................... 34
4 Compactness for cylinders $R \times Y$ ..................... 36
  4.1 The heat flow on cobordisms ....................... 41
  4.2 Uniform elliptic regularity .......................... 46
  4.3 Proof of Lemma 4.2 ................................ 65
  4.4 Proof of Theorem 4.1 ............................... 67
5 Perturbations ........................................... 73

1 Introduction

In Donaldson theory one obtains invariants of 4-manifolds $Z$ by counting instantons on a suitable bundle over $Z$. On the other hand, in symplectic geometry one can obtain invariants of $Z$ by counting $J$-holomorphic curves in a suitable symplectic manifold associated to $Z$. For example, when $Z = S \times \Sigma$ is a product of surfaces, then the relevant $J$-holomorphic curves are maps from $S$ into the moduli space of
flat connections on $\Sigma$. There is a heuristic argument, dating back to Atiyah [1], that suggests these invariants are the same. Indeed, consider the case $Z = S \times \Sigma$, and fix a metric in which the $S$-fibers are very large (equivalently, the $\Sigma$-fibers are very small). Motivated by Atiyah’s terminology, we refer to this type of metric as a ‘neck-stretching metric’, and use $e^{-2}$ to denote the volume of $S$. Then with respect to such a metric, the instanton equation splits into two equations: the first is essentially the holomorphic curve equation, and the second shows that the curvature of the instanton is bounded by $e^2$ in the $\Sigma$-directions. In particular, taking $e$ to zero, the instanton equation formally recovers the $J$-holomorphic curve equation.

In this paper we formalize the heuristic of the previous paragraph in the case when $Z = S \times \Sigma$, and also when $Z = \mathbb{R} \times Y$, where $Y$ has positive first Betti number. In each case, we prove that if $e_\nu$ is a sequence converging to zero and $A_\nu$ is an instanton with respect to the $e_\nu$-metric, then a subsequence of the $A_\nu$ converges, in a suitable sense, to a $J$-holomorphic curve. The main convergence results are stated precisely in Theorem 3.3 for $S \times \Sigma$, and Theorem 4.1 for $\mathbb{R} \times Y$.

When $Z = \mathbb{R} \times Y$ our main theorem extends results of Dostoglou-Salamon in [8] [9], where they consider the special case where $Y$ is a mapping torus. The case $Z = S \times \Sigma$ serves as a model for the technically more difficult analysis necessary for $\mathbb{R} \times Y$. The results of this paper can be extended quite naturally to more general 4-manifolds, however we leave a full treatment of such extensions to future work.

The proofs of the main results proceed roughly as follows: As mentioned, when $e$ is small each instanton $A$ has curvature that is small in the $\Sigma$-direction. Motivated by Donaldson [5], our strategy is to use an analytic Narasimhan-Seshadri correspondence on $\Sigma$ to map each restriction $A|_{\{s\} \times \Sigma}$ to some nearby flat connection $\text{NS}(A|_{\{s\} \times \Sigma})$ on $\Sigma$. This correspondence preserves the equations in the sense that the limiting connection $\text{NS}(A|_{\{s\} \times \Sigma})$ is holomorphic. Then the convergence result for the case $Z = S \times \Sigma$ essentially follows from Gromov’s compactness theorem for holomorphic curves.

In the case when $Z = \mathbb{R} \times Y$, the situation is considerably more difficult. To describe the difficulty we need to digress to discuss the neck-stretching metric used. Fix a circle-valued Morse function for $Y$. Then we stretch the metric only in a fixed region that is bounded away from the critical points. In this set-up, we expect that the limiting $J$-holomorphic curve will now have Lagrangian boundary conditions, where the Lagrangians are associated to the Morse critical points in a natural way. Now suppose $A$ is an instanton with respect to the metric just described. Then $\text{NS}(A|_{\{s\} \times \Sigma})$ continues to be holomorphic in this case, but the essential difficulty is that it only has approximate Lagrangian boundary conditions. The standard Gromov compactness theorem breaks down when one does not have Lagrangian boundary conditions on the nose. We get around this using the following ingredients. First, we establish several $C^1$-estimates for the map $\text{NS}$ (see Section 3.1.3), and then $W^{2,2}$-estimates for instantons $A$ (see Section 4.2). These allow us to control the behavior of the holomorphic curves $\text{NS}(A|_{\{s\} \times \Sigma})$ near the boundary. The last ingredient is provided by the Yang-Mills heat flow on 3-manifolds with boundary. This gives us candidates for what the boundary conditions should be, if they were Lagrangian. Combining this with the $C^1$- and $W^{2,2}$-estimates allows us to reprove a version of
Gromov’s compactness theorem for almost Lagrangian boundary conditions.

In Section 2 we introduce our notation and conventions. We begin Section 3 by precisely stating our compactness result for $S \times \Sigma$. We then discuss the Narasimhan-Seshadri correspondence and develop the necessary estimates. We conclude Section 3 with the proof of the compactness result in the case of $S \times \Sigma$. Section 4 opens by stating the compactness result for $\mathbb{R} \times Y$. We then discuss the heat flow on 3-manifolds with boundary, and prove the compactness theorem for $\mathbb{R} \times Y$.

The choice of neck-stretching metric used for $\mathbb{R} \times Y$ is by no means new. Indeed, the use of a (real-valued) Morse function on $Y$ and the associated metric is again due to Atiyah [1]. In this case, the conjecture relating the instanton and symplectic invariants of $\mathbb{R} \times Y$ came to be known as the Atiyah-Floer conjecture (due to considerations on the instanton side, one typically requires that $Y$ is a homology 3-sphere). Unfortunately, at the time of writing, the Atiyah-Floer conjecture is ill-posed due to complications on the symplectic side arising from reducible connections; however, see [28]. Circle-valued Morse functions on $Y$ were introduced as a mechanism for ruling out these troublesome reducible connections (the existence of a suitable circle-valued Morse function requires that $Y$ have positive first Betti number). This approach appears, in various ways, in work by Dostoglou-Salamon [7], and then later in Wehrheim [33] and Wehrheim-Woodward [35]. The upshot with this circle-valued approach is that the symplectic Floer homology and instanton Floer homology of $Y$ are both well-defined; see [15], [35], [13], [14] and [3]. The relevant conjecture identifying these homology theories has come to be known as the quilted Atiyah-Floer conjecture [10], with the term ‘quilted’ reflecting the possible existence of Morse critical points.

Note: Upon completion of this project, the author learned of an earlier work [25] wherein Nishinou proves the result of Theorem 3.3 using essentially the same techniques used here. Included here are many details that are not present in [25]. Moreover, the technically more difficult Theorem 4.1 appears to be new.

Acknowledgments: The author is grateful to his thesis advisor Chris Woodward for his insight and valuable suggestions. In addition, the author benefited greatly from discussions with Katrin Wehrheim, via Chris Woodward, of her unpublished work [33] on the Atiyah-Floer conjecture. This unpublished work analyzes various bubbles that appear in the limit of instantons with Lagrangian boundary conditions with degenerating metrics, and it outlines the remaining problems in such an approach. The present paper takes a different approach that avoids instantons with Lagrangian boundary conditions. This work was partially supported by NSF grants DMS 0904358 and DMS 1207194.

2 Background, notation and conventions

Throughout this section $X$ will denote an oriented manifold. Given a fiber bundle $F \to X$, we will use $\Gamma(F)$ to denote the space of smooth sections. If $V \to X$ is a vector bundle, then we will write $\Omega^k(X, V) := \Gamma(\Lambda^k T^* X \otimes V)$ for the space of $k$-forms with values in $V$, and we set
\[ \Omega^*(X, V) := \bigoplus_k \Omega^k(X, V). \]

If \( V \to X \) is the trivial rank-1 bundle then we will write \( \Omega^k(X) \) for \( \Omega^k(X, V) \).

Given a metric on \( X \) and a connection on \( V \), we can define Sobolev norms on \( \Gamma(V) \) in the usual way. We will use \( W^{k,p}(X, V) \) to denote the completion of \( \Gamma(V) \) with respect to the \( W^{k,p} \)-norm. We note that the usual Sobolev embedding and compactness statements for \( W^{k,p}(X, \mathbb{R}) \) hold equally well for \( W^{k,p}(X, V) \) when \( V \) is finite-dimensional. If \( V \) is infinite-dimensional then we assume \( V \) is a Banach bundle and note that the usual embeddings hold here as well, but the compactness of these embeddings is very subtle question. We will typically not keep track of bundle \( V \) in the notation for the Sobolev norms. For example, we will use the same symbol \( \| \cdot \|_{L^2(X)} \) for the norm on \( \Gamma(TX) \) as for the norm on \( \Gamma(T^*X \otimes V) \).

Now suppose \( P \to X \) is a principal \( G \)-bundle, where \( G \) is a Lie group with Lie algebra \( \mathfrak{g} \). Given a manifold \( M \) and homomorphism \( G \to \text{Diff}(M) \), we can define the associated bundle \( P \times_G M := (P \times M) / G \). This is naturally a fiber bundle over \( X \) with fiber \( M \). If \( M \) has additional structure, and the image of \( G \to \text{Diff}(M) \) respects this structure, then this additional structure is passed to the fiber bundle \( P \times_G M \). For example, when \( V \) is a vector space and \( G \to \text{GL}(V) \subset \text{Diff}(V) \) is a representation, then \( P(V) := P \times_G V \) is a vector bundle. The most important example for us comes from the adjoint representation \( G \to \text{GL}(\mathfrak{g}) \subset \text{Diff}(\mathfrak{g}) \). This respects the Lie algebra structure, and so the adjoint bundle \( P(\mathfrak{g}) := P \times_G \mathfrak{g} \) is a vector bundle with a Lie bracket on the fibers. This fiber-Lie bracket combines with the wedge product to determine a graded Lie bracket on the vector space \( \Omega^*(X, P(\mathfrak{g})) \), and we denote this by \( \mu \Box v \mapsto [\mu \wedge v] \). Similarly, if \( \mathfrak{g} \) is equipped with an Ad-invariant inner product \( \langle \cdot, \cdot \rangle \), then this determines an inner product on the fibers of \( P(\mathfrak{g}) \) and moreover combines with the wedge to form a graded bilinear map \( \Omega^*(X, P(\mathfrak{g})) \otimes \Omega^*(X, P(\mathfrak{g})) \to \Omega^*(X) \) that we denote by \( \mu \Box v \mapsto \langle \mu \wedge v \rangle \).

### 2.1 Gauge theory

Let \( P \to X \) be a principal \( G \)-bundle. We will write

\[ \mathcal{A}(P) = \left\{ A \in \Omega^1(P, \mathfrak{g}) \middle| (g_p)^* A = \text{Ad}(g^{-1})A, \quad \forall g \in G \right\} \]

for the space of connections on \( P \). Here \( g_p \) (resp. \( \xi_p \)) is the image of \( g \in G \) (resp. \( \xi \in \mathfrak{g} \)) under the map \( G \to \text{Diff}(P) \) (resp. \( \mathfrak{g} \to \text{Vect}(P) \)) afforded by the group action. It follows that \( \mathcal{A}(P) \) is an affine space modeled on \( \Omega^1(X, P(\mathfrak{g})) \), and we denote the affine action by \( (V, A) \mapsto A + V \). In particular, \( \mathcal{A}(P) \) is a smooth (infinite dimensional) manifold with tangent space \( \Omega^1(X, P(\mathfrak{g})) \). Each connection \( A \in \mathcal{A}(P) \) determines a covariant derivative \( d_A : \Omega^*(X, P(\mathfrak{g})) \to \Omega^{*+1}(X, P(\mathfrak{g})) \) and a curvature (2-form) \( F_A \in \Omega^2(X, P(\mathfrak{g})) \). These satisfy \( d_{A+V} = d_A + [V \wedge \cdot] \) and \( F_{A+V} = F_A + d_AV + \frac{1}{2} [V \wedge V] \). We say that a connection \( A \) is irreducible if \( d_A \) is injective on 0-forms.
Given a metric on $X$, we can define the formal adjoint $d_A^*$, which satisfies
\[
(d_A V, W)_{L^2} = (V, d_A^* W)_{L^2}
\]
for all compactly supported $V, W \in \Omega^*(X, P(g))$. Here $(\cdot, \cdot)_{L^2}$ is the $L^2$-inner product coming from the metric on $X$.

A connection $A$ is flat if $F_A = 0$. We will denote the set of flat connections on $P$ by $\mathcal{A}_{\text{flat}}(P)$. If $A$ is flat then $\text{im} \, d_A \subseteq \ker d_A$ and we can form the harmonic spaces
\[
H^k_A := H^k_A(X, P(g)) := \frac{\ker (d_A|\Omega^k(X, P(g)))}{\text{im} (d_A|\Omega^{k-1}(X, P(g)))}, \quad H^*_A := \bigoplus_k H^k_A.
\]
Suppose $X$ is compact with (possibly empty) boundary, and let $\partial$ denote the restriction $\Omega^*(X, P(g)) \to \Omega^*(\partial X, P(g)|_{\partial X})$. Then the Hodge isomorphism \cite[Theorem 6.8] says
\[
H^* = \ker (d_A \oplus d_A^* \oplus \partial*), \quad \Omega^*(X, P(g)) \cong H^* \oplus \text{im} (d_A) \oplus \text{im} (d_A^*|_{\partial*}), \tag{1}
\]
for any flat connections $A$ on $X$, where the summands on the right are $L^2$-orthogonal. We will treat these isomorphisms as identifications. From the first isomorphism in \cite{1} we see that $H^* = \ker d_A$ is finite dimensional since $d_A \oplus d_A^*$ is elliptic. We will use $\text{proj}_A : \Omega^*(X, P(g)) \to H^*$ to denote the projection; see Lemma 3.10 for an extension to the case where $A$ is not flat.

**Example 2.1.** Suppose $X = \Sigma$ is a closed, oriented surface equipped with a metric. Then the pairing $\langle \mu, \nu \rangle := \int_{\Sigma} \langle \mu \wedge \nu \rangle$ is a symplectic form on the vector space $\Omega^1(X, P(g))$. Note that changing the orientation on $\Sigma$ replaces $\omega$ by $-\omega$. On surfaces, the Hodge star squares to $-1$ on 1-forms and so defines a complex structure on $\Omega^1(X, P(g))$. It follows that the triple $(\Omega^1(X, P(g)), \star, \omega)$ is Kähler. If $A \in \mathcal{A}(P)$ is flat, then $H^1 = \Omega^1(X, P(g))$ is a finite-dimensional Kähler subspace.

Now suppose $X$ is 4-manifold. Then on 2-forms the Hodge star squares to the identity, and it has eigenvalues $\pm 1$. Denoting by $\Omega^\pm(X, P(g))$ the $\pm 1$ eigenspace of $\star$, we have an $L^2$-orthogonal decomposition
\[
\Omega^2(X, P(g)) = \Omega^+(X, P(g)) \oplus \Omega^-(X, P(g)).
\]
The elements of $\Omega^-(X, P(g))$ are called anti-self dual 2-forms. A connection $A \in X$ is said to be anti-self dual (ASD) or an instanton if its curvature $F_A \in \Omega^-(X, P(g))$ is an anti-self dual 2-form; that is, if $F_A + \star F_A = 0$. Similarly, a connection $A$ is self dual if $F_A \in \Omega^+(X, P(g))$.

The space of connections $\mathcal{A}(P)$ admits a function
\[
\mathcal{YM}_P : \mathcal{A}(P) \to \mathbb{R}, \quad A \mapsto \int \frac{1}{2} |F_A|^2_{L^2}
\]
called the Yang-Mills functional. Obviously, the global minimizers of this function are the flat connections, when they exist. However, in high dimensions the existence
of flat connections is rare due to topological obstructions. In dimension four the minimizers are the self dual or ASD connections.

A gauge transformation is an equivariant bundle map \( U : P \to P \) covering the identity. The set of gauge transformations on \( P \) forms a Lie group, called the gauge group, and is denoted \( \mathcal{G}(P) \). This is naturally a Lie group with Lie algebra \( \Omega^0(X, \mathcal{P}(\mathfrak{g})) \) under the map \( R \mapsto \exp(-R) \), where \( \exp : \mathfrak{g} \to G \) is the Lie-theoretic exponential.

The gauge group acts on the left on the space \( \mathcal{A}(P) \) by pulling back by the inverse:

\[
(U, A) \mapsto U A := (U^{-1})^* A,
\]

for \( U \in \mathcal{G}(P), A \in \mathcal{A}(P) \). In terms of a faithful matrix representation of \( G \) we can write this as \( U^* A = U^{-1} A U + U^{-1} dU \), where the concatenation appearing on the right is just matrix multiplication, and \( dU \) is the linearization of \( U \) when viewed as a \( G \)-equivariant map \( P \to G \). The infinitesimal action of \( \mathcal{G}(P) \) at \( A \in \mathcal{A}(P) \) is

\[
\Omega^0(X, \mathcal{P}(\mathfrak{g})) \to \Omega^1(X, \mathcal{P}(\mathfrak{g})), \quad R \mapsto d_A R
\]

More generally, the derivative of \( [\cdot] \) at \( (U, A) \) is

\[
U \Omega^0(X, \mathcal{P}(\mathfrak{g})) \times \Omega^1(X, \mathcal{P}(\mathfrak{g})) \to \Omega^1(M, \mathcal{P}(\mathfrak{g}))
\]

\[
(U R, V) \mapsto \text{Ad}(U) d_A R + \text{Ad}(U) V.
\]

where, in writing this expression, we have chosen a faithful matrix representation \( G \to \text{GL}(V) \). This allows us to view \( \mathcal{G}(P) \) and its Lie algebra \( \Omega^0(X, \mathcal{P}(\mathfrak{g})) \) as subspaces of the same algebra \( \Gamma(\mathcal{P}(\text{End}(V))) \), and so it makes sense to identify the tangent space of \( \mathcal{G}(P) \) at \( U \) with \( U \Omega^0(X, \mathcal{P}(\mathfrak{g})) \), as we have done in \( [4] \).

The gauge group also acts on the left on \( \Omega^*(X, \mathcal{P}(\mathfrak{g})) \) by the pointwise adjoint action. The curvature of \( A \in \mathcal{A}(P) \) transforms under \( U \in \mathcal{G}(P) \) by \( F_{U^* A} = \text{Ad}(U^{-1}) F_A \). This shows that \( \mathcal{G}(P) \) restricts to an action on \( \mathcal{A}_{\text{flat}}(P) \) and, in 4-dimensions, the instantons.

In general, we will tend to use capital letters \( A, U \) to denote connections and gauge transformations on 4-manifolds \( Z \), lower case letters \( a, u \) to denote connections and gauge transformations on 3-manifolds \( Y \), and lower case Greek letter \( a, \mu \) to denote gauge transformations on surfaces \( \Sigma \).

We will be interested in the case \( G = \text{PU}(r) \) for \( r \geq 2 \). We equip the Lie algebra \( \mathfrak{g} = su(r) \subset \text{End}(\mathbb{C}^r) \) with the inner product \( \langle \mu, \nu \rangle := -\kappa_r \text{tr}(\mu \cdot \nu) \); here \( \kappa_r > 0 \) is arbitrary, but fixed. On manifolds \( X \) of dimension at most 4, the principal \( \text{PU}(r) \)-bundles \( P \to X \) are classified, up to bundle isomorphism, by two characteristic classes \( t_2(P) \in H^2(X, \mathbb{Z}) \) and \( q_4(P) \in H^4(X, \mathbb{Z}) \). These generalize the 2nd Stiefel-Whitney class and 1st Pontryagin class, respectively, to the group \( \text{PU}(r) \); see [38].

When \( X \) is a closed, oriented 4-manifold, there is also a Chern-Weil formula

\[
q_4(P) = -\frac{r}{4\pi^2 \kappa_r} \int_X \langle F_A \wedge F_A \rangle
\]

which holds for any connection \( A \in \mathcal{A}(P) \).
Consider a principal $PU(r)$-bundle $P \to X$ where we assume $\dim(X) \leq 3$. Then there are maps

$$\eta : \mathcal{G}(P) \to H^1(X, \mathbb{Z}_r), \quad \deg : \mathcal{G}(P) \to H^3(X, \mathbb{Z})$$

called the \textbf{parity} and \textbf{degree}. These detect the connected components of $\mathcal{G}(P)$ in the sense that $u$ can be connected to $u'$ by a path if and only if $\eta(u) = \eta(u')$ and $\deg(u) = \deg(u')$. We denote by $\mathcal{G}_0(P)$ the identity component of $\mathcal{G}(P)$. See [11].

Suppose $X = \Sigma$ is a closed, connected, oriented surface, and $P \to \Sigma$ is a principal $PU(r)$-bundle with $\tau_2(P) \mid \Sigma \in \mathbb{Z}_r$ a generator. It can be shown that all flat connections on $P$ on irreducible, and that $\mathcal{G}_0(P)$ acts freely on $\mathcal{A}_{\text{flat}}(P)$. Moreover, one has that

$$M(P) := \mathcal{A}_{\text{flat}}(P) / \mathcal{G}_0(P)$$

is a compact, simply-connected, smooth manifold with tangent space at $[a] \in M(P)$ canonically identified with $H^1_\alpha$ for any choice of representative $\alpha \in [a]$. It follows from Example 2.1 that $M(P)$ is a symplectic manifold, and any metric on $\Sigma$ determines an almost complex structure $J_\Sigma$ on $M(P)$ that is compatible with the symplectic form. See [35] for more details regarding these assertions.

Now suppose $Y_{ab}$ is an oriented elementary cobordism between closed, connected, oriented surfaces $\Sigma_a$ and $\Sigma_b$. Fix a $PU(r)$-bundle $Q_{ab} \to Y_{ab}$ with $\tau_2(Q_{ab}) \mid \Sigma_a$ a generator of $\mathbb{Z}_r$. Then the flat connections on $Q_{ab}$ are irreducible, and the quotient $\mathcal{A}_{\text{flat}}(Q_{ab}) / \mathcal{G}_0(Q_{ab})$ is a finite-dimensional, simply-connected, smooth manifold. Restricting to the two boundary components induces an embedding

$$\mathcal{A}_{\text{flat}}(Q_{ab}) / \mathcal{G}_0(Q_{ab}) \hookrightarrow M(Q_{\mid \Sigma_a}) \times M(Q_{\mid \Sigma_b}),$$

and we let $L(Q_{ab})$ denote the image. It follows that $L(Q_{ab}) \subset M(Q_{\mid \Sigma_a})^{-} \times M(Q_{\mid \Sigma_b})$ is a smooth Lagrangian submanifold [35], where the superscript in $M(Q_{\mid \Sigma_a})^{-}$ means that we have replaced the symplectic form with its negative.

More generally, if $G$ is a compact Lie group, and $P \to X$ is a principal $G$-bundle, then we can consider the space $\mathcal{A}_{\text{flat}}(P) / \mathcal{G}(P)$. (Note that we are not quotienting by the identity component here, so this will not be $M(P)$ when $G = PU(r)$ and $X = \Sigma$.) This space has a natural topology, and it follows from Uhlenbeck’s compactness theorem that this space is compact when $X$ is compact. However, it is rarely a smooth manifold.

### 2.2 The complexified gauge group

Let $G$ be a compact, connected Lie group and fix a faithful Lie group embedding $\rho : G \hookrightarrow U(n)$ for some $n$. We identify $G$ with its image in $U(n)$. Using the standard representation of $U(n)$ on $\mathbb{C}^n$, define $E := P \times_G \mathbb{C}^n$. The standard (real) inner product and complex structure on $\mathbb{C}^n$ are preserved by $U(n)$, and hence by $G$. It follows that there is an induced (real) inner product $\langle \cdot, \cdot \rangle$ and a complex structure $J_E$ on $E$, and that these are compatible in the sense that $\langle J_E \xi, J_E \eta \rangle = \langle \xi, \eta \rangle$ for all $\xi, \eta \in \Omega^0(\Sigma, E)$. 

7
Suppose \((\Sigma, j_\Sigma)\) is a Riemann surface. We will write \(\Omega^{k,l}(\Sigma, E)\) for the smooth \(E\)-valued forms of type \((k,l)\). Observe that \(j_\Sigma\) acts by the Hodge star on 1-forms. Consider the space

\[
C(E) := \left\{ \mathcal{D} : \Omega^0(\Sigma, E) \to \Omega^{0,1}(\Sigma, E) \mid \mathcal{D}(f \xi) = f(\mathcal{D}(\xi)) + (\partial f) \xi, \quad \text{for } \xi \in \Omega^0(\Sigma, E), f \in \Omega^0(\Sigma) \right\}
\]

of Cauchy-Riemann operators on \(E\). This can be naturally identified with the space of holomorphic structures on \(E\) (see [23, Appendix C]). Each element \(\mathcal{D} \in C(E)\) has a unique extension to an operator \(\mathcal{D} : \Omega^k(\Sigma, E) \to \Omega^{k+1}(\Sigma, E)\) satisfying the Leibniz rule.

Consider the space of \(C\)-linear covariant derivatives on \(E\):

\[
\left\{ D : \Omega^0(\Sigma, E) \to \Omega^1(\Sigma, E) \mid D(f \xi) = f(D\xi) + (df)\xi, \quad \text{for } \xi \in \Omega^0(\Sigma, E), f \in \Omega^0(\Sigma) \right\}
\]

There is a natural map from this space onto \(C(E)\) defined by

\[
D \mapsto \frac{1}{2} (D + J_E D \circ j_\Sigma)
\]

(6)

Here and below we are using the symbol \(\circ\) to denote composition of operators. For example, if \(M : \Omega(\Sigma, E) \to \Omega(\Sigma, E)\) is a derivation we define \(M \circ j_\Sigma : \Omega(\Sigma, E) \to \Omega(\Sigma, E)\) to be the derivation given by the formula \(i_X ((M \circ j_\Sigma) \xi) := i_{j_X}(M \xi)\); here \(i_X\) is contraction with a vector \(X\). The map (6) is surjective, and restricts to a \(C\)-linear isomorphism on the set

\[
\mathcal{A}(E) := \left\{ D : \Omega^0(\Sigma, E) \to \Omega^1(\Sigma, E) \mid D(f \xi) = f(D\xi) + (df)\xi, \quad D(\xi, \eta) = \{D\xi, \eta\} + \{\xi, D\eta\}, \quad \text{for } \xi, \eta \in \Omega^0(\Sigma, E), f \in \Omega^0(\Sigma) \right\}
\]

of Hermitian \(C\)-linear covariant derivatives on \(E\).

Let \(P(g)\) denote the complexification of the vector bundle \(P(g)\). Then we have bundle inclusions

\[
P(g) \subset P(g)^C \subset \text{End}(E),
\]

where \(\text{End}(E)\) is the bundle of complex linear automorphisms of \(E\) and the latter inclusion is induced by the embedding \(\rho\). Each connection \(\alpha \in \mathcal{A}(P)\) induces a covariant derivative \(d_{\alpha,\rho} : \Omega^k(\Sigma, E) \to \Omega^{k+1}(\Sigma, E)\), and so we have a map \(\mathcal{A}(P) \to \mathcal{A}(E)\). Furthermore, this map is an embedding of \(\Omega^1(\Sigma, P(g))\)-affine spaces, where \(\Omega^1(\Sigma, P(g))\) acts on \(\mathcal{A}(E)\) via the inclusion \(\Omega^1(\Sigma, P(g)) \subset \Omega^1(\Sigma, \text{End}(E))\). In particular, restricting to the image of \(\mathcal{A}(P)\) in \(\mathcal{A}(E)\), the map (5) becomes an embedding

\[
\mathcal{A}(P) \longrightarrow C(E), \quad \alpha \longmapsto \delta_\alpha := \frac{1}{2} (d_{\alpha,\rho} + J_E d_{\alpha,\rho} \circ j_\Sigma)
\]

(7)

The image of (7) is the set of covariant derivatives that preserve the \(G\)-structure, and we denote this image by \(\mathcal{C}(P)\). See [23, Appendix C] for the case when \(G = U(n)\). The space \(\mathcal{C}(P)\) is an affine space modeled on \(\Omega^{0,1}(\Sigma, P(g)^C)\). Similarly, \(\mathcal{A}(P)\) is
an affine space modeled on $\Omega^1(\Sigma, P(g))$, and \( \tilde{\alpha} \) intertwines these affine actions, where we identify $\Omega^1(\Sigma, P(g))$ with $\Omega^1(\Sigma, P(g)^C)$ by sending $\mu$ to its anti-linear part $\mu^{0,1} := \frac{1}{2} (\mu + J_E \mu \circ j_\Sigma)$. To summarize, we have a commutative diagram

$$
\begin{array}{ccc}
\mathcal{A}(P) & \xrightarrow{\cong} & C(P) \\
\downarrow & & \downarrow \\
\mathcal{A}(E) & \xrightarrow{\cong} & C(E)
\end{array}
$$

where the vertical arrows are inclusions and everything is equivariant with respect to the action of $\Omega^1(\Sigma, P(g))$.

Let $\alpha \in \mathcal{A}(P)$ be a connection on $P$ with curvature $F_\alpha \in \Omega^2(\Sigma, P(g))$. Consider the associated covariant derivative $d_{\alpha, \rho} \in \mathcal{A}(E)$ as well as its curvature $F_{\alpha, \rho} = d_{\alpha, \rho} \circ d_{\alpha, \rho} \in \Omega^2(\Sigma, \text{End}(E))$. Since the representation $\rho$ is faithful, we have pointwise estimates of the form

$$
c |F_{\alpha, \rho}| \leq |F_\alpha| \leq C |F_{\alpha, \rho}|;
$$

this allows us to discuss curvature bounds in terms of either $F_\alpha$ or $F_{\alpha, \rho}$.

To define the complexified gauge group, we first recall some basic properties of the complexification of compact Lie groups. See [19] or [17] for more details on this material. Since $G$ is compact and connected, there is a connected complex group $G^C$ and an embedding $G \hookrightarrow G^C$ such that $G$ is a maximal compact subgroup of $G^C$, and the Lie algebra $g^C = \text{Lie}(G^C)$ is the complexification of $g = \text{Lie}(G)$. This group $G^C$ is unique up to natural isomorphism and is called the complexification of $G$.

We may assume that the representation $\rho : G \to \text{U}(n)$ from above extends to an embedding $G^C \hookrightarrow \text{GL}(C^n)$, and we identify $G^C$ with its image (see [17] Proof of Theorem 1.7]). Then we have $G = \{ u \in G^C | u^t u = \text{Id} \}$, where $u^t$ denotes the conjugate transpose on $\text{GL}(C^n)$. It follows from the standard polar decomposition in $\text{GL}(C^n)$ that we can write $G^C = \{ g \exp(i\zeta) | g \in G, \zeta \in g \}$, and this decomposition is unique. The same holds true if we replace $g \exp(i\zeta)$ by $\exp(i\zeta) g$. It is then immediate that

$$
g \exp(i\zeta) = \exp(i \text{Ad}(g) \zeta) g 
$$

for all $g \in G$ and all $\zeta \in g$.

We can now define the complexified gauge group on $P$ to be

$$
G^C(P) := \Gamma(\text{P} \times_G G^C).
$$

As in the real case, we may identify $\Omega^0(\Sigma, P(g)^C)$ with the Lie algebra of $(G^C(P))^C$ via the map

$$
\zeta \mapsto \exp(-\zeta),
$$

hence the Lie group theoretic exponential map on $G^C(P)$ is given pointwise by the exponential map on $G^C$. It follows by the analogous properties of $G^C$ that each element of $G^C(P)$ can be written uniquely in the form
\[ g \exp(i\zeta) \]

for some \( g \in \mathcal{G}(P) \) and \( \zeta \in \Omega^0(\Sigma, P(g)) \), and \( (\mathcal{G} \ominus \mathcal{G}) \) continues to hold with \( g, \zeta \) interpreted as elements of \( \mathcal{G}(P), \Omega^0(\Sigma, P(g)) \), respectively.

The complexified gauge group acts on \( C(P) \) by

\[
\mathcal{G}(P)^C \times C(P) \longrightarrow C(P), \quad (\mu, D) \longmapsto \mu \circ D \circ \mu^{-1}
\]

(11)

Viewing \( \mathcal{G}(P) \) as a subgroup of \( \mathcal{G}(P)^C \) in the obvious way, then the identification \( (\mathcal{G} \ominus \mathcal{G}) \) is \( \mathcal{G}(P) \)-equivariant. We can then use \((\mathcal{G} \ominus \mathcal{G})\) and \((\mathcal{G} \ominus \mathcal{G})\) to define an action of the larger group \( \mathcal{G}(P)^C \) on \( A(P) \), extending the \( \mathcal{G}(P) \)-action. We denote the action of \( \mu \in \mathcal{G}(P)^C \) on \( a \) by \( (\mu^{-1})^*a \) or simply \( \mu a \) when there is no room for confusion. Explicitly, the action on \( A(P) \) takes the form

\[
d_{(\mu^{-1})^*a, \rho} = (\mu^*)^{-1} \circ \partial_a \circ \mu^* + \mu \circ \partial_a \circ \mu^{-1}.
\]

where the dagger is applied pointwise. In particular, the infinitesimal action at \( a \in A(P) \) is

\[
\Omega^0(\Sigma, P(g))^C \to \Omega^1(\Sigma, P(g)), \quad (\zeta + i\xi) \mapsto d_{a, \rho} \zeta + *d_{a, \rho} \xi
\]

(12)

More generally, the derivative of the map \( (\mu, a) \mapsto (\mu^{-1})^*a \) at \( (\mu, a) \) with \( \mu \in \mathcal{G}(P) \) (an element of the real gauge group) is the map

\[
\mu \left( \Omega^0(\Sigma, P(g))^C \oplus i\Omega^0(\Sigma, P(g)) \right) \times \Omega^1(\Sigma, P(g)) \longrightarrow \Omega^1(\Sigma, P(g))
\]

given by

\[
(\mu(\xi + i\zeta), \eta) \mapsto \text{Ad}(\mu) (d_{a, \rho} \zeta + *d_{a, \rho} \xi) + \eta = \left\{ d_{(\mu^{-1})^*a, \rho}(\mu^* \xi) + *d_{(\mu^{-1})^*a, \rho}(\mu^* \zeta) \right\} \mu^{-1} + \text{Ad}(\mu) \eta
\]

(13)

Compare with \((\mathcal{G} \ominus \mathcal{G})\). Here we are using the fact that \( \mathcal{G}(P)^C \) and its Lie algebra both embed into the space \( \Gamma(P \times G \text{ End}(C^n)) \), and so it makes sense to multiply Lie group and Lie algebra elements. The curvature transforms under \( \mu \in \mathcal{G}(P)^C \) by

\[
\mu^{-1} \circ F_{(\mu^{-1})^*a, \rho} \circ \mu = F_{a, \rho} + \partial_a \left( h^{-1} \partial_a h \right),
\]

(14)

where we have set \( h = \mu^* \mu \). We will mostly be interested in this action when \( \mu = \exp(i\zeta) \) for \( \zeta \in \Omega^0(\Sigma, P(g)) \), in which case the action can be written as

\[
\exp(-i\zeta) \circ F_{\exp(-i\zeta), \rho} \circ \exp(i\zeta) = *\mathcal{F}(a, \zeta),
\]

(15)

where we have set

\[
\mathcal{F}(a, \zeta) := *\left( F_{a, \rho} + \partial_a (\exp(-2i\zeta) \partial_a \exp(2i\zeta)) \right).
\]

(16)
It will be useful to define the (real) gauge group on $E$ and the complexified gauge group on $E$ by, respectively, 

\[ G(E) := \Gamma(\mathcal{P} \times G \mathbf{U}(n)), \quad G(E)^C := \Gamma(\mathcal{P} \times G \text{GL}(\mathbb{C}^n)). \]

(Note that the complexification of $\mathbf{U}(n)$ is $\text{GL}(\mathbb{C}^n)$, so this terminology is consistent, and in fact motivates, the terminology above.) These are both Lie groups with Lie algebras. Similarly, when we are in the continuous range we can form the Banach Lie group on $\mathbf{C}^k$, though, neither $G(E)^C$ nor $G(E)^C$ restrict to actions on $\mathcal{A}(P)$, unless $G = \mathbf{U}(n)$.

Finally, we mention that the vector spaces $\text{Lie}(G(E))$, $\text{Lie}(G(E)^C)$ and $\text{Lie}(G(P)^C)$ admit Sobolev completions. For example, the space \n
\[ \text{Lie}(G(P)^C)^{k,q} \]

is the $W^{k,q}$-completion of the vector space $\Gamma(P \times G P(\mathfrak{g})^C)$. When we are in the continuous range for Sobolev embedding (i.e., when $kq > 2$) then these are Banach Lie algebras. Similarly, when we are in the continuous range we can form the Banach Lie groups $G^{k,q}(E)$, $G^{k,q}(E)^C$ and $G^{k,q}(P)^C$ by taking the $W^{k,q}$-completions of the groups of smooth functions $G(E)$, $G(E)^C$ and $G(P)^C$, which we view as lying in the vector space $\Gamma(P \times G \text{End}(\mathbb{C}^n))^{k,q}$. The complexified gauge action extends to a smooth action of $G^{k,q}(E)^C$ on $A^{k-1,q}(E)$, and this restricts to a smooth action of $G^{k,q}(P)^C$ on $A^{k-1,q}(P)$. See [34, Appendix B] for more details regarding the Sobolev completions of these spaces.

## 3 Compactness for products $S \times \Sigma$

In this section we consider $Z = S \times \Sigma$, where $(S, g_S), (\Sigma, g_\Sigma)$ are closed, connected, oriented Riemannian surfaces. We also assume that $\Sigma$ has positive genus. We will work with the metric $g = \text{proj}_S^* g_S + \text{proj}_\Sigma^* g_\Sigma$ on $Z$; from now on we will typically drop the projections from the notation. For $\epsilon > 0$, define a new metric by 

\[ g_\epsilon := \epsilon^2 g_S + g_\Sigma. \]

Fix a principal $\text{PU}(r)$-bundle $P \to \Sigma$ such that $t_2(P)[\Sigma] \in \mathbb{Z}_r$ is a generator, and let $R \to Z$ be the pullback bundle under $Z \to \Sigma$. 

11
Given orthonormal coordinates \((U, x = (s, t))\) for \(S\), any connection \(A\) on \(R\) can be written as

\[
A|_{\{s,t\} \times \Sigma} = a(s, t) + \phi(s, t) \, ds + \psi(s, t) \, dt,
\]

where \(a(s, t)\) is a connection on \(P\) and \(\phi(s, t), \psi(s, t) \in \Omega^0(\Sigma, P(\mathfrak{g}))\). In fact, \(a\) can be defined in a coordinate-independent way as \(x \mapsto a(x) := \iota_x^* A \in \mathcal{A}(P)\), where \(\iota_x : \Sigma = \{x\} \times \Sigma \to Z\) is the inclusion. We will say that a connection \(A\) on \(R\) is \(\epsilon\text{-ASD}\) or an \(\epsilon\text{-instanton}\) if it is an instanton with respect to the metric \(g_\epsilon\); that is, if \(F_A = -*_\epsilon F_A\), where \(*_\epsilon\) is the Hodge star on \(Z\) coming from \(g_\epsilon\). This can be written explicitly in terms of local coordinates as

\[
\begin{align*}
\partial_s a - d_a \phi + *_\Sigma (\partial_t a - d_a \psi) &= 0 \\
\partial_s \psi - \partial_t \phi - [\psi, \phi] + e^{-2} *_\Sigma F_a &= 0
\end{align*}
\] (18)

where \(*_\Sigma\) is the Hodge star associated to the \(S\)-dependent metric \(g_\Sigma\). The \(\epsilon\text{-energy}\) of \(A\) is

\[
E^{\text{inst}}_\epsilon(A) := \frac{1}{2} \int_Z |F_A|_\epsilon^2 \, dv_{\epsilon} = \frac{1}{2} \int_Z \langle F_A \wedge *_\epsilon F_A \rangle,
\]

where the norm and volume form are the ones induced by \(g_\epsilon\); this is exactly the Yang-Mills functional on \(\mathcal{A}(R)\) with the metric \(g_\epsilon\). If \(A\) is \(\epsilon\text{-ASD}\), then \(E^{\text{inst}}_\epsilon(A) = 2\pi^2 \kappa_\epsilon r^{-1} q_4(R)\) is a topological invariant by the Chern-Weil formula [5].

On the symplectic side, we will be considering maps \(v : S \to M(P)\), where \(M(P) = A_{\text{flat}}(P)/G_0(P)\). Any such map \(v\) has a lift \(\alpha : S \to A_{\text{flat}}(P)\) if and only if the pullback \(G_0(P)\)-bundle \(v^* A_{\text{flat}}(P) \to S\) is trivial. When \(S\) is not closed, all maps \(v\) have smooth lifts since \(G_0(P)\) is connected and \(S\) retracts to its 1-skeleton; however, when \(S\) is closed there are many maps \(v\) that do not lift.

Given \(\alpha : S \to A_{\text{flat}}(P)\) there is a unique section \(\chi\) of the bundle \(T^*S \otimes \Omega^1(\Sigma, P(\mathfrak{g}))\) over \(S\) such that, for all \(x \in S\) and \(v \in T_x S\), the 1-form

\[
(D_x \alpha)(v) - d_x \alpha - d_x \alpha = \in H^1_{\alpha(x)} \subset \Omega^1(\Sigma, P(\mathfrak{g}))
\]

is \(\alpha(x)\)-harmonic; here \(D_x \alpha : T_x S \to T_{\alpha(x)} A(P)\) is the push-forward of \(\alpha : S \to A(P)\). This data naturally determines a connection \(A_0 \in \mathcal{A}(R)\) by setting

\[
A_0|_{\{x\} \times S} := \alpha(x) + \chi(x).
\]

If \(\alpha\) is the lift of some \(v\), then we call \(A_0\) a \(\text{representative}\) of \(v\), and we call \(A_0\) a \(\text{holomorphic curve representative}\) if \(v\) is holomorphic.

**Example 3.1.** Let \((U, (s, t))\) be local orthonormal coordinates for \(S\). Then

\[
Da|_U = ds \wedge \partial_s a + dt \wedge \partial_t a \quad \text{and} \quad \chi|_U = \phi \, ds + \psi \, dt,
\]

where \(\phi, \psi : U \to \Omega^0(\Sigma, P(\mathfrak{g}))\) are the unique sections such that

\[
\partial_s a(s, t) - d_{a(s,t)} \phi(s, t), \quad \partial_t a(s, t) - d_{a(s,t)} \psi(s, t)
\]
are \( \alpha(s,t) \)-harmonic. It follows that \( A_0 \) is a holomorphic curve representative if and only if

\[
\begin{align*}
\partial_s \alpha - d_s \phi + \ast \Sigma (\partial_t \alpha - d_t \psi) &= 0 \\
F_\alpha &= 0,
\end{align*}
\]

where the equations should be interpreted as being pointwise in \((s,t) \in S\).

We define the energy \( E_{\text{sym}}(A_0) \) of a representative \( A_0 \) to be the energy of the associated map \( v : S \to M(P) \).

**Lemma 3.2.** Given any \( \epsilon > 0 \), the energy of a holomorphic curve representative \( A_0 \) is

\[
E_{\text{sym}}(A_0) = 2\pi^2 \kappa_r r^{-1} q_4(R).
\]

**Proof.** Suppose \( A_0 \) represents \( v \). Then the energy of \( v \) is

\[
\frac{1}{2} \int_S |Dv|^2 \, dv,
\]

where \( Dv \) is the push-forward of \( v : S \to M(P) \). In local coordinates, write \( A_0 = \alpha + \phi \, ds + \psi \, dt \).

Then

\[
Dv = ds \otimes \partial_s v + dt \otimes \partial_t v = ds \otimes \beta_s + dt \otimes \beta_t,
\]

where we have set \( \beta_s := \partial_s \alpha - d_s \phi \) and \( \beta_t := \partial_t \alpha - d_t \psi \). On the other hand,

\[
F_{A_0} = ds \wedge \beta_s + dt \wedge \beta_t + ds \wedge dt \left( \partial_s \psi - \partial_t \phi - [\psi, \phi] \right)
\]

where we have used \( F_\alpha = 0 \). Then by (19) we have

\[
-\frac{1}{2} \int_\Sigma \langle F_{A_0} \wedge F_{A_0} \rangle = \int_\Sigma \langle \beta_s \wedge \beta_t \rangle \wedge ds \wedge dt = \|\beta_s\|_{L^2(\Sigma)}^2 \, ds \wedge dt = \frac{1}{2} |Dv|^2 \, dv.
\]

Integrating the right over \( S \) gives \( E_{\text{sym}}(A_0) \), and integrating the left over \( S \) gives \( 2\pi^2 \kappa_r r^{-1} q_4(R) \) by the Chern-Weil formula.

Now we can state the main theorem of this section.

**Theorem 3.3.** Fix \( 2 < q < \infty \) and let \( R \to Z \) be as above. Suppose \((\epsilon_v)_{v \in \mathbb{N}}\) is a sequence of positive numbers converging to \( 0 \), and assume that, for each \( v \), there is an \( \epsilon_v \)-ASD connection \( A_v \in A^1_{1,\text{loc}}(R) \). Then there is

(i) a finite set \( B \subset S \);

(ii) a subsequence of the \( A_v \) (still denoted \( A_v \));

(iii) a sequence of gauge transformations \( U_v \in G^2_{1,\text{loc}}(R) \); and

(iv) a holomorphic curve representative \( A_\infty \in A^1_{1,\text{loc}}(R) \).
such that the restrictions
\[ \sup_{x \in K} \| i_x^* (U_v^* A_v - A_\infty) \|_{C^0(\Sigma)} \xrightarrow{\nu} 0 \]
converge to zero for every compact \( K \subseteq S \setminus B \). The gauge transformations \( U_v \) can be chosen so that they restrict to the identity component \( G_0(P) \) on each slice \( \{ x \} \times \Sigma \subset Z \). Moreover, for each \( b \in B \) there is a positive integer \( m_b > 0 \) such that for any \( \nu \),
\[ E^{\text{sym}}(A_\infty) = E^{\text{inst}}_{\epsilon_v}(A_v) - 4\pi^2 \kappa_r^{-1} \sum_{b \in B} m_b. \tag{20} \]

Throughout we will use notation of the form
\[ A|_{\{ x \} \times \Sigma} = a(x) + \chi(x), \quad \text{and} \quad U|_{\{ x \} \times \Sigma} = \mu(x) \]
for connections and gauge transformations, respectively. Then the conclusion of the theorem says that \( \mu_v \mid A_v \) converges to \( a_\infty \) in \( C^0 \) on compact sets in \( S \setminus B \), where these are viewed as maps from \( S \) to \( \mathcal{A}(P) \), with the \( C^0 \)-topology on \( \mathcal{A}(P) \). Here, the action of \( \mu = \mu(x) \) is the gauge action on surfaces (not 4-manifolds).

Remark 3.4. (a) If one allows \( S \) to be a compact manifold-with-boundary, then the proof we give here remains valid, except the equality in (20) should be replaced by \( \leq \). This is due to holomorphic disk bubbles occurring at the boundary – the convergence we prove here is not strong enough to show that these are non-trivial; see Remark 3.18. In Section 4 we provide tools for recovering the equality in a certain setting.

(b) This theorem has a straightforward extension to the case where \( S \) has cylindrical ends (assuming one knows flat connections on the ends are non-degenerate; see Section 5). In this case, there can be energy loss at the ends, and so one obtains a limiting holomorphic curve on \( S \) with a finite number of broken cylindrical trajectories on the ends. Then Theorem 3.3 continues to hold provided one accounts for the energies of the broken trajectories on the left-hand side (20). See also Theorem 4.1.

(c) Suppose, for each \( \nu \), we have an open set \( S_\nu \subseteq S \) that is a deformation retract of \( S \), and with the further property that the \( S_\nu \) are increasing and exhausting: \( S_\nu \subset S_{\nu + 1} \) and \( S = \bigcup_\nu S_\nu \). Then the statement of Theorem 3.3 continues to hold if we assume that \( A_\nu \) is defined on \( S_\nu \times \Sigma \).

(d) It is possible to show that the \( C^0 \)-convergence of the \( a_\nu \) to \( a_\infty \) implies \( W^{1,2}_e \)-convergence of the \( A_\nu \) to \( A_\infty \), where \( W^{1,2}_e \) is the Sobolev norm defined with respect to the \( e \)-dependent metric. A similar statement holds in the case of Theorem 4.1 as well. We defer the details to a future paper.

As a stepping stone to Theorem 3.3, we first prove the following lemma; the assumptions allow us to rule out bubbling a priori.

Lemma 3.5. Fix \( 2 < q < \infty \) and a submanifold \( S_0 \subseteq S \) possibly with boundary. Let \( R_0 := R|_{S_0 \times \Sigma} \) denote the restriction. Suppose \( (\epsilon_v)_{v \in \mathbb{N}} \) is a sequence of positive numbers (not necessarily converging to zero), and suppose that for each \( v \) there is an \( \epsilon_v \)-ASD connection \( A_\nu \in \mathcal{A}^{1,2}(R_0) \) satisfying the following conditions for each compact \( K \subset S_0 \).
(i) The slice-wise curvatures converge to zero:

\[
\sup_{x \in K} \| F_{\alpha_{\nu}}(x) \|_{L^\infty(\Sigma)} \xrightarrow{\nu} 0.
\]

(ii) There is some constant \( C \) with

\[
\sup_{\nu} \sup_{x \in K} \| \text{proj}_{\alpha_{\nu}}(x) \circ D_x\alpha_{\nu} \|_{L^2(\Sigma)} \leq C,
\]

where \( \text{proj}_{\alpha_{\nu}}(x) \) is the harmonic projection, and \( D_x\alpha_{\nu} : T_xS \to T_{x(x)}A^{1,1}(P) \) is the push-forward.

Then there is a subsequence of the connections (still denoted \( A_{\nu} \)), a sequence of gauge transformations \( U_{\nu} \) on \( R_0 \), and a holomorphic curve representative \( A_\infty \) on \( R_0 \) such that

\[
\sup_{x \in K} \left\| \alpha_{\infty}(x) - \mu_{\nu}(x) ^* \alpha_{\nu}(x) \right\|_{C^0(\Sigma)} \xrightarrow{\nu} 0 \tag{21}
\]

for every compact \( K \subseteq S_0 \), and

\[
\sup_{x \in K} \left\| \text{proj}_{\alpha_\infty}(x) \circ D_x\alpha_\infty - \text{Ad}(\mu_{\nu}^{-1}(x))\text{proj}_{\alpha_{\nu}}(x) \circ D_x\alpha_{\nu} \right\|_{L^p(\Sigma)} \xrightarrow{\nu} 0 \tag{22}
\]

for any compact \( K \subseteq \text{int} S_0 \) and any \( 1 < p < \infty \).

The connections \( A_{\nu} \) from Theorem 3.3 satisfy the same type of convergence as in (22), for compact \( K \subseteq S \setminus B \). We also point out that the projection operator \( \text{proj}_{\alpha_{\nu}} \) appearing in (22) can be removed by weakening the \( C^0 \)-convergence to \( L^p \)-convergence.

The proofs of Lemma 3.5 and Theorem 3.3 will appear in Sections 3.2 and 3.3, respectively. First we develop some machinery that will allow us to pass from \( \epsilon \)-instantons to holomorphic curves.

### 3.1 Small curvature connections in dimension 2

In our proof of the various convergence results, we will encounter connections on surfaces that have small curvature. Here we develop a strategy for identifying nearby flat connections. The idea is to use the well-known fact that quotienting the subset of small curvature connections by the action of the complexified gauge group recovers the moduli space of flat connections (called a Narasimhan-Seshadri correspondence). The details of this procedure were originally carried out by Narasimhan and Seshadri [24]. They worked with unitary bundles, and this allowed them to use algebraic techniques. Later, their techniques were extended to more general structure groups by Ramanathan in his thesis [27]. (See Kirwan’s book [20] for a finite-dimensional version.)

In preparation for a boundary-value problem, we need to work in an analytic category. Consequently, we adopt an approach of Donaldson [5], and use an implicit function theorem argument to arrive at a Narasimhan-Seshadri correspondence in our setting. This allows us to establish several \( C^1 \)-estimates that will be needed for our proof of the main theorems.
3.1.1 The analytic Narasimhan-Seshadri correspondence

The goal of this section is to define a gauge-equivariant deformation retract $\text{NS}: A^{ss} \rightarrow A^{\text{flat}}$ and establish some of its properties. Here $A^{ss}$ is a suitable neighborhood of $A^{\text{flat}}$ (the superscript stands for semistable). The relevant properties of the map NS are laid out in Theorem 3.6 below. The proof will show that for each $\alpha \in A^{ss}$ there is a ‘purely imaginary’ complex gauge transformation $\mu$ such that $\mu^* \alpha$ a flat connection, and $\mu$ is unique provided it lies sufficiently close to the identity. We then define $\text{NS}(\alpha) := \mu^* \alpha$.

After proving Theorem 3.6 below, where the map NS is formally defined, we spend the remainder of this section establishing useful properties and estimates for NS. For example, in the proof of Lemma 3.15 we establish the Narasimhan-Seshadri correspondence $A^{ss}/G_C^0 \cong A^{\text{flat}}/G_0$, and in Proposition 3.14 and Corollary 3.17 we show that, to first order, the map NS is the identity plus the $L^2$-orthogonal projection to the tangent space of flat connections.

**Theorem 3.6.** Suppose $G$ is a compact, connected Lie group, $\Sigma$ is a closed Riemannian surface, and $P \rightarrow \Sigma$ is a principal $G$-bundle such that all flat connections are irreducible. Then for any $1 < q < \infty$, there are constants $C > 0$ and $\epsilon_0 > 0$, and a $G^{2,q}(P)$-equivariant deformation retract

$$\text{NS}_P : \{ \alpha \in A^{1,q}(P) \mid ||F_{\alpha}||_{L^q(S)} < \epsilon_0 \} \rightarrow A^{1,q}_{\text{flat}}(P)$$

(23)

that is smooth with respect to the $W^{1,q}$-topology on the domain and codomain. Moreover, the map $\text{NS}_P$ is also smooth with respect to the $L^p$-topology on the domain and codomain, for any $2 < p < \infty$.

**Remark 3.7.** The restriction in the second part of the theorem to $2 < p < \infty$ is merely an artifact of our proof, and it is likely that the conclusion holds for, say, $1 < p \leq 2$ as well. See Lemma 3.16.

**Proof of Theorem 3.6** Suppose we can define $\text{NS}_P$ on the set

$$\{ \alpha \in A^{1,q}(P) \mid \text{dist}_{W^{1,q}}(\alpha, A^{1,q}_{\text{flat}}(P)) < \epsilon_0 \}$$

(24)

for some $\epsilon_0 > 0$, and show that it satisfies the desired properties on this smaller domain. Then the $G^{2,q}$-equivariance will imply that it extends uniquely to the flow-out by the real gauge group:

$$\{ \mu^* \alpha \in A^{1,q}(P) \mid \mu \in G^{2,q}(P), \text{dist}_{W^{1,q}}(\alpha, A^{1,q}_{\text{flat}}(P)) < \epsilon_0 \}$$

and continues to have the desired properties on this larger domain. The next claim shows that this flow-out contains a neighborhood of the form appearing in the domain in (23), thereby reducing the problem to defining $\text{NS}_P$ on a set of the form (24).

**Claim:** For any $\overline{\epsilon}_0 > 0$, there is some $\epsilon_0 > 0$ with $\epsilon_0 < \overline{\epsilon}_0$.
\{ \alpha \in A^{1,q}(P) \mid \| F_\alpha \|_{L^q} < \epsilon_0 \} \\
\subseteq \left\{ \mu^* \alpha \in A^{1,q}(P) \mid \mu \in G^{2,q}(P), \ \text{dist}_{W^{1,q}} \left( \alpha, A_{\text{flat}}^{1,q}(P) \right) < \epsilon_0 \right\}.

For sake of contradiction, suppose that for every \( \epsilon > 0 \) there is some connection \( \alpha \) with \( \| F_\alpha \|_{L^q} < \epsilon \), but where

\[
\| \mu^* \alpha - \alpha_0 \|_{W^{1,q}} \geq \epsilon_0, \quad \forall \mu \in G^{2,q}(P), \quad \forall \alpha_0 \in A_{\text{flat}}^{1,q}(P)
\]  

(25)

Then we can find a sequence of connections \( \alpha_i \) with \( \| F_{\alpha_i} \|_{L^q} \to 0 \), but (25) holds with \( \alpha_i \) replacing \( \alpha \). By Uhlenbeck’s weak compactness theorem, there is a sequence of gauge transformations \( \mu_i \in G^{2,q}(P) \) such that, after possibly passing to a subsequence, \( \mu_i^* \alpha_i \) converges weakly in \( W^{1,q} \) to a limiting connection \( \alpha_0 \). The condition on the curvature implies that \( \alpha_0 \in A_{\text{flat}}^{1,q}(P) \) is flat. Moreover, the embedding \( W^{1,q} \hookrightarrow L^{2q} \) is compact, so the weak \( W^{1,q} \)-convergence of \( \mu_i^* \alpha_i \) implies that \( \mu_i^* \alpha_i \) converges strongly to \( \alpha_0 \) in \( L^{2q} \). By redefining \( \mu_i \) if necessary, we may suppose that \( \mu_i^* \alpha_i \) is in Coulomb gauge with respect to \( \alpha_0 \), \( d_{\alpha_0}^*(\mu_i^* \alpha_i - \alpha_0) = 0 \), and still retain the fact that \( \mu_i^* \alpha_i \) converges to \( \alpha_0 \) strongly in \( L^{2q} \). This gives

\[
\| \mu_i^* \alpha_i - \alpha_0 \|_{W^{1,q}}^q = \| \mu_i^* \alpha_i - \alpha_0 \|_{L^q}^q + \| d_{\alpha_i}^* (\mu_i^* \alpha_i - \alpha_0) \|_{L^q}^q + \| d_{\alpha_0}^* (\mu_i^* \alpha_i - \alpha_0) \|_{L^q}^q \\
\leq \| \mu_i^* \alpha_i - \alpha_0 \|_{L^q}^q + \| F_{\alpha_i} \|_{L^q}^q + \| F_{\alpha_0} \|_{L^q}^q + \frac{1}{2} \| \mu_i^* \alpha_i - \alpha_0 \|_{L^{2q}}^q \\
\leq C \left( \| \mu_i^* \alpha_i - \alpha_0 \|_{L^q}^q + \| F_{\alpha_0} \|_{L^q}^q \right)
\]

where we have used \( F_{\alpha_i + v} = d_{\alpha_0}(v) + \frac{1}{2} [v \wedge v] \). Hence \( \| \mu_i^* \alpha_i - \alpha_0 \|_{W^{1,q}}^q \to 0 \), in contradiction to (25). This proves the claim.

To define \( \text{NS}_P \), it therefore suffices to show that for \( \alpha \) sufficiently \( W^{1,q} \)-close to \( A_{\text{flat}}(P) \) there is a unique \( \Xi(\alpha) \in \Omega^0(\Sigma, P(g)) \) close to 0, with \( F_{\exp(i\Xi(\alpha))^* \alpha} = 0 \). Once we have shown this, then we will define

\[ \text{NS}_P(\alpha) := \exp(i\Xi(\alpha))^* \alpha. \]

Recall the definition of \( \mathcal{F} \) in (16). In light of (15), finding \( \Xi(\alpha) \) is equivalent to solving for \( \zeta \) in \( \mathcal{F}(\alpha, \zeta) = 0 \), for then \( \Xi(\alpha) := -\zeta \). To solve for \( \zeta \) we need to pass to suitable Sobolev completions.

It follows from (14), and the Sobolev embedding and multiplication theorems, that \( \mathcal{F} \) extends to a smooth map \( A^{1,q}(P) \times \text{Lie}(\mathcal{G}(P)) \to \text{Lie}(\mathcal{G}(P))\), whenever \( q > 1 \). Suppose \( \alpha_0 \) is a flat connection. The linearization of \( \mathcal{F} \) at \( (\alpha_0, 0) \) in the direction of \( (0, \zeta) \) is

\[ D_{(\alpha_0, 0)} \mathcal{F}(0, \zeta) = 2i E \ast \partial \partial_0^* \zeta, \]

where we have used the fact that \( d_{\alpha, \rho} \) commutes with \( E = i \) (this is because the complex structure \( E \) is constant and the elements of \( \mathcal{A}(P) \) are unitary). Observe that \( E \) acts by the Hodge star on vectors, so
\[ d_{a_0,\rho}(d_{a_0,\rho} \circ j_{\Sigma}) = d_{a_0,\rho} \ast d_{a_0,\rho}, \quad (d_{a_0,\rho} \circ j_{\Sigma})d_{a_0,\rho} = F_{a_0,\rho} \circ (j_{\Sigma}, \text{Id}). \]  

Using this and the fact that \( F_{a_0,\rho} = 0 \), we have

\[ D_{(a_0,\rho)} F(0, \xi) = \frac{1}{2} \Delta_{a_0,\rho} \xi \]

where \( \Delta_{a_0,\rho} = d_{a_0,\rho}^* d_{a_0,\rho} + d_{a_0,\rho} d_{a_0,\rho}^* \) is the Laplacian. By assumption, all flat connections are irreducible, so Hodge theory tells us that the operator \( \Delta_{a_0,\rho} : \text{Lie}(\mathcal{G}(P))^{2,q} \to \text{Lie}(\mathcal{G}(P))^{0,q} \) is an isomorphism. Since \( a_0 \) is flat, the pair \((a_0,0)\) is clearly a solution to \( F(a,\xi) = 0 \). It therefore follows by the implicit function theorem that there are \( \epsilon_{a_0}, \epsilon'_{a_0} > 0 \) such that, for any \( \alpha \in \mathcal{A}^{1,q} \) with \( \|\alpha - a_0\|_{W^{1,q}} < \epsilon_{a_0} \), there is a unique \( \Xi = \Xi(\alpha) \in \text{Lie}(\mathcal{G}(P))^{2,q} \) with \( \|\Xi(\alpha)\|_{W^{2,q}} < \epsilon'_{a_0} \) and \( F(\alpha, -\Xi(\alpha)) = 0 \). The implicit function theorem also implies that \( \Xi(\alpha) \) varies smoothly in \( \alpha \) in the \( W^{1,q} \)-topology. Moreover, by the uniqueness assertion, it follows that \( \Xi(\alpha) = 0 \) if \( \alpha \) is flat.

We need to show that \( \epsilon_{a_0} \) and \( \epsilon'_{a_0} \) can be chosen to be independent of \( a_0 \in \mathcal{A}^{1,q}_{\text{flat}}(P) \). Since the moduli space \( \mathcal{A}_{\text{flat}}/\mathcal{G} \) of flat connections on \( P \) is compact, it suffices to show that \( \epsilon_{a_0} = \epsilon_{\mu^{*}a_0} \) for all real gauge transformations \( \mu \in \mathcal{G}^{2,q}(P) \), and likewise for \( \epsilon'_{a_0} \). Fix \( \mu \in \mathcal{G}^{2,q}(P) \) and \( \alpha \) a connection that is \( W^{1,q} \)-close to \( a_0 \), then find \( \Xi(\alpha) \) as above. By (8) and the statement following (10) we have

\[ \exp(i\Xi(\alpha)) \mu = \mu \exp(i\text{Ad}(\mu^{-1})\Xi(\alpha)). \]  

Since the curvature is \( \mathcal{G}^{2,q}(P) \)-equivariant, we also have

\[ 0 = \text{Ad}(\mu^{-1}) F_{\exp(i\Xi)^{*}\alpha} = F_{(\exp(i\Xi)\mu)^{*}\alpha} = F_{\exp(i\text{Ad}(\mu^{-1})\Xi)^{*}(\mu^{*}\alpha)}. \]

It follows that \( \Xi(\mu^{*}\alpha) = \text{Ad}(\mu^{-1})\Xi(\alpha) \) since \( \Xi(\mu^{*}\alpha) \) is uniquely defined by \( F_{\exp(i\Xi(\mu^{*}\alpha))^{*}(\mu^{*}\alpha)} = 0 \).

We therefore have \( \epsilon_{\mu^{*}a_0} = \epsilon_{a_0} \) and \( \epsilon'_{\mu^{*}a_0} = \epsilon'_{a_0} \), so we can take \( \epsilon_{a_0} \) to be the minimum of

\[ \inf_{[a_0] \in \mathcal{A}_{\text{flat}}/\mathcal{G}} \epsilon_{a_0} > 0 \quad \text{and} \quad \inf_{[a_0] \in \mathcal{A}_{\text{flat}}/\mathcal{G}} \epsilon'_{a_0} > 0. \]

This argument also shows that \( \text{NS}_{P} \) is \( \mathcal{G}^{2,q}(P) \)-equivariant.

Finally, we show that \( \text{NS}_{P}(\alpha) \) depends smoothly on \( \alpha \) in the \( L^p \)-topology for \( p \geq 2 \). It suffices to show that \( \alpha \mapsto \Xi(\alpha) \) extends to a map \( \mathcal{A}^{0,p}(P) \to \text{Lie}(\mathcal{G}(P))^{1,p} \) that is smooth with respect to the specified topologies. To see this, note that \( F \) from (16) is well-defined as a map

\[ \mathcal{A}^{0,p}(P) \times \text{Lie}(\mathcal{G}(P))^{1,p} \to \text{Lie}(\mathcal{G}(P))^{-1,p}. \]

and is smooth with respect to the specified topologies (the restriction to \( p \geq 2 \) is required so that Sobolev multiplication is well-defined). Then the implicit function theorem argument we gave above holds verbatim to show that for each \( a \) sufficiently
Remark 3.8. Let \( \exp \) shows that the composition is smooth. The uniqueness of \( \tilde{\alpha} \) extension of the latter.

\( G = \mathbb{R} \) whenever they are both in the domain of \( NS \).

The following lemma addresses elliptic regularity for the operator \( d \), which are standard for flat connections, to connections with small curvature. The results extend several elliptic properties, which are standard for flat connections, to connections with small curvature. The following lemma addresses elliptic regularity for the operator \( d_{\alpha} \) on 0-forms.

**3.1.2 Analytic properties of almost flat connections**

This section is of a preparatory nature. The results extend several elliptic properties, which are standard for flat connections, to connections with small curvature. The following lemma addresses elliptic regularity for the operator \( d_{\alpha} \) on 0-forms.

**Lemma 3.9.** Suppose \( G \) is a compact Lie group, \( \Sigma \) is a closed oriented Riemannian surface, and \( P \to \Sigma \) is a principal \( G \)-bundle such that all flat connections are irreducible. Let \( 1 < q < \infty \). Then there are constants \( C > 0 \) and \( \epsilon_0 > 0 \) with the following significance.

(i) Suppose that either \( \alpha \in A^{1,q}(P) \) with \( \| F_{\alpha} \|_{L^q(\Sigma)} < \epsilon_0 \), or \( \alpha \in A^{0,q}(P) \) with

\( \| \alpha - \alpha_i \|_{L^2(\Sigma)} < \epsilon_0 \) for some \( \alpha_i \in A^{0,2}(P) \). Then the map \( d_{\alpha} : W^{1,q}(P(\mathfrak{g})) \to L^q(P(\mathfrak{g})) \) is a Banach space isomorphism onto its image. Moreover, for all \( f \in W^{1,q}(P(\mathfrak{g})) \) the following holds

\[
\| f \|_{L^q(\Sigma)} \leq C \| d_{\alpha} f \|_{L^q(\Sigma)}. \quad (28)
\]

(ii) For all \( \alpha \in A^{1,q}(P) \) with \( \| F_{\alpha} \|_{L^q(\Sigma)} < \epsilon_0 \), the Laplacian \( d_{\alpha}^* d_{\alpha} : W^{2,q}(P(\mathfrak{g})) \to L^q(P(\mathfrak{g})) \) is a Banach space isomorphism. Moreover, for all \( f \in W^{2,q}(P(\mathfrak{g})) \) the following holds

\[
\| f \|_{L^q(\Sigma)} \leq C \| d_{\alpha}^* d_{\alpha} f \|_{L^q(\Sigma)}. \quad (29)
\]
Proof. This is basically the statement of [8, Lemma 7.6], but adjusted a little to suit our situation. We prove (ii), the proof of (i) is similar. The assumption that all flat connections $\alpha_\flat$ are irreducible implies that the kernel and cokernel of the elliptic operator $d^*_\alpha d_\alpha : W^{2, \delta}(P(\mathfrak{g})) \to L^\delta(P(\mathfrak{g}))$ are trivial. In particular, we have an estimate $\| f \|_{W^{2, \delta}} \leq C \| d_\alpha * d_\alpha f \|_{L^\delta}$ for all $f \in W^{2, \delta}(P(\mathfrak{g}))$, so the statement of the lemma holds when $\alpha = \alpha_\flat$ is flat.

Next, fix $\alpha \in A^{1, \delta}(P)$ and $\alpha_\flat \in A^{1, \delta}_{\text{flat}}(P)$. Then, by the above discussion, and the relation $d_\alpha f = d_\alpha f + [\alpha - \alpha, f]$, we have $\| f \|_{W^{2, \delta}}$ is bounded by

$$C \| d^*_\alpha d_\alpha f \|_{L^\delta} \leq C \left\{ \| d_\alpha * d_\alpha f \|_{L^\delta} + \| d_\alpha [*(\alpha - \alpha_\flat), f] \|_{L^\delta} \right. $$

$$+ \left. \| [\alpha - \alpha_\flat \wedge [*(\alpha - \alpha_\flat), f]] \|_{L^\delta} \right\}$$

$$\leq C \left\{ \| d_\alpha * d_\alpha f \|_{L^\delta} \right. $$

$$+ \left. C' \| f \|_{W^{2, \delta}} (\| d_\alpha * (\alpha - \alpha_\flat) \|_{L^\delta} + \| [\alpha - \alpha_\flat] \||_{L^{2, \delta}}) \right\},$$

for all $f \in W^{2, \delta}(P(\mathfrak{g}))$, where we have used the embeddings $W^{2, \delta} \hookrightarrow W^{1, \delta}$ and $W^{2, \delta} \hookrightarrow L^\infty$ in the last step. Now suppose that $\| [\alpha - \alpha_\flat] \||_{L^{2, \delta}} < 1/2C'$. Then by composing $\alpha_\flat$ with a suitable gauge transformation, we may suppose $\alpha$ is in Coulomb gauge with respect to $\alpha_\flat$ and still retain the fact that $\| [\alpha - \alpha_\flat] \||_{L^{2, \delta}} < 1/2C'$. Then the above gives

$$\| f \|_{W^{2, \delta}} \leq C \| d_\alpha * d_\alpha f \|_{L^\delta} + \frac{1}{2} \| f \|_{W^{2, \delta}},$$

which shows that $d^*_\alpha d_\alpha$ is injective when sufficiently $L^{2, \delta}$-close to the space of flat connections.

Now we prove the lemma. Suppose (ii) in the statement of the lemma does not hold. Then there is some sequence of connections $\alpha_\nu$ with $\| F_{\alpha_\nu} \|_{L^\delta} \to 0$, but the estimate (29) does not hold for any $C > 0$. By Uhlenbeck’s weak compactness theorem, after possibly passing to a subsequence, there is some sequence of gauge transformations $u_\nu$, and a limiting flat connection $\alpha_\flat$, such that $\| [\alpha_\nu - u_\nu^* \alpha_\flat] \||_{L^{2, \delta}} \to 0$. So the discussion of the previous paragraph shows that, for $\nu$ sufficiently large, the estimate (29) holds with $\alpha$ replaced by $\alpha_\nu$. This is a contradiction, and it proves the lemma.

Now we move on to study the action of $d_\alpha$ on 1-forms. First we establish a Hodge-decomposition result for connections with small curvature. For $2 \leq q < \infty$ and $k \in \mathbb{Z}$, we will use the notation $V^{k, q}$ to denote the $W^{k, q}$-closure of a vector subspace $V \subseteq W^{k, q}(T^\ast \Sigma \otimes P(\mathfrak{g}))$. The standard Hodge decomposition says

$$W^{k, q}(T^\ast \Sigma \otimes P(\mathfrak{g})) = H^{1, q}_{\text{flat}}(\mathfrak{g}) \oplus \left( \text{im } d_\alpha \right)^{k, q} \oplus \left( \text{im } d^*_\alpha \right)^{k, q},$$

for any flat connection $\alpha_\flat$. Here $H^{1, q}_{\text{flat}}(\mathfrak{g})$ is finite dimensional (this dimension is independent of $\alpha_\flat \in A_\flat$), and so is equal to its own $W^{k, q}$-closure. Furthermore, the direct
Lemma 3.10. Assume that $P \to \Sigma$ satisfies the conditions of Lemma 3.9 and let $1 < q < \infty$ and $k \geq 0$. Then there are constants $\epsilon_0 > 0$ and $C > 0$ with the following significance. If $\alpha \in A^{1,q}(P)$ has $\|F_\alpha\|_{L^q(\Sigma)} < \epsilon_0$, then

$$H^1_\alpha := (\ker d_\alpha)^{k,q} \cap (\ker d^*_\alpha)^{k,q} \subseteq W^{k,q}(T^*\Sigma \otimes P(g))$$

has finite dimension equal to $\dim H^1_\alpha$, for any flat connection $\alpha$. Furthermore, the space $H^1_\alpha$ equals the $L^2$-orthogonal complement of the image of $d_\alpha \oplus d^*_\alpha$:

$$H^1_\alpha = \left( (\im d_\alpha)^{k,q} \oplus (\im d^*_\alpha)^{k,q} \right)^\perp,$$

and so we have a direct sum decomposition

$$W^{k,q}(T^*\Sigma \otimes P(g)) = H^1_\alpha \oplus \left( (\im d_\alpha)^{k,q} \oplus (\im d^*_\alpha)^{k,q} \right). \quad (31)$$

In particular, the $L^2$-orthogonal projection

$$\operatorname{proj}_\alpha : W^{k,q}(T^*\Sigma \otimes P(g)) \longrightarrow H^1_\alpha \quad (32)$$

is well-defined.

Remark 3.11. It follows by elliptic regularity that the space $(\ker d_\alpha)^{k,q} \cap (\ker d^*_\alpha)^{k,q}$ consists of smooth forms. Moreover, when $k - 2/q \geq k' - 2/q'$, the inclusion $W^{k,q} \subseteq W^{k',q'}$ restricts to an inclusion of finite-dimensional spaces

$$(\ker d_\alpha)^{k,q} \cap (\ker d^*_\alpha)^{k,q} \hookrightarrow (\ker d_\alpha)^{k',q'} \cap (\ker d^*_\alpha)^{k',q'},$$

and this map is onto by dimensionality. Hence, the definition of $H^1_\alpha$ is independent of the choice of Sobolev constants $k, q$.

Proof of Lemma 3.10. We first show that, when $\|F_\alpha\|_{L^q(\Sigma)}$ is sufficiently small, we have a direct sum decomposition

$$W^{k,q}(T^*\Sigma \otimes P(g)) = H^1_\alpha \oplus (\im d_\alpha)^{k,q} \oplus (\im d^*_\alpha)^{k,q}.$$ 

We prove this in the case $k = 0$; the case $k > 0$ is similar but slightly easier. By definition of $H^1_\alpha$, it suffices to show that the images of $d_\alpha$ and $*d_\alpha$ intersect trivially. Towards this end, write $d_\alpha f = *d_\alpha g$ for $0$-forms $f, g$ of Sobolev class $L^q = W^{0,q}$. Acting by $d_\alpha$ and then $d_\alpha \ast$ gives

$$[F_\alpha, f] = d_\alpha \ast d_\alpha g, \quad [F_\alpha, g] = -d_\alpha \ast d_\alpha f.$$ 

A priori, $d_\alpha \ast d_\alpha g$ and $d_\alpha \ast d_\alpha f$ are only of Sobolev class $W^{-2,q}$, however, the left-hand side of each of these equations is in $L'$, where $1/r = 1/q + 1/p$. So elliptic regularity implies that $f$ and $g$ are each $W^{2,q}$. (This bootstrapping can be continued to show
that \( f, g \) are smooth, but we will see in a minute that they are both zero.) By Lemma 3.9 and the embedding \( W^{2,q} \to L^\infty \), it follows that, whenever \( \|F_\alpha\|_{L^q} \) is sufficiently small, we have

\[
\|f\|_{L^\infty} \leq C\|d_\alpha \ast d_\alpha f\|_{L^q} = C\|F_\alpha \|_{L^q} \leq 2C\|F_\alpha\|_{L^q} \|g\|_{L^\infty}.
\]

Similarly, \( \|g\|_{L^\infty} \leq 2C\|F_\alpha\|_{L^q} \|g\|_{L^\infty} \), and hence

\[
\|f\|_{L^\infty} \leq 4C^2 \|F_\alpha\|_{L^q}^2 \|f\|_{L^\infty}.
\]

If \( \|F_\alpha\|_{L^q}^2 < (2C)^{-2} \), then this can happen only if \( f = g = 0 \). This establishes the direct sum (31).

Now we prove that the dimension of \( H^1_\alpha \) is finite and equals that of \( H^1_{\alpha_0} \) for any flat connection \( \alpha_0 \). It is well-known that the operator

\[
d_\alpha \oplus *d_\alpha : W^{k+1,q}(P(g)) \oplus W^{k+1,q}(P(g)) \to W^{k,q}(T^*\Sigma \oplus P(g))
\]

is elliptic, and hence Fredholm, whenever \( \alpha_0 \) is flat. The irreducibility condition implies that it has trivial kernel, and so has index given by \( -\dim(H^1_{\alpha_0}) \), which is a constant independent of \( \alpha_0 \). Then for any other connection \( \alpha \), the operator \( d_\alpha \oplus *d_\alpha \) differs from \( d_{\alpha_0} \oplus *d_{\alpha_0} \) by a compact operator, and so \( d_\alpha \oplus *d_\alpha \) is Fredholm with the same index \( -\dim(H^1_{\alpha_0}) \) [23, Theorem A.1.5]. It follows from Lemma 3.9 that the (bounded) operator

\[
d_\alpha \oplus *d_\alpha : W^{k+1,q}(P(g)) \oplus W^{k+1,q}(P(g)) \to W^{k,q}(T^*\Sigma \oplus P(g))
\]

is injective whenever \( \|F_\alpha\|_{L^q(\Sigma)} \) is sufficiently small, and hence the cokernel has finite dimension \( \dim(H^1_{\alpha_0}) \). That is, \( \dim(H^1_\alpha) = \dim(H^1_{\alpha_0}) \), so this finishes the proof of Lemma 3.10.

Next we show that the \( L^2 \)-orthogonal projection to \( H^1_\alpha = \ker d_\alpha \cap \ker d_\alpha^* \) depends smoothly on \( \alpha \) in the \( L^q \)-topology.

**Proposition 3.12.** Suppose that \( P \to \Sigma \) and \( \epsilon_0 > 0 \) are as in Lemma 3.10 and let \( 1 \leq q < \infty \). Then the assignment \( \alpha \mapsto \text{proj}_\alpha \) is affine-linear and bounded

\[
\|\text{proj}_\alpha - \text{proj}_{\alpha'}\|_{\text{op}, L^1} \leq C\|\alpha - \alpha'\|_{L^1(\Sigma)},
\]

provided \( \|F_{\alpha}\|_{L^1}, \|F_{\alpha'}\|_{L^1} < \epsilon_0 \), where \( \|\cdot\|_{\text{op}, L^1} \) is the operator norm on the space of linear maps \( L^1(T^*\Sigma \oplus P(g)) \to L^1(T^*\Sigma \oplus P(g)) \).

**Proof.** We will see that defining equations for \( \text{proj}_\alpha \) are affine linear, and so the statement will follow from the implicit function theorem in the affine-linear setting.

First, we introduce the following shorthand:

\[
W^{k,q}(\Omega^i) := W^{k,q} \left( \Lambda^i T^*\Sigma \oplus P(g) \right), \quad L^q(\Omega^i) := W^{0,q}(\Omega^i).
\]

Next, we note that for \( \mu \in L^{q}(\Omega^1) \), the \( L^2 \)-orthogonal projection \( \text{proj}_\alpha \mu \) is uniquely characterized by the following properties:
Property A: \( \exists (u, v) \in W^{1,q}(\Omega^0) \oplus W^{1,q}(\Omega^2), \mu - \text{proj}_a \mu = d_a u + d_a^* v, \)

Property B: \( \forall (a, b) \in W^{1, q^*}(\Omega^0) \oplus W^{1, q^*}(\Omega^2), \langle \text{proj}_a \mu, d_a a + d_a^* b \rangle = 0, \)

where \( q^* \) is the Sobolev dual to \( q \): \( 1/q + 1/q^* = 1 \). Here and below, we use the notation \( \langle \mu, \nu \rangle \) to denote the \( L^2 \)-pairing on forms. Note that by Lemma 3.9 the operators \( d_a \) and \( d_a^* \) are injective on 0- and 2-forms, respectively, so any pair \((u, v)\) satisfying Property A is unique.

Consider the map

\[
(\mathcal{A}^{0,q} \oplus L^q(\Omega^1)) \times (L^q(\Omega^1) \oplus W^{1,q}(\Omega^0) \oplus W^{1,q}(\Omega^2)) \rightarrow (W^{1,q^*}(\Omega^0))^* \oplus (W^{1,q^*}(\Omega^2))^* \oplus L^q(\Omega^1)
\]

(34)

defined by

\[
(a, \mu; v, u, v) \mapsto \left( \langle v, d_a (\cdot) \rangle, \langle v, d^*_a (\cdot) \rangle, \mu - v - d_a u - d_a^* v \right)
\]

The key point is that a tuple \((a, \mu; v, u, v)\) maps to zero under (34) if and only if this tuple satisfies Properties A and B above. By the identification \((W^{k,q^*})^* = W^{-k,q^*}\), we can equivalently view (34) as a map

\[
(\mathcal{A}^{0,q} \times L^q(\Omega^1)) \times (L^q(\Omega^1) \times W^{1,q}(\Omega^0) \times W^{1,q}(\Omega^2)) \rightarrow W^{-1,q}(\Omega^0) \oplus W^{-1,q}(\Omega^2) \oplus L^q(\Omega^1)
\]

(35)

defined by

\[
(a, \mu; v, u, v) \mapsto (d^*_a v, d_a v, \mu - v - d_a u - d_a^* v).
\]

Claim 1: The map (35) is bounded affine linear in the \( \mathcal{A}^{0,q} \)-variable, and bounded linear in the other 4 variables.

Claim 2: The linearization at \((a, 0, 0, 0, 0)\) of (35) in the last 3-variables is a Banach space isomorphism, provided \( \|a - a_j\|_{L^q} \) is sufficiently small for some flat connection \( a_j \).

Before proving the claims, we describe how they prove the lemma. Observe that \((a, 0, 0, 0, 0)\) is clearly a zero of (35) for any \( a \). Claim 1 implies that (35) is smooth, and so by Claim 2 we can use the implicit function theorem to show that, for each pair \((a, \mu) \in \mathcal{A}^{0,q} \oplus L^q(\Omega^1)\), with \( \|a - a_j\|_{L^q} \) sufficiently small, there is a unique \((v, u, v) \in L^q(\Omega^1) \oplus W^{1,q}(\Omega^0) \oplus W^{1,q}(\Omega^2)\) such that \((a, \mu; v, u, v)\) is a zero of (35). (A priori this only holds for \( \mu \) in a small neighborhood of the origin, but since (35) is linear in that variable, it extends to all \( \mu \).) It will then follow that \( v = \text{proj}_a \mu \) depends smoothly on \( a \) in the \( L^q \)-metric. In fact, (35) is affine linear in \( a \) and linear in the other variables, so the uniqueness assertion of the implicit function theorem implies...
that $\text{proj}_a$ depends affine-linearly on $\alpha$, and so it follows that $\|\text{proj}_a - \text{proj}_{a_0}\|_{\text{op}, L^q}$ is bounded by

$$\inf_{\|\mu\|_{L^q}=1} \left\| \left( \text{proj}_a - \text{proj}_{a_0} \right) \mu \right\|_{L^q} \leq C \inf_{\|\mu\|_{L^q}=1} \|\alpha - \alpha_0\|_{L^q} \|\mu\|_{L^q} = C\|\alpha - \alpha_0\|_{L^q}.$$ 

This proves the lemma for all $\alpha$ sufficiently $L^q$-close to $A_{\text{flat}}$. To extend it to all $\alpha$ with $\|F_\alpha\|_{L^q}$ sufficiently small, one argues by contradiction as in the proof of Lemma 3.9 using Uhlenbeck’s compactness theorem. It therefore remains to prove the claims.

**Proof of Claim 1:** It suffices to verify boundedness for each of the three (codomain) components separately. The first component is the map

$$A^{0,q} \times L^q(\Omega^1) \longrightarrow W^{-1,q}(\Omega^0), \quad (\alpha, \nu) \longmapsto d_\alpha^* \nu$$

(36)

It is a standard consequence from the principle of uniform boundedness that a bilinear map is continuous if it is continuous in each variable separately. The same holds if the map is linear in one variable and affine-linear in the second, so it suffices to show that (36) is bounded in each of the two coordinates separately. Fix $\alpha$ and a flat connection $\alpha_0$. Then

$$\|d_\alpha \nu\|_{W^{-1,q}} \leq \|d_\alpha \nu\|_{W^{-1,q}} + \|\alpha - \alpha_0 \wedge \nu\|_{W^{-1,q}} \leq \|d_\alpha \nu\|_{W^{-1,q}} + 2\|\alpha - \alpha_0\|_{L^q} \|\nu\|_{L^q} \leq C (1 + \|\alpha - \alpha_0\|_{L^q}) \|\nu\|_{L^q},$$

which shows that the map is bounded in the variable $\nu$, with $\alpha$ fixed. Next, fix $\nu$ and write

$$\|d_\alpha \nu - d_{\alpha_0} \nu\|_{W^{-1,q}} = \|\alpha - \alpha_0 \wedge \nu\|_{W^{-1,q}} \leq 2\|\nu\|_{L^q} \|\alpha - \alpha_0\|_{L^q},$$

which shows it is bounded in the $\alpha$-variable. This shows the first component of (36) is bounded. The other two components are similar.

**Proof of Claim 2:** The linearization of (35) at $(\alpha, 0; 0, 0, 0)$ in the last three variables is the map

$$L^q(\Omega^1) \times W^{1,q}(\Omega^0) \times W^{1,q}(\Omega^2) \longrightarrow W^{-1,q}(\Omega^0) \oplus W^{-1,q}(\Omega^2) \oplus L^q(\Omega^1), \quad (v, u, v) \longmapsto (d_\alpha^* v, d_\alpha^* u, -v - d_\alpha u - d_\alpha^* v)$$

By Claim 1, this is bounded linear, so by the open mapping theorem, it suffices to show that it is bijective. Suppose

$$(d_\alpha^* v, d_\alpha^* u, -v - d_\alpha u - d_\alpha^* v) = (0, 0, 0).$$

Then by Lemma 3.10 we can write $v$ uniquely as $v = v_H + d_\alpha a + d_\alpha^* b$ for $v_H \in H^1 = \ker d_\alpha \cap \ker d_\alpha^*$ and $(a, b) \in W^{1,q}(\Omega^0) \times W^{1,q}(\Omega^2)$, provided $\|F_\alpha\|_{L^q}$ is sufficiently small. This uniqueness, together with the first two components of (37), imply that $v = v_H$. The third component then reads

24
which is only possible if \( v_H = d_a u - d_a^* v = 0 \). By Lemma 3.9 this implies \((v, u, v) = (0, 0, 0)\), which proves injectivity.

To prove surjectivity, suppose the contrary. Then by the Hahn-Banach theorem, there are non-zero dual elements

\[
(f, g, \eta) \in W^{1,q} (\Omega^0) \oplus W^{1,q} (\Omega^1) \oplus L^q (\Omega^1)
\]

satisfying

\[
0 = \langle f, d_a^* v \rangle, \quad 0 = \langle g, d_a v \rangle, \quad 0 = \langle \eta, v + d_a u + d_a^* v \rangle
\]

for all \((v, u, v)\). The first two equations imply \( \langle d_a f, v \rangle = 0 \) and \( \langle d_a^* g, v \rangle = 0 \) for all \( v \). This implies \( d_a f = 0 \) and \( d_a^* g = 0 \), and so \( f = 0 \) and \( g = 0 \) by Lemma 3.9. For the third equation, take \((u, v) = (0, 0)\) and we get \( 0 = \langle \eta, v \rangle \) for all \( v \). But this can only happen if \( \eta = 0 \), which is a contradiction to the tuple \((f, g, \eta)\) being non-zero. \( \square \)

We end this preparatory section by establishing the analogue of Lemma 3.9 for 1-forms.

**Lemma 3.13.** Assume that \( P \rightarrow \Sigma \) satisfies the conditions of Lemma 3.9 and let \( 1 < q < \infty \). Then there are constants \( C > 0 \) and \( \epsilon_0 > 0 \) such that

\[
\| \eta - \text{proj}_{a_o} \eta \|_{W^{1,q} (\Sigma)} \leq C \left( \| d_{a_o} \eta \|_{L^q (\Sigma)} + \| d_{a_o} \ast \eta \|_{L^q (\Sigma)} \right)
\]

for all \( \eta \in W^{1,q} (T^* \Sigma \otimes P(g)) \) and all \( a \in A^{1,q} (\Sigma) \) with \( \| F_a \|_{L^q (\Sigma)} < \epsilon_0 \).

**Proof.** First note that this is just the standard elliptic regularity result if \( a = a_o \) is flat. To prove the lemma, suppose the conclusion does not hold. Then there is (1) a sequence of connections \( a_v \) with \( \| F_{a_v} \|_{L^q} \rightarrow 0 \), and (2) a sequence of 1-forms \( \eta_v \in \text{im} \ d_{a_v} \cap \text{im} \ast d_{a_v} \) with \( \| \eta_v \|_{W^{1,q}} = 1 \) and for which \( \| d_{a_v} \eta_v \|_{L^q} \) and \( \| d_{a_v} \ast \eta_v \|_{L^q} \) both converge to zero. By applying suitable gauge transformations to the \( a_v \), and by passing to a subsequence, it follows from Uhlenbeck compactness that the \( a_v \) converge strongly in \( L^{2q} \) to a limiting flat connection \( a_o \). Next, we have

\[
\| d_{a_v} \eta_v \|_{L^q} \leq \| d_{a_o} \eta_v \|_{L^q} + \| a_v - a_o \wedge \eta_v \|_{L^q},
\]

and so

\[
\| d_{a_v} \eta_v \|_{L^q} \leq \| d_{a_o} \eta_v \|_{L^q} + C_0 \| a_v - a_o \|_{L^{2q}} \| \eta_v \|_{W^{1,q}} \rightarrow 0,
\]

where in the last inequality we have used the embedding \( W^{1,q} \hookrightarrow L^{2q} \) for \( q > 1 \). Similarly \( \| d_{a_v} \ast \eta_v \|_{L^q} \rightarrow 0 \). By the elliptic estimate for the flat connection \( a_o \), we have

\[
\| \eta_v - \text{proj}_{a_o} \eta_v \|_{W^{1,q}} \leq C_1 \left( \| d_{a_o} \eta_v \|_{L^q} + \| d_{a_v} \ast \eta_v \|_{L^q} \right) \rightarrow 0.
\]

On the other hand, by Proposition 3.12 the projection operator \( \text{proj}_{a_v} \) is converging in the \( L^q \) operator norm to \( \text{proj}_{a_o} \); in particular,
\[
\| \text{proj}_{\mathcal{A}_v}(\eta_v) \|_{W^{1,\theta}} \leq C_2 \| \text{proj}_{\mathcal{A}_v}(\eta_v) \|_{L^\theta} = C_2 \| \text{proj}_{\mathcal{A}_v}(\eta_v) - \text{proj}_{\mathcal{A}_v}(\eta_v) \|_{L^\theta} \\
\leq C_2 \| \alpha_v - a_v \|_{L^\theta} \eta_v \|_{L^\theta} \longrightarrow 0,
\]

where the first inequality holds because \( H_{\mathcal{A}}^{1,\theta} \) is finite-dimensional (and so all norms are equivalent), the second inequality holds by Proposition 3.12, and the convergence to zero holds since \( \| \eta_v \|_{L^\theta} \leq C_3 \| \eta_v \|_{W^{1,\theta}} = C_3 \) is bounded. Combining this with (39) gives

\[
1 = \| \eta_v \|_{W^{1,\theta}} \leq \| \eta_v - \text{proj}_{\mathcal{A}_v} \eta_v \|_{W^{1,\theta}} + \| \text{proj}_{\mathcal{A}_v}(\eta_v) \|_{W^{1,\theta}} \longrightarrow 0,
\]

which is a contradiction, proving the lemma.

\[\Box\]

### 3.1.3 Analytic properties of \( \text{NS} \)

The next proposition will be used to obtain \( C^0 \)-estimates for convergence of instantons to holomorphic curves. It provides a quantitative version of the statement that \( \text{NS} \) is approximately the identity map on connections with small curvature.

**Proposition 3.14.** Let \( \text{NS}_p \) be the map (23), and \( 3/2 \leq q < \infty \). Then there are constants \( C > 0 \) and \( \epsilon_0 > 0 \) such that

\[
\| \text{NS}_p(\alpha) - \alpha \|_{W^{1,q}(\Sigma)} \leq C \| F_\alpha \|_{L^q(\Sigma)}
\]

for all \( \alpha \in \mathcal{A}^{1,q}(P) \) with \( \| F_\alpha \|_{L^q(\Sigma)} < \epsilon_0 \).

**Proof.** The basic idea is that \( \text{NS}_p(\alpha) = \exp(i \Xi(\alpha))^* \alpha \) can be expressed as a power series, with lowest order term given by \( \alpha \) (see the proof of Theorem 3.6 for the definition of \( \Xi(\alpha) \)). The goal is then to bound the higher order terms using the curvature. To describe this precisely, we digress to discuss the power series expansion for the exponential.

As discussed above, the space \( \text{Lie}(G(E)^C)^{2,q} \) can be viewed as the \( W^{2,q} \)-completion of the vector space \( \Gamma(P \times_{\mathcal{C}} \text{End}(\mathcal{C}^n)) \). Since \( q > 1 \), we are in the range in which pointwise matrix multiplication is well-defined, and \( \text{Lie}(G(E)^C)^{2,q} \) becomes a Banach algebra. Then for any \( \xi \in \text{Lie}(G(E)^C)^{2,q} \), the power series

\[
\sum_{k=0}^{\infty} \xi^k / k! \in \text{Lie}(G(E)^C)^{2,q}
\]

converges, where \( \xi^k \) is \( k \)-fold matrix multiplication on the values of \( \xi \).

As with finite-dimensional Lie theory, this power series represents the exponential map \( \exp : \text{Lie}(G(E)^C)^{2,q} \longrightarrow G^{2,q}(E)^C \), where we are using the inclusion \( G^{2,q}(E)^C \subset W^{2,q}(P \times_{\mathcal{C}} \text{End}(\mathcal{C}^n)) \). The power series defining \( \exp \) continues to hold on the restriction map \( \text{Lie}(G(P)^C)^{2,q} \longrightarrow G^{2,q}(P)^C \). Similarly, the usual power series definitions of \( \sin \) and \( \cos \) hold in this setting, and we have the familiar relation \( \exp(i \xi) = \cos(\xi) + i \sin(\xi) \). Note that by (12), for any real \( \xi \in \text{Lie}(G(E))^ {2,q} \) and \( \alpha \in \mathcal{A}^{1,q}(P) \), we have
\[
\exp(i\xi)^*a - a = -\left\{d_a(\cos(\xi)) - 1 + \ast d_a(\sin(\xi))\right\} \in T_aA^1A(P),
\] (41)

where the action of \(d_a\) on each of these power series is defined term by term.

Now we prove the proposition. We will show that there is some \(\epsilon > 0\) and \(C > 0\) such that, if \(a \in A^1A(P)\) satisfies \(\|F_a\|_{L^q(\Sigma)} < \epsilon_0\) and \(\|\Xi(a)\|_{W^{2,4}} < \epsilon_0\), then \(\|\Xi(a) - a\|_{W^{4,6}} \leq C\|F_a\|_{L^q}\). This is exactly the statement of the proposition, except we have an additional assumption on \(\Xi(a)\). However, we know \(\Xi(a)\) depends continuously on \(a\) and vanishes when \(a\) is flat, so this additional assumption is superfluous.

Set \(\Xi = \Xi(a)\) and \(\eta := NS_P(a) - a\). Using the power series expansion of \(\exp\), we have
\[
\eta = \exp(i\Xi)^*a - a = -\ast d_{a,\rho}\Xi + \frac{\Xi(d_{a,\rho}\Xi) + (d_{a,\rho}\Xi)\Xi}{2} + \ldots
\]
The \(n\)th term in the sum on the right has the form \(-\frac{\pi^n}{n!} \sum_{k=0}^n \Xi^k (d_{a,\rho}\Xi)^n - k - 1\). By assumption \(2q > 2\), and so
\[
\left\|\frac{\pi^n}{n!} \sum_{k=0}^n \Xi^k (d_{a,\rho}\Xi)^n \Xi \right\|_{W^{1,2q}} \leq \|d_{a,\rho}\Xi\|_{W^{1,2q}} \left(\frac{C_{2n}}{n!} \sum_{k=0}^n \|\Xi\|_{W^{1,2q}}^{n-1}\right)
\leq \|\Xi\|_{W^{2,2q}} \left(\frac{C_{2n}}{n!} \|\Xi\|_{W^{1,2q}}^{n-1}\right),
\]
where \(C_1\) is the constant from the Sobolev multiplication theorem. This gives \(\|\eta\|_{W^{1,q}} \leq C_2\|\eta\|_{W^{1,2q}} \leq C_2\|\Xi\|_{W^{2,2q}} \sum_{n=1}^\infty \|\Xi\|_{W^{1,2q}}^{n-1} C_{2n} / (n - 1)!\). Whenever \(\|\Xi\|_{W^{1,2q}} \leq 1\), we therefore have
\[
\|\eta\|_{W^{1,2q}} \leq C_2\|\Xi\|_{W^{2,2q}} \sum_{n=1}^\infty \frac{C_{2n}}{(n - 1)!} = C_3\|\Xi\|_{W^{2,2q}}.
\] (42)

We will be done if we can estimate \(\|\Xi\|_{W^{2,2q}}\) in terms of \(\eta\) and \(F_a\).

By (41) and the definition of \(NS_P\) we have that \(d_{a,\rho}\eta\) is equal to
\[
d_{a,\rho}(\exp(i\Xi)^*a - a) = -d_{a,\rho}(\ast d_{a,\rho}(\sin(\Xi)) + d_{a,\rho}(\cos(\Xi) - 1))
= -d_{a,\rho}(\ast d_{a,\rho}(\sin(\Xi)) + F_{a,\rho}(1 - \cos(\Xi)))
\] (43)

Now use the elliptic estimate from Lemma 3.9(ii):
\[
\|\sin(\Xi)\|_{L^{2,2q}} \leq C_4\|d_{a,\rho}(\ast d_{a,\rho}(\sin(\Xi)))\|_{L^{q},q}
\leq C_5 \{\|d_{a,\rho}\eta\|_{L^{2,q}} + \|F_{a,\rho}(1 - \cos(\Xi))\|_{L^{2,q}}\}
\leq C_6 \{\|\eta\|_{L^{q}}^2 + \|F_{a,\rho}\|_{L^{2,q}} (1 + \|1 - \cos(\Xi)\|_{L^{\infty}})\},
\]
where the second inequality is (43), and in the last inequality we used
\[ \|d_{\alpha, \rho}\eta\|_{L^2_\rho} \leq C \left( \|F_{\alpha, \rho}\|_{L^2_\rho} + \|\eta\|_{L^2_\rho}^2 \right) \]

coming from the identity \(0 = F_{\text{NS}_\rho(a), \rho} = F_{\alpha, \rho} + d_{\alpha, \rho}\eta + \frac{1}{2}|\eta \wedge \eta|\). Note also that the norm of \(F_{\alpha, \rho}\) is controlled by that of \(F_{\alpha, \rho}\), so we can drop the subscript \(\rho\) by picking up another constant:

\[ \|\sin(\Xi)\|_{W^{2,2}_{\rho}} \leq C \left\{ \|\eta\|_{L^2_{\rho}}^2 + \|F_{\alpha}\|_{L^2_\rho} (1 + \|1 - \cos(\Xi)\|_{L^\infty}) \right\} \]  \(44\)

For \(\|\Xi\|_{W^{2,2}_{\rho}}\) small we have \(\|\Xi\|_{W^{2,2}_{\rho}} \leq 2\|\sin(\Xi)\|_{W^{2,2}_{\rho}}\) and \(\|1 - \cos(\Xi)\|_{L^\infty} \leq 1\), so \(44\) gives \(\|\Xi\|_{W^{2,2}_{\rho}} \leq C_8 \left(\|\eta\|_{L^2_{\rho}}^2 + \|F_{\alpha}\|_{L^2_\rho}\right)\).

Returning to \((42)\), we conclude

\[ \|\eta\|_{W^{1,4}_{\rho}} \leq C_9 \left(\|\eta\|_{L^2_{\rho}}^2 + \|F_{\alpha}\|_{L^2_\rho}\right) \leq C_{10} \left(\|\eta\|_{W^{1,4}_{\rho}} + \|F_{\alpha}\|_{L^2_\rho}\right) \]

where we have used the embedding \(W^{1,4}_{\rho} \hookrightarrow L^{4\rho}\), which holds provided \(q \geq 3/2\). This gives \(\|\eta\|_{W^{1,4}_{\rho}} (1 - C_{10}\|\eta\|_{W^{1,4}_{\rho}}) \leq C_{10}\|F_{\alpha}\|_{L^2_\rho}\), and completes the proof since we can ensure that \(\|\eta\|_{W^{1,4}_{\rho}} \leq 1/2C_{10}\) by requiring that \(\|\Xi\|_{W^{2,2}_{\rho}}\) is sufficiently small (when \(\Xi = 0\), it follows that \(\eta = 0\), and everything is continuous in these norms).

Throughout the remainder of this section, we assume that \(\Sigma\) is closed, connected and orientable, \(G = PU(r)\), and \(P \to \Sigma\) is a bundle for which \(t_2(P) |\Sigma| \in \mathbb{Z}_r\) is a generator. Let \(\Pi : \mathcal{A}_{\text{flat}}^1(P) \to M(P) = \mathcal{A}_{\text{flat}}^1(P)/\mathcal{G}_{D}^2(P)\) denote the quotient map. The assumptions on \(G\) and \(P\) imply that \(M(P)\) and \(\Pi\) are smooth. We will be interested in estimating the derivative of the composition \(\Pi \circ \text{NS}_P\).

Any choice of orientation and metric on \(\Sigma\) determines complex structures on the tangent bundles \(T\mathcal{A}^1(P)\) and \(TM(P)\) that are induced by the Hodge star on 1-forms. Denote by

\[ D_{\alpha} (\Pi \circ \text{NS}_P) : T_{\alpha}\mathcal{A}^1(P) \to T_{\Pi \circ \text{NS}_P(a)} M(P) \]

the linearization of \(\Pi \circ \text{NS}_P\) at \(\alpha\), defined with respect to the \(W^{1,q}_{\rho}\)-topology on the domain. The following lemma will be used to show that holomorphic curves in \(\mathcal{A}^1(P)\) descend to holomorphic curves in \(M(P)\).

**Lemma 3.15.** Suppose \(G = PU(r)\), \(\Sigma\) is a closed, connected, oriented Riemannian surface, and \(P \to \Sigma\) is a principal \(G\)-bundle with \(t_2(P) |\Sigma| \in \mathbb{Z}_r\) a generator. Let \(1 < q < \infty\) and suppose \(\alpha\) is in the domain of \(\text{NS}_P\). Then the linearization \(D_{\alpha} (\Pi \circ \text{NS}_P)\) is complex-linear:

\[ *D_{\alpha} (\Pi \circ \text{NS}_P) = D_{\alpha} (\Pi \circ \text{NS}_P) * . \]

**Proof.** We refer to the notation of Section 2.2. The complex gauge group \(\mathcal{G}(P)^C\) acts on \(\mathcal{C}(P)\), and hence \(\mathcal{A}(P)\), in a way that preserves the complex structure, and this holds true in the Sobolev completions of these spaces. Indeed, let \(\mu \in \mathcal{G}^2(P)^C\), \(\alpha \in \mathcal{A}^1(P)\) and \(\eta \in W^{1,\frac{1}{2}}(T^*\Sigma \otimes P(\mathfrak{g}))\). Then by \(\Pi\) we have

28
Lemma 3.16. Let \( P \) be as in the statement of Lemma 3.15 and \( 1 < q < \infty \). There are constants \( C, \delta > 0 \) so that

\[
\left| D_\alpha (\Pi \circ \text{NS}_P) (\eta) \right|_{M(P)} \leq C \| \eta \|_{L^q(\Sigma)}
\]

(45)

for all 1-forms \( \eta \in W^{1,q}(T^* \Sigma \otimes \mathfrak{P}(\mathfrak{g})) \) and all connections \( \alpha \in \mathcal{A}(P) \) with \( \| F_\alpha \|_{L^q(\Sigma)} < \delta \).

Proof. Note that by Lemma 3.10 there is a decomposition

\[
T_\alpha A^{1,q}(P) = H_\alpha^1 \oplus (\text{im } d_\alpha \oplus \text{im } d_\alpha^*) ,
\]
whenever \( \alpha \) has sufficiently small curvature. Moreover, the first summand is \( L^2 \)-orthogonal. Denote by \( \text{proj}_\alpha : T_\alpha \mathcal{A}^{1,\bar{1}}(P) \to H^2_\alpha \) the projection to the \( d_\alpha \)-harmonic space, and note that this is continuous with respect to the \( L^q \)-norm on the domain and codomain. We claim that the operator

\[
D_\alpha (\Pi \circ \text{NS}_{P}) : T_\alpha \mathcal{A}^{1,\bar{1}} \to H_{\text{NS}_{P}(\alpha)}
\]

can be written as a composition

\[
T_\alpha \mathcal{A}^{1,\bar{1}} \to H_\alpha \xrightarrow{M_\alpha} H_{\text{NS}_{P}(\alpha)}
\]

for some bounded linear map \( M_\alpha \), where the first map is \( \text{proj}_\alpha \). Indeed, we have

\[
D_\alpha (\Pi \circ \text{NS}_{P}) (\mu) = D_\alpha (\Pi \circ \text{NS}_{P}) (\text{proj}_\alpha \mu)
\]

since the difference \( \mu - \text{proj}_\alpha \mu \) lies in \( \text{im} d_\alpha \oplus \text{im} * d_\alpha \), and this space is contained in the kernel of \( D_\alpha (\Pi \circ \text{NS}_{P}) \) by Remark 3.8. So the claim follows by taking

\[
M_\alpha := D_\alpha (\Pi \circ \text{NS}_{P}) |_{H_\alpha}
\]

to be the restriction. Since \( M_\alpha \) is a linear map between finite-dimensional spaces, it is bounded with respect to any norm. We take the \( L^q \)-norm on these harmonic spaces. Then \( D_\alpha (\Pi \circ \text{NS}_{P}) \) is the composition of two functions that are continuous with respect to the \( L^q \) norm:

\[
|D_\alpha (\Pi \circ \text{NS}_{P}) \mu|_{L^q(M(P))} = C \| D_\alpha (\Pi \circ \text{NS}_{P}) \mu \|_{L^q(\Sigma)} = C \| M_\alpha \circ \text{proj}_\alpha \mu \|_{L^q(\Sigma)} \leq C_a \| \mu \|_{L^q(\Sigma)}.
\]

That this constant can be taken independent of \( \alpha \), for \( F_\alpha \) sufficiently small, follows using an Uhlenbeck compactness argument similar to the one carried out at the beginning of the proof of Theorem 3.6. Here one needs to use the fact that \( D_\alpha (\Pi \circ \text{NS}_{P}) = \text{proj}_\alpha \) when \( \alpha \) is a flat connection, and so this has norm 1 (which is clearly independent of \( \alpha \)).

**Corollary 3.17.** Suppose \( 1 < q < \infty \), and let \( P \to \Sigma \) be as in the statement of Lemma 3.15. Then there is a constant \( \epsilon_0 > 0 \) and a bounded function \( f : \mathcal{A}^{1,\bar{1}}(P) \to \mathbb{R}^+ \) such that for each \( \alpha \in \mathcal{A}^{1,\bar{1}}(P) \) with \( \| F_\alpha \|_{L^2(\Sigma)} < \epsilon_0 \), the following estimate holds

\[
\| \text{proj}_\alpha \eta - D_\alpha (\Pi \circ \text{NS}_{P}) \eta \|_{L^q(\Sigma)} \leq f(\alpha) \| \text{proj}_\alpha \eta \|_{L^q(\Sigma)}
\]

for all \( \eta \in L^q(T^* \Sigma \otimes P(\mathbb{G})) \), where \( \text{proj}_\alpha \) is the map (32). Furthermore, \( f \) can be chosen so that \( f(\alpha) \to 0 \) as \( \| F_\alpha \|_{L^2(\Sigma)} \to 0 \).

**Proof.** Consider the operator \( \text{proj}_\alpha - D_\alpha (\Pi \circ \text{NS}_{P}) \). It is clear from Lemma 3.16 that its kernel contains \( \text{im}(d_\alpha) \oplus \text{im}(\text{im}(d_\alpha)) \), and so we have

\[
\| \text{proj}_\alpha \eta - D_\alpha (\Pi \circ \text{NS}_{P}) \eta \|_{L^q} \leq C_1 \| \text{proj}_\alpha \eta - D_\alpha (\Pi \circ \text{NS}_{P}) \eta \|_{L^2}.
\]

30
This last term is bounded by \( C_1 \| \text{proj}_\alpha - D_\alpha (\Pi \circ \text{NS}_P) \|_{\text{op}, L^2} \| \text{proj}_\alpha \eta \|_{L^2}, \) where \( \| \cdot \|_{\text{op}, L^2} \) is the operator norm. On the finite-dimensional space \( H_{\alpha} \), the \( L^q \)- and \( L^2 \)-norms are equivalent: \( \| \text{proj}_\alpha \eta \|_{L^2} \leq C_2 \| \text{proj}_\alpha \eta \|_{L^q} \). This constant \( C_2 \) is independent of \( \text{proj}_\alpha \eta \in H_{\alpha} \), however it may depend on \( \alpha \). On the other hand, Proposition 3.12 tells us that \( C_2 \) is independent of \( \alpha \) provided \( \| F_\alpha \|_{L^2} \) is sufficiently small. So we have

\[
\| \text{proj}_\alpha \eta - D_\alpha (\Pi \circ \text{NS}_P) \eta \|_{L^q} \leq f(\alpha) \| \text{proj}_\alpha \eta \|_{L^q},
\]

where we have set \( f(\alpha) := C_1 C_2 \| \text{proj}_\alpha - D_\alpha (\Pi \circ \text{NS}_P) \|_{\text{op}, L^2}. \) By Theorem 3.6 and Proposition 3.12 the function \( f(\alpha) \) depends continuously on \( \alpha \) in the \( L^2 \)-topology. If \( \alpha = \alpha_0 \) is flat, then \( D_\alpha (\Pi \circ \text{NS}_P) \) equals the projection \( \text{proj}_\alpha \), and so \( f(\alpha_0) = 0. \) In particular, \( f(\alpha) \to 0 \) as \( \alpha \) approaches \( A^{1,0}_{\flat}(\Sigma) \) in the \( L^2 \)-topology. That \( f(\alpha) \to 0 \) as \( \| F_\alpha \|_{L^q} \to 0 \) follows from a contradiction argument using Uhlenbeck compactness. 

\[
\square
\]

3.2 Proof of Lemma 3.5

Note that (i) implies that Theorem 3.6 applies to \( \alpha_v(x) \) for each \( x \in K \) and \( \nu \) sufficiently large. Let \( \text{NS} \) be the map constructed in Theorem 3.6 for the bundle \( P \to \Sigma ). Define \( \nu_v : S \to M(P) \) by \( \nu_v(x) := \Pi \circ \text{NS} (\alpha_v(x)). \) Then Lemmas 3.15 and 3.16 imply that \( \nu_v \) is a holomorphic map.

Claim 1: Fix a compact \( K \subset S_0 \). The \( \nu_v \) have uniformly bounded energy density:

\[
\sup_{\nu} \sup_{x \in K} |D\nu_v|_{M(P)} < \infty.
\]

Here \( | \cdot |_{M(P)} \) is the norm on \( M(P) \) given by the \( L^2 \)-inner product on the tangent (harmonic) spaces. It suffices to prove the claim in local orthonormal coordinates \( x = (s, t) \) on \( K. \) Since \( \nu_v \) is holomorphic, \( D\nu_v(x) \) is controlled by \( \partial_s \nu_v(x) \). Suppressing the point \( x \) we have

\[
|\partial_s \nu_v|_{M(P)} = \| \partial_s (\Pi \circ \text{NS}(\alpha_v)) \|_{L^2(\Sigma)} = \| D_{\alpha_v} (\Pi \circ \text{NS}) (\partial_s \alpha_v) \|_{L^2(\Sigma)} = \| D_{\alpha_v} (\Pi \circ \text{NS}) (\text{proj}_{\alpha_v} \partial_s \alpha_v) \|_{L^2(\Sigma)},
\]

where the last equality holds by Lemma 3.16 since \( \partial_s \alpha_v \) and \( \text{proj}_{\alpha_v} \partial_s \alpha_v \) differ by an element of \( \text{im} \partial_s \alpha_v \oplus \text{im} \ast \partial_s \alpha_v. \) By Lemma 3.16 we have

\[
\| D_{\alpha_v} (\Pi \circ \text{NS}) (\text{proj}_{\alpha_v} \partial_s \alpha_v) \|_{L^2(\Sigma)} \leq C_0 \| \text{proj}_{\alpha_v} \partial_s \alpha_v \|_{L^2(\Sigma)}.
\]

By assumption (ii) in the statement of the Lemma 3.5, this last term is controlled by a constant \( C \), and this proves the claim.

It follows from Claim 1 that \( \{ \nu_v \} \) is a \( C^1 \)-bounded sequence of maps \( K \to M(P) \) for each compact \( K \subset S_0. \) By the compactness of the embedding \( C^1(K) \to C^0(K), \) there is a subsequence, still denoted by \( \{ \nu_v \}, \) which converges weakly in \( C^1, \) and
strongly in $C^0$, to some limiting holomorphic map $v_\infty : K \to M(P)$. By repeating the above with a sequence $K_n$ of compact sets that exhaust $S_0$, and by taking a diagonal subsequence, one can show that $v_\infty$ is defined on all of $S_0$ and the $v_\nu$ converge to $v_\infty$ in $C^0$ on compact subsets of $S_0$.

**Remark 3.18.** We can actually say quite a bit more: The uniform energy bound given in Claim 1 implies that, after possibly passing to a further subsequence, we have that the $v_\nu$ converge to $v_\infty$ in $C^\infty$ on compact subsets of the interior of $S_0$ (see [23, Theorem 4.1.1]).

**Claim 2:** There exists a smooth lift $\alpha_\infty : S_0 \to A_{\text{flat}}(P)$ of $v_\infty : S_0 \to M(P)$.

This is only non-trivial when $S_0$ is closed, and in this case the result basically follows because (i) each $v_\nu A_{\text{flat}}(P) \to S_0$ is trivializable ($v_\nu$ admits a lift $\alpha_\nu$), and (ii) the $v_\nu$ converge to $v_\infty$ in $C^0$ so $v_\nu A_{\text{flat}}(P)$ is trivializable. To see this explicitly, we will show that, for large enough $\nu$, there is a map $\Gamma : I \times S_0 \to M(P)$ that restricts to $v_\infty$ on $\{0\} \times S_0$ and to $v_\nu$ on $\{1\} \times S_0$. Then the result will follow since the pullback over an elementary cobordism is trivial if and only if it is trivial over one of the boundary components. To construct $\Gamma$, pick $\nu$ large enough so $\text{dist}(v_\nu(x), v_\infty(x))$ is smaller than the injectivity radius of $M(P)$. Let $\gamma_t(x)$ be the geodesic from $v_\infty(x)$ to $v_\nu(x)$ parametrized so that it has length 1. Then defining $\Gamma(t, x) := \gamma_t(x)$ proves the claim.

Now we use Claim 2 to translate the convergence of Claim 1 into a statement about $\alpha_\nu$ and $\alpha_\infty$. Note that, because $M(P)$ is finite-dimensional, we can choose any metric we want. At this point it is convenient to choose the metric on the tangent space induced from the $C^0$-norm on the harmonic spaces. In particular, the $C^0$-convergence in the $S$-directions immediately implies that, for each $x \in S_0$, there are gauge transformations $\mu_\nu(x) \in G^\infty_0(P)$ such that

$$\sup_{x \in K} \|\alpha_\infty(x) - \mu_\nu(x)^* \text{NS}(\alpha_\nu(x))\|_{C^0(\Sigma)} \to 0. \quad (47)$$

By perturbing the gauge transformations, we may suppose that each $\mu_\nu(x)$ is smooth in $x$. This gives that $\|\alpha_\infty - \mu_\nu^* \alpha_\nu\|_{C^0(K \times \Sigma)}$ is bounded by

$$\sup_{x \in K} \left\{ \left\| \mu_\nu^* \alpha_\nu - \mu_\nu^* \text{NS}(\alpha_\nu) \right\|_{C^0(\Sigma)} + \left\| \alpha_\infty - \mu_\nu^* \text{NS}(\alpha_\nu) \right\|_{C^0(\Sigma)} \right\} \leq C \sup_{x \in K} \left\{ \left\| F_{\alpha_\nu} \right\|_{C^0(\Sigma)} + \left\| \alpha_\infty - \mu_\nu^* \text{NS}(\alpha_\nu) \right\|_{C^0(\Sigma)} \right\}$$

where the inequality follows from Proposition 3.14. This last term goes to zero by assumption (i) and (47). This proves (21).

Now we prove (22). We begin with $p = 2$. We will work in local coordinates $x = (s, t)$ on int $S_0$. By the holomorphic and $\epsilon$-ASD conditions, it suffices to prove that

$$\sup_{x \in K} \left\| \text{proj}_{\alpha_\infty} \partial_s \alpha_\infty - \text{Ad}(\mu_\nu^{-1}) \text{proj}_{\alpha_\nu} \partial_s \alpha_\nu \right\|_{L^2} \to 0$$
for each compact $K \subset \text{int} \, S_0$, where here and below the $L^2$-norms are on $\Sigma$. The key ingredient is that $\partial_s v_\nu$ converges to $\partial_s v_\infty$ in $C^0$ on $K$; this is coming from Remark 3.18. We first translate this to a statement about the connections: The appropriate lift of $\partial_s v_\infty$ to $T_{\alpha_\infty} A(P)$ is the harmonic projection $\text{proj}_{\alpha_\infty} \partial_s \alpha_\infty$. Similarly, the appropriate lift of $\partial_s v_\nu(x)$ is the harmonic projection of the linearization $D_{\alpha_\nu} \text{NS} (\partial_s \alpha_\nu (x))$. This harmonic projection is exactly $D_{\alpha_\nu} (\Pi \circ \text{NS} (\partial_s \alpha_\nu))$, since $D_{\alpha} \Pi = \text{proj}_\alpha$ whenever $\alpha$ is flat. Then the $C^0$ convergence $\partial_s v_\nu \to \partial_s v_\infty$ implies that

$$\sup_K \left\| \text{proj}_{\alpha_\infty} \partial_s \alpha_\infty - \text{Ad} (\mu_\nu^{-1}) D_{\alpha_\nu} (\Pi \circ \text{NS}) (\partial_s \alpha_\nu) \right\|_{L^2} \to 0. \quad (48)$$

The gauge transformations that appear here are exactly those from the previous paragraph; this is due to the fact that since the $\mu_\nu^* \alpha_\nu$ converge to $\alpha_\infty$, the harmonic spaces $\text{Ad}(\mu_\nu^{-1}(x)) H_{\alpha_\nu(x)}$ converge to $H_{\alpha_\infty(x)}$. For each $x \in K$, the triangle inequality gives that $\| \text{proj}_{\alpha_\infty} \partial_s \alpha_\infty - \text{Ad} (\mu_\nu^{-1}) \text{proj}_{\alpha_\nu} \partial_s \alpha_\nu \|_{L^2}$ is bounded by

$$\left\| \text{proj}_{\alpha_\infty} \partial_s \alpha_\infty - \text{Ad} (\mu_\nu^{-1}) D_{\alpha_\nu} (\Pi \circ \text{NS}) (\partial_s \alpha_\nu) \right\|_{L^2} + \left\| D_{\alpha_\nu} (\Pi \circ \text{NS}) (\partial_s \alpha_\nu) - \text{proj}_{\alpha_\nu} \partial_s \alpha_\nu \right\|_{L^2} \leq \left\| \text{proj}_{\alpha_\infty} \partial_s \alpha_\infty - \text{Ad} (\mu_\nu^{-1}) D_{\alpha_\nu} (\Pi \circ \text{NS}) (\partial_s \alpha_\nu) \right\|_{L^2} + f(\alpha_\nu) \| \text{proj}_{\alpha_\nu} \partial_s \alpha_\nu \|_{L^2}$$

where the inequality here is Corollary 3.17. It follows from (48) that the first term on the right goes to zero. For the second term, note that assumption (ii) implies that $\| \text{proj}_{\alpha_\nu} \partial_s \alpha_\nu \|_{L^2}$ is bounded by some constant $C$, uniformly in $x \in K$. Then Corollary 3.17 combines with assumption (i) in the statement of Lemma 3.5 to give that $f(\alpha_\nu) \to 0$ uniformly in $x \in K$. This finishes the proof of (22) with $p = 2$.

Now we prove (22) for $1 < p < \infty$. To simplify notation, replace $\alpha_\nu$ by $\mu_\nu^* \alpha_\nu$. Note that by Proposition 3.12, for any 1-forms $\eta, \eta'$ on $\Sigma$ we have

$$\| \text{proj}_{\alpha_\infty} \eta - \text{proj}_{\alpha_\nu} \eta' \|_{L^p} \leq C \| \alpha_\infty - \alpha_\nu \|_{L^\infty} \| \eta \|_{L^p} + \| \text{proj}_{\alpha_\nu} \eta - \text{proj}_{\alpha_\nu} \eta' \|_{L^p},$$

where all norms are on $\Sigma$ and we are working pointwise on $S_0$. The harmonic space $H^1_{\alpha_\nu}$ is finite-dimensional and converging to $H^1_{\alpha_\infty}$, so the $L^p$-norm on $H^1_{\alpha_\nu}$ is equivalent to the $L^2$-norm by a constant $C'$ that is independent of $\nu$. We therefore have that $\| \text{proj}_{\alpha_\infty} \eta - \text{proj}_{\alpha_\nu} \eta' \|_{L^p}$ is controlled by

$$C \| \alpha_\infty - \alpha_\nu \|_{L^\infty} \| \eta \|_{L^p} + C' \| \text{proj}_{\alpha_\nu} \eta - \text{proj}_{\alpha_\nu} \eta' \|_{L^2} \leq C \| \alpha_\infty - \alpha_\nu \|_{L^\infty} (\| \eta \|_{L^p} + C' \| \eta' \|_{L^2}) + C' \| \text{proj}_{\alpha_\infty} \eta - \text{proj}_{\alpha_\nu} \eta' \|_{L^2}$$

where we used Proposition 3.12 again. Apply this with $\eta = \text{proj}_{\alpha_\infty} \partial_s \alpha_\infty$ (which is obviously uniformly bounded in $L^p(\Sigma)$), and $\eta' = \text{proj}_{\alpha_\nu} \partial_s \alpha_\nu$ (which is bounded uniformly on $K$ in $L^2(\Sigma)$ by assumption (ii) in the statement of Lemma 3.5). Since $\| \alpha_\infty - \alpha_\nu \|_{L^\infty}$ and $\| \text{proj}_{\alpha_\infty} \eta - \text{proj}_{\alpha_\nu} \eta' \|_{L^2}$ converge to zero uniformly on $K$, it follows that $\| \text{proj}_{\alpha_\infty} \eta - \text{proj}_{\alpha_\nu} \eta' \|_{L^p}$ does as well.
3.3 Proof of Theorem 3.3
In light of Lemma 3.5, we consider each of the following cases:

**Case 1** \( \|F_{A_v}\|_{L^\infty(Z)} \to \infty \);

**Case 2** \( \|F_{A_v}\|_{L^\infty(Z)} \to \Delta > 0 \);

**Case 3** \( \|F_{A_v}\|_{L^\infty(Z)} \to 0 \), and \( \sup_{x \in S} \|\text{proj}_{A_v} \circ D_x A_v\|_{L^2(\Sigma)} \to \infty \).

To prove the theorem, we will show that each case leads to energy quantization. That is, we will show that there is a positive constant \( \tilde{h} > 0 \), depending only on the group \( \text{PU}(r) \), with the following significance: We will show that each case above implies there is some bubbling point \( x \in S \) and a set \( T_x \subset Z \) (\( T_x \) will be either a point in \( \{x\} \times \Sigma \), or the whole fiber \( \{x\} \times \Sigma \)), such that for every neighborhood \( U \) of \( T_x \), the energy \( E_{c_i}(A_v) \geq \tilde{h} \) is uniformly bounded from below for all \( v \). To see why this implies the theorem, let \( B \) denote the set of exceptional points \( x \in S \). Since each \( A_v \) is an \( \epsilon_v \)-instanton on the bundle \( R \), it follows that \( \epsilon_v \)-energies are equal to \( 2\pi^2 r^{-1} q_4(R) \geq 0 \). This implies that \( B \) must be a finite set. Then hypotheses of Lemma 3.5 hold on \( S_0 := S \setminus B \), so the convergence result in Theorem 3.3 follows from the lemma. That \( A_{\infty} \) extends over the bubbling set \( B \) follows from the removable singularities theorem for holomorphic curves: The set \( B \) projects to a finite set in \( S \); now apply the removable singularities theorem to \( v_{\infty} \) in the proof of Lemma 3.5 and lift using Claim 2 appearing in that proof. The verification of the bound (20) is standard.

It remains to prove that the three cases above lead to energy quantization. For the first two cases we follow [8], so we only sketch the details; see also Section 4.4.

**Case 1. (Instantons on \( S^4 \))** For each \( v \) identify a point \( z_v \in Z \) where \( |F_{A_v}(z_v)| \) is maximized. To simplify notation, we may assume \( z_v = z_{\infty} \) is fixed for all \( v \) (this is approximately true by compactness of \( Z \)). Restrict attention to a small fixed neighborhood of \( z_{\infty} \). Rescale each \( A_v \) by \( \epsilon_v \), but only in the \( S \)-directions (the \( \Sigma \)-direction remain unscaled). Then we obtain a sequence of connections \( \tilde{A}_v \) on increasing and exhausting subsets of \( C \times \Sigma \) that are instantons with respect to a fixed metric, and have curvature \( |F_{\tilde{A}_v}| \) blowing up at \( z_{\infty} \in C \times \Sigma \). Now we rescale again, this time in all four directions and by the maximum of \( |F_{\tilde{A}_v}| \). By Uhlenbeck’s strong compactness theorem to the twice-rescaled sequence of connections, these converge to a non-trivial instanton on a bundle \( R_{\infty} \) over \( S^4 \). The energy of such instantons is \( \tilde{h} = 2\pi^2 r^{-1} q_4(R_{\infty}) \). Note that since \( t_2(R_{\infty}) \in H^2(S^4, \Z_r) \) obviously vanishes, \( q_4(R_{\infty}) \) is divisible by \( 2r \) (see, e.g., [11] Equation (4))), and so \( \tilde{h} = 4\pi^2 r \).

**Case 2. (Instantons on \( C \times P \))** Exactly as in the previous case, we rescale by \( \epsilon_v \) in the \( S \)-directions to obtain a sequence \( \tilde{A}_v \) of instantons on exhausting subsets of \( C \times \Sigma \) with curvature maximized at some \( z_{\infty} \in C \times \Sigma \). If these maxima \( |F_{\tilde{A}_v}| \) diverge, then we repeat the analysis of Case 1 and get an instanton on \( S^4 \). Otherwise, the curvatures are \( L^\infty \) bounded and we can apply Uhlenbeck’s strong compactness theorem directly to the \( \tilde{A}_v \) to obtain a non-flat finite-energy instanton on
of its components by coordinates converge to some \( x \) in cases where there are boundary conditions, as we will encounter in Theorem 4.1. The key ingredient is (22). Our argument has the additional bonus that it applies even when the right-hand side. This diverges to \( \epsilon \) on bundles over the domain \( \Sigma \). The finite energy instanton \( A \) limits to flat connections on the cylindrical ends \( S^1 \times \Sigma \) at \( \pm \infty \), and the energy of \( A \) is given by the difference of the Chern-Simons functional applied to each of these limiting flat connections. Then it is shown in [8] that this difference is given by \( 4\pi^2 \gamma \epsilon^{-1} \deg(u) \), where \( u \) is a certain gauge transformation on \( S^1 \times P \) and necessarily has non-zero degree (otherwise \( A \) would be flat). We can therefore take \( \epsilon = 4\pi^2 \gamma \epsilon^{-1} \).

Case 3. (Holomorphic spheres in \( M(P) \)) In this case, we rescale around a blow-up point to find that a holomorphic sphere bubbles off in the moduli space of flat connections. These rescaled connections do not satisfy a fixed ASD equation, and so Uhlenbeck’s strong compactness theorem does not apply, as it did in Cases 1 and 2. One could try to use Uhlenbeck’s weak compactness theorem, but this theorem is too weak to conclude that the limiting bubbles are non-constant. Dostoglou and Salamon resolved this issue [9] by developing several intricate estimates for these types of \( \epsilon \)-instantons. We present an alternative argument using the heat flow (the key ingredient is [22]). Our argument has the additional bonus that it applies even in cases where there are boundary conditions, as we will encounter in Theorem 4.1.

Set \( c_v := \| \text{proj}_{\beta_{\epsilon,0}} \circ D_{\alpha_v} \|_{L^2(\Sigma)} \) where \( x_v \in S \) is chosen to maximize the right-hand side. This diverges to \( \infty \), by assumption. By compactness we may assume \( x_v \) converge to some \( x_\infty \in S \), and we fix a small ball \( B_{2\epsilon}(x_\infty) \subset S \) with holomorphic coordinates \( x = (s,t) \). We may also assume \( \nu \) is large enough so that \( x_\nu \in \delta \) of \( x_\infty \). Write \( A_V \|_{L^2(\Sigma)} = a_v + \phi_v \, ds + \psi_v \, dt \) and define a rescaled connection \( \hat{A}_v \) in terms of its components by

\[
\hat{a}_v(x) := a_v^{-1} x + x_v, \quad \hat{\phi}_v(x) := c_v^{-1} \phi_v(c_v^{-1} x + x_v), \quad \hat{\psi}_v(x) := c_v^{-1} \psi_v(c_v^{-1} x + x_v). \tag{49}
\]

View these as being defined on subsets \( C \times \Sigma \). By Hofer’s Lemma [23, Lemma 4.6.4], we can refine the choice of \( x_v \) to ensure that these subsets are increasing and exhaust \( C \times \Sigma \). Write \( F_{\hat{a}_v} = F_{\hat{a}_v} - \hat{\beta}_{s,v} \wedge ds - \hat{\beta}_{t,v} \wedge dt + \hat{\gamma}_v \, ds \wedge dt \) in terms of its components. These satisfy \( \hat{\beta}_{s,v} \ast \hat{\beta}_{t,v} = 0 \) and \( \hat{\gamma} = -\hat{\epsilon}_v^{-2} \ast F_{\hat{a}_v} \), where \( \hat{\epsilon}_v := c_v \epsilon_v \). It may not be the case that \( \hat{\epsilon}_v \) is decaying to zero; this is replaced by the assumption that the slice-wise curvatures are going to zero in \( L^\infty(C \times \Sigma) \): \( \| F_{\hat{a}_v} \|_{L^\infty} = \| F_{\hat{a}_v} \|_{L^\infty} \to 0 \). We also have

\[
\| \text{proj}_{\hat{a}_v(0)} \partial_t \hat{a}_v(0) \|_{L^2(\Sigma)} = \frac{1}{2c_v} \| \text{proj}_{\hat{a}_v(x_v)} \circ D_{\alpha_v} \|_{L^2(\Sigma)} = \frac{1}{2} \tag{50}
\]

where we have used the \( \hat{\epsilon}_v \)-ASD equation in the first equality. Then the \( \hat{A}_v \) satisfy the hypotheses of Lemma 3.5 so (after possibly passing to a subsequence) there exists a sequence of gauge transformations \( U_v \in G_{2\beta}^{1}(C \times P) \), and a limiting holomorphic curve representative \( \hat{A}_\infty \in A_{1\ell}^{1\beta}(C \times P) \) such that

\[ C \times P \to C \times \Sigma. \]
\[ \sup_{K} \| \text{proj}_{\hat{\alpha}_v} \partial_s \hat{\alpha}_\infty - \text{Ad}(\mu_{\nu}^{-1}) \text{proj}_{\hat{\alpha}_v} \partial_s \hat{\alpha}_\nu \|_{L^2(\Sigma)} \xrightarrow{v} 0, \]

for all compact \( K \subset \mathbb{C} \). Then this descends to a finite-energy holomorphic curve \( v_\infty : \mathbb{C} \to M(P) \), which is non-trivial by [50]. By removal of singularities [23] Theorem 4.1.2 (ii) it follows that \( v_\infty \) extends to a holomorphic sphere \( v_\infty : S^2 \to M(P) \). We have energy quantization with \( \bar{\hbar} = \frac{4\pi^2 \kappa_r}{r} - 1 \) for non-constant holomorphic spheres [6, Corollary 6.3], which completes Case 3.

4 Compactness for cylinders \( \mathbb{R} \times Y \)

Fix a broken circle fibration \( f : Y \to S^1 \) as in [10]. This means that \( Y \) is a closed, connected, oriented 3-manifold, and \( f \) is a Morse function with connected and nonempty fibers. We will assume that the number \( N \) of critical points of \( f \) is positive (otherwise we are essentially in the case of Section 3). We also assume the critical points of \( f \) have distinct critical values. This implies that \( N \) is even, and also allows us to view \( Y \) as a composition of cobordisms:

\[ Y_{01} \cup \Sigma_1 (I \times \Sigma_1) \cup \Sigma_2 Y_{12} \cup \Sigma_2 \ldots \cup \Sigma_{N-1} Y_{(N-1)0} \cup \Sigma_0 (I \times \Sigma_0) \cup \Sigma_0 \]

(51)

where \( I := [0,1] \) is the unit interval, each \( \Sigma_i \subset Y \) is a fixed regular fiber of \( f \), and each \( Y_{(i+1)} \subset Y \) is a cobordism from \( \Sigma_i \) to \( \Sigma_{i+1} \) such that \( f|_{Y_{(i+1)}} \) has exactly one critical point. Note that (51) is cyclic in the sense that the cobordism \( I \times \Sigma_0 \) on the right is glued to the cobordism \( Y_{01} \) on the left, reflecting the fact that \( f \) maps to the circle. We set \( Y_* := \sqcup Y_{(i+1)} \) and \( \Sigma_* := \sqcup \Sigma_i \).

Figure 1: An illustration of a broken circle fibration with \( N = 2 \). Each \( Y_{(i+1)} \) has a unique critical point \( p_i \), with corresponding critical value \( c_i \).

In this section we consider the 4-manifold \( Z := \mathbb{R} \times Y \). We equip \( Z \) with a product metric \( g = ds^2 + g_Y \), where \( g_Y \) is a metric on \( Y \). To simplify the exposition, we assume \( g_Y \) has been chosen so that \( g_Y|_{I \times \Sigma_*} = dt^2 + g_\Sigma \), where \( g_\Sigma \) is a metric on \( \Sigma_* \) that is independent of \( t \in I \). Then we define a metric \( g_\epsilon \) by setting

\[ g_\epsilon|_{\mathbb{R} \times Y_*} := ds^2 + \epsilon^2 g_Y|_{\mathbb{R} \times Y_*} \quad g_\epsilon|_{\mathbb{R} \times I \times \Sigma_*} := ds^2 + dt^2 + \epsilon^2 g_\Sigma. \]
Figure 2: Picture above are four copies of the manifold $Y$, viewed from the ‘top’ relative to the illustration in Figure 1. Each copy of $Y$ has $N = 4$ critical points, represented by stars *. Moving from left to right, the different copies represent the metric $g_\epsilon$ on $\tilde{Y} = \{pt\} \times Y \subset \mathbb{R} \times Y$ as $\epsilon$ decreases to zero. Notice that the volumes of the $\Sigma_i$ and the $Y_{t(i+1)}$ are going to zero. However, the length in the $I$-direction (the ‘neck’) of each $I \times \Sigma_i$ is remaining fixed. In the picture on the far right, the $Y_{t(i+1)}$ have collapsed entirely to the critical points of $f$.

See Figure 2 When $\epsilon \neq 1$, the metric $g_\epsilon$ is not smooth on $Z$ with respect to the given smooth structure on $Z$. However, there is a different smooth structure on $Z$ in which $g_\epsilon$ is smooth. We call this the $\epsilon$-dependent smooth structure, and say a function, tensor, connection, etc. is $\epsilon$-smooth if it is smooth with respect to the $\epsilon$-dependent smooth structure. See [10, Section 2.1] for more details.

We take $R \to Z$ to be the pullback of a PU($r$)-bundle $Q \to Y$ such that $t_2(Q)$ is Poincaré dual to $d[\gamma] \in H_1(Y, \mathbb{Z}_r)$, where $d \in \mathbb{Z}_r$ is a generator, and $\gamma : S^1 \to Y$ is a section of $f : Y \to S^1$. Set

$$Q_{t(j+1)} := Q|_{Y_{t(j+1)}}, \quad Q_\gamma := \sqcup_j Q_{t(j+1)}, \quad P_j := Q|_{\{0\} \times \Sigma_j}, \quad P_\gamma := \sqcup_j P_j.$$ 

The assumption on $t_2(Q)$ ensures $t_2(Q_{t(j+1)}) \left[Q_{t(j+1)}\right] = t_2(P_j) \left[P_j\right] = d$. In [11] we show that the fibers $\Sigma_\gamma \subset Y$ determine a connected component in $G(Q)$ consisting of degree 1 gauge transformations. We let $G_\gamma \subset G(Q)$ denote the subgroup generated by this component.

We say that a connection on $R$ is $\epsilon$-ASD or an $\epsilon$-instanton if it is ASD with respect to $g_\epsilon$. In coordinates over $\mathbb{R} \times I \times \Sigma_\gamma \subset Z$, this takes the form [13]. Over $\mathbb{R} \times Y_\gamma$, any connection can be written as $a(s) + p(s) \, ds$, where $a(s)$ is a path of connections on $Q$, and $p(s)$ is a path of $Q(g)$-valued 0-forms on $Y$. Then the $\epsilon$-ASD condition on $\mathbb{R} \times Y_\gamma$ takes the form

$$\partial_s a(s) - da(s) p(s) + \epsilon^{-1} \ast_Y F_{a(s)} = 0,$$

where $\ast_Y$ is the Hodge star coming from $g_Y$. As before, the $\epsilon$-energy is $E^{\text{inst}}_\epsilon(A) := \frac{1}{2} \int_Z \langle F_A \wedge \ast \epsilon F_A \rangle$ (this makes sense for any connection on $R$ that is $W^{1,2}$ with respect
to the $\epsilon$-smooth structure). If $A$ is $\epsilon$-ASD, then the $\epsilon$-energy is finite if and only if $A$ limits to flat connections at $\pm \infty$ in the sense that there are flat connections $a^\pm \in \mathcal{A}_{\text{flat}}(Q)$ such that $\lim_{s \to \pm \infty} a(s) = a^\pm$, and $\lim_{s \to \pm \infty} \psi(s) = 0$. When this is the case, $E^{\text{gast}}_\epsilon(A) = CS(a^+) - CS(a^-)$, where $CS$ is the Chern-Simons functional. In particular, this is independent of $\epsilon$ and $A$.

Now we describe the symplectic theory. Restricting to each of the two boundary components of $Y_{ij(j+1)}$ determines a Lagrangian embedding

$$L(Q_{ij(j+1)}) \hookrightarrow M(P_j)^- \times M(P_{j+1}).$$

Then we set

$$M := M(P_0)^- \times M(P_1) \times M(P_2)^- \times \ldots \times M(P_{N-1})$$

$$L_{(0)} := L(Q_{01}) \times L(Q_{23}) \times \ldots \times L(Q_{(N-2)(N-1)})$$

$$L_{(1)} := L(Q_{12}) \times L(Q_{34}) \times \ldots \times L(Q_{(N-1)N});$$

the superscript in $M(P_2)^-$ means that we are using the negative of the given symplectic structure on this manifold. Then there are natural Lagrangian embeddings $L_{(0)}, L_{(1)} \hookrightarrow M$. The Hodge star from $\xi$ determines a compatible almost complex structure $J$ on $M$. The intersection points $L_{(0)} \cap L_{(1)}$ can be canonically identified with the space $\mathcal{A}_{\text{flat}}(Q)/\mathcal{G}_\xi$ of flat connections on $Y$ modulo gauge; see [10].

Lagrangian intersection Floer homology considers strips $\mathbb{R} \times I \to M$ with Lagrangian boundary conditions. For our purposes, we are interested in certain lifts of these to the space of connections on $\mathbb{R}$. That is, we define a strip representative to be a connection $A_0 \in \mathcal{A}(\mathbb{R})$ satisfying the following: Over $\mathbb{R} \times I \times \Sigma$, $A_0$ has the form $a + \phi ds + \psi dt$, where $a : \mathbb{R} \times I \to \mathcal{A}_{\text{flat}}(\Sigma, P^\bullet(g))$, and $\phi, \psi : \mathbb{R} \times I \to \Omega^0(\Sigma, P^\bullet(g))$ are $\partial$-forms defined so that $\partial_a a - d_a \phi$ and $\partial_a a - d_a \psi$ are harmonic; see Example 3.1. Over $\mathbb{R} \times Y$, $A_0$ has the form $a + p ds$, where $a : \mathbb{R} \to \mathcal{A}_{\text{flat}}(\Sigma^\bullet)$ is determined uniquely (up to gauge) by the condition that $a(s)|_{\partial Y^\bullet} = a(s, \cdot)|_{\partial Y \times \Sigma}$, and $p : \mathbb{R} \to \Omega^0(Y, \nabla^\bullet(g))$ is determined by the condition that $\partial_s a(s) - d_{\partial Y}(s) p(s) = a(s)$-harmonic. It follows that any strip representative $A_0$ is a continuous connection on $Z$ that is $W^{1,p}$ with respect to the $\epsilon$-dependent smooth structure for any $\epsilon > 0$. We will often write $A_0 = a + p ds$ on all of $Z$, and so $a|_{\mathbb{R} \times I \times \Sigma} = a + \psi dt$. It follows immediately that any strip representative descends to give a map $\mathbb{R} \times I \to M$ with Lagrangian boundary conditions in the $L_{(j)}$. Moreover, this projection gives the space of strip representatives the structure of a principal bundle over the space of maps $\mathbb{R} \times I \to M$ with Lagrangian boundary conditions. We define the energy of a strip representative $A_0$ by the formula

$$E^{\text{symp}}(A_0) := \frac{1}{2} \int_{\mathbb{R} \times I \times \Sigma,} \left| \partial_a a - d_a \phi \right|^2 + \left| \partial_a a - d_a \psi \right|^2 \text{dvol}.$$ 

This is concocted so that it recovers the energy of the curve that $A_0$ represents. We will say that a strip representative $A_0$ is a holomorphic strip representative if, over
\( \mathbb{R} \times I \times \Sigma \), its components satisfy \([19]\). When \( A_0 \) is a holomorphic strip representative, then the energy \( E^{\text{sym}}(A_0) \) is finite if and only if \( A_0 \) limits to flat connections \( a^\pm \) at \( \pm \infty \); see \([10]\). In this case we have

\[
E^{\text{sym}}(A_0) = CS(a^+) - CS(a^-)
\]

and so the energy is again a topological quantity only depending on the limiting connections (this essentially follows because the symplectic action functional for \( M \) is given by the Chern-Simons functional of representatives; see also Lemma \([3.2]\).

For \( s_0 \in \mathbb{R} \), and \( A \in \mathcal{A}(\mathbb{R}) \), let \( \tau_{s_0}^* A \in \mathcal{A}(\mathbb{R}) \) be the connection defined by translating \( \tau_{s_0} A|_{\{s\} \times Y} := A|_{\{s+s_0\} \times Y} \). Given \( x \in \mathbb{R} \times I \), we will use \( t_x : \Sigma_\ast = \{x\} \times \Sigma \hookrightarrow Z \) to denote the inclusion, and so the pullback \( a(x) := \iota^*_x A \) can be viewed as an \( \mathbb{R} \times I \)-dependent connection on \( P_\ast \to \Sigma_\ast \). Similarly, if \( s \in \mathbb{R} \), then \( t_s : \{s\} \times Y \hookrightarrow Z \) is the inclusion, and the restriction \( a(s) := \iota^*_s A \) can be viewed as an \( \mathbb{R} \)-dependent connection on \( Q \to Y \). Now we can state the main theorem.

**Theorem 4.1.** Fix \( 2 < q < \infty \), and let \( R \to Z \) be as above. Assume all flat connections on \( Q \) are non-degenerate, and fix flat connections \( a^\pm \in A^{\text{flat}}(Q) \). Suppose \( (\epsilon_r)_{r \in \mathbb{N}} \) is a sequence of positive numbers converging to 0, and assume that, for each \( v \), there is an \( \epsilon_v \)-ASD connection \( A_v \in A^{1,q}_{\text{loc}}(R) \) that limits to \( a^\pm \) at \( \pm \infty \). Then there is

(i) a finite set \( B \subset \mathbb{R} \times I \);

(ii) a subsequence of the \( A_v \) (still denoted \( A_v \));

(iii) a sequence of gauge transformations \( U_v \in G^{2,q}_{\text{loc}}(R) \);

(iv) a finite sequence of flat connections \( \{a^0 = a^- , a^1 , \ldots , a^{l-1} , a^{l} = a^+\} \subseteq A^{\text{flat}}(Q) \);

(v) for each \( j \in \{1, \ldots , l\} \), a holomorphic strip representative \( A_j^l \in A^{1,q}_{\text{loc}}(R) \) limiting to \( u_{j-1}^l , a^{l-1} \) at \( -\infty \) and \( u_j^l , a^l \) at \( +\infty \), for some \( u_{j-1}^l , u_j^l \in \mathcal{G}_\Sigma \), possibly depending on \( A^l \);

(vi) for each \( j \), a sequence \( s_j^l \in \mathbb{R} \)

such that for each \( j \in \{1, \ldots , l\} \), the restrictions

\[
\sup_{x \in K} \left\| \iota_x^* (U_{s_{j}}^l \tau_{s_{j}}^* A_v - A^l) \right\|_{C^0(\Sigma)} \overset{\nu}{\longrightarrow} 0
\]

(52)

converge to zero for every compact \( K \subset \mathbb{R} \times I \setminus \overline{B} \). The gauge transformations \( U_v \) can be chosen so that they restrict to the identity component \( \mathcal{G}_0(P_i) \) on each \( \{x\} \times \Sigma_i \subset Z \). Moreover, for each \( b \in B \) there is a positive integer \( m_b > 0 \) such that for any \( \nu \),

\[
\sum_{j=1}^{l} E^{\text{sym}}(A^l_j) \leq E_{\epsilon_v}^{\text{inst}}(A_v) - 4\pi^2 \kappa r^{-1} \sum_{b \in B} m_b.
\]

(53)

Finally, \( E^{\text{sym}}(A^l_j) > 0 \) for each \( j \).
To simplify the exposition, we have assumed that all flat connections \( a \in A_{\text{flat}}(Q) \) are non-degenerate, meaning that \( d_a \) is injective on 1-forms. In general, this need not be the case. However, non-degeneracy can always be achieved by first performing a suitable perturbation to the defining equations. See Section 5.

For \( x \in \mathbb{R} \times I \), we will continue to use the notation \( \alpha_v(x) = t^*_x A_v, \alpha^j(x) = t^*_x A^j \) and \( \mu(x) = t^*_x U \). Then the conclusion of the theorem says that for each \( j \) and compact \( K \subset \mathbb{R} \times I \setminus B \), the sequence

\[
\sup_{(s,t) \in K} \| \mu(s,t)^* \alpha_v(s-t^j_v,t) - \alpha^j(s,t) \|_{C_0(\Sigma^\nu)} \to 0
\]

converges to zero; see Remark 3.4 (d). As in Section 3, it is convenient to first prove a modified version that a priori excludes bubbling.

**Lemma 4.2.** Let \( S_0 := \mathbb{R} \times I \setminus B \), where \( B \) is any finite set, and let \( 2 < q < \infty \). Suppose \((\epsilon_v)_{v \in \mathbb{N}}\) is a sequence of positive numbers (not necessarily converging to zero), and that for each \( v \) there is an \( \epsilon_v \)-ASD connection \( A_v \in \mathcal{A}^{1,q}(\mathbb{R}) \) with uniformly bounded \( \epsilon_v \)-energy, and satisfying the following conditions:

(i) For each compact \( K \subset S_0 \), the slice-wise curvatures on \( \Sigma^\nu \) converge to zero:

\[
\sup_{(s,t) \in K} \| F_{A_v(s,t)} \|_{L^\infty(\Sigma^\nu)} \to 0.
\]

(ii) For each compact \( L \subset \mathbb{R} \) with \( L \times \{0,1\} \cap B = \emptyset \), the slice-wise curvatures on \( Y^\nu \) converge to zero:

\[
\sup_{s \in L} \| F_{A_v(s)} \|_{L^\infty(Y^\nu)} \to 0, a_v(s) := t^*_s A_v.
\]

(iii) For each compact \( K \subset S_0 \), there is some constant \( C \) with

\[
\sup_v \sup_{(s,t) \in K} \| \partial_s a_v(s,t) - d_{a_v(s,t)} \phi_v(s,t) \|_{L^2(\Sigma^\nu)} \leq C.
\]

(iv) For each compact \( L \subset \mathbb{R} \) with \( L \times \{0,1\} \cap B = \emptyset \), there is some constant \( C \) with

\[
\sup_v \sup_{s \in L} \| \partial_s a_v(s) - d_{a_v(s)} p_v(s) \|_{L^2(Y^\nu)} \leq C.
\]

Then there is a subsequence of the connections (still denoted \( A_v \)), a sequence of gauge transformations \( U_v \) on \( R|_{S_0 \times \Sigma} \), and a holomorphic strip representative \( A_\infty \in \mathcal{A}^{1,q}_{\text{loc}}(\mathbb{R}) \) such that:

\[
\sup_{(s,t) \in K} \left\| \alpha_\infty(s,t) - \mu_v(s,t)^* \alpha_v(s,t) \right\|_{C^0(\Sigma^\nu)} \to 0, \quad (54)
\]

\[
\sup_{s \in L} \left\| a_\infty(s) - u_v(s)^* a_v(s) \right\|_{L^4(Y^\nu)} \to 0, \quad (55)
\]
\[ \sup_{(s,t) \in K} \left( \left\| \partial_s a_\infty(s,t) - d_{a_\infty(s,t)} \phi_\infty(s,t) \right\|_{L^2(\Sigma^*)} - \left\| \partial_s a_\nu(s,t) - d_{a_\nu(s,t)} \phi_\nu(s,t) \right\|_{L^2(\Sigma^*)} \right)^\nu \to 0, \]  

\[ \sup_{s \in L} \left( \left\| \partial_s a_\infty(s) - d_{a_\infty(s)} p_\infty(s) \right\|_{L^2(Y^*)} - \left\| \partial_s a_\nu(s) - d_{a_\nu(s)} p_\nu(s) \right\|_{L^2(Y^*)} \right)^\nu \to 0, \]

for any compact \( K \subseteq S_0 \), and any compact \( L \subset \mathbb{R} \) with \( L \times \{0,1\} \cap B = \emptyset \).

The key technical point of this lemma is that the convergence in (56) holds even for \( K \) that intersect the boundary of \( \mathbb{R} \times I \); compare with (22). We also mention that the connections \( \tau_\nu^s A_\nu \) from Theorem 4.1 satisfy the same type of convergence as in (56), for compact \( K \subset \mathbb{R} \times I \setminus B \).

We point out that Lemma 4.2 makes a stronger hypothesis in (iii) relative the analogous assumption (ii) in Lemma 3.5 (there is no \( \text{proj}_a \) appearing in Lemma 4.2 (iii)). We use the stronger assumption to prove the various elliptic estimates in Section 4.2. On a related note, the conclusion (56) is weaker than the analogous (22), since the former only claims that the values of the norms converge, whereas the latter makes a claim about the convergence of the vector values of the functions.

In the next section we review the heat flow on 3-manifolds. This will be used to obtain Lagrangian boundary conditions for the limiting holomorphic strip representatives appearing in these theorems. We then develop several estimates that allow us to obtain the convergence in (56) in the case when \( K \) intersects the boundary of \( \mathbb{R} \times I \). The proofs of Lemma 4.2 and Theorem 4.1 will appear in Sections 4.3 and 4.4, respectively.

### 4.1 The heat flow on cobordisms

Suppose \( Q \) is principal \( G \)-bundle over a Riemannian 3-manifold \( Y \). In his thesis \[26\], Råde studied the Yang-Mills heat flow; that is, he studied solutions \( \tau \mapsto a(\tau) \in \mathcal{A}(Q) \) to the gradient flow of the Yang-Mills functional

\[ \frac{d}{d\tau} a(\tau) = -d_{a(\tau)}^* F_{a(\tau)}, \quad a(0) = a, \]

where \( a \in \mathcal{A}(Q) \) is an initial condition. Specifically, Råde proved the following:

**Theorem 4.3.** Suppose \( G \) is compact and \( Y \) is a closed, oriented manifold of dimension 3. Let \( a \in \mathcal{A}^{1,2}(Q) \). Then (58) has a unique solution \( \{ \tau \mapsto a(\tau) \} \in C^0_{\text{loc}} ([0,\infty), \mathcal{A}^{1,2}(Q)) \), with the further property that \( F_{a(\cdot)} \in C^0_{\text{loc}} ([0,\infty), L^2) \cap L^2_{\text{loc}} ([0,\infty), W^{1,2}) \). Moreover, the limit \( \lim_{\tau \to \infty} a(\tau) \) exists, is a critical point of the Yang-Mills functional, and varies continuously with the initial data \( a \) in the \( W^{1,2} \)-topology.
Differentiating $\mathcal{YM}_Q(a(\tau))$ in $\tau$ and using\(^{[58]}\) shows that $\mathcal{YM}_Q(a(\tau))$ decreases in $\tau$. Moreover, it follows from Uhlenbeck’s compactness theorem together with\(^{[26, Proposition 7.2]}\) that the critical values of the Yang-Mills functional are discrete. Combining these two facts, it follows that there is some $\tilde{c}_Q > 0$ such that if $\mathcal{YM}_Q(a) < \tilde{c}_Q$, then the associated limiting connection $\lim_{\tau} a(\tau)$ is flat. The flow therefore defines a continuous gauge equivariant deformation retract

$$\text{Heat}_Q : \left\{ a \in A^{1,2}(Q) \mid \mathcal{YM}_Q(a) < \tilde{c}_Q \right\} \rightarrow A^{1,2}_\text{flat}(Q) \quad (59)$$

whenever $Y$ is a closed 3-manifold.

**Remark 4.4.** Råde’s theorem continues to hold, exactly as stated, in dimension 2 as well. Given a bundle $P \rightarrow \Sigma$ over a closed connected oriented surface, we therefore have that $\text{NS}_P$ and $\text{Heat}_P$ are both maps of the form $\left\{ a \in A^{1,2}(P) \mid \mathcal{YM}_P(a) < \epsilon_P \right\} \rightarrow A^{1,2}_\text{flat}(P)$, for some $\epsilon_P > 0$. It turns out these are equal, up to a gauge transformation. That is,

$$\Pi \circ \text{NS}_P = \Pi \circ \text{Heat}_P, \quad (60)$$

where $\Pi : A^{1,2}_\text{flat}(P) \rightarrow A^{1,2}_\text{flat}(P)/G_0^{2,2}(P)$ is the quotient map. Though we will not use this fact in this paper, we sketch a proof at the end of this section for completeness.

In the remainder of this section we prove a version of Råde’s Theorem\(^{4.3}\) but for bundles $Q$ over 3-manifolds with boundary. The most natural boundary condition for our application is of Neumann type. This will allow us to use a reflection principle and thereby appeal directly to Råde’s result for closed 3-manifolds.

Råde’s result holds with the $W^{1,2}$-topology. However, on 3-manifolds not all $W^{1,2}$-sections are continuous. This makes the issue of boundary conditions rather tricky. One way to get around this is to observe that, in dimension 3, restricting $W^{1,2}$-functions to codimension-1 subspaces is in fact well-defined. We take an equivalent approach by considering the space $A^{1,2}(Q, \partial Q)$, which we define to be the $W^{1,2}$-closure of the set of smooth $a \in A(Q)$ that satisfy

$$i_{\partial_n} a |_{U} = 0 \quad (61)$$

on some neighborhood $U$ of $\partial Q$ ($U$ may depend on $a$). Here $\partial_n \in \Gamma(TQ)$ is a fixed extension of the outward pointing unit normal to $\partial Q$; we may assume that the set $U$ is always contained in the region in which $\partial_n$ is non-zero. Use the normalized gradient flow of $\partial_n$ to write $U = [0, \epsilon) \times \partial Y$. Let $t$ denote the coordinate on $[0, \epsilon)$. Then in these coordinates we can write any connection as $a|_{(t) \times \partial Y} = a(t) + \psi(t) dt$. Then \(^{[61]}\) is equivalent to requiring $\psi(t) = 0$.

Set $A^{1,2}_\text{flat}(Q, \partial Q) := A^{1,2}(Q, \partial Q) \cap A^{1,2}_\text{flat}(Q)$. Both of the spaces $A^{1,2}(Q, \partial Q)$ and $A^{1,2}_\text{flat}(Q, \partial Q)$ admit the action of the subgroup $\mathcal{G}(Q, \partial Q) \subset \mathcal{G}(Q)$ consisting of gauge transformations that restrict to the identity in a neighborhood of $\partial Q$. (We are purposefully only working with the smooth gauge transformations here.)

**Theorem 4.5.** Let $G$ be a compact, connected Lie group, and $Q \rightarrow Y$ be a principal $G$-bundle over a compact, connected, oriented Riemannian 3-manifold $Y$ with non-empty boundary.
1. There is some \( \epsilon_Q > 0 \) and a continuous strong deformation retract

\[
\text{Heat}_Q : \{ a \in \mathcal{A}^{1,2}(Q, \partial Q) \mid \mathcal{Y}\mathcal{M}_Q(a) < \epsilon_Q \} \to \mathcal{A}_{\text{flat}}^{1,2}(Q, \partial Q).
\]

Furthermore, \( \text{Heat}_Q \) intertwines the action of \( \mathcal{G}(Q, \partial Q) \).

2. Suppose \( \Sigma \subset Y \) is an embedded surface that is closed and oriented. Suppose further that either \( \Sigma \subset \text{int} \, Y \), or \( \Sigma \subset \partial Y \). Then for every \( \epsilon > 0 \), there is some \( \delta > 0 \) such that if \( a \in \mathcal{A}^{1,2}(Q, \partial Q) \) satisfies \( \|F_a\|_{L^2(\Sigma)} < \delta \), then \( \|(\text{Heat}_Q(a) - a)|\Sigma\|_{L^2(\Sigma)} < \epsilon \), for every \( 1 \leq q \leq 4 \).

**Remark 4.6.** Recently, Charalambous \[4\] has proven similar results for manifolds with boundary.

**Proof.** Consider the double \( Y^{(2)} := Y \cup_{\partial Y} Y \), which is a closed 3-manifold. Denote by \( \iota_Y : Y \to Y^{(2)} \) the inclusion of the second factor. We will identify \( Y \) with its image under \( \iota_Y \). There is a natural involution \( \sigma : Y^{(2)} \to Y^{(2)} \) defined by switching the factors in the obvious way. Then \( Y^{(2)} \) has a natural smooth structure making \( \iota_Y \) smooth and \( \sigma \) a diffeomorphism (this is just the smooth structure obtained by choosing the same collar on each side of \( \partial Y \)). Clearly the map \( \sigma \) is orientation-reversing, satisfies \( \sigma^2 = \text{Id} \) and has fixed point set equal to \( \partial Y \). Similarly, we can form \( Q^{(2)} := Q \cup_{\partial Q} Q \) and an involution \( \tilde{\sigma} : Q^{(2)} \to Q^{(2)} \). Then \( Q^{(2)} \) is naturally a principal \( G \)-bundle over \( Y^{(2)} \) and \( \tilde{\sigma} \) is a bundle map covering \( \sigma \). Furthermore, \( \tilde{\sigma} \) commutes with the \( G \)-action on \( Q^{(2)} \).

Though \( \tilde{\sigma} \) is not a gauge transformation (it does not cover the identity), it behaves as one in many ways. For example, since \( \tilde{\sigma} \) a bundle map, the space of connections \( \mathcal{A}(Q^{(2)}) \) is invariant under pullback by \( \tilde{\sigma} \). The action on covariant derivatives takes the form \( d_{\tilde{\sigma}\ast} = \sigma^* \circ d_{\ast} \circ \sigma^* \), where \( \sigma^* : \Omega(Y^{(2)}, Q^{(2)}(g)) \to \Omega(Y^{(2)}, Q^{(2)}(g)) \) is pullback by \( \sigma \). The induced action on the tangent space \( T_{a} \mathcal{A}(Q^{(2)}) = \Omega^1(Y^{(2)}, Q^{(2)}(g)) \) is given by pullback by \( \sigma \). Likewise, the curvature satisfies \( F_{\tilde{\sigma}\ast} = \sigma^* F_{\ast} \). In particular, the flow equation \[58\] on the double \( Y^{(2)} \) is invariant under the action of \( \tilde{\sigma} \). We set \( \epsilon_{Q^{(2)}} := \epsilon_{Q^{(2)}}/2 \), where \( \epsilon_{Q^{(2)}} > 0 \) is as in \[59\].

Now suppose \( a \in \mathcal{A}^{1,2}(Q, \partial Q) \) has \( \mathcal{Y}\mathcal{M}_Q(a) < \epsilon_Q \). Then \( a \) has a unique extension \( a^{(2)} \) to all of \( Q^{(2)} \), satisfying \( \tilde{\sigma}\ast a^{(2)} = a^{(2)} \). We call \( a^{(2)} \) the double of \( a \), and we claim that \( a^{(2)} \in \mathcal{A}^{1,2}(Q^{(2)}) \). To see this, first suppose that \( a \) is smooth. Then the boundary condition on \( a \) implies that \( a^{(2)} \) is continuous on all of \( Q^{(2)} \) and smooth on the complement of \( \partial Q \). In particular, \( a^{(2)} \) is \( W^{1,2} \). (Note that in general \( a^{(2)} \) will not be smooth, even if \( a \) is. For example, the normal derivatives on each side of the boundary do not agree: \( \lim_{y \to \partial Y} \partial_n a = - \lim_{y \to \partial Y} \partial_n \tilde{\sigma}\ast a \), unless they are both zero, and this latter condition is not imposed by our boundary conditions.) More generally, every \( a \in \mathcal{A}^{1,2}(Q, \partial Q) \) is a \( W^{1,2} \)-limit of smooth connections \( a_j \) whose normal component vanishes in a neighborhood of the boundary. By the linearity of the integral it is immediate that the doubles of the \( a_j \) converge to \( a^{(2)} \) in \( W^{1,2} \), and this proves the claim.
By assumption, we have $\mathcal{Y}M_{Q^{(2)}}(a^{(2)}) < \bar{c}_{Q^{(2)}}$, so by the discussion at the beginning of this section, there is a unique solution $a^{(2)}(\tau)$ to the flow equation on the closed 2-manifold $Y^{(2)}$, with initial condition $a^{(2)}(0) = a^{(2)}$. Furthermore, the limit $\text{Heat}_{Q^{(2)}}(a^{(2)}):= \lim_{\tau \to \infty} a^{(2)}(\tau)$ exists and is flat. Since (58) is $\bar{\sigma}$-invariant, the uniqueness assertion guarantees that $\bar{\sigma} a^{(2)}(\tau) = a^{(2)}(\tau)$ for all $\tau$. In particular,

$$\bar{\sigma} \text{Heat}_{Q^{(2)}}(a^{(2)}) = \text{Heat}_{Q^{(2)}}(a^{(2)}).$$

Define $\text{Heat}_{Q}(a) := \text{Heat}_{Q^{(2)}}(a^{(2)})|_{Q}$. Then (62) shows that

$$\nu_{a} \text{Heat}_{Q}(a)|_{\partial Y} = 0,$$

so $\text{Heat}_{Q}$ does map into $A_{\text{flat}}^{1,2}(Q, \partial Q)$. Similarly, gauge transformation $u \in \mathcal{G}(Q, \partial Q)$ has a unique extension to a $\bar{\sigma}$-invariant gauge transformation in $\mathcal{G}(Q^{(2)})$. In particular, $\text{Heat}_{Q}(u^{*}a) = u^{*} \text{Heat}_{Q}(a)$ follows from the $\mathcal{G}(Q^{(2)})$-equivariance of $\text{Heat}_{Q^{(2)}}$. This finishes the proof of 1.

To prove 2, we will assume $\Sigma \subset \text{int } Y$. The remaining case $\Sigma \subset \partial Y$ follows by replacing $Y$ with its double, for then we have $\Sigma \subset \text{int } Y^{(2)}$ and the analysis carries over directly. For sake of contradiction, suppose there is some sequence $a_{v} \in A_{\text{flat}}^{1,2}(Q)$ with $\|F_{a_{v}}\|_{L^{2}} \to 0$, but

$$c_{0} \leq \|\text{Heat}_{Q}(a_{v}) - a_{v}\|_{L^{2}(\Sigma)}$$

for some fixed $c_{0} > 0$. By Uhlenbeck’s weak compactness theorem, there is a sequence of gauge transformations $u_{v} \in \mathcal{G}^{2,2}$ such that $u_{v}^{*}a_{v}$ converges weakly in $W^{1,2}$ (hence strongly in $L^{4}$) to a limiting connection $a_{\infty} \in A_{\text{flat}}^{1,2}(Q)$, after possibly passing to a subsequence. Then $a_{\infty}$ is necessarily flat. Be redefining $u_{v}$, if necessary, we may assume that each $u_{v}^{*}a_{v}$ is in Coulomb gauge with respect to $a_{\infty}$, and still retain the fact that $u_{v}^{*}a_{v}$ converges to $a_{\infty}$ strongly in $L^{4}$. Then

$$\|u_{v}^{*}a_{v} - a_{\infty}\|_{W^{1,2}}^{2} = \|u_{v}^{*}a_{v} - a_{\infty}\|_{L^{2}}^{2} + \|d_{a_{\infty}}(u_{v}^{*}a_{v} - a_{\infty})\|_{L^{2}}^{2} \leq C_{1} \left( \|u_{v}^{*}a_{v} - a_{\infty}\|_{L^{2}}^{2} + \|F_{a_{v}}\|_{L^{2}}^{2} + \|u_{v}^{*}a_{v} - a_{\infty}\|_{L^{4}}^{4} \right)$$

for some constant $C_{1}$. Observe that the right-hand side is going to zero, so $a_{v}$ is converging in $W^{1,2}$ to the space of flat connections, and so

$$\|a_{v} - (u_{v}^{-1})^{*}a_{\infty}\|_{W^{1,2}} \to 0.$$  

(64)

On the other hand, by the trace theorem [54, Theorem B.10], we have

$$c_{0} \leq \|\text{Heat}_{Q}(a_{v}) - a_{v}\|_{L^{q}(\Sigma)} \leq C_{2} \|\text{Heat}_{Q}(a_{v}) - a_{v}\|_{W^{1,2}(Y)}$$

(65)

for some $C_{2}$ depending only on $Y$ and $1 \leq q \leq 4$ (the inequality on the left is (63)). Since $\text{Heat}_{Q}$ is continuous in the $W^{1,2}$-topology, and restricts to the identity on the
space of flat connections, there is some \( e' > 0 \) such that if \( a_v \) is within \( e' \) of the space of flat connections, then \( C_2 \| \text{Heat}^\gamma \nu (a_v) - a_v \|_{W^{1,2}(\gamma)} \leq \frac{C}{2} \). By (64) \( a_v \) is within \( e' \) of \( A_{\text{flat}}(Q) \) for \( v \) large, and so we have a contradiction to (65).

The next lemma states that we can always put a connection \( a \in A(Q) \) in a gauge so that it is an element of \( A(Q, \partial Q) \). This is basically just a variation on the fact that connections can be put in temporal gauge, so we omit the proof. We state a version with an additional \( R \) parameter, since this is the context in which the lemma will be used.

**Lemma 4.7.** Let \( R \to Z \) be as in the introduction to Section 4. Then for every \( A \in \mathcal{A}_{loc}^{1,2}(R) \) there is an identity-component gauge transformation \( U \in \mathcal{G}_{loc}^*(R) \) with

\[
U^* A|_{s \times \gamma_{(i+1)}} \in \mathcal{A}^{1,2}(Q_{i(i+1)}, \partial Q_{i(i+1)}), \quad \forall i \in \{0, \ldots, N-1\}, \forall s \in \mathbb{R}.
\]

Furthermore, if \( A \) is smooth then \( U^* A \) is smooth as well.

**Proof of Remark 4.4.** We suppress the Sobolev exponents, unless they are relevant. By definition, \( \text{NS}_P(a) \) lies in the complex gauge orbit of \( a \). The key observation to the proof of (60) is that the Yang-Mills heat flow, and hence \( \text{Heat}^\gamma (a) \), always lies in the complexified gauge orbit of the initial condition \( a \). Indeed, in [5] Donaldson shows that for any \( a \in A(P) \) there is some path \( \tilde{c}(\tau) \in \Omega^0(\Sigma, P(g)) \) for which the equation

\[
\frac{d}{d\tau} \tilde{a}(\tau) = -d^*_{\tilde{a}(\tau)} F_{\tilde{a}(\tau)} + d\tilde{a}(\tau) \tilde{c}(\tau), \quad \tilde{a}(0) = a
\]

has a unique solution \( \tau \mapsto \tilde{a}(\tau) \) for all \( 0 \leq \tau < \infty \). The solution has the further property that it takes the form \( \tilde{a}(\tau) = \mu(\tau)^* a \), for some path of complex gauge transformations \( \mu(\tau) \in \mathcal{G}(P)^C \) starting at the identity. It is then immediate that \( a(\tau) := \exp \left( \int_0^\tau 1 \right)^* \tilde{a}(\tau) \) solves (68), and so

\[
\text{Heat}^\gamma (a) = \lim_{\tau \to \infty} \exp \left( \int_0^\tau 1 \right)^* \mu(\tau)^* a.
\]

Clearly \( \exp \left( \int_0^\tau 1 \right)^* \mu(\tau)^* a \) lies in the complex gauge orbit of \( a \) for all \( \tau \), and so \( \text{Heat}^\gamma (a) \) must as well. Now \( \text{NS}_P(a) \) lies in the complex gauge orbit by definition, so there is some complex gauge transformation \( \tilde{\mu} \in \mathcal{G}(P)^C \) (possibly depending on \( a \)) with \( \tilde{\mu}^* \text{NS}(a) = \text{Heat}(a) \). We will be done if we can show that \( \tilde{\mu} \) is a real gauge transformation that lies in the identity component. The former statement is equivalent to showing \( \tilde{h} := \tilde{\mu}^T \tilde{\mu} = \text{Id} \). By (14) we must have that \( \tilde{h} \) is a solution to

\[
\tilde{F}(h) := i \delta_{\text{Heat}(a)} \left( h^{-1} \partial_{\text{Heat}(a)} h \right) = 0.
\]

Clearly the identity gauge transformation is a solution as well. It suffices to show that \( \tilde{F}(h) \) has a unique solution, at least for \( a \) close to the space of flat connections. The map \( \tilde{F} \) is defined on (the \( W^{2,2} \)-completion of) \( \mathcal{G}(P)^C \), we can take its codomain
to be (the $L^2$-completion of) $\Omega^0 (\Sigma, P(g)^C)$. Similarly to our analysis of $F$ in the proof of Theorem 3.6, the derivative of $\tilde{\mathcal{F}}$ at the identity is the map sending $\eta \in W^{2,2} (P(g)^C)$ to $\frac{1}{2} \Delta_{\text{Heat}(\alpha)} \rho \eta \in L^2 (P(g)^C)$. This derivative is invertible, so by the inverse function theorem $\tilde{\mathcal{F}}$ is a diffeomorphism in a neighborhood of the identity. This is the uniqueness we are looking for, provided we can arrange so that $\tilde{h}$ lies in a suitably small neighborhood of the identity. However, this is immediate since the gauge transformation $\tilde{h}$ depends continuously on $\alpha$ in the $W^{1,2}$-topology, and $\tilde{h} = \text{Id}$ if $\alpha$ is flat.

To finish the proof of (60), we need to show that $\tilde{\mu} /G(\mathcal{P})$ is actually in the identity component $G_0 (\mathcal{P})$. However, this is also immediate from the continuous dependence of $\tilde{\mu}$ on $\alpha$. Indeed, pick any path from $\alpha$ to $A_{\text{flat}} (\mathcal{P})$ that never leaves a suitably small neighborhood of $A_{\text{flat}} (\mathcal{P})$. Applying the above construction to the values of this path of connections provides a path of real gauge transformations from $\tilde{\mu}$ to the identity.

### 4.2 Uniform elliptic regularity

In this section we establish several elliptic estimates that will be used in the proof of Lemma 4.2, below. We refer freely to the notation from the introduction to Section 4, and we will write

$$F_A|_{\mathbb{R} \times I \times \Sigma} = F_a + ds \land \beta_s + dt \land \beta_t + ds \land dt \gamma, \quad F_A|_{\mathbb{R} \times Y} = F_a + ds \land b_s$$

for the components of the curvature of a connection $A$. Throughout we assume $A$ is $\epsilon$-ASD and satisfies uniform estimates of the form

$$\sup_{\mathbb{R} \times I} \|\beta_s\|_{L^2(\Sigma)} + \sup_{\mathbb{R}} \|b_s\|_{L^2(Y)} \leq c_0$$

for some fixed constant $c_0$. We also assume the slice-wise curvatures on $\Sigma$ and $Y$ are sufficiently small in $L^\infty$:

$$\sup_{\mathbb{R} \times I \times \Sigma} |F_a| + \sup_{\mathbb{R} \times Y} |F_a| \leq \delta_0.$$  

Here $\delta_0 > 0$ is a constant chosen so that if $\alpha, a$ satisfy (69), then there is some $C$ for which

$$\|\rho\|_{L^p(\Sigma)} \leq C \|d_\alpha \rho\|_{L^2(\Sigma)}, \quad \|r\|_{L^q(Y)} \leq C \|d_a r\|_{L^2(Y)}$$

for all $1 \leq p < \infty, 1 \leq q \leq 6$, and all 0-forms $\rho, r$. That such a $\delta_0 > 0$ exists follows because all flat connections are irreducible, and because irreducibility is a gauge-invariant and open condition. Moreover, the constants $\delta_0, C$ depend only on the bundle and the fixed metric. See Lemma 3.9. Note that (68) and (69) are basically the hypotheses of Lemma 4.2. Intuitively, the conditions in (69) assert that $A$ almost represents a strip in $M$ with Lagrangian boundary conditions, while (68) can be viewed as an analogue of a uniform bound on the energy density for such strips.

The main result of this section is the following.
Theorem 4.8. Fix a constant $c_0 > 0$, and let $\delta_0 > 0$ be as above. Then there are constants $\epsilon_0, C > 0$ so that

$$\begin{align*}
\|\nabla_s \beta_s\|_{L^2(K \times I \times \Sigma_*)} + \|\nabla_t \beta_s\|_{L^2(K \times I \times \Sigma_*)} \\
+ \|\nabla^2_s \beta_s\|_{L^2(K \times I \times \Sigma_*)} + \|\nabla^2_t \beta_s\|_{L^2(K \times I \times \Sigma_*)} \\
+ \|\nabla_s \beta_s\|_{L^2(K \times Y_*)} \\
\leq C(1 + \text{vol}(K) + F_{\epsilon}^{\text{inst}}(A))
\end{align*}$$

for all $0 < \epsilon < \epsilon_0$, all compact $K \subset \mathbb{R}$, and all $\epsilon$-ASD connections $A$ satisfying (68) and (69).

In the next section this theorem will be combined with the embeddings

$$W^{2,2}(K \times I) \hookrightarrow C^0(K \times I), \quad W^{1,2}(K) \hookrightarrow C^0(K)$$

to conclude (56) and (57) in the statement of Lemma 4.2.

It is convenient to use the $\epsilon$-dependent norm

$$\|v\|^2_{L^2(U),\epsilon} := \int_U (v \wedge \ast_{\epsilon} v)$$

for measurable $U \subseteq \mathbb{R} \times Y$, where $v$ is a form, and $\ast_{\epsilon}$ is the Hodge star on $\mathbb{R} \times Y$ induced from the metric $g_\epsilon$. We will also refer to the $\epsilon$-dependent $L^2$-inner product $\langle \cdot, \cdot \rangle_{\epsilon}$ and $L^p$-norms $\| \cdot \|_{L^p(U),\epsilon}$ defined in the obvious way. We drop the $\epsilon$ in the notation when we are working with the $\epsilon$-independent metric $g = g_1$; that is,

$$\| \cdot \|_{L^p(U)} := \| \cdot \|_{L^p(U),1}.$$

(For example, all norms appearing in Theorem 4.8 are with respect to this $\epsilon$-independent metric.) In particular, if $v$ is a map from $\mathbb{R}$ to the space of $k$-forms on $Y_*$, then

$$\|d_s \wedge v\|^p_{L^p(\mathbb{R} \times Y_*),\epsilon} = \|v\|^p_{L^p(\mathbb{R} \times Y_*),\epsilon} = \epsilon^{3-pk} \|v\|^p_{L^p(\mathbb{R} \times Y_*)},$$

and if $v$ is a map from $\mathbb{R} \times I$ to the space of $k$-forms on $\Sigma_*$, then

$$\|d_s \wedge dt \wedge v\|^p_{L^p(\mathbb{R} \times I \times \Sigma_*),\epsilon} = \|v\|^p_{L^p(\mathbb{R} \times I \times \Sigma_*),\epsilon} = \epsilon^{2-pk} \|v\|^p_{L^p(\mathbb{R} \times I \times \Sigma_*)}.$$

We will prove Theorem 4.8 in four steps. The first is Proposition 4.9 and establishes a general estimate that bounds first derivatives on the interval $I$ and surface $\Sigma_*$ in terms of the 3-dimensional operators $d_s$ and $d^s_*$ on $Y$. The second step is Proposition 4.10 and bounds the $s$-derivative $\nabla_s F_A$ for an $\epsilon$-instanton $A$ in terms of the energy $\|F_A\|_{L^2,\epsilon}$. This represents a version of the standard elliptic bootstrapping estimate for instantons, where here we keep track of how the constants depend on $\epsilon$. We have restricted only to the $s$-derivative because this is the only direction in which no $\epsilon$-scaling occurs; in Corollary 4.12 we bound various other first derivatives of the curvature. As a third step we establish a second order version of the first step; this is stated in Proposition 4.13. The last step is Proposition 4.14 and is a certain second order version of the second step, in which we bound the second derivatives $\nabla^2_s F_A$ and $\nabla^s_\ast F_A$. Then Theorem 4.8 is an immediate corollary of these latter two propositions.
**Proposition 4.9.** (General elliptic estimates; 1st order) Suppose $A$ is an $e$-ASD connection satisfying (68) and (69) for some $c_0 > 0$. Then there are constants $c_0, C > 0$ such that the following holds for all $0 < \epsilon < c_0$ and all $e$-smooth 1-forms $v$ on $Y$ with $v|_{I \times \Sigma_\epsilon} = v + \theta dt$:

\[
\begin{align*}
\|d_a v\|_{L^2(I \times \Sigma_\epsilon),\epsilon} &+ \|d_a^\epsilon v\|_{L^2(I \times \Sigma_\epsilon),\epsilon} + \|\nabla_t v\|_{L^2(I \times \Sigma_\epsilon),\epsilon} \\
&+ \|d_a \theta\|_{L^2(I \times \Sigma_\epsilon),\epsilon} + \|\nabla_t \theta\|_{L^2(I \times \Sigma_\epsilon),\epsilon} \\
&\leq C \left( \|v\|_{L^2(Y),\epsilon} + \|d_a v\|_{L^2(Y),\epsilon} + \|d_a^\epsilon v\|_{L^2(Y),\epsilon} \right).
\end{align*}
\]

(70)

The proof will show that the bound (70) continues to hold with the same constants when $I \times \Sigma_\epsilon$ is replaced by any measurable subset in the complement of the critical fibers of the Morse function $f : Y \to S^1$. We also note that the bound in (70) can be integrated over any interval in $\mathbb{R}$ to obtain analogous $L^2$-bounds for 4-manifolds.

**Proof of Proposition 4.9.** To illustrate the basic argument we first suppose we are in the simpler situation in which $f : Y \to S^1$ has no critical points. Then the 1-form $dt$ is globally defined and so we have decompositions

\[
v = v + \theta dt, \quad a = a + \psi dt
\]

(71)

that are defined globally on $Y$. This gives

\[
d_a v = d_a v + dt \wedge (\nabla_t v - d_a \theta), \quad \text{and} \quad d_a^\epsilon v = d_a^\epsilon v - \nabla_t \theta.
\]

(72)

In this simplified situation we are continuing to use the rescaled metric taking the form $dt^2 + e^2 g_\Sigma$ on $I \times \Sigma_\epsilon$ and $e^2 g_Y$ on $Y_\epsilon$ (even though $Y_\epsilon$ is now a product cobordism). Throughout this proof, the variable $t$ will denote the coordinate defined with respect to the $e$-dependent smooth structure; for example, $|dt|_e = 1$ everywhere on $Y$.

In this simplified setting we will prove

\[
\begin{align*}
\|d_a v\|_{L^2(Y),\epsilon} &+ \|d_a^\epsilon v\|_{L^2(Y),\epsilon} + \|\nabla_t v\|_{L^2(Y),\epsilon} + \|d_a \theta\|_{L^2(Y),\epsilon} + \|\nabla_t \theta\|_{L^2(Y),\epsilon} \\
&\leq C \left( \|v\|_{L^2(Y),\epsilon} + \|d_a v\|_{L^2(Y),\epsilon} + \|d_a^\epsilon v\|_{L^2(Y),\epsilon} \right)
\end{align*}
\]

for a constant $C$ that is independent of $\epsilon$. We will then describe how to adjust the proof to accommodate the more general situation in which $f$ has critical points. Here and below all norms, inner products and Hodge stars are over $Y$, unless otherwise specified. One exception is that the notation $d_a^\epsilon$ is the adjoint taken with respect to the inner product on the surface $\Sigma_\epsilon$, whereas $d_a^\epsilon$ is that on the 3-manifold; the rule being that we view $d_a$ as an operator on the surface and $d_a$ as an operator on the 3-manifold.

Take the $L^2$-norm square of each term in (72), and then add to get
\[ \|d_a v\|_2^2 + \|d_a^\epsilon v\|_2^2 = \|d_\alpha v\|_2^2 + \|\nabla_\nu v - d_\alpha \theta\|_2^2 + \|\nabla_\theta - d_\alpha^\epsilon v\|_2^2 \]
\[ = \|d_\alpha v\|_2^2 + \|d_\alpha^\epsilon v\|_2^2 + \|\nabla_\nu v\|_2^2 \]
\[ + \|d_\alpha \theta\|_2^2 + \|\nabla_\theta\|_2^2 \]
\[ - 2 \left( \nabla_\nu \theta, d_\alpha^\epsilon v \right)_\epsilon - 2 \left( \nabla_\nu v, d_\alpha \theta \right)_\epsilon, \]  

where the norms and inner products are on \( Y \). It suffices to bound these last two terms. We integrate by parts in the second of these to get
\[ - 2 \left( \nabla_\nu \theta, d_\alpha^\epsilon v \right)_\epsilon - 2 \left( \nabla_\nu v, d_\alpha \theta \right)_\epsilon = -2 \left( \nabla_\nu \theta, d_\alpha^\epsilon v \right)_\epsilon + 2 \left( v, \nabla_\nu d_\alpha \theta \right)_\epsilon \]
\[ = -2 \left( \nabla_\nu \theta, d_\alpha^\epsilon v \right)_\epsilon + 2 \left( v, \nabla_\nu \theta \right)_\epsilon + 2 \left( v, [\beta_1, \theta] \right)_\epsilon \]
\[ = 2 \left( v, [\beta_1, \theta] \right)_\epsilon, \]

where we canceled the first two terms in the last step after a second integration by parts. To control this, write
\[ (v, [\beta_1, \theta])_\epsilon = \int_{Y_\epsilon} (v \wedge [\ast_\epsilon \beta_1, \theta]) + \int_{I \times \Sigma \epsilon} (v \wedge [\ast_\epsilon \beta_1, \theta]). \]  

We begin by estimating the second term on the right in (75). For this we note that pointwise on \( I \times \Sigma_\epsilon \) we have \( \ast_\epsilon \beta_1 = \ast \beta_1 \), since \( \beta_1 \) is a 1-form and the \( \epsilon \)-scaling is only in the \( \Sigma_\epsilon \)-direction. This shows that \( |\int_{I \times \Sigma_\epsilon} (v \wedge [\ast_\epsilon \beta_1, \theta])| \) is bounded by
\[ \left( \sup_I \| \beta_1 \|_{L^2(I \times \Sigma_\epsilon)} \right) \| v \|_{L^2(I, L^2(I \times \Sigma_\epsilon))} \| \theta \|_{L^2(I, L^2(I \times \Sigma_\epsilon))} \]
\[ \leq c_0 C_1 \left( \| v \|_{L^2(I \times \Sigma_\epsilon)}^2 + \| d_\alpha v \|_{L^2(I \times \Sigma_\epsilon)}^2 + \| d_\alpha^\epsilon v \|_{L^2(I \times \Sigma_\epsilon)}^2 \right) \left( \| d_\alpha \theta \|_{L^2(I \times \Sigma_\epsilon)}^2 \right) \]

where we have used the standard elliptic estimates for the operator \( d_\alpha \oplus d_\alpha^\epsilon \) (with respect to the fixed metric). Converting back to the \( \epsilon \)-dependent norms, we can continue this as follows:
\[ = c_0 C_1 \left( \| v \|_{L^2(I \times \Sigma_\epsilon), \epsilon}^2 + \epsilon \| d_\alpha v \|_{L^2(I \times \Sigma_\epsilon), \epsilon}^2 + \epsilon \| d_\alpha^\epsilon v \|_{L^2(I \times \Sigma_\epsilon), \epsilon}^2 \right) \left( \| d_\alpha \theta \|_{L^2(I \times \Sigma_\epsilon), \epsilon}^2 \right) \]
\[ \leq C_2 \| v \|_{L^2(I \times \Sigma_\epsilon), \epsilon}^2 + \frac{1}{2} \left( \| d_\alpha v \|_{L^2(I \times \Sigma_\epsilon), \epsilon}^2 + \| d_\alpha^\epsilon v \|_{L^2(I \times \Sigma_\epsilon), \epsilon}^2 + \| d_\alpha \theta \|_{L^2(I \times \Sigma_\epsilon), \epsilon}^2 \right), \]

for \( \epsilon \) sufficiently small. This finishes the bound on the \( I \times \Sigma_\epsilon \) part of (75) since the derivative terms appearing here can be absorbed by the analogous terms in (73).

For the first term on the right-hand side of (75) (\( Y_\epsilon \) part), we use the fact that \( \beta_1 \) is the \( dt \)-component of \( F_\alpha \) and we have a bound of the form \( \| F_\alpha \|_{L^2(Y_\epsilon)} \leq c_0 \). This gives
\[ \| \ast_\epsilon \beta_1 \|_{L^2(Y_\epsilon)} = \epsilon \| \beta_1 \|_{L^2(Y_\epsilon)} = \| dt \|_{L^2(Y_\epsilon)} \leq \| F_\alpha \|_{L^2(Y_\epsilon)} \leq c_0, \]
where in the first equality we used $\epsilon \beta_i = \epsilon \beta_i$, and in the second we used $\epsilon = e^{dt} e = |dt|$, which holds due to the $\epsilon$-scaling on $Y_*$. At this point the computation is similar to the one above. We begin by writing

$$
\left| \int_{Y_*} (v \wedge [\epsilon \beta_i, \theta]) \right| \leq \| \epsilon \beta_i \|_{L^2(Y_*)} \| v \|_{L^1(Y_*)} \| \theta \|_{L^2(I)} .
$$

As before we want to use the elliptic estimates for the operator $d_a \oplus d_a^*$, with respect to the fixed metric. We recall that the appropriate $\epsilon$-dependent $'t'$-derivative on $Y_*$ is actually $\epsilon \nabla t$, since $\nabla t$ is the object defined with respect to the $\epsilon$-dependent metric. This allows us to bound $| \int_{Y_*} (v \wedge [\epsilon \beta_i, \theta]) |$ by

$$
eq c_0 C_3 \left( \| v \|_{L^2(Y_*), \epsilon} + \| \epsilon \nabla t v \|_{L^2(Y_*), \epsilon} + \| d_a v \|_{L^2(Y_*), \epsilon} + \| d_a^* v \|_{L^2(Y_*), \epsilon} \right) \times \left( \| \epsilon \nabla t \theta \|_{L^2(Y_*), \epsilon} + \| d_a \theta \|_{L^2(Y_*), \epsilon} \right).
$$

Now we continue this by converting back to the $\epsilon$-dependent metric:

$$
\leq c_0 C_3 \left( e^{-1/2} \| v \|_{L^2(Y_*), \epsilon} + e^{-1/2} \| \epsilon \nabla t v \|_{L^2(Y_*), \epsilon} + e^{1/2} \| d_a v \|_{L^2(Y_*), \epsilon} + e^{1/2} \| d_a^* v \|_{L^2(Y_*), \epsilon} \right) \\
\leq c_0 C_4 \left( e^{1/2} \| \epsilon \nabla t \theta \|_{L^2(Y_*), \epsilon} + e^{1/2} \| d_a \theta \|_{L^2(Y_*), \epsilon} \right) .
$$

When $\epsilon^2$ is small, the derivative terms appearing here can be absorbed into the analogous terms in (73), so this completes the proof in the simplified situation with no critical points.

Now we describe how to adjust the above argument to accommodate the case where $f$ has $N$ critical points $\{ p_i \}$. We recall that $N$ is even and each $Y_{i(i+1)}$ contains a unique critical point $p_i$ with index 1 or 2. Note that the 1-form $dt = df / |df|_{\epsilon}$ is well-defined on the complement $Y \setminus \{ p_i \}$ of the critical points. Consequently, the component functions appearing in (71) are also defined in this region. Write $c_i = f(p_i)$ for the critical value associated to $p_i$. Fix $r > 0$ small, and set

$$
Y_r := f^{-1} \left( S^1 \setminus \bigcup_i B_r(c_i) \right) ,
$$

where $B_r(c_i) \subset S^1$ is the closed interval of radius $r$ around $c_i$. When $r = 0$ we declare $Y_0$ to be the complement in $Y$ of the critical fibers $\bigcup f^{-1}(c_i)$. Note that $Y_0$ has full measure in $Y$.

We will repeat the calculations above with $Y$ replaced by $Y_r$. Due to the integration by parts, this will result in some additional boundary terms, but we will see
that these cancel as \( r \) goes to zero. Explicitly, note that the computation of (73) holds with all norms interpreted as being over \( Y_r \) instead of \( Y \). The new feature occurs in (74) where the integration by parts in \( \nabla_t \) gives the aforementioned boundary terms:

\[
-2 \left( \nabla_t \theta, d_a^\alpha v \right)_{L^2(Y_r),\epsilon} - 2 \left( \nabla_t v, d_a^\theta \right)_{L^2(Y_r),\epsilon}
\]

\[
= 2 \left( v, [\beta_t, \theta] \right)_{L^2(Y_r),\epsilon} - 2 \int_{\partial Y_r} (v, d_a^\theta)_{\epsilon}.
\]

Our above estimates for the first term in the second line remains valid, so we therefore have

\[
\|d_a v\|_{L^2(Y_r),\epsilon}^2 + \|d^\epsilon_d v\|_{L^2(Y_r),\epsilon}^2 + \|\nabla_t v\|_{L^2(Y_r),\epsilon}^2 + \int_{\partial Y_r} (v, d_a^\theta)_{\epsilon}
\]

\[
\leq C \left( \|v\|_{L^2(Y_r),\epsilon}^2 + \|d_a v\|_{L^2(Y_r),\epsilon}^2 + \|d^\epsilon_d v\|_{L^2(Y_r),\epsilon}^2 \right)
\]

Since the right-hand side is independent of \( r \), we will be done if we can show

\[
\lim_{r \to 0^+} \int_{\partial Y_r} (v, d_a^\theta)_{\epsilon} = 0.
\]

This inner product is independent of \( \epsilon \) by the conformal scaling properties of 1-forms on surfaces. The manifold \( Y_r \) is a deformation retract of the cobordism \( I \times \Sigma_\star \), and so can be viewed as a cobordism between surfaces \( S^- \) to \( S^+ \), with \( S^\pm \cong \Sigma_\star \). See Figure 3.

In terms of the notation \( S^\pm \), to prove the proposition it suffices to show

\[
\left( \int_{S^+} (v, d_a^\theta) - \int_{S^-} (v, d_a^\theta) \right) \xrightarrow{r \to 0} 0.
\]

Since \( S^- \) and \( S^+ \) are both converging to the same fiber \( f^{-1}(\cup_i c_i) \) as \( r \) approaches 0, verifying the limit (77) would be straightforward if, for example, we knew the integrands were \( L^\infty \). Unfortunately, in general these integrands are not \( L^\infty \). The issue arises from the fact that the coordinate decomposition \( v = v + \theta dt \) is only valid away from the critical points of \( f \). Though this implies \( v, \theta \) are \( L^\infty \) on all of \( Y \) (e.g., \( |v| \leq |\theta| \)), the derivative \( d_a^\theta \) is typically not \( L^\infty \). However, we will see that it is \( L^1 \), and this will be enough to verify (77). To carry this out, we begin by isolating the problem. For \( \rho > 0 \), let \( N^\pm_{\rho, f} \subset S^\pm \) denote the set of points in \( S^\pm \) that are within a distance of \( \rho \) to a gradient trajectory of the restriction \( f : Y_r \to S^1 \). Then \( N^\pm_{\rho, f} \) is a collection of \( N/2 \) annuli and \( N/2 \) pairs of disks; a pair of disks arise from the unstable (resp. stable) manifold of an index 1 (resp. 2) critical point, and an annulus from the stable (resp. unstable) manifold of an index 1 (resp. 2) critical point. The same holds for \( N^\pm_{\rho, f} \). The neighborhoods \( N^\pm_{\rho, f} \) contain the regions in which \( d_a^\theta \) is poorly behaved. The next claim is the key estimate in controlling \( d_a^\theta \) in this region.

**Claim:** There is a constant \( C \) so that
The manifolds $Y_r S_r^-$ and $S_r^+$ each have $N$ connected components, where $N$ is the number of critical points of the Morse function $f : Y \to S^1$. The case $N = 2$ is illustrated on the left above. The surfaces $S_r^\pm$ are indicated by the solid lines, and these bound the cobordism $Y_r$. The dotted lines indicate where the cobordisms $Y_r$ and $I \times \Sigma$ meet. The figure on the right is a larger illustration of the region around the critical point $p_0$. Taking the standard orientation of the circle, this is a critical point of index 1. The annulus labeled $N_{p,r,0}^{-\rho}$ is the portion of the neighborhood $N_{p,r,0}$ that lies in the component of $S_r^-$ closest to $p_0$. Similarly, the pair of disks labeled $N_{p,r,0}^{+\rho}$ is the portion of $N_{p,r,0}$ in the component of $S_r^+$ closest to $p_0$. Though this is not illustrated, the intersection of $S_r^-$ with the stable manifold of $p_0$ is a loop lying in the middle of the annulus $N_{p,r,0}^{-\rho}$. Similarly, the unstable manifold of $p_0$ determines a pair of points in $S_r^+$, and these are the centers of the disks $N_{p,r,0}^{+\rho}$.

$$\int_{N_{p,r,0}^\pm} |(v, d_\alpha \theta)| \leq C(\rho + r)$$
for all $\rho > 0$, $r > 0$ sufficiently small. The constant depends on $f$, $v$, and the fixed metric.

Before proving the claim, we will show how it used to finish the proof of (77). Let $S_0 := f^{-1}(\cup_i \mathcal{C}_i)$ denote the critical fibers of $f$. Then the normalized gradient flow of $\mp f$ provides embeddings

$$\varphi_r^\pm : S_0 \setminus \{p_j\} \to S_r^\pm.$$ 

The image of $\varphi_r^\pm$ is the complement in $S_r^\pm$ of the stable and unstable manifolds of $f|_{Y \setminus Y}$. The family $\{\varphi_r^\pm\}$ varies continuously in $r$ and approaches the inclusion

$$\varphi_0 : S_0 \setminus \{p_j\} \hookrightarrow S_0$$

as $r$ approaches 0; the same holds for $\{\varphi_r^-\}$. Each function $(v, d_\alpha \theta)|_{S_r^\pm}$ is well-defined and smooth on the complement of the critical points $p_j$. These observations imply that the family of functions

$$(\varphi_r^+)^* \left((v, d_\alpha \theta)|_{S_r^+}\right) - (\varphi_r^-)^* \left((v, d_\alpha \theta)|_{S_r^-}\right) : S_0 \setminus \{p_j\} \to \mathbb{R}$$

52
is continuous in $r$ and converges pointwise, as $r$ approaches zero, to the zero function on $S_0 \setminus \{ p_i \}$. On any compact $K \subset S_0 \setminus \{ p_i \}$ this convergence is therefore uniform, and so we have

$$\int_{\varphi_{\rho}^+ (K)} (v, d_\alpha \theta) - \int_{\varphi_{\rho}^- (K)} (v, d_\alpha \theta) = \int_K (\varphi_{\rho}^+)^* \left( (v, d_\alpha \theta) \big|_{S_r^+} \right) - (\varphi_{\rho}^-)^* \left( (v, d_\alpha \theta) \big|_{S_r^-} \right) \to 0. \quad (78)$$

Combining this observation with the claim, we obtain the desired convergence in (77): Fix $\delta > 0$. By the claim we can ensure

$$\int_{\mathbb{R}_{\rho, r}} |(v, d_\alpha \theta)| \leq \delta/3$$

for all $\rho, r > 0$ sufficiently small, which we assume is the case; we will refine the choice of $r$ momentarily. Notice that the functions $\varphi_{\rho}^\pm$ only drastically change the metric near the stable/unstable manifold. In particular, there is some compact $K \subset S_0 \setminus \{ p_i \}$ so that $S_r^\pm \cap \varphi_{\rho}^\pm(K) \subset \mathbb{R}_{\rho, r}$ for all $r > 0$. By (78), we have

$$\left| \int_{\varphi_{\rho}^+ (K)} (v, d_\alpha \theta) - \int_{\varphi_{\rho}^- (K)} (v, d_\alpha \theta) \right| \leq \delta/3,$$

for any sufficiently small $r > 0$. Putting this all together gives

$$\left| \int_{S_r^+} (v, d_\alpha \theta) - \int_{S_r^-} (v, d_\alpha \theta) \right| \leq \int_{S_r^+ \setminus \varphi_{\rho}^+ (K)} |(v, d_\alpha \theta)| + \int_{S_r^- \setminus \varphi_{\rho}^- (K)} |(v, d_\alpha \theta)| + \int_{\varphi_{\rho}^+ (K)} (v, d_\alpha \theta) - \int_{\varphi_{\rho}^- (K)} (v, d_\alpha \theta) \leq \int_{\mathbb{R}_{\rho, r}} |(v, d_\alpha \theta)| + \int_{\mathbb{R}_{\rho, r}} |(v, d_\alpha \theta)| + \delta/3 \leq \delta.$$

Since this holds for all $r$ sufficiently small, this completes the proof of (77).

To finish the proof of the proposition it therefore suffices to prove the claim. Fix a critical point $p_i$ and denote by $N_{\rho, r, j}^\pm$, the component of $N_{\rho, r}^\pm$ that is closest to $p_i$; see Figure 3. With this notation, we then have

$$\int_{N_{\rho, r}^\pm} |(v, d_\alpha \theta)| \leq \|v\|_{L^\infty(Y)} \sum_i \|d_\alpha \theta\|_{L^1(N_{\rho, r, j}^\pm)}.$$

Since $\|v\|_{L^\infty(Y)}$ is independent of $\rho, r$, it suffices to bound each $\|d_\alpha \theta\|_{L^1(N_{\rho, r, j}^\pm)}$. We will do this by expressing $d_\alpha \theta$ in terms of smooth coordinates on $Y$, and then estimating to show that the coefficients in these coordinates are $L^1$. Fix a critical point $p_i$. Without loss of generality, we may assume $p_i$ has index 1 with respect to the positive
orientation of the circle. Then the Morse lemma allows us to identify a neighborhood of \( p_i \) with the set

\[
\left\{ (x_1, x_2, x_3, \zeta) \in U \times B_r(0) \subset \mathbb{R}^3 \times \mathbb{R} \mid -x_1^2 + x_2^2 + x_3^2 = \zeta \right\}
\]

(79)

for some open neighborhood \( U \) of the origin in \( \mathbb{R}^3 \). To simplify the discussion we assume the metric on \( Y \) near \( p_i \) agrees with the metric on (79) induced from the standard metric on \( \mathbb{R}^3 \times \mathbb{R} \); the more general situation can be reduced to this by noting that (i) two different metrics on \( Y \) induce equivalent \( L^1 \)-norms on \( N^+_{\rho, r, i} \) and (ii) the constants defining this equivalence can be chosen to depend only on the metrics (that is, the constants do not depend \( \rho, r \) since the neighborhoods \( N^ \pm_{\rho, r, i} \) are all contained in a compact region).

In terms of the Morse coordinates (79), we have the following:

- \( p_i \) is identified with \((0,0,0,0)\),
- the function \( f \) is the projection \((x_1, x_2, x_3, \zeta) \mapsto \zeta + f(p_i)\),
- the unstable manifold is the set of points \((x_1, 0, 0, -x_1^2)\), and
- the stable manifold is the set of points \((0, x_2, x_3, x_2^2 + x_3^2)\).

By choosing \( U \) appropriately, we can arrange so that \( N^ \pm_{\rho, r, i} \) is identified, under these coordinates, with the fiber \( \zeta = \mp r \). That is, \( N^+_{\rho, r, i} \) is the set of points \((x_1, x_2, x_3, -r)\) in (79) that are within a distance of \( \rho \) to the unstable manifold, and \( N^-_{\rho, r, i} \) is the set of points \((x_1, x_2, x_3, r)\) in (79) that are within a distance of \( \rho \) to the stable manifold. See Figure 4.

The restriction of the principal bundle \( Q \) to the coordinate patch (79) is trivializable, and we let \( d \) denote the trivial connection coming from a fixed trivialization. Then \( d\alpha \) and \( d\theta \) differ by the \( L^\infty \) form \([\alpha, \theta]\), so to prove the claim it suffices to bound the \( L^1 \)-norm of \( d\theta \). The coordinates \( x_1, x_2, x_3 \) are smooth functions near (and at) the critical points of \( f \). In particular, writing

\[
v = w_1 dx_1 + w_2 dx_2 + w_3 dx_3
\]

it follows that the \( w_i \) are smooth because \( v \) is smooth. We also have

\[
dt = \frac{1}{|df|} df = \frac{2}{|df|} (x_1 dx_1 - x_2 dx_2 - x_3 dx_3)
\]

and so comparing with the coordinates \( v = v + \theta dt \), we find \( \theta = \frac{1}{2x_1} |df| w_1 \). This gives

\[
d\theta = \frac{dx_2}{2x_1} ((\partial_2 |df|) w_1 + |df| \partial_2 w_1) + \frac{dx_3}{2x_1} ((\partial_3 |df|) w_1 + |df| \partial_3 w_1).
\]

Taking the norm, we obtain

\[
|d\theta| \leq \frac{C}{|x_1|} = \frac{C}{\sqrt{-\zeta + x_2^2 + x_3^2}},
\]

(80)
Figure 4: The figure on the left illustrates the pair of disks $N^+_{\rho,r,i}$ with center $\bullet$ and radius $\rho$. The vertical ($x_1$)-axis is the unstable manifold of $p_i$, and the intersection of this with $N^+_{\rho,r,i}$ consists of the two centers $\bullet$. The projection of $N^+_{\rho,r,i}$ to the $x_2x_3$-plane is illustrated as the circle of radius $R^+_{\rho,r}$ with center $\ast$, the origin. The figure on the right illustrates the annulus $N^-_{\rho,r,i}$ and its projection to the $x_2x_3$-plane. This projection is another annulus with inner radius $\sqrt{r}$ and outer radius $R^-_{\rho,r}$. The stable manifold of $p_i$ is the $x_2x_3$-plane and $\rho$ is the distance in $N^+_{\rho,r,i}$ from the stable manifold to the boundary of $N^+_{\rho,r,i}$.

where $C$ depends only on $v$ and $f$.

First we analyze the integral of $|d\theta|$ over $N^+_{\rho,r,i}$. Recall $N^+_{\rho,r,i}$ is a pair of disks corresponding to the fiber $\zeta = -r$. Projecting $N^+_{\rho,r,i}$ to the $x_2x_3$-plane is 2-1 with image $\{(x_2, x_3) \mid x_2^2 + x_3^2 \leq (R^+_{\rho,r})^2\}$, a disk with some radius $R^+_{\rho,r} > 0$. It is easy to check that $R^+_{\rho,r} \leq \rho$. Now integrating (80) over both disks in $N^+_{\rho,r,i}$ we get

$$\|d\theta\|_{L^1(N^+_{\rho,r,i})} \leq 4\pi C \left( \sqrt{r + (R^+_{\rho,r})^2} - \sqrt{r} \right) \leq 4\pi C \left( \sqrt{r + \rho^2} - \sqrt{r} \right),$$

which is the desired estimate for the region $N^+_{\rho,r,i}$.

Now we move on to $N^-_{\rho,r,i}$, which is an annulus corresponding to the fiber $\zeta = r$. Projecting this annulus to the $x_2x_3$-plane is a map that is 2-1 off of the intersection of this annulus with this plane. The image is the annulus

$$\{(x_2, x_3) \mid r \leq x_2^2 + x_3^2 \leq (R^-_{\rho,r})^2\}$$

for some radius $\sqrt{r} < R^-_{\rho,r} \leq \rho + \sqrt{r}$. The claim then follows by integrating (80) over $N^-_{\rho,r,i}$:

$$\|d\theta\|_{L^1(N^-_{\rho,r,i})} \leq 4\pi C \sqrt{-r + (R^-_{\rho,r})^2} \leq 4\pi C \sqrt{\rho^2 + 2\sqrt{r}\rho}.$$
Now we begin to estimate the derivatives of $F_A$ for an $\varepsilon$-ASD connection.

**Proposition 4.10.** (Instanton bootstrapping estimate; 1st order) Fix $c_0 > 0$, open $\Omega \subset \mathbb{R}$ and compact $K \subset \Omega$. Then there are $\varepsilon_0, C > 0$ so that

$$\|\nabla_s F_A\|_{L^2(K \times Y), \varepsilon} \leq C(1 + \text{vol}(K) + \|F_A\|_{L^2(\Omega \times Y), \varepsilon})$$

for all $0 < \varepsilon < \varepsilon_0$ and all $\varepsilon$-ASD connections $A$ satisfying (68) and (69).

**Remark 4.11.** The proof will show that constants $\varepsilon_0, C$ can be chosen to depend on $K, \Omega$ only through the value $1/\text{dist}(\partial K, \partial \Omega)$. In particular, if $\Omega = \mathbb{R}$, then they can be taken to be independent of $K$. Similarly, by taking $K = [n, n + 1], \Omega = (n - 1/2, n + 3/2)$, and then summing over $n$ we get

$$\|\nabla_s F_A\|_{L^2(\mathbb{R} \times Y), \varepsilon} \leq 2C(2 + \|F_A\|_{L^2(\mathbb{R} \times Y), \varepsilon})$$

Before proving Proposition 4.10, we note that combining Propositions 4.9 and 4.10 we obtain uniform bounds for various other first derivatives of the curvature.

**Corollary 4.12.** Under the assumptions of Proposition 4.10, the following are bounded by $C(1 + \|F_A\|_{L^2(\mathbb{R} \times Y), \varepsilon})$:

$$\begin{align*}
\|d_{\mathcal{A}} s b_s\|_{L^2(\mathbb{R} \times I \times \Sigma_s), \varepsilon} + \|d_{\mathcal{A}}^\alpha s b_s\|_{L^2(\mathbb{R} \times I \times \Sigma_s), \varepsilon} + \|\nabla_i s b_s\|_{L^2(\mathbb{R} \times I \times \Sigma_s), \varepsilon} \\
\|d_{\mathcal{A}} s \gamma\|_{L^2(\mathbb{R} \times I \times \Sigma_s), \varepsilon} + \|d_{\mathcal{A}}^\alpha s \gamma\|_{L^2(\mathbb{R} \times I \times \Sigma_s), \varepsilon} + \|\nabla_i s \gamma\|_{L^2(\mathbb{R} \times I \times \Sigma_s), \varepsilon}
\end{align*}$$

(81)

The constant $C$ depends only on $c_0$. If $K \subset \mathbb{R}$ is any compact set, then

$$\|b_s\|_{L^1(K \times Y_s)} \leq C(1 + \text{vol}(K) + \|F_A\|_{L^2(\mathbb{R} \times Y), \varepsilon})$$

**Proof of Corollary 4.12.** For the items in the first and second rows of (81), apply Proposition 4.9 with $\nu = b_s$, and use the identities

$$d_{\mathcal{A}} b_s = \nabla_s F_{\mathcal{A}}, \quad d_{\mathcal{A}}^\alpha b_s = 0,$$

together with $\nabla_s F_{\mathcal{A}} = \nabla_s F_A + ds \wedge \nabla_s F_A$.

We can estimate the items in the third row as follows:

$$\|b_s\|_{L^4(\mathbb{R} \times I \times \Sigma_s)} + \|b_s \gamma\|_{L^2(\mathbb{R} \times I \times \Sigma_s)} \leq c_0 \left(\|b_s\|_{L^2(L^4)} + \|\gamma\|_{L^2(L^4)}\right),$$

where we have set $L^2(L^4) := L^2(\mathbb{R} \times I, L^4(\Sigma_s))$ and used (68). Using the embedding $W^{1,2} \hookrightarrow L^4$ on $\Sigma_s$, we can bound this by

$$c_0 C \left(\|b_s\|_{L^2} + \|d_{\mathcal{A}} b_s\|_{L^2} + \|d_{\mathcal{A}}^\alpha b_s\|_{L^2} + \|d_{\mathcal{A}} \gamma\|_{L^2}\right)$$

$$= c_0 C \left(\|b_s\|_{L^2_{\varepsilon}} + C_{\varepsilon} \|d_{\mathcal{A}} b_s\|_{L^2_{\varepsilon}} + C_{\varepsilon} \|d_{\mathcal{A}}^\alpha b_s\|_{L^2_{\varepsilon}} + C_{\varepsilon} \|d_{\mathcal{A}} \gamma\|_{L^2_{\varepsilon}}\right)$$

where $L^2_{\varepsilon} := L^2(\mathbb{R} \times I \times \Sigma_s)$. We have already bounded these in the first and second row, so the result for the third row of (81) follows.

56
It remains to bound \( \|b_\delta\|_{L^4(K \times \Omega)} \). For this, write
\[
\|b_\delta\|_{L^4(K \times \Omega)} \leq c_0 \|b_\delta\|_{L^2(L^4)},
\]
where now \( L^2(L^4) := L^2(K, L^4(\Omega)) \) and we have used (68) again. This is bounded by
\[
c_0 C \left( \|b_\delta\|_{L^2(K \times \Omega)} + \|d_\delta b_\delta\|_{L^2(K \times \Omega)} \right) 
\leq c_0 C \left( \|b_\delta\|_{L^2(K \times \Omega)} + \epsilon^{1/2} \|\nabla_s F_A\|_{L^2(\Omega \times Y, \epsilon)} \right).
\]
Using (68) one last time, we have \( \|b_\delta\|_{L^2(K \times Y)} \leq \text{vol}(K)c_0 \) is bounded. Finally, use Proposition 4.10 to control \( \epsilon^{1/2} \|\nabla_s F_A\|_{L^2, \epsilon} \).

**Proof of Proposition 4.10.** In what follows, all unspecified integrals and Hodge stars are over \( \mathbb{R} \times Y \); for example, \( L^2 = L^2(\mathbb{R} \times Y) \). Observe that
\[
\nabla_s F_A = \partial_s F_A + [p, F_A] = d_A \partial_s A - d_A (d_A p) = d_A (\partial_s A - d_A p).
\]
and
\[
\partial_s A - d_A p = \partial_s a - d_a p + (\partial_s p - \nabla_s p) \, ds = \partial_s a - d_a p = b_\delta.
\]
In particular, we get
\[
\nabla_s F_A = d_A b_\delta,
\]
where \( d_A \) is the derivative on the 4-manifold \( \mathbb{R} \times Y \), and in this formula we are viewing \( b_\delta \) as a 1-form on this 4-manifold.

Let \( h : \mathbb{R} \rightarrow [0, 1] \) be a compactly supported bump function for \( K \subset \Omega \). Then there is a constant \( C_h \) so that \( |\partial_s h| \leq C_h \); the dependence on \( K, \Omega, h \) of all constants will only be through the values \( C_h \) and \( \text{vol}(\text{supp}(h)) \). Then
\[
\|h \nabla_s F_A\|_{L^2, \epsilon}^2 = \int h^2 \langle \nabla_s F_A \wedge *_\epsilon \nabla_s F_A \rangle = \int h^2 \langle d_A b_\delta \wedge *_\epsilon \nabla_s F_A \rangle.
\]
By Stokes’ theorem, we obtain
\[
\|h \nabla_s F_A\|_{L^2, \epsilon}^2 = \int h^2 \langle d_A b_\delta \wedge *_\epsilon \nabla_s F_A \rangle 
= -\int 2h \partial_s h \, ds \wedge \langle b_\delta \wedge *_\epsilon \nabla_s F_A \rangle + \int h^2 \langle b_\delta \wedge d_A \nabla_s *\epsilon F_A \rangle 
= -\int 2h \partial_s h \, ds \wedge \langle b_\delta \wedge *_\epsilon \nabla_s F_A \rangle - \int h^2 \langle b_\delta \wedge [b_\delta \wedge *\epsilon F_A] \rangle,
\]
where, in the last step, we used \( \nabla_s d_A = d_A \nabla_s + [b_\delta, \cdot] \) and the \( \epsilon \)-ASD condition.

Next, use the inequality
\[
2ab \leq \delta^{-1} a^2 + \delta b^2, \quad \delta > 0
\]
with \( \delta = 5 \) to get
\[ \| h \nabla_b F_A \|^2_{L^2, c} \leq 5C(h) \| b_s \|^2_{L^2, c} + \frac{1}{2} \| h \nabla_b F_A \|^2_{L^2, c} - \int h^2 \langle b_s \wedge [b_s \wedge *_c F_A] \rangle. \]

Subtract the term \( \frac{1}{2} \| h \nabla_b F_A \|^2_{L^2, c} \) from both sides to get
\[ \frac{4}{5} \| h \nabla_b F_A \|^2_{L^2, c} \leq 5C(h) \| b_s \|^2_{L^2, c} - \int h^2 \langle b_s \wedge [b_s \wedge *_c F_A] \rangle. \] (84)

It suffices to bound the second term on the right. For this, we note that in terms of the Hodge star * \( _c \) on \( Y \) we have
\[ *_c F_A = ds \wedge *_c F_a + *_c b_s, \]
so this second term is just
\[ \int h^2 ds \wedge \langle b_s \wedge [b_s \wedge *_c F_a] \rangle = \int_{R \times Y_*} h^2 ds \wedge \langle b_s \wedge [b_s \wedge *_c F_a] \rangle \]
\[ + \int_{R \times I \times \Sigma} h^2 ds \wedge \langle b_s \wedge [b_s \wedge *_c F_a] \rangle. \] (85)

We will be done if we can satisfactorily estimate (85); we begin by estimating the first integral on the right. Note that on \( Y_* \) we have \( *_c F_a = e^{-1} * Y F_a \), so by (68) the integral \( \int_{R \times Y_*} h^2 ds \wedge \langle b_s \wedge [b_s \wedge *_c F_a] \rangle \) is controlled by
\[ c_0 \| h b_s \|^2_{L^2(R, L^4(Y_*))} \leq c_0 C_1 \left( \| h b_s \|^2_{L^2(R, L^4(Y_*))} + \| h a b_s \|^2_{L^2(R, L^4(Y_*))} \right), \]
where we have used the Sobolev embedding \( W^{1, 2}(Y_*) \hookrightarrow L^4(Y_*) \) and the e-ASD condition \( d_a^c b_s = e^2 d_a^c b = 0 \) on \( Y_* \). Using (68) again we can bound \( \| h b_s \|^2_{L^2(R, L^4(Y_*))} \) by \( c_0^2 \) times the volume of the support of \( h \). To control \( \| h a b_s \|^2_{L^2(R, L^4(Y_*))} \), we convert back to the e-dependent norm to write
\[ \| h a b_s \|^2_{L^2(R, L^4(Y_*))} = e \| h \nabla_b F_a \|^2_{L^2(R, L^4(Y_*))} \leq e \| h \nabla_b F_A \|^2_{L^2, c}. \]
Taking \( e < 1/5 \), this can be absorbed by the left-hand side of (84).

It remains to estimate the second integral in (85). Expanding \( b_s \) and \( F_a \) into components on \( I \times \Sigma_* \), this becomes
\[ \int_{R \times I \times \Sigma_*} h^2 ds \wedge dt \wedge \langle b_s \wedge [b_s \wedge *_c F_a] \rangle + 2 \int_{R \times I \times \Sigma_*} h^2 ds \wedge dt \wedge \langle \gamma \wedge [b_s \wedge *_c \beta_t] \rangle. \]
Using (68), this is controlled by
\[ c_0 C_2 \| h b_s \|^2_{L^2(R, L^1, L^4(\Sigma_*))} \left( \| h e^{-2} F_a \|^2_{L^2(R, L^1, L^4(\Sigma_*))} + \| h \gamma \|^2_{L^2(R, L^1, L^4(\Sigma_*))} \right) \]
\[ \leq C_3 \left( \| h b_s \|^2_{L^2} + \| h a b_s \|^2_{L^2} + \| h a^c b_s \|^2_{L^2} \left( \| h e^{-2} d_a^c F_a \|^2_{L^2} + \| h a^c \gamma \|^2_{L^2} \right), \]

58
where the $L^2$-norms are over $\mathbb{R} \times I \times \Sigma_*$. Using (83) and converting back to the $\epsilon$-dependent norms, we can bound this by
\[
C_4\delta^{-1}\left(\|h\beta_s\|_{L^2}\epsilon^2 + \|hd^*_s\beta_s\|_{L^2}\epsilon^2 + \|hd^*_s\beta_s\|_{L^2}\epsilon^2\right)
+ \delta\|hd^*_sA\|_{L^2}\epsilon^2 + \delta\|hd^*_s\gamma\|_{L^2}\epsilon^2.
\]
By Proposition 4.9 applied to $v = b_s$ and $v = \ast F_A$, the last two terms are bounded by
\[
\delta C \left(\|F_A\|_{L^2}\epsilon^2 + \|h\nabla_s F_A\|_{L^2}\epsilon^2\right).
\]
By taking $\delta$ so that $\delta C < 1/5$, the derivative term can be absorbed into the left-hand side of (86), as usual, the non-derivative term is fine. Having fixed $\delta$, we focus now on the remaining derivative terms $C_4\delta^{-1}\epsilon^2\left(\|hd^*_s\beta_s\|_{L^2}\epsilon^2 + \|hd^*_s\beta_s\|_{L^2}\epsilon^2\right)$ in (86). By Proposition 4.9 these terms are controlled by
\[
CC_4\delta^{-1}\epsilon^2 \left(\|F_A\|_{L^2}\epsilon^2 + \|h\nabla_s F_A\|_{L^2}\epsilon^2\right).
\]
Now take $\epsilon$ small enough so $CC_4\delta^{-1}\epsilon^2 < 1/5$. Then the $\nabla_s F_A$ term can be subtracted and absorbed into the left-hand side of (84).

The following is a second order version of Proposition 4.9. To simplify the discussion we state it for the special case $v = b_s$ and $v = \ast F_A$. The point is that we can bound various second order derivatives by the norms of $\nabla_s^2 F_A$, $\nabla_t \nabla_s F_A$, plus lower order terms. We have singled out the derivatives $\nabla_s^2 F_A$, $\nabla_t \nabla_s F_A$ because these derivatives scale favorably in $\epsilon$; this will be estimated in Proposition 4.14 below.

**Proposition 4.13.** (General elliptic estimates; 2nd order) Fix $c_0 > 0$, open $\Omega \subset \mathbb{R}$ and compact $K \subset \Omega$. There are $c_0, C > 0$ so that the following holds for all $0 < \epsilon < c_0$ and all $\epsilon$-ASD connections $A$ satisfying (68) and (69):
\[
\|\nabla_s d_A \beta_s\|_{L^2(K \times \Sigma_\ast),\epsilon}^2 + \|\nabla_s d^*_A \beta_s\|_{L^2(K \times \Sigma_\ast),\epsilon}^2
+ \|\nabla_s \nabla_t d_A \beta_s\|_{L^2(K \times \Sigma_\ast),\epsilon}^2 + \|\nabla_s \nabla_t d^*_A \beta_s\|_{L^2(K \times \Sigma_\ast),\epsilon}^2
+ \|\nabla_t \nabla_s \nabla_t \gamma\|_{L^2(K \times \Sigma_\ast),\epsilon}^2 + \|\nabla_t \nabla_s \nabla_t \gamma\|_{L^2(K \times \Sigma_\ast),\epsilon}^2
+ \|\nabla_t \nabla_s \nabla_t \gamma\|_{L^2(K \times \Sigma_\ast),\epsilon}^2 + \|\nabla_t \nabla_s \nabla_t \gamma\|_{L^2(K \times \Sigma_\ast),\epsilon}^2
\leq C \left(1 + \text{vol}(K) + \|F_A\|_{L^2(\Omega \times Y),\epsilon}
+ \|\nabla_t^2 F_A\|_{L^2(\Omega \times Y),\epsilon} + \text{sup}_{r > 0} \|\nabla_t \nabla_s F_A\|_{L^2(\Omega \times Y),\epsilon}\right),
\]
where $Y_t$ is as in (76). The same result holds with $\beta_t$ (resp. $\ast F_A$) in place of $\beta_s$ (resp. $\gamma$) on the left.

As with Proposition 4.9, the proof will show that the bound (87) continues to hold with $I \times \Sigma_\ast$ replaced by any subset of $Y_t$. Also, the constants only depend on the choice of $K, \Omega$ through the distance from $\partial K$ to $\partial \Omega$; see Remark 4.11.
Proof of Proposition 4.13. The only reason we restrict to compact \( K \) (rather than all of \( \mathbb{R} \)) is so we can appeal to Proposition 4.10. To simplify notation, we will ignore this and work with \( K = \Omega = \mathbb{R} \), with the understanding that the ‘true’ computation would involve a bump function with compact support in \( \Omega \); the extension to this ‘true’ case is no more complicated than the situation appearing in Proposition 4.10.

It therefore suffices to bound the left-hand side of (87) by a constant times

\[
1 + \| F_A \|_{L^2(\mathbb{R} \times \mathbb{R})} + \| \nabla s d\bar{s} \|_{L^2(\mathbb{R} \times \mathbb{R})}^2 + \sup_{r > 0} \| \nabla_i d\bar{s} \|_{L^2(\mathbb{R} \times \mathbb{R})}^2.
\]

Let \( r > 0 \) be small. Then on \( Y_r \) we have

\[
d\bar{s} = d\bar{s} + dt \wedge (\nabla_t \bar{s} - d\alpha \gamma), \quad 0 = d\bar{s} = d\alpha \beta - \nabla_t \gamma.
\]

Applying \( \nabla_s \) and then \( \nabla_t \) to both equations, take the norm square and add everything to get

\[
\| \nabla_s d\bar{s} \|_{L^2(\mathbb{R} \times \mathbb{R})}^2 + \| \nabla_i d\bar{s} \|_{L^2(\mathbb{R} \times \mathbb{R})}^2 \quad \text{is greater than or equal to}
\]

\[
\| \nabla_s d\bar{s} \|_{L^2(\mathbb{R} \times \mathbb{R})}^2 + \| \nabla_i d\bar{s} \|_{L^2(\mathbb{R} \times \mathbb{R})}^2 + \| \nabla_i \nabla_t \gamma \|_{L^2(\mathbb{R} \times \mathbb{R})}^2
\]

\[=
2(\nabla_s \nabla_t \gamma) L^2(\mathbb{R} \times \mathbb{R}),
\]

It suffices to show that we can control the four cross terms at the end. We will work this out explicitly for the terms

\[
-2(\nabla_s \nabla_t \gamma) L^2(\mathbb{R} \times \mathbb{R}), \quad \text{the analysis for the remaining terms (i.e., the last two terms in (88)) is similar.}
\]

As in the proof of Proposition 4.9, the idea is to integrate by parts. These will then cancel, up to some boundary terms coming from \( \partial Y_r \) plus lower-order terms coming from the commutation relations

\[
\nabla_s \nabla_t \gamma = \gamma, \quad \nabla_s d\alpha - d\alpha \nabla_s = \beta s, \quad \nabla_i d\alpha - d\alpha \nabla_t = \beta t.
\]

Explicitly, we integrate by parts in \( \nabla_t \) and then in \( d\alpha \) to get that (89) is equal to a linear combination of the boundary term

\[
\int_{\mathbb{R} \times \partial Y_r} (\nabla_s \beta s, \nabla_s d\alpha \gamma)
\]

together with the following lower order cross terms

\[
(\nabla_s \beta s, \nabla_s \gamma) L^2(\mathbb{R} \times \mathbb{R}), \quad (\nabla_s \beta s, \nabla_t \gamma) L^2(\mathbb{R} \times \mathbb{R}), \quad (\nabla_s \beta s + [\gamma, d\alpha \gamma]) L^2(\mathbb{R} \times \mathbb{R}), \quad (\nabla_s \beta s, [\beta s, \nabla_t \gamma]) L^2(\mathbb{R} \times \mathbb{R}), \quad (* \sum \beta \wedge \beta \wedge \beta \wedge \beta) L^2(\mathbb{R} \times \mathbb{R})
\]

(95)
The star that appears in (95) is the Hodge star on surfaces. As in the proof of Proposition 4.9, the boundary term (90) goes to zero as \( r \) decreases to zero. It therefore suffices to show that the lower order terms (91-95) are suitably bounded with constants independent of \( r \) and \( \epsilon \).

- (91): This is bounded by

\[
\delta \| \nabla_s d_a \gamma \|^2_{L^2(\mathbb{R} \times Y_r),\epsilon} + \delta^{-1} C \| [\gamma, \beta_s] \|^2_{L^2(\mathbb{R} \times Y_r),\epsilon}.
\]

The first term on the right is good for \( \delta \) small since it can be absorbed by analogous terms. To bound the second term first notice that the portion of the integral over \( I \times \Sigma_r \) is controlled by Corollary 4.12. The portion over the complementary region \( Y_r \cap Y_\star \) can be bounded by the following similar argument: By the scaling properties of the Hodge star on 3-manifolds, we have

\[
\| [\gamma, \beta_s] \|^2_{L^2(\mathbb{R} \times Y_r \cap Y_\star),\epsilon} = \epsilon \| [\gamma, \beta_s] \|^2_{L^2(\mathbb{R} \times Y_r \cap Y_\star),\epsilon} \leq C \| \gamma \|^2_{L^2(\mathbb{R} \times L^4(\Sigma_r \cap Y_\star))},
\]

where we used that \( \beta_s \) is a component of \( b_s \), together with (68). The embedding \( W^{1,2} \hookrightarrow L^4 \) on \( Y_r \cap Y_\star \), together with the fact that \( a \) is irreducible implies that there is a bound of the form

\[
\| \gamma \|^2_{L^2(\mathbb{R} \times L^4(\Sigma_r \cap Y_\star))} \leq C \| d_a \gamma \|^2_{L^2(\mathbb{R} \times Y_r \cap Y_\star)}
\]

\[
= C \left( \epsilon^2 \| \nabla \gamma \|^2_{L^2(\mathbb{R} \times Y_r \cap Y_\star)} + \| d_a \gamma \|^2_{L^2(\mathbb{R} \times Y_r \cap Y_\star)} \right)
\]

\[
= \epsilon^{-1} C \left( \| \nabla \gamma \|^2_{L^2(\mathbb{R} \times Y_r \cap Y_\star),\epsilon} + \| d_a \gamma \|^2_{L^2(\mathbb{R} \times Y_r \cap Y_\star),\epsilon} \right),
\]

where in the second line we used \( |dt| = \epsilon \) on the \( \nabla \gamma \)-term. Apply Proposition 4.9 with \( \nu = b_s \) to bound \( \nabla \gamma \) and \( d_a \gamma \) in terms of \( d_a b_s \) and \( d_a^* b_s \), which gives

\[
\| [\gamma, \beta_s] \|^2_{L^2(\mathbb{R} \times Y_r),\epsilon} \leq c_0 C \left( \| b_s \|^2_{L^2(\mathbb{R} \times Y_r)} + \| d_a b_s \|^2_{L^2(\mathbb{R} \times Y_r)} \right),
\]

which, by Proposition 4.10, is bounded by a constant times \( \| F_a \|^2_{L^2(\mathbb{R} \times Y_r),\epsilon} \).

- (92): Integrate by parts in \( \nabla_s \). Then this is bounded by

\[
\delta \| \nabla_s^2 \beta_s \|^2_{L^2(\mathbb{R} \times Y_r),\epsilon} + \delta^{-1} \| [\beta_s, \gamma] \|^2_{L^2(\mathbb{R} \times Y_r),\epsilon}.
\]

The first of these can be absorbed for small \( \delta \). The second is controlled as in (91), here note that \( \beta_s \) is a component of \( F_a \), and in particular satisfies

\[
\sup_{\mathbb{R}} \| \beta_s \|_{L^2(Y_r \cap Y_\star)} \leq \epsilon^{-1} \sup_{\mathbb{R}} \| F_a \|_{L^2(Y_\star)} \leq c_0.
\]

- (93): Integrate by parts in \( \nabla_s \) to get that (93) is equal to

\[
- (\beta_{sy} [\nabla_s \gamma, d_a \gamma])_{L^2(\mathbb{R} \times Y_r),\epsilon} - (\beta_{sy} [\gamma, \nabla_s d_a \gamma])_{L^2(\mathbb{R} \times Y_r),\epsilon}.
\]
The second term is exactly (91). The first term in (96) can be controlled as follows:

The portion of the integral over \( I \times \Sigma \) is bounded by

\[
c_0(\delta \| \partial_s \nabla s \gamma \|^2_{L^2(R \times I \times \Sigma, \epsilon)} + \delta^{-1} \epsilon^2 \| d^e_\epsilon \partial_s \nabla s \gamma \|^2_{L^2(R \times I \times \Sigma, \epsilon)}).
\]

For the first of these terms, use \( d^e_\epsilon \partial_s \nabla s \gamma = \nabla s d^e_\epsilon \beta \gamma \), then Corollary 4.12 and take \( \delta \) small to absorb the second order terms. For the second term, use the identity \( d^e_\epsilon \partial_s \nabla s \gamma = -\nabla s \beta t + \nabla t \beta s \) to produce terms of the form \( \nabla s d^e_\epsilon \beta \gamma \), \( \nabla t d^e_\epsilon \beta s \) and lower order terms. Having chosen \( \delta \), now pick \( \epsilon \) small enough to absorb the second order terms.

For the first of these terms, use \( \partial_s \nabla s \gamma = \nabla s \partial_s \beta \gamma \), then Corollary 4.12 and take \( \delta \) small to absorb the second order terms. For the second term, use the identity \( d^e_\epsilon \partial_s \nabla s \gamma = -\nabla s \beta t + \nabla t \beta s \) to produce terms of the form \( \nabla s d^e_\epsilon \beta \gamma \), \( \nabla t d^e_\epsilon \beta s \) and lower order terms. Having chosen \( \delta \), now pick \( \epsilon \) small enough to absorb the second order terms.

The portion of the integral over \( Y_r \cap Y \) is bounded similarly. Here one should use \( \sup_R \| \partial_s \beta \|^2_{L^2(Y \epsilon)} = \epsilon \sup_R \| \partial_s \beta \|^2_{L^2(Y \epsilon)} \leq c_0 \epsilon \) as well as the embedding \( W^{1,2} \hookrightarrow L^4 \) on the 3-manifold \( Y_r \cap Y \).

**Proposition 4.14.** (Instanton bootstrapping estimates; 2nd order) Fix \( c_0 > 0 \), open \( \Omega \subset R \) and compact \( K \subset \Omega \). Then there are \( \epsilon_0, \epsilon, C > 0 \) so that

\[
\| \nabla^2 s F_A \|_{L^2(K \times Y, \epsilon)} + \sup_{r > 0} \| \nabla_r \nabla s F_A \|_{L^2(K \times Y_r, \epsilon)} \leq C \left( 1 + \| F_A \|_{L^2(\Omega \times Y, \epsilon)} \right)
\]

for all \( 0 < \epsilon < \epsilon_0 \) and all \( \epsilon \)-ASD connections \( A \) satisfying (68) and (69).

As usual, the constants only depend on the choice of \( K, \Omega \) through the distance from \( \partial K \) to \( \partial \Omega \).

**Proof of Proposition 4.14.** Fix a compactly supported bump function \( h \) for \( K \subset \Omega \). All integrals, inner products, norms, etc. are over \( R \times Y \) unless otherwise specified.

We begin with \( \nabla^2 s F_A \), but in the end we need to compute this simultaneously with \( \nabla_r \nabla s F_A \). The proof is in many ways quite similar to that of Proposition 4.10 so we will be brief, putting most emphasis on the new features. Use integration by parts and the \( \epsilon \)-ASD relation \( d^e_\epsilon F_A = 0 \) to get that the quantity \( \| h \nabla^2 s F_A \|_{L^2 \epsilon}^2 \) is given by a linear combination of the following terms

\[
\int h^2 ds \wedge \langle [\nabla s b_\epsilon \wedge \nabla s b_\epsilon] \wedge b_\epsilon \rangle \quad (97)
\]

\[
\int h(\partial_s h) ds \wedge \langle \nabla s b_\epsilon \wedge \nabla^\epsilon \nabla s b_\epsilon \rangle, \quad (98)
\]

where \( \nabla^\epsilon \) is the \( \epsilon \)-dependent Hodge star on \( Y \). We will estimate these in the bullets below (we separate (97) into two integrals, one over \( R \times Y \epsilon \) and one over \( R \times I \times \Sigma \)).

• (97) on \( R \times Y \epsilon \): Use (68) to control this by
Note that by the $\epsilon$-ASD relation on $\mathbb{R} \times Y_\star$, $\nabla_s b_s = \epsilon^{-1} d_a^\star F_A$ is coexact. Combining this observation with the embedding $W^{1,2} \hookrightarrow L^4$ for $Y_\star$, we can bound this by a constant times

$$c_0 \| h \nabla_s b_s \|^2_{L^2(\mathbb{R} \times Y_\star)} .$$

For the first term in (99), we use (68) on $\nabla s^2 d a b_s$ for some $\delta$ by Proposition 4.10. For the remaining terms we commute $\delta$ for $\Sigma$ becomes a linear combination of the two terms

$$\| h d_a \nabla_s b_s \|^2_{L^2(\mathbb{R} \times Y_\star)} \leq \| h \nabla_s d a b_s \|^2_{L^2(\mathbb{R} \times Y_\star)} + \| h b_s \|^2_{L^2(\mathbb{R} \times Y_\star)} ,$$

where we used $[\nabla s, d a] = [b_s, \cdot]$ in the inequality. Corollary 4.12 provides a uniform bound for the $L^4$-norm. For the derivative term, convert to the $\epsilon$-dependent norm to get

$$\| h \nabla_s d a b_s \|^2_{L^2(\mathbb{R} \times Y_\star)} = \epsilon \| h \nabla_s d a b_s \|^2_{L^2(\mathbb{R} \times Y_\star), \epsilon} \leq \epsilon \| h \nabla_s F_A \|^2_{L^2, \epsilon} .$$

This can absorbed for $\epsilon$ small.

- On $\mathbb{R} \times I \times \Sigma_\star$: In coordinates we have $\nabla_s b_s = \nabla_s \beta_s + dt \wedge \nabla_s \gamma$, so becomes a linear combination of the two terms

$$\int_{\mathbb{R} \times I \times \Sigma_\star} h^2 ds \wedge dt \wedge ([\nabla_s \beta_s \wedge \nabla_s \gamma], \beta_s) ,$$

$$\int_{\mathbb{R} \times I \times \Sigma_\star} h^2 ds \wedge dt \wedge ([\nabla_s \beta_s \wedge \nabla_s \beta_s], \gamma) .$$

For the first term in (99), we use (68) on $\beta_s$, together with the embedding $W^{1,2} \hookrightarrow L^4$ for $\Sigma_\star$ to control this by a constant times

$$\delta^{-1} \left( \| h \nabla_s \beta_s \|^2_{L^2(\mathbb{R} \times I \times \Sigma_\star)} + \| h d_a \nabla_s \beta_s \|^2_{L^2(\mathbb{R} \times I \times \Sigma_\star)} + \| h d_a^\star \nabla_s \beta_s \|^2_{L^2(\mathbb{R} \times I \times \Sigma_\star)} \right)$$

$$+ \delta \| h d_a \nabla_s \gamma \|^2_{L^2(\mathbb{R} \times I \times \Sigma_\star)}$$

for some $\delta > 0$ that we will determine momentarily. The norm of $h \nabla_s \beta_s$ is controlled by Proposition 4.10. For the remaining terms we commute $\nabla_s$ and $d a$ and then convert to the $\epsilon$-dependent norms to get that these remaining terms are bounded by

$$\delta^{-1} \left( \epsilon \| h \nabla_s d a \beta_s \|^2_{L^2(\mathbb{R} \times I \times \Sigma_\star), \epsilon} + \| h \nabla_s d a^\epsilon \beta_s \|^2_{L^2(\mathbb{R} \times I \times \Sigma_\star), \epsilon} \right)$$

$$+ \delta \| h \nabla_s d a \gamma \|^2_{L^2(\mathbb{R} \times I \times \Sigma_\star), \epsilon}$$

plus terms of the form $\| h [\beta_s \wedge \beta_s] \|^2_{L^2}$ and $\| h [\beta_s, \gamma] \|^2_{L^2}$, which are uniformly bounded by Corollary 4.12. By Proposition 4.13 the quantity (100) is controlled by

$$C_1 (\delta^{-1} \epsilon + \delta) \left( 1 + \| h F_A \|^2_{L^2, \epsilon} + \| h \nabla_s F_A \|^2_{L^2, \epsilon} + \| h \nabla_s F_A \|^2_{L^2, \epsilon} \right) .$$

By taking $\delta$ small, and then $\epsilon$ smaller, we can absorb these second derivative terms (recall we should really be estimating $\nabla_s^2 F_A$ and $\nabla_s \nabla_s F_A$ simultaneously).

To bound the second term in (99), integrate by parts in $\nabla_s$ to get that this second term equals a linear combination of
\[
\int_{\mathbb{R} \times I \times \Sigma_\gamma} h^2 ds \wedge dt \wedge \langle [\nabla_s \beta_s \wedge \beta_s], \nabla_s \gamma \rangle, \quad \int_{\mathbb{R} \times I \times \Sigma_\gamma} h^2 ds \wedge dt \wedge \langle [\nabla_s^2 \beta_s \wedge \beta_s], \gamma \rangle, \quad \int_{\mathbb{R} \times I \times \Sigma_\gamma} h(\partial_s h) ds \wedge dt \wedge \langle [\nabla_s \beta_s \wedge \beta_s], \gamma \rangle.
\]

The first of these is exactly the first term in (99) and was already bounded. The second of these is controlled by
\[
\delta \| h \nabla_s^2 \beta_s \|_{L^2(\mathbb{R} \times I \times \Sigma_\gamma)} + \delta^{-1} \| h [\beta_s, \gamma] \|_{L^2(\mathbb{R} \times I \times \Sigma_\gamma)}
\]
which is fine for \( \delta \) small by Corollary 4.12. The last of these terms is bounded by a constant times
\[
\| h \nabla_s \beta_s \|_{L^2(\mathbb{R} \times I \times \Sigma_\gamma), e} + \| h d_t \gamma \|_{L^2(\mathbb{R} \times I \times \Sigma_\gamma), e'}
\]
which is controlled by Proposition 4.10

- (98) on \( \mathbb{R} \times Y \): This is bounded by
\[
\delta \| h \nabla_s^2 b_s \|_{L^2, e} + \delta^{-1} \| \nabla_s b_s \|_{L^2(\text{supp}(h) \times Y), e'}.
\]
The first of these can be absorbed for small \( \delta \); the second is bounded by Proposition 4.10 since \( h \) has compact support.

This completes the argument for \( \nabla_s^2 F_A \), so we move on to the quantity \( \nabla_t \nabla_s F_A \).

Integration by parts as above shows that \( \| h \nabla_t \nabla_s F_A \|_{L^2(\mathbb{R} \times Y_r)} \) is equal to a linear combination of terms of the form
\[
\int_{\mathbb{R} \times Y_r} h^2 ds \wedge dt \wedge \langle [\nabla_t \nabla_s \beta_s \wedge \beta_s], \gamma \rangle, \quad \int_{\mathbb{R} \times Y_r} h^2 ds \wedge dt \wedge \langle [\nabla_s \beta_s \wedge \beta_t], \nabla_t \gamma \rangle, \quad \int_{\mathbb{R} \times Y_r} h^2 ds \wedge dt \wedge \langle [\nabla_t F_A, \gamma], \nabla_t \gamma \rangle, \quad \int_{\mathbb{R} \times Y_r} h(\partial_s h) ds \wedge dt \wedge \langle [\beta_s \wedge \beta_t], \nabla_t \gamma \rangle
\]

\]
together with the terms
\[
\int_{\mathbb{R} \times \partial Y_r} h^2 ds \wedge \langle [\nabla_t \nabla_s \beta_s \wedge \nabla_t \beta_s], \gamma \rangle, \quad \int_{\mathbb{R} \times \partial Y_r} h^2 ds \wedge \langle [\nabla_s \beta_t \wedge \beta_s], \gamma \rangle, \quad \int_{\mathbb{R} \times \partial Y_r} h^2 ds \wedge \langle [\nabla_t \beta_s \wedge \beta_t], \gamma \rangle, \quad \int_{\mathbb{R} \times \partial Y_r} h^2 ds \wedge \langle [\beta_s \wedge \beta_t], \nabla_t \gamma \rangle
\]
coming from integration in the $t$-direction. The first group can be bounded as we did in the case of $\nabla_k^2 F_A$ (recall $|dt| = \epsilon$ on $Y_r \cap Y_\epsilon$), and the second group goes to zero as $r$ decreases to 0, just as with Proposition 4.9.

4.3 Proof of Lemma 4.2

Throughout we write $A_v|_{\{s\}_\times Y} = a_v(s) + p(s) \, ds$, $A_v|_{\{s,t\}_\times \Sigma_k} = a_v(s,t) + \phi_v(s,t) \, ds + \psi(s,t) \, dt$, $\beta_{s,v} = \partial_s a_v - d_v \phi_v$, and $\beta_{t,v} = \partial_t a_v - d_v \psi_v$. We use similar notation for $A_v|_{\{s\}_\times Y}$.

By the assumption on the curvature of the $a_v$, it follows that Theorem 3.6 applies to $a_v(s,t)$ for all $s, t$, provided $v$ is sufficiently large. Let $\NS_k$ be the map constructed in Theorem 3.6 for the bundle $P_i \to \Sigma_k$. Let $M$ be the product symplectic manifold defined in the beginning of Section 4. Define a map $v_v : \mathbb{R} \times I \to M$ by sending $(s,t) \in \mathbb{R} \times I$ to

$$(\Pi \circ \NS_{k_0} (a_v(s,1-t)|_{\Sigma_k}) , \ldots , \Pi \circ \NS_{p_{N-1}} (a_v(s,t)|_{\Sigma_{N-1}})) ;$$

(101)

the terms on $\Sigma_k$ with $k$ even have $1 - t$, and those with $k$ odd have $t$. Then Lemmas 3.15 and 3.16 imply that $v_v$ is holomorphic with respect to the complex structure $\sum_{i=0}^{\infty} (-1)^{r+1} \cdot \Sigma_i$ on $M$. We will denote by $| \cdot |_M$ the norm induced by this almost complex structure. Recall (iii) from the statement of Lemma 4.2. This implies that the energy density of $v_v$ is uniformly bounded on compact sets in $S_0$. Since $S_0 = \mathbb{R} \times I \setminus B$, where $B$ is a finite set, it follows from the removal of singularities theorem for holomorphic maps [23 Theorem 4.1.2 (ii)] that each $v_v$ extends to a holomorphic map defined on all of the interior $\mathbb{R} \times (0,1)$, we do not have Lagrangian boundary conditions, so we it may not extend over the bad points at the boundary $B \cap \mathbb{R} \times \{0,1\}$.

It follows exactly as in Claim 1 appearing in Section 3.2 that for each compact $K \subset S_0$ there is a uniform bound of the form $\sup_K |\partial_t v_v|_M^2 \leq C_K$. In particular, there is a subsequence, still denoted by $\{v_v\}$, that converges weakly in $C^1$, and strongly in $C^0$, on compact subsets of $S_0$, including the boundary. Let $v_\infty \in C^1(S_0, M)$ denote the limiting holomorphic curve. As with the $v_v$, the curve $v_\infty$ extends to $\mathbb{R} \times (0,1)$ and is $C^\infty$ in this region [23 Theorem B.4.1].

**Remark 4.15.** We can actually say quite a bit more: The uniform energy bound implies that, after possibly passing to a further subsequence, we have that the $v_v : \mathbb{R} \times I \to M$ converge to $v_\infty$ in $C^\infty$ on compact subsets of the interior, $S_0 \cap (\mathbb{R} \times (0,1))$ (see [23 Theorem 4.1.1]). In particular, this automatically proves [56] for $K \subset S_0 \cap (\mathbb{R} \times (0,1))$. However, for the proof of Theorem 4.7, we will need this for $K$ intersecting the boundary $\partial S_0$; we address this in Claim 2, below.

Claim 1 below states that $v_\infty$ actually does have Lagrangian boundary conditions. This will follow because the $v_v$ have approximate Lagrangian boundary conditions. To state this precisely, let $L_{(0)} , L_{(1)} \subset M$ be the Lagrangians from the beginning of Section 4. Let $B$ be as in the statement of Lemma 4.2 and let $S_\infty$ denote the
set of $s \in \mathbb{R}$ such that $(s, 0), (s, 1) \notin B$. Note that $\mathbb{R} \setminus S_R$ is a finite set because $B$ is finite.

Claim 1: For $j \in \{0, 1\}$ and for each $s \in S_R$, the sequence $(v(s, j))_v$ converges (in the metric on $M$) to a point in $L_{(j)}$. In particular, since $\mathbb{R} \setminus S_R$ is finite, the map $v_\infty$ extends to a map defined on all of $\mathbb{R} \times I$ with Lagrangian boundary conditions: $v_\infty(\cdot, j) : \mathbb{R} \to L_{(j)}$.

By applying suitable gauge transformations, we may assume each $A_v$ satisfies the conclusions of Lemma 4.7. Fix a compact $K \subset S_R$. Consider the hypothesis in Lemma 4.2 stating that the norms $\|F_{a_v(s)}\|_{L^\infty}$ decay to zero uniformly on $K$. This implies that Theorem 4.5 applies to each $a_v(s)$ for each $s \in K$ and for all $v$ sufficiently large (exactly how large will depend on $K$). For each $i$, let $\Pi \circ \text{Heat}_{Q_i}$ be the map constructed in Theorem 4.5. By restriction to $\partial Y_s$, the assignment sending $s \in K$ to

$$(\Pi \circ \text{Heat}_{Q_i} (a_v(s)|\gamma_0) , \Pi \circ \text{Heat}_{Q_i} (a_v(s)|\gamma_2) , \ldots , \Pi \circ \text{Heat}_{Q_i} (a_v(s)|\gamma_{(N-2)(N-1)}) )$$

determines a map $\ell_{v, 0 : K \to L_{(0)}}$. Similarly, we obtain a map $\ell_{v, 1 : K \to L_{(1)}}$ by using $\Pi \circ \text{Heat}_{Q_i} (a_v(s)|\gamma_{(i+1)})$ with $i$ odd. We will show

$$\sup_{s \in K} \text{dist}_M (\ell_{v, j}(s), v_v(s, j)) \xrightarrow{v \to 0} 0. \quad (102)$$

Then Claim 1 follows by repeating this argument for a sequence of compact $K$ that exhaust $S_R$. This will also establish (55) via the second assertion of Theorem 4.5.

The proof of (102) is just a computation. It is useful to keep track of the bundle $P$ in the notation for the projection $\Pi_P : \mathcal{A}_{flat}(P) \to M(P)$. Let $\text{dist}_M$ denote the distance function on $M$ coming from the metric induced by the $L^2$-inner product on the harmonic spaces $H^1_{a} \cong T_{[a]}M$. Then fixing $s \in K$, we find that $\text{dist}_M (\ell_{v, j}(s), v_v(s, j))^2$ is equal to

$$\sum_i \text{dist}_M(P_{i+j}) \left( \left\{ \Pi_{Q_i} \circ \text{Heat}_{Q_i} (a_v(s)|\gamma_i) \right\} \right)_{\Sigma_{i+j}}^2$$

$$= \sum_i \text{dist}_M(P_{i+j}) \left( \Pi_{P_{i+j}} \left( \left\{ \text{Heat}_{Q_i} (a_v(s)|\gamma_i) \right\} \right)_{\Sigma_{i+j}}^2 \right.$$}

$$\left. \left. = \sum_i \left\{ \text{Heat}_{Q_i} (a_v(s)|\gamma_i) \right\} \right)_{\Sigma_{i+j}}^2 - \Pi_{P_{i+j}} \left( \left\{ \text{NS}_{P_{i+j}} (a_v(s)|\Sigma_{i+j}) \right\} \right)_{\Sigma_{i+j}}^2 \right)$$

The first equality holds because restricting a flat connection on $Q_i$ to the boundary commutes with harmonic projections $\Pi_{Q_i}$ and $\Pi_{P_{i+j}}$; the inequality holds by the definition of the distance on the $M(P_i)$, and because $\Pi_{i+j}$ has operator norm equal to one. Taking the supremum over $s \in K$ and using the triangle inequality, we can continue this to get that $\sup_s \text{dist}_M (\ell_{v, j}(s), v_v(s, j))^2$ is bounded by
A1. the summand goes to zero by Proposition 3.14. This verifies (102) and proves Claim 4.5. It remains to prove (56) and (57). For this, consider the real-valued functions

\[ a_v \] lifts to a smooth \( F \) since \( \nu \) of Theorem 4.5 shows that the first term in the summand goes to zero as \( v \to \infty \), since \( F_{n_v} \) converges to zero in \( L^\infty \) (uniformly in \( v \)). Similarly, the second term in the summand goes to zero by Proposition 3.14. This verifies (102) and proves Claim 1.

Just as in Claim 2 appearing in Section 3.2, the holomorphic strip \( u_\infty : \mathbb{R} \times I \to M \) lifts to a smooth \( \alpha_\infty : \mathbb{R} \times I \to A_{\text{flat}}(F \ast) \), and hence determines a holomorphic strip representative \( A_\infty \in A_{\text{loc}}^{1, q}(\mathbb{R} \times Q) \) (this follows because \( \mathbb{R} \times S^1 \) retracts to its 1-skeleton). Then the convergence statement in (54) follows exactly as the proof of (21) in Lemma 3.5. It remains to prove (56) and (57). For this, consider the real-valued functions

\[ e_v := \| \beta_{s,v} \|_{L^2(Y_\ast)} : S_0 \to \mathbb{R}, \quad f_v := \| b_{s,v} \|_{L^2(Y_\ast)} : S_\mathbb{R} \to \mathbb{R}. \]

In light of the Sobolev embeddings \( W^{2,2} \hookrightarrow C^0 \) and \( W^{1,2} \hookrightarrow C^0 \) for compact sets in dimensions 2 and 1, respectively, the convergence statements (56) and (57) follow immediately from the next claim (after passing to a suitable subsequence).

Claim 2: There is a constant \( C \) so that

\[ \sup_v \| e_v \|_{W^{2,2}(\mathbb{R} \times I)} + \| f_v \|_{W^{1,2}(\mathbb{R})} \leq C. \]

To prove the bound for \( e_v \), apply Kato’s inequality

\[ |d|V| \leq |\nabla V|, \]

with \( V = \beta_{s,v} \) and \( \nabla = ds \otimes \nabla_s + dt \otimes \nabla_t \). Now the result follows immediately from Theorem 4.8 together with the uniform energy bound on the \( A_v \). The bound for \( f_v \) is similar, but use Kato’s inequality with \( V = b_{s,v} \) and \( \nabla = ds \otimes \nabla_s \). This completes the proof of Lemma 4.2.

4.4 Proof of Theorem 4.1

We follow the proof of Theorem 3.3 quite closely, with Lemma 4.2 used in place of Lemma 3.5. In particular, we assume there is some compact \( K \subset \mathbb{R} \) for which one of the following cases holds.
Case 1 \( \|F_{\alpha_v}\|_{L^\infty(K \times I \times \Sigma^1)} + \|F_{\alpha_0}\|_{L^\infty(K \times Y_0)} \to \infty; \)

Case 2 \( \|F_{\alpha_v}\|_{L^\infty(K \times I \times \Sigma^1)} + \|F_{\alpha_0}\|_{L^\infty(K \times Y_0)} \to \Delta > 0; \)

Case 3 \( \sup_{(s,t) \in K} \|b_{\alpha v}(s,t)\|_{L^2(\Sigma^1)} + \sup_{s \in K} \|b_{\alpha v}(s)\|_{L^2(Y_0)} \to \infty, \)

Most of the work is in showing that each case leads to energy quantization. Supposing we have shown this, it would then follow (from Lemma 4.2 applied to the complement of the bubbling set) that a subsequence of the \( A_v \) converges in the sense of the statement of Theorem 4.1 (with \( j = 1 \) and \( s^1 = 0 \)) to a limiting holomorphic strip representative \( A^1 \) on \( R \to Z \). At this stage, the only real difference between this non-compact case and the compact case of Theorem 3.3 is that here one does not know that \( A^1 \) has non-zero energy. For example, it could be the case that all of the energy has escaped to \( \infty \), and so \( A^1 \) is a flat connection. To rectify this, one incorporates time translations; we defer a complete discussion of this until after of our case analysis for energy quantization.

Case 1. (Instantons on \( S^4 \)) By passing to a subsequence, we may assume the \( L^\infty \)-norm of each curvature is always achieved on \( \Sigma_{j+1} \) or \( Y_{j+1} \) for some \( j \). That is, one of the following holds for all \( v \):

\[
\|F_{\alpha_v}\|_{L^\infty(K \times I \times \Sigma^1)} \geq \max \left\{ \|F_{\alpha_v}\|_{L^\infty(K \times I \times \Sigma^1)}, \|F_{\alpha_0}\|_{L^\infty(K \times Y_0)} \right\} \quad (103)
\]

\[
\|F_{\alpha_v}\|_{L^\infty(K \times Y_{j+1})} \geq \max \left\{ \|F_{\alpha_v}\|_{L^\infty(K \times I \times \Sigma^1)}, \|F_{\alpha_0}\|_{L^\infty(K \times Y_0)} \right\}. \quad (104)
\]

If (103) holds, find a point \( (s_v, t_v) \in K \times I \) so that

\[
\|F_{\alpha_v(s_v, t_v)}\|_{L^\infty(\Sigma_{j+1})} = \|F_{\alpha_0}\|_{L^\infty(K \times I \times \Sigma_{j+1})}.
\]

Similarly, if (104) holds, then find a point \( s_v \in K \) with

\[
\|F_{\alpha_v(s_v)}\|_{L^\infty(Y_{j+1})} = \|F_{\alpha_0}\|_{L^\infty(K \times Y_{j+1})}.
\]

By passing to a subsequence, we may suppose the \( s_v \) converge to some element of \( K \subset \mathbb{R} \). Similarly, we may assume \( t_v \to t_\infty \in I \) converges. Strictly speaking, we need to distinguish between whether \( t_\infty \) lies in the interior of \( I \) or on the boundary. However, the analysis for when \( t_\infty \) lies in the boundary can be incorporated to the analysis for when (104) holds. Precisely, we find ourselves considering the following subcases:

Subcase 1 (103) holds and \( t_\infty \neq 0, 1; \)

Subcase 2 (104) holds, or (103) holds and \( t_\infty = 0, 1. \)

Without loss of generality, we may suppose \( j = 0 \) and \( t_\infty \in [0, 1) \).

In Subcase 1, for each \( v \), define a rescaled connection \( \tilde{A}_v \) in terms of its components as follows:
\[
\begin{align*}
\tilde{a}_\nu(s,t) & := a_\nu(\epsilon_\nu s + s_\nu, \epsilon_\nu t + t_\nu)|_{\Sigma_1} \\
\tilde{\phi}_\nu(s,t) & := \epsilon_\nu \phi(\epsilon_\nu s + s_\nu, \epsilon_\nu t + t_\nu)|_{\Sigma_1} \\
\tilde{\psi}_\nu(s,t) & := \epsilon_\nu \psi(\epsilon_\nu s + s_\nu, \epsilon_\nu t + t_\nu)|_{\Sigma_1}.
\end{align*}
\]

We view these as connections and 0-forms defined on \(B_{r^{-1}} \times \Sigma_1 \subseteq C \times \Sigma_1\), where \(\eta = \frac{1}{2} \min\{t_\infty, 1 - t_\infty\}\), \(B_r \subseteq C\) is the ball of radius \(r\) centered at zero, and we assume \(\nu\) is large enough so \(t_\nu \leq \eta\).

In Subcase 2, for each \(\nu\) we define a connection \(\tilde{A}_\nu\) on a neighborhood of \(\mathbb{R} \times X_{01}\) as follows:

\[
\begin{align*}
\tilde{a}_\nu(s) & := a_\nu(\epsilon_\nu s + s_\nu)|_{Y_{01}}, \\
\tilde{\phi}_\nu(s) & := \epsilon_\nu \phi(\epsilon_\nu s + s_\nu)|_{Y_{01}}, \\
\tilde{\psi}_\nu(s) & := \epsilon_\nu \psi(\epsilon_\nu s + s_\nu)|_{Y_{01}},
\end{align*}
\]

which we view as a connection defined on \(\mathbb{R} \times X_{01}(\epsilon_\nu^{-1})\), where \(Y_{01}(r)\) is the manifold \(Y_{01}\) with a cylinder of length \(r\) attached to each boundary component. This construction will come up repeatedly, so we introduce some notation: Given \(r > 0\) and a smooth \(X\), we set

\[
X(r) := X \cup \partial X [0, r) \times \partial X, \quad X^\infty := X \cup \partial X [0, \infty) \times \partial X.
\]

**Remark 4.16.** There exist smooth structures on these spaces that are compatible in the sense that the inclusions

\[
X(r) \subseteq X(r') \subseteq X^\infty
\]

are smooth embeddings for \(r \leq r'\). If \(X\) has a metric \(g\), then we will consider the metric on \(X(r)\) and \(X^\infty\) that is given by \(g\) on \(X\) and \(dt^2 + g|_{\partial X}\) on the end. In particular, the embeddings become metric embeddings. This will be called the fixed metric on the given manifold, and we denote its various norms by \(\| \cdot \|, \| \cdot \|_{L^p}, \text{etc.}\) If \(X\) is equipped with a bundle \(B \to X\) then we define bundles \(B(r) \to X(r)\) and \(B^\infty \to X^\infty\) in the obvious way. Note also that we have the following decomposition

\[
\mathbb{R} \times X^\infty = \left( \mathbb{R} \times X \right) \cup \left( \mathbb{H} \times \partial X \right).
\]

In both Subcases the connections \(\tilde{A}_\nu\) are ASD with respect to the fixed metric, and have uniformly bounded energy \(\frac{1}{2} \| F_{\tilde{A}_\nu} \|_{L^2}^2 \leq CS(a^+) - CS(a^-)\); here the norm should be taken on the domain on which the connection is defined. Furthermore, the energy densities are bounded from below:
In particular, the condition of Case 1 implies that \( \|F_{\overline{\alpha}}\|_{L^\infty} \to \infty \). Following the usual rescaling argument [30] [8] Section 9] (see also [23] Theorem 4.6.1) for the closely-related case of \( J \)-holomorphic curves we can conformally rescale in a small neighborhood \( U \) of the blow-up point to obtain a sequence of finite-energy instantons with energy density bounded by 1 and defined on increasing balls in \( \mathbb{R}^4 \). By Uhlenbeck’s strong compactness theorem, there is a subsequence that converges, modulo \( \overline{\alpha} \) and with energy density bounded by 1 and defined on increasing balls in \( \mathbb{R}^4 \).

By Uhlenbeck’s removable singularities theorem this extends to a non-constant instanton, also denoted by \( \overline{\alpha} \), on a \( PU(r) \)-bundle \( R_\infty \to S^4 \). Since \( \overline{\alpha} \) is ASD and non-constant we have

\[
0 < \frac{1}{2} \int_{S^4} |F_{\overline{\alpha}}|^2 = -\frac{1}{2} \int_{S^4} (F_{\overline{\alpha}} \wedge F_{\overline{\alpha}}) = 2\pi^2 k_r r^{-1} q_4(R_\infty).
\]

As in Case 1 from Section 3.3 this means that we have energy quantization with \( h = 4\pi^2 k_r \).

**Case 2. (Instantons on non-compact domains)** This case is much the same as the previous, in that instantons near the blow-up point bubble off. However, this time the geometry of the underlying spaces on which these bubbles form can be more exotic. Define \( \overline{\alpha} \) exactly as in Case 1 above. Everything up to and including equation (110) continues to hold. In particular, \( \lim \inf \|F_{\overline{\alpha}}\|_{L^\infty} \) is bounded from below by \( \Delta > 0 \). After possibly passing to a subsequence, we may assume the energy densities \( \|F_{\overline{\alpha}}\|_{L^\infty} \) converge to some \( \Delta' \in [\Delta, \infty] \). If \( \Delta' = \infty \) then we are done by precisely the same analysis as in Case 1. So we may assume \( 0 < \Delta' < \infty \), in which case we can apply Uhlenbeck’s strong compactness theorem directly to the sequence \( \overline{\alpha} \). We may therefore assume this sequence converges to a non-flat finite-energy instanton \( \overline{\alpha} \) on a bundle over one of the spaces \( \mathbb{R} \times Y_{01}^\infty \) or \( C \times \Sigma_1 \), depending on whether we are in Subcase 1 or 2 (see the discussion above Remark 4.16 for a definition of \( Y_{01}^\infty \)).

We show in [12] that the energy of any such instanton \( \overline{\alpha} \) is \( 4\pi^2 k_r r^{-1} k \) for some positive \( k \in \mathbb{N} \).

**Case 3. (Holomorphic spheres and disks in \( M(P) \))** For each \( \nu \), let \( c_\nu \) be the supremum of the numbers

\[
\|b_{s,t}(s)\|_{L^2(\Sigma_\nu)}, \quad \|b_{s,t}(s)\|_{L^2(Y_\nu)}
\]

over \( s \in K \) and \( t \in I \). The conditions of this case imply that \( c_\nu \to \infty \). Find \( j_\nu \in \{0, \ldots, N-1\} \) and points \( (s_j,t_\nu) \in K \times I \) for which

\[
c_\nu = \|b_{s_j,t_\nu}(s)\|_{L^2(\Sigma_\nu)}, \quad \text{or} \quad c_\nu = \|b_{s_j,t_\nu}(s)\|_{L^2(Y_{(j+1)\nu})}.
\]

such points exist since \( b_{s_j}, b_{s_j} \) decay at \( \pm \infty \), due to the finite energy assumption; alternatively, one could replace \( c_\nu \) by \( c_\nu/2 \), without changing the argument below. If \( c_\nu = \|b_{s_j}(s)\|_{L^2(Y_{(j+1)\nu})} \), then we just declare \( t_\nu = 0 \). By passing to a subsequence,
we can assume that \( j_v = 1 \) for all \( v \), and that the \((s_v, t_v)\) converge to some \((s_\infty, t_\infty) \in K \times I\). The two relevant subcases to consider are as follows:

**Subcase 1** \( t_\infty \in (0, 1) \)

**Subcase 2** \( t_\infty \in \{0, 1 \} \)

We may assume, without loss of generality, that if Subcase 2 holds then \( t_\infty = 0 \). Define rescaled connections \( \hat{A}_v \) using \((105)\) and \((106)\), except replace every \( \epsilon_v \) by \( \epsilon_v^{-1} \) (the subcases here correspond to those from Case 1 in the obvious way).

Subcase 1 can be treated exactly as Case 3 from Section 3.3. For Subcase 2 we will use an argument similar to the one appearing in that section. The argument will show that a holomorphic disk bubble appears in \( M(P_0) \times M(P_1) \) with Lagrangian boundary conditions in \( L(Q_{01}) \subset M(P_0) \times M(P_1) \).

We view the rescaled connections \( \hat{A}_v \) as being defined on \( \mathbb{R} \times Y_{01}(c_v) \) (see the discussion above Remark 4.16). The components of \( F_{\hat{A}_v} \) satisfy

\[
\hat{\beta}_{s,v} + * \hat{\beta}_{b,v} = 0, \quad \hat{\gamma} = -\hat{\epsilon}_v^{-2} * F_{\hat{A}_v}, \quad \hat{b}_{s,v} = -\hat{\epsilon}_v^{-1} F_{\hat{b}_s},
\]

where \( \hat{\epsilon}_v := c_v \epsilon_v \), and the Hodge star is the one on the surface \( \Sigma_0 \sqcup \Sigma_1 \). It may not be the case that the \( \hat{\epsilon}_v \) are decaying to zero; this is replaced by the assumption in this case that the slice-wise curvatures converge to zero in \( L^\infty \):

\[
\| F_{\hat{b}_s} \|_{L^\infty} = \| F_{\hat{A}_v} \|_{L^\infty} \longrightarrow 0, \quad \| \hat{b}_{s,v} \|_{L^\infty} = \| F_{\hat{A}_v} \|_{L^\infty} \longrightarrow 0.
\]

Our choice of rescaling also gives

\[
1 \leq \| \hat{\beta}_{s,v}(0,0) \|_{L^2(\Sigma_1)} + \| \hat{\beta}_{b,v}(0) \|_{L^2(Y_{01})} \leq 2. \tag{111}
\]

By arguing as in Lemma 4.2 it follows that, after possibly passing to a subsequence, there exists a sequence of gauge transformations \( \mu_v : \mathbb{H} \rightarrow G(P_0 \sqcup P_1) \), and a limiting connection \( \hat{A}_\infty \in A(\mathbb{R} \times Q_{01}^\infty) \) that is a holomorphic representative

\[
\hat{\beta}_{s,\infty} + * \hat{\beta}_{b,\infty} = 0, \quad F_{\hat{b}_s,\infty} = 0, \quad F_{\hat{A},\infty} = 0,
\]

and satisfies \((56)\) and \((57)\); here we are using \((109)\) to, e.g., view \( a_\infty \) as a map defined on \( \mathbb{H} \). Let \( \Pi_{R_0} : A_{\text{flat}}(P_0) \rightarrow M(P_0) \) and \( \Pi_{Q_{01}} : A_{\text{flat}}(Q_{01}) \rightarrow L(Q_{01}) \) be the projections to the moduli spaces. Then

\[
v_\infty := (\Pi_{R_0}(\hat{A}_\infty|_{\Sigma_0}), \Pi_{P_1}(\hat{A}_\infty|_{\Sigma_1})) : \mathbb{H} \longrightarrow M(P_0) \times M(P_1)
\]

is a holomorphic curve with Lagrangian boundary conditions \( \mathbb{R} \rightarrow L(Q_{01}) \subset M(P_0) \times M(P_1) \) determined by \( a_\infty : \mathbb{R} \rightarrow A_{\text{flat}}(Q_{01}) \). Furthermore, \( v_\infty \) has bounded energy

\[
\int_\mathbb{H} |\partial_t v_\infty|^2 = \int_{\mathbb{H} \times [\Sigma_1] \cup \Sigma_2} |\hat{\beta}_{s,\infty}|^2 \leq \lim \inf v \| \beta_{s,v} \|^2_{L^2(\mathbb{R} \times Y),c_v} \leq \lim \inf v \frac{1}{2} \| F_{\hat{A}_v} \|^2_{L^2(\mathbb{R} \times Y),c_v} = (CS(a^+) - CS(a^-)). \tag{112}
\]
In particular, the removal of singularities theorem [23, Theorem 4.1.2 (ii)] applies and so \( v_\infty \) extends to a holomorphic disk \( v_\infty : D \to M(P_1) \times M(P_2) \) with Lagrangian boundary conditions. Then (56) and (57) combine with (111) to give

\[
2|\partial sv_\infty(0,0)| \geq \liminf_v \|\dot{\beta}_{sv}(0,0)\|_{L^2(\Sigma_1)} + \|\dot{\beta}_{sv}(0)\|_{L^2(Y_{01})} \geq 1.
\]

In particular, \( v_\infty \) is non-constant. Since \( v_\infty \) is a disk with boundary conditions in a simply-connected Lagrangian, it follows that \( v_\infty \) has an extension to a map of the form \( \dot{v}_\infty : S^2 \to M(P_1) \times M(P_2) \) that agrees with \( v_\infty \) on one hemisphere \( D \subseteq S^2 \) and lies in the Lagrangian in the other hemisphere. Then the energy of \( v_\infty \) is given by

\[
\int_D |\partial_s v_\infty|^2 = -\int_D \omega(\partial_s v_\infty, \partial_t v_\infty) = -\int_{S^2} \omega(\partial_s \dot{v}_\infty, \partial_t \dot{v}_\infty) = (\dot{v}_\infty^* \omega) \left[ S^2 \right],
\]

where the second equality holds since \( \omega \) vanishes on the Lagrangian. By monotonicity, it follows that the energy is an integer multiple of \( h = 4\pi^2 Kr^{-1} \). This finishes the analysis for Case 3.

Finally, we address translations; we follow the strategy of [29]. The moduli space of flat connections on \( Q \) is canonically identified with the set of Lagrangian intersection points \( L(0) \cap L(1) \), and the non-degeneracy assumption on the elements of \( \mathcal{A}_{\text{flat}}(Q) \) implies that \( L(0) \cap L(1) \) is a finite set in \( M \); see [10, Section 4]. In particular, there is some \( \epsilon_0 > 0 \) so that \( B_\epsilon(p) \cap B_\epsilon(p') = \emptyset \), for all \( p, p' \in L(0) \cap L(1) \). Define \( v_v \) as in (101). By assumption, each \( A_v \) converges at \( \pm \infty \) to the flat connection \( a^\pm \). Since the maps \( NS_P \) preserve flat connections, it follows that each \( v_v \) converges at \( \pm \infty \) to the Lagrangian intersection point \( p^\pm \in L(0) \cap L(1) \) associated to \( a^\pm \). Define

\[
s_v^1 := \sup \left\{ s \in \mathbb{R} \mid \text{dist}_{M(P_v)}(p^-, v_v(s,t)) \leq \epsilon_0, \quad \forall t \in I \right\}.
\]

(We may assume \( p^- \neq p^+ \), otherwise all instantons are trivial and all holomorphic curves are constant; in particular, the set defining \( s_v^1 \) is non-empty.) Then for each \( v \) we have

\[
\text{dist}_{M(P_v)} \left( p^-, (\tau_{s_v^1} v_v)(s,t) \right) \leq \epsilon_0 \quad \forall t \in I, \forall s \leq 0, \tag{113}
\]

\[
\text{dist}_{M(P_v)} \left( (\tau_{s_v^1} v_v)(0,t), p^+ \right) = \epsilon_0 \quad \text{for some } t \in I. \tag{114}
\]

Then the case analysis above combines with Lemma [4,2] to show that, after passing to a subsequence, the translates \( \tau_{s_v^1} v_v \) converge on compact sets off of a finite bubbling set to a limiting holomorphic strip \( v^1 \). This limits to some Lagrangian intersection point \( p^0 \) at \( -\infty \) and \( p^1 \) at \( \infty \). By (113) and the definition of \( \epsilon_0 \), we must have \( p^0 = p^- \) and \( p^1 = p^+ \). On the other hand, the equalities expressed in (114) show that \( v^1(0,t) \) is not at a Lagrangian intersection point for some \( t \in I \). In particular, \( v^1 \) is non-constant and so \( p^1 \neq p^0 \). This proves that \( v^1 \) has positive energy, and also shows that the \( \tau_{s_v^1} v_v \) become arbitrarily close to \( p^1 \).
Continue inductively with $p^1$ replacing $p^0$, etc. to obtain a sequence of limiting holomorphic strips $\nu'$ that limit to Lagrangian intersection points $p^j$ and $p'$. The theorem follows by lifting the $\nu'$ and $p'$ to representatives, and converting the convergence of the $\nu'_j \nu\nu_\nu$ to statements about the representatives, as we did in the proof of Lemma 4.2.

5 Perturbations

Theorems 3.3 and 4.1 both have extensions to the case where the ASD equations are perturbed by a suitable function. We describe this now, freely referring to the notation established above.

We begin with Theorem 3.3. Suppose $H$ is a section of the bundle

$$T^*S \otimes \text{Maps}(\mathcal{A}(P), \mathbb{R})^G(P) \to S,$$

where $\text{Maps}(\mathcal{A}(P), \mathbb{R})^G(P)$ is the space of gauge invariant real-valued maps on $\mathcal{A}(P)$. Fixing $x \in S$ and $v \in T_x S$, then the differential of the contraction $t_v H(x) : \mathcal{A}(P) \to \mathbb{R}$ can be represented by some $X^H_{x,v} : \mathcal{A}(P) \to \Omega^1(\Sigma, \mathcal{P}(g))$ in the sense that

$$d(t_v H(x))_\alpha(v) = \int_{\Sigma} \langle X^H_{x,v}(\alpha) \wedge v \rangle$$

for all $\alpha \in \mathcal{A}(P)$ and all $v \in \Omega^1(\Sigma, \mathcal{P}(g))$. We assume $H$ has been chosen so that for each $x, v,$

(a) $\sup_{\alpha \in \mathcal{A}(P)} \|X^H_{x,v}(\alpha)\|_{L^\infty(\Sigma)} < C_{x,v}$, for some $C_{x,v}$ that depends continuously on $x, v$, and

(b) if $\{A_v\}$ is any sequence of connections on $\mathcal{R}$ such that $\|t_v (A_v - A_0)\|_{L^p(Z)}$ is bounded, then $X^H_{x,v}(t_v A_v)$ has an $L^p(Z)$-convergent subsequence.

The main example we have in mind is when $H$ is defined by the holonomy around loops in $\Sigma$; Kronheimer shows [21] Lemma 10] that these conditions are always satisfied for such $H$.

Allowing $x$ and $v$ to vary, $X^H_{x,v}$ naturally determines a map $X = X^H : \mathcal{A}(\mathcal{R}) \to \Omega^2(Z, \mathcal{P}(g))$ by declaring $X(A)$ to be the 2-form defined for $v \in T_x S, w \in T\Sigma$ by

$$X(A)(v, w) := X^H_{x,v}(a(x))(w), \quad X(A)(w, v) := -X(A)(v, w)$$

and defined to be zero otherwise. This formula combines with the $G(P)$-invariance of $H$ to imply that $X$ satisfies $X(U^* A) = \text{Ad}(U^{-1}) X(A)$ for all $U \in G(\mathcal{R})$. In local coordinates $x = (s, t)$ on $\Sigma$, $X$ has the form $X(A) = ds \wedge F_s(x) + dt \wedge G_t(x)$ for $x$-dependent $F_s(x), G_t(x) \in \Omega^1(\Sigma, \mathcal{P}(g))$.

The relevant perturbed $e$-instanton equation is

$$(F_A - X(A))^+ := \frac{1}{2} (1 + *_e) (F_A - X(A)) = 0.$$
In local coordinates, this condition has the form \( \textbf{15} \), where the zero on the right-hand side of the top equation is replaced by \( F_x(a) + *_X G_x(a) \). On the symplectic side, the gauge invariance means that \( X \) determines a 1-form on \( S \) with values in the vertical bundle in \( TM(P) \); we denote this 1-form by the same symbol \( X \). The relevant holomorphic curve equation is
\[
(\nu - \text{proj}_\alpha X(a)) + *_\Sigma (\nu - \text{proj}_\alpha X(a)) \circ j = 0,
\]
where \( \alpha \) is a lift of \( v \), and \( \text{proj}_\alpha \) is the \( L^2 \)-orthogonal projection to the \( \alpha \)-harmonic space (the gauge invariance of \( H \) implies that this is independent of the lift \( \alpha \)). Suppose \( A \in \mathcal{A}(R) \) represents a solutions of this equation. In local coordinates, this has the form \( \textbf{19} \), where the zero in the right-hand side of the first equation is again replaced by \( F_x(a) + *_X G_x(a) \). Then Theorem \textbf{3.3} continues to hold in this X-perturbed setting, provided one replaces the energies \( \mathcal{E}^{\text{sym}} \) and \( \mathcal{E}^{\text{inst}} \) with the X-perturbed energies obtained by replacing \( F_A \) with \( \bar{F}_A = X(A) \).

The adjustments to the proofs are as follows: One should replace every appearance of \( D_x \alpha_v \) with \( D_x \alpha_v - X(\alpha_v) \), and likewise with \( D_x \alpha_\infty \). Then Claim 1 in the proof of Lemma \textbf{3.5} would show \( Dv_v - Dv_0 (\Pi \circ \text{NS}) X(\alpha_v) \) is uniformly bounded on compact subsets of \( S_0 \) (in local coordinates, the \( ds \)-component of this is \( D_v v - D_v (\Pi \circ \text{NS}) F_x(\alpha_v) \)). It follows from assumption (a) on \( H \) that the \( Dv_v \) are uniformly bounded, and so converge weakly in \( C^1 \) on compact sets to a limiting \( v_\infty \). The \( v_v \) satisfy \( \bar{D} v_v = X^{(0,1)}_v \), where \( X^{(0,1)}_v \) is the \((0,1)\)-component of \( D_v (\Pi \circ \text{NS}) (X(\alpha_v)) \). It follows from assumption (b) on \( H \) that the limiting map \( v_\infty \) satisfies \( \textbf{115} \). Then the proofs of Lemma \textbf{3.5} and Theorem \textbf{3.3} continue as in the unperturbed case.

Now we address Theorem \textbf{4.1}. Let \( H \) be a function as above, with \( S := \mathbb{R} \times I \) and \( P := P_* \). We also assume that \( X \) and its first derivatives are bounded in \( C^0 \) on \( \mathbb{R} \times I \times \Sigma_* \) (see the next paragraph for more details). In addition to these assumptions, we assume that, for any connection \( A \) over \( \mathbb{R} \times I \times \Sigma_* \), the induced 2-form \( X(A) \) vanishes to all orders near \( \mathbb{R} \times \{0,1\} \times \Sigma_* \). For example, all of these conditions are satisfied if \( H \) is a suitable holonomy perturbation; see \[10\] Section 5.3. Then \( X \) admits a canonical smooth extension as a map \( X : \mathcal{A}(R) \rightarrow \Omega^2(Z, R(g)) \), by declaring \( X(A) \) to be zero on \( \mathbb{R} \times \Sigma_* \).

The modifications to the proofs of Theorem \textbf{4.1} and Lemma \textbf{4.2} are as follows: Of course, one needs to make the modifications discussed in the \( S \times \Sigma \) case above. In addition to these, one needs to extend the analysis of Section \textbf{4.2}. The only real difference from the unperturbed case shows up in \[82\], where we used the \( \epsilon \)-ASD relations. In the perturbed case there are terms of the form \( \int h^2 (\nabla_s F_A \wedge \nabla_s X(A)) \) and \( \int h^2 (\nabla_s F_A \wedge *_{\epsilon} \nabla_s X(A)) \) appearing on the right-hand side of \[82\]. Working in temporal gauge, for simplicity, we have

\[
\nabla_s X(A) = ds \wedge \partial_t F(a) + dt \wedge \partial s G(a) = ds \wedge (\partial_t F(a) + DF(\beta_s)) + dt \wedge (\partial s G(a) + DG(\beta_s)),
\]

where \( DF, DG \) are derivatives of \( F,G \) in the \( \alpha \)-variable, and \( \partial_t F, \partial s G \) are the derivatives in \( s \). Then these additional terms are all controlled by
\[
C (\| \nabla_s \beta_s \|_\epsilon + \| \nabla_s \beta_t \|) \| \beta_s \|_\epsilon,
\]

\[74\]
where the norms are $L^2$-norms, and $C$ depends linearly on the $C^0$ norm of the derivatives of $X$ (which we assumed are bounded). It follows that (116) is bounded by $\delta \| \nabla s F_A \|^2 + \delta^{-1} C \| F_A \|^2$ for any $\delta > 0$, and so the $\nabla s F_A$ term can be absorbed into the left-hand side of (82). Then the proof of Lemma 4.2 goes through as before. If we assume $H$ is independent of $R$, then Theorem 4.1 holds with the obvious modifications made to the statement, analogous to the case of $S \times \Sigma$ above; one also needs to replace every occurrence of the word ‘flat’ with ‘$H$-flat’ (see [10] for a definition).

Theorem 4.1 has an extension to the case when $H$ does depend on $R$ as well. In this case we assume that $H$ is independent of $s$ when $|s|$ is large. One has to be a little careful working with the translations as we did at the end of Section 4.4. Let $X_s$ be the perturbation 1-form coming from $H$, but translated by $s$, and let $X_{\pm}$ be the perturbation 1-form that is constant in the $R$-direction and agrees with $X_s$ for $s$ near $\pm \infty$. Since we do not have translational invariance, the translated holomorphic curves now satisfy a perturbed equation of the form $\partial_s v + J \partial_t v = X_{s^j}(v)$. However, since $H$ is independent of $s$ for large $|s|$, these $v$-dependent perturbations can be controlled. In particular, for each $j$ a subsequence of $\tau^* s^j v$ converges to a curve $v^j$ that satisfies the perturbed equation with $X_-, X_+^j$ or $X_+$ on the right-hand side, depending on whether $s^j$ converges to $-\infty, s^j \in R$ or $+\infty$.

References


[29] D. Salamon. *Morse theory, the Conley index and Floer homology.*


