Kyoto University

Kyoto-Math 2014-02

Large dynamics of Yang—Mills theory: mean dimension formula

by

Masaki Tsukamoto

July 2014

京都大学理学部数学教室
Department of Mathematics
Faculty of Science
Kyoto University
Kyoto 606-8502, JAPAN

LARGE DYNAMICS OF YANG–MILLS THEORY: MEAN DIMENSION FORMULA

MASAKI TSUKAMOTO

Abstract. This paper studies the Yang–Mills ASD equation over the cylinder as a non-linear evolution equation. We consider a dynamical system consisting of bounded orbits of this evolution equation. This system contains many chaotic orbits, and moreover it becomes an infinite dimensional and infinite entropy system. We study the mean dimension of this huge dynamical system. Mean dimension is a topological invariant of dynamical systems introduced by Gromov. We prove the exact formula of the mean dimension by developing a new technique based on the metric mean dimension theory of Lindenstrauss–Weiss.

1. Introduction

1.1. **Main result.** This paper explores a large chaotic dynamics of **Yang–Mills gauge theory**. Yang–Mills theory is the study of special connections (Yang–Mills connections, ASD connections and its perturbations) on principal fiber bundles over manifolds. Its origin is quantum physics, and it has been intensively studied in differential/algebraic geometry, low-dimensional topology and representation theory. Many astonishing results have been obtained for more than 30 years. But its *dynamical* aspect has been largely neglected. The purpose of the paper is to reveal a new rich dynamical structure of Yang–Mills theory.

Traditionally most researchers in Yang–Mills theory have been interested in highly concentrated special connections called **instantons**. Probably this is a reason why dynamical aspect of the theory has not attract their attentions for a long time. When we look at only concentrated solutions, we don't need a dynamical point of view. Dynamics appears only when we are interested in a very long term phenomena. For example, calculating geodesics on Riemannian manifolds is the simplest problem in calculus of variations. But when we look at very long geodesics (i.e. geodesic flow), we face a complicated dynamical problem.

To explain our viewpoint more concretely, we recall a familiar picture of instanton Floer homology (Floer [8] and Donaldson [4]). Let Y be a closed oriented Riemannian

Date: July 8, 2014.

2010 *Mathematics Subject Classification.* 58D27, 53C07, 37B99.

Key words and phrases. Yang–Mills gauge theory, mean dimension, metric mean dimension.

This paper was supported by Grant-in-Aid for Young Scientists (B) 25870334 from JSPS.

3-manifold, and we consider the cylinder $\mathbb{R} \times Y$ with the product metric. We denote its $\mathbb{R}$-coordinate by t. Let E be a principal $SU(2)$ bundle over $\mathbb{R} \times Y$. A connection A on E is said to be **anti-self-dual** (ASD) if its curvature F_A is anti-self-dual with respect to the Hodge star operation:

$$*F_A = -F_A.$$

It is a crucial point in Floer theory that this equation can be expressed as a non-linear *evolution equation*. Suppose A is expressed in the temporal gauge, i.e. it has no dt-part. Then the ASD equation becomes

$$\frac{\partial A(t)}{\partial t} = - *_3 F(A(t)), \tag{1.1}$$

where $A(t)$ is the restriction of A to the section $\{t\} \times Y$. Fixed points of the equation (1.1) are flat connections, and connecting orbits between fixed points correspond to instantons. Floer homology is constructed by using these objects. Generators of Floer chain complex are flat connections, and the differentials involve instanton counting. Therefore we can say that Floer homology uses some dynamics of the evolution equation (1.1).

But the equation (1.1) also contains more complicated dynamical objects other than fixed points and connecting orbits. Firstly the equation (1.1) admits many periodic orbits. Periodic points of period $T > 0$ correspond to instantons over $(\mathbb{R}/T\mathbb{Z}) \times Y$, and a lot of such solutions can be constructed by using the gluing theorem of Taubes [23]. Secondly, and more importantly, the above evolution equation contains many chaotic orbits similar to ones in the **Bernoulli shift** $\{0,1\}^{\mathbb{Z}}$. This can be shown by using *infinite gluing technique* [25, 27] as follows. Pick up two sufficiently concentrated instantons A_0 and A_1 over the Euclidean space $\mathbb{R}^4$. We consider the gluing of infinitely many copies of A_0 and A_1 over $\mathbb{R} \times Y$. Take a point $x = (x_n)_n$ in the Bernoulli shift $\{0,1\}^{\mathbb{Z}}$. For each $n \in \mathbb{Z}$ we glue A_0 or A_1 in a neighborhood of $\{t = n\}$ depending on whether $x_n = 0$ or $x_n = 1$. Then, in a rough expression, the resulting ASD connection A_x looks like

$$A_x = \cdots \sharp A_{x_{-1}} \sharp A_{x_0} \sharp A_{x_1} \sharp \cdots .$$

The dynamical behavior of A_x imitates that of the point x in the Bernoulli shift, and it is generically chaotic.

Indeed the dynamics of (1.1) is much more complicated than the Bernoulli shift. Suppose A_0 and A_1 admit non-trivial deformation. Then each A_{x_n} can be deformed. So the ASD connection A_x has infinitely many deformation parameters. This means that the equation (1.1) contains a dynamics like $[0,1]^{\mathbb{Z}}$ (the shift action on the **Hilbert cube**). $[0,1]^{\mathbb{Z}}$ is an *infinite dimensional* dynamical system of *infinite topological entropy*. So this is much larger than the Bernoulli shift.

We have explained that the ASD equation (1.1) contains a huge dynamics. The purpose of the paper is to develop this unexplored aspect of gauge theory. One motivation of this study comes from the work of Gromov [11]. He introduced a new topological invariant of dynamical systems called **mean dimension**. This provides a non-trivial information for infinite dimensional and infinite entropy systems. For example the $\mathbb{Z}$-action on the Hilbert cube $[0,1]^{\mathbb{Z}}$ has mean dimension 1. Mean dimension has been attracting researchers in several areas such as topological dynamics [19, 17, 12, 13, 18, 14], function theory [2, 21, 26] and operator algebra [16, 7]. We review the definition of mean dimension in Section 2.1.

While the idea of mean dimension is related to various subjects, Gromov's original motivation is geometric. When we study geometric PDE (holomorphic/harmonic maps, complex/minimal subvarieties, etc.) in a non-compact manifold without any asymptotic boundary condition, we often encounter a very large dynamical system (as we have seen above). Gromov proposed the study of such large dynamical systems from the viewpoint of mean dimension. Very little has been known in this direction yet. But here we report one progress of this program in the case of Yang–Mills theory: We get the exact formula of the mean dimension. Probably our method can be also applied to other equations. We will discuss this point again in the end of this subsection.

From now on we concentrate on the simplest case: the 3-manifold Y is the sphere $S^3 = \{x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1\}$ with the standard metric. Set $X := \mathbb{R} \times S^3$. The important point is that X (endowed with the product metric) is an anti-self-dual manifold with a uniformly positive scalar curvature. Here the anti-self-duality means that the Weyl conformal curvature of X is ASD. This metrical condition will be used via a certain Weitzenböck formula. Let $E = X \times SU(2)$ be the product principal $SU(2)$ bundle. All principal $SU(2)$ bundles over X are isomorphic to the product bundle E. Let A be a connection on E. Its curvature F_A is a 2-form valued in the adjoint bundle $\mathrm{ad}E = X \times su(2)$. Hence for each point $p \in X$ we can identify $(F_A)_p$ as a linear map

$$(F_A)_p : \Lambda^2(T_pX) \to su(2).$$

Let $|(F_A)_p|_{\mathrm{op}}$ be its operator norm, and set $\|F_A\|_{\mathrm{op}} := \sup_{p\in X} |(F_A)_p|_{\mathrm{op}}$.

Let d be a non-negative real number. We define $\mathcal{M}_d$ as the space of the gauge equivalence classes of ASD connections A on E satisfying

$$\|F_A\|_{\mathrm{op}} \le d. \tag{1.2}$$

This condition (1.2) means that we consider only *bounded* orbits of the evolution equation (1.1). The space $\mathcal{M}_d$ is endowed with the topology of C^∞ convergence over compact subsets: The sequence $[A_n]$ in $\mathcal{M}_d$ converges to $[A]$ if and only if there exist gauge transformations g_n satisfying $g_n(A_n) \to A$ in C^∞ over every compact subset of X. The space $\mathcal{M}_d$ is compact and metrizable by the Uhlenbeck compactness (Uhlenbeck [30], Wehrheim [31]).

We introduce a continuous action of $\mathbb{R}$ on $\mathcal{M}_d$. This corresponds to the natural time-shift $A(t) \mapsto A(t+s)$ in the evolution equation (1.1). $\mathbb{R}$ acts on $X = \mathbb{R} \times S^3$ by the shift on the $\mathbb{R}$-factor : $\mathbb{R} \times X \to X$, $(s,(t,\theta)) \mapsto (t+s,\theta)$. This lifts to the action on

$E = X \times SU(2)$ by $\mathbb{R} \times E \to E$, $(s, (t, \theta, u)) \mapsto (t + s, \theta, u)$. Then $\mathbb{R}$ acts on $\mathcal{M}_d$ by

$$\mathbb{R} \times \mathcal{M}_d \to \mathcal{M}_d, \quad (s, [A]) \mapsto [s^*(A)], \tag{1.3}$$

where $s^*(A)$ is the pull-back of A by $s : E \to E$. We study the dynamics of this action. This means that we are interested in the asymptotic behavior (as $t \to \pm\infty$) of bounded orbits of the evolution equation (1.1).

It is known that $\mathcal{M}_d$ for $d < 1$ is the one-point space consisting only of the flat connection (Tsukamoto [29]). So this is uninteresting. But when $d > 1$, $\mathcal{M}_d$ becomes an infinite dimensional and infinite topological entropy system (Matsuo–Tsukamoto [22]). So this is a relevant object of mean dimension theory. We denote the mean dimension of the action (1.3) by $\dim(\mathcal{M}_d : \mathbb{R})$. The mean dimension $\dim(\mathcal{M}_d : \mathbb{R})$ is a non-negative real number. Its rough intuitive meaning is as follows. Suppose we try to store on computer the orbits of $\mathcal{M}_d$ over the time $-T < t < T$ up to an error $\varepsilon > 0$. How many memory (/bit) do we need? It can be estimated by the mean dimension (more precisely **metric mean dimension**): We need at least

$$|\log_2 \varepsilon|\,(2T)\dim(\mathcal{M}_d : \mathbb{R}) + o(T) \quad (T \to +\infty).$$

This is one expression of a fundamental theorem of Lindenstrauss–Weiss [19]. See Theorem 2.3 and discussions around it for more precise explanations.

Our main result is the formula of the mean dimension $\dim(\mathcal{M}_d : \mathbb{R})$. Our formula involves an **energy density** $\rho(d)$ introduced by Matsuo–Tsukamoto [20]. For $[A] \in \mathcal{M}_d$ we define the energy density $\rho(A)$ by

$$\rho(A) := \lim_{T \to +\infty} \left(\frac{1}{8\pi^2 T} \sup_{t \in \mathbb{R}} \int_{(t,t+T)\times S^3} |F_A|^2 d\mathrm{vol} \right). \tag{1.4}$$

This limit always exists (Section 2.2). We denote by $\rho(d)$ the supremum of $\rho(A)$ over $[A] \in \mathcal{M}_d$. The energy density $\rho(d)$ is always non-negative and finite. It is positive for $d > 1$ and goes to infinity as $d \to +\infty$ ([22]).

The main task of the paper is to prove the upper bound estimate on the mean dimension:

Theorem 1.1.

$$\dim(\mathcal{M}_d : \mathbb{R}) \le 8\rho(d).$$

The lower bound on the mean dimension was already proved by Matsuo–Tsukamoto [22, Theorem 1.2]. Let $\mathcal{D} \subset [0, +\infty)$ be the set of left-discontinuous points of the function $\rho(d)$:

$$\mathcal{D} = \{d \in [0, +\infty) |\, \lim_{\varepsilon \to +0} \rho(d - \varepsilon) \ne \rho(d)\}.$$

This set is at most countable because $\rho(d)$ is a monotone function. From [22, Theorem 1.2] (see also Remark 1.3 below)

$$\dim(\mathcal{M}_d : \mathbb{R}) \ge 8\rho(d), \quad (d \in [0, +\infty) \setminus \mathcal{D}). \tag{1.5}$$

Therefore we get:

Corollary 1.2. *For* $d \in [0, +\infty) \setminus \mathcal{D}$*,*

$$\dim(\mathcal{M}_d : \mathbb{R}) = 8\rho(d).$$

Since $\mathcal{D}$ is at most countable, we get the formula of the mean dimension $\dim(\mathcal{M}_d : \mathbb{R})$ for almost every $d \geq 0$. This formula can be seen as a dynamical analogue of the pioneering work of Atiyah–Hitchin–Singer [1, Theorem 6.1]. Here we briefly recall their result. Let A be an irreducible ASD connection on a principal $SU(2)$ bundle P over a *compact* ASD 4-manifold M of positive scalar curvature. Atiyah–Hitchin–Singer calculated the number of the deformation parameters of A by using the Atiyah–Singer index theorem. The answer is given by

$$8c_2(P) - 3(1 - b_1(M)) \ \text{ where } c_2(P) = \frac{1}{8\pi^2}\int_M |F_A|^2 d\text{vol}.$$

Corollary 1.2 is clearly analogous to this dimension formula. The energy density (1.4) is an "averaged" second Chern number.

Remark 1.3. [22, Theorem 1.2] asserts

$$\dim_{loc}(\mathcal{M}_d : \mathbb{R}) = 8\rho(d), \quad (d \in [0, +\infty) \setminus \mathcal{D}).$$

Here $\dim_{loc}(\mathcal{M}_d : \mathbb{R})$ is the **local mean dimension** of $\mathcal{M}_d$. Local mean dimension is a variant of mean dimension, and it is always a lower bound on the original mean dimension. Therefore we get (1.5).

Corollary 1.2 is the second success of non-trivial calculation of mean dimension in geometric analysis. The first one was found by Matsuo–Tsukamoto [21, Corollary 1.2]. They proved the formula of the mean dimension of the system of Lipschitz holomorphic curves in the Riemann sphere. In the case of holomorphic curves the Nevanlinna theory provides a very simple method for obtaining the upper bound on mean dimension ([26]). So the difficult part of [21, Corollay 1.2] is the proof of the lower bound. But, in the Yang–Mills case, the upper bound (Theorem 1.1) is also difficult because we don't have a "Nevanlinna theory" for ASD equation. We need to develop a entirely new technique to obtain the upper bound, and this is the main task of the paper. The outline of the proof is explained in Section 1.3. Here we emphasize a key idea of the proof; using **metric mean dimension**. Metric mean dimension is a notion introduced by Lindenstrauss–Weiss [19]. It is a bridge between topological entropy theory and mean dimension theory. We review its definition in Section 2.1. In this paper we show that metric mean dimension is a very flexible tool for obtaining a good upper bound on mean dimension. Probably no one has expected that metric mean dimension is useful in geometric analysis. So this is the most important point of the paper. Hopefully this idea has a potential to be applied to many other problems. For example, Gromov [11, Chapter 4] studied a dynamical system consisting of complex subvarieties in $\mathbb{C}^n$. He proved an upper bound on the mean dimension [11, p. 408, Corollary]. But his estimate is very crude. So he proposed the problem of proving a better bound [11, p. 409, Remarks and open questions (a)]. It seems difficult to reach a good estimate by improving Gromov's argument directly. Metric mean dimension might shed a new light on this problem.

1.2. **Application to dynamical embedding problem.** Here we discuss one application of Theorem 1.1 in order to illustrate a dynamical importance of mean dimension. In this subsection we restrict the $\mathbb{R}$-action (1.3) to the subgroup $\mathbb{Z} \subset \mathbb{R}$, and we consider $\mathcal{M}_d$ as a space endowed with a continuous action of $\mathbb{Z}$. The mean dimension $\dim(\mathcal{M}_d : \mathbb{Z})$ of this $\mathbb{Z}$-action is equal to $\dim(\mathcal{M}_d : \mathbb{R})$. So we get (Theorem 1.1)

$$\dim(\mathcal{M}_d : \mathbb{Z}) \leq 8\rho(d).$$

Let D be a natural number, and let $([0,1]^D)^{\mathbb{Z}}$ be the $\mathbb{Z}$-shift on the D-dimensional cube (i.e. the "D-dimensional version" of the Hilbert cube). $\mathbb{Z}$ naturally acts on this space, and its mean dimension is D. The following *embedding problem* is a long-standing question in topological dynamics.

Problem 1.4. Let M be a $\mathbb{Z}$-system, i.e. a compact metric space endowed with a continuous action of $\mathbb{Z}$. Decide whether there exists a $\mathbb{Z}$-equivariant topological embedding from M into the shift $([0,1]^D)^{\mathbb{Z}}$.

This problem goes back to the Ph.D. thesis of Jaworski [15] in 1974. But here we skip the history and present only a current development. If we can equivariantly embed M into $([0,1]^D)^{\mathbb{Z}}$ then the mean dimension $\dim(M : \mathbb{Z})$ is less than or equal to D. Lindenstrauss–Tsukamoto [18] conjectured that the following partial converse holds.

Conjecture 1.5. *Let M_n $(n \geq 1)$ be the space of periodic points of period n in M. Suppose*

$$\dim(M : \mathbb{Z}) < \frac{D}{2}, \quad \frac{\dim M_n}{n} < \frac{D}{2} \quad (\forall n \geq 1).$$

Then we can embed M into $([0,1]^D)^{\mathbb{Z}}$ equivariantly.

Roughly speaking, we conjectured that mean dimension and periodic points are the only essential obstructions to the embedding. This conjecture itself is widely open, but Gutman–Tsukamoto [14] found that we can solve the problem if we sightly extend the system M by using an aperiodic symbolic subshift. Let $\{1, 2, \ldots, l\}^{\mathbb{Z}}$ be the symbolic shift, and let $Z \subset \{1, 2, \ldots, l\}^{\mathbb{Z}}$ be a subsystem without periodic points. We consider the product system $M \times Z$, which naturally admits a $\mathbb{Z}$-action and becomes an extension of the original system M. The mean dimension of $M \times Z$ is equal to the mean dimension of M. From [14, Corollary 1.8], we get:

Theorem 1.6. *If the mean dimension* $\dim(M:\mathbb{Z})$ *is strictly smaller than* $D/2$*, then we can embed the product system* $M\times Z$ *into* $([0,1]^D)^{\mathbb{Z}}$ *equivariantly.*

Here the condition $\dim(M:\mathbb{Z}) < D/2$ is known to be optimal ([14, Proposition 4.2]). By applying this theorem to $\mathcal{M}_d$, we get the following corollary.

Corollary 1.7. *Suppose* $\rho(d) < D/16$*. Then* $\mathcal{M}_d \times Z$ *can be* $\mathbb{Z}$*-equivariantly embedded into* $([0,1]^D)^{\mathbb{Z}}$*.*

This is a manifestation that the energy density $\rho(d)$ properly controls the size of $\mathcal{M}_d$. If Conjecture 1.5 is proved, then we will be able to show that $\mathcal{M}_d$ itself can be embedded into $([0,1]^D)^{\mathbb{Z}}$ under the same condition $\rho(d) < D/16$. Here it is worth to point out that we have no idea how to construct concretely the embedding given in Corollary 1.7. The above is a pure existence theorem. It is very interesting to find an explicit construction of such an embedding because it will give a new way to obtain an upper bound on the mean dimension; if $\mathcal{M}_d \times Z$ can be equivariantly embedded into $([0,1]^D)^{\mathbb{Z}}$, then we get $\dim(\mathcal{M}_d:\mathbb{Z}) \le D$.

1.3. **Ideas of the proof.** In this subsection we explain a rough strategy of the proof of Theorem 1.1. Our argument here is intuitive and non-rigorous.

The most important idea is the use of metric mean dimension as we explained in the end of Section 1.1. Metric mean dimension is always an upper bound on mean dimension (Theorem 2.3). So we want to estimate the metric mean dimension of $\mathcal{M}_d$. Intuitively this means that we estimate how many memory (/bit) we need in order to store on computer the orbits of $\mathcal{M}_d$ over the time $-T < t < T$ up to an error $\varepsilon > 0$. We want to know its asymptotics as $T \to \infty$ and $\varepsilon \to 0$. Our argument has the following three steps.

Step 1: Decomposition of $\mathcal{M}_d$. We decompose the space $\mathcal{M}_d$ into appropriately small pieces:

$$\mathcal{M}_d = U_1 \cup \cdots \cup U_n.$$

We try to memorize each U_i separately. This is an advantage of metric mean dimension over original mean dimension. Mean dimension does not behave smoothly for a decomposition of a space. Metric mean dimension is flexible for such a decomposition if we appropriately control the number n of the pieces. So we can *localize* the argument by using metric mean dimension.

Step 2: Instanton approximation. The above U_i are infinite dimensional in general. We construct their finite dimensional approximations by using the technique of **instanton approximation**. Instanton approximation is an analogue of the famous Runge theorem in complex analysis; for any meromorphic function in $\mathbb{C}$ and any compact subset $K \subset \mathbb{C}$ we can construct a rational function which approximates the given function over K. In the same spirit, for any ASD connection A on E and any compact subset $K \subset X$, we can construct an instanton (finite energy ASD connection) which approximates A over K.

Instanton approximation technique was first introduced by Taubes [24] and Donaldson [3], and it was used by Matsuo–Tsukamoto [20] in the context of mean dimension. Here we apply this technique to our present situation. For each U_i we construct a map

$$U_i \to V_i, \quad [A] \mapsto [A'],$$

such that A' is an instanton which approximates A over $-T < t < T$. We can control the energy of A' so that V_i becomes a finite dimensional space. V_i is a good approximation of U_i over $-T < t < T$. So we only need to memorize V_i instead of U_i.

Step 3: Quantitative deformation theory. We investigate V_i by constructing a deformation theory of instantons. Instanton deformation theory is a quite standard subject, but our main emphasis is on its quantitative aspect. We need to develop a deformation theory with estimates independent of several parameters (e.g. second Chern number, etc.). A key ingredient is a decomposition of $\mathbb{R}$ into "good intervals" and "bad intervals". (Indeed this decomposition will be also important in Step 1.) We fix a sufficiently small number $\nu > 0$. Take an ASD connection A on E, and let $n \in \mathbb{Z}$. If the L^∞-norm of the curvature F_A over $n < t < n+1$ is greater than or equal to ν, then we call the interval $(n, n+1)$ good. Otherwise we call it bad. If A is an instanton, then there are only finitely many good intervals. The meaning of this good/bad dichotomy is as follows. If $(n, n+1)$ is good, then for any gauge transformation g of E over $n < t < n+1$ we have

$$\min_{\pm} \|g \pm 1\|_{L^\infty((n,n+1)\times S^3)} \leq \mathrm{const}(\nu) \cdot \|d_A g\|_{L^2_{2,A}((n,n+1)\times S^3)} .$$

(See Lemma 4.2.) This means that we have a good control of gauge transformations over good intervals. If $(n, n+1)$ is bad, then A is close to a trivial flat connection (which is reducible) over $n < t < n+1$. So we lose the above control of gauge transformations there. This apparently causes a difficulty. But if A is close to a trivial flat connection, then its structure is simple. So A has little information over bad intervals. (This means that bad intervals are "not so bad".) We need to analyze these two different behaviors separately. This can be done by introducing appropriate weighted norms, and we alway have to care effects of the weight on our estimates.

Our quantitative deformation theory tells us how many memory we need in order to memorize V_i. Then we combine this with the results in the previous steps, and we can get the desired estimate on the metric mean dimension.

Organization of the paper: In Section 2.1 we explain the basic definitions of mean dimension and metric mean dimension. In Section 2.2 we prepare a lemma on the energy density $\rho(d)$. In Section 2.3 we explain some notations which are used in the rest of the paper.

In Section 3.1 we introduce weighted norms which reflect the good/bad decomposition structure. In Section 3.2 we state three main propositions (Propositions 3.2, 3.3 and 3.4) and prove Theorem 1.1 by assuming them. Propositions 3.2, 3.3 and 3.4 correspond to the above three steps respectively, and their proofs occupy the rest of the paper.

In Section 4 we prove Proposition 3.2. In Section 5 we prepare several estimates on instanton approximation and prove Proposition 3.3. In Section 6 we develop a quantitative study of instanton deformation theory in detail and prove Proposition 3.4.

Acknowledgement. I wish to thank Professor Kenji Fukaya and Professor Elon Lindenstrauss. I came up with the idea of using metric mean dimension through conversations with them.

2. Some preliminaries

2.1. Review of mean dimension.

In this subsection we review the basic facts on the mean dimension theory. For the details, see Gromov [11] and Lindenstrauss–Weiss [19].

Let (M, dist) be a compact metric space. Here dist is a distance function of M. We introduce some metric invariants of (M, dist). Let N be a topological space. For $\varepsilon > 0$, a continuous map $f : M \to N$ is called an ε-**embedding** if $\mathrm{Diam} f^{-1}(y) < \varepsilon$ for all $y \in N$. We define the ε-**width dimension** $\mathrm{Widim}_\varepsilon(M, \mathrm{dist})$ as the minimum integer $n \geq 0$ such that there exist an n-dimensional finite polyhedron P and an ε-embedding $f : M \to P$. The covering dimension $\dim M$ is obtained by

$$\dim M = \lim_{\varepsilon \to 0} \mathrm{Widim}_\varepsilon(M, \mathrm{dist}).$$

For $\varepsilon > 0$ we set

$$\#(M, \mathrm{dist}, \varepsilon) = \min\{\, |\alpha| \,|\, \alpha \text{ is an open covering of } M \text{ with } \mathrm{Diam} U < \varepsilon \text{ for all } U \in \alpha\},$$

$$\#_{\mathrm{sep}}(M, \mathrm{dist}, \varepsilon) = \max\{n \geq 1 \,|\, \exists x_1, \dots, x_n \in M \text{ with } \mathrm{dist}(x_i, x_j) > \varepsilon \ (i \neq j)\}.$$

These are almost equivalent to each other: For $0 < \delta < \varepsilon/2$

$$\#_{\mathrm{sep}}(M, \mathrm{dist}, \varepsilon) \leq \#(M, \mathrm{dist}, \varepsilon) \leq \#_{\mathrm{sep}}(M, \mathrm{dist}, \delta).$$

The next lemma will be useful.

Lemma 2.1. *Let (M, dist) and (N, dist') be metric spaces. Let $\varepsilon > 0$ and $\delta > 0$. Suppose there exists a map (not necessarily continuous) $f : M \to N$ satisfying*

$$\mathrm{dist}'(f(x), f(y)) \leq \delta \Rightarrow \mathrm{dist}(x, y) \leq \varepsilon.$$

Then $\#_{\mathrm{sep}}(M, \mathrm{dist}, \varepsilon) \leq \#_{\mathrm{sep}}(N, \mathrm{dist}', \delta)$.

Proof. Obvious. □

The following example is important. This was used by Li–Liang [16, Lemma 7.4].

Example 2.2. Let $(V, \|\cdot\|)$ be an n-dimensional Banach space over $\mathbb{R}$. Let $B_r(V)$ be the closed r-ball of V around the origin. For any $\varepsilon > 0$

$$\#_{\mathrm{sep}}(B_r(V), \|\cdot\|, \varepsilon) \leq \left(\frac{\varepsilon + 2r}{\varepsilon}\right)^n.$$

Proof. Let μ be the translation invariant measure (i.e. Haar measure) on V normalized so that $\mu(B_1(V)) = 1$. Then for any $r > 0$ we have $\mu(B_r(V)) = r^n$. Choose $\{x_1, \dots, x_N\} \subset B_r(V)$ with $\|x_i - x_j\| > \varepsilon$ for $i \neq j$. Let B_i be the closed $\varepsilon/2$-ball centered at x_i. These B_i are disjoint and their union is contained in $B_{r+\varepsilon/2}(V)$. Hence

$$N(\varepsilon/2)^n = \mu\left(\bigcup_{i=1}^N B_i\right) \leq \mu(B_{r+\varepsilon/2}(V)) = (r+\varepsilon/2)^n.$$

□

Suppose the Lie group $\mathbb{R}$ continuously acts on a compact metric space (M, dist). For a subset $\Omega \subset \mathbb{R}$ we define a new distance dist_Ω on M by

$$\mathrm{dist}_\Omega(x, y) := \sup_{t\in\Omega} \mathrm{dist}(t.x, t.y).$$

We define the mean dimension $\dim(M : \mathbb{R})$ by

$$\dim(M : \mathbb{R}) := \lim_{\varepsilon\to 0}\left(\lim_{T\to+\infty} \frac{\mathrm{Widim}_\varepsilon(M, \mathrm{dist}_{(-T,T)})}{2T}\right).$$

This is independent of the choice of a distance function dist. So the mean dimension is a topological invariant. If $\dim M < +\infty$, then the mean dimension $\dim(M : \mathbb{R})$ is zero.

Next we introduce metric mean dimension (Lindenstrauss–Weiss [19, Section 4]). For $\varepsilon > 0$ we define $S(M, \mathrm{dist}, \varepsilon)$ by

$$S(M, \mathrm{dist}, \varepsilon) = \lim_{T\to+\infty} \frac{\log \#(M, \mathrm{dist}_{(-T,T)}, \varepsilon)}{2T}.$$

This is the entropy of M "at the scale ε". The above limit always exists because of the natural subadditivity:

$$\#(M, \mathrm{dist}_{\Omega_1\cup\Omega_2}, \varepsilon) \leq \#(M, \mathrm{dist}_{\Omega_1}, \varepsilon) + \#(M, \mathrm{dist}_{\Omega_2}, \varepsilon), \quad (\Omega_1, \Omega_2 \subset \mathbb{R}).$$

The topological entropy of M is defined by $h_{\mathrm{top}}(M : \mathbb{R}) = \lim_{\varepsilon\to 0} S(M, \mathrm{dist}, \varepsilon)$. We define the metric mean dimension $\dim_{\mathrm{M}}(M, \mathrm{dist} : \mathbb{R})$ by

$$\dim_{\mathrm{M}}(M, \mathrm{dist} : \mathbb{R}) := \liminf_{\varepsilon\to 0} \frac{S(M, \mathrm{dist}, \varepsilon)}{|\log \varepsilon|}. \tag{2.1}$$

The metric mean dimension $\dim_{\mathrm{M}}(M, \mathrm{dist} : \mathbb{R})$ depends on the choice of a distance. If the topological entropy is finite, then the metric mean dimension is zero. Lindenstrauss–Weiss [19, Theorem 4.2] proved the following fundamental theorem.

Theorem 2.3. *Metric mean dimension is always an upper bound on mean dimension:*

$$\dim(M : \mathbb{R}) \leq \dim_{\mathrm{M}}(M, \mathrm{dist} : \mathbb{R}).$$

In particular if the topological entropy is finite, then the mean dimension is zero.

2.2. **Energy density.** In this subsection we prepare a lemma on the energy density $\rho(d)$ introduced in (1.4). First of all, the limit in the definition (1.4) always exists because we have the natural subadditivity:

$$\sup_{t\in\mathbb{R}}\int_{(t,t+T_1+T_2)\times S^3}|F_A|^2 d\mathrm{vol} \le \sup_{t\in\mathbb{R}}\int_{(t,t+T_1)\times S^3}|F_A|^2 d\mathrm{vol} + \sup_{t\in\mathbb{R}}\int_{(t,t+T_2)\times S^3}|F_A|^2 d\mathrm{vol}.$$

Lemma 2.4.

$$\rho(d) = \lim_{T\to+\infty}\left(\frac{1}{16\pi^2 T}\sup_{[A]\in\mathcal{M}_d}\int_{(-T,T)\times S^3}|F_A|^2 d\mathrm{vol}\right). \tag{2.2}$$

The limit of the right-hand-side exists because of the subadditivity.

Proof. This can be proved by the method of [28, Theorem 1.3]. But here we give a simpler proof based on the ergodic theorem. In this proof we restrict the $\mathbb{R}$-action (1.3) to the subgroup $\mathbb{Z}\subset\mathbb{R}$ as in Section 1.2. We denote by $\rho_1(d)$ the right-hand-side of (2.2). $\rho(d)\le\rho_1(d)$ is obvious. We define a continuous function $\varphi:\mathcal{M}_d\to\mathbb{R}$ by

$$\varphi([A]) = \frac{1}{8\pi^2}\int_{(0,1)\times S^3}|F(A)|^2 d\mathrm{vol}.$$

Then for $[A]\in\mathcal{M}_d$ and positive integers n we have the following equation:

$$\frac{1}{8\pi^2 n}\int_{(0,n)\times S^3}|F(A)|^2 d\mathrm{vol} = \frac{1}{n}\sum_{k=0}^{n-1}\varphi(k[A]).$$

Here $k[A]=[k^*A]$ is the pull-back of $[A]$ by $(t,\theta)\mapsto(t+k,\theta)$. We can choose a sequence $[A_1],[A_2],\dots$ in $\mathcal{M}_d$ so that

$$\frac{1}{n}\sum_{k=0}^{n-1}\varphi(k[A_n]) = \frac{1}{8\pi^2 n}\int_{(0,n)\times S^3}|F(A_n)|^2 d\mathrm{vol}\to\rho_1(d)\quad(n\to\infty).$$

We define a Borel probability measure μ_n on $\mathcal{M}_d$ by

$$\mu_n := \frac{1}{n}\sum_{k=0}^{n-1}\delta_{k[A_n]}$$

where $\delta_{k[A_n]}$ is the delta measure concentrated at the point $k[A_n]$. Then

$$\int_{\mathcal{M}_d}\varphi\, d\mu_n = \frac{1}{n}\sum_{k=0}^{n-1}\varphi(k[A_n])\to\rho_1(d).$$

The space of Borel probability measures is weak*-compact. So we can pick up an accumulation point μ_∞ of $\{\mu_n\}$. μ_∞ is a $\mathbb{Z}$-invariant Borel probability measure (Einsiedler–Ward [6, Theorem 4.1]) and satisfies

$$\int_{\mathcal{M}_d}\varphi\, d\mu_\infty = \rho_1(d).$$

By the ergodic decomposition [6, Theorem 4.8], we can choose an ergodic component μ of μ_∞ satisfying

$$\int_{\mathcal{M}_d} \varphi\, d\mu \geq \rho_1(d).$$

By the pointwise ergodic theorem [6, Theorem 2.30], for μ-a.e. $[A] \in \mathcal{M}_d$

$$\frac{1}{n}\sum_{k=0}^{n-1} \varphi(k[A]) \to \int_{\mathcal{M}_d} \varphi\, d\mu \geq \rho_1(d).$$

This implies $\rho(A) \geq \rho_1(d)$ for μ-a.e. $[A] \in \mathcal{M}_d$. In particular we get $\rho(d) \geq \rho_1(d)$. $\square$

2.3. **Notations.** • In most of the arguments the variable t means the natural projection $t : \mathbb{R} \times S^3 \to \mathbb{R}$.

• The value of d (which is the parameter of $\mathcal{M}_d$) is fixed in the rest of the paper. So we treat it as a constant and omit to write the dependence on d in various estimates. For two quantities x and y we write

$$x \lesssim y$$

if there exists a universal positive constant C (which might depend on d) satisfying $x \leq Cy$. We also use the following notation:

$$x \lesssim_{a,b,c,\dots,k} y$$

This means that there exists a positive constant $C(a, b, c, \dots, k)$ which depends only on parameters $a, b, c, \dots, k$ satisfying $x \leq C(a, b, c, \dots, k)y$.

• Let A be a connection on E. Let $k \geq 0$ be an integer, and let $p \geq 1$. For $\xi \in \Omega^i(\mathrm{ad}E)$ $(0 \leq i \leq 4)$ and a subset $U \subset X$, we define a norm $\|\xi\|_{L^p_{k,A}(U)}$ by

$$\|\xi\|_{L^p_{k,A}(U)} := \left(\sum_{j=0}^{k} \left\|\nabla_A^j \xi\right\|_{L^p(U)}^p\right)^{1/p}.$$

For $\alpha < \beta$ we often denote the norm $\|\xi\|_{L^p_{k,A}((\alpha,\beta)\times S^3)}$ by $\|\xi\|_{L^p_{k,A}(\alpha,\beta)}$.

3. Main propositions and the proof of Theorem 1.1

3.1. **Setting of the weighted norms.** The following lemma is a basis of our good/bad decomposition argument.

Lemma 3.1. *We can choose $\nu > 0$ so that the following statement holds. Let $T > 1$ (possibly $T = \infty$) and let A be an ASD connection on E over $(0, T) \times S^3$ satisfying $\|F_A\|_{L^\infty(0,T)} < \nu$. Then*
(1) $|F_A| \lesssim \exp(2|t - T/2| - T)$ over $1/3 < t < T - 1/3$. Moreover $\|F_A\|_{L^2(0,T)} < 1$.
(2) There exists a bundle trivialization g of E over $0 < t < T$ such that

- *g is a temporal gauge, i.e. the connection matrix $g(A)$ has no dt-component.*
- *$|\nabla^k g(A)| \lesssim_k \exp(2|t - T/2| - T)$ over $1/3 < t < T - 1/3$ for all integers $k \geq 0$.*

Proof. This can be proved in the same way as in Donaldson–Kronheimer [5, Chapter 7.3, Proposition 7.3.3] or Donaldson [4, Proposition 4.4]. But here we briefly explain how to deduce the above statement from these references.

(1) By [4, Proposition 4.4] we can find $L > 0$ and $\nu > 0$ such that if an ASD connection A over $-L < t < L$ satisfies $\|F_A\|_{L^\infty(-L,L)} < \nu$ then

$$\int_{-1<t<1} |F_A|^2 d\mathrm{vol} < \frac{1}{10}\left(\int_{-L<t<-L+1} |F_A|^2 d\mathrm{vol} + \int_{L-1<t<L} |F_A|^2 d\mathrm{vol}\right).$$

Using this estimate iteratively, we can show that the condition $\|F_A\|_{L^\infty(0,T)} < \nu \ll 1$ implies $\|F_A\|_{L^2(0,T)} \lesssim \nu$ (the implicit constant is independent of T). Then we can prove the exponential decay of the condition (1) by [5, Proposition 7.3.3].

(2) The derivatives of F_A also satisfy the same exponential decay condition. Then we can choose a bundle trivialization g of E over $\{t = T/2\}$ so that $|\nabla^k g(A)| \lesssim_k e^{-T}$. We extend it to $-T < t < T$ by the temporal gauge condition. This satisfies the required properties. □

For a real number t and a subset G of $\mathbb{Z}$ we define $|t - G|$ as the infimum of $|t - n|$ over $n \in G$. Let A be a connection on E. We set

$$G(A) = \{n \in \mathbb{Z} |\, \|F_A\|_{L^\infty(n,n+1)} \geq \nu\}.$$

Here ν is the positive constant introduced in Lemma 3.1. For a positive integer T we set $G(A, T) = G(A) \cup \{-T, T\}$. For $r > 0$ we define $U_r(A, T) \subset \mathcal{M}_d$ as the set of $[B] \in \mathcal{M}_d$ such that there exists a gauge transformation g of E over $-T < t < T$ satisfying

$$e^{|n-G(A,T)|} \|g(B) - A\|_{L^2_{10,A}(n,n+1)} \leq r \text{ for all integers } -T \leq n \leq T - 1.$$

Let A be a non-flat instanton (finite energy ASD connection) on E. Here "finite energy" means

$$\int_X |F_A|^2 d\mathrm{vol} < +\infty.$$

By [4, Theorem 4.2] the curvature F_A decays exponentially as $t \to \pm\infty$. We define $G'(A)$ as the set of integers n satisfying $\|F_A\|_{L^\infty(n,n+1)} \geq \nu/2$. This is a non-empty finite set. Fix $0 < \alpha < 1$ and we define a smooth function $W_A : \mathbb{R} \to (0, +\infty)$ as a smoothing of the function

$$\exp(\alpha|t - G'(A)|).$$

The function $\exp(\alpha|t - G'(A)|)$ has finitely many non-differentiable points. So we smooth them out. Details of the smoothing are not important. We construct W_A so that it satisfies

$$e^{\alpha|t-G'(A)|} \lesssim W_A(t) \lesssim e^{\alpha|t-G'(A)|}, \quad W_A^{(k)} \lesssim_k W_A,$$

where the implicit constants are independent of $t \in \mathbb{R}$. $W_A^{(k)}$ is the k-th derivative of W_A.

Let $G'(A) = \{n_1 < n_2 < \cdots < n_G\}$, and set $n_0 = -\infty$ and $n_{G+1} = +\infty$. For $u \in \Omega^i(\mathrm{ad}E)$ and $k \geq 0$ we define a norm

$$\|u\|_{k,A} = \max_{0\leq j\leq G} \|W_A u\|_{L^2_{k,A}(n_j,n_{j+1})}. \tag{3.1}$$

For $r > 0$ we define $V_r(A)$ as the set of gauge equivalence classes of ASD connections B on E such that there exists a gauge transformation g of E satisfying

$$\|g(B) - A\|_{2,A} \leq r.$$

3.2. **Main propositions and the proof of Theorem 1.1.**

Proposition 3.2. *For any $\delta > 0$ and any integer $T > 1$ there exist $[A_1], \dots, [A_n] \in \mathcal{M}_d$ satisfying*

$$\log n \lesssim_\delta T, \quad \mathcal{M}_d = \bigcup_{i=1}^n U_\delta(A_i, T).$$

Proposition 3.3. *For any $r > 0$ we can choose $\delta_0 = \delta_0(r) > 0$ satisfying the following statement. For any $[A] \in \mathcal{M}_d$ and any integer $T > 1$ there exists a non-flat instanton A' on E and a map*

$$U_{\delta_0}(A,T) \to V_r(A'), \quad [B] \mapsto [B']$$

such that
(1)

$$\|F_{A'}\|_{L^\infty(X)} \leq D_0, \quad \left| \int_X |F_{A'}|^2 d\mathrm{vol} - \int_{(-T,T)\times S^3} |F_A|^2 d\mathrm{vol} \right| \lesssim 1.$$

Here D_0 is a universal constant independent of r.
(2) For any $[B] \in U_{\delta_0}(A,T)$ there exists a gauge transformation g of E over $|t| < T-1$ satisfying

$$|g(B') - B| \lesssim e^{-\sqrt{2}|t-T|} + e^{-\sqrt{2}|t+T|} \quad (|t| < T-1).$$

For two connections A_1 and A_2 on E, we set

$$\mathrm{dist}_{L^\infty}([A_1],[A_2]) = \inf_{g:E\to E} \|g(A_1) - A_2\|_{L^\infty(X)},$$

where g runs over all gauge transformations of E.

Proposition 3.4. *For any $D > 0$ there exist positive numbers $r_0 = r_0(D)$ and $C_0 = C_0(D)$ satisfying the following statement. Let A be a non-flat instanton on E with $\|F_A\|_{L^\infty(X)} \leq D$. Then for any $0 < \varepsilon < 1$*

$$\#_{\mathrm{sep}}(V_{r_0}(A), \mathrm{dist}_{L^\infty}, \varepsilon) \leq (C_0/\varepsilon)^{8c_2(A)+3},$$

where

$$c_2(A) = \frac{1}{8\pi^2}\int_X |F_A|^2 d\mathrm{vol}.$$

The proofs of the above three propositions occupy the rest of the paper. Here we prove Theorem 1.1, assuming them.

Proof of Theorem 1.1. We define a distance on $\mathcal{M}_d$ by

$$\mathrm{dist}([A],[B]) = \inf_{g:E\to E} \|g(A) - B\|_{L^\infty(0,1)},$$

where g runs over all gauge transformations of E. This is compatible with the given topology of $\mathcal{M}_d$. Recall that for a subset $\Omega \subset \mathbb{R}$ we denote by $\mathrm{dist}_\Omega([A],[B])$ the supremum of $\mathrm{dist}([s^*A],[s^*B])$ over $s \in \Omega$. We will prove the upper bound on the metric mean dimension: $\dim_{\mathrm{M}}(\mathcal{M}_d, \mathrm{dist} : \mathbb{R}) \le 8\rho(d)$. Then we get $\dim(\mathcal{M}_d : \mathbb{R}) \le 8\rho(d)$ since the metric mean dimension is an upper bound on the mean dimension (Theorem 2.3).

Let $D_0 > 0$ be the universal constant introduced in Proposition 3.3 (1), and let $r_0 = r_0(D_0)$ be the positive constant introduced in Proposition 3.4 with respect to D_0. Moreover let $\delta_0 = \delta_0(r_0(D_0))$ be the positive constant introduced in Proposition 3.3 with respect to $r_0(D_0)$.

Claim 3.5. *There exists $C_1 > 0$ satisfying the following statement. For any $0 < \varepsilon < 1$ there exists an integer $L_0 = L_0(\varepsilon) > 1$ such that for any integer $T > L_0$ and any $[A] \in \mathcal{M}_d$ we have*

$$\log \#_{\mathrm{sep}}(U_{\delta_0}(A,T), \mathrm{dist}_{(-T+L_0,T-L_0)}, \varepsilon) \le (|\log\varepsilon| + C_1)\left(\frac{1}{\pi^2}\int_{(-T,T)\times S^3} |F_A|^2 d\mathrm{vol} + C_1\right).$$

Proof. By Proposition 3.3 for any $[A] \in \mathcal{M}_d$ and any integer $T > 1$ there exist a non-flat instanton $[A']$ and a map

$$U_{\delta_0}(A,T) \to V_{r_0}(A'), \quad [B] \mapsto [B']$$

satisfying the conditions (1) and (2) of the statement there. If we choose $L_0 = L_0(\varepsilon) > 0$ sufficiently large, then (by the condition (2)) for any $[B] \in U_{\delta_0}(A,T)$ there exists a gauge transformation g of E over $|t| < T-1$ satisfying

$$|g(B') - B| < \varepsilon/3 \quad (|t| < T - L_0 + 1).$$

Then for any $[B_1],[B_2] \in U_{\delta_0}(A,T)$ with $T > L_0$ we get

$$\mathrm{dist}_{L^\infty}([B_1'],[B_2']) \le \varepsilon/3 \Longrightarrow \mathrm{dist}_{(-T+L_0,T-L_0)}([B_1],[B_2]) \le \varepsilon.$$

By Lemma 2.1

$$\begin{aligned}\#_{\mathrm{sep}}(U_{\delta_0}(A,T), \mathrm{dist}_{(-T+L_0,T-L_0)}, \varepsilon) &\le \#_{\mathrm{sep}}(V_{r_0}(A'), \mathrm{dist}_{L^\infty}, \varepsilon/3) \\ &\le (3C_0/\varepsilon)^{8c_2(A')+3} \quad \text{(by Proposition 3.4).}\end{aligned}$$

By the condition (1) of Proposition 3.3

$$8c_2(A') + 3 \le \frac{1}{\pi^2}\int_{(-T,T)\times S^3} |F_A|^2 d\mathrm{vol} + \mathrm{const},$$

where const is a universal constant. Thus we get the conclusion. □

Take $0 < \varepsilon < 1$ and let $L_0 = L_0(\varepsilon) > 0$ be the positive number introduced in the above claim. By Proposition 3.2 for any integer $T > 1$ there exist $[A_1], \dots, [A_n] \in \mathcal{M}_d$ satisfying

$$\log n \lesssim T + L_0, \quad \mathcal{M}_d = \bigcup_{i=1}^n U_{\delta_0}(A_i, T+L_0).$$

Then $\#(\mathcal{M}_d, \mathrm{dist}_{(-T,T)}, \varepsilon)$ is bounded by

$$\sum_{i=1}^n \#(U_{\delta_0}(A_i, T+L_0), \mathrm{dist}_{(-T,T)}, \varepsilon) \le \sum_{i=1}^n \#_{\mathrm{sep}}(U_{\delta_0}(A_i, T+L_0), \mathrm{dist}_{(-T,T)}, \varepsilon/3).$$

By Claim 3.5, $\log \#(\mathcal{M}_d, \mathrm{dist}_{(-T,T)}, \varepsilon)$ is bounded by

$$\log n + (|\log \varepsilon| + \log 3 + C_1)\left(\frac{1}{\pi^2} \sup_{[A]\in\mathcal{M}_d} \int_{(-T-L_0, T+L_0)\times S^3} |F_A|^2 d\mathrm{vol} + C_1\right).$$

Since $\log n \lesssim T + L_0$ and L_0 does not depend on T, we get (by using Lemma 2.4)

$$S(\mathcal{M}_d, \mathrm{dist}, \varepsilon) = \lim_{T\to\infty} \frac{\log \#(\mathcal{M}_d, \mathrm{dist}_{(-T,T)}, \varepsilon)}{2T} \le \mathrm{const} + (|\log\varepsilon| + \log 3 + C_1) 8\rho(d).$$

Here const and C_1 are independent of ε. Thus

$$\dim_{\mathrm{M}}(\mathcal{M}_d, \mathrm{dist} : \mathbb{R}) = \liminf_{\varepsilon\to 0} \frac{S(\mathcal{M}_d, \mathrm{dist}, \varepsilon)}{|\log \varepsilon|} \le 8\rho(d).$$

□

4. Decomposition of $\mathcal{M}_d$: proof of Proposition 3.2

We prove Proposition 3.2 in this section. A theme of this section is a problem of gluing gauge transformations. A simplified situation is the following: Let $[A], [B] \in \mathcal{M}_d$. Let $U_1, U_2 \subset X$ be open sets, and let g_i be gauge transformations of E over U_i ($i = 1, 2$). Suppose $|g_i(B) - A|$ are very small over U_i for both $i = 1, 2$. Can we find a gauge transformation h of E over $U_1 \cup U_2$ satisfying $|h(B) - A| \ll 1$? Unfortunately the answer is *No* in general. If A and B are very close to flat connections over $U_1 \cap U_2$, then we have to consider a *gluing parameter* over $U_1 \cap U_2$ and cannot find such a gauge transformation h. (This phenomena appears in constructions of gluing instantons. See [5, Chapter 7.2].) In Lemmas 4.4 and 4.5 below we formulate situations where the answer to the above question becomes *Yes*.

The following is a basis of the argument. This is proved in [20, Corollary 6.3].

Lemma 4.1. *Let A be a non-flat ASD connection on E with $\|F_A\|_{L^\infty} < \infty$. Then A is irreducible, i.e. if a gauge transformation g satisfies $g(A) = A$ then $g = \pm 1$.*

Proof. We give a sketch of the proof for the convenience of readers. Suppose A is reducible. Then A is reduced to a $U(1)$ connection. In particular F_A is a $u(1)$-valued anti-self-dual

2-form. Using the Yang–Mills equation $d_A^* F_A = 0$ and the Weitzenböck formula (see (5.2) in Section 5), we get

$$(\nabla^*\nabla + 2)F_A = 0.$$

Here we have used the fact that the curvature F_A does not contribute to the formula because it is $u(1)$-valued. Then $\|F_A\|_{L^\infty} < \infty$ implies $F_A = 0$ all over X. See discussions around (5.3). □

The next lemma means that we have a good control of gauge transformations over "good intervals".

Lemma 4.2. *Let $\kappa > 0$ and let $[A] \in \mathcal{M}_d$ with $\|F_A\|_{L^\infty(0,1)} \geq \kappa$. For any gauge transformation g of E over $0 < t < 1$ we have*

$$\min_{\pm} \|g \pm 1\|_{L^\infty(0,1)} \lesssim_\kappa \|d_A g\|_{L^2_{2,A}(0,1)}.$$

Proof. It is standard that we can deduce this kind of statement from the following linearized one. (For the detail, see [22, Lemma 3.2].)

Claim 4.3. *Let u be a section of* ad$E$ *over $0 < t < 1$. Then*

$$\|u\|_{L^\infty(0,1)} \lesssim_\kappa \|d_A u\|_{L^2_{2,A}(0,1)}.$$

Proof. Suppose the contrary. Then there exist $[A_n] \in \mathcal{M}_d$ with $\|F_{A_n}\|_{L^\infty(0,1)} \geq \kappa$ and $u_n \in \Gamma((0,1) \times S^3, \mathrm{ad}E)$ $(n \geq 1)$ satisfying

$$\|d_{A_n} u_n\|_{L^2_{2,A_n}(0,1)} < \frac{1}{n}, \quad \|u_n\|_{L^\infty(0,1)} = 1.$$

Since $\mathcal{M}_d$ is compact, we can assume that A_n converges to some A with $\|F_A\|_{L^\infty(0,1)} \geq \kappa$ in C^∞ over every compact subset. Then $\{u_n\}$ is bounded in $L^2_{3,A}((0,1) \times S^3)$. By choosing a subsequence, we can assume that u_n weakly converges to some u in $L^2_{3,A}((0,1) \times S^3)$ with $d_A u = 0$. We have $\|u\|_{L^\infty(0,1)} = 1$ because the Sobolev embedding $L^2_{3,A}((0,1) \times S^3) \to L^\infty((0,1) \times S^3)$ is compact. This means that A is reducible over $0 < t < 1$. By the unique continuation theorem (Donaldson–Kronheimer [5, Chapter 4, Lemma 4.3.21]) A is reducible all over X. This contradicts Lemma 4.1. □

□

In the next two lemmas we formulate situations where we can glue two gauge transformations. In the first lemma, an overlapping region is "good". The argument is straightforward. In the second lemma, an overlapping region is "bad". Our formulation have to be more involved.

Lemma 4.4. *For any $\kappa, \delta > 0$ we can choose $\varepsilon_1 = \varepsilon_1(\kappa, \delta) > 0$ so that the following statement holds. Let $[A], [B] \in \mathcal{M}_d$, and let g_1 and g_2 be gauge transformations of E over $0 < t < 2$ and $1 < t < 3$ respectively. Suppose*

$$\|F_A\|_{L^\infty(1,2)} \geq \kappa, \quad \|g_i(B) - A\|_{L^2_{10,A}(1,2)} < \varepsilon_1 \ (i = 1, 2).$$

Then there exists a gauge transformation h of E over $0 < t < 3$ such that $h = g_1$ over $0 < t < 1$, $h = \pm g_2$ over $2 < t < 3$ and

$$\|h(B) - A\|_{L^2_{10,A}(1,2)} < \delta.$$

Proof. Set $w = g_2 g_1^{-1}$ over $1 < t < 2$. We have $d_A w = w \cdot (g_1(B) - A) + (A - g_2(B)) \cdot w$. Hence $\|d_A w\|_{L^2_{10,A}(1,2)} \lesssim \varepsilon_1$. By Lemma 4.2 we get $\min_\pm \|w \pm 1\|_{L^\infty(1,2)} \lesssim_\kappa \varepsilon_1$. We can assume $\|w - 1\|_{L^\infty(1,2)} \leq \|w + 1\|_{L^\infty(1,2)}$. Then $\|w - 1\|_{L^\infty(1,2)} \lesssim_\kappa \varepsilon_1 \ll 1$. Thus w is expressed as $w = e^u$ with $\|u\|_{L^2_{11,A}(1,2)} \lesssim_\kappa \varepsilon_1$. Take a cut-off $\varphi : \mathbb{R} \to [0,1]$ such that $\mathrm{supp}(d\varphi) \subset (0,1)$, $\varphi(0) = 0$ and $\varphi(1) = 1$. We set $h = e^{\varphi u} g_1$. If we choose ε_1 sufficiently small, then this satisfies the statement. □

In the rest of this section we take and fix a point $\theta_0 \in S^3$. Recall that we introduced the positive constant ν in Lemma 3.1.

Lemma 4.5. *For any $\delta > 0$ we can choose positive numbers $\varepsilon_2 = \varepsilon_2(\delta)$ and $L_1 = L_1(\delta)$ so that the following statement holds. Take $[A], [B] \in \mathcal{M}_d$, an integer $T \geq 2L_1$ and gauge transformations g_1 and g_2 over $-1 < t < L_1$ and $T - L_1 < t < T + 1$ respectively. Suppose the following three conditions.*

- $\|F_A\|_{L^\infty(0,T)}, \|F_B\|_{L^\infty(0,T)} < \nu$, $\|F_A\|_{L^\infty(-1,0)}, \|F_A\|_{L^\infty(T,T+1)} \geq \nu$.
- $\|g_1(B) - A\|_{L^2_{10,A}(0,L_1)}, \|g_2(B) - A\|_{L^2_{10,A}(T-L_1,T)} < \varepsilon_2$.
- *Set $p = (L_1 - 1, \theta_0), q = (T - L_1 + 1, \theta_0) \in X$ and define $g_2'(q) : E_p \to E_p$ by the following commutative diagram:*

$$\begin{array}{ccc} E_p & \xrightarrow{\text{parallel translation by } B} & E_q \\ \Big\downarrow {\scriptstyle g_2'(q)} & & \Big\downarrow {\scriptstyle g_2(q)} \\ E_p & \xrightarrow{\text{parallel translation by } A} & E_q \end{array}$$

  Here the horizontal arrows are the parallel translations by B and A along the minimum geodesic between p and q. Under these settings, we have

$$\min_\pm \mathrm{dist}_{SU(2)}(g_1(p), \pm g_2'(q)) < \varepsilon_2,$$

  where $\mathrm{dist}_{SU(2)}$ is the distance on $SU(2)$ defined by the standard Riemannian structure.

Then there exists a gauge transformation h of E over $-1 < t < T + 1$ such that $h = g_1$ over $-1 < t < 0$, $h = \pm g_2$ over $T < t < T + 1$ and

$$e^{\min(n+1, T-n)} \|h(B) - A\|_{L^2_{10,A}(n,n+1)} < \delta$$

for all integers $0 \leq n \leq T - 1$.

Proof. Let g_A and g_B be the temporal gauges of A and B over $0 < t < T$ introduced in Lemma 3.1. The connection matrices $A' := g_A(A)$ and $B' := g_B(B)$ satisfy

$|\nabla^k A'|, |\nabla^k B'| \lesssim_k \exp(2|t-T/2|-T)$. Set $w_1 = g_A \circ g_1 \circ g_B^{-1}$ over $0 < t < L_1$ and $w_2 = g_A \circ g_2 \circ g_B^{-1}$ over $T - L_1 < t < T$. They satisfy $\|w_1(B') - A'\|_{L^2_{10,A'}(0,L_1)} < \varepsilon_2$ and $\|w_2(B') - A'\|_{L^2_{10,A'}(T-L_1,T)} < \varepsilon_2$. Moreover we can assume $\mathrm{dist}_{SU(2)}(w_1(p), w_2(q)) < \varepsilon_2$. Here we regard w_1 and w_2 as $SU(2)$-valued functions over $0 < t < L_1$ and $T - L_1 < t < T$ respectively.

We get $|dw_1| \lesssim \varepsilon_2 + e^{-2L_1}$ and $|dw_2| \lesssim \varepsilon_2 + e^{-2L_1}$ over $L_1 - 2 < t < L_1$ and $T - L_1 < t < T - L_1 + 2$ respectively. Then w_1 and w_2 are expressed as $w_1 = w_1(p)e^{u_1}$ over $L_1 - 2 < t < L_1$ and $w_2 = w_2(q)e^{u_2}$ over $T - L_1 < t < T - L_1 + 2$ such that

$$\|u_1\|_{L^2_{11}(L_1-2,L_1)} \lesssim \varepsilon_2 + e^{-2L_1}, \quad \|u_2\|_{L^2_{11}(T-L_1,T-L_1+2)} \lesssim \varepsilon_2 + e^{-2L_1}.$$

We take a path $v : \mathbb{R} \to SU(2)$ such that $v(t) = w_1(p)$ for $t \le L_1 - 1$, $v(t) = w_2(q)$ for $t \ge L_1$ and $|\nabla^k v| \lesssim_k \varepsilon_2$. We also take a cut-off $\varphi : \mathbb{R} \to [0,1]$ so that $\mathrm{supp}(d\varphi) \subset (L_1 - 2, L_1 - 1) \cup (T - L_1 + 1, T - L_1 + 2)$, $\varphi(t) = 1$ over $\{t \le L_1 - 2\} \cup \{t \ge T - L_1 + 2\}$ and $\varphi = 0$ over $L_1 - 1 \le t \le T - L_1 + 1$. We define a gauge transformation h of E over $-1 < t < T + 1$ by

$$h = \begin{cases} g_A^{-1} \circ (ve^{\varphi u_1}) \circ g_B & (t \le T/2), \\ g_A^{-1} \circ (ve^{\varphi u_2}) \circ g_B & (t > T/2). \end{cases}$$

Then $|\nabla_A^k(h(B) - A)| \lesssim_k \exp(2|t-T/2|-T)$ over $L_1 < t < T - L_1$, $\|h(B) - A\|_{L^2_{10,A}(0,L_1)} \lesssim \varepsilon_2 + e^{-2L_1}$ and $\|h(B) - A\|_{L^2_{10,A}(T-L_1,T)} \lesssim \varepsilon_2 + e^{-2L_1}$. We can choose L_1 and ε_2 so that h satisfies the statement. □

Using Lemmas 4.4 and 4.5, we can provide a sufficient condition for a given connection $[B]$ to be contained in $U_\delta(A, T)$:

Lemma 4.6. *For any $\delta > 0$ we can choose $\varepsilon_3 = \varepsilon_3(\delta) > 0$ and an integer $R_1 = R_1(\delta) > L_1(\delta)$ ($L_1(\delta)$ is the constant introduced in Lemma 4.5) so that the following statement holds. Take $[A], [B] \in \mathcal{M}_d$ and an integer $T > 1$. If they satisfy the following two conditions, then $[B] \in U_\delta(A, T)$.*

- *$G(A) \cap [-T - R_1, T + R_1] = G(B) \cap [-T - R_1, T + R_1]$. Let $n_1 < n_2 < \cdots < n_G$ be the elements of this set, and we set $p_k = (n_k + L_1, \theta_0)$ and $q_k = (n_k - L_1 + 1, \theta_0)$ for $1 \le k \le G$.*
- *For each $1 \le k \le G$ there exists a gauge transformation g_k of E over $n_k - R_1 < t < n_k + R_1$ satisfying*

$$\|g_k(B) - A\|_{L^2_{10,A}(n_k - R_1, n_k + R_1)} < \varepsilon_3 \quad (1 \le k \le G),$$
$$\min_{\pm} \mathrm{dist}_{SU(2)}(g_k(p_k), \pm g'_{k+1}(q_{k+1})) < \varepsilon_3 \quad (1 \le k < G).$$

Here $g'_{k+1}(q_{k+1})$ is defined by the following commutative diagram.

$$\begin{array}{ccc} E_{p_k} & \xrightarrow{\text{parallel translation by } B} & E_{q_{k+1}} \\ \Big\downarrow g'_{k+1}(q_{k+1}) & & \Big\downarrow g_{k+1}(q_{k+1}) \\ E_{p_k} & \xrightarrow{\text{parallel translation by } A} & E_{q_{k+1}} \end{array}$$

Proof. First let's consider the case $G(A)\cap[-T-R_1,T+R_1]=G(B)\cap[-T-R_1,T+R_1]=\emptyset$. By Lemma 3.1 we can choose trivializations g_A and g_B of E over $-T-R_1<t<T+R_1$ such that the connection matrices $g_A(A)$ and $g_B(B)$ satisfy

$$|\nabla^k g_A(A)|, |\nabla^k g_B(B)| \lesssim_k e^{2(|t|-T-R_1)} \quad (|t|<T+R_1-1).$$

Then $h := g_A^{-1}\circ g_B$ satisfies (if $R_1 \gg 1$)

$$e^{|n-G(A,T)|}\,\|h(B)-A\|_{L^2_{10,A}(n,n+1)} \le e^{|n-\{\pm T\}|}\,\|h(B)-A\|_{L^2_{10,A}(n,n+1)} < \delta$$

for all $-T\le n\le T-1$. Hence $[B]\in U_\delta(A,T)$.

Next suppose $G(A)\cap[-T-R_1,T+R_1]\neq\emptyset$. From the compactness of $\mathcal{M}_d$ we can find $\kappa>0$ so that if $[C]\in\mathcal{M}_d$ satisfies $\|F_C\|_{L^\infty(0,1)}\ge\nu$ then $\|F_C\|_{L^\infty(n,n+1)}\ge\kappa$ for all integers $|n|\le L_1+1$. Let $\varepsilon_1=\varepsilon_1(\kappa,\delta e^{-L_1-1})$ and $\varepsilon_2=\varepsilon_2(\delta)$ be the positive constants introduced in Lemmas 4.4 and 4.5. We take $\varepsilon_3>0$ and $R_1>0$ so that

$$\varepsilon_3<\min(\varepsilon_1,\varepsilon_2),\quad R_1>L_1+2,\quad \varepsilon_3 e^{R_1}<\delta.$$

We inductively define gauge transformations h_k of E over $n_1-R_1<t<n_k+R_1$ for $k=1,2,\dots,G$ so that the following two conditions hold:

- $h_k=g_1$ over $n_1-R_1<t<n_1$ and $h_k=\pm g_k$ over $n_k<t<n_k+R_1$.
- $e^{|n-G(A,T)|}\,\|h_k(B)-A\|_{L^2_{10,A}(n,n+1)}<\delta$ for all integers $n_1\le n<n_k$.

$h_1:=g_1$ obviously satisfies the conditions. Suppose we have constructed h_k $(k<G)$.

Case 1. Suppose $n_{k+1}-n_k-1<2L_1$. Set $m=\lfloor\frac{n_k+n_{k+1}}{2}\rfloor$. From the definition of κ we have $\|F_A\|_{L^\infty(m,m+1)}\ge\kappa$. We also have $\|h_k(B)-A\|_{L^2_{10,A}(m,m+1)}, \|g_{k+1}(B)-A\|_{L^2_{10,A}(m,m+1)}<\varepsilon_3<\varepsilon_1$. Then we can glue h_k and g_{k+1} over $m<t<m+1$ by Lemma 4.4 and get h_{k+1}. This satisfies the required conditions.

Case 2. Suppose $n_{k+1}-n_k-1\ge 2L_1$. Then we can apply Lemma 4.5. We glue h_k and g_{k+1} over $n_k+1<t<n_{k+1}$ and get h_{k+1}.

Therefore we get h_G over $n_1-R_1<t<n_G+R_1$. If $(-T,T)\subset(n_1-R_1,n_G+R_1)$, then it satisfies

$$e^{|n-G(A,T)|}\,\|h_G(B)-A\|_{L^2_{10,A}(n,n+1)}<\delta \tag{4.1}$$

for all integers $-T\le n<T$. Hence $[B]\in U_\delta(A,T)$.

So the remaining case is $(-T,T)\not\subset(n_1-R_1,n_G+R_1)$. Suppose $-T<n_1-R_1$. Then $\|F_A\|_{L^\infty(-T-R_1,n_1)}<\nu$ and $\|F_B\|_{L^\infty(-T-R_1,n_1)}<\nu$. Hence by Lemma 3.1 there are trivializations g_A and g_B of E over $-T-R_1<t<n_1$ such that the connection matrices

$g_A(A)$ and $g_B(B)$ satisfy appropriate exponential decay conditions. We glue $g_A^{-1} \circ g_B$ to h_G as in the proof of Lemma 4.5. In the case of $T > n_G + R_1$, we proceed in the same way over $n_G + 1 < t < T + R_1$. Then we get a gauge transformation h of E over $-T < t < T$ satisfying (4.1) for all integers $-T \le n < T$. Thus $[B] \in U_\delta(A, T)$ □

By using Lemma 4.6 we prove Proposition 3.2. We write the statement again for the convenience of readers.

Proposition 4.7 (= Proposition 3.2)**.** *For any $\delta > 0$ and any integer $T > 1$ there exist $[A_1], \dots, [A_n] \in \mathcal{M}_d$ satisfying*

$$\log n \lesssim_\delta T, \quad \mathcal{M}_d = \bigcup_{i=1}^{n} U_\delta(A_i, T).$$

Proof. Let $\varepsilon_3 = \varepsilon_3(\delta)$ and $R_1 = R_1(\delta)$ be the positive constants introduced in Lemma 4.6. Let $\varepsilon = \varepsilon(\delta) < \varepsilon_3$ be a small positive number which will be fixed later. For each subset $\Omega \subset \mathbb{Z} \cap [-T - R_1, T + R_1]$ we define

$$\mathcal{M}_d^\Omega = \{[A] \in \mathcal{M}_d |\, G(A) \cap [-T - R_1, T + R_1] = \Omega\}.$$

$\mathcal{M}_d$ is decomposed into these $\mathcal{M}_d^\Omega$, and the number of the choices of $\Omega \subset \mathbb{Z} \cap [-T - R_1, T + R_1]$ is equal to $2^{2(T+R_1)+1} \lesssim_\delta 4^T$.

We choose an open cover α of $\mathcal{M}_d$ such that if $[A], [B] \in \mathcal{M}_d$ is contained in the same open set $U \in \alpha$ then there exists a gauge transformation g of E over $-R_1 < t < R_1$ satisfying

$$\|g(B) - A\|_{L^2_{10,A}(-R_1,R_1)} < \varepsilon.$$

Note that the choice of α depends on δ and ε.

Take $\Omega = \{n_1 < n_2 < \cdots < n_G\} \subset \mathbb{Z} \cap [-T - R_1, T + R_1]$. We define an open covering $\mathcal{U}$ of $\mathcal{M}_d$ by

$$\mathcal{U} = \bigvee_{k=1}^{G} (-n_k) \cdot \alpha.$$

Here $(-n_k)\cdot\alpha$ is the translation of α by $(-n_k)$, and $\mathcal{U}$ is the set of open subsets $U_1 \cap \cdots \cap U_G$ ($U_k \in (-n_k) \cdot \alpha$). The cardinality of $\mathcal{U}$ is bounded by $|\alpha|^G \le |\alpha|^{2T+2R_1+1}$.

We choose $V \in \mathcal{U}$ and consider $\mathcal{M}_d^\Omega \cap V$. Let $\mathcal{A}$ be the set of connections A on E satisfying $[A] \in \mathcal{M}_d^\Omega \cap V$. Take and fix one $A_0 \in \mathcal{A}$. For every $A \in \mathcal{A}$ and $1 \le k \le G$ there exists a gauge transformation $g_{A,k}$ over $n_k - R_1 < t < n_k + R_1$ satisfying

$$\|g_{A,k}(A_0) - A\|_{L^2_{10,A}(n_k-R_1,n_k+R_1)} < \varepsilon.$$

Let $L_1 = L_1(\delta) > 0$ be the positive constant introduced in Lemma 4.5, and set $p_k = (n_k + L_1, \theta_0)$ and $q_k = (n_k - L_1 + 1, \theta_0)$ for $1 \le k \le G$. We consider the map:

$$\mathcal{A} \to SU(2)^{G-1}, \quad A \mapsto (g_{A,k}(p_k)^{-1} g'_{A,k+1}(q_{k+1}))_{k=1}^{G-1}.$$

Here $g'_{A,k+1}(q_{k+1})$ is defined by the commutative diagram:

$$\begin{array}{ccc} E_{p_k} & \xrightarrow{\text{parallel translation by } A_0} & E_{q_{k+1}} \\ \Big\downarrow g'_{A,k+1}(q_{k+1}) & & \Big\downarrow g_{A,k+1}(q_{k+1}) \\ E_{p_k} & \xrightarrow{\text{parallel translation by } A} & E_{q_{k+1}} \end{array}$$

Considering a covering of $SU(2)$ by ε-balls, we can construct a decomposition $\mathcal{A} = \mathcal{A}_1 \cup \cdots \cup \mathcal{A}_N$ such that

- $\log N \lesssim_\varepsilon G \lesssim_\delta T$.
- If $A, B \in \mathcal{A}$ is contained in the same $\mathcal{A}_i$ then

$$\mathrm{dist}_{SU(2)}(g_{A,k}(p_k)^{-1} g'_{A,k+1}(q_{k+1}),\ g_{B,k}(p_k)^{-1} g'_{B,k+1}(q_{k+1})) < \varepsilon \quad (\forall 1 \le k \le G).$$

Claim 4.8. *For any* $1 \le i \le G$ *and* $A, B \in \mathcal{A}_i$ *we get* $[B] \in U_\delta(A, T)$.

Proof. We check the conditions of Lemma 4.6. The condition $G(A) \cap [-T - R_1, T + R_1] = G(B) \cap [-T - R_1, T + R_1]$ is satisfied. For each $1 \le k \le G$ we set $g_k = g_{A,k} \circ g_{B,k}^{-1}$ over $n_k - R_1 < t < n_k + R_1$. They satisfy

$$\mathrm{dist}_{SU(2)}(g_k(p_k), g'_{k+1}(q_{k+1})) < \varepsilon < \varepsilon_3 \quad (\forall 1 \le k \le G - 1).$$

We have

$$g_k(B) - A = g_k(B - g_{B,k}(A_0)) + g_{A,k}(A_0) - A.$$

Hence we can choose $\varepsilon = \varepsilon(\delta) > 0$ so small that

$$\|g_k(B) - A\|_{L^2_{10,A}(n_k - R_1, n_k + R_1)} < \varepsilon_3.$$

Then we can apply Lemma 4.6 to A and B, and we get $[B] \in U_\delta(A, T)$. □

Pick up $A_1 \in \mathcal{A}_1, \ldots, A_N \in \mathcal{A}_N$. Then by the above claim

$$\mathcal{M}_d^\Omega \cap V \subset U_\delta(A_1, T) \cup \cdots \cup U_\delta(A_N, T).$$

We have the following bounds on several parameters: $\log N \lesssim_\delta T$. The number of the choices of $V \in \mathcal{U}$ is $\lesssim_\delta |\alpha|^{2T}$. Note that $|\alpha|$ is now a constant depending only on δ. The number of the choices of $\Omega \subset \mathbb{Z} \cap [-T - R_1, T + R_1]$ is $\lesssim_\delta 4^T$. Combining these estimates, we get the conclusion. □

5. Instanton approximation: Proof of Proposition 3.3

We develop instanton approximation technique and prove Proposition 3.3 in this section. First we prepare some facts concerning a Green kernel function. Let $\Delta = \nabla^*\nabla$ be the Laplacian on functions in X. Our sign convention of Δ is geometric ($\Delta = -\partial^2/\partial x_1^2 - \partial^2/\partial x_2^2 - \partial^2/\partial x_3^2 - \partial^2/\partial x_4^2$ over $\mathbb{R}^4$). Let $g(x, y)$ be the Green kernel of $\Delta + 2$ over X. This satisfies

$$(\Delta_y + 2) g(x, y) = \delta_x(y)$$

in the distributional sense, i.e. for any compactly supported smooth function φ over X

$$\varphi(x) = \int_X g(x,y)(\Delta_y + 2)\varphi(y) d\mathrm{vol}(y).$$

$g(x,y)$ is positive everywhere. It is smooth outside the diagonal, and its singularity along the diagonal is $\mathrm{dist}(x,y)^{-2}$:

$$\mathrm{dist}(x,y)^{-2} \lesssim g(x,y) \lesssim \mathrm{dist}(x,y)^{-2} \quad (\mathrm{dist}(x,y) \le 1).$$

It decays exponentially in a long range:

$$g(x,y) \lesssim e^{-\sqrt{2}\,\mathrm{dist}(x,y)} \quad (\mathrm{dist}(x,y) > 1). \tag{5.1}$$

A detailed construction of $g(x,y)$ is explained in [20, Appendix].

For $u \in \Omega^i(\mathrm{ad}E)$ we define its **Taubes norm** $\|u\|_{\mathrm{Tau}}$ by

$$\|u\|_{\mathrm{Tau}} = \sup_{x\in X} \int_X g(x,y)|u(y)| d\mathrm{vol}(y).$$

This was introduced by Taubes [24] and Donaldson [3]. An importance of this norm is linked to the following Weitzenböck formula. Let A be a connection on E. For $\phi \in \Omega^+(\mathrm{ad}E)$ we have ([9, Chapter 6]):

$$d_A^+ d_A^* \phi = \frac{1}{2}\nabla_A^* \nabla_A \phi + \left(\frac{S}{6} - W^+\right)\phi + F_A^+ \cdot \phi,$$

where S is the scalar curvature of X and W^+ is the self-dual part of the Weyl curvature. Since $X = \mathbb{R}\times S^3$ is conformally flat, we have $W^+ = 0$. The scalar curvature S is constantly equal to 6. So we get

$$d_A^+ d_A^* \phi = \frac{1}{2}(\nabla_A^* \nabla_A + 2)\phi + F_A^+ \cdot \phi. \tag{5.2}$$

For any smooth $\eta \in \Omega^+(\mathrm{ad}E)$ with $\|\eta\|_{L^\infty(X)} < \infty$ there uniquely exists smooth $\phi \in \Omega^+(\mathrm{ad}E)$ satisfying

$$(\nabla_A^* \nabla_A + 2)\phi = \eta, \quad \|\phi\|_{L^\infty(X)} < \infty.$$

We sometimes denote ϕ by $(\nabla_A^* \nabla_A + 2)^{-1}\eta$. It satisfies

$$|\phi(x)| \le \int_X g(x,y)|\eta(y)| d\mathrm{vol}(y), \quad \|\phi\|_{L^\infty(X)} \le \|\eta\|_{\mathrm{Tau}}. \tag{5.3}$$

Moreover it satisfies the following. (Indeed this is the most spectacular property of the Taubes norm).

$$\left\|(d_A^*\phi \wedge d_A^*\phi)^+\right\|_{\mathrm{Tau}} \le 10\, \|\eta\|_{\mathrm{Tau}}^2. \tag{5.4}$$

For the detailed proofs of the above estimates, see [20, Section 4, Appendix].

We define $\mathcal{A}$ as the set of connections A on E such that

$$F_A^+ \text{ is compactly supported}, \quad \left\|F_A^+\right\|_{\mathrm{Tau}} \le \frac{1}{1000}, \quad \|F_A\|_{C_A^5} := \max_{0\le k\le 5} \left\|\nabla_A^k F_A\right\|_{L^\infty(X)} < \infty.$$

Here 1/1000 has no special meaning. Any sufficiently small number will do. The last condition is connected to the following fact: Take any point $p \in X$. Let g be the exponential gauge of radius $\pi/2$ around p. (The injectivity radius of X is equal to π.) Then the connection matrix $g(A)$ satisfies

$$|\nabla^k g(A)| \lesssim \|F_A\|_{C^k_A}.$$

We summarize the results of [20, Sections 4 and 5] in the following proposition.

Proposition 5.1. *We can construct a gauge equivariant map*

$$\mathcal{A} \ni A \mapsto \phi_A \in \Omega^+(\mathrm{ad}E)$$

satisfying the following conditions.
(1) $A + d_A^*\phi_A$ *is an ASD connection.*
(2) ϕ_A *is smooth and*

$$|\phi_A(x)| \lesssim \int_X g(x,y)|F_A^+(y)|d\mathrm{vol}(y), \quad \|\phi_A\|_{L^\infty(X)} \lesssim \left\|F_A^+\right\|_{\mathrm{Tau}}, \quad \|\nabla_A\phi_A\|_{L^\infty(X)} < \infty.$$

(3) If F_A *is compactly supported, then*

$$\int_X |F(A + d_A^*\phi_A)|^2 d\mathrm{vol} = \int_X \mathrm{tr}(F_A^2).$$

(4) For any $A, B \in \mathcal{A}$, $\|\phi_A - \phi_B\|_{L^\infty(X)} \lesssim \|A - B\|_{C^1_A}$.

Proof. We roughly explain the construction of ϕ_A for the convenience of readers. Let $\Omega^+(\mathrm{ad}E)_0$ be the set of smooth $\eta \in \Omega^+(\mathrm{ad}E)$ satisfying $\lim_{x\to\pm\infty}|\eta(x)| = 0$. Take $\eta \in \Omega^+(\mathrm{ad}E)_0$ and set $\phi = (\nabla_A^*\nabla_A + 2)^{-1}\eta \in \Omega^+(\mathrm{ad}E)_0$. We want to solve the equation $F^+(A + d_A^*\phi) = 0$. This is equivalent to

$$\eta = -2F_A - 2F_A^+ \cdot \phi - 2(d_A^*\phi \wedge d_A^*\phi)^+.$$

We denote the right-hand-side by $\Phi(\eta)$. By using the estimates (5.3) and (5.4), we can prove that Φ becomes a contraction map with respect to the Taubes norm over

$$\left\{\eta \in \Omega^+(\mathrm{ad}E)_0 \middle| \, \|\eta\|_{\mathrm{Tau}} \le \frac{3}{1000}\right\}.$$

Therefore the sequence η_n defined by

$$\eta_0 = 0, \quad \eta_{n+1} = \Phi(\eta_n)$$

is a Cauchy sequence with respect to the Taubes norm. Then $\phi_n := (\nabla_A^*\nabla_A + 2)^{-1}\eta_n$ is a convergent sequence in $L^\infty(X)$. Let ϕ_A be the limit of ϕ_n. We can prove that ϕ_A is smooth and ϕ_n converges to ϕ_A in C^∞ over every compact subset of X. Then it satisfies $F^+(A + d_A^*\phi_A) = 0$. The conditions (2), (3) and (4) can be checked by a detailed investigation of the above construction. $\square$

We need some more detailed estimates on ϕ_A. They are established in the next two lemmas.

Lemma 5.2. *We can choose* $0 < \tau < 1/1000$ *so that the following statement holds. If* $A \in \mathcal{A}$ *satisfies* $\left\|F_A^+\right\|_{\mathrm{Tau}} \leq \tau$ *then* ϕ_A *satisfies*

$$\|\nabla_A \phi_A\|_{L^\infty(X)} \leq 1 + \|F_A\|_{C^1_A}.$$

Proof. Suppose the statement is false. Then for any $n > 0$ there exists $A_n \in \mathcal{A}$ such that $\|F^+(A_n)\|_{\mathrm{Tau}} \leq 1/n$ and

$$R_n := \|\nabla_{A_n}\phi_{A_n}\|_{L^\infty(X)} > 1 + \|F(A_n)\|_{C^1_{A_n}}.$$

Take $p_n \in X$ satisfying $|\nabla_{A_n}\phi_{A_n}(p_n)| > R_n/2$. We consider the geodesic coordinate and the exponential gauge (w.r.t. A_n) of radius $\pi/2$ around p_n. Then the connection matrix of A_n in this gauge (also denoted by A_n) satisfies

$$|A_n| + |\nabla A_n| \lesssim \|F(A_n)\|_{C^1_{A_n}} < R_n.$$

We have the ASD equation

$$(\nabla_{A_n}^*\nabla_{A_n} + 2)\phi_{A_n} = -2F^+(A_n) - 2F^+(A_n)\cdot\phi_{A_n} - 2(d_{A_n}^*\phi_{A_n} \wedge d_{A_n}^*\phi_{A_n})^+$$

and the estimates $\|\phi_{A_n}\|_{L^\infty} \lesssim \|F^+(A_n)\|_{\mathrm{Tau}} \leq 1/n$ and $\|F^+(A_n)\|_{L^\infty} < R_n$. Then

$$\left|\sum_{i,j} g^{ij}(x)\partial_i\partial_j\phi_{A_n}\right| \lesssim R_n^2 \quad (|x| \leq \pi/2).$$

Here x is the geodesic coordinate around p_n. Set $\phi_n(y) = \phi_{A_n}(y/R_n)$ for $|y| \leq \pi/2$. This satisfies

$$|\nabla\phi_n(0)| > 1/2, \quad \left|\sum_{i,j} g^{ij}(y/R_n)\partial_i\partial_j\phi_n\right| \lesssim 1.$$

From the latter condition and $\|\phi_n\|_{L^\infty} \lesssim 1/n$, ϕ_n converges to 0 in C^1 over $|y| \leq \pi/3$. But this contradicts $|\nabla\phi_n(0)| > 1/2$. □

For $T > 1$ and $K > 0$ we define $\mathcal{A}(T,K) \subset \mathcal{A}$ as the set of connections A on E satisfying

$$\left\|F_A^+\right\|_{\mathrm{Tau}} \leq \tau, \quad \mathrm{supp}(F_A^+) \subset \{(t,\theta) \in \mathbb{R}\times S^3 \,|\, T-1 < |t| < T\}, \quad \|F_A\|_{C^5_A} \leq K.$$

Here τ is the positive constant introduced in Lemma 5.2. For $x = (t,\theta) \in \mathbb{R}\times S^3$ we set

$$g_T(x) = g_T(t) = e^{-\sqrt{2}|t-T|} + e^{-\sqrt{2}|t+T|},$$
$$\hat{g}_T(x) = \hat{g}_T(t) = (1+|t-T|)e^{-\sqrt{2}|t-T|} + (1+|t+T|)e^{-\sqrt{2}|t+T|}.$$

From the exponential decay estimate (5.1), the Green kernel $g(x,y)$ satisfies

$$\int_{T-1<|t|<T} g(x,y)d\mathrm{vol}(y) \lesssim g_T(x), \quad \int_X g(x,y)g_T(y)d\mathrm{vol}(y) \lesssim \hat{g}_T(x).$$

Lemma 5.3. *(1) For any $A \in \mathcal{A}(T,K)$ and $0 \le k \le 5$, $|\nabla_A^k \phi_A(x)| \lesssim_K g_T(x)$.*
(2) There exists $L_2 = L_2(K) > 1$ such that every $A \in \mathcal{A}(T,K)$ satisfies

$$\left| \int_{T-L_2<t<T+L_2} |F(A+d_A^*\phi_A)|^2 d\mathrm{vol} - \int_{T-L_2<t<T+L_2} \mathrm{tr}(F_A^2) \right| \le 1/10,$$

$$\left| \int_{-T-L_2<t<-T+L_2} |F(A+d_A^*\phi_A)|^2 d\mathrm{vol} - \int_{-T-L_2<t<-T+L_2} \mathrm{tr}(F_A^2) \right| \le 1/10.$$

(3) For any $A, B \in \mathcal{A}(T,K)$ and $0 \le k \le 5$

$$|\nabla_A^k \phi_A(x) - \nabla_B^k \phi_B(x)| \lesssim_K \hat{g}_T(x) \, \|A - B\|_{C^5_A}.$$

Proof. (1) From Proposition 5.1 (2), $|\phi_A(x)| \lesssim_K g_T(x)$. By Lemma 5.2, $\|\nabla_A \phi_A\|_{L^\infty} \lesssim_K 1$. Set $R = \sup_{t\in\mathbb{R}} g_T(t)^{-1} \|\phi_A\|_{L^2_{2,A}(t,t+1)}$. We have the ASD equation

$$(\nabla_A^* \nabla_A + 2)\phi_A = -2F_A^+ - 2F_A^+ \cdot \phi_A - 2(d_A^*\phi_A \wedge d_A^*\phi_A)^+.$$

From the elliptic estimate

$$\begin{aligned} \|\phi_A\|_{L^2_{2,A}(t,t+1)} &\lesssim_K \|\phi_A\|_{L^2(t-1,t+2)} + \|(\nabla_A^*\nabla_A + 2)\phi_A\|_{L^2(t-1,t+2)} \\ &\lesssim_K g_T(t) + \|d_A^*\phi_A \wedge d_A^*\phi_A\|_{L^2(t-1,t+2)} \\ &\lesssim_K g_T(t) + \|d_A^*\phi_A\|_{L^2(t-1,t+2)} \quad (\|\nabla_A\phi_A\|_{L^\infty} \lesssim_K 1). \end{aligned}$$

Let $\varepsilon = \varepsilon(K) > 0$ be a small number which will be fixed later. From the interpolation (Gilbarg–Trudinger [10, Theorem 7.28]),

$$\|d_A^*\phi_A\|_{L^2(t-1,t+2)} \le C(\varepsilon, K) \|\phi_A\|_{L^2(t-1,t+2)} + \varepsilon \|\phi_A\|_{L^2_{2,A}(t-1,t+2)}.$$

Hence

$$\|\phi_A\|_{L^2_{2,A}(t,t+1)} \le C'(\varepsilon,K) g_T(x) + C''(K)\varepsilon \|\phi_A\|_{L^2_{2,A}(t-1,t+2)}.$$

Then

$$R \le C'(\varepsilon, K) + C'''(K)\varepsilon R.$$

We choose ε so that $C'''(K)\varepsilon < 1/2$. Then $R \lesssim_K 1$, i.e. $\|\phi_A\|_{L^2_{2,A}(t,t+1)} \lesssim_K g_T(x)$. The rest of the argument is a standard bootstrapping.

(2) Set $a = d_A^*\phi_A$ and $cs_A(a) = \mathrm{tr}(2a\wedge F_A + a \wedge d_A a + \frac{2}{3}a^3)$. We have $\mathrm{tr}(F(A+a)^2) - \mathrm{tr}F_A^2 = dcs_A(a)$. Then by the Stokes theorem

$$\int_{T-L_2<t<T+L_2} |F(A+a)|^2 d\mathrm{vol} - \int_{T-L_2<t<T+L_2} \mathrm{tr}F_A^2 = \int_{t=T+L_2} cs_A(a) - \int_{t=T-L_2} cs_A(a).$$

By (1), the right-hand-side goes to zero (uniformly in A and T) as $L_2 \to \infty$.

(3) From (1), $|\nabla_A^k\phi_A(x)|, |\nabla_B^k\phi_B(x)| \lesssim_K g_T(x)$ for $0 \le k \le 5$. Set $a = B - A$. It is enough to prove the statement under the assumption $\|a\|_{C^5_A} < 1$. From the ASD equation,

(5.5)

$$\begin{aligned} &(\nabla_A^*\nabla_A + 2)(\phi_A - \phi_B) = 2(F_B^+ - F_A^+) + 2(F_B^+ \cdot \phi_B - F_A^+ \cdot \phi_A) \\ &+ 2\left\{(d_B^*\phi_B \wedge d_B^*\phi_B)^+ - (d_A\phi_A^* \wedge d_A^*\phi_A)^+\right\} + a * \nabla_B\phi_B + (\nabla_A a) * \phi_B + a * a * \phi_B. \end{aligned}$$

For any $t \in \mathbb{R}$, by the elliptic estimate
(5.6)
$$\|\phi_A - \phi_B\|_{L^2_{2,A}(t,t+1)} \lesssim_K \|\phi_A - \phi_B\|_{L^2(t-1,t+2)} + g_T(t)\|a\|_{C^1_A} + g_T(t)\|d^*_A\phi_A - d^*_A\phi_B\|_{L^2(t-1,t+2)}.$$
From Proposition 5.1 (4) we have $\|\phi_A - \phi_B\|_{L^\infty} \lesssim \|a\|_{C^1_A}$. So we get
$$\|\phi_A - \phi_B\|_{L^2_{2,A}(t,t+1)} \lesssim_K \|a\|_{C^1_A} + \|d^*_A\phi_A - d^*_A\phi_B\|_{L^2(t-1,t+2)}.$$
By using the interpolation as in (1), we get
$$\|\phi_A - \phi_B\|_{L^2_{2,A}(t,t+1)} \lesssim_K \|a\|_{C^1_A}.$$
Then the bootstrapping shows $\|\phi_A - \phi_B\|_{C^1_A} \lesssim_K \|a\|_{C^1_A}$. By this estimate, the modulus of the right-hand-side of (5.5) is $\lesssim_K g_T(x)\|a\|_{C^1_A}$. Then by the Green kernel estimate (5.3)
$$|\phi_A(x) - \phi_B(x)| \lesssim_K \hat{g}_T(x)\|a\|_{C^1_A}.$$
Using this and $\|\phi_A - \phi_B\|_{C^1_A} \lesssim_K \|a\|_{C^1_A}$ in (5.6), we get $\|\phi_A - \phi_B\|_{L^2_{2,A}(t,t+1)} \lesssim_K \hat{g}_T(t)\|a\|_{C^1_A}$. The rest of the proof is a bootstrapping. □

The next lemma is a preliminary version of Proposition 3.3. Here we connect the set $U_\delta(A,T)$ to $\mathcal{A}(T,K)$ above.

Lemma 5.4. *There exist positive numbers δ_1 and K such that for any $[A] \in \mathcal{M}_d$, any integer $T > 1$ and $0 < \delta \le \delta_1$ we can construct a (not necessarily continuous) map*
$$U_\delta(A,T) \to \mathcal{A}(T,K), \quad [B] \mapsto \hat{B},$$
satisfying the following conditions.
(1) There exists a gauge transformation g of E over $|t| < T-1$ satisfying $g(\hat{B}) = B$.
(2) There exists a gauge transformation h of E satisfying
$$\sup_{n\in\mathbb{Z}} e^{|n-G(\hat{A})|}\left\|h(\hat{B}) - \hat{A}\right\|_{L^2_{10,\hat{A}}(n,n+1)} \lesssim \delta.$$
(3) The curvature $F(\hat{A})$ is supported in $|t| < T$. Moreover
$$\left|\int_X \operatorname{tr}(F(\hat{A})^2) - \int_{-T<t<T}|F_A|^2 d\mathrm{vol}\right| \lesssim 1,$$
$$\int_{T-1<t<T}\operatorname{tr}(F(\hat{A})^2) \ge 10, \quad \int_{-T<t<-T+1}\operatorname{tr}(F(\hat{A})^2) \ge 10.$$

Proof. Choose a representative A of $[A]$. First we define $\hat{A}$. We take a cut-off $\varphi : \mathbb{R} \to [0,1]$ such that $\mathrm{supp}(d\varphi) \subset \{T - 1/2 < |t| < T\}$, $\varphi = 1$ over $|t| \le T - 1/2$ and $\varphi = 0$ over $|t| \ge T$. We can choose a trivialization u of E over $T-1 < |t| < T$ so that the connection matrix $u(A)$ satisfies $\|u(A)\|_{C^{10}} \lesssim 1$. We define a connection A_0 by $A_0 = u^{-1}(\varphi u(A))$. $A_0 = A$ over $|t| \le T - 1/2$, and A_0 is flat over $|t| \ge T$. The self-dual curvature $F^+(A_0)$ is supported in $T - 1/2 < |t| < T$. We try to reduce its Taubes norm by gluing sufficiently many concentrated instantons to A_0 over $T - 1/2 < |t| < T$. This is a rather standard technique for specialists of gauge theory. For the detail, see Donaldson [3, pp. 190-199]. After this gluing procedure, we get a connection $\hat{A}$ such that $\hat{A} = A$ over $|t| \leq T - 1/2$, $F(\hat{A})$ is supported in $|t| < T$ and

$$\mathrm{supp}(F^+(\hat{A})) \subset \{T - 1/2 < |t| < T\}, \quad \left\| F^+(\hat{A}) \right\|_{\mathrm{Tau}} \leq \tau/2, \quad \left\| F(\hat{A}) \right\|_{C^5_{\hat{A}}} \lesssim 1.$$

We can also assume that $\hat{A}$ satisfies the condition (3) of the statement. The last condition of (3) can be achieved by increasing the number of gluing instantons. Moreover, by the same reasoning, we can assume $\left\| F(\hat{A}) \right\|_{L^\infty(T-1,T)}, \left\| F(\hat{A}) \right\|_{L^\infty(-T,-T+1)} \geq \nu$. Hence $-T, T-1 \in G(\hat{A})$. This fact together with $\hat{A} = A$ over $|t| \leq T - 1/2$ implies

$$|n - G(\hat{A})| \leq |n - G(A,T)| \quad (-T \leq n \leq T-1). \tag{5.7}$$

Next we take $[B] \in U_\delta(A,T)$ ($\delta \leq \delta_1$) different from $[A]$. We can choose a representative B of $[B]$ satisfying

$$e^{|n-G(A,T)|} \|B - A\|_{L^2_{10,A}(n,n+1)} \leq \delta \quad (-T \leq n \leq T-1).$$

We take a cut-off $\psi : \mathbb{R} \to [0,1]$ such that $\mathrm{supp}(d\psi) \subset \{T-1 < |t| < T - 1/2\}$, $\psi = 1$ over $|t| \leq T-1$ and $\psi = 0$ over $|t| \geq T - 1/2$. Set $\hat{B} = \psi B + (1-\psi)\hat{A}$. This satisfies the condition (1) because $\hat{B} = B$ over $|t| \leq T-1$. $F^+(\hat{B})$ is supported in $\{T-1 < |t| < T\}$ and

$$\left\| F^+(\hat{B}) \right\|_{\mathrm{Tau}} \leq \mathrm{const} \cdot \delta_1 + \left\| F^+(\hat{A}) \right\|_{\mathrm{Tau}} \leq \tau$$

if we choose δ_1 sufficiently small. We can find a universal constant $K > 0$ so that $\left\| F(\hat{B}) \right\|_{C^5_{\hat{B}}} \leq K$ for all $[B] \in U_{\delta_1}(A,T)$. Then $\hat{B} \in \mathcal{A}(T,K)$.

We want to check the condition (2). $\hat{B} - \hat{A} = 0$ over $|t| \geq T - 1/2$. For $|t| < T - 1/2$ we have $\hat{A} = A$ and $\hat{B} - \hat{A} = \psi(B - A)$. Using (5.7), for $-T \leq n \leq T-1$

$$e^{|n-G(\hat{A})|} \left\| \hat{B} - \hat{A} \right\|_{L^2_{10,\hat{A}}(n,n+1)} \leq e^{|n-G(A,T)|} \|\psi(B-A)\|_{L^2_{10,A}(n,n+1)} \lesssim \delta.$$

If $n < -T$ or $n > T-1$ then $e^{|n-G(\hat{A})|} \left\| \hat{B} - \hat{A} \right\|_{L^2_{10,\hat{A}}(n,n+1)}$ is zero. This shows (2). □

Then we can prove the main result of this section.

Proposition 5.5 (= Proposition 3.3). *For any $r > 0$ we can choose $\delta_0 = \delta_0(r) > 0$ satisfying the following statement. For any $[A] \in \mathcal{M}_d$ and any integer $T > 1$ there exists a non-flat instanton A' on E and a (not necessarily continuous) map*

$$U_{\delta_0}(A,T) \to V_r(A'), \quad [B] \mapsto [B']$$

such that

(1)

$$\|F_{A'}\|_{L^\infty(X)} \leq D_0, \quad \left| \int_X |F_{A'}|^2 d\mathrm{vol} - \int_{(-T,T)\times S^3} |F_A|^2 d\mathrm{vol} \right| \lesssim 1.$$

Here D_0 is a universal constant independent of r.
(2) For any $[B] \in U_{\delta_0}(A,T)$ there exists a gauge transformation h of E over $|t| < T-1$ satisfying

$$|h(B') - B| \lesssim g_T(t) \quad (|t| < T-1).$$

Proof. Let $0 < \delta_0 = \delta_0(r) \le \delta_1$ (δ_1 is the positive constant introduced in Lemma 5.4). δ_0 will be fixed later. Take $[B] \in U_{\delta_0}(A,T)$ and set $B' = \hat{B} + d^*_{\hat{B}}\phi_{\hat{B}}$. Here $\hat{B}$ is constructed by Lemma 5.4, and $\phi_{\hat{B}}$ is constructed by Proposition 5.1. B' is an ASD connection. $F(\hat{A})$ is compactly supported, and hence Proposition 5.1 (3) implies

$$\int_X |F(A')|^2 d\mathrm{vol} = \int_X \mathrm{tr}(F(\hat{A})^2) < \infty. \tag{5.8}$$

Thus A' is an instanton. We will show $[B'] \in V_r(A')$ and the above conditions (1) and (2).

First we check (1). We have $F(A') = F(\hat{A}) + d_{\hat{A}}d^*_{\hat{A}}\phi_{\hat{A}} + (d_{\hat{A}}\phi_{\hat{A}})^2$. Since $\hat{A} \in \mathcal{A}(T,K)$, we get $\|F(A')\|_{L^\infty(X)} \lesssim 1$ by Lemma 5.3 (1). Moreover by (5.8) and Lemma 5.4 (3)

$$\left|\int_X |F_{A'}|^2 d\mathrm{vol} - \int_{-T<t<T} |F_A|^2 d\mathrm{vol}\right| = \left|\int_X \mathrm{tr}(F(\hat{A})^2) - \int_{-T<t<T} |F_A|^2 d\mathrm{vol}\right| \lesssim 1$$

Thus we have proved the condition (1).

Next we check (2). From Lemma 5.4 (1) we can assume $\hat{B} = B$ over $|t| < T-1$. Then $B' - B = d^*_{\hat{B}}\phi_{\hat{B}}$ over $|t| < T-1$. By Lemma 5.3 (1) we have $|d^*_{\hat{B}}\phi_{\hat{B}}| \lesssim g_T(t)$. Thus $|B' - B| \lesssim g_T(t)$ over $|t| < T-1$. This shows the condition (2).

The rest of the task is to show that A' is non-flat and $[B'] \in V_r(A')$. From lemma 5.3 (2) and Lemma 5.4 (3)

$$\int_{T-L_2<t<T+L_2} |F(A')|^2 d\mathrm{vol} > 9, \quad \int_{-T-L_2<t<-T+L_2} |F(A')|^2 d\mathrm{vol} > 9.$$

This implies that A' is not flat. Moreover by Lemma 3.1 the L^∞-norms of $F(A')$ over $T - L_2 < t < T + L_2$ and $-T - L_2 < t < -T + L_2$ are both bounded from below by ν. Hence

$$G'(A') \cap [T - L_2, T + L_2] \neq \emptyset, \quad G'(A') \cap [-T - L_2, -T + L_2] \neq \emptyset. \tag{5.9}$$

From Lemma 5.3 (1) $A' = \hat{A} + d^*_{\hat{A}}\phi_{\hat{A}}$ satisfies

$$|F(A') - F(\hat{A})| \lesssim g_T(t).$$

Then we can find a universal constant $L > L_2$ so that

$$t \in G(\hat{A}) \Longrightarrow (t - L, t + L) \cap G'(A') \neq \emptyset.$$

Then for all $n \in \mathbb{Z}$

$$|n - G'(A')| \le |n - G(\hat{A})| + L. \tag{5.10}$$

From Lemma 5.4 (2) we can assume

$$\sup_{n\in\mathbb{Z}} e^{|n-G(\hat{A})|}\left\|\hat{B}-\hat{A}\right\|_{L^2_{10,\hat{A}}(n,n+1)} \lesssim \delta_0. \tag{5.11}$$

$B'-A' = \hat{B}-\hat{A}+d^*_{\hat{B}}\phi_{\hat{B}} - d^*_{\hat{A}}\phi_{\hat{A}}$. From Lemma 5.3 (1) we have $\left\|A'-\hat{A}\right\|_{C^4_{\hat{A}}} \lesssim 1$. Then

$$e^{|n-G'(A')|}\left\|B'-A'\right\|_{L^2_{2,A'}(n,n+1)} \lesssim e^{|n-G'(A')|}\left\|\hat{B}-\hat{A}\right\|_{L^2_{2,\hat{A}}(n,n+1)} + e^{|n-G'(A')|}\left\|d^*_{\hat{B}}\phi_{\hat{B}} - d^*_{\hat{A}}\phi_{\hat{A}}\right\|_{L^2_{2,\hat{A}}(n,n+1)}.$$

From (5.10) and (5.11)

$$e^{|n-G'(A')|}\left\|\hat{B}-\hat{A}\right\|_{L^2_{2,\hat{A}}(n,n+1)} \lesssim \delta_0.$$

From Lemma 5.3 (3), $\left\|d^*_{\hat{B}}\phi_{\hat{B}} - d^*_{\hat{A}}\phi_{\hat{A}}\right\|_{L^2_{2,\hat{A}}(n,n+1)} \lesssim \hat{g}_T(n)\left\|\hat{B}-\hat{A}\right\|_{C^5_{\hat{A}}}$. By (5.11) and the Sobolev embedding,

$$e^{|n-G'(A')|}\left\|d^*_{\hat{B}}\phi_{\hat{B}} - d^*_{\hat{A}}\phi_{\hat{A}}\right\|_{L^2_{2,\hat{A}}(n,n+1)} \lesssim e^{|n-G'(A')|}\hat{g}_T(n)\delta_0.$$

Recall $\hat{g}_T(n) = (1+|n-T|)e^{-\sqrt{2}|n-T|} + (1+|n+T|)e^{-\sqrt{2}|n+T|}$ and (5.9). So $e^{|n-G'(A')|}\hat{g}_T(n) \lesssim e^{|n-\{\pm T\}|}\hat{g}_T(n) \lesssim 1$. Combining the above estimates, we conclude

$$\sup_{n\in\mathbb{Z}} e^{|n-G'(A')|}\left\|B'-A'\right\|_{L^2_{2,A'}(n,n+1)} \lesssim \delta_0.$$

Recall the definition of the norm $|||\cdot|||_{2,A'}$ in (3.1). It uses the weight function $W_{A'}$, and this satisfies $W_{A'}(t) \lesssim e^{\alpha|t-G'(A')|}$. Since $\alpha<1$ we get

$$|||B'-A'|||_{2,A'} \lesssim \sup_{n\in\mathbb{Z}} e^{|n-G'(A')|}\left\|B'-A'\right\|_{L^2_{2,A'}(n,n+1)} \lesssim \delta_0.$$

Thus we can choose $\delta_0 \ll r$ so that $|||B'-A'|||_{2,A'} \le r$ and hence $[B'] \in V_r(A')$. □

6. Quantitative deformation theory: proof of Proposition 3.4

The purpose of this section is to prove Proposition 3.4. Let $D>0$ be a positive number, and let A be a non-flat instanton on E satisfying $\|F_A\|_{L^\infty(X)} \le D$. First we recall some notations. We denote

$$G'(A) = \{n\in\mathbb{Z} | \, \|F_A\|_{L^\infty(n,n+1)} \ge \nu/2\} = \{n_1 < n_2 < \cdots < n_G\}.$$

Let W_A be the weight function introduced in Section 3.1. It is a smoothing of the function $e^{\alpha|t-G'(A)|}$ $(0<\alpha<1)$. For $u\in\Omega^i(\mathrm{ad}E)$ we define ($n_0=-\infty$ and $n_{G+1}=+\infty$)

$$|||u|||_{k,A} = \max_{0\le j\le G}\|W_A u\|_{L^2_{k,A}(n_j,n_{j+1})}.$$

The connection A is fixed throughout this section. So we usually abbreviate $|||u|||_{k,A}$ and $|||u|||_{0,A}$ to $|||u|||_k$ and $|||u|||$ respectively. We also abbreviate the weight function W_A to W. We define $L^{2,W}_k(\Omega^i(\mathrm{ad}E))$ as the Banach space of locally L^2_k sections $u \in \Omega^i(\mathrm{ad}E)$ satisfying $\|u\|_k < \infty$. Our main object is the space

$$V_r(A) = \{[B] : \text{ ASD on } E \mid \exists g : E \to E \text{ s.t. } \|g(B) - A\|_2 \le r\} \quad (r > 0).$$

First we prepare a lemma concerning $\Omega^0(\mathrm{ad}E)$. Here we essentially use our good/bad decomposition structure.

Lemma 6.1. *(1) For $u \in L^{2,W}_3(\Omega^0(\mathrm{ad}E))$,*

$$\|u\|_{L^\infty(X)} \lesssim_D \|d_A u\|_2 .$$

(2) For $u \in L^{2,W}_k(\Omega^0(\mathrm{ad}E))$ with $k \ge 1$, $\|u\|_k \lesssim_A \|d_A u\|_{k-1}$. Note that the implicit constant here depends on A. Hence this is less effective than (1).

Proof. (1) This follows from the Sobolev embedding and

$$\|u\|_{L^2(t,t+1)} \lesssim_D \|d_A u\| \quad (\forall t \in \mathbb{R}). \tag{6.1}$$

By the same argument as in Claim 4.3,

$$\|u\|_{L^2(n,n+1)} \lesssim_D \|d_A u\|_{L^2(n,n+1)} \quad (\forall n \in G'(A)). \tag{6.2}$$

Take $t \in (n_1, n_2)$ with $|t - n_1| \le |t - n_2|$ (other cases can be treated in the same way). For each $n_1 < s < n_1 + 1$

$$|u(t,\theta)| \le |u(s,\theta)| + \left|\int_s^t |\nabla_A u| d\tau\right| \le |u(s,\theta)| + \int_{n_1}^t |\nabla_A u| d\tau.$$

$$\int_{n_1}^t |\nabla_A u| d\tau = \int_{n_1}^t e^{-\alpha(\tau-n_1)} e^{\alpha(\tau-n_1)} |\nabla_A u| d\tau \le \sqrt{\int_{n_1}^t e^{-2\alpha(\tau-n_1)} d\tau} \sqrt{\int_{n_1}^t e^{2\alpha(\tau-n_1)} |\nabla_A u|^2 d\tau}.$$

Since $e^{\alpha|t-G'(A)|} \lesssim W(t)$, we get

$$\int_{n_1}^t |\nabla_A u| d\tau \lesssim \sqrt{\int_{n_1}^t W^2 |\nabla_A u|^2 d\tau},$$

$$|u(t,\theta)|^2 \lesssim |u(s,\theta)|^2 + \int_{n_1}^t W^2 |\nabla_A u|^2 d\tau.$$

Integrating over $(s,\theta) \in (n_1, n_1+1) \times S^3$,

$$\int_{S^3} |u(t,\theta)|^2 d\mathrm{vol}_{S^3}(\theta) \lesssim \int_{(n_1,n_1+1)\times S^3} |u|^2 d\mathrm{vol} + \int_{(n_1,t)\times S^3} W^2 |\nabla_A u|^2 d\mathrm{vol}.$$

Using (6.2)

$$\int_{S^3} |u(t,\theta)|^2 d\mathrm{vol}_{S^3}(\theta) \lesssim \|d_A u\|^2 .$$

The desired estimate (6.1) follows from this.

(2) It is enough to prove $\|u\| \lesssim_A \|d_A u\|$, and this follows from (6.1) and

$$\int_{\{t<n_1\}\cup\{t>n_G\}} W^2 |u|^2 d\mathrm{vol} \lesssim \|d_A u\|^2 .$$

For simplicity we assume $n_G = 0$ and prove

$$\int_{t>0} W^2 |u|^2 d\mathrm{vol} \lesssim \|d_A u\|^2 .$$

We can assume that u is smooth and compactly supported. Let $t > 0$.

$$|u(t,\theta)| \le \int_t^\infty |\nabla_A u(s,\theta)| ds = \int_t^\infty W(s)^{-1} W(s) |\nabla_A u(s,\theta)| ds.$$

For $0 < t < s$ we have $W(t)W(s)^{-1} \lesssim e^{\alpha(t-s)}$. Hence

$$W(t)|u(t,\theta)| \lesssim \int_t^\infty e^{\alpha(t-s)} W(s) |\nabla_A u(s,\theta)| ds,$$

$$\begin{aligned} W(t)^2 |u(t,\theta)|^2 &\lesssim \int_t^\infty e^{\alpha(t-s)} ds \int_t^\infty e^{\alpha(t-s)} W(s)^2 |\nabla_A u(s,\theta)|^2 ds \\ &= \frac{1}{\alpha} \int_t^\infty e^{\alpha(t-s)} W(s)^2 |\nabla_A u(s,\theta)|^2 ds. \end{aligned}$$

Therefore

$$\begin{aligned} \int_0^\infty W(t)^2 |u(t,\theta)|^2 dt &\lesssim \int_0^\infty \left(\int_0^s e^{\alpha(t-s)} dt \right) W(s)^2 |\nabla_A u(s,\theta)|^2 ds \\ &\le \frac{1}{\alpha} \int_0^\infty W(s)^2 |\nabla_A u(s,\theta)|^2 ds. \end{aligned}$$

Thus

$$\int_{t>0} W^2 |u|^2 d\mathrm{vol} \lesssim \int_{t>0} W^2 |\nabla_A u|^2 d\mathrm{vol} \le \|d_A u\|^2 .$$

□

Let $d_A^{*,W} : \Omega^1(\mathrm{ad}E) \to \Omega^0(\mathrm{ad}E)$ be the formal adjoint of $d_A : \Omega^0(\mathrm{ad}E) \to \Omega^1(\mathrm{ad}E)$ with respect to the weighted inner product: For compactly supported smooth $u \in \Omega^0(\mathrm{ad}E)$ and $a \in \Omega^1(\mathrm{ad}E)$

$$\int_X W^2 \langle d_A u, a \rangle d\mathrm{vol} = \int_X W^2 \langle u, d_A^{*,W} a \rangle d\mathrm{vol}.$$

The following lemma studies the Coulomb gauge condition.

Lemma 6.2. *(1) For $u \in L_1^{2,W}(\Omega^0(\mathrm{ad}E))$ and $a \in L^{2,W}(\Omega^1(\mathrm{ad}E))$ with $d_A^{*,W} a = 0$ (in the distributional sense)*

$$\|d_A u\| + \|a\| \lesssim_D \|d_A u + a\| .$$

(2) Let $k \ge 0$. For $u \in L_{k+1}^{2,W}(\Omega^0(\mathrm{ad}E))$ and $a \in L_k^{2,W}(\Omega^1(\mathrm{ad}E))$ with $d_A^{*,W} a = 0$

$$\|d_A u\|_k + \|a\|_k \lesssim_{k,D} \|d_A u + a\|_k .$$

Proof. (1) We can suppose that u is smooth and compactly supported. Let $N = N(D) > 0$ be a sufficiently large integer which will be fixed later. It is very important that several implicit constants below do not depend on N. Recall $G'(A) = \{n_1 < \cdots < n_G\}$. Let $G = qN + r$ with $0 \le r < N$. We decompose $\mathbb{R}$ as follows:

$$\mathbb{R} = (-\infty, n_N] \cup [n_N, n_{2N}] \cup \cdots \cup [n_{(q-1)N}, n_{qN}] \cup [n_{qN}, \infty).$$

We call these intervals $I_0, I_1, \dots, I_q$ respectively. If $q = 0$, then we simply set $I_0 = \mathbb{R}$. We set $I_{-1} = I_{q+1} = \emptyset$. For $0 \le k \le q$ we take a cut-off function $\varphi_k : \mathbb{R} \to [0, 1]$ such that $\varphi_k = 1$ on I_k, $\mathrm{supp}(\varphi_k) \subset I_{k-1} \cup I_k \cup I_{k+1} =: J_k$ and

$$\mathrm{supp}(d\varphi_k) \subset \bigcup_{n \in G'(A)} (n, n+1), \quad |d\varphi_k| \lesssim \frac{1}{N}. \tag{6.3}$$

From $d_A^{*,W} a = 0$

$$\int_X W^2 \langle d_A u + a, d_A(\varphi_k u) \rangle d\mathrm{vol} = \int_X W^2 \langle d_A u, d_A(\varphi_k u) \rangle d\mathrm{vol}.$$

$$\begin{aligned} \left| \int_X W^2 \langle d_A u, d_A(\varphi_k u) \rangle d\mathrm{vol} \right| &\gtrsim \int_{I_k} W^2 |d_A u|^2 d\mathrm{vol} - \frac{1}{N} \int_{\mathrm{supp}(d\varphi_k)} W^2 |d_A u||u| d\mathrm{vol} \\ &\ge \int_{I_k} W^2 |d_A u|^2 d\mathrm{vol} - \frac{1}{N} \sqrt{\int_{J_k} W^2 |d_A u|^2 d\mathrm{vol}} \sqrt{\int_{\mathrm{supp}(d\varphi_k)} W^2 |u|^2 d\mathrm{vol}}. \end{aligned}$$

$$\begin{aligned} \left| \int_X W^2 \langle d_A u + a, d_A(\varphi_k u) \rangle d\mathrm{vol} \right| &\le \sqrt{\int_{J_k} W^2 |d_A u + a|^2 d\mathrm{vol}} \sqrt{\int_X W^2 |d_A(\varphi_k u)|^2 d\mathrm{vol}} \\ &\lesssim \sqrt{\int_{J_k} W^2 |d_A u + a|^2 d\mathrm{vol}} \sqrt{\int_{\mathrm{supp}(d\varphi_k)} W^2 |u|^2 d\mathrm{vol} + \int_{J_k} W^2 |d_A u|^2 d\mathrm{vol}}. \end{aligned}$$

From (6.2) in the proof of Lemma 6.1 and the above (6.3),

$$\int_{\mathrm{supp}(d\varphi_k)} W^2 |u|^2 d\mathrm{vol} \lesssim \int_{\mathrm{supp}(d\varphi_k)} |u|^2 d\mathrm{vol} \lesssim_D \int_{J_k} |d_A u|^2 d\mathrm{vol} \lesssim \int_{J_k} W^2 |d_A u|^2 d\mathrm{vol}.$$

Combining these estimates,

$$\begin{aligned} &\int_{I_k} W^2 |d_A u|^2 d\mathrm{vol} \\ &\quad \lesssim_D \sqrt{\int_{J_k} W^2 |d_A u|^2 d\mathrm{vol}} \left(\sqrt{\int_{J_k} W^2 |d_A u + a|^2 d\mathrm{vol}} + \frac{1}{N} \sqrt{\int_{J_k} W^2 |d_A u|^2 d\mathrm{vol}} \right). \end{aligned}$$

Set

$$R = \max_k \sqrt{\int_{I_k} W^2 |d_A u|^2 d\mathrm{vol}}, \quad S = \max_k \sqrt{\int_{I_k} W^2 |d_A u + a|^2 d\mathrm{vol}}.$$

Then we get

$$R^2 \le C(D) \left(S + \frac{R}{N} \right) R, \text{ i.e. } R \le C(D) S + \frac{C(D)}{N} R.$$

We choose $N = N(D)$ so that $C(D)/N < 1/2$. Then $R \le 2C(D)S$. We have $\|d_A u\| \le R$ and $S \lesssim_D \|d_A u + a\|$. Thus $\|d_A u\| \lesssim_D \|d_A u + a\|$. Then $\|a\| \lesssim_D \|d_A u + a\|$.

(2) Let $k \ge 1$. By the elliptic regularity of the operator $d_A^{*,W} + d_A^+$,

$$\|d_A u\|_k \lesssim_{k,D} \|d_A u\| + \left\| (d_A^{*,W} + d_A^+) d_A u \right\|_{k-1}.$$

We have $(d_A^{*,W}+d_A^+)d_Au = d_A^{*,W}d_Au = d_A^{*,W}(d_Au+a)$. Hence by (1)

$$|||d_Au|||_k \lesssim_{k,D} \|d_Au\| + |||d_Au+a|||_k \lesssim_D |||d_Au+a|||_k\,.$$

□

Recall the Weitzenböck formula (5.2):

$$d_A^+d_A^*\phi = \frac{1}{2}(\nabla_A^*\nabla_A+2)\phi \quad (\phi\in\Omega^+(\mathrm{ad}E)).$$

Here d_A^* and ∇_A^* are the formal adjoints of d_A and ∇_A with respect to the standard (non-weighted) inner products. For any smooth $\eta\in\Omega^+(\mathrm{ad}E)$ with $\|\eta\|_{L^\infty(X)}<\infty$ there uniquely exists a smooth $\phi\in\Omega^+(\mathrm{ad}E)$ satisfying $\|\phi\|_{L^\infty(X)}<\infty$ and $d_A^+d_A^*\phi=\eta$. We denote this ϕ by $(d_A^+d_A^*)^{-1}\eta$. (See Section 5 and [20, Appendix].) We need to study the behavior of $(d_A^+d_A^*)^{-1}$ under the weighted norms.

Lemma 6.3. *For any $k\geq 0$ and any compactly supported smooth $\eta\in\Omega^+(\mathrm{ad}E)$*

$$\left|\left|\left|(d_A^+d_A^*)^{-1}\eta\right|\right|\right|_{k+2} \lesssim_{k,D} |||\eta|||_k\,.$$

So we can uniquely extend the operator $(d_A^+d_A^*)^{-1}$ to a bounded linear map from $L_k^{2,W}(\Omega^+(\mathrm{ad}E))$ to $L_{k+2}^{2,W}(\Omega^+(\mathrm{ad}E))$. We set

$$P_A := d_A^*(d_A^+d_A^*)^{-1} : L_k^{2,W}(\Omega^+(\mathrm{ad}E)) \to L_{k+1}^{2,W}(\Omega^1(\mathrm{ad}E)).$$

This satisfies $\|P_A\eta\|_{k+1}\lesssim_{k,D} |||\eta|||_k$.

Proof. Set $\phi=(d_A^+d_A^*)^{-1}\eta$. It is enough to prove $|||\phi|||\lesssim|||\eta|||$. By the Green kernel estimate (5.3)

$$|\phi(x)| \lesssim \int_X g(x,y)|\eta(y)|d\mathrm{vol}(y).$$

We have $\int_X g(x,y)d\mathrm{vol}(y)\lesssim 1$ (uniformly in x) and $g(x,y)\lesssim e^{-\sqrt{2}\,\mathrm{dist}(x,y)}$ for $\mathrm{dist}(x,y)>1$. Set $h(x,y)=W(x)W(y)^{-1}g(x,y)$.

$$W(x)|\phi(x)| \lesssim \int_X h(x,y)W(y)|\eta(y)|d\mathrm{vol}(y).$$

Since $e^{\alpha|t-G'(A)|}\lesssim W(t)\lesssim e^{\alpha|t-G'(A)|}$

$$W(x)W(y)^{-1}\lesssim e^{\alpha\,\mathrm{dist}(x,y)}.$$

Hence (noting $\alpha<1<\sqrt{2}$)

$$\int_X h(x,y)d\mathrm{vol}(y)\lesssim 1 \text{ (uniformly in } x), \quad h(x,y)\lesssim e^{(\alpha-\sqrt{2})\mathrm{dist}(x,y)} \quad (\mathrm{dist}(x,y)>1).$$

From the former condition

$$W(x)^2|\phi(x)|^2 \lesssim \int_X h(x,y)W(y)^2|\eta(y)|^2 d\mathrm{vol}(y).$$

We denote by t and s the $\mathbb{R}$-coordinates of $x, y \in \mathbb{R}\times S^3$ respectively.

$$\begin{aligned}\int_{n_i<t<n_{i+1}} W(x)^2|\phi(x)|^2 d\mathrm{vol}(x) &\lesssim \int_X \left(\int_{n_i<t<n_{i+1}} h(x,y) d\mathrm{vol}(x)\right) W(y)^2|\eta(y)|^2 d\mathrm{vol}(y)\\ &= \underbrace{\int_{n_i-1\le s\le n_{i+1}+1}\left(\int_{n_i<t<n_{i+1}} h(x,y)d\mathrm{vol}(x)\right) W(y)^2|\eta(y)|^2 d\mathrm{vol}(y)}_{(I)}\\ &\quad + \underbrace{\int_{\{s<n_i-1\}\cup\{s>n_{i+1}+1\}}\left(\int_{n_i<t<n_{i+1}} h(x,y)d\mathrm{vol}(x)\right) W(y)^2|\eta(y)|^2 d\mathrm{vol}(y)}_{(II)}.\end{aligned}$$

We have $\int_{n_i<t<n_{i+1}} h(x,y)d\mathrm{vol}(x) \lesssim 1$. So the term (I) is $\lesssim \|\eta\|$. When $s < n_i - 1$ or $s > n_{i+1}+1$,

$$\int_{n_i<t<n_{i+1}} h(x,y)d\mathrm{vol}(x) \lesssim \int_{n_i}^{n_{i+1}} e^{(\alpha-\sqrt{2})|t-s|}dt \lesssim \max\left(e^{(\alpha-\sqrt{2})|s-n_i|}, e^{(\alpha-\sqrt{2})|s-n_{i+1}|}\right).$$

Then the term (II) is also $\lesssim \|\eta\|$. Thus we conclude $\|\phi\| \lesssim \|\eta\|$. □

We define $H^{1,W}_A$ as the space of $a \in \Omega^1(\mathrm{ad}E)$ satisfying $d^{*,W}_A a = d^+_A a = 0$ and $\|a\| < \infty$. All the norms $\|\cdot\|_{k,A}$ $(k \ge 0)$ are equivalent over $H^{1,W}_A$ by the elliptic regularity.

Lemma 6.4.

$$\dim H^{1,W}_A = 8c_2(A) + 3, \quad c_2(A) := \frac{1}{8\pi^2}\int_X |F_A|^2 d\mathrm{vol}.$$

Proof. We set $\mathcal{D}_A = d^{*,W}_A + d^+_A : L^{2,W}_1(\Omega^1(\mathrm{ad}E)) \to L^{2,W}(\Omega^0(\mathrm{ad}E)\oplus\Omega^+(\mathrm{ad}E))$. $H^{1,W}_A$ is the kernel of $\mathcal{D}_A$. We will show that $\mathcal{D}_A$ is surjective. The map

$$d^{*,W}_A d_A : L^{2,W}_2(\Omega^0(\mathrm{ad}E)) \to L^{2,W}(\Omega^0(\mathrm{ad}E))$$

is injective and has a closed range by Lemma 6.1 (2). So it is an isomorphism by the principle of orthogonal projection. (See the proof of Lemma 6.5 (2) below.) Let $(u,\eta) \in L^{2,W}(\Omega^0(\mathrm{ad}E)\oplus\Omega^+(\mathrm{ad}E))$. We can find $v \in L^{2,W}_2(\Omega^0(\mathrm{ad}E))$ satisfying $d^{*,W}_A d_A v = u - d^{*,W}_A P_A\eta$. By $d^+_A P_A = 1$

$$\mathcal{D}_A(d_A v + P_A\eta) = (d^{*,W}_A d_A v + d^{*,W}_A P_A\eta, \eta) = (u,\eta).$$

Thus $\mathcal{D}_A$ is surjective. Therefore $\dim H^{1,W}_A = \dim \mathrm{Ker}(\mathcal{D}_A)$ is equal to the index of $\mathcal{D}_A$. The calculation of $\mathrm{index}(\mathcal{D}_A)$ is standard, and we get $\mathrm{index}(\mathcal{D}_A) = 8c_2(A)+3$ by Donaldson [4, Proposition 3.19]. □

Lemma 6.5. *(1) Let $k \ge 1$. For any $u \in L^{2,W}_{k+1}(\Omega^0(\mathrm{ad}E))$, $a \in H^{1,W}_A$ and $\eta \in L^{2,W}_{k-1}(\Omega^+(\mathrm{ad}E))$*

$$\|d_A u\|_k + \|a\| + \|\eta\|_{k-1} \lesssim_{k,D} \|d_A u + a + P_A\eta\|_k.$$

(2) Let $k \ge 1$. We define a map

$$\Phi : L^{2,W}_{k+1}(\Omega^0(\mathrm{ad}E)) \oplus H^{1,W}_A \oplus L^{2,W}_{k-1}(\Omega^+(\mathrm{ad}E)) \to L^{2,W}_k(\Omega^1(\mathrm{ad}E))$$

by $\Phi(u,a,\eta) = -d_A u + a + P_A\eta$. Then Φ is an isomorphism.

Proof. (1) Set $b = d_A u + a + P_A\eta$. $d_A^+ b = \eta$. So $\|\eta\|_{k-1} \lesssim_{k,D} \|b\|_k$. By Lemma 6.2 (2)

$$\|d_A u\|_k + \|a\| \lesssim_{k,D} \|d_A u + a\|_k \leq \|b\|_k + \|P_A\eta\|_k \lesssim_{k,D} \|b\|_k + \|\eta\|_{k-1} \lesssim_{k,D} \|b\|_k .$$

(2) It is enough to prove that Φ is surjective. Take $b \in L_k^{2,W}(\Omega^1(\mathrm{ad}E))$. Set $\eta = d_A^+ b$ and $b' = b - P_A\eta$. This satisfies $d_A^+ b' = 0$. By Lemma 6.1 (2), the space $d_A(L_1^{2,W}(\Omega^0(\mathrm{ad}E)))$ is closed in $L^{2,W}(\Omega^1(\mathrm{ad}E))$. So let $b' = -d_A u + a$ $(u \in L_1^{2,W}(\Omega^0(\mathrm{ad}E)))$ be the orthogonal decomposition with respect to the weighted inner product:

$$\int_X W^2 \langle d_A v, a\rangle d\mathrm{vol} = 0 \quad (\forall v \in L_1^{2,W}(\Omega^0(\mathrm{ad}E))).$$

Then $d_A^{*,W} a = 0$. Moreover $d_A^+ a = d_A^+(b' + d_A u) = 0$. Hence $a \in H_A^{1,W}$. We have $d_A u = a - b' \in L_k^{2,W}$. So $u \in L_{k+1}^{2,W}$. $b = -d_A u + a + P_A\eta = \Phi(u,a,\eta)$. Thus Φ is surjective. □

Let $a \in H_A^{1,W}$. The connection $A + a$ is an approximate solution of the ASD equation. In the next lemma, we perturb it and construct a genuine solution.

Lemma 6.6. *We can choose $r_1 = r_1(D) > 0$ so that the following statements hold.*
(1) For any $a \in H_A^{1,W}$ with $\|a\| \leq r_1$ there uniquely exists $\eta \in L_1^{2,W}(\Omega^+(\mathrm{ad}E))$ satisfying

$$F^+(A + a + P_A\eta) = 0, \quad \|\eta\|_1 \leq r_1.$$

We denote this η by η_a and set $\tilde{a} = a + P_A\eta_a$.
(2) For any $a, b \in H_A^{1,W}$ with $\|a\|, \|b\| \leq r_1$

$$\left\|\tilde{a} - \tilde{b}\right\|_{L^\infty(X)} \lesssim_D \|a - b\| .$$

Proof. (1) $F^+(A + a + P_A\eta) = \eta + \{(a + P_A\eta)^2\}^+$. Set $Q(\eta) = -\{(a + P_A\eta)^2\}^+$ for $\eta \in L_1^{2,W}(\Omega^+(\mathrm{ad}E))$. If $\eta_1, \eta_2 \in L_1^{2,W}(\Omega^+(\mathrm{ad}E))$ satisfy $\|\eta_1\|_1, \|\eta_2\|_1 \leq r_1$, then

$$\|Q(\eta_1)\|_1 \lesssim_D r_1^2, \quad \|Q(\eta_1) - Q(\eta_2)\|_1 \lesssim_D r_1 \|\eta_1 - \eta_2\|_1 .$$

Here we have used $L_2^{2,W} \times L_2^{2,W} \to L_1^{2,W}$. So if we choose $r_1 > 0$ sufficiently small, then Q becomes a contraction map over $\{\eta \in L_1^{2,W}(\Omega^+(\mathrm{ad}E)) | \|\eta\|_1 \leq r_1\}$. Thus the statement (1) follows.

(2) We have $\eta_a = -\{(a + P_A\eta_a)^2\}^+$ and $\eta_b = -\{(b + P_A\eta_b)^2\}^+$. Hence

$$\|\eta_a - \eta_b\|_1 \lesssim_D r_1 (\|a - b\| + \|\eta_a - \eta_b\|_1) .$$

If r_1 is sufficiently small, then $\|\eta_a - \eta_b\|_1 \lesssim_D \|a - b\|$. The rest of the argument is a bootstrapping. □

The next lemma is a conclusion of analytic arguments in this section. This is a non-linear version of Lemma 6.5.

Lemma 6.7. *We can choose $r_0 = r_0(D) > 0$ so that the following statement holds. For any connection B on E with $\|B - A\|_2 \le r_0$ there exists $(u, a, \eta) \in L^{2,W}_3(\Omega^0(\mathrm{ad}E)) \oplus H^{1,W}_A \oplus L^{2,W}_1(\Omega^+(\mathrm{ad}E))$ satisfying*

$$B = e^u(A + a + P_A\eta), \quad \|d_A u\|_2 + \|a\| + \|\eta\|_1 < r_1.$$

Here r_1 is the positive constant introduced in Lemma 6.6.

Proof. Let $r_0 = r_0(D)$ and $r_2 = r_2(D)$ be two positive numbers which will be fixed later. They will satisfy $0 < r_0 \ll r_2 < r_1$ We use a continuity method. The crucial point is that by Lemma 6.1 (1)

$$\|u\|_{L^\infty(X)} \lesssim_D \|d_A u\|_2 \quad (u \in L^{2,W}_3(\Omega^0(\mathrm{ad}E))). \tag{6.4}$$

Set $B = A + b$ with $\|b\|_2 \le r_0$. We define $\mathcal{T} \subset [0, 1]$ as the set of $0 \le t \le 1$ such that there exists $(u_t, a_t, \eta_t) \in L^{2,W}_3(\Omega^0(\mathrm{ad}E)) \oplus H^{1,W}_A \oplus L^{2,W}_1(\Omega^+(\mathrm{ad}E))$ satisfying

$$A + tb = e^{u_t}(A + a_t + P_A(\eta_t)), \quad \|d_A u_t\|_2 + \|a_t\| + \|\eta_t\|_1 < r_2. \tag{6.5}$$

The origin 0 is contained in $\mathcal{T}$. We will shows that $\mathcal{T}$ is closed and open. Then $1 \in \mathcal{T}$ and the proof is completed.

Step 1. We show that $\mathcal{T}$ is closed. Take $t \in \mathcal{T}$ and (u_t, a_t, η_t) satisfying the above (6.5). We want to derive a priori bound.

$$\begin{aligned} tb = &- d_A u_t + a_t + P_A\eta_t - (d_A e^{u_t})(e^{-u_t} - 1) - d_A(e^{u_t} - 1 - u_t) \\ &+ (e^{u_t} - 1)(a_t + P_A\eta_t)e^{-u_t} + (a_t + P_A\eta_t)(e^{-u_t} - 1). \end{aligned}$$

By (6.4) we get $\|-d_A u_t + a_t + P_A\eta_t\|_2 \lesssim_D r_0 + r_2^2$. By Lemma 6.5 (1), we can choose r_0 and r_2 so that

$$\|d_A u_t\|_2 + \|a_t\| + \|\eta_t\|_1 \le \frac{r_2}{2}. \tag{6.6}$$

Then the rest of the argument is standard. Suppose $\{t_i\} \subset \mathcal{T}$ is a sequence converging to $t_\infty \in [0, 1]$. Then by Lemma 6.1 (2) the sequence $(u_{t_i}, a_{t_i}, \eta_{t_i})$ is bounded in $L^{2,W}_3 \oplus H^{1,W}_A \oplus L^{2,W}_1$. So we can assume that it weakly converges to some $(u_{t_\infty}, a_{t_\infty}, \eta_{t_\infty})$. From the above bound (6.6) we get

$$\|d_A u_{t_\infty}\|_2 + \|a_{t_\infty}\| + \|\eta_{t_\infty}\|_1 \le \frac{r_2}{2} < r_2.$$

Hence it satisfies (6.5) for $t = t_\infty$. Thus $t_\infty \in \mathcal{T}$.

Step 2. We show that $\mathcal{T}$ is open in $[0, 1]$. Take $t \in \mathcal{T}$. We want to show that t is an inner point. Consider the map

$$f : L^{2,W}_3(\Omega^0(\mathrm{ad}E)) \oplus H^{1,W}_A \oplus L^{2,W}_1(\Omega^+(\mathrm{ad}E)) \to L^{2,W}_2(\Omega^1(\mathrm{ad}E)) \tag{6.7}$$

defined by $f(u, a, \eta) = e^u(A + a + P_A\eta) - A$. It is enough to prove that the derivative $(df)_{(0,a_t,\eta_t)}$ is an isomorphism.

$$(df)_{(0,a_t,\eta_t)}(u, a, \eta) = -d_A u + a + P_A\eta - [a_t + P_A\eta_t, u].$$

Here it is convenient to consider that the left-hand-side of (6.7) is endowed with the norm $\|d_A u\|_2 + \|a\| + \|\eta\|_1$. By Lemma 6.5 the map $\Phi(u, a, \eta) := -d_A u + a + P_A \eta$ is an isomorphism from $L^{2,W}_3 \oplus H^{1,W}_A \oplus L^{2,W}_1$ to $L^{2,W}_2$ with $\|d_A u\|_2 + \|a\| + \|\eta\|_1 \lesssim_D \|\Phi(u, a, \eta)\|_2$. By (6.4) and (6.5)

$$\|[a_t + P_A \eta_t, u]\|_2 \lesssim_D r_2 \|d_A u\|_2 .$$

So if r_2 is chosen sufficiently small, then the derivative $(df)_{(0, a_t, \eta_t)}$ is isomorphic. □

Then we can prove Proposition 3.4. Recall that for connections B_1 and B_2 on E we defined

$$\mathrm{dist}_{L^\infty}([B_1], [B_2]) = \inf_{g: E \to E} \|g(B_1) - B_2\|_{L^\infty(X)} .$$

Proposition 6.8 (= Proposition 3.4)**.** *There exists $C_0 = C_0(D) > 0$ such that for any $0 < \varepsilon < 1$*

$$\#_{\mathrm{sep}}(V_{r_0}(A), \mathrm{dist}_{L^\infty}, \varepsilon) \leq (C_0/\varepsilon)^{8c_2(A)+3}.$$

Here $r_0 = r_0(D)$ is the positive constant introduced in Lemma 6.7.

Proof. Set $B_{r_1}(H^{1,W}_A) = \{a \in H^{1,W}_A \mid \|a\| \leq r_1\}$.

Claim 6.9. *There exist $C_2 = C_2(D) > 0$ and a map $f : V_{r_0}(A) \to B_{r_1}(H^{1,W}_A)$ such that for any $[B_1], [B_2] \in V_{r_0}(A)$*

$$\mathrm{dist}_{L^\infty}([B_1], [B_2]) \leq C_2 \|f([B_1]) - f([B_2])\| .$$

Proof. Take $[B] \in V_{r_0}(A)$. By Lemma 6.7 we can find $(a, \eta) \in H^{1,W}_A \oplus L^{2,W}_1(\Omega^+(\mathrm{ad}E))$ satisfying

$$[B] = [A + a + P_A \eta], \quad \|a\| + \|\eta\|_1 < r_1.$$

Since B is ASD, $F^+(A + a + P_A \eta) = 0$. Then by Lemma 6.6 (1) we have $\eta = \eta_a$ and $[B] = [A + \tilde{a}]$. We set $f([B]) = a$.

Take $[B_1], [B_2] \in V_{r_0}(A)$ and set $a_1 = f([B_1])$ and $a_2 = f([B_2])$. We have $[B_1] = [A + \tilde{a}_1]$ and $[B_2] = [A + \tilde{a}_2]$. By Lemma 6.6 (2)

$$\mathrm{dist}_{L^\infty}([B_1], [B_2]) \leq \|\tilde{a}_1 - \tilde{a}_2\|_{L^\infty(X)} \lesssim_D \|a_1 - a_2\| .$$

□

By Lemma 2.1 and Example 2.2

$$\#_{\mathrm{sep}}(V_{r_0}(A), \mathrm{dist}_{L^\infty}, \varepsilon) \leq \#_{\mathrm{sep}}(B_{r_1}(H^{1,W}_A), \|\cdot\|, \varepsilon/C_2) \leq \left(\frac{1 + 2r_1 C_2}{\varepsilon}\right)^{\dim H^{1,W}_A} .$$

By Lemma 6.4, $\dim H^{1,W}_A = 8c_2(A) + 3$. Thus we get the conclusion. □

We have completed all the proofs of Theorem 1.1.

Remark 6.10. By the same argument we can prove the following more general result: Let $\mathcal{M} \subset \mathcal{M}_d$ be an $\mathbb{R}$-invariant closed subset. Then

$$\dim(\mathcal{M} : \mathbb{R}) \leq 8 \sup_{[A] \in \mathcal{M}} \rho(A).$$

But we don't have any reasonable lower bound on the mean dimension for general $\mathcal{M}$.

References

[1] M.F. Atiyah, N.J. Hitchin, I.M. Singer, Self-duality in four-dimensional Riemannian geometry, Proc. R. Soc. Lond. A. **362** (1978) 425-461.

[2] B.F.P. Da Costa, Deux exemples sur la dimension moyenne d'un espace de courbes de Brody, arXiv:1110.6082.

[3] S.K. Donaldson, The approximation of instantons, Geom. Funct. Anal. **3** (1993) 179-200.

[4] S.K. Donaldson, Floer homology groups in Yang-Mills theory, with the assistance of M. Furuta and D. Kotschick, Cambridge University Press, Cambridge (2002).

[5] S.K. Donaldson, P.B. Kronheimer, The geometry of four-manifolds, Oxford University Press, New York (1990).

[6] M. Einsiedler, T. Ward, Ergodic theory with a view towards number theory, Graduate Texts in Mathematics **259**, Springer, London.

[7] G.A. Elliott, Z. Niu, The C^*-algebra of a minimal homeomorphism of zero mean dimension, arXiv:1406.2382.

[8] A. Floer, An instanton-invariant for 3-manifolds, Comm. Math. Phys., **118** (1988) 215-240.

[9] D.S. Freed, K.K. Uhlenbeck, Instantons and four-manifolds, Second edition, Springer-Verlag, New York (1991).

[10] D. Gilbarg, N. S. Trudinger, Elliptic partial differential equations of second order, Reprint of the 1998 edition, Classics in Mathematics, Springer-Verlag, Berlin (2001).

[11] M. Gromov, Topological invariants of dynamical systems and spaces of holomorphic maps: I, Math. Phys. Anal. Geom. **2** (1999) 323-415.

[12] Y. Gutman, Mean dimension & Jaworski-type theorem, arXiv:1208.5248.

[13] Y. Gutman, Dynamical embedding in cubical shifts & the topological Rokhlin and small boundary properties, arXiv:1301.6072.

[14] Y. Gutman, M. Tsukamoto, Mean dimension and a sharp embedding theorem: extensions of aperiodic subshifts, Ergodic Theory Dynam. Systems, DOI: `http://dx.doi.org/10.1017/etds.2013.30` (to appear in print).

[15] A. Jaworski, Ph.D. Thesis, University of Maryland (1974).

[16] H. Li, B. Liang, Mean dimension, mean rank, and von Neumann-Lück rank, arXiv:1307.5471.

[17] E. Lindenstrauss, Mean dimension, small entropy factors and an embedding theorem, Inst. Hautes Études Sci. Publ. Math. **89** (1999) 227-262.

[18] E. Lindenstrauss, M. Tsukamoto, Mean dimension and embedding problem: an example, Israel J. Math. **199** (2014) 573-584.

[19] E. Lindenstrauss, B. Weiss, Mean topological dimension, Israel J. Math. **115** (2000) 1-24.

[20] S. Matsuo, M. Tsukamoto, Instanton approximation, periodic ASD connections, and mean dimension, J. Funct. Anal. **260** (2011) 1369-1427.

[21] S. Matsuo, M. Tsukamoto, Brody curves and mean dimension, J. Amer. Math. Soc. DOI: `http://dx.doi.org/10.1090/S0894-0347-2014-00798-0#sthash.RzSxNpa7.dpuf` (to appear in print).

[22] S. Matsuo, M. Tsukamoto, Local mean dimension of ASD moduli spaces over the cylinder, to appear in Israel J. Math, arXiv:1302.5977.

[23] C.H. Taubes, Self-dual Yang–Mills connections on non-self-dual 4-manifolds, J. Differential Geom. **17** (1982) 139-170.

[24] C.H. Taubes, Path-connected Yang-Mills moduli spaces, J.Differential Geom. 19 (1984), 337-392.

[25] M. Tsukamoto, Gluing an infinite number of instantons, Nagoya Math. J. **188** (2007) 107-131.

[26] M. Tsukamoto, Moduli space of Brody curves, energy and mean dimension, Nagoya Math. J. **192** (2008) 27-58.

[27] M. Tsukamoto, Gauge theory on infinite connected sum and mean dimension, Math. Phys. Anal. Geom. **12** (2009) 325-380.

[28] M. Tsukamoto, Remark on energy density of Brody curves, Proc. Japan Acad. Ser. A **88** (2012) 127-131.

[29] M. Tsukamoto, Sharp lower bound on the curvatures of ASD connections over the cylinder, to appear in J. Math. Soc. Japan, arXiv:1204.1143.

[30] K.K. Uhlenbeck, Connections with L^p bounds on curvature, Commun. Math. Phys. **83** (1982) 31-42.

[31] K. Wehrheim, Uhlenbeck compactness, EMS Series of Lectures in Mathematics, European Mathematical Society, Zürich (2004).

Masaki Tsukamoto
Department of Mathematics, Kyoto University, Kyoto 606-8502, Japan
E-mail address: `tukamoto@math.kyoto-u.ac.jp`