Uniqueness of two-bubble wave maps

Jacek Jendrej, Andrew Lawrie

UNIQUENESS OF TWO-BUBBLE WAVE MAPS

JACEK JENDREJ AND ANDREW LAWRIE

Abstract. This is the second part of a two-paper series that establishes the uniqueness and regularity of a threshold energy wave map that does not scatter in both time directions.

Consider the $\mathbb{S}^2$-valued equivariant energy critical wave maps equation on $\mathbb{R}^{1+2}$, with equivariance class $k \geq 4$. It is known that every topologically trivial wave map with energy less than twice that of the unique k-equivariant harmonic map $\boldsymbol{Q}_k$ scatters in both time directions. We study maps with precisely the threshold energy $\mathcal{E} = 2\mathcal{E}(\boldsymbol{Q}_k)$.

In the first part of the series we gave a refined construction of a threshold wave map that asymptotically decouples into a superposition of two harmonic maps (bubbles), one of which is concentrating in scale. In this paper, we show that this solution is the *unique* (up to the natural invariances of the equation) two-bubble wave map. Combined with our earlier work [11] we can now give an exact description of every threshold wave map.

1. Introduction

This paper concerns wave maps from the Minkowski space $\mathbb{R}^{1+2}_{t,x}$ into the two-sphere $\mathbb{S}^2$, with k-equivariant symmetry. These are formal critical points of the Lagrangian action,

$$\mathcal{A}(\Psi) = \frac{1}{2}\int_{\mathbb{R}^{1+2}_{t,x}} \Big(-|\partial_t \Psi(t,x)|^2 + |\nabla \Psi(t,x)|^2 \Big)\, \mathrm{d}x\mathrm{d}t,$$

restricted to the class of maps $\Psi : \mathbb{R}^{1+2}_{t,x} \to \mathbb{S}^2 \subset \mathbb{R}^3$ that take the form,

$$\Psi(t,r,\theta) = (u(t,r), k\theta) \hookrightarrow (\sin u(t,r)\cos k\theta, \sin u(t,r)\sin k\theta, \cos u(t,r)) \in \mathbb{S}^2 \subset \mathbb{R}^3,$$

for some fixed $k \in \mathbb{N}$. Here u is the colatitude measured from the north pole of the sphere and the metric on $\mathbb{S}^2$ is given by $ds^2 = du^2 + \sin^2 u\, d\omega^2$. We note that (r,θ) are polar coordinates on $\mathbb{R}^2$, and $u(t,r)$ is radially symmetric.

Wave maps are known as nonlinear σ-models in high energy physics literature, see for example, [6, 25]. They satisfy a canonical example of a geometric wave equation – it simultaneously generalizes the free scalar wave equation to manifold valued maps and the classical harmonic maps equation to Lorentzian domains. The $2d$ case considered here is of particular interest, as the static solutions given by finite energy harmonic maps are amongst the simplest examples of topological solitons; other examples include kinks in scalar field equations, vortices in Ginzburg-Landau equations, magnetic monopoles, Skyrmions, and Yang-Mills instantons; see [25]. Wave maps under k-equivariant symmetry possess intriguing features from the point of view of nonlinear dynamics, for example, bubbling harmonic maps, multi-soliton solutions, etc., in the relatively simple setting of a geometrically natural scalar semilinear wave equation. For a more thorough presentation of the physical or geometric content of wave maps, see e.g., [6, 25, 36].

The Cauchy problem for k-equivariant wave maps is given by

$$\begin{aligned} \partial_t^2 u - \partial_r^2 u - \frac{1}{r}\partial_r u + k^2 \frac{\sin 2u}{2r^2} &= 0, \\ (u(t_0), \partial_t u(t_0)) &= (u_0, \dot{u}_0), \quad t_0 \in \mathbb{R}. \end{aligned} \tag{1.1}$$

J.Jendrej was supported by ANR-18-CE40-0028 project ESSED. A. Lawrie was supported by NSF grant DMS-1700127 and a Sloan Research Fellowship.

The conserved energy is

$$\mathcal{E}(\boldsymbol{u}(t)) := 2\pi \int_0^\infty \frac{1}{2}\left((\partial_t u)^2 + (\partial_r u)^2 + k^2 \frac{\sin^2 u}{r^2} \right) r \, \mathrm{d}r, \tag{1.2}$$

where we have used bold font to denote the vector $\boldsymbol{u}(t) := (u(t), \partial_t u(t))$. We will write vectors with two components as $\boldsymbol{v} = (v, \dot{v})$, noting that the notation $\dot{v}$ will not, in general, refer to a time derivative of v but rather just to the second component of $\boldsymbol{v}$. With this notation (1.1) can be rephrased as the Hamiltonian system

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{u}(t) = J \circ \mathrm{D}\,\mathcal{E}(\boldsymbol{u}(t)), \tag{1.3}$$

where

$$J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \mathrm{D}\,\mathcal{E}(\boldsymbol{u}(t)) = \begin{pmatrix} -\Delta u(t) + r^{-2} f(u(t)) \\ \partial_t u(t) \end{pmatrix}. \tag{1.4}$$

Note that above we have introduced the notation,

$$f(u) := k^2 \sin(2u).$$

We remark that both (1.3) and (1.2) are invariant under the scaling

$$\boldsymbol{u}(t,\cdot) \mapsto \boldsymbol{u}(t/\lambda, \cdot)_\lambda = (u(t/\lambda, \cdot/\lambda), \lambda^{-1}\partial_t u(t/\lambda, \cdot/\lambda)), \qquad \lambda > 0.$$

which makes this problem energy critical.

It follows from (1.2) that any regular k-equivariant initial data $\boldsymbol{u}_0$ of finite energy must satisfy $\lim_{r\to 0} u_0(r) = m\pi$ and $\lim_{r\to\infty} u_0(r) = n\pi$ for some $m, n \in \mathbb{Z}$. Since the smooth wave map flow depends continuously on the initial data these integers are fixed over any time interval $t \in I$ on which the solution is defined. This splits the energy space into disjoint classes indexed by the pair (m, n) and it is natural to consider the Cauchy problem (1.1) within a fixed class. These classes are related to the topological degree of the full map $\Psi(t) : \mathbb{R}^2 \to \mathbb{S}^2$. In particular, k-equivariant wave maps with $(m, n) = (0, 0)$ correspond to topologically trivial maps Ψ, whereas those with $(m, n) = (0, 1)$ are degree $= k$ maps.

The unique (up to scaling) k-equivariant harmonic map is given explicitly by

$$Q(r) := 2\arctan(r^k),$$

and we write, $\boldsymbol{Q} := (Q, 0)$. We note that $Q(r)$ has degree $= k$ and it is a standard fact that $\boldsymbol{Q}$ minimizes the energy amongst all degree k maps (see, e.g., [11]) and in particular amongst k-equivariant maps with $(m, n) = (0, 1)$. It is not hard to show that $\mathcal{E}(\boldsymbol{Q}) = 4\pi k$.

In this paper we consider topologically trivial k-equivariant wave maps, i.e., those with data $\boldsymbol{u}_0$ that satisfies $\lim_{r\to 0} u_0(r) = \lim_{r\to\infty} u_0(r) = 0$. The natural function space in which to consider such solutions in the energy space, which comes with the norm,

$$\|\boldsymbol{u}_0\|_{\mathcal{H}}^2 := \|u_0\|_H^2 + \|\dot{u}_0\|_{L^2}^2 := \int_0^\infty \left((\partial_r u_0(r))^2 + k^2 \frac{u_0(r)^2}{r^2} \right) r \, \mathrm{d}r + \int_0^\infty \dot{u}_0(r)^2 \, r \, \mathrm{d}r.$$

Denoting by $\mathcal{L}_0 := -\Delta + k^2 r^{-2}$ we remark that the H norm of a smooth function u_0 can also expressed as $\|u_0\|_H^2 = \langle \mathcal{L}_0 u_0 \mid u_0 \rangle$, where $\langle f \mid g \rangle := (2\pi)^{-1} \langle f \mid g \rangle_{L^2(\mathbb{R}^2)}$ is the L^2 inner product. We use $\mathcal{L}_0$ to define spaces of higher regularity, and we let $\mathcal{H}^2$ denote the norm

$$\|\boldsymbol{u}_0\|_{\mathcal{H}^2}^2 := \|u_0\|_{H^2}^2 + \|\dot{u}_0\|_H^2 := \langle \mathcal{L}_0 u_0 \mid \mathcal{L}_0 u_0 \rangle + \langle \mathcal{L}_0 \dot{u}_0 \mid \dot{u}_0 \rangle.$$

We also require the following weighted norm,

$$\|\boldsymbol{u}_0\|_{\boldsymbol{\Lambda}^{-1}\mathcal{H}} := \|(r\partial_r u_0, (r\partial_r + 1)\dot{u}_0)\|_{\mathcal{H}}.$$

While $\boldsymbol{Q} \notin \mathcal{H}$, this solution to (1.1) still plays a significant role in the dynamics of solutions in $\mathcal{H}$; for example, superpositions of two bubbles, i.e., $Q(r/\lambda) - Q(r/\mu)$ for $\lambda \neq \mu$, are elements of H.

1.1. **Sub-threshold theorems and bubbling.** The regularity theory for energy critical wave maps has been extensively studied; [1, 2, 14–20, 37, 38, 42, 43, 45, 46]. Recently, the focus has been on the nonlinear dynamics of solutions with large energy. A remarkable *sub-threshold theorem* was established in [22, 39, 40, 44]: every wave map with energy less than that of the first nontrivial harmonic map is globally regular on $\mathbb{R}^{1+2}$ and scatters to a constant map. The role of the minimal harmonic map in the formulation of the sub-threshold theoem was first clarified by fundamental work of Struwe [41], who showed that the smooth equivariant wave map flow can only develop a singularity by concentrating energy at the tip of a light cone via the bubbling off of at least one non-trivial finite energy harmonic map. Bubbling wave maps were first constructed in a series of influential works by Krieger, Schlag, Tataru [23], Rodnianski, Sterbenz [34], and Raphaël, Rodnianski [32], with the latter work yielding a stable blow-up regime; see also the recent work [21] for stability properties of the solutions from [23], as well as [13] for a classification of blowup solutions with a given radiation profile, and [31] for a construction of a new class of singular solutions that blow up in infinite time. In particular, all of these works demonstrate that blow up by bubbling can occur for maps with energy slightly above the ground-state harmonic map, which shows the sharpness of the sub-threshold theorem.

The sub-threshold theorem can be refined by taking into account the topological degree of the map. Only topologically trivial maps can scatter to a constant map and it was shown in [3, 24] that the correct threshold that ensures scattering is $\mathcal{E} < 2\mathcal{E}(\boldsymbol{Q})$ (rather than $\mathcal{E}(\boldsymbol{Q})$). The reasoning behind the number $2\mathcal{E}(\boldsymbol{Q})$ is as follows. The topological degree counts (with orientation) the number of times a map 'wraps around' $\mathbb{S}^2$. If a harmonic map of degree k bubbles off from a wave map $\Psi(t)$, then, in order for $\Psi(t)$ to be degree zero, it must also 'unwrap' k times away from the bubble. The minimum energy required for each wrapping is $4\pi k = \mathcal{E}(\boldsymbol{Q})$. Thus the energy required for a degree zero map to form a bubble is $\mathcal{E} \geq 8\pi k = 2\mathcal{E}(\boldsymbol{Q})$.

1.2. **Main result: uniqueness of two-bubble wave maps.** We consider topologically trivial k-equivariant maps with precisely the threshold energy $\mathcal{E} = 2\mathcal{E}(\boldsymbol{Q})$. Building on the work [9] of the first author and our work [11], we can now give an exact description of every such map. We show that for equivariance classes $k \geq 4$, there is a *unique* (up to the natural invariances up the equation) threshold wave map that does not scatter in both time directions.

Let $\boldsymbol{u}(t) : [T_0, \infty) \to \mathcal{H}$ be a solution to (1.1) with $\mathcal{E}(\boldsymbol{u}) = 2\mathcal{E}(\boldsymbol{Q})$. We say $\boldsymbol{u}(t)$ is a *two-bubble in forward time* if there exist $\iota \in \{+1, -1\}$ and continuous functions $\lambda(t), \mu(t) > 0$ such that

$$\lim_{t\to\infty} \|(u(t) - \iota(Q_{\lambda(t)} - Q_{\mu(t)}), \partial_t u(t))\|_{\mathcal{H}} = 0, \quad \lambda(t) \ll \mu(t) \text{ as } t \to \infty. \tag{1.5}$$

A *two-bubble in the backward time direction* is defined similarly. Here Q_ν denotes the scaling $Q_\nu(r) := Q(r/\nu)$. In [9] the first author constructed a two-bubble in forward time. In [11] we showed that the solution from [9] must be global and scattering in backwards time. In the companion paper [12] we gave a refined construction of a two-bubble in forward time, showing that it possesses additional regularity and decay, i.e., it lies in the space $\mathcal{H} \cap \mathcal{H}^2 \cap \boldsymbol{\Lambda}^{-1}\mathcal{H}$. In this paper we show that *there is only one* 2*-bubble wave map* in each equivariance classes $k \geq 4$.

Theorem 1.1 (Uniqueness of 2-bubble wave maps)**.** *Let $k \geq 4$. There exists a global-in-time solution $\boldsymbol{u}_c : \mathbb{R} \to \mathcal{H} \cap \mathcal{H}^2 \cap \boldsymbol{\Lambda}^{-1}\mathcal{H}$ of* (1.1) *such that*

$$\big\|\boldsymbol{u}_c(t) - \big(-\boldsymbol{Q} + \boldsymbol{Q}_{q_k|t|^{-\frac{2}{k-2}}}\big)\big\|_{\mathcal{H}} \to 0 \quad \textit{as} \quad t \to \infty,$$

where $q_k > 0$ is an explicit constant depending on k (see (1.6)*).*

Moreover, if $\boldsymbol{u}(t) \in \mathcal{H}$ is any other 2*-bubble in forward time, then there exists $(t_0, \mu_0) \in \mathbb{R} \times (0, \infty)$ such that,*

$$\boldsymbol{u}(t) = \boldsymbol{u}_{c,t_0,\mu_0,\pm}(t) := \pm\Big(u_c(t - t_0, r/\mu_0), \frac{1}{\mu_0}\partial_t u_c(t - t_0, r/\mu_0)\Big),$$

i.e., $\boldsymbol{u}_c(t)$ is unique *up to sign, time translation, and scale.*

Remark 1.2. We note that [11, Theorem 1.6] ensures that $\boldsymbol{u}_c(t)$ is global and scatters freely in backwards time.

Remark 1.3. The solution $\boldsymbol{u}_c(t)$ from Theorem 1.1 was constructed in [9]. However, the proof of uniqueness given in Section 3 requires more detailed information about $\boldsymbol{u}_c(t)$ than what is obtained via the methods in [9]. This refined construction of $\boldsymbol{u}_c(t)$ is carried out in the companion paper [12], and is summarized in Theorem 1.12 below. Of course only after Theorem 1.1 is proved can we be sure that $\boldsymbol{u}_c(t)$ is the same solution found in [9]. We note that the companion paper [12] contains the proof that $\boldsymbol{u}_c(t) \in \mathcal{H} \cap \mathcal{H}^2 \cap \boldsymbol{\Lambda}^{-1}\mathcal{H}$, as well as an expansion of the solution into profiles that decay up to the rate $t^{-\frac{3k-2}{k-2}}$, along with a precise dynamical characterization of the modulation parameters associated to each bubble; see the beginning of Section 1.3 for a detailed statement.

This result can be combined with the main theorem in [11] to obtain the following complete classification.

Theorem 1.4 (Classification of $\mathcal{E} = 2\mathcal{E}(\boldsymbol{Q})$ wave maps)**.** *Fix any equivariance class $k \geq 4$. Let $\boldsymbol{u} : (T_-, T_+) \to \mathcal{H}$ be a solution to* (1.1) *such that*

$$\mathcal{E}(\boldsymbol{u}) = 2\mathcal{E}(\boldsymbol{Q}) = 8\pi k.$$

Then $T_- = -\infty$, $T_+ = +\infty$, and one the following alternatives holds:

- $\boldsymbol{u}(t)$ *scatters freely in both time directions*
- $\boldsymbol{u}(t) = \boldsymbol{u}_{c,t_0,\mu_0,\pm}(t)$*, for some $(t_0, \mu_0) \in \mathbb{R} \times (0, \infty)$. This solution is a two-bubble in forward time and freely scattering in backwards time*
- $\boldsymbol{u}(t) = (u_{c,t_0,\mu_0,\pm}(-t), -\partial_t u_{c,t_0,\mu_0,\pm}(-t))$*, for some $(t_0, \mu_0) \in \mathbb{R} \times (0, \infty)$. This solution is a two-bubble in backwards time and freely scattering in forwards time and is given by time-reversing the solution from Theorem 1.1*

Remark 1.5. Several of the conclusions in the statement of Theorem 1.4 were proved in [11, Theorem 1.6]. In that work we showed that any threshold solution that does not scatter in some direction *must* be a two-bubble in that direction as in (1.5), and with rates $\lambda(t), \mu(t)$ that are to leading order the same as the rates of $\boldsymbol{u}_c(t)$. Additionally, in [11] we solved the so-called *collision problem* for this equation. We showed that any two bubble in forward time must scatter freely in backwards time, i.e., when scales of the bubbles become comparable, this 'collision' completely annihilates the 2-bubble structure and the entire solution becomes free radiation; see also [27–29]. Viewing the evolution of $\boldsymbol{u}_c(t)$ in forward time, this means that the 2-bubble emerges from pure radiation, and constitutes an orbit connecting two different dynamical behaviors.

Remark 1.6. Theorem 1.1 fits in a broader program to classify solutions to nonlinear wave equations via their linear radiation. Note that the solution $\boldsymbol{u}_c(t)$ emits zero linear radiation as $t \to \infty$. For (1.1) the conjecture (soliton resolution) is that the only solutions with this property are the trivial solution $\boldsymbol{u}(t) \equiv 0$ and pure multi-bubbles such as $\boldsymbol{u}_c(t)$. Theorem 1.1 says that $\boldsymbol{u}_c(t)$ is the only solution with two bubbles that emits zero radiation forward-in-time. More generally, one can fix a linear forward radiation profile $\boldsymbol{u}_L(t)$ and ask if there are solutions $\boldsymbol{u}(t)$ to (1.1) that asymptotically decouple into a sum of bubbles plus the radiative wave $\boldsymbol{u}_L(t)$, and more ambitiously, for which $\boldsymbol{u}_L(t)$ can these be classified?

This same type of perspective can be taken in the context of solutions that develop a singularity in finite time via bubbling, see [13] where a classification is given in terms of a given finite time radiation profile $\boldsymbol{u}^*$, which is a weak limit of the solution as $t \to T_+ < \infty$.

Remark 1.7. We expect identical theorems to hold for the equivariance classes $k = 2, 3$. In fact, the argument used to prove uniqueness in this paper adapts easily to these cases. However, we only carry out the refined construction in [12] for $k \geq 4$. The proof given in that paper can be readily adapted to cover the $k = 3$ case, but we avoided this due to a technical inconvenience to keep the exposition as simple as possible, see [12, Remark 1.4]. The $k = 2$ case is more delicate due to the

failure $r\partial_r Q(r) \notin H^*$. This introduces the need for cut-offs in the modulation analysis. This issue was confronted in [11], but we also avoided it in [12] to keep the analysis as straightforward as possible. Finally, the dynamics of non-scattering threshold solutions in the case $k = 1$ is different – there is blow up in finite time; see the recent paper by Rodriguez [35]. However, we still expect an analogous uniqueness statement to hold in that setting; see e.g., [13, Conjecture 1.9].

Remark 1.8. One can compare/contrast Theorem 1.4 with the classification of $E = E(W)$ threshold solutions of the focusing energy critical wave equation by Duyckaerts and Merle [4] with data (u_0, u_1) in the subset $(\dot{H}^1 \cap L^2) \times L^2$ of the energy space; see also [5] for the corresponding theorem for NLS. There W is the ground state Aubin-Talenti solution and it is shown that every threshold solution either scatters in both time directions, exhibits ODE blow up in both directions, is equal to W, or is one of two solutions $W^{\pm}$; W^- scatters freely in one direction and scatters to W in the other, and W^+ exhibits finite time ODE blow up in one direction and scatters to W in the other. One main difference here is that the non-scattering threshold solution $\boldsymbol{u}_c(t)$ contains 2 bubbles, one of which is concentrating, which significantly complicates the analysis.

Remark 1.9 (Strong vs. weak soliton interactions). The first uniqueness result for multi-solitons is due to Martel [26] who constructed and proved uniqueness of N-soliton solutions to g-KdV with distinct, nontrivial velocities. We refer to the multi-solitons in that work as *weakly interacting* since the leading order dynamics are given/determined by the internal motion of each individual soliton. We emphasize here a distinction with Theorem 1.1: the bubbles in $\boldsymbol{u}_c(t)$ are *strongly interacting* in the sense that the dynamics are driven by nonlinear interactions between the two bubbles.

Remark 1.10 (Unique strongly interacting topological solitons). One may compare Theorem 1.1 with the authors' recent work with Kowalczyk, [10], which establishes the existence and *uniqueness* of strongly interacting kink-antikink solutions to scalar field equations on the line (e.g., sine-Gordon and ϕ^4-model). While the Theorem 1.1 and the main result in [10] are quite similar in nature (albeit for different equations), we develop a completely different technique in this paper to establish uniqueness. We explain the difference in more detail in Remark 1.15 below.

Remark 1.11 (Uniqueness theorems in the blow up setting). Finally, we mention two other uniqueness results for solutions with non-trivial dynamics in the blow-up setting, namely the pioneering work of Merle [30] which proved the existence and uniqueness (up to phase) of minimal mass blow up for the mass critical NLS, and the remarkable paper by Raphaël and Szeftel [33] which proved an analogous result for same equation, but with an inhomogeneous nonlinearity (which precludes the use of the psuedo-conformal symmetry in the proof). Several techniques used in this series of papers were inspired by [33], although we emphasize the method we use to prove uniqueness is novel.

1.3. **The existence and regularity of two bubble wave maps.** The starting point for the proof of Theorem 1.1 is the existence of the two-bubble wave map $\boldsymbol{u}_c(t)$ from Theorem 1.1 with a precise description of its dynamics and regularity. This is the content of the companion paper [12]. For the reader's convenience we review the main conclusions here.

We begin by introducing some notation needed to state Theorem 1.12 below. We define,

$$\rho_k := \Big(\frac{8k}{\pi}\sin(\pi/k)\Big)^{\frac{1}{2}}, \quad \gamma_k := \frac{k}{2}\rho_k^2, \quad q_k := \Big(\frac{k-2}{2}\rho_k\Big)^{-\frac{2}{k-2}} \tag{1.6}$$

We remark that $\rho_k^2 = 16k\|\Lambda Q\|_{L^2}^{-2}$. Given a radial function $w : \mathbb{R}^2 \to \mathbb{R}$ we denote the H and L^2 re-scalings as follows

$$w_\lambda(r) := w(r/\lambda), \quad w_{\underline{\lambda}}(r) := \frac{1}{\lambda} w(r/\lambda)$$

The corresponding infinitesimal generators are given by

$$\begin{aligned}\Lambda w &:= -\frac{\partial}{\partial\lambda}\Big|_{\lambda=1} w_\lambda = r\partial_r w \quad (H\,\text{scaling})\\ \Lambda_0 w &:= -\frac{\partial}{\partial\lambda}\Big|_{\lambda=1} w_{\underline{\lambda}} = (1+r\partial_r)w \quad (L^2\text{scaling})\end{aligned} \tag{1.7}$$

Next, we define $C^\infty(0,\infty)$ functions $A, B, \widetilde{B}$ as the unique solutions to the equations,

$$\begin{aligned}&\mathcal{L}A = -\Lambda_0\Lambda Q, \quad 0 = \langle A \mid \Lambda Q\rangle\\ &\mathcal{L}B = \gamma_k \Lambda Q - 4r^{k-2}[\Lambda Q]^2, \quad 0 = \langle B \mid \Lambda Q\rangle\,,\\ &\mathcal{L}\widetilde{B} = -\gamma_k\Lambda Q + 4r^{-k-2}[\Lambda Q]^2, \quad 0 = \left\langle \widetilde{B} \mid \Lambda Q\right\rangle\end{aligned} \tag{1.8}$$

where here $\mathcal{L} := -\Delta + r^{-2}f'(Q)$ is the operator obtained via linearization about Q. These are constructed in [12, Lemma 3.3], and here we note that

$$\begin{aligned}A(r), B(r) = O(r^k) \ \text{ as } \ r\to 0, \quad \widetilde{B}(r) = O(r^k\,|\log r|) \ \text{ as } \ r\to 0\\ A(r), B(r), \widetilde{B}(r) = O(r^{-k+2}) \ \text{ as } \ r\to\infty\end{aligned}$$

Next, given a time interval $J\subset\mathbb{R}$ and a quadruplet of C^1 functions $(\mu(t),\lambda(t),a(t),b(t))$ on J we define the 2-bubble ansatz,

$$\mathbf{\Phi}(\mu(t),\lambda(t),a(t),b(t),r) = (\Phi(\mu(t),\lambda(t),a(t),b(t),r), \dot\Phi(\mu(t),\lambda(t),a(t),b(t),r))$$

by

$$\begin{aligned}\Phi(\mu,\lambda,a,b) &:= (Q_\lambda + b^2A_\lambda + \nu^k B_\lambda) - (Q_\mu + a^2A_\mu + \nu^k\widetilde{B}_\mu)\\ \dot\Phi(\mu,\lambda,a,b) &:= b\Lambda Q_{\underline{\lambda}} + b^3\Lambda A_{\underline{\lambda}} - 2\gamma_k b\nu^k A_{\underline{\lambda}} + b\nu^k\Lambda B_{\underline{\lambda}} - kb\nu^k B_{\underline{\lambda}} - ka\nu^{k+1}B_{\underline{\lambda}}\\ &\quad + a\Lambda Q_{\underline{\mu}} + a^3\Lambda A_{\underline{\mu}} + 2\widetilde{\gamma}_k a\nu^k A_{\underline{\mu}} + a\nu^k\Lambda\widetilde{B}_{\underline{\mu}} + kb\nu^{k-1}\widetilde{B}_{\underline{\mu}} + ka\nu^k\widetilde{B}_{\underline{\mu}}\end{aligned} \tag{1.9}$$

where we have introduced the notation, $\nu := \lambda/\mu$. To ensure that $\dot\Phi\in L^2$, we now restrict to the setting $k\geq 4$. See [12, Remark 1.4] for a discussion of the cases $k=2,3$. The main result from [12] is the following theorem.

Theorem 1.12 (A refined two-bubble construction)**.** [12] *Fix any equivariance class $k\geq 4$. There exists a global-in-time solution $\boldsymbol{u}_c(t)\in\mathcal{H}$ to* (1.1) *that is a* two-bubble in forward time *with the following additional properties:*

- *The solution $\boldsymbol{u}_c(t)$ lies in the space $\mathcal{H}\cap\mathcal{H}^2\cap\mathbf{\Lambda}^{-1}\mathcal{H}$, and scatters freely in negative time.*
- *There exists $T_0>0$, a quadruplet of $C^1([T_0,\infty)$ functions $(\mu_c(t),\lambda_c(t),a_c(t),b_c(t))$, and $\boldsymbol{w}_c(t)\in\mathcal{H}\cap\mathcal{H}^2\cap\mathbf{\Lambda}^{-1}\mathcal{H}$ so that on the time interval $[T_0,\infty)$ the solution $\boldsymbol{u}_c(t)$ admits a decomposition,*

$$\boldsymbol{u}_c(t) = \mathbf{\Phi}(\mu_c(t),\lambda_c(t),a_c(t),b_c(t)) + \boldsymbol{w}_c(t) \tag{1.10}$$

where $\mathbf{\Phi}$ is defined in (1.9) *and the functions $(\mu_c(t),\lambda_c(t),a_c(t),b_c(t))$ satisfy,*

$$\begin{aligned}\lambda_c(t) &= q_k t^{-\frac{2}{k-2}}(1+O(t^{-\frac{4}{k-2}+\epsilon})) \quad as \quad t\to\infty,\\ \mu_c(t) &= 1 - \frac{k}{2(k+2)}q_k^2 t^{-\frac{4}{k-2}} + O(t^{-\frac{6}{k-2}+\epsilon}) \quad as \quad t\to\infty,\\ b_c(t) &= q_k\frac{2}{k-2}t^{-\frac{k}{k-2}}(1+O(t^{-\frac{4}{k-2}+\epsilon})), \quad as \quad t\to\infty,\\ a_c(t) &= \frac{2k}{(k-2)(k+2)}q_k^2 t^{-\frac{k+2}{k-2}}(1+O(t^{-\frac{4}{k-2}+\epsilon})) \quad as \quad t\to\infty,\end{aligned}$$

where $\epsilon > 0$ is any fixed small constant. We also have,

$$|\lambda_c'(t) + b_c(t)| \lesssim t^{-\frac{2}{k-2}(2k-1)} \quad as \quad t \to \infty,$$
$$|\mu_c'(t) - a_c(t)| \lesssim t^{-\frac{2}{k-2}(2k-1)} \quad as \quad t \to \infty.$$

Finally, $\boldsymbol{w}_c(t)$ satisfies,

$$\|\boldsymbol{w}_c(t)\|_{\mathcal{H}}^2 \lesssim \lambda_c(t)^{3k-2},$$
$$\|\boldsymbol{w}_c(t)\|_{\mathcal{H}^2}^2 \lesssim \lambda_c(t)^{3k-4},$$
$$\|\boldsymbol{\Lambda w}_c(t)\|_{\mathcal{H}}^2 \lesssim \lambda_c(t)^{2k-2},$$

uniformly in $t \geq T_0$.

1.4. **An outline of the proof of uniqueness: method of refined modulation parameters.** The goal of this paper is to prove that $\boldsymbol{u}_c(t)$ from (1.10) is the unique 2-bubble in forward time up to a change of sign, a fixed time translation, and rescaling. We introduce a dynamical method to accomplish this. We will highlight below where the need for the refined construction in Theorem 1.12 appears in the proof.

The first observation is that $\boldsymbol{u}_c(t)$ yields an invariant 2-dimensional sub-manifold $\mathcal{M}$ of the energy space $\mathcal{H}$ via time translation and scaling. For large times, it is natural to endow this manifold with coordinates related to the 2-bubble structure of $\boldsymbol{u}_c(t)$, i.e., for all $t \geq T_0$ we use the C^1 functions $\lambda_c(t), \mu_c(t)$ given by Theorem 1.12 such that

$$\boldsymbol{u}_c(t) = \boldsymbol{Q}_{\lambda_c(t)} + \boldsymbol{Q}_{\mu_c(t)} + o_{\mathcal{H}}(1)$$

where $\lambda_c(t) = q_k t^{-\frac{2}{k-2}}(1 + o(1))$ and $\mu_c = 1 + o(1)$ as $t \to \infty$. Because $\lambda_c(t)$ is monotonic in time, it is natural to reparamaterize time via the inverse function, i.e., $t = \lambda_c^{-1}(\sigma)$ for $\sigma \in (0, \sigma_0]$ where $\sigma_0 = \lambda_c(T_0)$. We define,

$$\boldsymbol{U}(\mu, \sigma) := \boldsymbol{u}_c(\lambda_c^{-1}(\sigma))_\mu$$

i.e., (μ, σ) give coordinates on $\mathcal{M}$ in the large time regime.

Now let $\boldsymbol{u}(t) \in \mathcal{H}$ be *any other* 2-bubble solution in forward time (with the sign $\iota = +1$ in (1.5)). The idea is to modulate about $\boldsymbol{U}(\mu, \sigma)$. Via a standard argument, we show that there exist C^1-functions $\mu(t), \sigma(t)$ and $\boldsymbol{g}(t) \in \mathcal{H}$ such that for large enough times t we have

$$\begin{aligned} \boldsymbol{u}(t) &= \boldsymbol{U}(\mu(t), \sigma(t)) + \boldsymbol{g}(t), \\ 0 &= \big\langle \Lambda Q_{\mu(t)\sigma(t)} \mid g(t) \big\rangle = \big\langle \Lambda Q_{\mu(t)} \mid g(t) \big\rangle \end{aligned} \tag{1.11}$$

and that $\|\boldsymbol{g}(t)\|_{\mathcal{H}}, \sigma(t) \to 0$ as $t \to \infty$ (for the simple reason that $\mathcal{M}$ also asymptotically approaches the set of two bubble configurations $\{\boldsymbol{Q}_\lambda - \boldsymbol{Q}_\mu : (\lambda, \mu) \in (0, \infty) \times (0, \infty), \lambda/\mu \ll 1\}$). Note that the desired uniqueness would follow from showing that $\boldsymbol{g}(t) = 0$ for some time $t \geq T_0$.

We now make use of the fact that $\boldsymbol{u}(t)$ and $\boldsymbol{U}(\mu, \sigma)$ both have energy $\mathcal{E} = 2\mathcal{E}(\boldsymbol{Q})$. For each time $t \geq T_0$ we consider a Taylor expansion of the energy,

$$\begin{aligned} &\mathcal{E}(\boldsymbol{U}(\mu(t), \sigma(t))) = \mathcal{E}(\boldsymbol{u}(t)) = \mathcal{E}(\boldsymbol{U}(\mu(t), \sigma(t)) + \boldsymbol{g}(t)) \\ &\quad = \mathcal{E}(\boldsymbol{U}(\mu(t), \sigma(t))) + \langle D\mathcal{E}(\boldsymbol{U}(\mu(t), \sigma(t))) \mid \boldsymbol{g}(t) \rangle + \big\langle D^2\mathcal{E}(\boldsymbol{U}(\mu(t), \sigma(t)))\boldsymbol{g}(t) \mid \boldsymbol{g}(t) \big\rangle + o(\|\boldsymbol{g}\|_{\mathcal{H}}^2) \end{aligned}$$

Subtracting $\mathcal{E}(\boldsymbol{U}(\mu(t), \sigma(t)))$ from both sides, establishing a coercivity estimate for the quadratic term (which is a consequence of the orthogonality conditions (1.11)), and making the "little oh" term above smaller than half the coercivity constant $c_1 > 0$ (which is possible by taking $T_0 > 0$ large enough) we arrive at the inequality

$$0 \geq \langle D\mathcal{E}(\boldsymbol{U}(\mu(t), \sigma(t))) \mid \boldsymbol{g}(t) \rangle + \frac{1}{2} c_1 \|\boldsymbol{g}(t)\|_{\mathcal{H}}^2 \tag{1.12}$$

We next turn to studying the dynamics of the term $\langle D\mathcal{E}(\boldsymbol{U}(\mu(t), \sigma(t))) \mid \boldsymbol{g}(t) \rangle$ with the objective of finding a contradiction above in the case that $\boldsymbol{g}(t) \not\equiv 0$. This is not an unnatural object to study,

as one can observe that $D\mathcal{E}(\boldsymbol{U}(\mu,\sigma)) = -\lambda_c'(\lambda_c^{-1}(\sigma))\mu^{-1}J \circ \partial_\sigma \boldsymbol{U}(\mu,\sigma)$, i.e., it is a renormalized 90-degree rotation of the tangent vector $\partial_\sigma \boldsymbol{U}(\mu,\sigma)$ and moreover to leading order we have,

$$\langle D\mathcal{E}(\boldsymbol{U}(\mu(t),\sigma(t))) \mid \boldsymbol{g}(t)\rangle \approx -\lambda_c'(\lambda_c^{-1}(\sigma))(\mu\sigma)^{-1}\,\langle \Lambda Q_{\mu\sigma} \mid \dot{g}\rangle$$

In other words, $\langle D\mathcal{E}(\boldsymbol{U}(\mu(t),\sigma(t))) \mid \boldsymbol{g}(t)\rangle$ is deeply related to the dynamics of the modulation parameters as can be seen from differentiating the terms in the second line in (1.11). However, a naive second differentiation of the orthogonality conditions (1.11) does not directly reveal useful information on the dynamics since terms of critical size but indeterminate sign arise. Here we use a technique similar to the one developed in [7,9,11] – we perform an ad hoc correction *to the modulation parameters themselves* using a localized virial functional. After proving that $-\lambda_c'(\lambda_c^{-1}(\sigma)) = \rho_k \sigma^{\frac{k}{2}}(1+o(1))$ in Theorem 1.12, we define,

$$b(t) := \frac{1}{\rho_k \sigma^{\frac{k}{2}}}\,\langle D\mathcal{E}(\boldsymbol{U}(\mu(t),\sigma(t))) \mid \boldsymbol{g}(t)\rangle + \langle \mathcal{A}_0(\mu(t)\sigma(t))g(t) \mid \dot{g}(t)\rangle \tag{1.13}$$

where $\mathcal{A}_0(\mu\sigma)$ is the same localized and rescaled version of the virial operator used in the companion paper [12], i.e., $\mathcal{A}_0(\mu\sigma) \approx (\mu\sigma)^{-1}\Lambda_0$ up to scale $\mu\sigma$. While the correction is small (order $\|\boldsymbol{g}\|_{\mathcal{H}}^2$) as compared to the first term, its derivative is large and designed to cancel terms with critical size but indeterminate sign.

The heart of the argument is an almost monotonicity formula for $b(t)$, proved in Proposition 4.3, which readily leads to a contradiction in (1.12). It is in the proof of Proposition 4.3 where the need for refined asymptotics and refined regularity estimates for $\boldsymbol{U}(\mu,\sigma)$ arises – indeed, one can observe from (1.13) and (1.4) that the equation for b' will involve estimates on the second derivatives of $U(\mu,\sigma)$, given that $\boldsymbol{g}(t)$ can only be assumed to lie in $\mathcal{H}$. The proof also requires weighted energy estimates. The list of estimates on $\boldsymbol{U}(\mu,\sigma)$ needed for the argument is given in Corollaries 3.3 and 3.4, and Theorem 1.12 is proved with these in mind. Of course in [11] the same type of higher regularity and weighted estimates arise as well, but there we modulated around the 2-bubble family $\boldsymbol{Q}_\lambda - \boldsymbol{Q}_\mu$, rather than the *constructed solution* $\boldsymbol{U}(\mu,\sigma)$, and thus the analogous estimates there followed trivially from the formula for $Q(r)$.

Remark 1.13. The basic outline above draws inspiration from the first author's work [8] in a different context. There one uses a combination of the energy expansion with a modulation analysis to rule out two bubble configurations with opposing signs for the critical NLW, albeit without the virial correction to the modulation parameters, which is a crucial ingredient here.

Remark 1.14. Note that the argument does not use that $\boldsymbol{u}_c(t)$ is a threshold solution in a crucial way, and thus should be applicable in other settings.

Remark 1.15. Together with the proof of uniqueness of the strongly interacting kink-antikink pair in [10] we have now introduced two quite different techniques to prove uniqueness (and existence) of solutions to dispersive PDEs exhibiting nontrivial dynamics under some qualitative assumption – here the assumption is the solution has threshold energy but is non-scattering, and in [10] we look for asymptotically stationary 2-kink solutions to scalar field equations.

The methods differ as follows. In [10] we first establish a quantitative classification of the dynamics for any kink-antikink pair. Then we find a single, unique kink-antikink solution in a time-weighted function space via a contraction mapping argument – in fact this is done in two-steps by way of a novel implementation of Liapunov-Schmidt reduction. The preliminary quantitative classification result is then used to show that any finite energy kink-antikink must also lie in this weighted function space, which proves uniqueness.

In contrast, here *we do not make any use of the dynamical classification of non-scattering threshold solutions obtained in our previous paper [11]* to prove uniqueness. We resort instead to the novel modulation technique that we just outlined above. The steep cost however, is that this modulation method requires very refined information on the constructed solution $\boldsymbol{u}_c(t)$ (including $\mathcal{H}^2$-estimates), which leads to the lengthy computations in the companion paper [12].

In summary, one can say quite roughly that the method here is inspired by the general principle of weak-strong uniqueness whereas in [10] the method uses the contraction mapping principle to deduce uniqueness. We note that the method from [10] should be adaptable to the present setting and vice versa. Both methods should be applicable in other settings as well.

2. Preliminaries

For radial functions u, v on $L^2(\mathbb{R}^2)$, we write $u = u(r), v = v(r)$ and we use the notation,

$$\langle u \mid v\rangle := \frac{1}{2\pi}\langle u \mid v\rangle_{L^2(\mathbb{R}^2)} = \int_0^\infty u(r)\overline{v(r)}\, r\,\mathrm{d}r.$$

Let $\mathcal{L}_0$ denote the operator

$$\mathcal{L}_0 w := -\Delta w + \frac{k^2}{r^2} w. \tag{2.1}$$

We define the function space H as the completion of $C_0^\infty((0,\infty))$ functions w under the norm

$$\|w\|_H^2 := \langle \mathcal{L}_0 w \mid w\rangle = \int_0^\infty \Big((\partial_r w(r))^2 + k^2\frac{w(r)^2}{r^2}\Big)\, r\mathrm{d}r$$

For the vector pair $\boldsymbol{w} = (w, \dot w)$ we define the norm $\mathcal{H}$ by

$$\|\boldsymbol{w}\|_{\mathcal{H}}^2 := \|w\|_H^2 + \|\dot w\|_{L^2}^2$$

Next, we define the space H^2 via the norm,

$$\|w\|_{H^2}^2 := \langle \mathcal{L}_0 w \mid \mathcal{L}_0 w\rangle = \int_0^\infty \Big((\partial_r^2 w(r))^2 + (2k^2+1)\frac{(\partial_r w(r))^2}{r^2} + (k^4-4k^2)\frac{w(r)^2}{r^4}\Big)\, r\,\mathrm{d}r$$

And for the pair $\boldsymbol{w} = (w, \dot w)$ we define $\mathcal{H}^2$ by

$$\|\boldsymbol{w}\|_{\mathcal{H}^2}^2 := \|w\|_{H^2}^2 + \|\dot w\|_H^2$$

We also require the following weighted norm,

$$\|\boldsymbol{w}\|_{\boldsymbol{\Lambda}^{-1}\mathcal{H}} := \|(\Lambda w, \Lambda_0 \dot w)\|_{\mathcal{H}}$$

where Λ, Λ_0 are defined in (1.7). It is a standard fact that the regularity of a solution $\boldsymbol{u}(t)$ to (1.1) in the space $\mathcal{H}\cap\mathcal{H}^2\cap\boldsymbol{\Lambda}^{-1}\mathcal{H}$ is propagated by the flow.

The infinitesimal generators Λ, Λ_0 defined in (1.7) satisfy the integration by parts identities,

$$\langle \Lambda f \mid g\rangle = -\langle f \mid \Lambda g\rangle - 2\langle f \mid g\rangle, \quad \langle \Lambda_0 f \mid g\rangle = -\langle f \mid \Lambda_0 g\rangle$$

The operator $\mathcal{L}_U$ obtained by linearization of (1.1) about the first component of finite energy map $\boldsymbol{U} = (U, \dot U)$ plays an important role in the analysis. Given $g \in H$ we have,

$$\mathcal{L}_U g := -\Delta g + k^2\frac{\cos 2U}{r^2} g$$

In fact, given any $\boldsymbol{g} = (g, \dot g) \in \mathcal{H}$ we have

$$\big\langle \mathrm{D}^2\,\mathcal{E}(\boldsymbol{U})\boldsymbol{g} \mid \boldsymbol{g}\big\rangle = \langle \mathcal{L}_U g \mid g\rangle_{L^2} + \langle \dot g \mid \dot g\rangle_{L^2} = \int_0^\infty \Big(\dot g^2 + (\partial_r g)^2 + k^2\frac{\cos 2U}{r^2} g^2\Big) r\mathrm{d}r$$

The most important instance of the operator $\mathcal{L}_U$ is given by linearizing (1.1) about $U = Q_\lambda$. In this case we use the short-hand notation,

$$\mathcal{L}_\lambda := \mathcal{L}_{Q_\lambda} = (-\Delta + \frac{k^2}{r^2}) + \frac{1}{r^2}(f'(Q) - k^2)$$

We write $\mathcal{L} := \mathcal{L}_1$. We often use the notation $\mathcal{L} = \mathcal{L}_0 + P$, where $\mathcal{L}_0$ is as in (2.1) and

$$P(r) := \frac{1}{r^2}(f'(Q) - k^2) = -\frac{2k^2\sin^2 Q}{r^2} = -4k^2\frac{r^{2k-2}}{(1+r^{2k})^2} \tag{2.2}$$

We recall that

$$\Lambda Q(r) = k \sin Q = \frac{2kr^k}{1+r^{2k}}$$

is a zero energy eigenfunction for $\mathcal{L}$, that is,

$$\mathcal{L}\Lambda Q = 0, \quad \text{and} \quad \Lambda Q \in L^2_{\mathrm{rad}}(\mathbb{R}^2).$$

for all $k \geq 2$. When $k = 1$, $\mathcal{L}\Lambda Q = 0$ holds but $\Lambda Q \notin L^2$ due to slow decay as $r \to \infty$ and 0 is referred to as a threshold resonance. In fact, ΛQ spans the kernel of $\mathcal{L}$; see [11] for more.

We require the following localized coercivity result for functions in the orthogonal complement to the kernel of $\mathcal{L}$. This was proved in detail in [9]; see also [12].

Lemma 2.1 (Localized coercivity for $\mathcal{L}$)**.** [9, Lemma 5.4] *There exists a uniform constant $c_1 > 0$ with the following property. Suppose $w \in H$ is such that*

$$\langle w \mid \Lambda Q \rangle = 0. \tag{2.3}$$

Then,

$$\langle \mathcal{L}w \mid w \rangle \geq c_1 \|w\|_H^2.$$

In addition, for any $c > 0$, there exists $R_1 > 0$ large enough so that for all $w \in H$ as in (2.3)*, we have*

$$\int_0^{R_1} \Big((\partial_r w(r))^2 + k^2 \frac{w(r)^2}{r^2} \Big) r \mathrm{d}r + \langle Pw \mid w \rangle \geq -c \|w\|_H^2 \tag{2.4}$$

Lastly, for any $c > 0$, there exists $r_1 > 0$ small enough so that for all $w \in H$ as in (2.3)*, we have*

$$\int_{r_1}^{\infty} \Big((\partial_r w(r))^2 + k^2 \frac{w(r)^2}{r^2} \Big) r \mathrm{d}r + \langle Pw \mid w \rangle \geq -c \|w\|_H^2$$

2.1. **The truncated virial operators.** We define truncated virial operators $\mathcal{A}(\lambda)$ and $\mathcal{A}_0(\lambda)$, and state related estimates. Nearly identical operators were introduced by the first author in [9] and used crucially by the authors in [11]. We require a slight modification, which was established in [12].

Lemma 2.2. [9, Lemma 4.6] [12, Lemma 4.1] *For each $c, R > 0$ there exists a function $p(r) = p_{c,R}(r) \in C^{5,1}((0,+\infty))$ with the following properties:*

(1) $p(r) = \frac{1}{2} r^2$ for $r \leq R$,
(2) there exists $\widetilde{R} = \widetilde{R}(R,c) > R$ such that $p(r) \equiv \text{const}$ for $r \geq \widetilde{R}$,
(3) $|p'(r)| \lesssim r$ and $|p''(r)| \lesssim 1$ for all $r > 0$, with constants independent of c, R,
(4) $p''(r) \geq -c$ and $\frac{1}{r} p'(r) \geq -c$, for all $r > 0$,
(5) $|r \partial_r \Delta p| \leq c$ for all $r > 0$,
(6) $\Delta^2 p(r) \leq c \cdot r^{-2}$, for all $r > 0$,
(7) $\Delta^3 p(r) \geq -c \cdot r^{-4}$ for all $r > 0$,
(8) $\Big| r \big(\frac{p'(r)}{r} \big)' \Big| \leq c$, for all $r > 0$,
(9) $\left| r \left(r \big(\frac{p'(r)}{r} \big)' \right)' \right| \leq c$ for all $r > 0$.

For each $\lambda > 0$ define $\mathcal{A}(\lambda)$ and $\mathcal{A}_0(\lambda)$ as follows,

$$[\mathcal{A}(\lambda) w](r) := p'\big(\frac{r}{\lambda}\big) \cdot \partial_r w(r), \tag{2.5}$$

$$[\mathcal{A}_0(\lambda) w](r) := \big(\frac{1}{2\lambda} p''\big(\frac{r}{\lambda}\big) + \frac{1}{2r} p'\big(\frac{r}{\lambda}\big) \big) w(r) + p'\big(\frac{r}{\lambda}\big) \cdot \partial_r w(r). \tag{2.6}$$

Note the similarity between $\mathcal{A}$ and $\frac{1}{\lambda}\Lambda$ and between $\mathcal{A}_0$ and $\frac{1}{\lambda}\Lambda_0$. Recall the notation, $\mathcal{L}_0 := -\Delta + \frac{k^2}{r^2}$.

Lemma 2.3. [9, Lemma 5.5] [12, Lemma 4.2] *Let $c_0 > 0$ be arbitrary. There exists $c > 0$ small enough and $R, \widetilde{R} > 0$ large enough in Lemma 2.2 so that the operators $\mathcal{A}(\lambda)$ and $\mathcal{A}_0(\lambda)$ defined in* (2.5) *and* (2.6) *have the following properties:*

- *the families $\{\mathcal{A}(\lambda) : \lambda > 0\}$, $\{\mathcal{A}_0(\lambda) : \lambda > 0\}$, $\{\lambda\partial_\lambda\mathcal{A}(\lambda) : \lambda > 0\}$ and $\{\lambda\partial_\lambda\mathcal{A}_0(\lambda) : \lambda > 0\}$ are bounded in $\mathscr{L}(H; L^2)$, with the bound depending only on the choice of the function $p(r)$,*
- *In addition, the operators $\mathcal{A}_0(\lambda)$ and $\lambda\partial_\lambda\mathcal{A}_0(\lambda)$ satisfy the bounds*

$$\|\partial_r\mathcal{A}_0(\lambda)w\|_{L^2} + \|r^{-1}\mathcal{A}_0(\lambda)w\|_{L^2} \lesssim \|\partial_r w\|_H + \frac{1}{\lambda}\|w\|_H$$
$$\|\partial_r\lambda\partial_\lambda\mathcal{A}_0(\lambda)w\|_{L^2} + \|r^{-1}\lambda\partial_\lambda\mathcal{A}_0(\lambda)w\|_{L^2} \lesssim \|\partial_r w\|_H + \frac{1}{\lambda}\|w\|_H$$

  with a constant that depends only on the choice of the function $p(r)$,
- *For all $w \in H \cap H^2$ we have*

$$\langle \mathcal{A}_0(\lambda)w \mid \mathcal{L}_0 w\rangle \geq -\frac{c_0}{\lambda}\|w\|_H^2 + \frac{1}{\lambda}\int_0^{R\lambda}\left((\partial_r w)^2 + \frac{k^2}{r^2}w^2\right) r\mathrm{d}r, \tag{2.7}$$

- *Moreover, for $\lambda, \mu > 0$ with $\lambda/\mu \ll 1$,*

$$\|\Lambda_0\Lambda Q_{\underline{\lambda}} - \mathcal{A}_0(\lambda)\Lambda Q_\lambda\|_{L^2} \leq c_0, \tag{2.8}$$

- *Finally, let $P_\lambda(r)$ denote the potential, $P_\lambda(r) := \frac{1}{r^2}(f'(Q_\lambda) - k^2)$. We have,*

$$\left|\langle \mathcal{A}_0(\lambda)w \mid P_\lambda(r)w\rangle - \left\langle \frac{1}{\lambda}\Lambda_0 w \mid P_\lambda(r)w\right\rangle\right| \leq \frac{c_0}{\lambda}\|w\|_H^2 \tag{2.9}$$

2.2. **Properties of the ansatz $\mathbf{\Phi}(\mu, \lambda, a, b)$.** The 2-bubble ansatz $\mathbf{\Phi}(\mu, \lambda, a, b)$ is defined in (1.9). The arguments in [12] required detailed information about $\mathbf{\Phi}$ and we recall several formulas and estimates proved there. First recall that $A, B, \widetilde{B}$ are defined so that,

$$\begin{aligned}
\mathcal{L}_\lambda(b^2 A_\lambda + \nu^k B_\lambda) &= -\frac{b^2}{\lambda}\Lambda_0\Lambda Q_{\underline{\lambda}} + \gamma_k\frac{\nu^k}{\lambda}\Lambda Q_{\underline{\lambda}} - 4\frac{(r/\mu)^k(\Lambda Q_\lambda)^2}{r^2} \\
\mathcal{L}_\mu(a^2 A_\mu + \nu^k \widetilde{B}_\mu) &= -\frac{a^2}{\mu}\Lambda_0\Lambda Q_{\underline{\mu}} - \widetilde{\gamma}_k\frac{\nu^k}{\mu}\Lambda Q_{\underline{\mu}} + 4\frac{(r/\lambda)^{-k}(\Lambda Q_\mu)^2}{r^2}
\end{aligned} \tag{2.10}$$

The existence of such $A, B, \widetilde{B}$ is made precise in the following lemma.

Lemma 2.4. [12, Lemma 3.3] *Let $k \geq 4$. There exist $C^\infty(0, \infty)$ functions $A, B, \widetilde{B}$ satisfying* (1.8)*. Moreover $A, B, \widetilde{B}$ satisfy the estimates*

$$\begin{aligned}
&A(r) = O(r^k), \quad \partial_r A(r) = O(r^{k-1}), \quad \partial_r^2 A(r) = O(r^{k-2}) \quad as \quad r \to 0 \\
&A(r) = O(r^{-k+2}), \quad \partial_r A(r) = O(r^{-k+1}), \quad \partial_r^2 A(r) = O(r^{-k}) \quad as \quad r \to \infty
\end{aligned}$$

$$\begin{aligned}
&B(r) = O(r^k), \quad \partial_r B(r) = O(r^{k-1}), \quad \partial_r^2 B(r) = O(r^{k-2}) \quad as \quad r \to 0 \\
&B(r) = O(r^{-k+2}), \quad \partial_r B(r) = O(r^{-k+1}), \quad \partial_r^2 B(r) = O(r^{-k}) \quad as \quad r \to \infty
\end{aligned}$$

$$\begin{aligned}
&\widetilde{B}(r) = O(r^k\,|\log r|), \ \partial_r\widetilde{B}(r) = O(r^{k-1}\,|\log r|), \ \partial_r^2\widetilde{B}(r) = O(r^{k-2}\,|\log r|) \quad as \quad r \to 0 \\
&\widetilde{B}(r) = O(r^{-k+2}), \quad \partial_r\widetilde{B}(r) = O(r^{-k+1}), \quad \partial_r^2\widetilde{B}(r) = O(r^{-k}) \quad as \quad r \to \infty
\end{aligned}$$

Moreover we have $A, \Lambda A, \Lambda_0\Lambda A, B, \Lambda B, \Lambda_0\Lambda B \in L^2(\mathbb{R}^2)$ and $\widetilde{B}, \Lambda\widetilde{B}, \Lambda_0\Lambda\widetilde{B} \in L^2(\mathbb{R}^2)$.

We have the following technical lemmas proved in [12].

Lemma 2.5. [12, Lemma 3.8] *Let $A, B, \widetilde{B}$ be as in Lemma 2.4, and let $\nu = \lambda/\mu \ll 1$. Then,*

$$\begin{aligned} \|r^{-1}[\Lambda Q_\lambda]^2 \Lambda Q_\mu\|_{L^2} + \|r^{-1}\Lambda Q_\lambda [\Lambda Q_\mu]^2\|_{L^2} + \|r^{-1}[\Lambda Q_\lambda]^2 A_\mu\|_{L^2} &\lesssim \nu^k \\ \|r^{-1}[\Lambda Q_\lambda]^2 \widetilde{B}_\mu\|_{L^2} &\lesssim \nu^{k-o(1)} \\ \|r^{-1}[\Lambda Q_\mu]^2 A_\lambda\|_{L^2} + \|r^{-1}[\Lambda Q_\mu]^2 B_\lambda\|_{L^2} &\lesssim \nu^{k-2} \\ \|r^{-1}\Lambda Q_\lambda [A_\mu]^2\|_{L^2} + \|r^{-1}\Lambda Q_\lambda [\widetilde{B}_\mu]^2\|_{L^2} + \|r^{-1}\Lambda Q_\mu [A_\lambda]^2\|_{L^2} + \|r^{-1}\Lambda Q_\mu [B_\lambda]^2\|_{L^2} &\lesssim \nu^k \end{aligned} \tag{2.11}$$

where the $o(1)$ above can be replaced with any small constant.

With $\mathbf{\Phi} = (\Phi, \dot{\Phi})$ defined as in (1.9) we have,

$$\begin{aligned} -\Delta\Phi + \frac{1}{r^2} f(\Phi) &= \gamma_k \frac{\nu^k}{\lambda} \Lambda Q_{\underline{\lambda}} - \frac{b^2}{\lambda} \Lambda_0 \Lambda Q_{\underline{\lambda}} + \gamma_k \frac{\nu^k}{\mu} \Lambda Q_{\underline{\mu}} + \frac{a^2}{\mu} \Lambda_0 \Lambda Q_{\underline{\mu}} \\ &\quad - \frac{1}{r^2}\Big(f(Q_\lambda - Q_\mu) - f(Q_\lambda) + f(Q_\mu) - 4\big(\frac{r}{\mu}\big)^k (\Lambda Q_\lambda)^2 - 4\big(\frac{r}{\lambda}\big)^{-k} (\Lambda Q_\mu)^2 \Big) \\ &\quad - \frac{1}{r^2}\Big(f(\Phi) - f(Q_\lambda - Q_\mu) - f'(Q_\lambda - Q_\mu)((b^2 A_\lambda + \nu^k B_\lambda) - (a^2 A_\mu + \nu^k \widetilde{B}_\mu)) \Big) \\ &\quad - \frac{1}{r^2}\Big(f'(Q_\lambda - Q_\mu)((b^2 A_\lambda + \nu^k B_\lambda) - (a^2 A_\mu + \nu^k \widetilde{B}_\mu)) \\ &\qquad - f'(Q_\lambda)(b^2 A_\lambda + \nu^k B_\lambda) + f'(Q_\mu)(a^2 A_\mu + \nu^k \widetilde{B}_\mu) \Big) \end{aligned} \tag{2.12}$$

Next we recall the estimates proved in [12].

Lemma 2.6. [12, Lemma 3.11] *Let $\mathbf{\Phi} = (\Phi, \dot{\Phi})$ be defined as in (1.9) and let (μ, λ, a, b) satisfy $\nu := \lambda/\mu \ll 1$ and $|a|, |b| \ll 1$. Then, for $\alpha = 1, 2, 3$ we have*

$$\begin{aligned} &\Big\| r^{-\alpha}\Big(f(Q_\lambda - Q_\mu) - f(Q_\lambda) + f(Q_\mu) - 4\big(\frac{r}{\mu}\big)^k (\Lambda Q_\lambda)^2 - 4\big(\frac{r}{\lambda}\big)^{-k} (\Lambda Q_\mu)^2 \Big) \Big\|_{L^2} \\ &\qquad \lesssim \nu^{2k} \lambda^{-\alpha+1} \\ &\Big\| r^{-\alpha}\Big(f(\Phi) - f(Q_\lambda - Q_\mu) - f'(Q_\lambda - Q_\mu)((b^2 A_\lambda + \nu^k B_\lambda - (a^2 A_\mu + \nu^k \widetilde{B}_\mu)) \Big) \Big\|_{L^2} \\ &\qquad \lesssim b^4 \lambda^{-\alpha+1} + a^4 \nu^{\alpha-1} \lambda^{-\alpha+1} + \nu^{2k} \lambda^{-\alpha+1} \\ &\Big\| r^{-\alpha}\Big(f'(Q_\lambda - Q_\mu)((b^2 A_\lambda + \nu^k B_\lambda - (a^2 A_\mu + \nu^k \widetilde{B}_\mu)) \\ &\qquad\quad - f'(Q_\lambda)(b^2 A_\lambda + \nu^k B_\lambda + f'(Q_\mu)(a^2 A_\mu + \nu^k \widetilde{B}_\mu) \Big) \Big\|_{L^2} \\ &\qquad \lesssim b^2 \nu^{k-1} \lambda^{-\alpha+1} + a^2 \nu^{k+\alpha-2} \lambda^{-\alpha+1} + \nu^{2k-1} \lambda^{-\alpha+1} \end{aligned} \tag{2.13}$$

Lemma 2.7. [12, Lemma 3.13] *Let $\boldsymbol{w} \in \mathcal{H} \cap \mathcal{H}^2$, let $\mathbf{\Phi} = (\Phi, \dot{\Phi})$ be defined as in (1.9), and let (μ, λ, a, b) satisfy $\nu := \lambda/\mu \ll 1$ and $|a|, |b| \ll 1$. Then,*

$$\Big\| \frac{1}{r}\Big(f(\Phi + w) - f(\Phi) - f'(\Phi) w \Big) \Big\|_{L^2} \lesssim \|w\|_H^2 \tag{2.14}$$

3. Refined Modulation Analysis

Let $\boldsymbol{u}_c(t)$ be the solution constructed in Theorem 1.12 that approaches a 2-bubble in forward time. Let $\boldsymbol{u}(t) \in \mathcal{H}$ be *any two-bubble in forward time* as in (1.5). The goal of this paper is to prove Theorem 1.1 by showing that $\boldsymbol{u}(t) = \boldsymbol{u}_c(t)$.

Since $\boldsymbol{u}(t)$ is a 2-bubble in forward time we know from (1.5) that

$$\widehat{\mathbf{d}}(\boldsymbol{u}(t)) := \inf_{\sigma, \mu > 0} \|\boldsymbol{u}(t) - (\boldsymbol{Q}_{\sigma\mu} - \boldsymbol{Q}_\mu)\|_{\mathcal{H}}^2 + \sigma^k,$$

satisfies

$$\widehat{\mathbf{d}}(\boldsymbol{u}(t)) \to 0 \ \text{ as } \ t \to \infty. \tag{3.1}$$

Note that here we have fixed the sign $\iota = +1$ in (1.5) without loss of generality. In fact, we know more than this (see [11, Theorem 1.6 and Remark 1.7]), but the above is all we require in the sequel. In [11] we used the smallness of $\mathbf{d}(\boldsymbol{u})$ to modulate around the 2-parameter family of pure two bubbles $\boldsymbol{Q}_{\sigma\mu} - \boldsymbol{Q}_{\mu}$, that, is we imposed orthogonality conditions on the difference

$$\boldsymbol{u}(t) - \boldsymbol{Q}_{\mu\sigma} - \boldsymbol{Q}_{\mu}$$

by modulating in σ and μ. In contrast, here we modulate around the 2-parameter family of maps given by the rescaled *trajectory* of the solution $\boldsymbol{u}_c(t)$ from Theorem 1.12, the two parameters being time t and the scale μ. This is natural in that this trajectory is *invariant* in the sense of dynamical systems.

Let $\boldsymbol{u}_c(t)$ be the solution given by Theorem 1.12. Note that that $\lambda_c(t)$ is monotone decreasing on the interval $[T_0, \infty)$. By Theorem 1.12 we have that

$$\widehat{\mathbf{d}}(\boldsymbol{u}_c(t)) \leq \eta_0(T_0), \quad \forall t \in [T_0, \infty)$$

where $\eta_0 = \eta_0(T_0) \to 0$ as $T_0 \to \infty$ is a constant that we can fix later to be as small as we like.

Definition 3.1. Let $\sigma_0 = \lambda_c(T_0)$ and define the inverse function

$$\lambda_c^{-1} : (0, \sigma_0] \to [T_0, \infty)$$

i.e., we can express each $t \in [T_0, \infty)$ uniquely by $t = \lambda_c^{-1}(\sigma)$ for some $\sigma \in (0, \sigma_0]$.

Define

$$\boldsymbol{U}(\mu, \sigma, \cdot) := \boldsymbol{u}_c(\lambda_c^{-1}(\sigma), \cdot)_\mu = (u_c(\lambda_c^{-1}(\sigma), \cdot/\mu), \frac{1}{\mu}\partial_t u_c(\lambda_c^{-1}(\sigma), \cdot/\mu)) \tag{3.2}$$

Then $\boldsymbol{U}$ defines a mapping,

$$(0, \infty) \times (0, \sigma_0] \ni (\mu, \sigma) \mapsto \boldsymbol{U}(\mu, \sigma, \cdot) \in \mathcal{H}$$

In fact, since the constructed solution has the threshold energy $\mathcal{E} = 2\mathcal{E}(\boldsymbol{Q})$, we can can view $\boldsymbol{U}(\mu, \sigma)$ as a 2-dimensional invariant (under the wave map flow) sub-manifold of $\{\mathcal{E} = 2\mathcal{E}(\boldsymbol{Q})\} \subset \mathcal{H}$, i.e,

$$(\mu, \sigma) \mapsto \boldsymbol{U}(\mu, \sigma) \subset \{\mathcal{E} = 2\mathcal{E}(\boldsymbol{Q})\} \subset \mathcal{H}$$

3.1. **Consequences of Theorem 1.12.** In this section we establish a collection of estimates on $\boldsymbol{U}(\mu, \sigma)$ that will be needed in the proof of Theorem 1.1. We write

$$\boldsymbol{U}(\mu, \sigma) =: (U(\mu, \sigma), \dot{U}(\mu, \sigma)).$$

Introducing the notation,

$$\xi(\sigma) := -\lambda_c'(\lambda_c^{-1}(\sigma))$$

we record formulas for the μ, σ derivatives of $\boldsymbol{U}(\mu, \sigma)$.

Lemma 3.2. *Let $\boldsymbol{U} : (0, \infty) \times (0, \eta) \to \mathcal{H}$ be defined as above with $\eta \leq \sigma_0$. Then,*

$$\partial_\mu \boldsymbol{U}(\mu, \sigma) = -\frac{1}{\mu}\boldsymbol{\Lambda}\boldsymbol{U}(\mu, \sigma) := -\frac{1}{\mu}\left([\Lambda u_c(\lambda_c^{-1}(\sigma), \cdot)]_\mu, [\Lambda_0 \partial_t u_c(\lambda_c^{-1}(\sigma), \cdot)]_{\underline{\mu}}\right) \tag{3.3}$$

$$\partial_\sigma \boldsymbol{U}(\mu, \sigma) = -\frac{\mu}{\xi(\sigma)} J \circ \mathrm{D}\,\mathcal{E}(\boldsymbol{U}(\mu, \sigma)) = -\frac{\mu}{\xi(\sigma)}\begin{pmatrix} \dot{U}(\mu, \sigma) \\ \Delta U(\mu, \sigma) - \frac{1}{r^2} f(U(\mu, \sigma)) \end{pmatrix} \tag{3.4}$$

Proof. The proof of (3.3) is a direct computation using the definition (3.2) and the definition of $\mathbf{\Lambda}$, i.e, for $\boldsymbol{V} = (V, \dot{V})$ we have

$$\mathbf{\Lambda}\boldsymbol{V} = (\Lambda V, \Lambda_0 \dot{V}).$$

To prove (3.4) we note that for any $\mu > 0$ the mapping

$$\widetilde{\boldsymbol{u}}(t) := \boldsymbol{U}(\mu, \lambda_c(t/\mu)) = [\boldsymbol{u}_c(t/\mu, \cdot)]_\mu$$

solves (1.3) on the interval $[\mu T_0, \infty)$, i.e,

$$\partial_t \widetilde{\boldsymbol{u}}(t) = J \circ \mathrm{D}\,\mathcal{E}(\widetilde{\boldsymbol{u}}(t)) \quad \forall\, t \in [\mu T_0, \infty).$$

In particular for $t = \mu\lambda_c^{-1}(\sigma)$ we have

$$J \circ \mathrm{D}\,\mathcal{E}(\widetilde{\boldsymbol{u}}(\mu\lambda_c^{-1}(\sigma))) = J \circ \mathrm{D}\,\mathcal{E}(\boldsymbol{U}(\mu, \sigma)),$$

which is the right-hand-side of (3.4) up to the factor $-\frac{\mu}{\xi(\sigma)}$. On the other hand, using the chain-rule

$$\begin{aligned}\partial_t \widetilde{\boldsymbol{u}}(t)\restriction_{t=\mu\lambda_c^{-1}(\sigma)} &= \left[\partial_\sigma \boldsymbol{U}(\mu, \lambda_c(t/\mu))\lambda_c'(t/\mu)\frac{1}{\mu}\right]\restriction_{t=\mu\lambda_c^{-1}(\sigma)} \\ &= \partial_\sigma \boldsymbol{U}(\mu, \sigma)\lambda_c'(\lambda_c^{-1}(\sigma))\frac{1}{\mu} = -\frac{\xi(\sigma)}{\mu}\partial_\sigma \boldsymbol{U}(\mu, \sigma)\end{aligned}$$

which establishes the claim. □

Importantly, Theorem 1.12 yields refined regularity and decay information about $\boldsymbol{U}(\mu, \sigma)$ that will be crucial in the proof of Theorem 1.1. First we fix notation. We write,

$$\boldsymbol{U}(\mu, \sigma) = \mathbf{\Phi}(\mu, \sigma) + \boldsymbol{w}_c(\mu, \sigma) \tag{3.5}$$

where we have defined,

$$\begin{aligned}\mathbf{\Phi}(\mu, \sigma) &:= \mathbf{\Phi}(\mu_c(\lambda_c^{-1}(\sigma)), \sigma, a_c(\lambda_c^{-1}(\sigma)), b_c(\lambda_c^{-1}(\sigma)))_\mu \\ \boldsymbol{w}_c(\mu, \sigma) &:= \boldsymbol{w}_c(\lambda_c^{-1}(\sigma))_\mu\end{aligned}$$

Here $\boldsymbol{w}_c(\mu, \sigma) := \boldsymbol{w}_c(\lambda_c^{-1}(\sigma))_\mu$ is as in Theorem 1.12 and $\mathbf{\Phi}(\mu, \lambda, a, b)$ is defined in (1.9). Next, define $\boldsymbol{g}_c(\mu, \sigma)$ via,

$$\begin{aligned}\boldsymbol{g}_c(\mu, \sigma) &:= \mathbf{\Phi}(\mu, \sigma) + \boldsymbol{w}_c(\mu, \sigma) \\ &\quad - \boldsymbol{Q}_{\mu\sigma} + \boldsymbol{Q}_{\mu_c(\lambda_c^{-1}(\sigma))\mu} - (0, b_c(\lambda_c^{-1}(\sigma))\Lambda Q_{\underline{\mu\sigma}} + a_c(\lambda_c^{-1}(\sigma))\Lambda Q_{\underline{\mu_c(\lambda_c^{-1}(\sigma))\mu}})\end{aligned}$$

so that we may also write,

$$\begin{aligned}\boldsymbol{U}(\mu, \sigma) &= (Q_{\mu\sigma}, b_c(\lambda_c^{-1}(\sigma))\Lambda Q_{\underline{\mu\sigma}}) - (Q_{\mu_c(\lambda_c^{-1}(\sigma))\mu}, -a_c(\lambda_c^{-1}(\sigma))\Lambda Q_{\underline{\mu_c(\lambda_c^{-1}(\sigma))\mu}}) \\ &\quad + \boldsymbol{g}_c(\mu, \sigma)\end{aligned} \tag{3.6}$$

First we translate the main estimates from Theorem 1.12 to estimates for the error $\boldsymbol{w}_c(\mu, \sigma)$ and for the parameters $\xi(\sigma), b_c(\lambda_c^{-1}(\sigma)), \mu_c(\lambda_c^{-1}(\sigma))$ and $a_c(\lambda_c^{-1}(\sigma))$.

Corollary 3.3. *Let $\boldsymbol{U}(\mu, \sigma)$, $\boldsymbol{w}_c(\mu, \sigma)$, and $\boldsymbol{g}_c(\mu, \sigma)$ be defined as above. Then, $\boldsymbol{U}(\mu, \sigma) \in \mathcal{H} \cap \mathcal{H}^2 \cap \Lambda\mathcal{H}^2$ and we have the estimates,*

$$\|\boldsymbol{w}_c(\mu, \sigma)\|_{\mathcal{H}} \lesssim \sigma^{\frac{3}{2}k-1} \tag{3.7}$$

$$\|\boldsymbol{w}_c(\mu, \sigma)\|_{\mathcal{H}^2} \lesssim \mu^{-1}\sigma^{\frac{3}{2}k-2} \tag{3.8}$$

$$\|\mathbf{\Lambda}\boldsymbol{w}_c(\mu, \sigma)\|_{\mathcal{H}} \lesssim \sigma^{k-1} \tag{3.9}$$

uniformly in $\sigma \in (0, \sigma_0]$ and $\mu \in (0, \infty)$. We have the asymptotics,

$$\lim_{\sigma\to 0^+} \frac{\xi(\sigma)}{\rho_k \sigma^{\frac{k}{2}}} = \lim_{\sigma\to 0^+} \frac{b_c(\lambda_c^{-1}(\sigma))}{\rho_k \sigma^{\frac{k}{2}}} = 1, \tag{3.10}$$

$$\left|\mu_c(\lambda_c^{-1}(\sigma)) - 1\right| = o(1)\sigma \quad as \quad \sigma \to 0^+, \tag{3.11}$$

and

$$\left|a_c(\lambda_c^{-1}(\sigma))\right| = o(1)\sigma^{\frac{k}{2}} \quad as \quad \sigma \to 0^+. \tag{3.12}$$

In particular, the above additionally yield the less-refined estimates,

$$\|\boldsymbol{g}_c(\mu,\sigma)\|_{\mathcal{H}} \lesssim \sigma^k \tag{3.13}$$

$$\|\boldsymbol{\Lambda g}_c(\mu,\sigma)\|_{\mathcal{H}} \lesssim \sigma^k \tag{3.14}$$

where the $o(1)$ in all of the above denotes a constant that can be made as small as we like by taking T_0 large enough (and hence σ_0 small enough).

Corollary 3.4. *Let $\boldsymbol{U}(\mu,\sigma)$ be as above. Then,*

$$\|\frac{1}{\rho_k\sigma^{\frac{k}{2}}}\dot{U}(\mu,\sigma) - \Lambda Q_{\underline{\mu\sigma}}\|_{L^2} = o(1) \quad as \quad \sigma \to 0 \tag{3.15}$$

$$\|\frac{1}{\rho_k\sigma^{\frac{k}{2}}}\dot{U}(\mu,\sigma) - \Lambda Q_{\underline{\mu\sigma}}\|_{H} = \frac{o(1)}{\mu\sigma} \quad as \quad \sigma \to 0 \tag{3.16}$$

$$\|r(-\Delta U(\mu,\sigma) + r^{-2}f(U(\mu,\sigma)))\|_{L^2} \lesssim \sigma^{k-1} \quad as \quad \sigma \to 0 \tag{3.17}$$

$$\| - \Delta U(\mu,\sigma) + r^{-2}f(U(\mu,\sigma))\|_{L^2} \lesssim \frac{\sigma^k}{\mu\sigma} \quad as \quad \sigma \to 0 \tag{3.18}$$

Moreover, for any $\boldsymbol{h} \in \mathcal{H}$ we have,

$$\left|\left\langle D^2\mathcal{E}(\boldsymbol{U}(\mu,\sigma))\partial_\mu \boldsymbol{U}(\mu,\sigma) \mid \boldsymbol{h}\right\rangle\right| \leq o(1)\sigma^{\frac{k}{2}}\frac{\|\boldsymbol{h}\|_{\mathcal{H}}}{\mu\sigma} \tag{3.19}$$

$$\left|\left\langle D^2\mathcal{E}(\boldsymbol{U}(\mu,\sigma))\partial_\sigma \boldsymbol{U}(\mu,\sigma) - (0, \frac{\sigma^{\frac{k}{2}}}{\rho_k\sigma}(\gamma_k\Lambda Q_{\underline{\mu\sigma}} - \rho_k^2\Lambda_0\Lambda Q_{\underline{\mu\sigma}})) \mid \boldsymbol{h}\right\rangle\right| \leq o(1)\sigma^{\frac{k}{2}}\frac{\|\boldsymbol{h}\|_{\mathcal{H}}}{\sigma} \tag{3.20}$$

where $o(1)$ can be replaced with any small constant by taking σ small enough.

Before proving Corollary 3.4 it will be convenient to first translate estimates for the ansatz $\boldsymbol{\Phi}(\mu,\lambda,a,b)$ into estimates for $\boldsymbol{\Phi}(\mu,\sigma)$.

Lemma 3.5. *Let $\boldsymbol{\Phi}(\mu,\sigma)$ be defined as above. Then,*

$$\|\dot{\Phi}(\mu,\sigma)\|_{L^2} + \|\Lambda_0\dot{\Phi}(\mu,\sigma)\|_{L^2} \lesssim \sigma^{\frac{k}{2}} \tag{3.21}$$

$$\begin{gathered}\|\dot{\Phi}(\mu,\sigma) - b_c(\lambda_c^{-1}(\sigma))\Lambda Q_{\underline{\mu\sigma}}\|_{L^2} \lesssim o(1)\sigma^{\frac{k}{2}} \\ \left|\left\langle \Lambda Q_{\underline{\mu\sigma}} \mid \dot{\Phi}(\mu,\sigma) - b_c(\lambda_c^{-1}(\sigma))\Lambda Q_{\underline{\mu\sigma}}\right\rangle\right| \lesssim o(1)\sigma^{\frac{3}{2}k-1}\end{gathered} \tag{3.22}$$

$$\|\dot{\Phi}(\mu,\sigma) - b_c(\lambda_c^{-1}(\sigma))\Lambda Q_{\underline{\mu\sigma}}\|_{H} \lesssim o(1)\frac{\sigma^{\frac{k}{2}}}{\mu\sigma} \tag{3.23}$$

$$\begin{gathered}\|\dot{\Phi}(\mu,\sigma) - a_c(\lambda_c^{-1}(\sigma))\Lambda Q_{\underline{\mu\mu_c(\lambda_c^{-1}(\sigma))}}\|_{L^2} \lesssim \sigma^{\frac{k}{2}} \\ \left|\left\langle \Lambda Q_{\underline{\mu}} \mid \dot{\Phi}(\mu,\sigma) - a_c(\lambda_c^{-1}(\sigma))\Lambda Q_{\underline{\mu\mu_c(\lambda_c^{-1}(\sigma))}}\right\rangle\right| \lesssim \sigma^{\frac{3}{2}k-1}\end{gathered} \tag{3.24}$$

$$\|r(\Delta\Phi(\mu,\sigma) - r^{-2}f(\Phi(\mu,\sigma)))\|_{L^2} \lesssim \sigma^k \tag{3.25}$$

$$\|\Delta\Phi(\mu,\sigma) - r^{-2}f(\Phi(\mu,\sigma))\|_{L^2} \lesssim \frac{\sigma^k}{\mu\sigma} \tag{3.26}$$

where the $o(1)$ in all of the above denotes a constant that can be made as small as we like by taking T_0 large enough (and hence σ_0 small enough).

Proof of Lemma 3.5. The estimate (3.21) the estimate (3.23) and the first estimates in (3.22) and (3.24) follow directly from the definition of $\dot{\Phi}(\mu,\sigma)$ along with the estimates (3.10), (3.11), and (3.12). The second estimates in (3.22) and (3.24) follow from the same considerations along with [12, Lemma 3.5], which contains the standard estimates regarding the pairings of $A, B, \widetilde{B}, \Lambda Q$ in L^2 at different scales.

To prove (3.25) and (3.26) we record the formula,

$$\begin{aligned}
&-\Delta\Phi(\mu,\sigma)+\frac{1}{r^2}f(\Phi(\mu,\sigma))\\
&\qquad=\gamma_k\frac{\sigma^k}{\mu_c(\lambda_c^{-1}(\sigma))^k\mu\sigma}\Lambda Q_{\underline{\mu\sigma}}-\frac{b_c(\lambda_c^{-1}(\sigma))^2}{\mu\sigma}\Lambda_0\Lambda Q_{\underline{\mu\sigma}}\\
&\qquad+\gamma_k\frac{\sigma^k}{\mu_c(\lambda_c^{-1}(\sigma))^{k+1}\mu}\Lambda Q_{\underline{\mu_c(\lambda_c^{-1}(\sigma))\mu}}+\frac{a_c(\lambda_c^{-1}(\sigma))^2}{\mu_c(\lambda_c^{-1}(\sigma))\mu}\Lambda_0\Lambda Q_{\underline{\mu_c(\lambda_c^{-1}(\sigma))\mu}}\\
&\qquad-\frac{1}{r^2}\Big(f(Q_{\mu\sigma}-Q_{\mu_c(\lambda_c^{-1}(\sigma))\mu})-f(Q_{\mu\sigma})+f(Q_{\mu_c(\lambda_c^{-1}(\sigma))\mu})\\
&\qquad\qquad-4\Big(\frac{r}{\mu_c(\lambda_c^{-1}(\sigma))\mu}\Big)^k(\Lambda Q_{\mu\sigma})^2-4\Big(\frac{r}{\mu\sigma}\Big)^{-k}(\Lambda Q_{\mu_c(\lambda_c^{-1}(\sigma))\mu})^2\Big)\\
&\qquad-\frac{1}{r^2}\Big(f(\Phi)-f(Q_{\mu\sigma}-Q_{\mu_c(\lambda_c^{-1}(\sigma))\mu})\\
&\qquad\qquad-f'(Q_{\mu\sigma}-Q_{\mu_c(\lambda_c^{-1}(\sigma))\mu})\big(b_c(\lambda_c^{-1}(\sigma))^2T_{\mu\sigma}-a_c(\lambda_c^{-1}(\sigma))^2\widetilde{T}_{\mu_c(\lambda_c^{-1}(\sigma))\mu}\big)\Big)\\
&\qquad-\frac{1}{r^2}\Big(f'(Q_{\mu\sigma}-Q_{\mu_c(\lambda_c^{-1}(\sigma))\mu})\big(b_c(\lambda_c^{-1}(\sigma))^2T_{\mu\sigma}-a_c(\lambda_c^{-1}(\sigma))^2\widetilde{T}_{\mu_c(\lambda_c^{-1}(\sigma))\mu}\big)\\
&\qquad\qquad-f'(Q_{\mu\sigma})b_c(\lambda_c^{-1}(\sigma))^2T_{\mu\sigma}+f'(Q_{\mu_c(\lambda_c^{-1}(\sigma))\mu})a_c(\lambda_c^{-1}(\sigma))^2\widetilde{T}_{\mu_c(\lambda_c^{-1}(\sigma))\mu}\Big)
\end{aligned}$$

which is a rescaling of (2.12). The estimate (3.25) now follows from (2.10) and (2.13) with $\alpha=1$ along with (3.10), (3.11), and (3.12). The estimate (3.26) follows from the same considerations using (2.10) and (2.13) with $\alpha=2$. □

Proof. First we note that (3.15) is a direct consequence of (3.22) along with (3.10) and (3.7). The estimate (3.16) is a direct consequence of (3.23) along with (3.10) and (3.8).

To prove (3.17) and (3.18) we use the decomposition (3.5) to write,

$$\begin{aligned}
-\Delta U(\mu,\sigma)+r^{-2}f(U(\mu,\sigma))&=-\Delta\Phi(\mu,\sigma)+r^{-2}f(\Phi(\mu,\sigma))\\
&+r^{-2}(f(\Phi(\mu,\sigma)+w_c(\mu,\sigma))-f(\Phi(\mu,\sigma))-f'(\Phi(\mu,\sigma))w_c(\mu,\sigma))\\
&-\Delta w_c(\mu,\sigma)+r^{-2}f'(\Phi(\mu,\sigma))w_c(\mu,\sigma)
\end{aligned}\tag{3.27}$$

First we prove (3.17). For the first line in (3.27) we apply (3.25). For the second line we use (2.14) along with (3.7) to obtain,

$$\|r^{-1}(f(\Phi(\mu,\sigma)+w_c(\mu,\sigma))-f(\Phi(\mu,\sigma))-f'(\Phi(\mu,\sigma))w_c(\mu,\sigma))\|_{L^2}\lesssim\|w_c(\mu,\sigma)\|_H^2\lesssim\sigma^{3k-2}$$

Finally, for the last line in (3.27) we have the estimates,

$$\|r\Delta w_c(\mu,\sigma)\|_{L^2}\lesssim\|\Lambda w_c(\mu,\sigma)\|_H\lesssim\sigma^{k-1}$$

which follows from (3.9) and

$$\|r^{-1}f'(\Phi(\mu,\sigma))w_c(\mu,\sigma)\|_{L^2}\lesssim\|w_c(\mu,\sigma)\|_H\lesssim\sigma^{\frac{3}{2}k-1}$$

which follows from (3.7). This proves (3.17). The proof of (3.18) is similar.

Next, we prove (3.19). First, using (3.3) we have

$$\big\langle D^2\mathcal{E}(\boldsymbol{U}(\mu,\sigma)\partial_\mu\boldsymbol{U}(\mu,\sigma)\mid\boldsymbol{h}\big\rangle=-\frac{1}{\mu}\big\langle\mathcal{L}_{U(\mu,\sigma)}\Lambda U(\mu,s)\mid h\big\rangle-\frac{1}{\mu}\Big\langle\Lambda_0\dot{U}(\mu,\sigma)\mid\dot{h}\Big\rangle\tag{3.28}$$

To treat the second term on the right above we use (3.5) to estimate,

$$\left|\frac{1}{\mu}\left\langle \Lambda_0\dot{U}(\mu,\sigma) \mid \dot{h}\right\rangle\right| \lesssim \frac{1}{\mu}(\|\Lambda_0\dot{\Phi}(\mu,\sigma)\|_{L^2}+\|\Lambda_0\dot{w}_c(\mu,\sigma)\|_{L^2})\|\dot{h}\|_{L^2} \lesssim \frac{\sigma^{\frac{k}{2}}}{\mu}\|\dot{h}\|_{L^2}$$

where we have used the estimates (3.21) and (3.9) in the second inequality above. To handle the first term on the right of (3.28) we make use of the fact that $\mathcal{L}\Lambda Q = 0$ along with the decomposition (3.6) to write,

$$\begin{aligned}\mathcal{L}_{U(\mu,\sigma)}\Lambda U(\mu,\sigma) = (\mathcal{L}_{U(\mu,\sigma)}-\mathcal{L}_{\mu\sigma})\Lambda Q_{\mu\sigma} + (\mathcal{L}_{U(\mu,\sigma)}-\mathcal{L}_{\mu_c(\lambda_c^{-1}(\sigma))\mu})\Lambda Q_{\mu_c(\lambda_c^{-1}(\sigma))\mu}\\ + \mathcal{L}_{U(\mu,\sigma)}\Lambda g_c(\mu,\sigma)\end{aligned}$$

We use the weighted estimate (3.14) to treat the contribution of last term above,

$$\left|\left\langle \frac{1}{\mu}\mathcal{L}_{U(\mu,\sigma)}\Lambda g_c(\mu,\sigma) \mid h\right\rangle\right| \lesssim \frac{1}{\mu}\|\Lambda g_c(\mu,\sigma)\|_H\|h\|_H \lesssim \frac{\sigma^k}{\mu}\|h\|_H$$

To treat the first term note that,

$$\begin{aligned}(\mathcal{L}_{U(\mu,\sigma)}-\mathcal{L}_{\mu\sigma})\Lambda Q_{\mu\sigma} &= r^{-2}\Big(f'(U(\mu,\sigma))-f'(Q_{\mu\sigma})\Big)\Lambda Q_{\mu\sigma}\\ &= r^{-2}\Big(f'(Q_{\mu\sigma}-Q_{\mu_c(\lambda_c^{-1}(\sigma))\mu})-f'(Q_{\mu\sigma})\Big)\Lambda Q_{\mu\sigma} + r^{-2}O(g_c(\mu,\sigma)\Lambda Q_{\mu\sigma})\end{aligned} \tag{3.29}$$

Using the pointwise estimate,

$$\left|f'(Q_{\mu\sigma}-Q_{\mu_c(\lambda_c^{-1}(\sigma))\mu})-f'(Q_{\mu\sigma})\right|\Lambda Q_{\mu\sigma} \lesssim \Lambda Q^2_{\mu_c(\lambda_c^{-1}(\sigma))\mu}\Lambda Q_{\mu\sigma} + \Lambda Q_{\mu_c(\lambda_c^{-1}(\sigma))\mu}\Lambda Q^2_{\mu\sigma}$$

we deduce that,

$$\begin{aligned}&\left|\frac{1}{\mu}\left\langle r^{-2}\Big(f'(Q_{\mu\sigma}-Q_{\mu_c(\lambda_c^{-1}(\sigma))\mu})-f'(Q_{\mu\sigma})\Big)\Lambda Q_{\mu\sigma} \mid h\right\rangle\right|\\ &\qquad\lesssim \frac{1}{\mu}\Big(\|r^{-1}\Lambda Q^2_{\mu_c(\lambda_c^{-1}(\sigma))}\Lambda Q_\sigma\|_{L^2}+\|\Lambda Q_{\mu_c(\lambda_c^{-1}(\sigma))}\Lambda Q^2_\sigma\|_{L^2}\Big)\|h\|_H \lesssim \frac{\sigma^k}{\mu}\|h\|_H\end{aligned}$$

We can also use (3.13) to deduce that,

$$\left|\frac{1}{\mu}\left\langle r^{-2}g_c(\mu,\sigma)\Lambda Q_{\mu\sigma} \mid h\right\rangle\right| \lesssim \frac{\sigma^k}{\mu}\|h\|_H$$

where in the last line we used (3.11). And thus,

$$\left|\frac{1}{\mu}\left\langle (\mathcal{L}_{U(\mu,\sigma)}-\mathcal{L}_{\mu\sigma})\Lambda Q_{\mu\sigma} \mid h\right\rangle\right| \lesssim \frac{\sigma^k}{\mu}\|h\|_H$$

The second term in (3.28) is treated in the same way. This completes the proof of (3.19).

Lastly, we prove (3.20). Using (3.4) we have

$$\begin{aligned}\left\langle D^2\mathcal{E}(\boldsymbol{U}(\mu,\sigma))\partial_\sigma\boldsymbol{U}(\mu,\sigma) \mid \boldsymbol{h}\right\rangle &= -\frac{\mu}{\xi(\sigma)}\left\langle \Delta U(\mu,\sigma)-r^{-2}f(U(\mu,\sigma)) \mid \dot{h}\right\rangle\\ &\quad -\frac{\mu}{\xi(\sigma)}\left\langle \mathcal{L}_{U(\mu,\sigma)}\dot{U}(\mu,\sigma) \mid h\right\rangle\end{aligned} \tag{3.30}$$

For the first term on the right above we recall the decomposition (3.27) to obtain,

$$\begin{aligned}-\frac{\mu}{\xi(\sigma)}\left\langle \Delta U(\mu,\sigma)-r^{-2}f(U(\mu,\sigma)) \mid \dot{h}\right\rangle &= -\frac{\mu}{\xi(\sigma)}\left\langle -\Delta\Phi(\mu,\sigma)+r^{-2}f(\Phi(\mu,\sigma)) \mid \dot{h}\right\rangle\\ &\quad -\frac{\mu}{\xi(\sigma)}\left\langle r^{-2}(f(\Phi(\mu,\sigma)+w_c(\mu,\sigma))-f(\Phi(\mu,\sigma))-f'(\Phi(\mu,\sigma))w_c(\mu,\sigma)) \mid \dot{h}\right\rangle\\ &\quad -\frac{\mu}{\xi(\sigma)}\left\langle -\Delta w_c(\mu,\sigma) \mid \dot{h}\right\rangle - \frac{\mu}{\xi(\sigma)}\left\langle r^{-2}f'(\Phi(\mu,\sigma))w_c(\mu,\sigma) \mid \dot{h}\right\rangle\end{aligned}$$

For the first line above we use the expansion (3.17) along with the estimates (2.13) with $\alpha = 2$ and the asymptotics (3.10),(3.11), and (3.12) to obtain,

$$\begin{aligned}
&-\frac{\mu}{\xi(\sigma)}\Big\langle -\Delta\Phi(\mu,\sigma) + r^{-2}f(\Phi(\mu,\sigma)) \mid \dot h\Big\rangle \\
&\qquad = \gamma_k \frac{\sigma^{\frac{k}{2}}}{\rho_k \sigma}\Big\langle \Lambda Q_{\underline{\mu\sigma}} \mid \dot h\Big\rangle - \frac{\rho_k \sigma^{\frac{k}{2}}}{\sigma}\Big\langle \Lambda_0 \Lambda Q_{\underline{\mu\sigma}} \mid \dot h\Big\rangle + o(1)\frac{\sigma^{\frac{k}{2}}}{\sigma}\|\dot h\|_{L^2}
\end{aligned}$$

which reveals the leading order terms that appear on the left-hand side of (3.20). For the next line we have,

$$\begin{aligned}
&|\frac{\mu}{\xi(\sigma)}\Big\langle r^{-2}(f(\Phi(\mu,\sigma)+w_c(\mu,\sigma)) - f(\Phi(\mu,\sigma)) - f'(\Phi(\mu,\sigma))w_c(\mu,\sigma)) \mid \dot h\Big\rangle| \\
&\qquad \lesssim \frac{\mu}{\xi(\sigma)}\|r^{-2}w_c(\mu,\sigma)^2\|_{L^2}\|\dot h\|_{L^2} \lesssim \sigma^{-\frac{k}{2}}\mu\|w_c(\mu,\sigma)\|_H\|w_c(\mu,\sigma)\|_{H^2}\|\dot h\|_{L^2} \lesssim \sigma^{\frac{5}{2}k-3}\|\dot h\|_{L^2}
\end{aligned}$$

where in the last line we used (3.7) and (3.8). Finally, for the last line we again use (3.8) to obtain,

$$\begin{aligned}
&\Big|\frac{\mu}{\xi(\sigma)}\Big\langle -\Delta w_c(\mu,\sigma) \mid \dot h\Big\rangle\Big| \lesssim \mu\sigma^{-\frac{k}{2}}\|w_c(\mu,\sigma)\|_{H^2}\|\dot h\|_{L^2} \lesssim \sigma^{k-1}\frac{\|\dot h\|_{L^2}}{\sigma} \\
&\Big|\frac{\mu}{\xi(\sigma)}\Big\langle r^{-2}f'(\Phi(\mu,\sigma))w_c(\mu,\sigma) \mid \dot h\Big\rangle\Big| \lesssim \sigma^{k-1}\frac{\|\dot h\|_{L^2}}{\sigma}
\end{aligned}$$

which completes the estimates for the first term in (3.30). To treat the second term in (3.30) we expand using the fact that $\mathcal{L}\Lambda Q = 0$ as follows,

$$\begin{aligned}
\mathcal{L}_{U(\mu,\sigma)}\dot U(\mu,\sigma) &= b_c(\lambda_c^{-1}(\sigma))(\mathcal{L}_{U(\mu,\sigma)} - \mathcal{L}_{\mu\sigma})\Lambda Q_{\underline{\mu\sigma}} \\
&\quad + a_c(\lambda_c^{-1}(\sigma))(\mathcal{L}_{U(\mu,\sigma)} - \mathcal{L}_{\mu\mu_c(\lambda_c^{-1}(\sigma))})\Lambda Q_{\underline{\mu_c(\lambda_c^{-1}(\sigma))\mu}} \\
&\quad + \mathcal{L}_{U(\mu\sigma)}\Big(\dot\Phi(\mu,\sigma) - b_c(\lambda_c^{-1}(\sigma))\Lambda Q_{\underline{\mu\sigma}} - a_c(\lambda_c^{-1}(\sigma))\Lambda Q_{\underline{\mu_c(\lambda_c^{-1}(\sigma))\mu}}\Big) \\
&\quad + \mathcal{L}_{U(\mu\sigma)}\dot w_c(\mu,\sigma)
\end{aligned} \tag{3.31}$$

The contribution of the last term above to (3.20) is controlled by the estimate (3.8),

$$\frac{\mu}{\xi(\sigma)}\left|\left\langle \mathcal{L}_{U(\mu,s)}\dot w_c(\mu,\sigma) \mid h\right\rangle\right| \lesssim \mu\sigma^{-\frac{k}{2}}\|\dot w_c(\mu,\sigma)\|_H\|h\|_H \lesssim \sigma^{k-1}\frac{\|h\|_H}{\sigma}$$

Next, consider the first line in (3.31). Using (3.29), (3.10) we have,

$$\begin{aligned}
\frac{\mu}{\xi(\sigma)}\Big|b_c(\lambda_c^{-1}(\sigma))\langle(\mathcal{L}_{U(\mu,\sigma)} - \mathcal{L}_{\mu\sigma})\Lambda Q_{\underline{\mu\sigma}} \mid h\rangle\Big| &\lesssim \frac{\mu}{\mu\sigma}\|r^{-1}\Lambda Q^2_{\mu_c(\lambda_c^{-1}(\sigma))\mu}\Lambda Q_{\mu\sigma}\|_{L^2}\|h\|_H \\
&\quad + \frac{\mu}{\mu\sigma}\|r^{-1}\Lambda Q_{\mu_c(\lambda_c^{-1}(\sigma))\mu}\Lambda Q^2_{\mu\sigma}\|_{L^2}\|h\|_H + \frac{\mu}{\mu\sigma}\|g_c(\mu,\sigma)\|_H\|h\|_H \\
&\lesssim \sigma^k\frac{\|h\|_H}{\sigma}
\end{aligned}$$

where the last inequality follows from the estimates (2.11) from Lemma 2.5 along with the estimate (3.13). The contribution of the second line in (3.31) to (3.20) is handled similarly. Finally, the estimate,

$$\frac{\mu}{\xi(\sigma)}\Big|\Big\langle \mathcal{L}_{U(\mu\sigma)}\Big(\dot\Phi(\mu,\sigma) - b_c(\lambda_c^{-1}(\sigma))\Lambda Q_{\underline{\mu\sigma}} - a_c(\lambda_c^{-1}(\sigma))\Lambda Q_{\underline{\mu_c(\lambda_c^{-1}(\sigma))\mu}}\Big) \mid h\Big\rangle\Big| \lesssim \sigma^k\frac{\|h\|_H}{\sigma}$$

follows directly from the definition of $\dot\Phi(\mu,\sigma)$ along with (3.10), (3.11), and (3.12). Plugging the preceding estimates back into (3.30) we obtains,

$$\big\langle D^2\mathcal{E}(\boldsymbol{U}(\mu,\sigma))\partial_\sigma \boldsymbol{U}(\mu,\sigma) \mid \boldsymbol{h}\big\rangle = \gamma_k\frac{\sigma^{\frac{k}{2}}}{\rho_k\sigma}\Big\langle \Lambda Q_{\underline{\mu\sigma}} \mid \dot h\Big\rangle - \frac{\rho_k\sigma^{\frac{k}{2}}}{\sigma}\Big\langle \Lambda_0\Lambda Q_{\underline{\mu\sigma}} \mid \dot h\Big\rangle + o(1)O(\frac{\sigma^{\frac{k}{2}}}{\sigma}\|\boldsymbol{h}\|_{\mathcal{H}})$$

as claimed. □

3.2. **Modulation around $\boldsymbol{U}(\mu,\sigma)$.** Next we modulate around $\boldsymbol{U}(\mu,\sigma)$.

Lemma 3.6 (Modulation Lemma)**.** *There exists an $\eta_0 > 0$ small enough so that the following statement holds true. Let J be a time interval, and let $\boldsymbol{u} : J \to \mathcal{H}$ be a solution to* (1.1) *such that*

$$\widehat{\mathbf{d}}(\boldsymbol{u}(t)) < \eta \leq \eta_0 \quad \forall t \in J$$

Then there exist $C^1(J;(0,\infty))$ functions $\mu(t), \sigma(t)$ such that defining $\boldsymbol{g}(t) \in \mathcal{H}$ by

$$\boldsymbol{g}(t) = \boldsymbol{u}(t) - \boldsymbol{U}(\mu(t),\sigma(t)) \tag{3.32}$$

we have, for each $t \in J$,

$$\Big\langle \Lambda Q_{\underline{\mu}} \mid g \Big\rangle = 0 \tag{3.33}$$

$$\Big\langle \Lambda Q_{\underline{\sigma\mu}} \mid g \Big\rangle = 0 \tag{3.34}$$

In addition, there exists a uniform constant $c_1 > 0$ such that

$$\big\langle D^2\mathcal{E}(\boldsymbol{U}(\mu(t),\sigma(t))\boldsymbol{g}(t) \mid \boldsymbol{g}(t)\big\rangle \geq c_1 \|\boldsymbol{g}(t)\|_{\mathcal{H}_0}^2 \tag{3.35}$$

Finally, we have the estimates,

$$|\mu'(t)| \lesssim \|\dot g\|_{L^2} + \sigma^{\frac{k}{2}}\|g\|_H \lesssim \|\boldsymbol{g}\|_{\mathcal{H}} \tag{3.36}$$

$$\Big|\mu(t)\sigma'(t) + \xi(\sigma(t)) + \frac{\Big\langle \Lambda Q_{\underline{\lambda(t)}} \mid \dot g \Big\rangle}{\|\Lambda Q\|_{L^2}^2}\Big| \lesssim o(1)\|\dot g\|_{L^2} + \sigma^{\frac{k}{2}}\|g\|_H \lesssim o(1)\|\boldsymbol{g}\|_{\mathcal{H}} \tag{3.37}$$

where the $o(1)$ term above can be taken as small as we like by taking $\eta > 0$ small.

Remark 3.7. By (3.1) we may apply Lemma 3.6 to the arbitrary 2-bubble solution $\boldsymbol{u}(t)$ on the time interval $J = [T_0, \infty)$ for large enough $T_0 > 0$, and we obtain a decomposition

$$\boldsymbol{u}(t) = \boldsymbol{U}(\mu(t),\sigma(t)) + \boldsymbol{g}(t), \quad \forall\ t \in J$$

Note that by (3.1) and the proof of Lemma 3.6 we obtain the qualitative behavior,

$$\|\boldsymbol{g}(t)\|_{\mathcal{H}}^2 + \sigma(t)^k \to 0 \ \text{ as } \ t \to \infty.$$

Before proving Lemma 3.6 we record several identities that will be used throughout the rest of the paper. We use the notation

$$\boldsymbol{g}(t) =: (g(t), \dot g(t)), \quad \boldsymbol{U}(\mu,\sigma) := (U(\mu,\sigma), \dot U(\mu,\sigma))$$

The equation satisfied by $\boldsymbol{g}(t)$ as defined in (3.32) is

$$\begin{aligned}
\partial_t \boldsymbol{g}(t) &= \partial_t \boldsymbol{u}(t) - \partial_t \boldsymbol{U}(\mu(t),\sigma(t)) \\
&= J \circ \mathrm{D}\,\mathcal{E}(\boldsymbol{u}(t)) - \mu'(t)\partial_\mu \boldsymbol{U}(\mu,\sigma) - \sigma'(t)\partial_\sigma \boldsymbol{U}(\mu,\sigma) \\
&= J \circ \Big(\mathrm{D}\,\mathcal{E}(\boldsymbol{U}(\mu,\sigma)+\boldsymbol{g}) - \mathrm{D}\,\mathcal{E}(\boldsymbol{U}(\mu,\sigma))\Big) \\
&\quad - \mu'\partial_\mu \boldsymbol{U}(\mu,\sigma) - \big(\sigma'\partial_\sigma \boldsymbol{U}(\mu,\sigma) - J \circ \mathrm{D}\,\mathcal{E}(\boldsymbol{U}(\mu,\sigma))\big)
\end{aligned}$$

We use (3.4) to rewrite the last line above, obtaining,

$$\begin{aligned}
\partial_t \boldsymbol{g}(t) &= J \circ \Big(\mathrm{D}\,\mathcal{E}(\boldsymbol{U}(\mu,\sigma)+\boldsymbol{g}) - \mathrm{D}\,\mathcal{E}(\boldsymbol{U}(\mu,\sigma))\Big) \\
&\quad - \mu'\partial_\mu \boldsymbol{U}(\mu,\sigma) - \big(\sigma' + \frac{\xi(\sigma)}{\mu}\big)\partial_\sigma \boldsymbol{U}(\mu,\sigma) \\
&= J \circ \Big(\mathrm{D}\,\mathcal{E}(\boldsymbol{U}(\mu,\sigma)+\boldsymbol{g}) - \mathrm{D}\,\mathcal{E}(\boldsymbol{U}(\mu,\sigma))\Big) \\
&\quad - \mu'\partial_\mu \boldsymbol{U}(\mu,\sigma) - \Big(\sigma' + \frac{\rho_k \sigma^{\frac{k}{2}}}{\mu}\Big)\partial_\sigma \boldsymbol{U}(\mu,\sigma) + \frac{1}{\mu}\big(\rho_k\sigma^{\frac{k}{2}} - \xi(\sigma)\big)\partial_\sigma \boldsymbol{U}(\mu,\sigma)
\end{aligned} \tag{3.38}$$

In components this reads,

$$\partial_t \begin{pmatrix} g \\ \dot g \end{pmatrix} = \begin{pmatrix} \dot g - \mu' \partial_\mu U(\mu,\sigma) - (\sigma' + \frac{\xi(\sigma)}{\mu})\partial_\sigma U(\mu,\sigma) \\ \Delta g - \frac{1}{r^2}\Big(f(U(\mu,\sigma)+g) - f(U(\mu,\sigma))\Big) - \mu' \dot{\partial_\mu} U(\mu,\sigma) - (\sigma' + \frac{\xi(\sigma)}{\mu})\dot{\partial_\sigma} U(\mu,\sigma) \end{pmatrix} \tag{3.39}$$

Proof of Lemma 3.6. The proof of the existence of $(\boldsymbol{g}(t), \mu(t), \sigma(t))$ as in the statement of the lemma is nearly identical to [11, Proof of Lemma 3.1] so we give only a brief sketch here, highlighting the differences. First note that since $\widehat{\mathbf{d}}(\boldsymbol{u}(t))$ is small, we can find $\mu_1(t), \sigma_1(t)$ with $\sigma_1^k(t) \leq \eta^2$ so that for $g_1(t)$ defined by

$$g_1(t) := u(t) - (Q_{\mu_1\sigma_1} - Q_{\mu_1}) \tag{3.40}$$

we have

$$\|g_1(t)\|_H^2 + \sigma_1(t)^k \leq 2\eta^2 \tag{3.41}$$

To simplify notation we will suppress the time-dependency in the expressions below. Define a mapping $F : H \times (0,\infty) \times (0,\infty) \to H$ by

$$F(g,\mu,\sigma) := g - U(\mu,\sigma) + (Q_{\mu_1\sigma_1} - Q_{\mu_1})$$

and recall that by definition of $\boldsymbol{g}_c(\mu,\sigma)$ and the estimate (3.13),

$$U(\mu,\sigma) = Q_{\mu\sigma} - Q_{\mu\mu_c(\lambda_c^{-1}(\sigma))} + g_c(\mu,\sigma), \quad \|g_c(\mu,\sigma)\|_H \lesssim \sigma^k \tag{3.42}$$

It follows that

$$F(g_c(\lambda_c^{-1}(\sigma_1))_{\mu_1} + Q_{\mu_1\mu_c(\lambda_c^{-1}(\sigma_1))} - Q_{\mu_1}, \mu_1, \sigma_1) = 0 \tag{3.43}$$

and we have

$$\begin{aligned} \|F(g,\mu,\sigma)\|_H &\leq \|g\|_H + \|U(\mu,\sigma) - (Q_{\mu\sigma} - Q_{\mu\mu_c(\lambda_c^{-1}(\sigma))})\|_H + \|Q_{\mu\sigma} - Q_{\mu_1\sigma_1}\|_H \\ &\quad + \|Q_{\mu\mu_c(\lambda_c^{-1}(\sigma))} - Q_{\mu_1\mu_c(\lambda_c^{-1}(\sigma))}\|_H + \|Q_{\mu_1\mu_c(\lambda_c^{-1}(\sigma))} - Q_{\mu_1}\|_H \\ &\lesssim \|g\|_H + \|g_c(\lambda_c^{-1}(\sigma))\|_H + \left|\frac{\mu\sigma}{\mu_1\sigma_1} - 1\right|^{\frac12} + \left|\frac{\mu}{\mu_1} - 1\right|^{\frac12} + \left|\mu_c(\lambda_c^{-1}(\sigma)) - 1\right|^{\frac12} \end{aligned}$$

Next, define a mapping $G : H \times (0,\infty) \times (0,\infty) \to \mathbb{R}^2$ by

$$G(g,\mu,\sigma) = \begin{pmatrix} \frac{1}{\mu}\Big\langle \Lambda Q_{\underline{\mu}} \mid F(g,\mu,\sigma)\Big\rangle \\ \frac{1}{\mu\sigma}\Big\langle \Lambda Q_{\underline{\mu\sigma}} \mid F(g,\mu,\sigma)\Big\rangle \end{pmatrix}$$

Using (3.43) we have

$$G(g_c(\lambda_c^{-1}(\sigma_1))_{\mu_1} + Q_{\mu_1\mu_c(\lambda_c^{-1}(\sigma_1))} - Q_{\mu_1}, \mu_1, \sigma_1) = 0$$

Moreover, for any $h \in H$ we have the estimates

$$\begin{aligned} \frac{1}{\mu}\left|\Big\langle \Lambda Q_{\underline{\mu}} \mid h\Big\rangle\right| &\lesssim \|r/\mu \Lambda Q_{\underline{\mu}}\|_{L^2}\|r^{-1}h\|_{L^2} \lesssim \|h\|_H \\ \frac{1}{\mu\sigma}\left|\Big\langle \Lambda Q_{\underline{\mu\sigma}} \mid h\Big\rangle\right| &\lesssim \|r(\mu\sigma)^{-1}\Lambda Q_{\underline{\mu\sigma}}\|_{L^2}\|r^{-1}h\|_{L^2} \lesssim \|h\|_H \end{aligned}$$

which ensures that G is well defined and continuous. As in [11, Proof of Lemma 3.1] one can now readily check that the implicit function theorem can applied to G, meaning that for each g_0 in a small enough neighborhood (of size $\simeq \eta_0$) of $g_c(\lambda_c^{-1}(\sigma_1))_{\mu_1} + Q_{\mu_1\mu_c(\lambda_c^{-1}(\sigma_1))} - Q_{\mu_1}$, we can find unique $(\mu_0,\sigma_0) = \varsigma(g_0)$ (for the function ς given by the implicit function theorem) in a neighborhood of (μ_1,σ_1) (we note that it is convenient here to work in the variables, $s = \log\sigma, m := \log\mu$) and so that

$$G(g_0,\mu_0,\sigma_0) = 0$$

We refer the reader to [11, Lemma 3.1 and Remark 3.2] for precise details on the version and implementation of the implicit function in this setting. The desired triple (g, μ, σ) as in the lemma is then given by

$$(\mu, \sigma) := \varsigma(g_1), \quad g := F(g_1, \mu, \sigma)$$

where g_1 is as in (3.40), as long as g_1 is close enough in H to $g_c(\lambda_c^{-1}(\sigma_1))_{\mu_1} + Q_{\mu_1\mu_c(\lambda_c^{-1}(\sigma_1))} - Q_{\mu_1}$. To see this we measure,

$$\begin{aligned}\|g_1 - g_c(\lambda_c^{-1}(\sigma_1))_{\mu_1} + Q_{\mu_1\mu_c(\lambda_c^{-1}(\sigma_1))} - Q_{\mu_1}\|_H &\leq \|g_1\|_H + \|g_c(\lambda_c^{-1}(\sigma_1))\|_H \\ &\quad + \|Q_{\mu_1\mu_c(\lambda_c^{-1}(\sigma_1))} - Q_{\mu_1}\|_H \\ &\lesssim \sigma_1^k + \big|\mu_c(\lambda_c^{-1}(\sigma_1)) - 1\big|^{\frac{1}{2}} \lesssim \eta\end{aligned}$$

where the last line above follows from (3.42), (3.41), and (3.11). Thus μ, σ and g are well-defined. To conclude, we note that it follows from the definition of F that

$$g = F(g_1, \varsigma(g_1)) = F(g_1, \mu, \sigma) = u - U(\mu, \sigma)$$

and from the definition of G that

$$\Big\langle \Lambda Q_{\underline{\mu\sigma}} \mid g \Big\rangle = 0 \quad \text{and} \quad \Big\langle \Lambda Q_{\underline{\mu}} \mid g \Big\rangle = 0$$

as desired.

The coercivity estimate follows from a standard argument using the orthogonality conditions (3.33) and (3.34) together with the localized coercivity Lemma 2.1. Indeed, the smallness of σ yields a uniform constant $c_1 > 0$ for which

$$\int_0^\infty (\partial_r g)^2 + k^2 \frac{\cos(2Q_{\mu\sigma} - \cos 2Q_\mu)}{r^2} g^2 \, r \, \mathrm{d}r \geq c_1 \|g\|_H^2$$

For a detailed proof of the above see [9, Lemma 5.4]. Next, we argue perturbatively. Note that

$$\big\langle D^2\mathcal{E}(\boldsymbol{U}(\mu, \sigma)\boldsymbol{g} \mid \boldsymbol{g}\big\rangle = \int_0^\infty (\partial_r g)^2 + k^2 \frac{\cos(2U(\mu, \sigma))}{r^2} g^2 \, r \, \mathrm{d}r$$

and,

$$\begin{aligned}\cos 2U(\mu, \sigma) &= \cos\big(2Q_{\mu\sigma} - 2Q_\mu + (2Q_\mu - 2Q_{\mu_c(\lambda_c^{-1}(\sigma))\mu}) + 2g_c(\lambda_c^{-1}(\sigma))\big) \\ &= \cos\big(2Q_{\mu\sigma} - 2Q_\mu\big) + O(\Big|Q_\mu - Q_{\mu_c(\lambda_c^{-1}(\sigma))\mu}\Big|) + O(\big|g_c(\lambda_c^{-1}(\sigma))\big|)\end{aligned}$$

Thus,

$$\begin{aligned}\big\langle D^2\mathcal{E}(\boldsymbol{U}(\mu, \sigma)\boldsymbol{g} \mid \boldsymbol{g}\big\rangle &\geq c_1\|g\|_H^2 - O\Big(\big|\mu_c(\lambda_c^{-1}(\sigma)) - 1\big|^{\frac{1}{2}}) + \|g_c(\lambda_c^{-1}(\sigma))\|_H\Big)\|g\|_H^2 \\ &\geq c_0\|g\|_H^2\end{aligned}$$

where the last line follows by taking $\sigma > 0$ small enough. This completes the proof of (3.35).

Next we prove the estimates (3.36) and (3.37). We differentiate the modulation equations, beginning with (3.33),

$$\begin{aligned}0 = \frac{\mathrm{d}}{\mathrm{d}t}\Big\langle \Lambda Q_{\underline{\mu}} \mid g \Big\rangle &= -\frac{\mu'}{\mu}\Big\langle [\Lambda_0 \Lambda Q]_{\underline{\mu}} \mid g \Big\rangle + \Big\langle \Lambda Q_{\underline{\mu}} \mid \partial_t g \Big\rangle \\ &= -\frac{\mu'}{\mu}\Big\langle [\Lambda_0 \Lambda Q]_{\underline{\mu}} \mid g \Big\rangle + \Big\langle \Lambda Q_{\underline{\mu}} \mid \dot{g} \Big\rangle - \mu'\Big\langle \Lambda Q_{\underline{\mu}} \mid \partial_\mu U(\mu, \sigma) \Big\rangle \\ &\quad - \Big(\sigma' + \frac{\xi(\sigma)}{\mu}\Big)\Big\langle \Lambda Q_{\underline{\mu}} \mid \partial_\sigma U(\mu, \sigma) \Big\rangle\end{aligned}$$

Rearranging the above gives

$$\Big\langle \Lambda Q_{\underline{\mu}} \mid \dot g \Big\rangle = \mu' \Big(\Big\langle \Lambda Q_{\underline{\mu}} \mid \partial_\mu U(\mu,\sigma) \Big\rangle + \frac{1}{\mu} \Big\langle [\Lambda_0 \Lambda Q]_{\underline{\mu}} \mid g \Big\rangle \Big) + \Big(\mu\sigma' + \xi(\sigma) \Big) \frac{1}{\mu} \Big\langle \Lambda Q_{\underline{\mu}} \mid \partial_\sigma U(\mu,\sigma) \Big\rangle$$

Next write $\lambda := \sigma\mu$, and note that by the chain rule we have

$$\frac{\lambda'}{\lambda} = \frac{\sigma'}{\sigma} + \frac{\mu'}{\mu} \tag{3.44}$$

Differentiating (3.34) gives

$$\begin{aligned} 0 = \frac{\mathrm{d}}{\mathrm{d}t} \big\langle \Lambda Q_{\underline{\lambda}} \mid g \big\rangle &= -\frac{\lambda'}{\lambda} \big\langle [\Lambda_0 \Lambda Q]_{\underline{\lambda}} \mid g \big\rangle + \big\langle \Lambda Q_{\underline{\lambda}} \mid \partial_t g \big\rangle \\ &= -\frac{\lambda'}{\lambda} \big\langle [\Lambda_0 \Lambda Q]_{\underline{\lambda}} \mid g \big\rangle + \big\langle \Lambda Q_{\underline{\lambda}} \mid \dot g \big\rangle - \mu' \big\langle \Lambda Q_{\underline{\lambda}} \mid \partial_\mu U(\mu,\sigma) \big\rangle \\ &\quad - \Big(\sigma' + \frac{\xi(\sigma)}{\mu} \Big) \big\langle \Lambda Q_{\underline{\lambda}} \mid \partial_\sigma U(\mu,\sigma) \big\rangle \end{aligned}$$

which, using (3.44) yields,

$$\begin{aligned} \big\langle \Lambda Q_{\underline{\lambda}} \mid \dot g \big\rangle + \frac{\xi(\sigma)}{\mu\sigma} \langle [\Lambda_0 \Lambda Q]_{\underline{\lambda}} \mid g \rangle &= \mu' \Big(\big\langle \Lambda Q_{\underline{\lambda}} \mid \partial_\mu U(\mu,\sigma) \big\rangle + \frac{1}{\mu} \big\langle [\Lambda_0 \Lambda Q]_{\underline{\lambda}} \mid g \big\rangle \Big) \\ &\quad + \big(\mu\sigma' + \xi(\sigma) \big) \Big(\frac{1}{\mu} \big\langle \Lambda Q_{\underline{\lambda}} \mid \partial_\sigma U(\mu,\sigma) \big\rangle + \lambda^{-1} \big\langle [\Lambda_0 \Lambda Q]_{\underline{\lambda}} \mid g \big\rangle \Big) \end{aligned}$$

We obtain the following system of equations,

$$\begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \begin{pmatrix} \mu' \\ \mu\sigma' + \xi(\sigma) \end{pmatrix} = \begin{pmatrix} \Big\langle \Lambda Q_{\underline{\mu}} \mid \dot g \Big\rangle \\ \big\langle \Lambda Q_{\underline{\lambda}} \mid \dot g \big\rangle + \frac{\xi(\sigma)}{\mu\sigma} \big\langle [\Lambda_0 \Lambda Q]_{\underline{\lambda}} \mid g \big\rangle \end{pmatrix} =: \begin{pmatrix} B_1 \\ B_2 \end{pmatrix} \tag{3.45}$$

where

$$\begin{aligned} M_{11} &:= \Big\langle \Lambda Q_{\underline{\mu}} \mid \partial_\mu U(\mu,\sigma) \Big\rangle + \mu^{-1} \Big\langle \Lambda_0 \Lambda Q_{\underline{\mu}} \mid g \Big\rangle \\ M_{22} &:= \mu^{-1} \big\langle \Lambda Q_{\underline{\lambda}} \mid \partial_\sigma U(\mu,\sigma) \big\rangle + \lambda^{-1} \big\langle \Lambda_0 \Lambda Q_{\underline{\lambda}} \mid g \big\rangle \\ M_{12} &:= \mu^{-1} \Big\langle \Lambda Q_{\underline{\mu}} \mid \partial_\sigma U(\mu,\sigma) \Big\rangle \\ M_{21} &:= \big\langle \Lambda Q_{\underline{\lambda}} \mid \partial_\mu U(\mu,\sigma) \big\rangle + \mu^{-1} \big\langle [\Lambda_0 \Lambda Q]_{\underline{\lambda}} \mid g \big\rangle \end{aligned}$$

Note the estimates,

$$\begin{aligned} &|B_1| \lesssim \|\dot g\|_{L^2} \\ &|B_2| \lesssim \|\dot g\|_{L^2} + \sigma^{\frac{k}{2}} \|g\|_H \quad \text{and} \quad \big| B_2 - \big\langle \Lambda Q_{\underline{\lambda}} \mid \dot g \big\rangle \big| \lesssim \sigma^{\frac{k}{2}} \|g\|_H \end{aligned} \tag{3.46}$$

We claim the bounds

Claim 3.8. The following estimates hold true.

$$\big| M_{11} - \|\Lambda Q\|_{L^2}^2 \big| \lesssim \sigma^{\frac{1}{2}} + \|g\|_H \tag{3.47}$$

$$\big| M_{22} + (1 - o(1)) \|\Lambda Q\|_{L^2}^2 \big| \lesssim \sigma^k + \sigma \|g\|_H \tag{3.48}$$

$$|M_{12}| \lesssim o(1) \tag{3.49}$$

$$|M_{21}| \lesssim \sigma \tag{3.50}$$

$$\det M = M_{11} M_{22} + O(\sigma)$$

where $o(1)$ can be replaced by a constant that can be made as small as we like by taking η small enough.

Proof of Claim 3.8. First we prove (3.47). Recall that $\partial_\mu U(\mu,\sigma) = -\frac{1}{\mu}\Lambda U(\mu,\sigma)$. Hence,

$$\left|M_{11} - \|\Lambda Q\|_{L^2}^2\right| \leq \left|\left\langle \Lambda Q_{\underline{\mu}} \mid \frac{1}{\mu}(\Lambda U(\mu,\sigma) + \Lambda Q_\mu)\right\rangle\right| + \left|\mu^{-1}\left\langle [\Lambda_0 \Lambda Q]_{\underline{\mu}} \mid g\right\rangle\right|$$

The second term on the right above can be bounded as follows:

$$\left|\mu^{-1}\left\langle [\Lambda_0\Lambda Q]_{\underline{\mu}} \mid g\right\rangle\right| \lesssim \|(r/\mu)[\Lambda_0\Lambda Q]_{\underline{\mu}}\|_{L^2}\|r^{-1}g\|_{L^2} \lesssim \|g\|_H$$

To control the first term on the right, we first write

$$\Lambda U(\mu,\sigma) = \Lambda Q_{\sigma\mu} - \Lambda Q_{\mu\mu_c(\lambda^{-1}(\sigma))} + \Lambda g_c(\mu,\sigma)$$

so after rescaling we have,

$$\begin{aligned}\left\langle \Lambda Q \mid (\Lambda U(\mu,\sigma)_{\mu^{-1}} + \Lambda Q\right\rangle \lesssim |\langle \Lambda Q \mid \Lambda Q_\sigma\rangle| + \left|\left\langle \Lambda Q \mid \Lambda Q - \Lambda Q_{\mu_c(\lambda^{-1}(\sigma))}\right\rangle\right| \\ + \left|\left\langle \Lambda Q \mid \Lambda g_c(\lambda^{-1}(\sigma))\right\rangle\right|\end{aligned}$$

For the first term we have $|\langle \Lambda Q \mid \Lambda Q_\sigma\rangle| \ll \sigma^{\frac{1}{2}}$. Next, observe that by (3.13),

$$\left|\left\langle \Lambda Q \mid \Lambda g_c(\lambda_c^{-1}(\sigma))\right\rangle\right| \lesssim \|r\Lambda Q\|_{L^2}\|r^{-1}\Lambda g_c(\lambda_c^{-1}(\sigma)\|_{L^2} \lesssim \|g_c(\lambda_c^{-1}(\sigma)\|_H \lesssim \sigma^k$$

Lastly, we use (3.11) to deduce that

$$\left|\left\langle \Lambda Q \mid \Lambda Q - \Lambda Q_{\mu_c(\lambda_c^{-1}(\sigma))}\right\rangle\right| \lesssim \left|\mu_c(\lambda_c^{-1}(\sigma)) - 1\right|^{\frac{1}{2}} \lesssim \sigma^{\frac{1}{2}}$$

Combining these estimates proves (3.47). Next we treat the term M_{22}. We have

$$M_{22} = \left\langle \Lambda Q_{\underline{\lambda}} \mid \frac{1}{\mu}\partial_\sigma U(\mu,\sigma)\right\rangle + \lambda^{-1}\left\langle [\Lambda_0\Lambda Q]_{\underline{\lambda}} \mid g\right\rangle$$

For the second term above, we have

$$\left|\lambda^{-1}\left\langle [\Lambda_0\Lambda Q]_{\underline{\lambda}} \mid g\right\rangle\right| \lesssim \|(r/\lambda)[\Lambda_0\Lambda Q]_{\underline{\lambda}}\|_{L^2}\|r^{-1}g\|_{L^2} \lesssim \|g\|_H$$

By (3.4) and the definition of $\dot{U}(\mu,\sigma)$ we have

$$\begin{aligned}\frac{1}{\mu}\partial_\sigma U(\mu,\sigma) = -\frac{1}{\xi(\sigma)}\dot{U}(\mu,\sigma) = -\Lambda Q_{\underline{\mu\sigma}} + \Big(1 - \frac{b_c(\lambda_c^{-1}(\sigma))}{\xi(\sigma)}\Big)\Lambda Q_{\underline{\mu\sigma}} \\ - \frac{1}{\xi(\sigma)}\Big(\dot{\Phi}(\mu,\sigma) - b_c(\lambda_c^{-1}(\sigma))\Lambda Q_{\underline{\mu\sigma}}\Big) - \frac{1}{\xi(\sigma)}\dot{w}_c(\mu,\sigma)\end{aligned}$$

It then follows from (3.10), (3.22), and (3.7) that,

$$\left\langle \Lambda Q_{\underline{\lambda}} \mid \frac{1}{\mu}\partial_\sigma U(\mu,\sigma)\right\rangle = -(1-o(1))\|\Lambda Q\|_{L^2} + O(\sigma^{\frac{k}{2}})$$

This proves (3.48). To prove (3.49) we write,

$$\begin{aligned}\frac{1}{\mu}\partial_\sigma U(\mu,\sigma) = -\frac{a_c(\lambda_c^{-1}(\sigma))}{\xi(\sigma)}\Lambda Q_{\underline{\mu\mu_c(\lambda_c^{-1}(\sigma))}} - \frac{1}{\xi(\sigma)}\Big(\dot{\Phi}(\mu,\sigma) - a_c(\lambda_c^{-1}(\sigma))\Lambda Q_{\underline{\mu\mu_c(\lambda_c^{-1}(\sigma))}}\Big) \\ - \frac{1}{\xi(\sigma)}\dot{w}_c(\mu,\sigma)\end{aligned}$$

and thus, using (3.10) (3.11), (3.12), (3.7) and (3.24) we arrive at the estimate,

$$|M_{12}| = \left|\left\langle \Lambda Q_{\underline{\mu}} \mid \frac{1}{\mu}\partial_\sigma U(\mu,\sigma)\right\rangle\right| \lesssim o(1)$$

Finally, we estimate (3.50).

$$|M_{21}| \lesssim \left|\left\langle \Lambda Q_{\underline{\lambda}} \mid \partial_\mu U(\mu,\sigma)\right\rangle\right| + \left|\frac{1}{\mu}\left\langle [\Lambda_0\Lambda Q]_{\underline{\lambda}} \mid g\right\rangle\right|$$

The second term above is controlled as follows,

$$\left|\frac{1}{\mu}\left\langle [\Lambda_0 \Lambda Q]_{\underline{\lambda}} \mid g\right\rangle\right| \lesssim \sigma \|r/\lambda \Lambda_0 \Lambda Q_{\underline{\lambda}}\|_{L^2} \|r^{-1} g\|_{L^2} \lesssim \sigma \|g\|_H$$

To estimate the first recall that

$$\partial_\mu U(\mu,\sigma) = -\frac{1}{\mu}\Lambda U(\mu,\sigma) = -\sigma \Lambda Q_{\underline{\sigma\mu}} + \mu_c(\lambda_c^{-1}(\sigma))\Lambda Q_{\underline{\mu_c(\lambda_c^{-1}(\sigma))\mu}} - (\Lambda g_c)(\lambda_c^{-1}(\sigma))_{\underline{\mu}}$$

and hence

$$\begin{aligned}\left|\left\langle \Lambda Q_{\underline{\lambda}} \mid \partial_\mu U(\mu,\sigma)\right\rangle\right| &\lesssim \sigma \|\Lambda Q\|_{L^2}^2 + \left|\left\langle \Lambda Q_{\underline{\sigma}} \mid \Lambda Q_{\underline{\mu_c(\lambda_c^{-1}(\sigma))}}\right\rangle\right| + \left|\left\langle \Lambda Q_{\underline{\sigma}} \mid (\Lambda g_c)(\lambda_c^{-1}(\sigma))\right\rangle\right| \\ &\lesssim \sigma\end{aligned}$$

as claimed. □

With the estimates in Claim 3.8 in hand, we see that we can invert M as long as $\|\boldsymbol{g}\|_{\mathcal{H}}$ and σ are small enough and solve for $(\mu', \mu\sigma' + \xi(\sigma))$ in (3.45). This yields,

$$\mu' = \left[\frac{1}{M_{11}M_{22}} + O(\sigma)\right](M_{22}B_1 - M_{12}B_2)$$

From Claim 3.8 and (3.46) we conclude that

$$|\mu'| \lesssim \|\dot g\|_{L^2} + \sigma^{\frac{k}{2}}\|g\|_H$$

which proves (3.36). Similarly,

$$\mu\sigma' + \xi(\sigma) = \left[\frac{1}{M_{11}M_{22}} + O(\sigma)\right](M_{11}B_2 - M_{21}B_1) \tag{3.51}$$

Therefore, on the one hand we can conclude from Claim 3.8 and (3.46) that

$$|\mu\sigma' + \xi(\sigma)| \lesssim \|\dot g\|_{L^2} + \sigma^{\frac{k}{2}}\|g\|_H$$

In fact, extracting the leading order from the right-hand-side of (3.51) we deduce that

$$\left|\mu\sigma' + \xi(\sigma) + \frac{1}{\|\Lambda Q\|_{L^2}^2}\left\langle \Lambda Q_{\underline{\lambda}} \mid \dot g\right\rangle\right| \lesssim o(1)\|\dot g\|_{L^2} + \sigma^{\frac{k}{2}}\|g\|_H$$

proving (3.37). □

4. The poof of uniqueness

In this section we complete the proof of Theorem 1.1.

4.1. **An outline of the proof of Theorem 1.1.**

We begin with a short outline of end of the proof of Theorem 1.1. The purpose is to motivate the computations performed in the next subsection.

Let $\boldsymbol{u}(t) \in \mathcal{H}$ be any 2-bubble in forward time as in (1.5) on the time interval $[T_0, \infty)$. By taking $T_0 > 0$ large enough we may apply Lemma 3.6 on the time interval $[T_0, \infty)$, obtaining a decomposition

$$\boldsymbol{u}(t) = \boldsymbol{U}(\mu(t), \sigma(t)) + \boldsymbol{g}(t)$$

as in Lemma 3.6. By the local Cauchy theory, it will suffice to find a single time $t \geq T_0$ for which we have $\|\boldsymbol{g}(t)\|_{\mathcal{H}} = 0$. The starting point is the following Taylor expansion of the conserved energy

about the constructed trajectory $\boldsymbol{U}(\mu,\sigma)$. For each time $t \geq T_0$ we have

$$\begin{aligned} 2\mathcal{E}(\boldsymbol{Q}) = \mathcal{E}(\boldsymbol{u}(t)) &= \mathcal{E}(\boldsymbol{U}(\mu(t),\sigma(t)) + \boldsymbol{g}(t)) \\ &= \mathcal{E}(\boldsymbol{U}(\mu(t),\sigma(t))) + \langle D\mathcal{E}(\boldsymbol{U}(\mu(t),\sigma(t))) \mid \boldsymbol{g}(t)\rangle \\ &\quad + \left\langle D^2\mathcal{E}(\boldsymbol{U}(\mu(t),\sigma(t)))\boldsymbol{g}(t) \mid \boldsymbol{g}(t)\right\rangle + o(\|\boldsymbol{g}\|_{\mathcal{H}}^2) \\ &= 2\mathcal{E}(\boldsymbol{Q}) + \langle D\mathcal{E}(\boldsymbol{U}(\mu(t),\sigma(t))) \mid \boldsymbol{g}(t)\rangle \\ &\quad + \left\langle D^2\mathcal{E}(\boldsymbol{U}(\mu(t),\sigma(t)))\boldsymbol{g}(t) \mid \boldsymbol{g}(t)\right\rangle + o(\|\boldsymbol{g}\|_{\mathcal{H}}^2) \end{aligned}$$

Subtracting $2\mathcal{E}(\boldsymbol{Q})$ from both sides, recalling the coercivity estimate from Lemma 3.6, i.e., (3.35), and making the "little oh" term above smaller than half the coercivity constant $c_1 > 0$ (which is possible by taking $T_0 > 0$ large enough) we arrive at the inequality

$$0 \geq \langle D\mathcal{E}(\boldsymbol{U}(\mu(t),\sigma(t))) \mid \boldsymbol{g}(t)\rangle + \frac{1}{2}c_1\|\boldsymbol{g}(t)\|_{\mathcal{H}}^2 \tag{4.1}$$

We will show there is necessarily a time $T_1 \geq T_0$ such that

$$\langle D\mathcal{E}(\boldsymbol{U}(\mu(T_1),\sigma(T_1))) \mid \boldsymbol{g}(T_1)\rangle \geq -\frac{1}{4}c_1\|\boldsymbol{g}(T_1)\|_{\mathcal{H}}^2 \tag{4.2}$$

which together with (4.1) would imply that $\|\boldsymbol{g}(T_1)\|_{\mathcal{H}} = 0$ and thus

$$\boldsymbol{u}(T_1) = \boldsymbol{U}(\mu(T_1),\sigma(T_1)) = (u_c(\lambda_c^{-1}(\sigma(T_1)),\cdot/\mu(T_1)), \mu(T_1)^{-1}\partial_t u_c(\lambda_c^{-1}(\sigma(T_1)),\cdot/\mu(T_1))),$$

which would prove Theorem 1.1. In the next section we analyze the dynamics of

$$\langle D\mathcal{E}(\boldsymbol{U}(\mu(t),\sigma(t))) \mid \boldsymbol{g}(t)\rangle\,.$$

with the goal of proving (4.2).

4.2. **Analysis of the refined instability component.** We now come to the heart of the argument. For each $\mu,\sigma > 0$ define

$$\boldsymbol{\beta}(\mu,\sigma) := \frac{1}{\rho_k\sigma^{\frac{k}{2}}}\,\mathrm{D}\,\mathcal{E}(\boldsymbol{U}(\mu,\sigma)) = \frac{1}{\rho_k\sigma^{\frac{k}{2}}}\begin{pmatrix} -\Delta U(\mu,\sigma) + r^{-2}f(U(\mu,\sigma)) \\ \dot{U}(\mu,\sigma)\end{pmatrix} \tag{4.3}$$

We make a few comments on how to think of $\boldsymbol{\beta}(\mu,\sigma)$. Recall that

$$\partial_\sigma \boldsymbol{U}(\mu,\sigma) = -\frac{\mu}{\xi(\sigma)}J\circ\mathrm{D}\,\mathcal{E}(\boldsymbol{U}(\mu,\sigma))$$

Since $\rho_k\sigma^{\frac{k}{2}} \simeq \xi(\sigma)$, we see that $\boldsymbol{\beta}(\mu,\sigma)$ is essentially a 90-degree rotation of $\partial_\sigma\boldsymbol{U}(\mu,\sigma)$, rescaled by μ^{-1} i.e.,

$$\boldsymbol{\beta}(\mu,\sigma) = \frac{1}{\rho_k\sigma^{\frac{k}{2}}}\,\mathrm{D}\,\mathcal{E}(\boldsymbol{U}(\mu,\sigma)) \simeq \frac{1}{\mu}J\circ\partial_\sigma\boldsymbol{U}(\mu,\sigma)$$

Next, consider the coefficient of the projection of $\boldsymbol{g}(t)$ onto $\boldsymbol{\beta}(\mu,\sigma)$, modified by a small "virial" type correction term.

$$\begin{aligned} b(t) :&= \langle\boldsymbol{\beta}(\mu(t),\sigma(t)) \mid \boldsymbol{g}(t)\rangle + \langle\mathcal{A}_0(\mu(t)\sigma(t))g(t) \mid \dot{g}(t)\rangle \\ &= \frac{1}{\rho_k\sigma^{\frac{k}{2}}}\langle D\mathcal{E}(\boldsymbol{U}(\mu(t),\sigma(t))) \mid \boldsymbol{g}(t)\rangle + \langle\mathcal{A}_0(\mu(t)\sigma(t))g(t) \mid \dot{g}(t)\rangle \end{aligned} \tag{4.4}$$

The correction is intended to produce cancellations of terms of critical size, but indeterminate sign, when we compute $b'(t)$ below.

The basic lemma for the family $\boldsymbol{\beta}(\mu,\sigma)$ is the following.

Lemma 4.1. *The family of functionals $\beta(\mu,\sigma)$ is uniformly bounded in $\mathcal{H}^*$. In fact, we have the estimates,*

$$\left|\left\langle \boldsymbol{\beta}(\mu,\sigma) - (0,\Lambda Q_{\underline{\mu\sigma}}) \mid \boldsymbol{h} \right\rangle\right| = o(1)\|\boldsymbol{h}\|_{\mathcal{H}} \quad as \quad \sigma \to 0. \tag{4.5}$$

for all $\boldsymbol{h} \in \mathcal{H}$. In particular,

$$\left|b(t) - \left\langle \Lambda Q_{\underline{\sigma\mu}}, \mid \dot{g} \right\rangle\right| = o(1)\|\boldsymbol{g}\|_{\mathcal{H}} \quad as \quad \sigma \to 0. \tag{4.6}$$

We also have the estimate,

$$\|\boldsymbol{\beta}(\mu,\sigma) - (0,\Lambda Q_{\underline{\mu\sigma}})\|_{L^2\times H} = o(1)\frac{1}{\mu\sigma} \quad as \quad \sigma \to 0 \tag{4.7}$$

Remark 4.2. From (4.5) we see that to leading order

$$\boldsymbol{\beta}(\mu,\sigma) \simeq (0,\Lambda Q_{\underline{\mu\sigma}})$$

and is thus $b(t)$ is closely related to the quantity that is also called $b(t)$ in [11]. It is also related to the *refined unstable component* from [8].

Proof of Lemma 4.1. The proof of (4.5) and hence also of (4.6) are direct consequences of the definition of $\boldsymbol{\beta}(\mu,\sigma)$ in (4.3) and the estimates (3.15) and (3.17) from Corollary 3.4. The estimate (4.7) follows from (3.16) and (3.18). □

Proposition 4.3. *Let $\boldsymbol{u}(t) \in \mathcal{H}$ be a two-bubble in forward time and define $b(t)$ as in* (4.4)*. For any $c_0 > 0$ there exists $T_0 > 0$ such that*

$$b'(t) \le \frac{k\rho_k}{2\mu(t)\sigma(t)}\sigma(t)^{\frac{k}{2}}b(t) + c_0\frac{1}{\mu(t)\sigma(t)}\left(|b(t)|\,\sigma(t)^{\frac{k}{2}} + \|\boldsymbol{g}(t)\|_{\mathcal{H}}^2\right) \tag{4.8}$$

holds uniformly on the time interval $[T_0,\infty)$.

The main application of Proposition 4.3 is the following corollary.

Corollary 4.4. *Suppose that $\boldsymbol{u}(t) \in \mathcal{H}$ is a two-bubble in forward time. There exists $T_0 > 0$ with the following property. For every $\epsilon_1 > 0$ there exists $T_1 > 0$ with $T_1 \in [T_0,\infty)$ such that*

$$\sigma(T_1)^{\frac{k}{2}}b(T_1) \ge -\epsilon_1\|\boldsymbol{g}(T_1)\|_{\mathcal{H}}^2$$

Proof of Corollary 4.4 assuming Proposition 4.3. Note that if $\|\boldsymbol{g}(t)\|_{\mathcal{H}} = 0$ for any t, then we have $\boldsymbol{u}(t) \equiv \boldsymbol{U}(\mu(t),\sigma(t))$ as claimed by Theorem 1.1 and there is nothing to do. So we may assume that $\|\boldsymbol{g}(t)\|_{\mathcal{H}} > 0$ for all t for which Lemma 3.6 applies (i.e., all sufficiently large $t > 0$). Suppose Corollary 4.4 fails. Fixing a sufficiently large T_0' as in Lemma 3.6 there there exists $c_2 > 0$ so that for all $t \ge T_0'$ we have

$$\sigma(t)^{\frac{k}{2}}b(t) \le -c_2\,\|\boldsymbol{g}(t)\|_{\mathcal{H}}^2 \tag{4.9}$$

By Proposition 4.3 we can choose $T_0 \ge T_0'$ sufficiently large in order to find a uniform constant $c_3 > 0$ for which

$$b'(t) \le -c_3\frac{\|\boldsymbol{g}(t)\|_{\mathcal{H}}^2}{\mu(t)\sigma(t)} \quad \forall\, t \in [T_0,\infty)$$

But this implies that $b'(t) < 0$ on the entire interval $[T_0,\infty)$. By (4.9) we also have $b(t) < 0$ for all $t \in [T_0,\infty)$. But these two conditions are impossible since we know that $b(t) \to 0$ as $t \to \infty$. □

Proof of Proposition 4.3. We compute

$$\begin{aligned} b'(t) &= \frac{\mathrm{d}}{\mathrm{d}t}\left\langle \boldsymbol{\beta}(\mu(t),\sigma(t)) \mid \boldsymbol{g}(t)\right\rangle + \frac{\mathrm{d}}{\mathrm{d}t}\left\langle \mathcal{A}_0(\mu(t)\sigma(t))g(t) \mid \dot g(t)\right\rangle \\ &= \left(\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{\rho_k\sigma^{\frac{k}{2}}}\right)\left\langle D\mathcal{E}(\boldsymbol{U}(\mu,\sigma)) \mid \boldsymbol{g}(t)\right\rangle + \frac{1}{\rho_k\sigma^{\frac{k}{2}}}\frac{\mathrm{d}}{\mathrm{d}t}\left\langle D\mathcal{E}(\boldsymbol{U}(\mu,\sigma)) \mid \boldsymbol{g}(t)\right\rangle \\ &\quad + \frac{\mathrm{d}}{\mathrm{d}t}\left\langle \mathcal{A}_0(\mu(t)\sigma(t))g(t) \mid \dot g(t)\right\rangle \end{aligned}$$

The first leading term on the right-hand-side of (4.8) comes from differentiating $\sigma^{-\frac{k}{2}}$ above. Indeed by (3.37) we have

$$\begin{aligned} \left(\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{\rho_k\sigma^{\frac{k}{2}}}\right)\left\langle D\mathcal{E}(\boldsymbol{U}(\mu,\sigma)) \mid \boldsymbol{g}(t)\right\rangle &= -\frac{k}{2}\frac{\sigma'}{\sigma}\frac{1}{\rho_k\sigma^{\frac{k}{2}}}\left\langle D\mathcal{E}(\boldsymbol{U}(\mu,\sigma)) \mid \boldsymbol{g}(t)\right\rangle = -\frac{k}{2}\frac{\sigma'}{\sigma}b(t) \\ &= \frac{k}{2}\left(\frac{\xi(\sigma)}{\mu\sigma} + \frac{\left\langle \Lambda Q_{\underline{\mu\sigma}} \mid \dot g\right\rangle}{\mu\sigma\|\Lambda Q\|_{L^2}^2}\right)b(t) + o(1)O\left(\frac{\|\boldsymbol{g}(t)\|_{\mathcal{H}}^2}{\mu\sigma}\right) \end{aligned}$$

Using (3.10) and (4.6), and the fact that $\sigma(t)\to 0$ as $t\to\infty$ we conclude that

$$\begin{aligned} \left(\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{\rho_k\sigma^{\frac{k}{2}}}\right)\left\langle D\mathcal{E}(\boldsymbol{U}(\mu,\sigma)) \mid \boldsymbol{g}(t)\right\rangle &= \frac{k\rho_k}{2\mu(t)\sigma(t)}\sigma(t)^{\frac{k}{2}}b(t) + \frac{k}{2}\frac{\left\langle \Lambda Q_{\underline{\mu\sigma}} \mid \dot g\right\rangle^2}{\mu(t)\sigma(t)\|\Lambda Q\|_{L^2}^2} \\ &\quad + \frac{1}{\mu(t)\sigma(t)}o\left(|b(t)|\,\sigma(t)^{\frac{k}{2}} + \|\boldsymbol{g}(t)\|_{\mathcal{H}}^2\right) \end{aligned}$$

Hence Proposition 4.3 follows from the estimate

$$\frac{k\left\langle \Lambda Q_{\underline{\mu\sigma}} \mid \dot g\right\rangle^2}{2\mu\sigma\|\Lambda Q\|_{L^2}^2} + \frac{1}{\rho_k\sigma^{\frac{k}{2}}}\frac{\mathrm{d}}{\mathrm{d}t}\left\langle D\mathcal{E}(\boldsymbol{U}(\mu,\sigma)) \mid \boldsymbol{g}\right\rangle + \frac{\mathrm{d}}{\mathrm{d}t}\left\langle \mathcal{A}_0(\mu\sigma)g \mid \dot g\right\rangle \le o(1)O\Big(\frac{\|\boldsymbol{g}\|_{\mathcal{H}}^2}{\mu\sigma}\Big) \tag{4.10}$$

We begin the proof of (4.10) by expanding the second term on the left above using (3.38).

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\left\langle D\mathcal{E}(\boldsymbol{U}(\mu,\sigma)) \mid \boldsymbol{g}\right\rangle &= \left\langle D^2\mathcal{E}(\boldsymbol{U}(\mu,\sigma))\partial_t[\boldsymbol{U}(\mu,\sigma)] \mid \boldsymbol{g}\right\rangle + \left\langle D\mathcal{E}(\boldsymbol{U}(\mu,\sigma)) \mid \partial_t\boldsymbol{g}\right\rangle \\ &= \left\langle D^2\mathcal{E}(\boldsymbol{U}(\mu,\sigma))[\partial_t\boldsymbol{U}(\mu,\sigma)] \mid \boldsymbol{g}\right\rangle \\ &\quad + \left\langle D\mathcal{E}(\boldsymbol{U}(\mu,\sigma)) \mid \partial_t\boldsymbol{u}\right\rangle - \left\langle D\mathcal{E}(\boldsymbol{U}(\mu,\sigma)) \mid \partial_t[\boldsymbol{U}(\mu,\sigma)]\right\rangle \\ &= \left\langle D^2\mathcal{E}(\boldsymbol{U}(\mu,\sigma))[\partial_t\boldsymbol{U}(\mu,\sigma)] \mid \boldsymbol{g}\right\rangle \\ &\quad + \left\langle D\mathcal{E}(\boldsymbol{U}(\mu,\sigma)) \mid J\circ D\mathcal{E}(\boldsymbol{U}(\mu,\sigma)+\boldsymbol{g})\right\rangle \\ &\quad - \mu'\left\langle D\mathcal{E}(\boldsymbol{U}(\mu,\sigma)) \mid \partial_\mu\boldsymbol{U}(\mu,\sigma)\right\rangle - \sigma'\left\langle D\mathcal{E}(\boldsymbol{U}(\mu,\sigma)) \mid \partial_\sigma\boldsymbol{U}(\mu,\sigma)\right\rangle \end{aligned}$$

The fact that $\mathcal{E}(\boldsymbol{U}(\mu,\sigma))$ is constant in μ,σ implies the last two lines above $\equiv 0$ since

$$\begin{aligned} 0 &= \frac{\mathrm{d}}{\mathrm{d}\mu}\mathcal{E}(\boldsymbol{U}(\mu,\sigma)) = \left\langle D\mathcal{E}(\boldsymbol{U}(\mu,\sigma) \mid \partial_\mu\boldsymbol{U}(\mu,\sigma)\right\rangle \\ 0 &= \frac{\mathrm{d}}{\mathrm{d}\sigma}\mathcal{E}(\boldsymbol{U}(\mu,\sigma)) = \left\langle D\mathcal{E}(\boldsymbol{U}(\mu,\sigma) \mid \partial_\mu\boldsymbol{U}(\mu,\sigma)\right\rangle \end{aligned}$$

Then, subtracting $0 = \langle D\mathcal{E}(\boldsymbol{U}(\mu,\sigma)) \mid J\circ D\mathcal{E}\boldsymbol{U}(\mu,\sigma)\rangle$ we obtain

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\left\langle D\mathcal{E}(\boldsymbol{U}(\mu,\sigma)) \mid \boldsymbol{g}\right\rangle &= \left\langle D^2\mathcal{E}(\boldsymbol{U}(\mu,\sigma))[\partial_t\boldsymbol{U}(\mu,\sigma)] \mid \boldsymbol{g}\right\rangle \\ &\quad + \left\langle D\mathcal{E}(\boldsymbol{U}(\mu,\sigma)) \mid J\circ\big[D\mathcal{E}(\boldsymbol{U}(\mu,\sigma)+\boldsymbol{g}) - D\mathcal{E}(\boldsymbol{U}(\mu,\sigma))\big]\right\rangle \end{aligned} \tag{4.11}$$

Next, we re-write the first term above as follows. Recall that by (3.4) we have

$$\begin{aligned}\partial_t \boldsymbol{U}(\mu,\sigma) &= \mu' \partial_\mu \boldsymbol{U}(\mu,\sigma) + \sigma' \partial_\sigma \boldsymbol{U}(\mu,\sigma) \\ &= \mu' \partial_\mu \boldsymbol{U}(\mu,\sigma) + \left(\sigma' + \frac{\xi(\sigma)}{\mu}\right) \partial_\sigma \boldsymbol{U}(\mu,\sigma) + J \circ D\mathcal{E}(\boldsymbol{U}(\mu,\sigma))\end{aligned}$$

Hence, using also the self-adjointness of $D^2\mathcal{E}(\boldsymbol{U}(\mu,\sigma))$ and the skew-symmetry of J we have

$$\begin{aligned}\big\langle D^2\mathcal{E}(\boldsymbol{U}(\mu,\sigma))[\partial_t \boldsymbol{U}(\mu,\sigma)] \mid \boldsymbol{g}\big\rangle &= \big\langle D^2\mathcal{E}(\boldsymbol{U}(\mu,\sigma))[\partial_t \boldsymbol{U}(\mu,\sigma) - J \circ D\mathcal{E}(\boldsymbol{U}(\mu,\sigma))] \mid \boldsymbol{g}\big\rangle \\ &\quad + \big\langle D^2\mathcal{E}(\boldsymbol{U}(\mu,\sigma)) J \circ D\mathcal{E}(\boldsymbol{U}(\mu,\sigma)) \mid \boldsymbol{g}\big\rangle \\ &= \mu' \big\langle D^2\mathcal{E}(\boldsymbol{U}(\mu,\sigma))\partial_\mu \boldsymbol{U}(\mu,\sigma) \mid \boldsymbol{g}\big\rangle \\ &\quad + \left(\sigma' + \frac{\xi(\sigma)}{\mu}\right) \big\langle D^2\mathcal{E}(\boldsymbol{U}(\mu,\sigma)\partial_\sigma \boldsymbol{U}(\mu,\sigma) \mid \boldsymbol{g}\big\rangle \\ &\quad - \big\langle D\mathcal{E}(\boldsymbol{U}(\mu,\sigma)) \mid J \circ D^2\mathcal{E}(\boldsymbol{U}(\mu,\sigma))\boldsymbol{g}\big\rangle\end{aligned}$$

Inserting this back into (4.11) we obtain

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t} \big\langle D\mathcal{E}(\boldsymbol{U}(\mu,\sigma)) \mid \boldsymbol{g}\big\rangle \\ &\qquad = \big\langle D\mathcal{E}(\boldsymbol{U}(\mu,\sigma)) \mid J \circ \big[D\mathcal{E}(\boldsymbol{U}(\mu,\sigma)+\boldsymbol{g}) - D\mathcal{E}(\boldsymbol{U}(\mu,\sigma)) - D^2\mathcal{E}(\boldsymbol{U}(\mu,\sigma)\boldsymbol{g}\big]\big\rangle \\ &\qquad\quad + \mu' \big\langle D^2\mathcal{E}(\boldsymbol{U}(\mu,\sigma))\partial_\mu \boldsymbol{U}(\mu,\sigma) \mid \boldsymbol{g}\big\rangle + \left(\sigma' + \frac{\xi(\sigma)}{\mu}\right) \big\langle D^2\mathcal{E}(\boldsymbol{U}(\mu,\sigma)\partial_\sigma \boldsymbol{U}(\mu,\sigma) \mid \boldsymbol{g}\big\rangle\end{aligned}$$

Finally multiplying by $\sigma^{-\frac{k}{2}}\rho_k^{-1}$ and preparing for an application of (4.7) we obtain

$$\begin{aligned}&\frac{1}{\rho_k \sigma^{\frac{k}{2}}} \frac{\mathrm{d}}{\mathrm{d}t} \big\langle D\mathcal{E}(\boldsymbol{U}(\mu,\sigma)) \mid \boldsymbol{g}\big\rangle \\ &= \Big\langle \boldsymbol{\beta}(\mu,\sigma) - (0, \Lambda Q_{\underline{\mu\sigma}}) \mid J \circ \big[D\mathcal{E}(\boldsymbol{U}(\mu,\sigma)+\boldsymbol{g}) - D\mathcal{E}(\boldsymbol{U}(\mu,\sigma)) - D^2\mathcal{E}(\boldsymbol{U}(\mu,\sigma))\boldsymbol{g}\big]\Big\rangle \\ &\quad + \Big\langle (0, \Lambda Q_{\underline{\mu\sigma}}) \mid J \circ \big[D\mathcal{E}(\boldsymbol{U}(\mu,\sigma)+\boldsymbol{g}) - D\mathcal{E}(\boldsymbol{U}(\mu,\sigma)) - D^2\mathcal{E}(\boldsymbol{U}(\mu,\sigma)\boldsymbol{g}\big]\Big\rangle \\ &\quad + \frac{1}{\rho_k \sigma^{\frac{k}{2}}} \mu' \big\langle D^2\mathcal{E}(\boldsymbol{U}(\mu,\sigma))\partial_\mu \boldsymbol{U}(\mu,\sigma) \mid \boldsymbol{g}\big\rangle \\ &\quad + \frac{1}{\rho_k \sigma^{\frac{k}{2}}} \left(\sigma' + \frac{\xi(\sigma)}{\mu}\right) \big\langle D^2\mathcal{E}(\boldsymbol{U}(\mu,\sigma)\partial_\sigma \boldsymbol{U}(\mu,\sigma) \mid \boldsymbol{g}\big\rangle\end{aligned} \tag{4.12}$$

Consider the first line in (4.12). Using (4.7) we have,

$$\begin{aligned}&\Big| \Big\langle \boldsymbol{\beta}(\mu,\sigma) - (0, \Lambda Q_{\underline{\mu\sigma}}) \mid J \circ \big[D\mathcal{E}(\boldsymbol{U}(\mu,\sigma)+\boldsymbol{g}) - D\mathcal{E}(\boldsymbol{U}(\mu,\sigma)) - D^2\mathcal{E}(\boldsymbol{U}(\mu,\sigma))\boldsymbol{g}\big]\Big\rangle \Big| \\ &\qquad\qquad \lesssim o(1) \frac{\|\boldsymbol{g}\|_{\mathcal{H}}^2}{\mu\sigma}\end{aligned}$$

Next, consider the second term on the right in (4.12). Note that,

$$\begin{aligned}&\Big\langle (0, \Lambda Q_{\underline{\mu\sigma}}) \mid J \circ \big[D\mathcal{E}(\boldsymbol{U}(\mu,\sigma)+\boldsymbol{g}) - D\mathcal{E}(\boldsymbol{U}(\mu,\sigma)) - D^2\mathcal{E}(\boldsymbol{U}(\mu,\sigma)\boldsymbol{g}\big]\Big\rangle \\ &\qquad = -\frac{1}{\mu\sigma} \big\langle \Lambda Q_{\mu\sigma} \mid r^{-2}\big(f(U(\mu,\sigma)+g) - f(U(\mu,\sigma)) - f'(U(\mu,\sigma))g\big)\big\rangle\end{aligned} \tag{4.13}$$

We write,

$$\begin{aligned}&f(U(\mu,\sigma)+g) - f(U(\mu,\sigma)) - f'(U(\mu,\sigma))g \\ &\qquad\qquad = \frac{1}{2} f''(Q_{\mu\sigma})g^2 + \frac{1}{2}\big(f''(U(\mu,\sigma)) - f''(Q_{\mu\sigma})\big)g^2 + O(|g|^3)\end{aligned}$$

One can readily show using $U(\mu,\sigma) = \Phi(\mu,\sigma) + w_c(\mu,\sigma)$, the definition of $\Phi(\mu,\sigma)$ and the estimates (3.10), (3.11), (3.12) and (3.7) that the last two terms above contribute negligible errors, i.e., errors of size $o(1)(\mu\sigma)^{-1}\|g\|_H^2$. Hence,

$$-\frac{1}{\mu\sigma}\left\langle \Lambda Q_{\mu\sigma} \mid r^{-2}\big(f(U(\mu,\sigma)+g) - f(U(\mu,\sigma)) - f'(U(\mu,\sigma))g\big)\right\rangle \\ = -\frac{1}{\mu\sigma}\frac{1}{2}\left\langle \Lambda Q_{\mu\sigma} \mid r^{-2}f''(Q_{\mu\sigma})g^2\right\rangle + o(1)O\Big(\frac{\|g\|_H^2}{\mu\sigma}\Big)$$

We integrate by parts in the first term as follows,

$$\begin{aligned}-\frac{1}{2}\left\langle \Lambda Q_{\mu\sigma} \mid r^{-2}f''(Q_{\mu\sigma})g^2\right\rangle &= -\frac{1}{2}\int_0^\infty \partial_r(f'(Q_{\mu\sigma}) - k^2)g^2\,\mathrm{d}r \\ &= \int_0^\infty r^{-2}(f'(Q_{\mu\sigma}) - k^2)g\Lambda g\, r\,\mathrm{d}r = \langle \Lambda g \mid P_{\mu\sigma}g\rangle \\ &= \langle \Lambda_0 g \mid P_{\mu\sigma}g\rangle - \langle g \mid P_{\mu\sigma}g\rangle\end{aligned}$$

where $P_{\mu\sigma}$ is as in (2.2). Plugging all of this back into (4.13) we have show that

$$\Big\langle (0,\Lambda Q_{\underline{\mu\sigma}}) \mid J\circ\big[D\mathcal{E}(\boldsymbol{U}(\mu,\sigma)+\boldsymbol{g}) - D\mathcal{E}(\boldsymbol{U}(\mu,\sigma)) - D^2\mathcal{E}(\boldsymbol{U}(\mu,\sigma)\boldsymbol{g}\big]\Big\rangle \\ = \frac{1}{\mu\sigma}\langle \Lambda_0 g \mid P_{\mu\sigma}g\rangle - \frac{1}{\mu\sigma}\langle g \mid P_{\mu\sigma}g\rangle + o(1)O\Big(\frac{\|g\|_H^2}{\mu\sigma}\Big)$$

Next, applying the estimates (3.36) and (3.19) we have,

$$\Big|\frac{\mu'}{\rho_k\sigma^{\frac{k}{2}}}\left\langle D^2\mathcal{E}(\boldsymbol{U}(\mu,\sigma))\partial_\mu \boldsymbol{U}(\mu,\sigma) \mid \boldsymbol{g}\right\rangle\Big| \lesssim o(1)\frac{\|\boldsymbol{g}\|_{\mathcal{H}}^2}{\mu\sigma}$$

which takes care of the third term on the right-hand side of (4.12). Next, consider the last line of (4.12). Using the estimate (3.20) followed by (3.37) we have,

$$\begin{aligned}\frac{1}{\rho_k\sigma^{\frac{k}{2}}}\left(\sigma' + \frac{\xi(\sigma)}{\mu}\right)\left\langle D^2\mathcal{E}(\boldsymbol{U}(\mu,\sigma)\partial_\sigma\boldsymbol{U}(\mu,\sigma) \mid \boldsymbol{g}\right\rangle &= \frac{1}{\rho_k\sigma^{\frac{k}{2}}}\left(\sigma' + \frac{\xi(\sigma)}{\mu}\right)\gamma_k\frac{\sigma^{\frac{k}{2}}}{\rho_k\sigma}\left\langle \Lambda Q_{\underline{\mu\sigma}} \mid \dot g\right\rangle \\ &\quad - \frac{1}{\rho_k\sigma^{\frac{k}{2}}}\left(\sigma' + \frac{\xi(\sigma)}{\mu}\right)\frac{\rho_k\sigma^{\frac{k}{2}}}{\sigma}\left\langle \Lambda_0\Lambda Q_{\underline{\mu\sigma}} \mid \dot g\right\rangle + o(1)O(\frac{\sigma^{\frac{k}{2}}}{\sigma}\|\boldsymbol{g}\|_{\mathcal{H}})\frac{1}{\rho_k\sigma^{\frac{k}{2}}}\left(\sigma' + \frac{\xi(\sigma)}{\mu}\right) \\ &= -\frac{1}{\mu\sigma}\frac{k}{2}\frac{\left\langle \Lambda Q_{\underline{\mu\sigma}} \mid \dot g\right\rangle^2}{\|\Lambda Q\|_{L^2}^2} - \left(\frac{\sigma'}{\sigma} + \frac{\xi(\sigma)}{\mu\sigma}\right)\left\langle \Lambda_0\Lambda Q_{\underline{\mu\sigma}} \mid \dot g\right\rangle + o(1)O\Big(\frac{\|\boldsymbol{g}\|_{\mathcal{H}}^2}{\mu\sigma}\Big)\end{aligned}$$

where in the last line we used (1.6), i.e., $\gamma_k = \rho_k^2\frac{k}{2}$. To recap, by inserting the previous three estimates into (4.12) we have now shown that,

$$\begin{aligned}\frac{1}{\rho_k\sigma^{\frac{k}{2}}}\frac{\mathrm{d}}{\mathrm{d}t}\langle D\mathcal{E}(\boldsymbol{U}(\mu,\sigma)) \mid \boldsymbol{g}\rangle &= -\frac{1}{\mu\sigma}\frac{k}{2}\frac{\left\langle \Lambda Q_{\underline{\mu\sigma}} \mid \dot g\right\rangle^2}{\|\Lambda Q\|_{L^2}^2} \\ &\quad - \left(\frac{\sigma'}{\sigma} + \frac{\xi(\sigma)}{\mu\sigma}\right)\left\langle \Lambda_0\Lambda Q_{\underline{\mu\sigma}} \mid \dot g\right\rangle + \frac{1}{\mu\sigma}\langle \Lambda_0 g \mid P_{\mu\sigma}g\rangle - \frac{1}{\mu\sigma}\langle g \mid P_{\mu\sigma}g\rangle \\ &\quad + o(1)O\Big(\frac{\|\boldsymbol{g}\|_{\mathcal{H}}^2}{\mu\sigma}\Big)\end{aligned} \tag{4.14}$$

Note that the first term on the right above exactly cancels the first term on the left of (4.10). The terms on the second line are of critical size, and we now show that the differentiated virial correction

will cancel these terms up to admissible errors and a coercive term. Indeed, we claim the estimate,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\langle \mathcal{A}_0(\mu\sigma)g \mid \dot g\rangle + \frac{1}{\mu\sigma}\langle \Lambda_0 g \mid P_{\mu\sigma} g\rangle - \Big(\frac{\sigma'}{\sigma}+\frac{\xi(\sigma)}{\mu\sigma}\Big)\langle \mathcal{A}_0(\mu\sigma)\Lambda Q_{\mu\sigma} \mid \dot g\rangle \\ \leq c_0\frac{\|g\|_H^2}{\mu\sigma} - \frac{1}{\mu\sigma}\int_0^{R\mu\sigma}\Big((\partial_r g)^2 + k^2\frac{g^2}{r^2}\Big)\, r\,\mathrm{d}r\end{aligned} \tag{4.15}$$

where, $c_0>0$, $R>0$ are as in Lemma 2.3 and $c_0>0$ can be taken as small as we like independent of $\mu\sigma$, and $P_{\mu\sigma}$ is as in (2.2). To see this, we expand the derivative of the virial correction as follows. Using the notation, $\lambda=\mu\sigma$ we have,

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle \mathcal{A}_0(\mu\sigma)g \mid \dot g\rangle = (\frac{\mu'}{\mu}+\frac{\sigma'}{\sigma})\langle [\lambda\partial_\lambda \mathcal{A}_0](\lambda)g \mid \dot g\rangle + \langle \mathcal{A}_0(\mu\sigma)\partial_t g \mid \dot g\rangle + \langle \mathcal{A}_0(\mu\sigma)g \mid \partial_t \dot g\rangle \tag{4.16}$$

Note that the first term on the right above contributes an admissible error. Indeed, using the estimates (3.36), (3.37), (3.10), and the first bullet point in Lemma 2.3 we have,

$$\Big|(\frac{\mu'}{\mu}+\frac{\sigma'}{\sigma})\langle [\lambda\partial_\lambda \mathcal{A}_0](\lambda)g \mid \dot g\rangle\Big| \lesssim \Big|\frac{\mu'}{\mu}+\frac{\sigma'}{\sigma}\Big|\,\|\boldsymbol{g}\|_{\mathcal{H}}^2 \lesssim (\sigma^{\frac{k}{2}}+\|\boldsymbol{g}\|_{\mathcal{H}})\frac{\|\boldsymbol{g}\|_{\mathcal{H}}^2}{\mu\sigma}$$

Next, we expand the second two terms on the right of (4.16) using the equation satisfied by $\boldsymbol{g}$ in (3.39). For the second term we have,

$$\begin{aligned}\langle \mathcal{A}_0(\mu\sigma)\partial_t g \mid \dot g\rangle = \langle \mathcal{A}_0(\mu\sigma)\dot g \mid \dot g\rangle - \mu'\langle \mathcal{A}_0(\mu\sigma)\partial_\mu U(\mu,\sigma) \mid \dot g\rangle \\ - (\sigma'+\frac{\xi(\sigma)}{\mu})\langle \mathcal{A}_0(\mu\sigma)\partial_\sigma U(\mu,\sigma) \mid \dot g\rangle\end{aligned} \tag{4.17}$$

Since $\mathcal{A}_0(\mu\sigma)$ is antisymmetric, we have, $\langle \mathcal{A}_0(\mu\sigma)\dot g \mid \dot g\rangle = 0$. For the second term on the right above we use (3.3) and the first bullet point in Lemma 2.3 to deduce that

$$|\mu'\langle \mathcal{A}_0(\mu\sigma)\partial_\mu U(\mu,\sigma) \mid \dot g\rangle| \lesssim \frac{|\mu'|}{\mu}(\|\Lambda\Phi(\mu,\sigma)\|_H + \|\Lambda w_c(\mu,\sigma)\|_H)\|\dot g\|_{L^2} \lesssim \frac{\|\boldsymbol{g}\|_{\mathcal{H}}^2}{\mu} \lesssim o(1)\frac{\|\boldsymbol{g}\|_{\mathcal{H}}^2}{\mu\sigma}$$

where the last inequality follows from (3.9) and (3.36). Next, we treat the last term in (4.17). Using (3.4) we write,

$$\begin{aligned}-(\sigma'+\frac{\xi(\sigma)}{\mu})\langle \mathcal{A}_0(\mu\sigma)\partial_\sigma U(\mu,\sigma) \mid \dot g\rangle &= \Big(\frac{\sigma'}{\sigma}+\frac{\xi(\sigma)}{\mu\sigma}\Big)\Big\langle \mathcal{A}_0(\mu\sigma)\frac{\mu\sigma}{\xi(\sigma)}\dot U(\mu,\sigma) \mid \dot g\Big\rangle \\ &= \Big(\frac{\sigma'}{\sigma}+\frac{\xi(\sigma)}{\mu\sigma}\Big)\langle \mathcal{A}_0(\mu\sigma)\Lambda Q_{\mu\sigma} \mid \dot g\rangle + (\mu\sigma'+\xi(\sigma))\Big\langle \mathcal{A}_0(\mu\sigma)\Big(\frac{b_c(\lambda_c^{-1}(\sigma))}{\xi(\sigma)}-1\Big)\Lambda Q_{\underline{\mu\sigma}} \mid \dot g\Big\rangle \\ &\quad + (\mu\sigma'+\xi(\sigma))\Big\langle \mathcal{A}_0(\mu\sigma)\frac{1}{\xi(\sigma)}\Big(\dot\Phi(\mu,\sigma) - b_c(\lambda_c^{-1}(\sigma))\Lambda Q_{\underline{\mu\sigma}}\Big) \mid \dot g\Big\rangle \\ &\quad + (\mu\sigma'+\xi(\sigma))\Big\langle \mathcal{A}_0(\mu\sigma)\frac{1}{\xi(\sigma)}\dot w_c(\mu,\sigma) \mid \dot g\Big\rangle \\ &= \Big(\frac{\sigma'}{\sigma}+\frac{\xi(\sigma)}{\mu\sigma}\Big)\langle \mathcal{A}_0(\mu\sigma)\Lambda Q_{\mu\sigma} \mid \dot g\rangle + o(1)O\Big(\frac{\|\boldsymbol{g}\|_{\mathcal{H}}^2}{\mu\sigma}\Big)\end{aligned}$$

where the last line follows from the first bullet point in Lemma 2.3 with (3.10), (3.11), (3.12), and (3.23), and finally (3.8). Plugging the previous three estimates back into (4.17) we obtain,

$$\langle \mathcal{A}_0(\mu\sigma)\partial_t g \mid \dot g\rangle = \Big(\frac{\sigma'}{\sigma}+\frac{\xi(\sigma)}{\mu\sigma}\Big)\langle \mathcal{A}_0(\mu\sigma)\Lambda Q_{\mu\sigma} \mid \dot g\rangle + o(1)O\Big(\frac{\|\boldsymbol{g}\|_{\mathcal{H}}^2}{\mu\sigma}\Big)$$

Lastly, we expand the term in (4.16) involving $\partial_t\dot g$ using (3.39). Preparing for a near identical argument to the one used to treat the virial correction in the companion paper [12, Proof of Lemma 4.6]

we write,

$$\begin{aligned}\langle \mathcal{A}_0(\mu\sigma)g \mid \partial_t \dot{g}\rangle &= \langle \mathcal{A}_0(\mu\sigma)g \mid (-\mathcal{L}_{\mu\sigma})g\rangle \\ &\quad - \Big\langle \mathcal{A}_0(\mu\sigma)g \mid r^{-2}\Big(f(U(\mu,\sigma)+g) - f(U(\mu,\sigma)) - f'(Q_{\mu\sigma})g\Big)\Big\rangle \\ &\quad - \mu' \Big\langle \mathcal{A}_0(\mu\sigma)g \mid \dot{\partial_\mu U}(\mu,\sigma)\Big\rangle - (\sigma' + \frac{\xi(\sigma)}{\mu}) \Big\langle \mathcal{A}_0(\mu\sigma)g \mid \dot{\partial_\sigma U}(\mu,\sigma)\Big\rangle\end{aligned} \tag{4.18}$$

For the first term on the right of (4.18) we recall the notation $\mathcal{L}_{\mu\sigma} = \mathcal{L}_0 + P_{\mu\sigma}$ and write,

$$\langle \mathcal{A}_0(\mu\sigma)g \mid (-\mathcal{L}_{\mu\sigma})g\rangle = -\langle \mathcal{A}_0(\mu\sigma)g \mid \mathcal{L}_0 g\rangle - \langle \mathcal{A}_0(\mu\sigma)g \mid P_{\mu\sigma}g\rangle$$

It then follows from (2.7) from Lemma 2.5 along with the estimate (2.9) (to treat the second term above) that,

$$\langle \mathcal{A}_0(\mu\sigma)g \mid (-\mathcal{L}_{\mu\sigma})g\rangle \leq c_0 \frac{\|g\|_H^2}{\mu\sigma} - \frac{1}{\mu\sigma}\int_0^{R\mu\sigma} \Big((\partial_r g)^2 + k^2\frac{g^2}{r^2}\Big)\, r\,\mathrm{d}r - \frac{1}{\mu\sigma}\langle \Lambda_0 g \mid P_{\mu\sigma}g\rangle$$

Note that $c_0 > 0$, $R > 0$ are as in Lemma 2.3 and $c_0 > 0$ can be taken as small as we like independent of $\mu\sigma$. Next, we estimate the second term on the right of (4.18) via an analysis nearly identical to the one used to estimate the second term in [12, Eqn. (4.30)]. The difference is that here we can only make use of the $\mathcal{H}$ regularity of g. First, note that by Lemma 2.3 we have,

$$\begin{aligned}&\Big|\Big\langle \mathcal{A}_0(\mu\sigma)g \mid r^{-2}\Big(f(U(\mu,\sigma)+g) - f(U(\mu,\sigma)) - f'(Q_{\mu\sigma})g\Big)\Big\rangle\Big| \\ &\qquad \lesssim \|g\|_H \|r^{-2}\big(f(U(\mu,\sigma)+g) - f(U(\mu,\sigma)) - f'(Q_{\mu\sigma})g\big)\|_{L^2}\end{aligned}$$

Hence it suffices to establish the estimate,

$$\|r^{-2}\big(f(U(\mu,\sigma)+g) - f(U(\mu,\sigma)) - f'(Q_{\mu\sigma})g\big)\|_{L^2} \lesssim o(1)\frac{\|g\|_H}{\mu\sigma} \tag{4.19}$$

To see this, we write,

$$\begin{aligned}&f(U(\mu,\sigma)+g) - f(U(\mu,\sigma)) - f'(Q_{\mu\sigma})g \\ &\qquad = f(\Phi(\mu,\sigma) + w_c(\mu,\sigma) + g) - f(\Phi(\mu,\sigma) + w_c(\mu,\sigma)) - f'(\Phi(\mu,\sigma) + w_c(\mu,s))g \\ &\qquad\quad + \Big(f'(\Phi(\mu,\sigma) + w_c(\mu,s)) - f'(\Phi(\mu,\sigma))\Big)g + \Big(f'(\Phi) - f'(Q_{\mu\sigma})\Big)g\end{aligned}$$

The contribution of the first line is handled using the pointwise estimate,

$$\frac{1}{r^2}\Big|f(\Phi(\mu,\sigma) + w_c(\mu,\sigma) + g) - f(\Phi(\mu,\sigma) + w_c(\mu,\sigma)) - f'(\Phi(\mu,\sigma) + w_c(\mu,s))g\Big| \lesssim \frac{1}{\sigma\mu} r^{-1} g^2$$

which follows from the definition of $\Phi(\mu,\sigma)$, (3.10) (3.11), (3.12) and (3.7). For the second term we use the pointwise estimate,

$$r^{-2}\left|\Big(f'(\Phi(\mu,\sigma) + w_c(\mu,s)) - f'(\Phi(\mu,\sigma))\Big)g\right| \lesssim \frac{1}{\mu\sigma} r^{-1}\,|w_c(\mu,\sigma)|\,|g|$$

together with (3.7). Finally, to treat the last term we note the pointwise estimate,

$$r^{-2}\,|f'(\Phi) - f'(Q_{\mu\sigma})| \lesssim \frac{1}{\mu} r^{-1} \lesssim o(1)\frac{1}{\mu\sigma} r^{-1}$$

This is sufficient to prove (4.19). Next, we use (3.36) and (3.3) to estimate,

$$\Big|\mu' \Big\langle \mathcal{A}_0(\mu\sigma)g \mid \dot{\partial_\mu U}(\mu,\sigma)\Big\rangle\Big| \lesssim \frac{\|\boldsymbol{g}\|_H^2}{\mu}\Big(\|\Lambda_0 \dot{\Phi}(\mu,\sigma)\|_{L^2} + \|\Lambda_0 w_c(\mu,\sigma)\|_{L^2}\Big) \lesssim o(1)\frac{\|\boldsymbol{g}\|_{\mathcal{H}}^2}{\mu\sigma}$$

where in the last inequality we used (3.8). Lastly, using the first bullet point in Lemma 2.3 (3.37) and (3.4) we have,

$$\begin{aligned}\left|(\sigma'+\frac{\xi(\sigma)}{\mu})\left\langle \mathcal{A}_0(\mu\sigma)g \mid \partial_\sigma \dot{U}(\mu,\sigma)\right\rangle\right| &\lesssim \frac{1}{\mu}\|\boldsymbol{g}\|_{\mathcal{H}}^2\frac{\mu}{\xi(\sigma)}\|\Delta U(\mu,\sigma)-\frac{1}{r^2}f(U(\mu,\sigma))\|_{L^2}\\ &\lesssim \sigma^{\frac{k}{2}}\frac{\|\boldsymbol{g}\|_{\mathcal{H}}^2}{\mu\sigma}\end{aligned}$$

where the last inequality is by (3.18) and (3.10). This completes the proof of (4.15).

We can now complete the proof of (4.10). Combining the estimates (4.14) and (4.15) we have,

$$\begin{aligned}&\frac{k\left\langle \Lambda Q_{\underline{\mu\sigma}} \mid \dot{g}\right\rangle^2}{2\mu\sigma\|\Lambda Q\|_{L^2}^2}+\frac{1}{\rho_k\sigma^{\frac{k}{2}}}\frac{\mathrm{d}}{\mathrm{d}t}\left\langle D\mathcal{E}(\boldsymbol{U}(\mu,\sigma)) \mid \boldsymbol{g}\right\rangle+\frac{\mathrm{d}}{\mathrm{d}t}\left\langle \mathcal{A}_0(\mu\sigma)g \mid \dot{g}\right\rangle\\ &\qquad\le c_0\frac{\|\boldsymbol{g}\|_{\mathcal{H}}^2}{\mu\sigma}+\left(\frac{\sigma'}{\sigma}+\frac{\xi(\sigma)}{\mu\sigma}\right)\left(\left\langle \mathcal{A}_0(\mu\sigma)\Lambda Q_{\mu\sigma} \mid \dot{g}\right\rangle-\left\langle \Lambda_0\Lambda Q_{\underline{\mu\sigma}} \mid \dot{g}\right\rangle\right)\\ &\qquad-\frac{1}{\mu\sigma}\int_0^{R\mu\sigma}\left((\partial_r g)^2+k^2\frac{g^2}{r^2}\right)r\,\mathrm{d}r-\frac{1}{\mu\sigma}\left\langle g \mid P_{\mu\sigma}g\right\rangle\end{aligned}$$

Using the estimate (2.8) along with (3.37) we have,

$$\begin{aligned}\left|\left(\frac{\sigma'}{\sigma}+\frac{\xi(\sigma)}{\mu\sigma}\right)\left\langle (\mathcal{A}_0(\mu\sigma)\Lambda Q_{\mu\sigma}-\Lambda_0\Lambda Q_{\underline{\mu\sigma}}) \mid \dot{g}\right\rangle\right| &\le \frac{1}{\mu\sigma}\|\mathcal{A}_0(\mu\sigma)\Lambda Q_{\mu\sigma}-\Lambda_0\Lambda Q_{\underline{\mu\sigma}}\|_{L^2}\|\boldsymbol{g}\|_{\mathcal{H}}^2\\ &\le c_0\frac{\|\boldsymbol{g}\|_{\mathcal{H}}^2}{\mu\sigma}\end{aligned}$$

Finally, the localized coercivity estimate (2.4) from Lemma 2.1 yields,

$$-\frac{1}{\mu\sigma}\int_0^{R\mu\sigma}\left((\partial_r g)^2+k^2\frac{g^2}{r^2}\right)r\,\mathrm{d}r-\frac{1}{\mu\sigma}\left\langle g \mid P_{\mu\sigma}g\right\rangle\le c_0\frac{\|\boldsymbol{g}\|_{\mathcal{H}}^2}{\mu\sigma}$$

by taking $R>0$ large enough. We conclude that,

$$\frac{k\left\langle \Lambda Q_{\underline{\mu\sigma}} \mid \dot{g}\right\rangle^2}{2\mu\sigma\|\Lambda Q\|_{L^2}^2}+\frac{1}{\rho_k\sigma^{\frac{k}{2}}}\frac{\mathrm{d}}{\mathrm{d}t}\left\langle D\mathcal{E}(\boldsymbol{U}(\mu,\sigma)) \mid \boldsymbol{g}\right\rangle+\frac{\mathrm{d}}{\mathrm{d}t}\left\langle \mathcal{A}_0(\mu\sigma)g \mid \dot{g}\right\rangle\le c_0\frac{\|\boldsymbol{g}\|_{\mathcal{H}}^2}{\mu\sigma}$$

where c_0 is a constant that can be taken arbitrarily small, independently of μ,σ. This proves (4.10) and completes the proof of the proposition. □

4.3. **The proof of Theorem 1.1.** We put the finishing touches on the the proof of Theorem 1.1.

Proof. We pick up where we left off in Section 4.1. Let $\boldsymbol{u}(t)\in\mathcal{H}$ be any forward-in-time 2-bubble solution to (1.1) on the time interval $[T_0,\infty)$ where $T_0>0$ is chosen sufficiently large so that Corollary 4.4 holds, as well as (4.1). Assume for contradiction that $\|\boldsymbol{g}(t)\|_{\mathcal{H}}>0$ for all $t\ge T_0$. So on the one hand, by (4.1) we have,

$$0\ge\left\langle D\mathcal{E}(\boldsymbol{U}(\mu(t),\sigma(t))) \mid \boldsymbol{g}(t)\right\rangle+\frac{1}{2}c_1\|\boldsymbol{g}(t)\|_{\mathcal{H}}^2. \tag{4.20}$$

for all $t\in[T_0,\infty)$ for a uniform constant $c_1>0$. One the other hand, by Corollary 4.4, the definition of $b(t)$ in (4.4), and by possibly taking T_0 larger so that $\sigma(t)$ is sufficiently small, we can find $T_1\ge T_0$ so that,

$$\left\langle D\mathcal{E}(\boldsymbol{U}(\mu(T_1),\sigma(T_1))) \mid \boldsymbol{g}(T_1)\right\rangle\ge-\frac{c_1}{4}\|\boldsymbol{g}(T_1)\|_{\mathcal{H}}^2$$

which yields a contradiction in (4.20) at time T_1. Thus, there exists $T\ge T_0$ for which $\|\boldsymbol{g}(T)\|_{\mathcal{H}}=0$. But this means that,

$$\boldsymbol{u}(T)=\boldsymbol{U}(\mu(T),\sigma(T))=(u_c(\lambda_c^{-1}(\sigma(T)),\cdot/\mu(T)),\mu(T)^{-1}\partial_t u_c(\lambda_c^{-1}(\sigma(T)),\cdot/\mu(T))),$$

i.e. $\boldsymbol{u}(t)$ agrees with $\boldsymbol{u}_c(t)$ up to a fixed time translation and rescaling. This completes the proof. □

References

[1] D. Christodoulou and A. S. Tahvildar-Zadeh. On the asymptotic behavior of spherically symmetric wave maps. *Duke Math. J.*, 71(1):31–69, 1993.

[2] D. Christodoulou and A. S. Tahvildar-Zadeh. On the regularity of spherically symmetric wave maps. *Comm. Pure Appl. Math.*, 46(7):1041–1091, 1993.

[3] R. Côte, C. Kenig, A. Lawrie, and W. Schlag. Characterization of large energy solutions of the equivariant wave map problem: I. *Amer. J. Math.*, 137(1):139–207, 2015.

[4] T. Duyckaerts and F. Merle. Dynamics of threshold solutions for energy-critical wave equation. *Int. Math. Res. Pap. IMRP*, 2008.

[5] T. Duyckaerts and F. Merle. Dynamic of threshold solutions for energy-critical NLS. *GAFA*, 18(6):1787–1840, 2009.

[6] D.-A. Geba and M. G. Grillakis. *An introduction to the theory of wave maps and related geometric problems*. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2017.

[7] J. Jendrej. Construction of two-bubble solutions for the energy-critical NLS. *Anal. PDE*, 10(8):1923–1959, 2017.

[8] J. Jendrej. Nonexistence of two-bubbles with opposite signs for the radial energy-critical wave equation. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)*, 18(2):735–778, 2018.

[9] J. Jendrej. Construction of two-bubble solutions for energy-critical wave equations. *Amer. J. Math.*, 141(1):55–118, 2019.

[10] J. Jendrej, M. Kowalczyk, and Lawrie A. Dynamics of strongly interacting kink-antikink pairs for scalar fields on a line. *arXiv E-Prints*, 2019.

[11] J. Jendrej and A. Lawrie. Two-bubble dynamics for threshold solutions to the wave maps equation. *Invent. Math.*, 213(3):1249–1325, 2018.

[12] J. Jendrej and A. Lawrie. Regularity of two-bubble wave maps. *ArXiv e-prints*, 2020.

[13] J. Jendrej, A. Lawrie, and C. Rodriguez. Bubbling dynamics for wave maps with prescribed radiation. *arXiv E-Prints*, 2019.

[14] S. Klainerman and M. Machedon. Space-time estimates for null forms and the local existence theorem. *Comm. Pure Appl. Math.*, 46(9):1221–1268, 1993.

[15] S. Klainerman and M. Machedon. Smoothing estimates for null forms and applications. *Internat. Math. Res. Notices*, (9), 1994.

[16] S. Klainerman and M. Machedon. Smoothing estimates for null forms and applications. *Duke Math. J.*, 81(1):99–133, 1995.

[17] S. Klainerman and M. Machedon. On the regularity properties of a model problem related to wave maps. *Duke Math. J.*, 87(3):553–589, 1997.

[18] S. Klainerman and S. Selberg. Remark on the optimal regularity for equations of wave maps type. *Comm. Partial Differential Equations*, 22(5-6):901–918, 1997.

[19] S. Klainerman and S. Selberg. Bilinear estimates and applications to nonlinear wave equations. *Commun. Contemp. Math.*, 4(2):223–295, 2002.

[20] J. Krieger. Global regularity of wave maps from $\mathbf{R}^{2+1}$ to H^2. Small energy. *Comm. Math. Phys.*, 250(3):507–580, 2004.

[21] J. Krieger and S. Miao. On stability of blow up solutions for the critical co-rotational wave maps problem. *To appear in Duke Math.*, preprint 2018.

[22] J. Krieger and W. Schlag. *Concentration Compactness for critical wave maps*. EMS Monographs. European Mathematical Society, Zürich, 2012.

[23] J. Krieger, W. Schlag, and D. Tataru. Renormalization and blow up for charge one equivariant wave critical wave maps. *Invent. Math.*, 171(3):543–615, 2008.

[24] A. Lawrie and S.-J. Oh. A refined threshold theorem for $(1 + 2)$-dimensional wave maps into surfaces. *Comm. Math. Phys.*, 342(3):989–999, 2016.

[25] N. Manton and P. Sutcliffe. *Topological solitons*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, Cambridge, 2004.

[26] Y. Martel. Asymptotic N-soliton-like solutions of the subcritical and critical generalized Korteweg-de Vries equations. *Amer. J. Math.*, 127(5):1103–1140, 2005.

[27] Y. Martel and F. Merle. Description of two soliton collision for the quartic gKdV equation. *Ann. of Math. (2)*, 174(2):757–857, 2011.

[28] Y. Martel and F. Merle. Inelastic interaction of nearly equal solitons for the quartic gKdV equation. *Invent. Math.*, 183(3):563–648, 2011.

[29] Y. Martel and F. Merle. Inelasticity of soliton collisions for the 5D energy critical wave equation. *Invent. Math.*, 214(3):1267–1363, 2018.

[30] F. Merle. Determination of blow-up solutions with minimal mass for nonlinear Schrödinger equations with critical power. *Duke Math. J.*, 69(2):427–454, 1993.
[31] M. Pillai. A continuum of infinite time blow-up solutions to the energy critical wave maps equation. *arXiv E-Prints*, 2019.
[32] P. Raphaël and I. Rodnianski. Stable blow up dynamics for the critical co-rotational wave maps and equivariant Yang-Mills problems. *Publ. Math. Inst. Hautes Études Sci.*, pages 1–122, 2012.
[33] P. Raphaël and J. Szeftel. Existence and uniqueness of minimal mass blow up solutions to an inhomogeneous L^2-critical NLS. *J. Amer. Math. Soc.*, 24(2):471–546, 2011.
[34] I. Rodnianski and J. Sterbenz. On the formation of singularities in the critical $O(3)$ σ-model. *Ann. of Math.*, 172:187–242, 2010.
[35] C. Rodriguez. Threshold dynamics for corotational wave maps. *arXiv E-Prints*, 2018.
[36] J. Shatah and M. Struwe. *Geometric Wave Equations*, volume 2 of *Courant Lecture Notes in Mathematics*. AMS, 2000.
[37] J. Shatah and A. Tahvildar-Zadeh. Regularity of harmonic maps from the Minkowski space into rotationally symmetric manifolds. *Comm. Pure Appl. Math.*, 45(8):947–971, 1992.
[38] J. Shatah and A. S. Tahvildar-Zadeh. On the Cauchy problem for equivariant wave maps. *Comm. Pure Appl. Math.*, 47(5):719–754, 1994.
[39] J. Sterbenz and D. Tataru. Energy dispersed large data wave maps in $2 + 1$ dimensions. *Comm. Math. Phys.*, (1):139–230, 2010.
[40] J. Sterbenz and D. Tataru. Regularity of wave maps in $2+1$ dimensions. *Comm. Math. Phys.*, (1):231–264, 2010.
[41] M. Struwe. Equivariant wave maps in two space dimensions. *Comm. Pure Appl. Math.*, 56(7):815–823, 2003.
[42] T. Tao. Global regularity of wave maps. I. Small critical Sobolev norm in high dimension. *Internat. Math. Res. Notices*, (6):299–328, 2001.
[43] T. Tao. Global regularity of wave maps II: Small energy in two dimensions. *Comm. Math. Phys.*, 224(2):443–544, 2001.
[44] T. Tao. Global regularity of wave maps III–VII. *Preprints*, 2008–2009.
[45] D. Tataru. On global existence and scattering for the wave maps equation. *Amer. J. Math.*, 123(1):37–77, 2001.
[46] Daniel Tataru. Local and global results for wave maps. I. *Comm. Partial Differential Equations*, 23(9-10):1781–1793, 1998.

Jacek Jendrej
CNRS and LAGA, Universit Sorbonne Paris Nord
99 av Jean-Baptiste Clément, 93430 Villetaneuse, France
jendrej@math.univ-paris13.fr

Andrew Lawrie
Department of Mathematics, Massachusetts Institute of Technology
77 Massachusetts Ave, 2-267, Cambridge, MA 02139, U.S.A.
alawrie@mit.edu