# ANALYSIS & PDE

Volume 8 No. 2 2015

msp

# Analysis & PDE

msp.org/apde




See inside back cover or msp.org/apde for submission instructions.

The subscription price for 2015 is US $205/year for the electronic version, and $390/year (+$55, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues from the last three years and changes of subscribers address should be sent to MSP.

Analysis & PDE (ISSN 1948-206X electronic, 2157-5045 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840, is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

APDE peer review and production are managed by EditFLOW® from MSP.

PUBLISHED BY
mathematical sciences publishers
nonprofit scientific publishing
http://msp.org/

ANALYSIS AND PDE
Vol. 8, No. 2, 2015
dx.doi.org/10.2140/apde.2015.8.257

# SMOOTH PARAMETRIC DEPENDENCE OF ASYMPTOTICS OF THE SEMICLASSICAL FOCUSING NLS

SERGEY BELOV AND STEPHANOS VENAKIDES

We consider the one-dimensional focusing (cubic) nonlinear Schrödinger equation (NLS) in the semiclassical limit with exponentially decaying complex-valued initial data, whose phase is multiplied by a real parameter. We prove smooth dependence of the asymptotic solution on the parameter. Numerical results supporting our estimates of important quantities are presented.

## 1. Introduction

We consider the semiclassical focusing nonlinear Schrödinger (NLS) equation

$$i\varepsilon\,\partial_t q + \varepsilon^2\,\partial_x^2 q + 2|q|^2 q = 0 \tag{1}$$

with the initial data

$$q(x,0) = A(x)e^{\frac{i\mu}{\varepsilon}S(x)}, \quad A(x), S(x) \in \mathbb{R}, \quad \mu \geq 0, \tag{2}$$

in the asymptotic limit $\varepsilon \to 0$. Equation (1) is a well-known integrable system [Zakharov and Shabat 1972], and a lot of work has been done on this initial value problem (see below). The focus of the present study is on the parameter $\mu$ in the exponent of the initial data. For the specific data

$$A(x) = -\operatorname{sech} x, \quad S'(x) = -\tanh x, \quad \mu \geq 0, \tag{3}$$

studied in [Kamvissis et al. 2003; Tovbis et al. 2004], the solution undergoes a transition at $\mu = 2$. When $\mu < 2$, the Lax spectrum contains discrete eigenvalues numbering $O\big(\frac{1}{\varepsilon}\big)$, each eigenvalue giving rise to a soliton in the solution, which thus consists of both a radiative and a solitonic part. When $\mu \geq 2$, the spectrum is purely continuous and the solution is purely radiative (absence of solitons). We prove that the local wave parameters (branch points of the Riemann surface that represents the asymptotic solution locally in space-time) vary smoothly with $\mu$, even at the critical value $\mu = 2$. Indeed, numerical experiments have shown absence of any noticeable transition in the behavior of the branch points at the critical value [Miller and Kamvissis 1998]. Theorem 4.5 establishes this fact rigorously.

The reason $\mu$ deserves special attention as a perturbation parameter is twofold. First, at the value of $\mu = 2$ there is a phase transition in the nature of the solution (there is a solitonic part when $\mu < 2$; see below). Perturbing $\mu$ across this value allows continuation of the validity of rigorously derived asymptotics [Tovbis et al. 2006] from the region $\mu \geq 2$ to the region $\mu < 2$. Ab initio derivation of such

Venakides thanks the NSF for supporting this work under grants NSF DMS-0707488 and NSF DMS-1211638.
*MSC2010:* primary 37K15, 37K40; secondary 35P25.
*Keywords:* NLS, semiclassical limit, Riemann–Hilbert problems.

asymptotics in the region $\mu < 2$ would be technically more demanding. Also, $\mu$ is a singularity of the RH contour in a way that cannot be remedied by contour deformations. Such a difficulty is absent when the perturbation parameters are space and time variables $x$ and $t$. Indeed, the methods of [Tovbis and Venakides 2009; 2010] are applied in this work, and amongst the surprises which allow the methods to apply is a collection of explicit formulae for dependence on $\mu$ summarized in Lemmas 4.1–4.3.

Essential mathematical difficulties are encountered in the solution of the initial value problem (1) and (2) in general, and (1) and (3) in particular.

(1) The calculation of the scattering data at $t = 0$ is extremely delicate, as seen from the work of Klaus and Shaw [2002].

(2) The linearizing Zakharov–Shabat eigenvalue problem [1972] is *not self-adjoint*. This is in contrast to the self-adjointness of the initial value problem for the small-dispersion Korteweg–de Vries (KdV) equation, in which a systematic steepest descent procedure was developed by Deift, Venakides and Zhou [Deift et al. 1997] for calculating the asymptotic solution (see also [Deift et al. 1994]). The approach in [Deift et al. 1997] extended the original steepest descent analysis of Deift and Zhou [1993] for oscillatory Riemann–Hilbert problems by adding to it the $g$-function mechanism. A systematic procedure then obtained the KdV solution, which consists of waves that are fully nonlinear. These waves are typically modulated. In other words, the oscillations are rapid, exhibiting wavenumbers and frequencies in the small spatiotemporal scale that vary in the large scale in accordance with *modulation equations*.

(3) The system of modulation equations in the form of a set of PDEs for the NLS equation exhibits complex characteristics [Forest and Lee 1986]. Posed naturally as an initial value problem, the system is thus *ill-posed* and modulated NLS waves are unstable. The instability to the large-scale spatio-temporal variation of the wave parameters (modulational instability) is the primary source of problems in nonlinear fiber optical transmission, which is governed by the NLS equation.

In spite of the modulational instability, there exist initial data with a particular combination of $A$ and $S$ that evolve into a profile of modulated waves. The ordered structure of modulated nonlinear waves was first observed numerically by Miller and Kamvissis [1998], for the initial data of (3) with $\mu = 0$ and values of $\varepsilon$ that allowed them to implement the multisoliton NLS formulae. Miller and Kamvissis observed the phenomenon of wave breaking (see below) and the formation of more complex wave structures past the break in this work. Later numerical findings by Ceniseros and Tian [2002], as well as by Cai, D. W. McLaughlin, and K. D. T.-R. McLaughlin [Cai et al. 2002], also detected the ordered structures.

These studies were followed by analytic work of Kamvissis, K. D. T.-R. McLaughlin and Miller [Kamvissis et al. 2003] for the same initial data ($\mu = 0$), the corresponding initial scattering data having been earlier calculated explicitly by Satsuma and Yajima [1974]. This work set forth a procedure aiming at the analytic determination of the observed phenomena that would practically extend the steepest descent procedure cited above to the non-self-adjoint case. Following a similar approach, Tovbis, Venakides and Zhou derived these phenomena rigorously from the initial data (3) with $\mu > 0$ [Tovbis et al. 2004]; the initial scattering data were previously obtained by Tovbis and Venakides [2000]. The asymptotic calculation of the wave solution was with error of order $\varepsilon$ for points in space-time that are off the break point and off the caustic curves (see below). In further work [Tovbis et al. 2006; 2007], the same authors also derived the long-time behavior of the asymptotic solution for $\mu \geq 2$ and generalized the prebreak analysis to a wide class of initial data. The detailed asymptotic behavior in a neighborhood of the first break point was derived recently by Bertola and Tovbis [2013].

The rigorous derivation of the mechanism of the second break remains an open problem. Using a combination of theoretical and numerical arguments, Lyng and Miller [2007] obtained significant insights for initial data (3) with $\mu = 0$ when the solution is an $N$-soliton, where $N = O\left(\frac{1}{\varepsilon}\right)$. In particular, they identified a mathematical mechanism for the second break, which depends essentially on the discrete nature of the spectrum of the $N$-soliton and turns out to differ from the mechanism of the first break.

The asymptotic solution for *shock initial data*,

$$A = \text{constant}, \quad S'(x) = \operatorname{sign} x, \quad \mu > 0, \tag{4}$$

was derived globally in time by Buckingham and Venakides [2007].

The work of the present paper relies on the determinant form of the modulation equations of the NLS obtained by Tovbis and Venakides [2009]. The modulation equations are transcendental equations, not differential equations, and thus the modulational instability does not hinder the analysis. Tovbis and Venakides utilized the determinant form to study the variation of the asymptotic procedure as parameters of the Riemann–Hilbert problem, in particular the spatial and the time variables that are parameters in the Riemann–Hilbert problem analysis, change. They proved [2010] that, in the case of a regular break, the nonlinear steepest descent asymptotics can be "automatically" continued through the breaking curve (however, the expressions for the asymptotic solution will be different on different sides of the curve). Although the results are stated and proven for the focusing NLS equation, they can be reformulated for AKNS systems, as well as for the nonlinear steepest descent method in a more general setting. The present paper examines the variation of the procedure with respect to the parameter $\mu$ and proves that the variation is smooth even as $\mu$ crosses the critical value $\mu = 2$.

**1A.** ***Background: n-phase waves, inverse scattering, and the Riemann–Hilbert Problem.*** In order to make the study accessible beyond the group of experts in the subject, we give an overview of our understanding of the phenomenology of the time evolution of the semiclassical NLS equation and the mathematics that represents this phenomenology.

In the ideal (and necessarily unstable) scenario, in which modulated wave profiles persist in space-time, so does the separation in two space-time scales. In the large scale, a set of boundaries (breaking or caustic curves) divides the space-time half-plane $(x, t)$, $t > 0$, into regions. Inside each region, and in the leading order as $\varepsilon \to 0$, the solution is an $n$-phase wave ($n = 0, 1, 2, 3, \dots$), with wave periods and wavelengths in the small scale. The wave parameters vary in the large scale. The increase in $n$ occurs typically as a new phase is generated at a point in space-time, due, for example, to wave-breaking (more precisely, to avert wave-breaking) or to two existing phases coming together. The newly generated oscillatory phase spreads in space with finite speed and the trace of its fronts in space-time constitutes the set of breaking curves.

An $n$-phase NLS wave is a solution of (1) which exhibits a "carrier" plane wave and $n$ nonlinearly interacting wave-phases that control its oscillating amplitude. The wave is characterized by a set of $2n+2$ real wave parameters: $n+1$ frequencies and $n+1$ wavenumbers. In the scenario discussed above, waves with periods and wavelengths of order $O(\varepsilon)$ constitute the small space-time scale. The boundaries separating phases in space-time exist in the large scale, which is of order $O(1)$. These boundaries play the role of nonlinear caustic curves. The analytic wave profile of an $n$-phase wave is given explicitly in terms of an elliptic ($n=1$) or hyperelliptic ($n>1$) Riemann theta function, derived from a compact Riemann surface of genus $n$. This is true not only for the NLS but for most of the integrable wave equations studied. The $2n+2$ branch points of the Riemann surface are the wave parameters of choice that determine the $n+1$ frequencies and $n+1$ wavenumbers. In the case of the NLS, the 0-phase wave is simply a plane wave.

The initial data (2) have the structure of a modulated 0-phase wave. As $t$, increasing from zero, reaches a value $t=t_{\text{break}}$, the $n=0$ initial phase breaks at a caustic point in space-time. As described above, a wave-phase of higher $n$ emerges then and spreads in space. As time increases, the endpoints of the spatial interval of existence of the new phase define the two caustic curves in space-time that emanate from the break point. The eventual breaking of this new phase is called the *second break*. The mechanism of the second break is fundamentally different from that of the first.

The *analytic description of n-phase waves* [Belokolos et al. 1994] is in terms of an $n$-phase Riemann theta function. This is an $n$-fold Fourier series obtained by summing $\exp\{2\pi i\,\boldsymbol{z}\cdot\boldsymbol{m}+\pi i(B\boldsymbol{m},\boldsymbol{m})\}$ over the multi-integer $\boldsymbol{m}$. $B$ is an $n\times n$ matrix with positive definite imaginary part that gives the series exponential quadratic convergence. In the case of NLS waves, the matrix $B$ arises from periods of the Riemann surface of the radical

$$R(z)=\left(\prod_{i=0}^{2n+1}(z-\alpha_i)\right)^{\frac{1}{2}},\quad \text{where } \alpha_{2j+1}=\overline{\alpha}_{2j}; \tag{5}$$

the elements of $B$ are linear combinations of the hyperelliptic integrals

$$\oint\frac{z^k}{R(z)},\quad k=0,1,\dots,n-1,$$

along appropriate closed contours on the Riemann surface of $R$ [Belokolos et al. 1994]. The series has a natural quasiperiodic structure in the $n$ complex arguments $\boldsymbol{z}=(z_1,\dots,z_n)$. Each $z_j$ is linear in $x$ and $t$ and represents a nonlinear phase of the wave; the wavenumbers and frequencies are expressed in terms of hyperelliptic integrals of the radical $R$ and are thus functions of the $\alpha_{2j}$, whose status as preferred parameters is obvious from (5).

The semiclassical limit procedure gives the emergent wave structure described above *without* any a priori ansatz of such structure. The radical $R(z)$ and the wave parameters arise naturally in the procedure. As mentioned above, these wave structures are modulated in space-time. In the large space-time scale, the branch points $\alpha_i$ vary and their number experiences a jump across the breaking curves. The branch points are calculated from the modulation equations in determinant form (they are transcendental equations, not partial differential equations, and thus there is no ill-posedness at hand). The number of branch points is obtained with the additional help of sign conditions that are obtained in the procedure.

*Overview of scattering, inverse scattering, and the Riemann–Hilbert problem (RHP).* The NLS was solved by Zakharov and Shabat [1972], who discovered a Lax pair that linearizes it. The Lax pair consists of two ordinary differential operators, one in the spatial variable $x$ and the second in the time variable $t$.

The first operator of the Lax pair is a Dirac-type operator that is not self-adjoint. The corresponding eigenvalue problem (Zakharov–Shabat) is a $2\times 2$ first-order *linear* ODE, with independent variable $x$. The NLS solution $q(x,t,\varepsilon)$ plays the role of a scatterer, entering in the off-diagonal entries of the ODE matrix. *Scattering data* are defined for those values of the spectral parameter $z$ that produce bounded (Zakharov–Shabat) solutions. This happens when $z$ is real (these solutions are called scattering states) and at the discrete set of proper eigenvalues $z_j$ (the eigenvalues come in complex-conjugate pairs; the normalized $L^2$ solutions are called bound states). The *reflection coefficient* $r(z)$, $z\in\mathbb{R}$ provides a connection between the asymptotic behaviors of the scattering states as $x\to\pm\infty$. The *norming constants* $c_j$, corresponding to the proper eigenvalues $z_j$, provide the asymptotic behavior of the bound states as $x\to+\infty$.

The second operator of the Lax pair evolves the scattering states and the bound states in time and is again a $2\times 2$ linear ODE system. The holding of the NLS equation guarantees that this evolution involves the action of a time-dependent unitary operator. As a result, the spectrum of the first Lax operator remains constant in time and the scattering data evolve in time through multiplication by simple explicit exponential propagators. The continuous spectrum contributes *radiation* to the solution of the NLS. The bound states contribute *solitons*.

Zakharov and Shabat [1972] developed the inverse scattering procedure for deriving $q(x,t)$ at any $(x,t)$ given the scattering data at $t=0$. In the modern approach initiated by Shabat [1976], the procedure is recast into a *matrix Riemann–Hilbert problem* for a $2\times 2$ matrix on the complex plane of the spectral variable $z$. One needs to determine the matrix $m(z)$ that is analytic on the closed complex plane, off an oriented contour $\Sigma$, that consists of the real axis and of small circles surrounding the eigenvalues. Modulo multiplication of its columns by normalizing factors $e^{\pm ixz/\varepsilon}$, the matrix $m(z)$ (the unknown of the problem at $t>0$) is a judiciously specified fundamental matrix solution of the eigenvalue problem of the first operator of the Lax pair (Zakharov–Shabat eigenvalue problem). In order to determine the matrix $m(z)$, one is given a jump condition on the contour $\Sigma$, and a normalization condition at $z\to\infty$,

$$m(z)=\begin{pmatrix} m_{11} & m_{12}\\ m_{21} & m_{22}\end{pmatrix}\to \text{identity, as } z\to\infty;\quad m_+(z)=m_-(z)V,\ \text{ when } z\in\Sigma. \tag{6}$$

The subscripts $\pm$ indicate limits taken from the left/right of the contour. The $2\times 2$ matrix $V=V(z,x,t,\varepsilon)$, defined on the jump contour and referred to as the *jump matrix*, is nonsingular and encodes the scattering data (see below). The space-time variables $x,t$ (and the semiclassical parameter $\varepsilon$) are parameters in the problem.

The solution to (1) is given by the simple formula

$$q(x,t,\varepsilon)=\lim_{z\to\infty} z m_{12}(z,x,t,\varepsilon). \tag{7}$$

The results and the calculations of this study are in the asymptotic limit of the semiclassical parameter $\varepsilon\to 0$.

**1B.** ***Background: the semiclassical limit $\varepsilon \to 0$.*** The Riemann–Hilbert approach is a major tool in the asymptotic analysis of integrable systems, as established with the discovery of the steepest descent method [Deift and Zhou 1993; 1995] and its extension through the $g$-function mechanism [Deift et al. 1997; Tovbis et al. 2004]. The asymptotic methods via the RHP approach also apply to orthogonal polynomial asymptotics [Deift 1999; Deift et al. 1999a] and to random matrices [Baik et al. 1999; Deift et al. 1999b; Duits and Kuijlaars 2009; Ercolani and McLaughlin 2003].

The semiclassical asymptotic analysis of the highly oscillatory RHP is similar in spirit to the steepest descent method for integrals in the complex plane. Throughout the analysis, the quantities $x$, $t$, and $\mu$ enter as parameters. The semiclassical analysis, performed with the aid of the $g$-function mechanism [Deift et al. 1997; Tovbis et al. 2004], is constructive. An undetermined function $g(z)$ is introduced in the RHP through a simple transformation of the independent matrix variable of the RHP. The contour of the RHP, itself an unknown, is partitioned (in a way to be determined) into two types of interlacing subarcs. The jump matrix is manipulated differently in the two subarc types. In one of them (main arcs), the jump matrix is factored in a certain way. In the other type (complementary arcs), it is factored differently or is left as is. The $g$-function mechanism then imposes on appropriate entries of the jump matrix factors the condition of constancy in the complex spectral variable combined with boundedness as $\varepsilon \to 0$, while imposing decay as $\varepsilon \to 0$ on other entries. The constancy conditions are equalities and the decay conditions are sign conditions that act on exponents, forcing the decay of the corresponding exponential entries. Put together, these conditions constitute a scalar RHP for the function $g(z)$ (or, as below, on its sister function $h(z)$). The contour of the RHP, its partitioning, and finally the functions $g(z)$ and $h(z)$ follow from the analysis of this scalar RHP. This procedure allows the peeling-off of the leading order solution of the original matrix RHP and leaves behind a matrix RHP for the error. This RHP is solvable with the aid of a Neumann series.

The formulae for the conditions obtained through the $g$-function mechanism have an intuitive interpretation that arises from (2D) potential theory in the complex plane of the spectral parameter $z$. The main question is to determine the equilibrium measure for an energy functional [Deift 1999; Kamvissis and Rakhmanov 2005; Simon 2011] that depends parametrically on the variables $x$ and $t$. The support of the measure depends on $x$ and $t$. For problems with self-adjoint Lax operators (e.g., the Korteweg–de Vries equation in the small dispersion limit), the support is on the real line (typically a set of intervals as in [Lax and Levermore 1983a; 1983b; 1983c; Venakides 1985]). In a general non-self-adjoint case the supports are in the complex plane. This is the case with the (focusing) NLS under study, whose spatial Lax operator, the Zakharov–Shabat system, is of Dirac type and is non-self-adjoint.

The conditions obtained through the $g$-function mechanism are exactly the variational conditions for the equilibrium measure. Deriving these conditions rigorously as such is highly taxing, especially in the non-self-adjoint case. This is not needed though. The conditions are used essentially as an ansatz in the RHP. As long as the calculation confirms the ansatz, the whole procedure is rigorous.

In the cases of the (focusing) NLS in the semiclassical limit that have been worked out so far [Boutet de Monvel et al. 2011; Buckingham and Venakides 2007; Lyng and Miller 2007; Tovbis et al. 2004; 2007], the support of the equilibrium measure is a finite union of arcs in the complex plane with complex-conjugate symmetry. Denote the endpoints of the "main arcs" (see below) by $\{\alpha_j\}_{j=0}^{N'}$, with some finite $N' \in \mathbb{N}$. The analysis leads naturally to the representation of these points as the roots of a monic polynomial, the square root of which is exactly the radical in (5). This radical, a finite-genus Riemann surface, constitutes the passage to the periodic structure of the local waveform. We refer to these endpoints henceforth as "branch points". It is necessary to establish the existence and the number of the branch points for each pair $(x,t)$ as well as the existence of the arcs, which provide the leading contribution to the expression of the local waveform.

The approach to obtaining the asymptotic solution from the initial data is to analyze the RHP for fixed $x$ and $t$, thus treating the space and time variables as parameters. The smooth dependence of the branch points $\alpha_j$ on the parameters $x$ and $t$ for the semiclassical NLS with $q(x,0) = \operatorname{sech}(x)$, $\mu = 0$ was studied in [Kamvissis et al. 2003, pp. 148–162] by considering moment conditions. A different approach, to start with local behavior near the $\alpha_j$, was applied in [Tovbis and Venakides 2009], leading to formulae of the form

$$\frac{\partial \alpha_j}{\partial x}(x,t) = -\frac{2\pi i \, \dfrac{\partial K}{\partial x}(\alpha_j, \vec{\alpha}, x, t)}{D(\vec{\alpha}) \displaystyle\oint_{\hat{\gamma}} \frac{f'(\zeta, x, t)}{(\zeta - \alpha_j(x,t)) R(\zeta, \vec{\alpha})} d\zeta}, \tag{8}$$

$$\frac{\partial \alpha_j}{\partial t}(x,t) = -\frac{2\pi i \, \dfrac{\partial K}{\partial t}(\alpha_j, \vec{\alpha}, x, t)}{D(\vec{\alpha}) \displaystyle\oint_{\hat{\gamma}} \frac{f'(\zeta, x, t)}{(\zeta - \alpha_j(x,t)) R(\zeta, \vec{\alpha})} d\zeta}, \tag{9}$$

where $\vec{\alpha} = \vec{\alpha}(x,t)$.

In addition to [Tovbis and Venakides 2009; 2010], this work is similar in spirit to [Kuijlaars and McLaughlin 2000], which pertains to random matrix theory. Those authors put a strong topology on the set of allowable potentials (a potential is analogous to the function $f$ in the present work) and demonstrate that the so-called "regular case" is generic — i.e., if you find a potential with fixed genus and all other side-conditions are satisfied in a strict sense, then the same genus holds for all potentials in a neighborhood.

In this study, we obtain the following main results.

(1) We extend formulae (8) and (9) to include the dependence of the branch points and of the contour $\hat{\gamma} = \hat{\gamma}(\mu)$ on the external parameter $\mu$,

$$\frac{\partial \alpha_j}{\partial \mu}(x,t,\mu) = -\frac{2\pi i \, \dfrac{\partial K}{\partial \mu}(\alpha_j, \vec{\alpha}, x, t, \mu)}{D(\vec{\alpha}) \displaystyle\oint_{\hat{\gamma}(\mu)} \frac{f'(\zeta, x, t, \mu)}{(\zeta - \alpha_j(x,t,\mu)) R(\zeta, \vec{\alpha})} d\zeta}, \tag{10}$$

where $\vec{\alpha} = \vec{\alpha}(x,t,\mu)$. We show that the dependence is smooth, meaning that the contour, the jump matrix, and the solution of the scalar RHP evolve smoothly (they are continuously differentiable) in $\mu$.

(2) We simplify the expression for $\dfrac{\partial K}{\partial \mu}$ in (43).

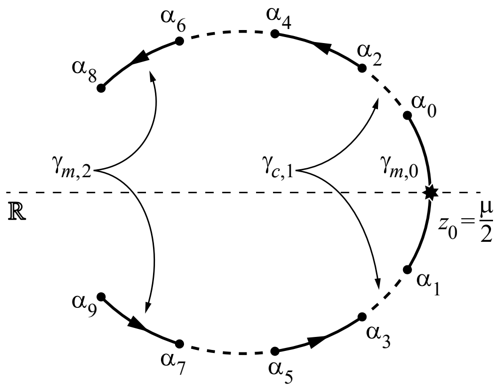



**Figure 1.** The RHP jump contour in the case of genus 4 with complex-conjugate symmetry in the notation of [Tovbis et al. 2004].

(3) We show good agreement of formula (10) with the dependence of the branch points on the parameter $\mu$, obtained by the direct numerical solution of the system (19) (see Figure 3).

(4) We prove the preservation of genus of the asymptotic solution in an open interval of parameter $\mu$. In particular, the genus is preserved (0 or 2) for all $x$ and $t > 0$ for some open interval (which depends on $x$ and $t$) for $\mu < 2$.

This paper is organized as follows. Section 2 contains definitions and prior results. Section 3 discusses the analyticity of $f$ in $\mu$ and the differentiability of the branch points $\alpha_j = \alpha_j(\mu)$. Section 4 studies the $\mu$-dependence of the quantities that appear in Theorem 4.5. Section 5 is devoted to sign conditions and the preservation of genus (Theorem 5.8). Section 6 provides numerical support for the main result of the paper. The Appendix supplies explicit formulations of all relevant expressions and the main result in the genus 0 and genus 2 cases.

## 2. Preliminaries

We consider a model scalar Riemann–Hilbert problem which arises in the process of the asymptotic solution of the semiclassical focusing NLS (1) with the initial condition (3). The input to the problem is a given function $f(z)$ that derives originally from the asymptotic limit of the scattering data for this initial value problem. The function $f(z)$ (see (21)) depends parametrically on the space and time variables, $x$ and $t$. It also depends on the real parameter $\mu$ in the initial data of the NLS (3). The following properties of the function $f(z)$ are crucial for our calculations.

(1) $f(z)$ is analytic at all points of $\mathbb{C} \setminus \mathbb{R}$ except for branch cuts.

(2) $f(z)$ has a $z \ln z$ singularity at the point $z = \mu/2$.

(3) $f(z)$ is Schwarz-symmetric.

Other functions that satisfy these conditions are a priori admissible as inputs to our model problem; whether they too lead to solutions is a matter of calculation.

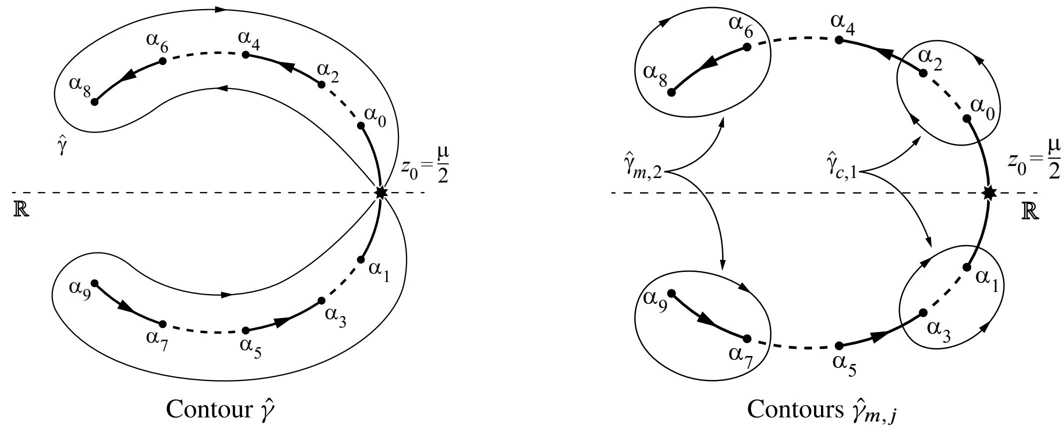



**Figure 2.** Contours of integration for the function $h(z)$ of (14). The point $z_0 = \mu/2$ is a point of nonanalyticity of $f(z)$ on $\gamma$.

The unknown of the problem is a function $h(z)$ that has the following properties.

(1) The function $h(z)$ is Schwarz-reflexive.

(2) The function $h(z)$ compensates for the points of nonanalyticity of $f(z)$ in the sense that $h + f$ is analytic at these points with only one exception, the point $z = \mu/2$.

(3) The function $h + f$ is analytic in $\overline{\mathbb{C}} \setminus \gamma$, where $\gamma$ is a contour to be determined that passes through point $z = \mu/2$ and is symmetric with respect to the real axis.

(4) The function $h(z)$ exhibits constant (independent of $z$) jumps across subarcs of the contour $\gamma$. This is made more specific below.

**Remark.** A cleaner formulation of the RHP could be achieved with the unknown $f + h$ instead of $h$. Indeed, the function

$$g(z) = \tfrac{1}{2}(h(z) + f(z))$$

jumps only across the contour $\gamma$, providing a cleaner RHP formulation. Yet the results are in terms of the function $h(z)$; hence our choice in its favor.

**Remark.** Based on the previous remark, we still refer to the contour $\gamma$ as the contour of the RHP. Since the contour itself is one of the unknowns, we refer to the problem as a "free contour RHP", in analogy to the well-known "free boundary value problems".

We now set the precise conditions for the contour $\gamma$ and the jumps across it.

(1) The contour $\gamma$ is a finite-length, non-self-intersecting arc that is symmetric with respect to the real axis. It intersects the real axis at a point $z = \mu/2$ (we refer to this point as $z_0$) to be discussed below.

(2) The contour $\gamma$ is oriented from its endpoint in the lower complex half-plane to its endpoint in the upper half-plane.

(3) For some integer $N$, we consider $2N+1$ distinct points of the contour in the upper half-plane, including the contour endpoint; we also consider their complex conjugates in the lower half-plane. We label the points in the upper half-plane with even indices $\{\alpha_{2i}\}_{i=0}^{2N}$ that increase in the direction of orientation of the contour and we label the points in the lower half-plane with odd indices $\{\alpha_{2i+1}\}_{i=0}^{2N}$ that decrease in the direction of orientation. Clearly, the sequence of points in the direction of orientation are

$$\underbrace{\alpha_{4N+1},\alpha_{4N-1},\alpha_{4N-3},\dots,\alpha_3,\alpha_1}_{\text{lower half-plane}},\ \underbrace{\alpha_0,\alpha_2,\dots,\alpha_{4N-4},\alpha_{4N-2},\alpha_{4N}}_{\text{upper half-plane}},$$

$\alpha_{4N+1}$ and $\alpha_{4N}$ are the endpoints of the contour $\gamma$, and

$$\alpha_{2i+1}=\overline{\alpha_{2i}},\quad i=0,1,2,\dots,2N. \tag{11}$$

The jumps of the RHP are defined on the arcs into which the contour is partitioned by these points. Two alternative types of RHP jumps are imposed; each arc is labeled as a main arc or a complementary arc, respectively. The two arc types interlace along the contour and the contour has main arcs at both ends. All arcs inherit their orientation from the contour $\gamma$.

It is trivial to check that arcs which are complex-conjugate to each other are either both main or both complementary. It is convenient to lump such an arc pair into one entity; in the following definitions, we retain the terms *main arc* and *complementary arc* for such arc pairs, by abuse of vocabulary.

(1) We define as main arcs $\gamma_{m,j}$, where $j=0,1,\dots,N$:

$$\gamma_{m,0}=[\alpha_1,\alpha_0],\quad \gamma_{m,j}=[\alpha_{4j-2},\alpha_{4j}]\cup[\alpha_{4j+1},\alpha_{4j-1}],\quad j=1,\dots,N.$$

Thus, a main arc consists of a single arc when $j=0$ and a union of two arcs when $j>0$.

(2) We define as complementary arcs $\gamma_{c,j}$, where $j=1,\dots,N$:

$$\gamma_{c,j}=[\alpha_{4j-4},\alpha_{4j-2}]\cup[\alpha_{4j-1},\alpha_{4j-3}],\quad j=1,\dots,N.$$

The jump conditions of the RHP are given by

$$\begin{cases} h_+(z)+h_-(z)=2W_j & \text{on } \gamma_{m,j},\ j=0,1,\dots,N,\\ h_+(z)-h_-(z)=2\Omega_j & \text{on } \gamma_{c,j},\ j=1,\dots,N,\\ h(z)+f(z) \text{ is analytic in } \overline{\mathbb{C}}\setminus\gamma, \end{cases} \tag{12}$$

where $W_j$ and $\Omega_j$ are real constants with a normalization $W_0=0$.

We end our formulation of the free contour RHP for the function $h(z)$ by reiterating what the knowns and what the unknowns of the problem are. The contour $\gamma$ is unknown, except for the requirement of passing through the singular point $z_0=\mu/2$. The positions of the $4N+2$ partitioning points are unknown. As we have formulated the problem so far, the value of the integer $N$ is free. This freedom is lifted if important additional conditions ("sign conditions") are imposed on the RHP, as occurs in the case of NLS [Tovbis et al. 2004]. The sign conditions guarantee the decay of certain jump matrix entries. In the presence of these conditions, the number of points turns into an important unknown; the $n$-phase wave represented has $n = 2N$. Finally, the real constants in the jump conditions are unknown. To summarize, the only known data is the function $f(z; \mu, \beta)$.

Our main concern is the dependence of the solution of the problem on the parameter $\mu$. Any other parameters (space and time if the RHP arises from the focusing NLS) are collectively labeled $\beta$. A multiparameter family of such functions $f$ will be discussed below. The specific function $f$ that corresponds to the focusing NLS equation with initial data (3) is given in the beginning of Section 3.

**Definition 2.1.** Let

$$\vec{\alpha} = \{\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_{4N+1}\}. \tag{13}$$

We say

$$\gamma \in \Gamma(\vec{\alpha}, \mu)$$

if a contour $\gamma$ satisfies all the conditions set above and shown in Figure 1.

Note that for fixed $\vec{\alpha}$ and $\mu$, the contour $\gamma$, aside from passing through $z = \alpha_j$ and $z = \mu/2$, is free to deform continuously within a domain of analyticity of $f$. Thus for a fixed $f$, the notation $\gamma = \gamma(\vec{\alpha}, \mu)$ may indicate the general element of the set $\Gamma(\vec{\alpha}, \mu)$. The following lemma is an immediate consequence of our definitions.

**Lemma 2.2.** *Consider* $\gamma_0 = \gamma_0(\vec{\alpha}_0, \mu_0) \in \Gamma(\vec{\alpha}_0, \mu_0)$.

*There exist open neighborhoods of* $\vec{\alpha}_0$ *and* $\mu_0$ *such that for all* $\vec{\alpha}$ *in the neighborhood of* $\vec{\alpha}_0$*, for all* $\mu$ *in the neighborhood of* $\mu_0$ *there is a contour* $\gamma(\vec{\alpha}, \mu) \in \Gamma(\vec{\alpha}, \mu)$.

**Definition 2.3.** We say that

$$\hat{\gamma} \in \hat{\Gamma}(\gamma, \vec{\alpha}, \mu)$$

if $\hat{\gamma}$ is a non-self-intersecting closed contour around $\gamma \in \Gamma(\vec{\alpha}, \mu)$ within the domain of analyticity of $f$ except at $z_0 = \mu/2$, with complex-conjugate symmetry $\overline{\hat{\gamma}} = \hat{\gamma}$.

We define $\hat{\gamma}_{m,j}$ and $\hat{\gamma}_{c,j}$ similarly.

**Remark 2.4.** By considering the loop contours $\hat{\gamma}$, $\hat{\gamma}_{m,j}$, $\hat{\gamma}_{c,j}$, the explicit dependence of the contours on the end points $\vec{\alpha}$ is removed (for example in (32)–(35)). So even though $\gamma = \gamma(\vec{\alpha}, \mu)$, in all our evaluations below $\hat{\gamma} = \hat{\gamma}(\mu)$.

**Remark 2.5.** Lemma 2.2 implies that if $\hat{\gamma}_0 \in \hat{\Gamma}(\gamma_0, \vec{\alpha}_0, \mu_0)$, then there is a contour $\gamma \in \Gamma(\vec{\alpha}, \mu)$ such that $\hat{\gamma} \in \hat{\Gamma}(\gamma, \vec{\alpha}, \mu)$ for all $\vec{\alpha}$ and $\mu$ in some open neighborhoods of $\vec{\alpha}_0$ and $\mu_0$.

**Definition 2.6.** We denote the RHP (12) as

$$\mathrm{RHP}(\gamma, \vec{\alpha}, \mu, f),$$

where $\gamma \in \Gamma(\vec{\alpha}, \mu)$.

The solution of the RHP (12), $h(z)$, can be found explicitly [Tovbis et al. 2004]:

$$h(z)=\frac{R(z)}{2\pi i}\left[\oint_{\hat\gamma}\frac{f(\zeta)}{(\zeta-z)R(\zeta)}\,d\zeta+\sum_{j=1}^{N}\oint_{\hat\gamma_{m,j}}\frac{W_j}{(\zeta-z)R(\zeta)}\,d\zeta+\sum_{j=1}^{N}\oint_{\hat\gamma_{c,j}}\frac{\Omega_j}{(\zeta-z)R(\zeta)}\,d\zeta\right],\tag{14}$$

or, in determinant form [Tovbis and Venakides 2009],

$$h(z)=\frac{R(z)}{D}K(z),\tag{15}$$

where $z$ lies inside of $\hat\gamma$ and outside all $\hat\gamma_{c,j}$ and $\hat\gamma_{m,j}$, and where

$$K(z)=\frac{1}{2\pi i}\begin{vmatrix}\oint_{\hat\gamma_{m,1}}\frac{d\zeta}{R(\zeta)} & \cdots & \oint_{\hat\gamma_{m,1}}\frac{\zeta^{N-1}d\zeta}{R(\zeta)} & \oint_{\hat\gamma_{m,1}}\frac{d\zeta}{(\zeta-z)R(\zeta)}\\ \vdots & \ddots & \vdots & \vdots\\ \oint_{\hat\gamma_{m,N}}\frac{d\zeta}{R(\zeta)} & \cdots & \oint_{\hat\gamma_{m,N}}\frac{\zeta^{N-1}d\zeta}{R(\zeta)} & \oint_{\hat\gamma_{m,N}}\frac{d\zeta}{(\zeta-z)R(\zeta)}\\ \oint_{\hat\gamma_{c,1}}\frac{d\zeta}{R(\zeta)} & \cdots & \oint_{\hat\gamma_{c,1}}\frac{\zeta^{N-1}d\zeta}{R(\zeta)} & \oint_{\hat\gamma_{c,1}}\frac{d\zeta}{(\zeta-z)R(\zeta)}\\ \vdots & \ddots & \vdots & \vdots\\ \oint_{\hat\gamma_{c,N}}\frac{d\zeta}{R(\zeta)} & \cdots & \oint_{\hat\gamma_{c,N}}\frac{\zeta^{N-1}d\zeta}{R(\zeta)} & \oint_{\hat\gamma_{c,N}}\frac{d\zeta}{(\zeta-z)R(\zeta)}\\ \oint_{\hat\gamma}\frac{f(\zeta)\,d\zeta}{R(\zeta)} & \cdots & \oint_{\hat\gamma}\frac{\zeta^{N-1}f(\zeta)\,d\zeta}{R(\zeta)} & \oint_{\hat\gamma}\frac{f(\zeta)\,d\zeta}{(\zeta-z)R(\zeta)}\end{vmatrix}\tag{16}$$

and

$$D=\det(A),\tag{17}$$

with

$$A=\begin{pmatrix}\oint_{\hat\gamma_{m,1}}\frac{d\zeta}{R(\zeta)} & \cdots & \oint_{\hat\gamma_{m,1}}\frac{\zeta^{N-1}d\zeta}{R(\zeta)}\\ \vdots & \ddots & \vdots\\ \oint_{\hat\gamma_{m,N}}\frac{d\zeta}{R(\zeta)} & \cdots & \oint_{\hat\gamma_{m,N}}\frac{\zeta^{N-1}d\zeta}{R(\zeta)}\\ \oint_{\hat\gamma_{c,1}}\frac{d\zeta}{R(\zeta)} & \cdots & \oint_{\hat\gamma_{c,1}}\frac{\zeta^{N-1}d\zeta}{R(\zeta)}\\ \vdots & \ddots & \vdots\\ \oint_{\hat\gamma_{c,N}}\frac{d\zeta}{R(\zeta)} & \cdots & \oint_{\hat\gamma_{c,N}}\frac{\zeta^{N-1}d\zeta}{R(\zeta)}\end{pmatrix}.\tag{18}$$

The arc end points $\{\alpha_j\}$ satisfy the system

$$K(\alpha_j)=0,\quad j=0,1,\dots,4N+1.\tag{19}$$

The dependence on $x$ and $t$ was considered in [Tovbis and Venakides 2009]. This is a simpler situation when the jump contour $\gamma$ in the RHP (12) is independent of the parameters.

The main related results in [Tovbis and Venakides 2009] are the determinant form (15) and:

**Theorem 2.7.** *Let $f(z,\vec{\beta})$, where $\vec{\beta}\in B\subset\mathbb{R}^m$. For all $\vec{\beta}\in B$ assume $f(z,\vec{\beta})$ is analytic on $S\in\mathbb{C}$. Moreover, $\gamma\setminus S$ consists of no more than finitely many points and $f$ is continuous on $\gamma$. The modulation equations* (19) *imply the system of* $4N+2$ *differential equations*

$$(\alpha_j)_{\beta_k}=-\frac{2\pi i\frac{\partial}{\partial\beta_k}K(\alpha_j)}{D\oint_{\hat{\gamma}}\frac{f'(\zeta)}{(\zeta-\alpha_j)R(\zeta)}d\zeta}. \tag{20}$$

In particular, one gets (8) and (9) for parameters $x$ and $t$. Note that the contour $\gamma$ is assumed independent of parameters $x$ and $t$ explicitly. The dependence on these parameters comes in through the branch points $\vec{\alpha}=\vec{\alpha}(x,t)$.

The main related result in [Tovbis and Venakides 2010] is:

**Theorem 2.8.** *Let the nonlinear steepest descent asymptotics for solution $q(x,t,\varepsilon)$ of the NLS* (1) *be valid at some point $(x_0,t_0)$. If $(x_*,t_*)$ is an arbitrary point, connected with $(x_0,t_0)$ by a piecewise-smooth path $\Sigma$, if the contour $\gamma(x,t)$ of the RHP* (12) *does not interact with singularities of $f(z)$ as $(x,t)$ varies from $(x_0,t_0)$ to $(x_*,t_*)$ along $\Sigma$, and if all the branch points are bounded and stay away from the real axis, then the nonlinear steepest descent asymptotics* (*with the proper choice of genus*) *are also valid at $(x_*,t_*)$.*

We extend Theorem 2.7 and make partial progress in the direction of Theorem 2.8 in the case when the jump contour explicitly depends on the parameter $\mu$. We require that the point of logarithmic singularity of $f=f(z,\mu)$, $z_0=\mu/2$ be always on $\gamma$. Additionally, we prove preservation of genus for all $x>0$, $t>0$, $\mu>0$, under certain conditions which guarantee that the parameters are away from asymptotic solution breaks (see Theorem 5.8). In particular, the genus is preserved in a neighborhood of the special value of the parameter $\mu=2$. Thus we obtain that for all $x>0$, $t>0$ (except on the first breaking curve), there is a small neighborhood such that for all $\mu<2$ in the neighborhood, the genus is the same as for $\mu=2$, where it is known to be 0 or 2.

## 3. $\mu$-dependence in the semiclassical focusing NLS

To apply the methods from [Tovbis and Venakides 2009] we need analyticity of $f(z,\mu)$ in the parameter $\mu$.

The function $f(z)$, obtained in [Tovbis et al. 2004] from a semiclassical approximation of the exactly derived scattering data for the NLS with initial condition (3) [Tovbis and Venakides 2000], is given by

$$f(z,\mu,x,t)=\left(\frac{\mu}{2}-z\right)\left[\frac{\pi i}{2}+\ln\left(\frac{\mu}{2}-z\right)\right]+\frac{z+T}{2}\ln(z+T)+\frac{z-T}{2}\ln(z-T)$$
$$-T\tanh^{-1}\frac{T}{\mu/2}-xz-2tz^2+\frac{\mu}{2}\ln 2\quad\text{when } \Im z\geq 0, \tag{21}$$

and

$$f(z)=\overline{f(\overline{z})}\quad\text{when } \Im z<0, \tag{22}$$

where the branch cuts are chosen as follows: for $0<\mu<2$, the logarithmic branch cuts are from $z=\mu/2$ along the real axis to $+\infty$, from $z=T$ to 0 and along the real axis to $+\infty$, and from $z=-T$ to 0 and along the real axis to $-\infty$; for $\mu\geq 2$, the branch cuts are from $z=T$ to $+\infty$ and from $z=-T$ to $-\infty$ along the real axis, where

$$T=T(\mu)=\sqrt{\frac{\mu^2}{4}-1},\quad \Im T\geq 0. \tag{23}$$

For $\mu\geq 2$, $T\geq 0$ is real and for $0<\mu<2$, $T$ is purely imaginary with $\Im T>0$.

**Lemma 3.1.** *$f(z,\mu)$ and $f'(z,\mu)$ are analytic in $\mu$ for $\mu>0$, $x>0$, $t>0$, for all $z$, $\Im z\neq 0$, $z\notin[-T,T]$.*

*Proof.* Consider

$$f'(z,\mu)=-\frac{\pi i}{2}-\ln\left(\frac{\mu}{2}-z\right)+\frac{1}{2}\ln\left(z^2-\frac{\mu^2}{4}+1\right)-x-4tz, \tag{24}$$

which is analytic in $\mu>0$, for $\Im z\neq 0$, $z\notin[-T,T]$.

For $\mu>0$, $\mu\neq 2$, $f(z,\mu)$ is clearly analytic in $\mu$ for $\Im z\neq 0$. At $\mu=2$ ($T=0$) we find the power series of $f(z,\mu)$ in $T$ and show that it contains only even powers. Since

$$T^{2k}=\left(\frac{\mu^2}{4}-1\right)^k=\frac{(\mu+2)^k}{4^k}(\mu-2)^k, \tag{25}$$

it will show analyticity of $f(z,\mu)$ in $\mu$.

Start with expanding basic terms in series at $T=0$:

$$\frac{1}{\mu/2}=\sqrt{1+T^2}^{\,-1}=\sum_{k=0}^{\infty}c_kT^{2k},\quad \ln(z\pm T)=z\ln z\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n}\left(\pm\frac{T}{z}\right)^n. \tag{26}$$

Then the logarithmic terms in (21) become

$$\begin{aligned}\frac{z+T}{2}\ln(z+T)+\frac{z-T}{2}\ln(z-T)&=z\ln z-z\sum_{n\text{ is even}}\frac{1}{n}\left(\frac{T}{z}\right)^n+T\sum_{n\text{ is odd}}\frac{1}{n}\left(\frac{T}{z}\right)^n\\&=z\ln z+\sum_{k=1}^{\infty}\frac{1}{2k(2k-1)z^{2k-1}}T^{2k},\end{aligned} \tag{27}$$

which has only even powers of $T$ and is analytic in $\mu$ for $\Im z\neq 0$. Next we consider the inverse hyperbolic tangent term in (21), and taking into account that $\tanh^{-1}z$ is an odd function,

$$\begin{aligned}T\tanh^{-1}\frac{T}{\mu/2}&=T\tanh^{-1}\frac{T}{\sqrt{1+T^2}}=T\tanh^{-1}\sum_{k=0}^{\infty}c_kT^{2k+1}\\&=T\sum_{k=0}^{\infty}\tilde{c}_kT^{2k+1}=\sum_{k=0}^{\infty}\tilde{c}_kT^{2k+2},\end{aligned} \tag{28}$$

which also has only even powers of $T$.

So $f(z,x,t,\mu)$ is analytic in $\mu$ for $\mu>0$, $x>0$, $t>0$, $\Im z\neq 0$, $z\notin[-T(\mu),T(\mu)]$. □

Then, for $\Im z > 0$,

$$\frac{\partial f}{\partial \mu}(z,\mu) = \frac{\pi i}{4} + \frac{1}{2}\ln\left(\frac{\mu}{2} - z\right) + \ln 2 + \frac{\mu}{8T}\left[\ln(z+T) - \ln(z-T) - 2\tanh^{-1}\frac{2T}{\mu}\right], \tag{29}$$

where $\tanh^{-1} x = x + O(x^3)$, as $x \to 0$; then

$$\frac{\partial f}{\partial \mu}(z,\mu) = \frac{\pi i}{4} + \frac{1}{2}\ln\left(\frac{\mu}{2} - z\right) + \ln 2 + \frac{\mu}{4z} - \frac{1}{2} + O(T),\ T \to 0. \tag{30}$$

So $\mu = 2$ is a removable singularity for $f_\mu(z,\mu)$ and

$$\lim_{\substack{\mu \to 2 \\ T \to 0}} \frac{\partial f}{\partial \mu}(z,\mu) = \frac{\pi i}{4} + \frac{1}{2}\ln(1-z) + \ln 2 + \frac{1}{2z} - \frac{1}{2}, \tag{31}$$

which is analytic in $z$ for $\Im z \neq 0$.

**Remark 3.2.** The jump of $f(z)$ is caused by the Schwarz reflection (22) on the real axis and is linear in $z$ since $\Im f$ is a linear function on the real axis (as a limit) near $\mu/2$ with $\Im f(\mu/2) = 0$ [Tovbis et al. 2004].

## 4. Parametric dependence of the scalar RHP

The main difficulty is the dependence of $f(z)$ (thus the RHP (12)) and the modulation equations (19) on parameter $\mu$, which also controls the logarithmic branch point $z = \mu/2$ on the contour $\hat{\gamma}$. We show that the dependence on $\mu$ is smooth.

To solve $\vec{K}(\vec{\alpha},\mu) = \vec{0}$, we need nondegeneracy of the system and apply the implicit function theorem. The following technical lemma simplifies expressions for partial derivatives in $\mu$ of (14) and (16).

**Lemma 4.1.** *Let the function $f$ be given by* (21), *and consider a contour $\gamma_0 = \gamma(\vec{\alpha},\mu_0) \in \Gamma(\vec{\alpha},\mu_0)$ having fixed arc end points $\vec{\alpha}$. There is an open neighborhood of $\mu_0$ such that for all $\mu$ in the neighborhood of $\mu_0$, there is $\hat{\gamma}(\mu) \in \hat{\Gamma}(\gamma,\vec{\alpha},\mu)$, and for all $j = 0,1,\ldots,4N+1$, $n \in \mathbb{N}$,*

$$\frac{\partial}{\partial \mu}\oint_{\hat{\gamma}(\mu)} \frac{\zeta^n f(\zeta,\mu)\,d\zeta}{R(\zeta,\vec{\alpha})} = \oint_{\hat{\gamma}(\mu)} \frac{\zeta^n \frac{\partial f(\zeta,\mu)}{\partial \mu}\,d\zeta}{R(\zeta,\vec{\alpha})}, \tag{32}$$

$$\frac{\partial}{\partial \mu}\oint_{\hat{\gamma}(\mu)} \frac{f(\zeta,\mu)\,d\zeta}{(\zeta-\alpha_j)R(\zeta,\vec{\alpha})} = \oint_{\hat{\gamma}(\mu)} \frac{\frac{\partial f(\zeta,\mu)}{\partial \mu}\,d\zeta}{(\zeta-\alpha_j)R(\zeta,\vec{\alpha})}, \tag{33}$$

$$\frac{\partial}{\partial \mu}\oint_{\hat{\gamma}_{m,k}} \frac{\zeta^n\,d\zeta}{R(\zeta,\vec{\alpha})} = 0, \quad \frac{\partial}{\partial \mu}\oint_{\hat{\gamma}_{m,k}} \frac{d\zeta}{(\zeta-\alpha_j)R(\zeta,\vec{\alpha})} = 0,\ k = 1,2,\ldots,N, \tag{34}$$

$$\frac{\partial}{\partial \mu}\oint_{\hat{\gamma}_{c,k}} \frac{\zeta^n\,d\zeta}{R(\zeta,\vec{\alpha})} = 0, \quad \frac{\partial}{\partial \mu}\oint_{\hat{\gamma}_{c,k}} \frac{d\zeta}{(\zeta-\alpha_j)R(\zeta,\vec{\alpha})} = 0,\ k = 1,2,\ldots,N. \tag{35}$$

*Proof.* The idea of the proof is to consider finite differences and take the limit as $\Delta\mu \to 0$. The complication is that both the integrands and the contours of integration depend on $\mu$.

Denote the integral on the left in (32) by $I_1$:

$$I_1(\mu) = \oint_{\hat{\gamma}(\mu)} \frac{\zeta^n f(\zeta,\mu)}{R(\zeta,\vec{\alpha})}\, d\zeta, \tag{36}$$

where $\hat{\gamma}(\mu) \in \hat{\Gamma}(\gamma,\vec{\alpha},\mu)$. Consider

$$\frac{I_1(\mu+\Delta\mu) - I_1(\mu)}{\Delta\mu}, \tag{37}$$

with small real $\Delta\mu \neq 0$. There are two logarithmic branch cuts near the contours of integration, in $f(z,\mu)$ and in $f(z,\mu+\Delta\mu)$, with both branch cuts chosen from $z_0(\mu) = \mu/2$ and $z_0(\mu+\Delta\mu)$ horizontally to the right along the real axis. Additionally, these functions have a jump on the real axis for $z < \mu/2$ from Schwarz symmetry.

We choose some fixed points $\delta_1$ and $\delta_2$ to be real, satisfying

$$\delta_1 < \frac{\mu}{2} - \frac{|\Delta\mu|}{2} < \frac{\mu}{2} + \frac{|\Delta\mu|}{2} < \delta_2. \tag{38}$$

Both contours of integration $\hat{\gamma}(\mu)$, $\hat{\gamma}(\mu+\Delta\mu)$ are pushed to the real axis near $z_0$ and split into

$$[\delta_1,\delta_2] := [\delta_1 + i0, \delta_2 + i0] \cup [\delta_2 - i0, \delta_1 - i0] \tag{39}$$

and its complement. On the complement, we can also deform both contours to coincide. So $\hat{\gamma}(\mu+\Delta\mu) = \hat{\gamma}(\mu)$.

Across $[\delta_1,\delta_2]$, $f(z,\mu)$ has a jump $\pi i|z_0(\mu) - z|$ and $f(z,\mu+\Delta\mu)$ has a jump $\pi i|z_0(\mu+\Delta\mu) - z|$. So contributions near $z_0$ in both cases are small.

Then

$$\frac{I_1(\mu+\Delta\mu) - I_1(\mu)}{\Delta\mu} = \frac{1}{\Delta\mu}\left(\oint_{\hat{\gamma}(\mu+\Delta\mu)} \frac{\zeta^n f(\zeta,\mu+\Delta\mu)}{R(\zeta,\vec{\alpha})}\, d\zeta - \oint_{\hat{\gamma}(\mu)} \frac{\zeta^n f(\zeta,\mu)}{R(\zeta,\vec{\alpha})}\, d\zeta\right); \tag{40}$$

we add and subtract $\oint_{\hat{\gamma}(\mu)} \frac{\zeta^n f(\zeta,\mu+\Delta\mu)}{R(\zeta,\vec{\alpha})}\, d\zeta$, giving

$$(40) = \frac{1}{\Delta\mu}\left(\oint_{\hat{\gamma}(\mu+\Delta\mu)-\hat{\gamma}(\mu)} \frac{\zeta^n f(\zeta,\mu+\Delta\mu)}{R(\zeta,\vec{\alpha})}\, d\zeta + \oint_{\hat{\gamma}(\mu)} \frac{\zeta^n (f(\zeta,\mu+\Delta\mu) - f(\zeta,\mu))}{R(\zeta,\vec{\alpha})}\, d\zeta\right). \tag{41}$$

The first integral is 0, because $\hat{\gamma}(\mu+\Delta\mu) = \hat{\gamma}(\mu)$.

Thus

$$\frac{I_1(\mu+\Delta\mu) - I_1(\mu)}{\Delta\mu} = \oint_{\hat{\gamma}(\mu)} \frac{\zeta^n \frac{f(\zeta,\mu+\Delta\mu) - f(\zeta,\mu)}{\Delta\mu}}{R(\zeta,\vec{\alpha})}\, d\zeta. \tag{42}$$

The last step is to take the limit as $\Delta\mu \to 0$ and to interchange it with the integral. The contour of integration is split into two: a small neighborhood near $z_0$ and its complement. For the integral near $z_0$, by a direct computation it can be shown that the limit can be passed under the integral. The integral over the second part of the contour has the integrand uniformly bounded in $\mu$, since $\log(\zeta-\mu/2)$ in $\partial f/\partial\mu$ is uniformly bounded away from $\mu/2$, so the limit and the integral can be interchanged. This completes the proof for the first integral (32).

The second integral (33) is done similarly. The rest of the integrals (34), (35) are independent of $\mu$ since the only dependence on $\mu$ sits in $z_0(\mu)\in\gamma_{m,0}$. ☐

Using Lemma 4.1,

$$\frac{\partial K}{\partial\mu}(\alpha_j,\vec\alpha,\mu)=\frac{1}{2\pi i}\begin{vmatrix} \oint_{\hat\gamma_{m,1}}\frac{d\zeta}{R(\zeta)} & \cdots & \oint_{\hat\gamma_{m,1}}\frac{\zeta^{N-1}d\zeta}{R(\zeta)} & \oint_{\hat\gamma_{m,1}}\frac{d\zeta}{(\zeta-\alpha_j)R(\zeta)}\\ \vdots & \ddots & \vdots & \vdots\\ \oint_{\hat\gamma_{m,N}}\frac{d\zeta}{R(\zeta)} & \cdots & \oint_{\hat\gamma_{m,N}}\frac{\zeta^{N-1}d\zeta}{R(\zeta)} & \oint_{\hat\gamma_{m,N}}\frac{d\zeta}{(\zeta-\alpha_j)R(\zeta)}\\ \oint_{\hat\gamma_{c,1}}\frac{d\zeta}{R(\zeta)} & \cdots & \oint_{\hat\gamma_{c,1}}\frac{\zeta^{N-1}d\zeta}{R(\zeta)} & \oint_{\hat\gamma_{c,1}}\frac{d\zeta}{(\zeta-\alpha_j)R(\zeta)}\\ \vdots & \ddots & \vdots & \vdots\\ \oint_{\hat\gamma_{c,N}}\frac{d\zeta}{R(\zeta)} & \cdots & \oint_{\hat\gamma_{c,N}}\frac{\zeta^{N-1}d\zeta}{R(\zeta)} & \oint_{\hat\gamma_{c,N}}\frac{d\zeta}{(\zeta-\alpha_j)R(\zeta)}\\ \oint_{\hat\gamma}\frac{f_\mu(\zeta)\,d\zeta}{R(\zeta)} & \cdots & \oint_{\hat\gamma}\frac{\zeta^{N-1}f_\mu(\zeta)\,d\zeta}{R(\zeta)} & \oint_{\hat\gamma}\frac{f_\mu(\zeta)\,d\zeta}{(\zeta-\alpha_j)R(\zeta)} \end{vmatrix}, \tag{43}$$

where $f_\mu$ is given by (29).

**Lemma 4.2.** *Let $f$ be given by* (21) *and consider a contour $\gamma_0=\gamma(\vec\alpha_0,\mu_0)\in\Gamma(\vec\alpha_0,\mu_0)$ having arc end points $\vec\alpha_0$. Then there are open neighborhoods of $\vec\alpha_0$ and $\mu_0$ such that for all $\vec\alpha$ and $\mu$ in the neighborhoods of $\vec\alpha_0$ and $\mu_0$, respectively, there is a contour $\gamma=\gamma(\vec\alpha,\mu)\in\Gamma(\vec\alpha,\mu)$ and*

$$K_j(\vec\alpha,\mu):=K(\alpha_j,\vec\alpha,\mu),\quad j=0,1,\ldots,4N+1, \tag{44}$$

*is continuously differentiable in $\vec\alpha$ and in $\mu$.*

*Proof.* Since $\gamma_0\in\Gamma(\vec\alpha_0,\mu_0)$ by Lemma 2.2, there are neighborhoods of $\vec\alpha_0$ and $\mu_0$ such that for all $\vec\alpha$ and $\mu$ in the neighborhoods of $\vec\alpha_0$ and $\mu_0$, respectively, there is a contour $\gamma=\gamma(\vec\alpha,\mu)\in\Gamma(\vec\alpha,\mu)$.

$K_j(\vec\alpha,\mu)$ is analytic in $\vec\alpha$ by the determinant structure and the integral entries (16), where explicit dependence on $\vec\alpha$ is only in the $R(z,\vec\alpha)$ term, which is analytic away from $z=\alpha_j$, $j=0,\ldots,4N+1$.

The integrals in the last row of the matrix in (43) involve the function $f_\mu$ given by (29), which is integrable near $z=\mu/2$, and hence the whole determinant is continuous in $\mu$. Thus $K_j(\alpha_j,\vec\alpha,\mu)$ is continuously differentiable in $\vec\alpha$ and in $\mu$. ☐

By Lemma 4.2, the modulation equations (19)

$$K_j(\vec\alpha,\mu)=K(\alpha_j,\vec\alpha,\mu)=0 \tag{45}$$

are smooth in $\vec\alpha$ and in the parameter $\mu$. Next we want to solve this system for $\vec\alpha=\vec\alpha(\mu)$ and derive smoothness in $\mu$ by the implicit function theorem.

For the next lemma we need $K'(z,\vec{\alpha},\mu) = \dfrac{dK}{dz}(z,\vec{\alpha},\mu)$:

$$K'(z,\vec{\alpha},\mu) = \frac{1}{2\pi i}\begin{vmatrix} \oint_{\hat{\gamma}_{m,1}} \frac{d\zeta}{R(\zeta)} & \cdots & \oint_{\hat{\gamma}_{m,1}} \frac{\zeta^{N-1} d\zeta}{R(\zeta)} & \oint_{\hat{\gamma}_{m,1}} \frac{d\zeta}{(\zeta-z)^2 R(\zeta)} \\ \vdots & \ddots & \vdots & \vdots \\ \oint_{\hat{\gamma}_{m,N}} \frac{d\zeta}{R(\zeta)} & \cdots & \oint_{\hat{\gamma}_{m,N}} \frac{\zeta^{N-1} d\zeta}{R(\zeta)} & \oint_{\hat{\gamma}_{m,N}} \frac{d\zeta}{(\zeta-z)^2 R(\zeta)} \\ \oint_{\hat{\gamma}_{c,1}} \frac{d\zeta}{R(\zeta)} & \cdots & \oint_{\hat{\gamma}_{c,1}} \frac{\zeta^{N-1} d\zeta}{R(\zeta)} & \oint_{\hat{\gamma}_{c,1}} \frac{d\zeta}{(\zeta-z)^2 R(\zeta)} \\ \vdots & \ddots & \vdots & \vdots \\ \oint_{\hat{\gamma}_{c,N}} \frac{d\zeta}{R(\zeta)} & \cdots & \oint_{\hat{\gamma}_{c,N}} \frac{\zeta^{N-1} d\zeta}{R(\zeta)} & \oint_{\hat{\gamma}_{c,N}} \frac{d\zeta}{(\zeta-z)^2 R(\zeta)} \\ \oint_{\hat{\gamma}} \frac{f(\zeta)\, d\zeta}{R(\zeta)} & \cdots & \oint_{\hat{\gamma}} \frac{\zeta^{N-1} f(\zeta)\, d\zeta}{R(\zeta)} & \oint_{\hat{\gamma}} \frac{f(\zeta)\, d\zeta}{(\zeta-z)^2 R(\zeta)} \end{vmatrix}, \tag{46}$$

where $z$ is inside of $\hat{\gamma}(\mu)$ and inside of $\hat{\gamma}_{m,j}$ and $\hat{\gamma}_{c,j}$ or $\hat{\gamma}_{c,j+1}$.

**Lemma 4.3.** *Let $f$ be given by* (21) *and consider a contour $\gamma_0 \in \Gamma(\vec{\alpha}_0,\mu_0)$, where $\vec{\alpha}_0$ and $\mu_0$ satisfy*

$$\vec{K}(\vec{\alpha}_0,\mu_0) = \vec{0}.$$

*Assume that for $\vec{\alpha}_0 = \{\alpha_j^0\}_{j=0}^{4N+1}$,*

$$\lim_{z\to\alpha_j^0} K'(z,\vec{\alpha}_0,\mu_0) \neq 0, \quad j = 0,1,\dots,4N+1.$$

*Then the modulation equations*

$$\vec{K}(\vec{\alpha},\mu) = \vec{0}$$

*can be uniquely solved for $\vec{\alpha} = \vec{\alpha}(\mu)$, which is continuously differentiable for all $\mu$ in some open neighborhood of $\mu_0$ and $\vec{\alpha}(\mu_0) = \vec{\alpha}_0$.*

*Proof.* $\vec{K}$ is continuously differentiable in $\vec{\alpha}$ and in $\mu$ by Lemma 4.2.

As shown in [Tovbis and Venakides 2009], the matrix

$$\left\{\frac{\partial \vec{K}}{\partial \vec{\alpha}}\right\}_{j,l} = \left\{\frac{\partial K(\alpha_j)}{\partial \alpha_l}\right\}_{j,l}$$

is diagonal and

$$\frac{\partial K(\alpha_j)}{\partial \alpha_j} = \frac{3}{2} D \lim_{z\to\alpha_j^0}\left(\frac{h(z)}{R(z)}\right)' = \frac{3}{2}\lim_{z\to\alpha_j^0} K'(z,\vec{\alpha},\mu) \neq 0. \tag{47}$$

So

$$\det\left|\frac{\partial \vec{K}}{\partial \vec{\alpha}}(\vec{\alpha}_0)\right| = \prod_j \frac{\partial K(\alpha_j)}{\partial \alpha_j} \neq 0, \tag{48}$$

under the assumptions. By the implicit function theorem, $\vec{\alpha}(\mu)$ are uniquely defined in some neighborhood of $\mu_0$ and smooth in $\mu$. Note that $\vec{\alpha}(\mu_0) = \vec{\alpha}_0$ by assumption. □

**Remark 4.4.** The condition $\lim_{z\to\alpha_j^0} K'(z,\vec{\alpha}_0,\mu_0)\neq 0$, $j=0,1,\dots,4N+1$, in Lemma 4.3 is equivalent to

$$\lim_{z\to\alpha_j^0}\frac{h'(z,\vec{\alpha}_0,\mu_0)}{R(z,\vec{\alpha}_0)}\neq 0,\quad j=0,1,\dots,4N+1.$$

All quantities below depend on parameters $x$ and $t$. We assume that for the rest of the paper $x$ and $t$ are fixed.

**Theorem 4.5** ($\mu$-perturbation in genus $N$). *Consider a finite-length non-self-intersecting contour $\gamma_0$ in the complex plane consisting of a finite union of oriented arcs*

$$\gamma_0=\left(\bigcup\gamma_{m,j}\right)\cup\left(\bigcup\gamma_{c,j}\right)\in\Gamma(\vec{\alpha}_0,\mu_0)$$

*with the distinct arc end points $\vec{\alpha}_0$ and depending on parameter $\mu$ (see Figure 1). Assume $\vec{\alpha}_0$ and $\mu_0$ satisfy a system of equations*

$$\vec{K}(\vec{\alpha}_0,\mu_0)=\vec{0},$$

*and $f$ is given by* (21). *Let $\gamma=\gamma(\vec{\alpha},\mu)$ be a contour of an RHP which seeks a function $h(z)$ which satisfies the conditions*

$$\begin{cases}h_+(z)+h_-(z)=2W_j & \text{on } \gamma_{m,j},\ j=0,1,\dots,N,\\ h_+(z)-h_-(z)=2\Omega_j & \text{on } \gamma_{c,j},\ j=1,\dots,N,\\ h(z)+f(z) \text{ is analytic in } \overline{\mathbb{C}}\setminus\gamma,\end{cases}\tag{49}$$

*where $\Omega_j=\Omega_j(\vec{\alpha},\mu)$ and $W_j=W_j(\vec{\alpha},\mu)$ are real constants (with normalization $W_0=0$) whose numerical values will be determined from the RH conditions. Assume that there is a function $h(z,\vec{\alpha}_0,\mu_0)$ which satisfies* (49) *and suppose $h'(z,\vec{\alpha}_0,\mu_0)/R(z,\vec{\alpha}_0)\neq 0$ for all $z$ on $\gamma_0$.*

*Then there is a contour $\gamma(\vec{\alpha},\mu)\in\Gamma(\vec{\alpha},\mu)$ such that the solution $\vec{\alpha}=\vec{\alpha}(\mu)$ of the system*

$$\vec{K}(\vec{\alpha},\mu)=\vec{0}\tag{50}$$

*and $h(z,\vec{\alpha}(\mu),\mu)$ which solves* (49) *are uniquely defined and continuously differentiable in $\mu$ in some open neighborhood of $\mu_0$.*

*Moreover,*

$$\frac{\partial\alpha_j}{\partial\mu}(\mu)=-\frac{2\pi i\,\dfrac{\partial K}{\partial\mu}(\alpha_j(\mu),\vec{\alpha}(\mu),\mu)}{D(\vec{\alpha}(\mu),\mu)\displaystyle\oint_{\hat{\gamma}(\mu)}\frac{f'(\zeta,\mu)}{(\zeta-\alpha_j(\mu))R(\zeta,\vec{\alpha}(\mu))}\,d\zeta},\tag{51}$$

$$\frac{\partial h}{\partial\mu}(z,\mu)=\frac{R(z,\vec{\alpha}(\mu))}{2\pi i}\oint_{\hat{\gamma}(\mu)}\frac{\dfrac{\partial f}{\partial\mu}(\zeta,\mu)}{(\zeta-z)R(\zeta,\vec{\alpha}(\mu))}\,d\zeta,\tag{52}$$

*where $z$ is inside of $\hat{\gamma}$.*

*Furthermore, $\Omega_j(\mu) = \Omega_j(\vec{\alpha}(\mu), \mu)$ and $W_j(\mu) = W_j(\vec{\alpha}(\mu), \mu)$ are defined and continuously differentiable in $\mu$ in some open neighborhood of $\mu_0$, and*

$$
\frac{\partial \Omega_j}{\partial \mu}(\mu) = -\frac{1}{D}
\begin{vmatrix}
\oint_{\hat{\gamma}_{m,1}} \frac{d\zeta}{R(\zeta)} & \cdots & \oint_{\hat{\gamma}_{m,1}} \frac{\zeta^{N-1} d\zeta}{R(\zeta)} \\
\vdots & \ddots & \vdots \\
\oint_{\hat{\gamma}_{m,N}} \frac{d\zeta}{R(\zeta)} & \cdots & \oint_{\hat{\gamma}_{m,N}} \frac{\zeta^{N-1} d\zeta}{R(\zeta)} \\
\oint_{\hat{\gamma}_{c,1}} \frac{d\zeta}{R(\zeta)} & \cdots & \oint_{\hat{\gamma}_{c,1}} \frac{\zeta^{N-1} d\zeta}{R(\zeta)} \\
\vdots & \ddots & \vdots \\
\oint_{\hat{\gamma}_{c,j-1}} \frac{d\zeta}{R(\zeta)} & \cdots & \oint_{\hat{\gamma}_{c,j-1}} \frac{\zeta^{N-1} d\zeta}{R(\zeta)} \\
\oint_{\hat{\gamma}} \frac{f_\mu(\zeta)}{R(\zeta, \vec{\alpha})} d\zeta & \cdots & \oint_{\hat{\gamma}} \frac{\zeta^{N-1} f_\mu(\zeta)}{R(\zeta, \vec{\alpha})} d\zeta \\
\oint_{\hat{\gamma}_{c,j+1}} \frac{d\zeta}{R(\zeta)} & \cdots & \oint_{\hat{\gamma}_{c,j+1}} \frac{\zeta^{N-1} d\zeta}{R(\zeta)} \\
\vdots & \ddots & \vdots \\
\oint_{\hat{\gamma}_{c,N}} \frac{d\zeta}{R(\zeta)} & \cdots & \oint_{\hat{\gamma}_{c,N}} \frac{\zeta^{N-1} d\zeta}{R(\zeta)}
\end{vmatrix},
$$

$$
\frac{\partial W_j}{\partial \mu}(\mu) = -\frac{1}{D}
\begin{vmatrix}
\oint_{\hat{\gamma}_{m,1}} \frac{d\zeta}{R(\zeta)} & \cdots & \oint_{\hat{\gamma}_{m,1}} \frac{\zeta^{N-1} d\zeta}{R(\zeta)} \\
\vdots & \ddots & \vdots \\
\oint_{\hat{\gamma}_{m,j-1}} \frac{d\zeta}{R(\zeta)} & \cdots & \oint_{\hat{\gamma}_{m,j-1}} \frac{\zeta^{N-1} d\zeta}{R(\zeta)} \\
\oint_{\hat{\gamma}} \frac{f_\mu(\zeta)}{R(\zeta, \vec{\alpha})} d\zeta & \cdots & \oint_{\hat{\gamma}} \frac{\zeta^{N-1} f_\mu(\zeta)}{R(\zeta, \vec{\alpha})} d\zeta \\
\oint_{\hat{\gamma}_{m,j+1}} \frac{d\zeta}{R(\zeta)} & \cdots & \oint_{\hat{\gamma}_{m,j+1}} \frac{\zeta^{N-1} d\zeta}{R(\zeta)} \\
\vdots & \ddots & \vdots \\
\oint_{\hat{\gamma}_{m,N}} \frac{d\zeta}{R(\zeta)} & \cdots & \oint_{\hat{\gamma}_{m,N}} \frac{\zeta^{N-1} d\zeta}{R(\zeta)} \\
\oint_{\hat{\gamma}_{c,1}} \frac{d\zeta}{R(\zeta)} & \cdots & \oint_{\hat{\gamma}_{c,1}} \frac{\zeta^{N-1} d\zeta}{R(\zeta)} \\
\vdots & \ddots & \vdots \\
\oint_{\hat{\gamma}_{c,N}} \frac{d\zeta}{R(\zeta)} & \cdots & \oint_{\hat{\gamma}_{c,N}} \frac{\zeta^{N-1} d\zeta}{R(\zeta)}
\end{vmatrix}, \tag{53}
$$

*where $R(\zeta) = R(\zeta, \vec{\alpha})$, $f(\zeta) = f(\zeta, \mu)$, $f_\mu(\zeta) = \dfrac{\partial f}{\partial \mu}(\zeta, \mu)$, and $D = D(\vec{\alpha}(\mu))$.*

*Proof.* By Lemma 4.3, there is a contour $\gamma(\vec{\alpha},\mu)\in\Gamma(\vec{\alpha},\mu)$ for all $\mu$ in some neighborhood of $\mu_0$ and the $\alpha_j(\mu)$ are continuously differentiable in $\mu$. The formula for $\partial\alpha_j/\partial\mu$ is derived similarly as in [Tovbis and Venakides 2009]. We differentiate the modulation equations $K(\alpha_j)=K(\alpha_j,\vec{\alpha},\mu)=0$ which define $\vec{\alpha}=\vec{\alpha}(\mu)$ with respect to $\mu$,

$$\sum_{l=0}^{4N+1}\frac{\partial K(\alpha_j)}{\partial\alpha_l}\frac{\partial\alpha_l}{\partial\mu}+\frac{\partial K}{\partial\mu}(\alpha_j)=0, \tag{54}$$

where the matrix $\left\{\frac{\partial K(\alpha_j)}{\partial\alpha_l}\right\}_{j,l}$ is diagonal [Tovbis and Venakides 2009], so

$$\frac{\partial K(\alpha_j)}{\partial\alpha_j}\frac{\partial\alpha_j}{\partial\mu}=-\frac{\partial K(\alpha_j)}{\partial\mu}. \tag{55}$$

Since

$$\frac{\partial K(\alpha_j)}{\partial\alpha_j}=\frac{D(\vec{\alpha},\mu)}{2\pi i}\oint_{\hat{\gamma}(\mu)}\frac{f'(\zeta,\mu)}{(\zeta-\alpha_j)R(\zeta,\vec{\alpha})}\,d\zeta, \tag{56}$$

we arrive at the evolution equations for the $\alpha_j$:

$$\frac{\partial\alpha_j}{\partial\mu}=-\frac{2\pi i\,\frac{\partial K}{\partial\mu}(\alpha_j)}{D(\vec{\alpha},\mu)\oint_{\hat{\gamma}(\mu)}\frac{f'(\zeta,\mu)}{(\zeta-\alpha_j)R(\zeta,\vec{\alpha})}\,d\zeta},\quad j=0,\dots,4N+1. \tag{57}$$

Next we compute $\frac{\partial h}{\partial\mu}$, which satisfies the scalar RHP

$$\begin{cases}h_{\mu,+}(z)+h_{\mu,-}(z)=0,\ z\in\gamma_{m,j},\ j=0,1,\dots,N,\\ h_\mu(z)+f_\mu(z)\text{ is analytic in }\overline{\mathbb{C}}\setminus\gamma.\end{cases} \tag{58}$$

Then

$$\frac{\partial h}{\partial\mu}(z,\mu)=\frac{R(z,\vec{\alpha}(\mu))}{2\pi i}\oint_{\hat{\gamma}(\mu)}\frac{\frac{\partial f}{\partial\mu}(\zeta,\mu)}{(\zeta-z)R(\zeta,\vec{\alpha}(\mu))}\,d\zeta, \tag{59}$$

where $z$ is inside of $\hat{\gamma}$. The integrand $\frac{\partial f}{\partial\mu}(\zeta,\mu)$ behaves like $\log(\zeta-z_0)$ near $\zeta=z_0$, and therefore is integrable.

Constants $W_j$ and $\Omega_j$ are found from the linear system [Tovbis et al. 2004]

$$\oint_{\hat{\gamma}(\mu)}\frac{\zeta^n f(\zeta,\mu)}{R(\zeta,\vec{\alpha})}\,d\zeta+\sum_{j=1}^{N}\oint_{\hat{\gamma}_{c,j}}\frac{\zeta^n\Omega_j}{R(\zeta,\vec{\alpha})}\,d\zeta+\sum_{j=1}^{N}\oint_{\hat{\gamma}_{m,j}}\frac{\zeta^n W_j}{R(\zeta,\vec{\alpha})}\,d\zeta=0,\quad n=0,\dots,N-1. \tag{60}$$

Differentiating in $\mu$ and using Lemma 4.1 leads to

$$\oint_{\hat{\gamma}(\mu)}\frac{\zeta^n f_\mu(\zeta,\mu)}{R(\zeta,\vec{\alpha})}\,d\zeta+\sum_{j=1}^{N}\oint_{\hat{\gamma}_{c,j}}\frac{\zeta^n(\Omega_j)_\mu}{R(\zeta,\vec{\alpha})}\,d\zeta+\sum_{j=1}^{N}\oint_{\hat{\gamma}_{m,j}}\frac{\zeta^n(W_j)_\mu}{R(\zeta,\vec{\alpha})}\,d\zeta=0,\quad n=0,\dots,N-1, \tag{61}$$

or in matrix form,

$$\begin{pmatrix} \oint_{\hat\gamma_{m,1}} \frac{d\zeta}{R(\zeta)} & \cdots & \oint_{\hat\gamma_{m,1}} \frac{\zeta^{N-1}d\zeta}{R(\zeta)} \\ \vdots & \ddots & \vdots \\ \oint_{\hat\gamma_{m,N}} \frac{d\zeta}{R(\zeta)} & \cdots & \oint_{\hat\gamma_{m,N}} \frac{\zeta^{N-1}d\zeta}{R(\zeta)} \\ \oint_{\hat\gamma_{c,1}} \frac{d\zeta}{R(\zeta)} & \cdots & \oint_{\hat\gamma_{c,1}} \frac{\zeta^{N-1}d\zeta}{R(\zeta)} \\ \vdots & \ddots & \vdots \\ \oint_{\hat\gamma_{c,N}} \frac{d\zeta}{R(\zeta)} & \cdots & \oint_{\hat\gamma_{c,N}} \frac{\zeta^{N-1}d\zeta}{R(\zeta)} \end{pmatrix}^T \begin{pmatrix} \frac{\partial \vec W}{\partial\mu} \\ \frac{\partial\vec\Omega}{\partial\mu} \end{pmatrix} = -\begin{pmatrix} \oint_{\hat\gamma(\mu)} \frac{f_\mu(\zeta,\mu)}{R(\zeta,\vec\alpha)}\,d\zeta \\ \vdots \\ \oint_{\hat\gamma(\mu)} \frac{\zeta^{N-1} f_\mu(\zeta,\mu)}{R(\zeta,\vec\alpha)}\,d\zeta \end{pmatrix}. \tag{62}$$

So $\dfrac{\partial\Omega_j}{\partial\mu}$ and $\dfrac{\partial W_j}{\partial\mu}$ satisfy (53). Note that $D \neq 0$ for distinct $\alpha_j$'s [Tovbis and Venakides 2009]. □

**Remark 4.6.** In [Tovbis and Venakides 2009], the case was considered where the contour $\gamma$ is independent of external parameters $x$ and $t$ and the dependence of $f$ on these parameters is linear. Here we apply the methods of that paper to the case of a dependence on the parameter $\mu$ when the jump contour explicitly passes through $z = \mu/2$, a point of singularity of $f$. Despite this more complicated dependence on $\mu$, the resulting formulae are the same. The main reason is Lemma 4.1, which allows us to find partial derivatives with respect to $\mu$ of contour integrals involving dependence on $\mu$ in both integrands and contours of integration.

**Remark 4.7.** Theorem 4.5 guarantees that the solution of the RHP (49) is uniquely continued with respect to external parameters. Additional sign conditions on $\Im h$ need to be satisfied for $h$ to correspond to an asymptotic solution of the NLS as in [Tovbis et al. 2004]. The sign conditions have to be satisfied near $\gamma$ and additionally on semi-infinite complementary arcs connecting the arc end points of $\gamma$ to $\infty$.

## 5. Sign conditions and preservation of genus

If the scalar RHP (12) is implemented in the asymptotic solution of the semiclassical NLS, certain sign conditions must be satisfied. Specifically, $\Im h(z) = 0$ on $\gamma_{m,j}$, $\Im h(z) < 0$ on both sides of $\gamma_{m,j}$, and $\Im h(z) \geq 0$ on $\gamma_{c,j}$ (see Definition 5.3 below). In this section we investigate the preservation of the sign structure of $\Im h$ under perturbations of $\mu$.

**Definition 5.1.** Define $\gamma^\infty = \gamma^\infty(\vec\alpha, \mu)$ as an extension of a contour $\gamma(\vec\alpha,\mu) \in \Gamma(\vec\alpha,\mu)$ as $\gamma^\infty(\vec\alpha,\mu) = (\infty, \alpha_{4N+1}] \cup \gamma(\vec\alpha,\mu) \cup [\alpha_{4N}, \infty)$. Both additional arcs are considered as a complementary arc $\gamma_{c,N+1} = (\infty, \alpha_{4N+1}] \cup [\alpha_{4N}, \infty)$, and assume $\gamma_{c,N+1} = \overline{\gamma_{c,N+1}}$, so $\gamma^\infty = \overline{\gamma^\infty}$. With a slight abuse of notation we write $\gamma^\infty(\vec\alpha,\mu) \in \Gamma(\vec\alpha,\mu)$.

**Lemma 5.2.** *If the conditions of Theorem 4.5 hold on $\gamma^\infty(\vec\alpha_0,\mu_0) \in \Gamma(\vec\alpha_0,\mu_0)$ for $\vec K(\vec\alpha_0,\mu_0) = \vec 0$, the statement of the theorem holds on $\gamma^\infty(\vec\alpha,\mu)$, where $\vec K(\vec\alpha,\mu) = \vec 0$.*

*Proof.* The proof is unchanged since $f$ is analytic near the additional semi-infinite arcs in $\gamma_{c,N+1}$ and the jump condition on the additional complementary arc $\gamma_{c,N+1}$ is taken to be zero ($\Omega_{N+1} = 0$) [Tovbis et al. 2004]. □

Note that the conditions in Lemma 5.2 are more restrictive since $\gamma \subset \gamma^\infty$.

**Definition 5.3.** A function $h$ satisfies sign conditions on $\gamma^\infty$ if $\Im h(z) = 0$ if $z \in \gamma_{m,j}$, $\Im h(z) < 0$ on both sides of $\gamma_{m,j}$ for all $j = 0, \ldots, N$, and $\Im h(z) \geq 0$ if $z \in \gamma_{c,j}$ for all $j = 1, \ldots, N+1$. We then write $h \in \mathrm{SC}(\gamma^\infty)$.

Note that the zero sign conditions ($\Im h(z) = 0$) on $\gamma_{m,j}$ are satisfied automatically through the construction of $h(z)$ by (14) in the case of $h$ solving an RHP (49). We only need to check preservation of negative signs of $\Im h$ on both sides of the main arcs $\gamma_{m,j}$ and the nonnegativity of $\Im h$ on the complementary arcs $\gamma_{c,j}$, especially on the semi-infinite arcs $(\infty, \alpha_{4N+1}]$ and $[\alpha_{4N}, \infty)$.

**Remark 5.4.** Introducing the sign conditions in Definition 5.3 requires us to revisit Lemma 2.2, since the main arcs $\gamma_{m,j}$ are now rigid (nondeformable like the complementary arcs) due to the requirement for $\Im h$ to be negative on both sides of $\gamma(\vec{\alpha}(\mu), \mu)$. It has been established [Kamvissis et al. 2003, Lemma 5.2.1; Tovbis et al. 2007, Theorem 3.2] that all the contours persist under deformations of parameters $x$ and $t$ provided all sign inequalities are satisfied. This can be adapted to the deformations of $\mu$. The danger of a main arc splitting into several disconnected branches as we perturb $\mu$ is averted by the fact that in the limit as $\mu \to \mu_0$, nonlinear local behavior would be produced near the main arcs, while the condition $h'/R \neq 0$ on $\gamma$ implies linear local behavior. Thus Lemma 2.2 is valid even with the added new sign conditions. Thus we only need to show that the sign conditions are satisfied.

**Theorem 5.5.** *Let $f$ be defined by* (21). *Let $\vec{K}(\vec{\alpha}_0, \mu_0) = \vec{0}$, $\gamma_0^\infty \in \Gamma(\vec{\alpha}_0, \mu_0)$ and assume $h$ solves* $\mathrm{RHP}(\gamma_0^\infty, \vec{\alpha}_0, \mu_0, f)$ *with $h'(z, \mu_0)/R(z, \mu_0) \neq 0$ for all $z \in \gamma_0^\infty$, and $h \in \mathrm{SC}(\gamma_0^\infty)$.*

*Then there is an open neighborhood of $\mu_0$ where for all $\mu$, there is an $h$ which solves* $\mathrm{RHP}(\gamma^\infty, \vec{\alpha}, \mu, f)$ *with $\gamma^\infty = \gamma^\infty(\vec{\alpha}, \mu)$, $\vec{K}(\vec{\alpha}, \mu) = \vec{0}$, $h'(z, \mu)/R(z, \mu) \neq 0$ for all $z \in \gamma^\infty$, and $h \in \mathrm{SC}(\gamma^\infty)$.*

*Proof.* Take any $\mu$ in a small enough open neighborhood of $\mu_0$. There are two things we need to prove in addition to Lemma 5.2: $h'(z, \mu)/R(z, \vec{\alpha}(\mu)) \neq 0$ on $\gamma^\infty$ and the sign conditions of $\Im h$ on $\gamma^\infty$.

Assume $h'(z, \mu_0)/R(z, \vec{\alpha}(\mu_0)) \neq 0$ on $\gamma_0^\infty$. Then there is a constant $C > 0$ such that

$$\left| \frac{h'(z, \mu_0)}{R(z, \vec{\alpha}(\mu_0))} \right| > C$$

for all $z \in \gamma_0$. Consider the solution $h(z, \mu)$ of $\mathrm{RHP}(\gamma^\infty, \vec{\alpha}, \mu)$, where $\vec{K}(\vec{\alpha}, \mu) = \vec{0}$. By Theorem 4.5 and Lemma 5.2, such a function exists and is continuously differentiable in $\mu$. Moreover, $h'(z, \mu)$ is continuous in $\mu$. Since $\gamma$ is a compact set in $\mathbb{C}$ and $h'(z, \mu)/R(z, \vec{\alpha}(\mu))$ is continuous in $z$ and $\mu$, we have $h'(z, \mu)/R(z, \vec{\alpha}(\mu)) \neq 0$ for all $z \in \gamma$.

To show that $h'(z, \mu)/R(z, \vec{\alpha}(\mu)) \neq 0$ holds on $\gamma^\infty$, we will now make use of the following properties of $f(z)$. On the real axis (in the nontangential limit from the upper half-plane),

$$\Im f(z + i0) = \lim_{\delta \to 0^+} f(z + i\delta), \quad z \in \mathbb{R},$$

is a piecewise linear function [Tovbis et al. 2004]

$$\Im f(z+i0) = \begin{cases} \dfrac{\pi}{2}\left(\dfrac{\mu}{2} - |z|\right) & \text{if } z < \dfrac{\mu}{2}, \\ \dfrac{\pi}{2}\left(z - \dfrac{\mu}{2}\right) & \text{if } z \geq \dfrac{\mu}{2}, \end{cases} \tag{63}$$

and since $g(z)$ is real on the real axis, $\Im h(z+i0) = -\Im f(z+i0)$. It is important for us that $|\Im h(z)|$ can be bounded away from zero as $z \to \infty$.

Similarly,

$$\Im f'(z+i0) = \begin{cases} \dfrac{\pi}{2} & \text{if } z \leq 0, \\ -\dfrac{\pi}{2} & \text{if } 0 < z \leq \dfrac{\mu}{2}, \\ \dfrac{\pi}{2} & \text{if } z > \dfrac{\mu}{2}, \end{cases}$$

and since $g'(z)$ is real on the real axis, $\Im h'(z+i0) = -\Im f'(z+i0)$.

Recall that $\gamma^\infty = (\infty, \alpha_{4N+1}] \cup \gamma \cup [\alpha_{4N}, \infty)$. The semi-infinite arcs $(\infty, \alpha_{4N+1}]$ and $[\alpha_{4N}, \infty)$ can be pushed to the real axis as $(-\infty - i0, -\mu/2 - i0) \cup [-\mu/2 - i0, \alpha_{4N+1}]$ and $[\alpha_{4N}, -\mu/2 + i0] \cup (-\mu/2 + i0, -\infty + i0)$, respectively.

On $[-\mu/2 - i0, \alpha_{4N+1}]$ and $[\alpha_{4N}, -\mu/2 + i0]$, we have $h'(z,\mu)/R(z, \vec{\alpha}(\mu)) \neq 0$ by continuity on a compact set. Finally, $\Im h'(z,\mu) = -\pi/2$ and $R(z, \vec{\alpha}) \in \mathbb{R}$ for all $z \in (-\mu/2 + i0, -\infty + i0)$. So $h'(z,\mu)/R(z, \vec{\alpha}(\mu)) \neq 0$ for all $z \in (-\mu/2 + i0, -\infty + i0)$. The interval $(-\infty - i0, -\mu/2 - i0)$ is done similarly. So $h'(z,\mu)/R(z, \vec{\alpha}(\mu)) \neq 0$ for all $z \in \gamma^\infty$, for any $\mu$ in the neighborhood of $\mu_0$.

Let $h \in \mathrm{SC}(\gamma_0^\infty)$. Then $h \in \mathrm{SC}(\gamma(\mu))$ by continuity of $h$ in $z$ and $\mu$, compactness of $\gamma$, and harmonicity of $\Im h$ combined with $h'(z,\mu)/R(z, \vec{\alpha}(\mu)) \neq 0$ for all $z \in \gamma^\infty$, which guarantees that the (negative) signs near the main arcs $\hat{\gamma}_{m,j}$ are preserved. On the semi-infinite arcs $(-\infty - i0, -\mu/2 - i0)$ and $(-\mu/2 + i0, -\infty + i0)$, $\Im h(z) = (\pi/2)(|z| - \mu/2)$ is positive and $[\alpha_{4N}, -\mu/2]$ and $[-\mu/2, \alpha_{4N+1}]$ are compact. So $\Im h \geq 0$ on $\gamma^\infty(\mu)$, that is, $h \in \mathrm{SC}(\gamma^\infty(\mu))$. □

**Definition 5.6.** We define the (finite) genus $G = G(\mu)$ of the asymptotic solution of the semiclassical one-dimensional focusing NLS with initial condition defined through $f(z,\mu)$ as (finite) $N \in \mathbb{N}$ if there exists an asymptotic solution of the NLS through the solution $h(z,\mu)$ of $\mathrm{RHP}(\gamma^\infty, \vec{\alpha}, \mu, f)$ with $\vec{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_{4N+1})$, such that $h'(z,\mu)/R(z) \neq 0$ for all $z \in \gamma^\infty$ and the sign conditions of $h$ on $\gamma^\infty$ are satisfied: $h \in \mathrm{SC}(\gamma^\infty)$.

**Remark 5.7.** This definition of the genus of the asymptotic solution coincides with the genus of the (limiting) hyperelliptic Riemann surface of $R(z)$.

**Theorem 5.8** (preservation of genus). *Suppose that for $\mu_0$, the genus of the asymptotic solution of the NLS with initial condition defined through $f(z,\mu_0)$ in* (21) *is $G(\mu_0)$.*

*Then there is an open neighborhood of $\mu_0$ such that, for all $\mu$ in the neighborhood of $\mu_0$, the genus of the asymptotic solution of the NLS with initial condition defined through $f(z,\mu)$ is preserved*: $G(\mu) = G(\mu_0)$.

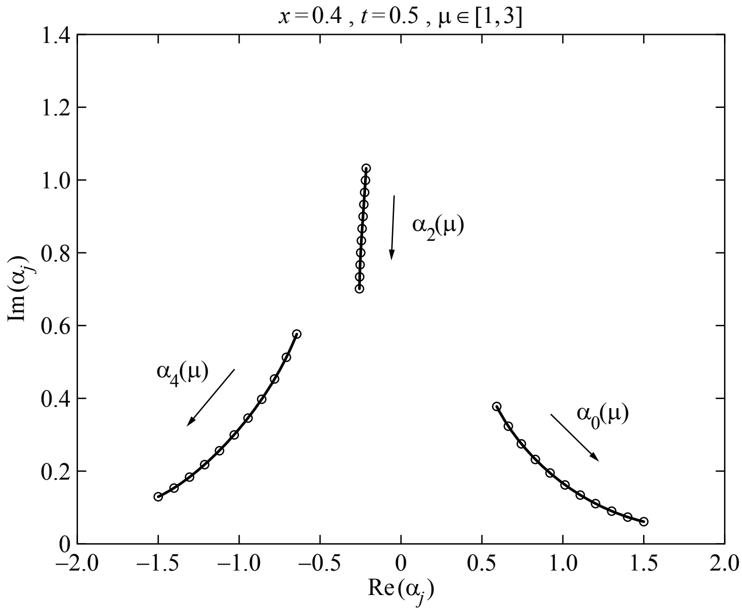



**Figure 3.** Comparison of $\mu$ evolution of $\vec{\alpha} = (\alpha_0, \alpha_2, \alpha_4)$ using (82) (solid lines) and (83) (circles).

*Proof.* Follows from Theorem 5.5 and Definition 5.6. □

**Corollary 5.9.** *Fix $x$ and $t > t_0$, where $t_0(x)$ is the time of the first break in the asymptotic solution. Then in some open neighborhood of $\mu = 2$, the genus of the solution is 2.*

*Proof.* For $\mu = 2$ and $t > t_0(x)$ the genus is 2 for all $x$ [Tovbis et al. 2004]. By Theorem 5.8, the genus is preserved in some open neighborhood of $\mu = 2$, including some open interval for $\mu < 2$. □

## 6. Numerics

Figure 3 compares solutions of (50) and (51) in genus 2 (see also (82) and (83) in the Appendix for more explicit expressions). The solutions are practically indistinguishable in the figure, with absolute difference less than $10^{-3}$ for $\mu \in [1, 3]$. This interval includes the critical value $\mu = 2$, which is the transition between the (solitonless) pure radiation case ($\mu \geq 2$) and the region with solitons ($0 < \mu < 2$). Computations are based on the code we developed for long-time studies of an obstruction in the $g$-function mechanism [Belov and Venakides 2015].

## Appendix

**A1. *Genus* 0 *region.*** It was shown in [Tovbis et al. 2004] that for all $\mu > 0$ and for all $x$, there is a breaking curve $t = t_0(x)$ in the $(x, t)$ plane. The region $0 \leq t < t_0(x)$ has genus 0 in the sense of genus of the underlying Riemann surface for the square root

$$R(z, \alpha_0) = \sqrt{(z - \alpha_0)(z - \alpha_1)}, \quad \alpha_1 = \overline{\alpha}_0,$$

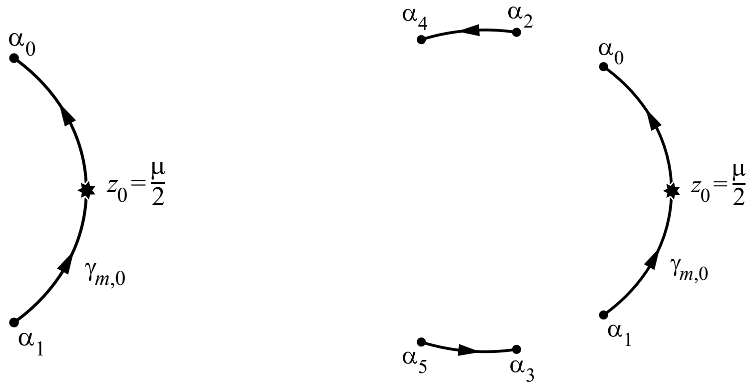



**Figure 4.** The jump contour in the case of genus 0 (left diagram) and genus 2 (right) with complex-conjugate symmetry in the notation of [Tovbis et al. 2004].

where the branch cut is chosen along the main arc connecting $\alpha_0$ and $\alpha_1 = \overline{\alpha}_0$ through $z = \mu/2$, and the branch is fixed by $R(z) \to -z$ as $z \to +\infty$. The asymptotic solution of the NLS is expressed in terms of $\alpha_0 = \alpha_0(x, t, \mu)$.

All expressions in the genus 0 region ($N = 0$) have a simpler form. In particular,

$$h(z, \alpha_0, \mu) = \frac{R(z, \alpha_0)}{2\pi i} \oint_{\hat{\gamma}(\mu)} \frac{f(\zeta, \mu)\, d\zeta}{(\zeta - z) R(\zeta, \alpha_0)}, \tag{64}$$

$$K(z, \alpha_0, \mu) = \frac{1}{2\pi i} \oint_{\hat{\gamma}(\mu)} \frac{f(\zeta, \mu)\, d\zeta}{(\zeta - z) R(\zeta, \alpha_0)}, \tag{65}$$

and with a slight abuse of notation,

$$K(\alpha_0, \mu) := K(\alpha_0, \alpha_0, \mu) = \frac{1}{2\pi i} \oint_{\hat{\gamma}(\mu)} \frac{f(\zeta, \mu) d\zeta}{(\zeta - \alpha_0) R(\zeta, \alpha_0)}, \tag{66}$$

$$\frac{\partial K}{\partial \mu}(\alpha_0, \mu) := \frac{\partial K}{\partial \mu}(\alpha_0, \alpha_0, \mu) = \frac{1}{2\pi i} \oint_{\hat{\gamma}(\mu)} \frac{f_\mu(\zeta, \mu) d\zeta}{(\zeta - \alpha_0) R(\zeta, \alpha_0)}. \tag{67}$$

**Theorem A.1** ($\mu$-perturbation in genus 0). *Consider a finite-length non-self-intersecting oriented arc $\gamma_0 = [\alpha_0(\mu_0), \overline{\alpha}_0(\mu_0)] \in \Gamma(\vec{\alpha}, \mu_0)$ in the complex plane with the distinct end points ($\alpha_0 \neq \overline{\alpha}_0$) and depending on a parameter $\mu$* (see Figure 4). *Assume $\alpha_0$ and $\mu_0$ satisfy the equation*

$$K(\alpha_0, \mu_0) = 0,$$

*and $f$ is given by* (21). *Let $\gamma = \gamma(\vec{\alpha}, \mu)$ be the contour of an RHP which seeks a function $h(z)$ which satisfies the conditions*

$$\begin{cases} h_+(z) + h_-(z) = 0 \textit{ on } \gamma, \\ h(z) + f(z) \textit{ is analytic in } \overline{\mathbb{C}} \setminus \gamma. \end{cases} \tag{68}$$

*Assume that there is a function $h(z, \alpha_0, \mu_0)$ which satisfies* (68) *and suppose $\dfrac{h'(z, \alpha_0, \mu_0)}{R(z, \alpha_0)} \neq 0$ for all $z$ on $\gamma_0$.*

*Then there is a contour $\gamma(\vec{\alpha}, \mu) \in \Gamma(\vec{\alpha}, \mu)$ such that the solution $\alpha_0(\mu)$ of the equation*

$$K(\alpha_0, \mu) = 0 \tag{69}$$

*and $h(z, \alpha(\mu), \mu)$ which solves* (68) *are uniquely defined and continuously differentiable in $\mu$ in some open neighborhood of $\mu_0$.*

*Moreover,*

$$\frac{\partial \alpha_0}{\partial \mu}(\mu) = -\frac{2\pi i \dfrac{\partial K}{\partial \mu}(\alpha_0(\mu), \mu)}{\displaystyle\oint_{\hat{\gamma}(\mu)} \frac{f'(\zeta, \mu)}{(\zeta - \alpha_0(\mu)) R(\zeta, \alpha_0(\mu))}\, d\zeta} \tag{70}$$

*and*

$$\frac{\partial h}{\partial \mu}(z, \mu) = \frac{R(z, \alpha_0(\mu))}{2\pi i} \oint_{\hat{\gamma}(\mu)} \frac{\dfrac{\partial f}{\partial \mu}(\zeta, \mu)}{(\zeta - z) R(\zeta, \alpha_0(\mu))}\, d\zeta, \tag{71}$$

*where $z$ is inside of $\hat{\gamma}$.*

**A2.** ***Genus 2 region.*** We now consider the genus 2 region ($N = 2$), with underlying Riemann surface for the square root

$$R(z) = \sqrt{(z-\alpha_0)(z-\alpha_1)(z-\alpha_2)(z-\alpha_3)(z-\alpha_4)(z-\alpha_5)},$$

where the branch cut is chosen along the main arcs connecting $\alpha_0$ and $\alpha_1$, $\alpha_2$ and $\alpha_4$, $\alpha_5$ and $\alpha_3$; and the branch is fixed by $R(z) \to -z^3$ as $z \to +\infty$.

Taking into account the complex-conjugate symmetry

$$\alpha_1 = \overline{\alpha}_0, \quad \alpha_3 = \overline{\alpha}_2, \quad \alpha_5 = \overline{\alpha}_4, \tag{72}$$

we have

$$h(z) = \frac{R(z)}{2\pi i} \left[ \oint_{\hat{\gamma}} \frac{f(\zeta)}{(\zeta - z) R(\zeta)}\, d\zeta + \oint_{\hat{\gamma}_m} \frac{W}{(\zeta - z) R(\zeta)}\, d\zeta + \oint_{\hat{\gamma}_c} \frac{\Omega}{(\zeta - z) R(\zeta)}\, d\zeta \right], \tag{73}$$

where $z$ is inside of $\hat{\gamma}$, $\hat{\gamma}_m$ is a loop around the main arc $\gamma_m = [\alpha_2, \alpha_4] \cup [\alpha_5, \alpha_3]$, and $\hat{\gamma}_c$ is a loop around the complementary arc $\gamma_c = [\alpha_0, \alpha_2] \cup [\alpha_3, \alpha_1]$ (see Figure 4). The real constants $W$ and $\Omega$ solve the system

$$\begin{cases} \displaystyle\oint_{\hat{\gamma}} \frac{f(\zeta)}{R(\zeta)}\, d\zeta + \Omega \oint_{\hat{\gamma}_c} \frac{d\zeta}{R(\zeta)} + W \oint_{\hat{\gamma}_m} \frac{d\zeta}{R(\zeta)} = 0, \\ \displaystyle\oint_{\hat{\gamma}} \frac{\zeta f(\zeta)}{R(\zeta)}\, d\zeta + \Omega \oint_{\hat{\gamma}_c} \frac{\zeta}{R(\zeta)}\, d\zeta + W \oint_{\hat{\gamma}_m} \frac{\zeta}{R(\zeta)}\, d\zeta = 0. \end{cases} \tag{74}$$

Other useful expressions written explicitly in the genus 2 region are

$$K(z)=\frac{1}{2\pi i}\begin{vmatrix}\oint_{\hat\gamma_m}\frac{d\zeta}{R(\zeta)} & \oint_{\hat\gamma_m}\frac{\zeta d\zeta}{R(\zeta)} & \oint_{\hat\gamma_m}\frac{d\zeta}{(\zeta-z)R(\zeta)}\\ \oint_{\hat\gamma_c}\frac{d\zeta}{R(\zeta)} & \oint_{\hat\gamma_c}\frac{\zeta d\zeta}{R(\zeta)} & \oint_{\hat\gamma_c}\frac{d\zeta}{(\zeta-z)R(\zeta)}\\ \oint_{\hat\gamma}\frac{f(\zeta)\,d\zeta}{R(\zeta)} & \oint_{\hat\gamma}\frac{\zeta f(\zeta)\,d\zeta}{R(\zeta)} & \oint_{\hat\gamma}\frac{f(\zeta)\,d\zeta}{(\zeta-z)R(\zeta)}\end{vmatrix} \tag{75}$$

and

$$K(z)=\frac{1}{2\pi i}\left[\oint_{\hat\gamma}\frac{f(\zeta)}{(\zeta-z)R(\zeta)}\,d\zeta+\oint_{\hat\gamma_m}\frac{W}{(\zeta-z)R(\zeta)}\,d\zeta+\oint_{\hat\gamma_c}\frac{\Omega}{(\zeta-z)R(\zeta)}\,d\zeta\right], \tag{76}$$

where $z$ is inside of $\hat\gamma$; and

$$\frac{\partial K}{\partial\mu}(\alpha_j,\vec\alpha,\mu)=\frac{1}{2\pi i}\begin{vmatrix}\oint_{\hat\gamma_m}\frac{d\zeta}{R(\zeta)} & \oint_{\hat\gamma_m}\frac{\zeta\, d\zeta}{R(\zeta)} & \oint_{\hat\gamma_m}\frac{d\zeta}{(\zeta-\alpha_j)R(\zeta)}\\ \oint_{\hat\gamma_c}\frac{d\zeta}{R(\zeta)} & \oint_{\hat\gamma_c}\frac{\zeta\, d\zeta}{R(\zeta)} & \oint_{\hat\gamma_c}\frac{d\zeta}{(\zeta-\alpha_j)R(\zeta)}\\ \oint_{\hat\gamma}\frac{f_\mu(\zeta)\,d\zeta}{R(\zeta)} & \oint_{\hat\gamma}\frac{\zeta f_\mu(\zeta)\,d\zeta}{R(\zeta)} & \oint_{\hat\gamma}\frac{f_\mu(\zeta)\,d\zeta}{(\zeta-\alpha_j)R(\zeta)}\end{vmatrix} \tag{77}$$

or

$$\frac{\partial K}{\partial\mu}(\alpha_j,\vec\alpha,\mu)=\frac{1}{2\pi i}\left[\oint_{\hat\gamma}\frac{f_\mu(\zeta)}{(\zeta-\alpha_j)R(\zeta)}\,d\zeta+\oint_{\hat\gamma_m}\frac{W_\mu}{(\zeta-\alpha_j)R(\zeta)}\,d\zeta+\oint_{\hat\gamma_c}\frac{\Omega_\mu}{(\zeta-\alpha_j)R(\zeta)}\,d\zeta\right], \tag{78}$$

where $f_\mu$ is given by (29). The real constants $W_\mu$ and $\Omega_\mu$ solve the system

$$\begin{cases}\oint_{\hat\gamma}\frac{f_\mu(\zeta)}{R(\zeta)}d\zeta+\Omega_\mu\oint_{\hat\gamma_c}\frac{d\zeta}{R(\zeta)}+W_\mu\oint_{\hat\gamma_m}\frac{d\zeta}{R(\zeta)}=0,\\ \oint_{\hat\gamma}\frac{\zeta f_\mu(\zeta)}{R(\zeta)}\,d\zeta+\Omega_\mu\oint_{\hat\gamma_c}\frac{\zeta}{R(\zeta)}\,d\zeta+W_\mu\oint_{\hat\gamma_m}\frac{\zeta}{R(\zeta)}\,d\zeta=0.\end{cases} \tag{79}$$

Also,

$$D=\begin{vmatrix}\oint_{\hat\gamma_m}\frac{d\zeta}{R(\zeta)} & \oint_{\hat\gamma_m}\frac{\zeta\, d\zeta}{R(\zeta)}\\ \oint_{\hat\gamma_c}\frac{d\zeta}{R(\zeta)} & \oint_{\hat\gamma_c}\frac{\zeta\, d\zeta}{R(\zeta)}\end{vmatrix}. \tag{80}$$

**Theorem A.2** ($\mu$-perturbation in genus 2). *Consider a finite-length non-self-intersecting contour $\gamma_0$ in the complex plane consisting of a union of oriented arcs $\gamma_0=\gamma_m\cup\gamma_c\cup[\alpha_0,\overline{\alpha}_0]$ with the distinct arc end points $\vec\alpha_0=(\alpha_0,\alpha_2,\alpha_4)$ in the upper half-plane and depending on a parameter $\mu$* (see Figure 4). *Assume $\vec\alpha_0$ and $\mu_0$ satisfy a system of equations*

$$\begin{cases} K(\alpha_0, \vec{\alpha}_0, \mu_0) = 0, \\ K(\alpha_2, \vec{\alpha}_0, \mu_0) = 0, \\ K(\alpha_4, \vec{\alpha}_0, \mu_0) = 0, \end{cases}$$

*and $f$ is given by* (21). *Let $\gamma = \gamma(\vec{\alpha}, \mu)$ be the contour of an RHP which seeks a function $h(z)$ which satisfies the conditions*

$$\begin{cases} h_+(z) + h_-(z) = 0 & \text{on } \gamma_{m,0} = [\alpha_0, \overline{\alpha}_0], \\ h_+(z) + h_-(z) = 2W & \text{on } \gamma_m, \\ h_+(z) - h_-(z) = 2\Omega & \text{on } \gamma_c, \\ h(z) + f(z) \text{ is analytic in } \overline{\mathbb{C}} \setminus \gamma, \end{cases} \tag{81}$$

*where $\Omega = \Omega(\vec{\alpha}, \mu)$ and $W = W(\vec{\alpha}, \mu)$ are real constants whose numerical values will be determined from the RH conditions. Assume that there is a function $h(z, \vec{\alpha}_0, \mu_0)$ which satisfies* (81) *and suppose $h'(z, \vec{\alpha}_0, \mu_0)/R(z, \vec{\alpha}_0) \neq 0$ for all $z$ on $\gamma_0$.*

*Then there is a contour $\gamma(\vec{\alpha}, \mu) \in \Gamma(\vec{\alpha}, \mu)$ such that the solution $\vec{\alpha} = \vec{\alpha}(\mu)$ of the system*

$$\begin{cases} K(\alpha_0, \vec{\alpha}, \mu) = 0, \\ K(\alpha_2, \vec{\alpha}, \mu) = 0, \\ K(\alpha_4, \vec{\alpha}, \mu) = 0 \end{cases} \tag{82}$$

*and $h(z, \vec{\alpha}(\mu), \mu)$ which solves* (81) *are uniquely defined and continuously differentiable in $\mu$ in some neighborhood of $\mu_0$.*

*Moreover,*

$$\begin{aligned} \frac{\partial \alpha_0}{\partial \mu}(x,t,\mu) &= -\frac{2\pi i \, \dfrac{\partial K}{\partial \mu}(\alpha_0, \vec{\alpha}, \mu)}{D \displaystyle\oint_{\hat{\gamma}(\mu)} \frac{f'(\zeta)}{(\zeta - \alpha_0) R(\zeta)} \, d\zeta}, \\ \frac{\partial \alpha_2}{\partial \mu}(x,t,\mu) &= -\frac{2\pi i \, \dfrac{\partial K}{\partial \mu}(\alpha_2, \vec{\alpha}, \mu)}{D \displaystyle\oint_{\hat{\gamma}(\mu)} \frac{f'(\zeta)}{(\zeta - \alpha_2) R(\zeta)} \, d\zeta}, \\ \frac{\partial \alpha_4}{\partial \mu}(x,t,\mu) &= -\frac{2\pi i \, \dfrac{\partial K}{\partial \mu}(\alpha_4, \vec{\alpha}, \mu)}{D \displaystyle\oint_{\hat{\gamma}(\mu)} \frac{f'(\zeta)}{(\zeta - \alpha_4) R(\zeta)} \, d\zeta}, \end{aligned} \tag{83}$$

$$\frac{\partial h}{\partial \mu}(z, x, t, \mu) = \frac{R(z)}{2\pi i} \oint_{\hat{\gamma}(\mu)} \frac{\dfrac{\partial f}{\partial \mu}(\zeta)}{(\zeta - z) R(\zeta)} \, d\zeta, \tag{84}$$

*where $z$ is inside of $\hat{\gamma}$.*

*Furthermore*, $\Omega(\mu) = \Omega(\vec{\alpha}(\mu), \mu)$ *and* $W(\mu) = W(\vec{\alpha}(\mu), \mu)$ *are defined and continuously differentiable in* $\mu$ *in some open neighborhood of* $\mu_0$, *and*

$$\frac{\partial \Omega}{\partial \mu}(x,t,\mu) = -\frac{1}{D}\begin{vmatrix} \oint_{\hat{\gamma}_m} \frac{d\zeta}{R(\zeta)} & \oint_{\hat{\gamma}_m} \frac{\zeta\, d\zeta}{R(\zeta)} \\ \oint_{\hat{\gamma}} \frac{f_\mu(\zeta)}{R(\zeta)} d\zeta & \oint_{\hat{\gamma}} \frac{\zeta f_\mu(\zeta)}{R(\zeta)} d\zeta \end{vmatrix}, \tag{85}$$

$$\frac{\partial W}{\partial \mu}(x,t,\mu) = -\frac{1}{D}\begin{vmatrix} \oint_{\hat{\gamma}} \frac{f_\mu(\zeta)}{R(\zeta)} d\zeta & \oint_{\hat{\gamma}} \frac{\zeta f_\mu(\zeta)}{R(\zeta)} d\zeta \\ \oint_{\hat{\gamma}_c} \frac{d\zeta}{R(\zeta)} & \oint_{\hat{\gamma}_c} \frac{\zeta\, d\zeta}{R(\zeta)} \end{vmatrix}, \tag{86}$$

*where* $\alpha_j = \alpha_j(x,t,\mu)$, $R(\zeta) = R(\zeta, \vec{\alpha}(x,t,\mu))$, $f(\zeta) = f(\zeta, x, t, \mu)$, $f_\mu(\zeta) = \dfrac{\partial f}{\partial \mu}(\zeta, x, t, \mu)$, *and* $D = D(\vec{\alpha}(x,t,\mu))$.

## References


[Baik et al. 1999] J. Baik, P. Deift, and K. Johansson, "On the distribution of the length of the longest increasing subsequence of random permutations", *J. Amer. Math. Soc.* **12**:4 (1999), 1119–1178. MR 2000e:05006 Zbl 0932.05001

[Belokolos et al. 1994] E. Belokolos, A. Bobenko, V. Enol'skij, A. Its, and V. Matveev, *Algebro-geometric approach to nonlinear integrable equations*, Springer, Berlin, 1994. Zbl 0809.35001

[Belov and Venakides 2015] S. Belov and S. Venakides, "Long-time limit studies of an obstruction in the $g$-function mechanism for semiclassical focusing NLS", preprint, 2015. arXiv 1504.02359

[Bertola and Tovbis 2013] M. Bertola and A. Tovbis, "Universality for the focusing nonlinear Schrödinger equation at the gradient catastrophe point: rational breathers and poles of the *tritronquée* solution to Painlevé I", *Comm. Pure Appl. Math.* **66**:5 (2013), 678–752. MR 3028484 Zbl 06155918

[Boutet de Monvel et al. 2011] A. Boutet de Monvel, V. P. Kotlyarov, and D. Shepelsky, "Focusing NLS equation: long-time dynamics of step-like initial data", *Int. Math. Res. Not.* **2011**:7 (2011), 1613–1653. MR 2012f:35498 Zbl 1220.35166

[Buckingham and Venakides 2007] R. Buckingham and S. Venakides, "Long-time asymptotics of the nonlinear Schrödinger equation shock problem", *Comm. Pure Appl. Math.* **60**:9 (2007), 1349–1414. MR 2008j:35167 Zbl 1125.35089

[Cai et al. 2002] D. Cai, D. W. McLaughlin, and K. D. T.-R. McLaughlin, "The nonlinear Schrödinger equation as both a PDE and a dynamical system", pp. 599–675 in *Handbook of dynamical systems*, vol. 2, edited by B. Fiedler, North-Holland, Amsterdam, 2002. MR 2004e:37002 Zbl 1004.35111

[Ceniceros and Tian 2002] H. D. Ceniceros and F.-R. Tian, "A numerical study of the semi-classical limit of the focusing nonlinear Schrödinger equation", *Phys. Lett. A* **306**:1 (2002), 25–34. MR 2004a:78001 Zbl 1005.81029

[Deift 1999] P. Deift, *Orthogonal polynomials and random matrices: a Riemann–Hilbert approach*, Courant Lecture Notes in Mathematics **3**, Courant Institute of Mathematical Sciences, New York, 1999. MR 2000g:47048 Zbl 0997.47033

[Deift and Zhou 1993] P. Deift and X. Zhou, "A steepest descent method for oscillatory Riemann–Hilbert problems: Asymptotics for the MKdV equation", *Ann. of Math.* (2) **137**:2 (1993), 295–368. MR 94d:35143 Zbl 0771.35042

[Deift and Zhou 1995] P. Deift and X. Zhou, "Asymptotics for the Painlevé II equation", *Comm. Pure Appl. Math.* **48**:3 (1995), 277–337. MR 96d:34004 Zbl 0869.34047

[Deift et al. 1994] P. Deift, S. Venakides, and X. Zhou, "The collisionless shock region for the long-time behavior of solutions of the KdV equation", *Comm. Pure Appl. Math.* **47**:2 (1994), 199–206. MR 95f:35220 Zbl 0797.35143

[Deift et al. 1997] P. Deift, S. Venakides, and X. Zhou, "New results in small dispersion KdV by an extension of the steepest descent method for Riemann–Hilbert problems", *Internat. Math. Res. Notices* 6 (1997), 286–299. MR 98b:35155 Zbl 0873.65111

[Deift et al. 1999a] P. Deift, T. Kriecherbauer, K. D. T.-R. McLaughlin, S. Venakides, and X. Zhou, "Strong asymptotics of orthogonal polynomials with respect to exponential weights", *Comm. Pure Appl. Math.* **52**:12 (1999), 1491–1552. MR 2001f:42037 Zbl 1026.42024

[Deift et al. 1999b] P. Deift, T. Kriecherbauer, K. D. T.-R. McLaughlin, S. Venakides, and X. Zhou, "Uniform asymptotics for polynomials orthogonal with respect to varying exponential weights and applications to universality questions in random matrix theory", *Comm. Pure Appl. Math.* **52**:11 (1999), 1335–1425. MR 2001g:42050 Zbl 0944.42013

[Duits and Kuijlaars 2009] M. Duits and A. B. J. Kuijlaars, "Universality in the two-matrix model: a Riemann–Hilbert steepest-descent analysis", *Comm. Pure Appl. Math.* **62**:8 (2009), 1076–1153. MR 2010f:82046 Zbl 1221.15052

[Ercolani and McLaughlin 2003] N. M. Ercolani and K. D. T.-R. McLaughlin, "Asymptotics of the partition function for random matrices via Riemann–Hilbert techniques and applications to graphical enumeration", *Int. Math. Res. Not.* **2003**:14 (2003), 755–820. MR 2005f:82048 Zbl 1140.82307

[Forest and Lee 1986] M. G. Forest and J. E. Lee, "Geometry and modulation theory for the periodic nonlinear Schrödinger equation", pp. 35–69 in *Oscillation theory, computation, and methods of compensated compactness* (Minneapolis, 1985), edited by C. Dafermos et al., IMA Vol. Math. Appl. **2**, Springer, New York, 1986. MR 869821 Zbl 0615.35003

[Kamvissis and Rakhmanov 2005] S. Kamvissis and E. A. Rakhmanov, "Existence and regularity for an energy maximization problem in two dimensions", *J. Math. Phys.* **46**:8 (2005), 083505, 24. MR 2006f:81069 Zbl 1110.81083

[Kamvissis et al. 2003] S. Kamvissis, K. D. T.-R. McLaughlin, and P. D. Miller, *Semiclassical soliton ensembles for the focusing nonlinear Schrödinger equation*, Annals of Mathematics Studies **154**, Princeton Univ. Press, 2003. MR 2004h:37115 Zbl 1057.35063

[Klaus and Shaw 2002] M. Klaus and J. K. Shaw, "Purely imaginary eigenvalues of Zakharov–Shabat systems", *Phys. Rev. E* (3) **65**:3 (2002), 036607. MR 2003c:34144

[Kuijlaars and McLaughlin 2000] A. B. J. Kuijlaars and K. D. T.-R. McLaughlin, "Generic behavior of the density of states in random matrix theory and equilibrium problems in the presence of real analytic external fields", *Comm. Pure Appl. Math.* **53**:6 (2000), 736–785. MR 2001f:31003 Zbl 1022.31001

[Lax and Levermore 1983a] P. D. Lax and C. D. Levermore, "The small dispersion limit of the Korteweg–de Vries equation, I", *Comm. Pure Appl. Math.* **36**:3 (1983), 253–290. MR 85g:35105a Zbl 0532.35067

[Lax and Levermore 1983b] P. D. Lax and C. D. Levermore, "The small dispersion limit of the Korteweg–de Vries equation, II", *Comm. Pure Appl. Math.* **36**:5 (1983), 571–593. MR 85g:35105b Zbl 0527.35073

[Lax and Levermore 1983c] P. D. Lax and C. D. Levermore, "The small dispersion limit of the Korteweg–de Vries equation, III", *Comm. Pure Appl. Math.* **36**:6 (1983), 809–829. MR 85g:35105c Zbl 0527.35074

[Lyng and Miller 2007] G. D. Lyng and P. D. Miller, "The $N$-soliton of the focusing nonlinear Schrödinger equation for $N$ large", *Comm. Pure Appl. Math.* **60**:7 (2007), 951–1026. MR 2008k:35454 Zbl 1185.35259

[Miller and Kamvissis 1998] P. D. Miller and S. Kamvissis, "On the semiclassical limit of the focusing nonlinear Schrödinger equation", *Phys. Lett. A* **247**:1-2 (1998), 75–86. MR 99e:81058 Zbl 0941.81029

[Satsuma and Yajima 1974] J. Satsuma and N. Yajima, "Initial value problems of one-dimensional self-modulation of nonlinear waves in dispersive media", *Progr. Theoret. Phys. Suppl. No. 55* (1974), 284–306. MR 57 #3675

[Shabat 1976] A. B. Shabat, "One-dimensional perturbations of a differential operator, and the inverse scattering problem", pp. 279–296 in Задачи механики и математической физики, посв. памяти И. Г. Петровского, edited by S. L. Sobolev and A. I. Suslov, Nauka, Moscow, 1976. In Russian; translated in *Sel. Math. Sov.* **4** (1975), 19–35. MR 56 #16043 Zbl 0569.34026

[Simon 2011] B. Simon, *Szegő's theorem and its descendants*, Princeton Univ. Press, 2011. MR 2012b:47080 Zbl 1230.33001

[Tovbis and Venakides 2000] A. Tovbis and S. Venakides, "The eigenvalue problem for the focusing nonlinear Schrödinger equation: new solvable cases", *Phys. D* **146**:1-4 (2000), 150–164. MR 2001i:35266 Zbl 0984.34072

[Tovbis and Venakides 2009] A. Tovbis and S. Venakides, "Determinant form of the complex phase function of the steepest descent analysis of Riemann–Hilbert problems and its application to the focusing nonlinear Schrödinger equation", *Int. Math. Res. Not.* **2009**:11 (2009), 2056–2080. MR 2010e:35271 Zbl 1175.30039

[Tovbis and Venakides 2010] A. Tovbis and S. Venakides, “Nonlinear steepest descent asymptotics for semiclassical limit of integrable systems: continuation in the parameter space”, *Comm. Math. Phys.* **295**:1 (2010), 139–160. MR 2011b:35494 Zbl 1194.35429

[Tovbis et al. 2004] A. Tovbis, S. Venakides, and X. Zhou, “On semiclassical (zero dispersion limit) solutions of the focusing nonlinear Schrödinger equation”, *Comm. Pure Appl. Math.* **57**:7 (2004), 877–985. MR 2005c:35269 Zbl 1060.35137

[Tovbis et al. 2006] A. Tovbis, S. Venakides, and X. Zhou, “On the long-time limit of semiclassical (zero dispersion limit) solutions of the focusing nonlinear Schrödinger equation: pure radiation case”, *Comm. Pure Appl. Math.* **59**:10 (2006), 1379–1432. MR 2007e:37041 Zbl 1115.35127

[Tovbis et al. 2007] A. Tovbis, S. Venakides, and X. Zhou, “Semiclassical focusing nonlinear Schrödinger equation, I: inverse scattering map and its evolution for radiative initial data”, *Int. Math. Res. Not.* **2007**:22 (2007), Art. ID rnm094, 54. MR 2009a:37146 Zbl 1142.35089

[Venakides 1985] S. Venakides, “The zero dispersion limit of the Korteweg–de Vries equation for initial potentials with nontrivial reflection coefficient”, *Comm. Pure Appl. Math.* **38**:2 (1985), 125–155. MR 87d:35129 Zbl 0571.35095

[Zakharov and Shabat 1972] V. E. Zakharov and A. B. Shabat, “Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media”, *J. Exper. Teoret. Phys.* **34**:1 (1972), 62–69. MR 53 #9966

Received 29 Nov 2012. Revised 13 Jun 2014. Accepted 9 Jan 2015.

SERGEY BELOV: sbelov@gmail.com
*Department of Mathematics, University of Houston, 4800 Calhoun Rd., Houston, TX 77204-2008, United States*

STEPHANOS VENAKIDES: ven@math.duke.edu
*Department of Mathematics, Duke University, Box 90320, Durham, NC 27708, United States*

ANALYSIS AND PDE
Vol. 8, No. 2, 2015
dx.doi.org/10.2140/apde.2015.8.289

# TUNNEL EFFECT FOR SEMICLASSICAL RANDOM WALKS

JEAN-FRANÇOIS BONY, FRÉDÉRIC HÉRAU AND LAURENT MICHEL

We study a semiclassical random walk with respect to a probability measure with a finite number $n_0$ of wells. We show that the associated operator has exactly $n_0$ eigenvalues exponentially close to 1 (in the semiclassical sense), and that the others are $\mathcal{O}(h)$ away from 1. We also give an asymptotic of these small eigenvalues. The key ingredient in our approach is a general factorization result of pseudodifferential operators, which allows us to use recent results on the Witten Laplacian.

## 1. Introduction

Let $\phi : \mathbb{R}^d \to \mathbb{R}$ be a smooth function and let $h \in ]0, 1]$ denote a small parameter throughout. Under suitable assumptions specified later, the density $e^{-\phi(x)/h}$ is integrable and there exists $Z_h > 0$ such that $d\mu_h(x) = Z_h e^{-\phi(x)/h} dx$ defines a probability measure on $\mathbb{R}^d$. We can associate to $\mu_h$ the Markov kernel $t_h(x, dy)$ given by

$$t_h(x, dy) = \frac{1}{\mu_h(B(x, h))} \mathbb{1}_{|x-y|<h} \, d\mu_h(y). \tag{1-1}$$

From the point of view of random walks, this kernel can be understood as follows: Assume that at step $n$ the walk is in $x_n$; then the point $x_{n+1}$ is chosen in the small ball $B(x_n, h)$ uniformly at random with respect to $d\mu_h$. The probability distribution at time $n \in \mathbb{N}$ of a walk starting from $x$ is given by the kernel $t_h^n(x, dy)$. The long-time behavior ($n \to \infty$) of the kernel $t_h^n(x, dy)$ carries information on the ergodicity of the random walk, and has many practical applications (we refer to [Lelièvre et al. 2010] for an overview of computational aspects). Observe that, if $\phi$ is a Morse function, then the density $e^{-\phi/h}$ concentrates at scale $\sqrt{h}$ around minima of $\phi$, whereas the moves of the random walk are at scale $h$.

Another point of view comes from statistical physics and can be described as follows: One can associate to the kernel $t_h(x, dy)$ an operator $\boldsymbol{T}_h$ acting on the space $C_0$ of continuous functions going to zero at infinity by the formula

$$\boldsymbol{T}_h f(x) = \int_{\mathbb{R}^d} f(y) t_h(x, dy) = \frac{1}{\mu_h(B(x, h))} \int_{|x-y|<h} f(y) \, d\mu_h(y).$$

This defines a bounded operator on $C_0$, enjoying the Markov property ($\boldsymbol{T}_h(1) = 1$).

Bony and Hérau are supported by the ANR project NOSEVOL, ANR 2011 BS01019 01. Michel is a member of the ERC project Semiclassical Analysis of Partial Differential Equations, ERC-2012-ADG, project number 320845.
*MSC2010:* 35S05, 35P15, 47A10, 60J05.
*Keywords:* analysis of PDEs, probability, spectral theory.

The transpose $\boldsymbol{T}_h^\star$ of $\boldsymbol{T}_h$ is defined by duality on the set of bounded positive measures $\boldsymbol{M}_b^+$ (resp. bounded measures $\boldsymbol{M}_b$). If $d\nu$ is a bounded measure, we have

$$\boldsymbol{T}_h^\star(d\nu) = \left( \int_{\mathbb{R}^d} \mathbb{1}_{|x-y|<h} \mu_h(B(y,h))^{-1} \, d\nu(y) \right) d\mu_h. \tag{1-2}$$

Assume that a particle in $\mathbb{R}^d$ is distributed according to a probability measure $d\nu$; then $\boldsymbol{T}_h^\star(d\nu)$ represents its distribution after a move according to $t_h(x, dy)$, and the distribution after $n$ steps is then given by $(\boldsymbol{T}_h^\star)^n(d\nu)$. The existence of a limit distribution is strongly related to the existence of an invariant measure. In the present context, one can easily see that $\boldsymbol{T}_h^\star$ admits the invariant measure

$$d\nu_{h,\infty}(x) = \widetilde{Z}_h \mu_h(B(x,h)) \, d\mu_h(x),$$

where $\widetilde{Z}_h$ is chosen so that $d\nu_{h,\infty}$ is a probability. The aim of the present paper will be to prove the convergence of $(\boldsymbol{T}_h^\star)^n(d\nu)$ towards $d\nu_{h,\infty}$ when $n$ goes to infinity for any probability measure $d\nu$, and to get precise information on the speed of convergence. Taking $d\nu(y) = \delta_x(y)$, it turns out that it is equivalent to study the convergence of $t_h^n(x, dy)$ towards $d\nu_{h,\infty}$. Note that, in the present setting, proving pointwise convergence ($h$ being fixed) of $t_h^n(x, dy)$ towards the invariant measure is an easy consequence of a general theorem (see [Feller 1971, Theorem 2, p. 272]). The purpose of our approach is to get convergence in a stronger topology and to obtain precise information on the behavior with respect to the semiclassical parameter $h$.

Before going further, let us recall some elementary properties of $\boldsymbol{T}_h$ that will be useful in the sequel. First, we can see easily from its definition that the operator $\boldsymbol{T}_h$ can be extended as a bounded operator both on $L^\infty(d\nu_{h,\infty})$ and $L^1(d\nu_{h,\infty})$. From the Markov property and the fact that $d\nu_{h,\infty}$ is stationary, it is clear that

$$\|\boldsymbol{T}_h\|_{L^\infty(d\nu_{h,\infty}) \to L^\infty(d\nu_{h,\infty})} = \|\boldsymbol{T}_h\|_{L^1(d\nu_{h,\infty}) \to L^1(d\nu_{h,\infty})} = 1.$$

Hence, by interpolation, $\boldsymbol{T}_h$ defines also a bounded operator of norm 1 on $L^2(\mathbb{R}^d, d\nu_{h,\infty})$. Finally, observe that $\boldsymbol{T}_h$ is selfadjoint on $L^2(d\nu_{h,\infty})$ (thanks again to the Markov property).

Let us go back to the study of the sequence $(\boldsymbol{T}_h^\star)^n$ and explain the topology we use to study the convergence of this sequence. Instead of looking at this evolution on the full set of bounded measures, we restrict the analysis by introducing the stable Hilbert space

$$\mathcal{H}_h = L^2(d\nu_{h,\infty}) = \left\{ f \text{ measurable on } \mathbb{R}^d \text{ such that } \int |f(x)|^2 \, d\nu_{h,\infty} < \infty \right\}, \tag{1-3}$$

for which we have a natural injection with norm 1, $\mathcal{J} : \mathcal{H}_h \hookrightarrow \boldsymbol{M}_b$, when identifying an absolutely continuous measure $d\nu_h = f(x) d\nu_{h,\infty}$ with its density $f$. Using (1-2), we can see easily that $\boldsymbol{T}_h^\star \circ \mathcal{J} = \mathcal{J} \circ \boldsymbol{T}_h$. From this identification, $\boldsymbol{T}_h^\star$ (acting on $\mathcal{H}_h$) inherits the properties of $\boldsymbol{T}_h$:

$$\boldsymbol{T}_h^\star : \mathcal{H}_h \to \mathcal{H}_h \text{ is selfadjoint and continuous with operator norm 1.} \tag{1-4}$$

Hence, its spectrum is contained in the interval $[-1, 1]$. Moreover, we will see later that $-1$ is sufficiently far from the spectrum. Since we are interested in the convergence of $(\boldsymbol{T}_h^\star)^n$ in the $L^2$ topology, it is then sufficient for our purpose to give a precise description of the spectrum of $\boldsymbol{T}_h$ near 1.

Convergence of Markov chains to stationary distributions is a wide area of research with many applications. Knowing that a computable Markov kernel converges to a given distribution may be very useful in practice. In particular, it is often used to sample a given probability in order to implement Monte Carlo methods (see [Lelièvre et al. 2010] for numerous algorithms and computational aspects). However, most results giving a priori bounds on the speed of convergence for such algorithms hold for discrete state space (we refer to [Diaconis 2009] for a state of the art on Monte Carlo Markov chain methods).

This point of view is also used to track extremal points of any function by simulated annealing procedure. For example, this was used in [Holley and Stroock 1988] on finite state space and in [Holley et al. 1989; Miclo 1992] on continuous state space.

Relatedly, let us recall that the study of time-continuous processes is of current interest in statistical physics (see for instance the work of Bovier, Eckhoff, Gayrard and Klein [Bovier et al. 2004; 2005] on metastable states).

More recently, Diaconis and Lebeau [2009] obtained first results on discrete time processes on continuous state space. This approach was then further developed in [Diaconis et al. 2011] to get convergence results on the Metropolis algorithm on bounded domains of Euclidean space. Similar results were also obtained in [Lebeau and Michel 2010; Guillarmou and Michel 2011] in various geometric situations. In all these papers, the probability $d\mu_h$ is independent of $h$, which leads ultimately to a spectral gap of order $h^2$. Here, the situation is quite different and somehow "more semiclassical". This permits us to exhibit situations with very small spectral gap of order $e^{-c/h}$. The precise asymptotic of this gap (and more generally of the eigenvalues close to 1) is driven by the tunnel effect between wells (see [Helffer and Sjöstrand 1984] for results in the case of Schrödinger operators). In this paper, we shall compute accurately this quantity under the following assumptions on $\phi$:

**Hypothesis 1.** *We suppose that $\phi$ is a Morse function with nondegenerate critical points and that there exist $c$, $R > 0$ and some constants $C_\alpha > 0$, $\alpha \in \mathbb{N}^d$ such that, for all $|x| \geq R$, we have*

$$|\partial_x^\alpha \phi(x)| \leq C_\alpha, \quad |\nabla\phi(x)| \geq c \quad \text{and} \quad \phi(x) \geq c|x| \qquad \text{for all } \alpha \in \mathbb{N}^d \setminus \{0\}.$$

*In particular, there is a finite number of critical points.*

Observe that functions $\phi$ satisfying this assumption are at most linear at infinity. It may be possible to relax this assumption to quadratic growth at infinity, and we guess our results hold true also in this context. However, it doesn't seem possible to get a complete proof with the class of symbols used in this paper.

Under the above assumption, it is clear that $d\mu_h(x) = Z_h e^{-\phi(x)/h}\,dx$ is a probability measure. For the following, we call $\mathcal{U}$ the set of critical points $\boldsymbol{u}$. We denote by $\mathcal{U}^{(0)}$ the set of minima of $\phi$ and by $\mathcal{U}^{(1)}$ the set of saddle points, i.e., the critical points with index 1 (note that this set may be empty). We also introduce $n_j = \sharp\,\mathcal{U}^{(j)}$, $j = 0, 1$, the number of elements of $\mathcal{U}^{(j)}$.

We shall first prove the following result:

**Theorem 1.1.** *There exist $\delta$, $h_0 > 0$ such that the following assertions hold true for $h \in\, ]0, h_0]$: First, $\sigma(\boldsymbol{T}_h^\star) \subset [-1+\delta, 1]$ and $\sigma_{\text{ess}}(\boldsymbol{T}_h^\star) \subset [-1+\delta, 1-\delta]$. Moreover, $\boldsymbol{T}_h^\star$ has exactly $n_0$ eigenvalues in $[1-\delta h, 1]$, which are in fact in $[1-e^{-\delta/h}, 1]$. Lastly, 1 is a simple eigenvalue for the eigenstate $\nu_{h,\infty} \in \mathcal{H}_h$.*

This theorem will be proved in the next section. The goal of this paper is to describe accurately the eigenvalues close to 1. We will see later that describing the eigenvalues of $\boldsymbol{T}_h^\star$ close to 1 has many common points with the spectral study of the so-called semiclassical Witten Laplacian (see Section 4). We introduce the following generic assumptions on the critical points of $\phi$:

**Hypothesis 2.** *We suppose that the values $\phi(\boldsymbol{s}) - \phi(\boldsymbol{m})$ are distinct for any $\boldsymbol{s} \in \mathcal{U}^{(1)}$ and $\boldsymbol{m} \in \mathcal{U}^{(0)}$.*

Note that this generic assumption could easily be relaxed at the cost of messy notation and less precise statements, following, e.g., [Hérau et al. 2011], and that we chose to focus in this article on other particularities of the problem.

Let us recall that, under the above assumptions, there exists a labeling of minima and saddle points, $\mathcal{U}^{(0)} = \{\boldsymbol{m}_k : k = 1, \dots, n_0\}$ and $\mathcal{U}^{(1)} = \{\boldsymbol{s}_j : j = 2, \dots, n_1 + 1\}$, which permits us to describe the low-lying eigenvalues of the Witten Laplacian (see [Helffer et al. 2004; Hérau et al. 2011], for instance). Observe that the enumeration of $\mathcal{U}^{(1)}$ starts with $j = 2$, since we will need a fictional saddle point $\boldsymbol{s}_1 = +\infty$. We shall recall this labeling procedure in the Appendix.

Let us denote by $1 = \lambda_1^\star(h) > \lambda_2^\star(h) \geq \dots \geq \lambda_{n_0}^\star(h)$ the $n_0$ largest eigenvalues of $\boldsymbol{T}_h^\star$. The main result of this paper is the following:

**Theorem 1.2.** *Under Hypotheses 1 and 2, there exists a labeling of minima and saddle points and constants $\alpha, h_0 > 0$ such that, for all $k = 2, \dots, n_0$ and for any $h \in ]0, h_0]$,*

$$1 - \lambda_k^\star(h) = \frac{h}{(2d+4)\pi} \mu_k \sqrt{\left| \frac{\det \phi''(\boldsymbol{m}_k)}{\det \phi''(\boldsymbol{s}_k)} \right|} e^{-2S_k/h} (1 + \mathcal{O}(h)),$$

*where $S_k := \phi(\boldsymbol{s}_k) - \phi(\boldsymbol{m}_k)$ (the Arrhenius number) and $-\mu_k$ denotes the unique negative eigenvalue of $\phi''$ at $\boldsymbol{s}_k$.*

**Remark 1.3.** The leading term in the asymptotic of $1 - \lambda_k^\star(h)$ above is exactly (up to the factor $(2d+4)$) the one of the $k$-th eigenvalue of the Witten Laplacian on the 0-forms obtained in [Helffer et al. 2004]. This relationship will be transparent from the proof below.

As an immediate consequence of these results and of the spectral theorem, we get that the convergence to equilibrium holds slowly and that the system has a metastable regime. More precisely, we have the following result, whose proof can be found at the end of Section 5.

**Corollary 1.4.** *Let $d\nu_h$ be a probability measure in $\mathcal{H}_h$ and assume first that $\phi$ has a unique minimum. Then, using that $\sigma(\boldsymbol{T}_h^\star) \subset [-1 + \delta, 1 - \delta h]$, it yields*

$$\|(\boldsymbol{T}_h^\star)^n (d\nu_h) - d\nu_{h,\infty}\|_{\mathcal{H}_h} = \mathcal{O}(h) \|d\nu_h\|_{\mathcal{H}_h} \tag{1-5}$$

*for all $n \gtrsim |\ln h| h^{-1}$, which corresponds to the Ehrenfest time. But, if $\phi$ has several minima, we can write*

$$(\boldsymbol{T}_h^\star)^n (d\nu_h) = \Pi d\nu_h + \mathcal{O}(h) \|d\nu_h\|_{\mathcal{H}_h} \tag{1-6}$$

*for all $h^{-1} |\ln h| \lesssim n \lesssim e^{2S_{n_0}/h}$. Here, $\Pi$ can be taken as the orthogonal projector on the $n_0$ functions $\chi_k(x) e^{-(\phi(x) - \phi(\boldsymbol{m}_k))/h}$, where $\chi_k$ is any cutoff function near $\boldsymbol{m}_k$.*

*On the other hand, we have, for any* $n \in \mathbb{N}$,

$$\|(\boldsymbol{T}_h^\star)^n(d\nu_h) - d\nu_{h,\infty}\|_{\mathscr{H}_h} \le (\lambda_2^\star(h))^n \|d\nu_h\|_{\mathscr{H}_h}, \tag{1-7}$$

*where* $\lambda_2^\star(h)$ *is described in Theorem 1.2. Note that this inequality is optimal. In particular, for* $n \gtrsim |\ln h| h^{-1} e^{2S_2/h}$, *the right-hand side of* (1-7) *is of order* $\mathbb{O}(h)\|d\nu_h\|_{\mathscr{H}_h}$.

*Thus, for a reasonable number of iterations* (*which guarantees* (1-5)), 1 *seems to be an eigenvalue of multiplicity* $n_0$; *whereas, for a very large number of iterations, the system returns to equilibrium. Then,* (1-6) *is a metastable regime.*

Since $t_h(x, dy)$ is absolutely continuous with respect to $d\nu_{h,\infty}$, then $(\boldsymbol{T}_h^\star)^n(\delta_{y=x}) = t_h^n(x, dy)$ belongs to $\mathscr{H}_h$ for any $n \ge 1$. Hence, the above estimate and the fact that $d\nu_{h,\infty}$ is invariant show that

$$\|t_h^n(x, dy) - d\nu_{h,\infty}\|_{\mathscr{H}_h} \le (\lambda_2^\star(h))^{n-1} \|t_h(x, dy)\|_{\mathscr{H}_h}.$$

Moreover, the prefactor $\|t_h(x, dy)\|_{\mathscr{H}_h}$ could be easily computed but depends on $x$ and $h$.

Throughout this paper, we use semiclassical analysis (see [Dimassi and Sjöstrand 1999; Martinez 2002; Zworski 2012] for expository books on this theory). Let us recall that a function $m : \mathbb{R}^d \to \mathbb{R}^+$ is an order function if there exists $N_0 \in \mathbb{N}$ and a constant $C > 0$ such that, for all $x, y \in \mathbb{R}^d$, $m(x) \le C\langle x-y\rangle^{N_0} m(y)$. Here and throughout we use the notation $\langle x\rangle = (1+|x|^2)^{\frac{1}{2}}$. This definition can be extended to functions $m : \mathbb{R}^d \times \mathbb{C}^{d'} \to \mathbb{R}^+$ by identifying $\mathbb{R}^d \times \mathbb{C}^{d'}$ with $\mathbb{R}^{d+2d'}$. Given an order function $m$ on $T^*\mathbb{R}^d \simeq \mathbb{R}^{2d}$, we will denote by $S^0(m)$ the space of semiclassical functions on $T^*\mathbb{R}^d$ whose derivatives are all bounded by $m$, and by $\Psi^0(m)$ the set of corresponding pseudodifferential operators. For any $\tau \in\, ]0, \infty]$ and any order function $m$ on $\mathbb{R}^d \times \mathbb{C}^d$, we will denote by $S^0_\tau(m)$ the set of symbols which are analytic with respect to $\xi$ in the strip $|\operatorname{Im} \xi| < \tau$ and bounded by some constant times $m(x, \xi)$ in this strip. We will denote by $S^0_\infty(m)$ the union over $\tau > 0$ of $S^0_\tau(m)$. We denote by $\Psi^0_\tau(m)$ the set of corresponding operators. Lastly, we say that a symbol $p$ is classical if it admits an asymptotic expansion $p(x, \xi; h) \sim \sum_{j\ge 0} h^j p_j(x, \xi)$. We will denote by $S^0_{\tau,\mathrm{cl}}(m)$ and $S^0_{\mathrm{cl}}(m)$ the corresponding classes of symbols.

We will also need some matrix-valued pseudodifferential operators. Let $\mathcal{M}_{p,q}$ denote the set of real-valued matrices with $p$ rows and $q$ columns, and $\mathcal{M}_p = \mathcal{M}_{p,p}$. Let $\mathscr{A} : T^*\mathbb{R}^d \to \mathcal{M}_{p,q}$ be a smooth function. We will say that $\mathscr{A}$ is a $(p, q)$-matrix weight if $\mathscr{A}(x, \xi) = (a_{i,j}(x, \xi))_{i,j}$ and $a_{i,j}$ is an order function for every $i = 1, \dots, p$ and $j = 1, \dots, q$. If $p = q$, we will simply say that $\mathscr{A}$ is a $q$-matrix weight.

Given a $(p, q)$-matrix weight $\mathscr{A}$, we will denote by $S^0(\mathscr{A})$ the set of symbols $p(x, \xi) = (p_{i,j}(x, \xi))_{i,j}$ defined on $T^*\mathbb{R}^d$ with values in $\mathcal{M}_{p,q}$ such that, for all $i, j$, $p_{i,j} \in S^0(a_{i,j})$, and by $\Psi^0(\mathcal{M}_{p,q})$ the set of corresponding pseudodifferential operators. Obvious extensions of these definitions leads to the definition of matrix-valued symbols analytic w.r.t. to $\xi$ and the corresponding operators, $S^0_\tau(\mathscr{A})$ and $\Psi^0_\tau(\mathscr{A})$. In the following, we shall mainly use the Weyl semiclassical quantization of symbols, defined by

$$\operatorname{Op}(p)u(x) = (2\pi h)^{-d} \int_{T^*\mathbb{R}^d} e^{ih^{-1}(x-y)\xi} p\big(\tfrac{1}{2}(x+y), \xi\big) u(y)\, dy\, d\xi \tag{1-8}$$

for $p \in S^0(\mathscr{A})$. We shall also use the following notations. Given two pseudodifferential operators $A$ and $B$, we shall write $A = B + \Psi^k(m)$ if the difference $A - B$ belongs to $\Psi^k(m)$. At the level of symbols, we shall write $a = b + S^k(m)$ instead of $a - b \in S^k(m)$.

The preceding theorem is close — in the spirit and in the proof — to the ones given for the Witten Laplacian in [Helffer et al. 2004] and for the Kramers–Fokker–Planck operators in [Hérau et al. 2011]. In those works, the results are deeply linked with some properties inherited from a so-called supersymmetric structure, which allow the operators to be written as twisted Hodge Laplacians of the form

$$P = d^*_{\phi,h} A d_{\phi,h},$$

where $\boldsymbol{d}$ is the usual differential, $d_{\phi,h} = h\boldsymbol{d} + d\phi(x)\wedge = e^{-\phi/h} h\boldsymbol{d}\, e^{\phi/h}$ is the differential twisted by $\phi$, and $A$ is a constant matrix in $\mathcal{M}_d$. Here we are able to recover a supersymmetric-type structure, and the main ingredients for the study of the exponentially small eigenvalues are therefore available. This is contained in the following theorem, that we give in rather general context since it may be useful in other situations.

Let us introduce the $d$-matrix weights $\Xi$, $\mathscr{A} : T^*\mathbb{R}^d \to \mathcal{M}_d$ given by $\mathscr{A}_{i,j}(x,\xi) = (\langle \xi_i \rangle \langle \xi_j \rangle)^{-1}$, $\Xi_{i,j} = \delta_{i,j} \langle \xi_i \rangle$, and observe that $(\Xi \mathscr{A})_{i,j} = \langle \xi_j \rangle^{-1}$. In the following theorem, we state an exact factorization result, which will be the key point in our approach.

**Theorem 1.5.** *Let $p(x,\xi;h) \in S^0_\infty(1)$ and let $P_h = \operatorname{Op}(p)$. Suppose that $p(x,\xi;h) = p_0(x,\xi) + S^0_\infty(h)$ and that, for all $(x,\xi) \in \mathbb{R}^{2d}$, $p(x,\xi;h)$ is real. Let $\phi$ satisfy Hypotheses 1 and 2 and assume that the following assumptions hold true*:

(i) $P_h(e^{-\phi/h}) = 0$;

(ii) *for all $x \in \mathbb{R}^d$, the function $\xi \in \mathbb{R}^d \mapsto p(x,\xi;h)$ is even*;

(iii) *for all $\delta > 0$, there exists $\alpha > 0$ such that, for all $(x,\xi) \in T^*\mathbb{R}^d$, $d(x,\mathcal{U})^2 + |\xi|^2 \geq \delta$ implies $p_0(x,\xi) \geq \alpha$*;

(iv) *for any critical point $\boldsymbol{u} \in \mathcal{U}$, we have*

$$p_0(x,\xi) = |\xi|^2 + |\nabla\phi(x)|^2 + r(x,\xi)$$

*with $r(x,\xi) = \mathcal{O}(|(x-\boldsymbol{u},\xi)|^3)$ near $(\boldsymbol{u},0)$.*

*Then, for $h > 0$ small enough, there exists a symbol $q \in S^0(\Xi\mathscr{A})$ satisfying the following properties*:

*First, $P_h = d^*_{\phi,h} Q^* Q d_{\phi,h}$ with $Q = \operatorname{Op}(q)$.*

*Next, $q(x,\xi;h) = q_0(x,\xi) + S^0(h\Xi\mathscr{A})$ and, for any critical point $\boldsymbol{u} \in \mathcal{U}$, we have*

$$q_0(x,\xi) = \operatorname{Id} + \mathcal{O}(|(x-\boldsymbol{u},\xi)|).$$

*If we assume additionally that $r(x,\xi) = \mathcal{O}(|(x-\boldsymbol{u},\xi)|^4)$, then $q_0(x,\xi) = \operatorname{Id} + \mathcal{O}(|(x-\boldsymbol{u},\xi)|^2)$ near $(\boldsymbol{u},0)$ for any critical point $\boldsymbol{u} \in \mathcal{U}$.*

*Lastly, if $p \in S^0_{\mathrm{cl}}(1)$ then $q \in S^0_{\mathrm{cl}}(\Xi\mathscr{A})$.*

As already mentioned, we decided in this paper not to give results in the most general case so that technical aspects do not hide the main ideas. Nevertheless, we would like to mention here some possible generalizations of the preceding result.

First, it should certainly be possible to use more general order functions and to prove factorization results for symbols in other classes (for instance $S^0(\langle (x,\xi)\rangle^2)$). This should allow us to see the supersymmetric structure of the Witten Laplacian as a special case of our result. In other words, the symbol $p(x,\xi;h) = |\xi|^2 + |\nabla\phi(x)|^2 - h\Delta\phi(x)$ would satisfy assumptions (i) to (iv) above.

The analyticity of the symbol $p$ with respect to the variable $\xi$ is certainly not necessary in order to get a factorization result (it suffices to take a nonanalytic $q$ in the conclusion to see it). Nevertheless, since our approach consists in conjugating the operator by $e^{-\phi/h}$, it seems difficult to deal with nonanalytic symbols. Moreover, using a regularization procedure in the proof the above theorem, it is certainly possible to prove that the symbol $q$ above can be chosen in a class $S^0_\tau(\Xi\mathscr{A})$ for some $\tau > 0$. Using this additional property, it may be possible to prove some Agmon estimates, construct more accurate quasimodes (on the 1-forms), and then to prove a full asymptotic expansion in Theorem 1.2.

A more delicate question should be to get rid of the parity assumption (ii). It is clear that this assumption is not necessary (take $q(x,\xi) = \langle\xi\rangle^{-2}\big(\mathrm{Id} + \mathrm{diag}(\xi_i/\langle\xi\rangle)\big)$ in the conclusion), but it seems difficult to prove a factorization result without it. For instance, if we consider the case $\phi = 0$ in dimension 1 (which doesn't fit exactly in our framework but enlightens the situation) then $P_h = hD_x$ cannot be smoothly factorized simultaneously on the left and on the right.

As will be seen in the proof below, the operator $Q$ (as well as $Q^*Q$) above is not unique. Trying to characterize the set of all possible $Q$ should be also a question of interest.

The optimality of assumption (iv) should be questioned. Expanding $q_0$ near $(\boldsymbol{u},0)$, we can see that we necessarily have

$$p_0(x,\xi) = |q_0(\boldsymbol{u},0)(\xi - i\nabla\phi)|^2 + \mathcal{O}(|(x-\boldsymbol{u},\xi)|^3)$$

near any critical point. In assumption (iv) we consider the case $q_0(\boldsymbol{u},0) = \mathrm{Id}$, but it could easily be relaxed to any invertible matrix $q_0(\boldsymbol{u},0)$.

Lastly, we shall mention that, for semiclassical differential operators of order 2, a supersymmetric structure (in the class of differential operators) was established by Hérau, Hitrik and Sjöstrand [Hérau et al. 2013]. This result requires fewer assumptions, but doesn't hold true in any good class of symbols.

The plan of the article is the following. In the next section we analyze the structure of the operator $\boldsymbol{T}_h^\star$ and prove the first results on the spectrum stated in (1-1). In Section 3 we prove Theorem 1.5 and apply it to the case of the random walk operator. In Section 4, we prove some preliminary spectral results, and in Section 5 we prove Theorem 1.2.

## 2. Structure of the operator and first spectral results

In this section, we analyze the structure of the spectrum of the operator $\boldsymbol{T}_h^\star$ on the space $\mathscr{H}_h = L^2(d\nu_{h,\infty})$ (see (1-3)). But it is more convenient to work with the standard Lebesgue measure than with the measure $d\nu_{h,\infty}$. We then introduce the Maxwellian $\mathcal{M}_h$, defined by

$$d\nu_{h,\infty} = \mathcal{M}_h(x)\,dx, \quad \text{so that} \quad \mathcal{M}_h = \widetilde{Z}_h \mu_h(B_h(x)) Z_h e^{-\phi(x)/h}, \tag{2-1}$$

and we make the change of function

$$\mathcal{U}_h u(x) := \mathcal{M}_h^{-1/2}(x) u(x),$$

where $\mathcal{U}_h$ is unitary from $L^2(\mathbb{R}^d) = L^2(\mathbb{R}^d, dx)$ to $\mathcal{H}_h$. Letting

$$T_h := \mathcal{U}_h^* \boldsymbol{T}_h^{\star} \mathcal{U}_h, \tag{2-2}$$

the conjugated operator acting in $L^2(\mathbb{R}^d)$, we have

$$\begin{aligned} T_h u(x) &= Z_h \mathcal{M}_h^{-1/2}(x) e^{-\phi(x)/h} \int_{\mathbb{R}^d} \mathbb{1}_{|x-y|<h} \mathcal{M}_h^{1/2}(y) \mu_h(B(y,h))^{-1} u(y)\,dy \\ &= \left( \frac{Z_h e^{-\phi(x)/h}}{\mu_h(B(x,h))} \right)^{\frac{1}{2}} \int_{|x-y|<h} u(y) \left( \frac{Z_h e^{-\phi(y)/h}}{\mu_h(B(y,h))} \right)^{\frac{1}{2}} dy. \end{aligned}$$

We let

$$a_h(x) = (\alpha_d h^d)^{1/2} \left( \frac{Z_h e^{-\phi(x)/h}}{\mu_h(B(x,h))} \right)^{\frac{1}{2}},$$

and define the operator $\mathbb{G}$ by

$$\mathbb{G}u(x) = \frac{1}{\alpha_d h^d} \int_{|x-y|<h} u(y)\,dy \tag{2-3}$$

where $\alpha_d = \operatorname{vol}(B(0,1))$ denotes the Euclidean volume of the unit ball, so that, with these notations, the operator $T_h$ is

$$T_h = a_h \mathbb{G} a_h, \tag{2-4}$$

i.e.,

$$T_h u(x) = a_h(x) \mathbb{G}(a_h u)(x).$$

We note that

$$a_h^{-2}(x) = \frac{\mu_h(B(x,h)) e^{\phi(x)/h}}{\alpha_d h^d Z_h} = \frac{1}{\alpha_d h^d} \int_{|x-y|<h} e^{(\phi(x)-\phi(y))/h}\,dy = e^{\phi(x)/h} \mathbb{G}(e^{-\phi/h})(x). \tag{2-5}$$

We now collect some properties of $\mathbb{G}$ and $a_h$.

One simple but fundamental observation is that $\mathbb{G}$ is a semiclassical Fourier multiplier, $\mathbb{G} = G(hD) = \operatorname{Op}(G)$, where

$$G(\xi) = \frac{1}{\alpha_d} \int_{|z|<1} e^{iz\cdot\xi}\,dz \quad \text{for all } \xi \in \mathbb{R}^d. \tag{2-6}$$

**Lemma 2.1.** *The function $G$ is analytic on $\mathbb{C}^d$ and enjoys the following properties*:

(i) $G : \mathbb{R}^d \to \mathbb{R}$.

(ii) *There exists* $\delta > 0$ *such that* $G(\mathbb{R}^d) \subset [-1+\delta, 1]$. *Near* $\xi = 0$, *we have*

$$G(\xi) = 1 - \beta_d |\xi|^2 + \mathcal{O}(|\xi|^4),$$

*where* $\beta_d = (2d+4)^{-1}$. *For any* $r > 0$, $\sup_{|\xi|\geq r} |G(\xi)| < 1$ *and* $\lim_{|\xi|\to\infty} G(\xi) = 0$.

(iii) *For all* $\tau \in \mathbb{R}^d$, *we have* $G(i\tau) \in \mathbb{R}$, $G(i\tau) \geq 1$ *and, for any* $r > 0$, $\inf_{|\tau|\geq r} G(i\tau) > 1$.

(iv) *For all* $\xi, \tau \in \mathbb{R}^d$ *we have* $|G(\xi + i\tau)| \leq G(i\tau)$.

*Proof.* The function $G$ is analytic on $\mathbb{C}^d$ since it is the Fourier transform of a compactly supported distribution. The fact that $G(\mathbb{R}^d) \subset \mathbb{R}$ is clear using the change of variable $z \mapsto -z$. The second item was shown in [Lebeau and Michel 2010].

We now prove (iii). The fact that $G(i\tau)$ is real for any $\tau \in \mathbb{R}^d$ is clear. Moreover, one can see easily that $\tau \mapsto G(i\tau)$ is radial, so that there exists a function $\Gamma : \mathbb{R} \to \mathbb{R}$ such that, for all $\tau \in \mathbb{R}^d$, $G(i\tau) = \Gamma(|\tau|)$. Simple computations show that $\Gamma$ enjoys the following properties:

- $\Gamma$ is even;
- $\Gamma$ is strictly increasing on $\mathbb{R}_+$;
- $\Gamma(0) = 1$.

This leads directly to the claimed properties for $G(i\tau)$.

Finally, the fact that $|G(\xi + i\tau)| \leq G(i\tau)$ for all $\xi, \tau \in \mathbb{R}^d$ is trivial, since $|e^{iz\cdot(\xi+i\tau)}| = e^{-z\cdot\tau}$ for all $z \in \mathbb{R}^d$. □

**Lemma 2.2.** *There exist* $c_1, c_2 > 0$ *such that* $c_1 < a_h(x) < c_2$ *for all* $x \in \mathbb{R}^d$ *and* $h \in ]0,1]$. *Moreover, the functions* $a_h$ *and* $a_h^{-2}$ *belong to* $S^0(1)$ *and have classical expansions* $a_h = a_0 + ha_1 + \cdots$ *and* $a_h^{-2} = a_0^{-2} + \cdots$. *In addition,*

$$\begin{aligned} a_0(x) &= G(i\nabla\phi(x))^{-1/2}, \\ a_1(x) &= G(i\nabla\phi(x))^{-3/2} \frac{1}{4\alpha_d} \int_{|z|<1} e^{-\nabla\phi(x)\cdot z} \langle \phi''(x)z, z\rangle \, dz. \end{aligned}$$

*Lastly, there exist* $c_0$, $R > 0$ *such that, for all* $|x| \geq R$, $a_h^{-2}(x) \geq 1 + c_0$ *for* $h > 0$ *small enough.*

*Proof.* By a simple change of variable, we have

$$a_h^{-2}(x) = \frac{1}{\alpha_d} \int_{|z|<1} e^{(\phi(x)-\phi(x+hz))/h} \, dz.$$

Since there exists $C > 0$ such that $|\nabla\phi(x)| \leq C$ for all $x \in \mathbb{R}^d$, we can find some constants $c_1, c_2 > 0$ such that $c_1 < a_h(x)^{-2} < c_2$ for all $x \in \mathbb{R}^d$ and $h \in ]0,1]$. Moreover, thanks to the bounds on the derivatives of $\phi$, we get easily that derivatives of $a_h^{-2}$ are also bounded. This shows that $a_h^{-2}$ belongs to $S^0(1)$ and, since it is bounded from below by $c_1 > 0$, we get immediately that $a_h \in S^0(1)$.

On the other hand, by simple Taylor expansion, we get that $a_h$ and $a_h^{-2}$ have classical expansions and the required expressions for $a_0$ and $a_1$. Since $|\nabla\phi(x)| \geq c > 0$ for $x$ large enough, it follows from

Lemma 2.1(iii) that there exist $c_0$, $R > 0$ such that, for all $|x| \geq R$, $G(i\nabla\phi(x)) \geq 1 + 2c_0$, and hence $a_h^{-2}(x) \geq 1 + c_0$ for $h > 0$ sufficiently small. □

Since we want to study the spectrum near 1, it will be convenient to introduce

$$P_h := 1 - T_h. \tag{2-7}$$

Using (2-4) and (2-5), we get

$$P_h = a_h(V_h(x) - G(hD_x))a_h \tag{2-8}$$

with $V_h(x) = a_h^{-2}(x) = e^{\phi/h}G(hD_x)(e^{-\phi/h})$. As a consequence of the previous lemmas, we get the following proposition for $P_h$:

**Proposition 2.3.** *The operator $P_h$ is a semiclassical pseudodifferential operator whose symbol $p(x, \xi; h)$ in $S^0_\infty(1)$ admits a classical expansion that reads $p = p_0 + hp_1 + \cdots$ with*

$$p_0(x,\xi) = 1 - G(i\nabla\phi(x))^{-1}G(\xi) \geq 0 \quad \textit{and} \quad p_1(x,\xi) = G_1(x)G(\xi),$$

*where*

$$G_1(x) = -G(i\nabla\phi(x))^{-2}\frac{1}{2\alpha_d}\int_{|z|<1} e^{-\nabla\phi(x)\cdot z}\langle\phi''(x)z, z\rangle\,dz = -\beta_d\Delta\phi(\boldsymbol{u}) + \mathcal{O}(|x - \boldsymbol{u}|),$$

*near any $\boldsymbol{u} \in \mathcal{U}$.*

*Proof.* The fact that $p$ belongs to $S^0_\infty(1)$ and admits a classical expansion is clear thanks to Lemma 2.1 and Lemma 2.2. From the standard pseudodifferential calculus in Weyl quantization, the symbol $p$ satisfies

$$\begin{aligned} p(x,\xi;h) &= 1 - a_0^2G - 2a_0a_1Gh - \frac{h}{2i}a_0\{G, a_0\} - \frac{h}{2i}\{a_0, a_0G\} + S^0(h^2) \\ &= 1 - a_0^2G - 2a_0a_1Gh + S^0(h^2). \end{aligned}$$

Combined with Lemma 2.2, this leads to the required expressions for $p_0$ and $p_1$.

Finally, the nonnegativity of $p_0$ comes from the formula

$$p_0 = G(i\nabla\phi(x))^{-1}\big((1 - G(\xi)) + \big(G(i\nabla\phi(x)) - 1\big)\big),$$

and Lemma 2.1, which implies that $1 - G(\xi) \geq 0$ and $G(i\nabla\phi(x)) - 1 \geq 0$. □

We finish this section with the following proposition, which is a part of Theorem 1.1.

**Proposition 2.4.** *There exist $\delta$, $h_0 > 0$ such that the following assertions hold true for $h \in\, ]0, h_0]$: First, $\sigma(T_h) \subset [-1+\delta, 1]$ and $\sigma_{\mathrm{ess}}(T_h) \subset [-1+\delta, 1-\delta]$. Second, 1 is a simple eigenvalue for the eigenfunction $\mathcal{M}_h^{1/2}$.*

*Proof.* We start by proving $\sigma(T_h) \subset [-1+\delta, 1]$. From (1-4), we already know that $\sigma(T_h) \subset [-1, 1]$. Moreover, Lemma 2.1(ii)–(iii) implies $0 \leq a_0(x) \leq 1$ and $G(\mathbb{R}^d) \subset [-1+\nu, 1]$ for some $\nu > 0$. Thus, we deduce that the symbol $\tau_h(x,\xi)$ of the pseudodifferential operator $T_h \in \Psi^0(1)$ satisfies

$$\tau_h(x,\xi) \geq -1 + \nu + \mathcal{O}(h).$$

Then, Gårding's inequality yields

$$T_h \geq -1 + \tfrac{1}{2}\nu$$

for $h$ small enough. Summing up, we obtain $\sigma(T_h) \subset [-1+\delta, 1]$.

Let us prove the assertion about the essential spectrum. Let $\chi \in C_0^\infty(\mathbb{R}^d; [0,1])$ be equal to 1 on $B(0, R)$, where $R > 0$ is as in Lemma 2.2. Since $\mathbb{G} = G(hD) \in \Psi^0(1)$ and $\lim_{|\xi|\to\infty} G(\xi) = 0$, the operator

$$T_h - (1-\chi)T_h(1-\chi) = \chi T_h + T_h \chi - \chi T_h \chi$$

is compact. Hence, $\sigma_{\mathrm{ess}}(T_h) = \sigma_{\mathrm{ess}}((1-\chi)T_h(1-\chi))$. Now, for all $u \in L^2(\mathbb{R}^d)$, we have

$$\begin{aligned}\langle (1-\chi)T_h(1-\chi)u, u\rangle &= \langle \mathbb{G} a_h(1-\chi)u, a_h(1-\chi)u\rangle \\ &\leq \|a_h(1-\chi)u\|^2 \leq (1+c_0)^{-1}\|u\|^2,\end{aligned}$$

since $\|\mathbb{G}\|_{L^2\to L^2} \leq 1$ and $|a_h(1-\chi)| \leq (1+c_0)^{-1/2}$, thanks to Lemma 2.1(iii) and Lemma 2.2. As a consequence, there exists $\delta > 0$ such that $\sigma_{\mathrm{ess}}(T_h) \subset [-1+\delta, 1-\delta]$.

To finish the proof, it remains to show that 1 is a simple eigenvalue. Let $k_h(x, y)$ denotes the distribution kernel of $T_h$. From (2-3), (2-4) and Lemma 2.2, there exists $\varepsilon > 0$ such that, for all $x, y \in \mathbb{R}^d$,

$$k_h(x, y) \geq \varepsilon h^{-d} \mathbb{1}_{|x-y|<h}. \tag{2-9}$$

We now consider $\widetilde{T}_h = T_h + 1$. Since $\|T_h\| = 1$, the operator $\widetilde{T}_h$ is bounded and nonnegative. Moreover, $\mathcal{M}_h^{1/2}$ is clearly an eigenvector associated to the eigenvalue $\|\widetilde{T}_h\| = 2$. On the other hand, (2-9) implies that $\widetilde{T}_h$ is positivity-preserving (this means that $u(x) \geq 0$ almost everywhere and $u \neq 0$ implies $\widetilde{T}_h u(x) \geq 0$ almost everywhere and $\widetilde{T}_h u \neq 0$). Furthermore, $\widetilde{T}_h$ is ergodic (in the sense that, for any $u, v \in L^2(\mathbb{R}^d)$ nonnegative almost everywhere and not the zero function, there exists $n \geq 1$ such that $\langle u, \widetilde{T}_h^n v\rangle > 0$). Indeed, let $u, v$ be two such functions. We have $\langle u, \widetilde{T}_h^n v\rangle \geq \langle u, T_h^n v\rangle$, where, by (2-9), the distribution kernel of $T_h^n$ satisfies

$$k_h^{(n)}(x, y) \geq \varepsilon_n h^{-d} \mathbb{1}_{|x-y|<(n-1)h}$$

with $\varepsilon_n > 0$. Thus, if $n \geq 1$ is chosen such that dist(ess-supp($u$), ess-supp($v$)) $< nh$, we have $\langle u, \widetilde{T}_h^n v\rangle > 0$. Lastly, the above properties of $\widetilde{T}_h$ and the Perron–Frobenius theorem (see Theorem XIII.43 of [Reed and Simon 1978]) imply that 1 is a simple eigenvalue of $T_h$. □

## 3. Supersymmetric structure

In this section, we prove that the operator $\mathrm{Id} - \boldsymbol{T}_h^\star$ admits a supersymmetric structure and prove Theorem 1.5. We showed in the preceding section that

$$\mathrm{Id} - \boldsymbol{T}_h^\star = \mathcal{U} P_h \mathcal{U}^*$$

and, before proving Theorem 1.5, we state and prove as a corollary the main result on the operator $P_h$. Recall here that $\beta_d = (2d+4)^{-1}$ and $\Xi\mathscr{A}$ is the matrix symbol defined by $\Xi\mathscr{A}_{i,j} = \langle \xi_j\rangle^{-1}$ for all $i, j = 1, \dots, d$.

**Corollary 3.1.** *There exists a classical symbol $q \in S^0_{\mathrm{cl}}(\Xi\mathscr{A})$ such that the following holds true: First, $P_h = L_\phi^* L_\phi$ with $L_\phi = Q d_{\phi,h} a_h$ and $Q = \mathrm{Op}(q)$. Second, $q = q_0 + \Psi^0(h\Xi\mathscr{A})$ with $q_0(x,\xi) = \beta_d^{1/2}\,\mathrm{Id} + \mathcal{O}(|(x-\boldsymbol{u},\xi)|^2)$ for any critical point $\boldsymbol{u} \in \mathcal{U}$.*

*Proof.* Since we know that $P_h = a_h(V_h(x) - \mathbb{G})a_h$, we only have to prove that $\beta_d^{-1}\widetilde{P}_h$ satisfies the assumptions of Theorem 1.5, where

$$\widetilde{P}_h = V_h(x) - G(hD). \tag{3-1}$$

Assumption (i) is satisfied by construction.

Observe that, thanks to Proposition 2.3, $\widetilde{P}_h$ is a pseudodifferential operator and, since the variables $x$ and $\xi$ are separated, its symbol in any quantization is given by $\tilde{p}_h(x,\xi) = V_h(x) - G(\xi)$. Moreover, Lemma 2.2 and Proposition 2.3 show that $\tilde{p}_h$ admits a classical expansion $\tilde{p} = \sum_{j=0}^\infty h^j \tilde{p}_j$ with $\tilde{p}_j$, $j \geq 1$, depending only on $x$, and $\tilde{p}_0(x,\xi) = G(i\nabla\phi(x)) - G(\xi)$. Hence, it follows from Lemma 2.1 that $\tilde{p}$ satisfies assumptions (ii) and (iii).

Finally, it follows from Lemma 2.1(ii) that, near $(\boldsymbol{u},0)$ (for any $\boldsymbol{u} \in \mathcal{U}$), we have

$$\tilde{p}(x,\xi) = \beta_d(|\xi|^2 + |\nabla\phi(x)|^2) + \mathcal{O}(|(x-\boldsymbol{u},\xi)|^4) + S^0(h),$$

so that we can apply Theorem 1.5 in the case where $r = \mathcal{O}(|(x-\boldsymbol{u},\xi)|^4)$. Taking into account the multiplication by $a_h$ completes the proof for $P_h$. □

*Proof of Theorem 1.5.* Given a symbol $p \in S^0(1)$ we recall first the well-known left and right quantizations

$$\mathrm{Op}^l(p)u(x) = (2\pi h)^{-d}\int_{T^*\mathbb{R}^d} e^{ih^{-1}(x-y)\xi} p(x,\xi)u(y)\,dy\,d\xi \tag{3-2}$$

and

$$\mathrm{Op}^r(p)u(x) = (2\pi h)^{-d}\int_{T^*\mathbb{R}^d} e^{ih^{-1}(x-y)\xi} p(y,\xi)u(y)\,dy\,d\xi. \tag{3-3}$$

If $p(x,y,\xi)$ belongs to $S^0(1)$, we define $\widetilde{\mathrm{Op}}(p)$, by

$$\widetilde{\mathrm{Op}}(p)(u)(x) = (2\pi h)^{-d}\int_{T^*\mathbb{R}^d} e^{ih^{-1}(x-y)\xi} p(x,y,\xi)u(y)\,dy\,d\xi, \tag{3-4}$$

We recall the formula allowing us to pass from one of these quantizations to the other. If $p(x,y,\xi)$ belongs to $S^0(1)$, then $\widetilde{\mathrm{Op}}(p) = \mathrm{Op}^l(p_l) = \mathrm{Op}^r(p_r)$ with

$$p_l(x,\xi) = (2\pi h)^{-d}\int_{T^*\mathbb{R}^d} e^{ih^{-1}z(\xi'-\xi)} p(x,x-z,\xi')\,d\xi'\,dz, \tag{3-5}$$

and

$$p_r(y,\xi) = (2\pi h)^{-d}\int_{T^*\mathbb{R}^d} e^{ih^{-1}z(\xi'-\xi)} p(y+z,y,\xi')\,d\xi'\,dz. \tag{3-6}$$

Recall that we introduced the $d$-matrix weight $\mathscr{A} : T^*\mathbb{R}^d \to \mathcal{M}_d$ given by $\mathscr{A}_{i,j}(x,\xi) = (\langle\xi_i\rangle\langle\xi_j\rangle)^{-1}$. Suppose that $p$ satisfies the hypotheses of Theorem 1.5: $P = \mathrm{Op}(p)$ with $p \in S^0_\infty(1)$, $p(x,\xi;h) = p_0(x,\xi) + S^0(h)$ such that:

(i) $P(e^{-\phi/h}) = 0$;

(ii) for all $x \in \mathbb{R}^d$, the function $\xi \in \mathbb{R}^d \mapsto p(x, \xi; h)$ is even;

(iii) for all $\delta > 0$, there exists $\alpha > 0$ such that, for all $(x, \xi) \in T^*\mathbb{R}^d$, $d(x, \mathcal{U})^2 + |\xi|^2 \geq \delta$ implies $p_0(x, \xi) \geq \alpha$;

(iv) near any critical points $\boldsymbol{u} \in \mathcal{U}$ we have

$$p_0(x, \xi) = |\xi|^2 + |\nabla\phi(x)|^2 + r(x, \xi)$$

with either $r = \mathcal{O}(|(x - \boldsymbol{u}, \xi)|^3)$ (assumption (A2)), or $r = \mathcal{O}(|(x - \boldsymbol{u}, \xi)|^4)$ (assumption (A2′)).

The symbol $p$ may depend on $h$, but we omit this dependence in order to lighten the notations.

The proof goes in several steps. First we prove that there exists a symbol $\hat{q} \in S^0_\infty(\mathscr{A})$ such that

$$P_h = d^*_{\phi,h} \widehat{Q} d_{\phi,h}, \quad \text{where } \widehat{Q} = \mathrm{Op}(\hat{q}).$$

In a moment we shall prove that the operator $\widehat{Q}$ can be chosen so that $\widehat{Q} = Q^* Q$ for some pseudodifferential operator $Q$ satisfying some good properties.

Let us start with the first step. For this purpose we need the following lemma:

**Lemma 3.2.** *Let $p \in S^0_\infty(1)$ and $P_h = \mathrm{Op}(p)$. Assume that, for all $x \in \mathbb{R}$, the function $\xi \mapsto p(x, \xi; h)$ is even. Suppose also that $P_h(e^{-\phi/h}) = 0$. Then, there exists $\hat{q} \in S^0_\infty(\mathscr{A})$ such that $P_h = d^*_{\phi,h} \widehat{Q} d_{\phi,h}$ with $\widehat{Q} = \mathrm{Op}(\hat{q})$. Moreover, if $p$ has a principal symbol then so does $\hat{q}$, and if $p \in S^0_{\infty,\mathrm{cl}}$ then $\hat{q} \in S^0_{\infty,\mathrm{cl}}$.*

**Remark 3.3.** Since $P_h(e^{-\phi/h}) = 0$, it is quite clear that $P_h$ can be factorized by $d_{\phi,h}$ on the right. On the other hand, the fact that $P_h$ can be factorized by $d^*_{\phi,h}$ on the left necessarily implies that $P^*_h(e^{-\phi/h}) = 0$. At first glance, there is no reason for this identity to hold true, since we don't suppose in the above lemma that $P_h$ is selfadjoint. This is actually verified for the following reason. Start from $\mathrm{Op}(p)(e^{-\phi/h}) = 0$; then, taking the conjugate and using the fact that $\phi$ is real, we get

$$\mathrm{Op}(\bar{p}(x, -\xi))(e^{-\phi/h}) = 0.$$

Hence, the parity assumption on $p$ implies that $\mathrm{Op}(p)^*(e^{-\phi/h}) = 0$.

*Proof of Lemma 3.2.* The fundamental, simple remark is that, if $a$ is a symbol such that $a(x, \xi) = b(x, \xi) \cdot \xi$, then the operator $\mathrm{Op}^l(a)$ can be factorized by $hD_x$ on the right: $\mathrm{Op}^l(a) = \mathrm{Op}^l(b) \cdot hD_x$, whereas the right quantization of $a$ can be factorized on the left: $\mathrm{Op}^r(a) = hD_x \cdot \mathrm{Op}^r(b)$. We have to implement this simple idea, dealing with the fact that our operator is twisted by $e^{\phi/h}$.

Introduce the operator $P_{\phi,h} = e^{\phi/h} P_h e^{-\phi/h}$. Then, for any $u \in \mathcal{S}(\mathbb{R}^d)$,

$$P_{\phi,h} u(x) = (2\pi h)^{-d} \int e^{ih^{-1}(x-y)\xi} e^{h^{-1}(\phi(x)-\phi(y))} p\big(\tfrac{1}{2}(x + y), \xi\big) u(y) \, dy \, d\xi.$$

We now use the Kuranishi trick. Let $\theta(x, y) = \int_0^1 \nabla\phi(tx + (1 - t)y) \, dt$. Then $\phi(x) - \phi(y) = (x - y) \cdot \theta(x, y)$ and

$$P_{\phi,h} u(x) = (2\pi h)^{-d} \int e^{ih^{-1}(x-y)(\xi - i\theta(x,y))} p\big(\tfrac{1}{2}(x + y), \xi\big) u(y) \, dy \, d\xi.$$

Since $p \in S^0_\infty$, a simple change of integration path shows that $P_{\phi,h}$ is a bounded pseudodifferential operator $P_{\phi,h} = \widetilde{\mathrm{O}}\mathrm{p}(\tilde{p}_\phi)$ with

$$\tilde{p}_\phi(x, y, \xi) = p\big(\tfrac{1}{2}(x+y), \xi + i\theta(x, y)\big).$$

To get the expression of $P_{\phi,h}$ in left quantization, it suffices then to apply (3-5) to get $P_{\phi,h} = \mathrm{Op}^l(p_\phi)$ with

$$\begin{aligned} p_\phi(x,\xi) &= (2\pi h)^{-d} \int_{\mathbb{R}^{2d}} e^{ih^{-1}(\xi'-\xi)(x-z)} p\big(\tfrac{1}{2}(x+z), \xi' + i\theta(x,z)\big)\, d\xi'\, dz \\ &= (2\pi h)^{-d} \int_{\mathbb{R}^{2d}} e^{ih^{-1}\xi'(x-z)} p\big(\tfrac{1}{2}(x+z), \xi' + \xi + i\theta(x,z)\big)\, d\xi'\, dz. \end{aligned}$$

Observe that for any smooth function $g : \mathbb{R}^d \to \mathbb{R}$ we have

$$g(\xi) - g(0) = \sum_{j=1}^{d} \int_0^1 \xi_j \partial_{\xi_j} g(\gamma_j^\pm(s,\xi))\, ds \tag{3-7}$$

with $\gamma_j^+(s,\xi) = (\xi_1, \dots, \xi_{j-1}, s\xi_j, 0, \dots, 0)$ and $\gamma_j^-(s,\xi) = (0, \dots, 0, s\xi_j, \xi_{j+1}, \dots, \xi_d)$. A very simple observation is that, for any $(x,\xi) \in T^*\mathbb{R}^d$ and any $s \in [0,1]$, we have $x \cdot \gamma_j^\pm(s,\xi) = \gamma_j^\pm(s,x) \cdot \xi$. This will be used often in the sequel.

Let us go back to the study of $p_\phi$. Since $P_h(e^{-\phi/h}) = 0$, we have $p_\phi(x,0) = 0$ and, by (3-7), we get

$$p_\phi(x,\xi) = \sum_{j=1}^{d} \xi_j \check{q}^\pm_{\phi,j}(x,\xi) = \sum_{j=1}^{d} \xi_j \check{q}_{\phi,j}(x,\xi)$$

with $\check{q}_{\phi,j} = \frac{1}{2}(\check{q}^+_{\phi,j} + \check{q}^-_{\phi,j})$ and

$$\check{q}^\pm_{\phi,j}(x,\xi) = (2\pi h)^{-d} \int_{\mathbb{R}^{2d}} e^{ih^{-1}\xi'(x-z)} \int_0^1 \partial_{\xi_j} p\big(\tfrac{1}{2}(x+z), \xi' + \gamma_j^\pm(s,\xi) + i\theta(x,z)\big)\, ds\, dz\, d\xi',$$

where the above integral has to be understood as an oscillatory integral. Since $\partial_\xi^\alpha p$ is bounded for any $\alpha$, integration by parts with respect to $\xi'$ and $z$ shows that $\check{q}^\pm_{\phi,j} \in S^0_\infty(1)$. Moreover, by definition of $\gamma_j^\pm$ we have

$$\xi_j \check{q}^\pm_{\phi,j} = (2\pi h)^{-d} \int_{\mathbb{R}^{2d}} e^{ih^{-1}\xi'(x-z)} c_j^\pm(x,z,\xi)\, dz\, d\xi$$

with $c_j^\pm(x,z,\xi) = p\big(\frac{1}{2}(x+z), \xi' + \gamma_j^\pm(1,\xi) + i\theta(x,z)\big) - p\big(\frac{1}{2}(x+z), \xi' + \gamma_j^\pm(0,\xi) + i\theta(x,z)\big)$. This symbol is clearly in $S^0_\infty(1)$, so that integration by parts as before shows that $\xi_j \check{q}^\pm_{\phi,j} \in S^0_\infty(1)$. Since $\xi_j$ and $\check{q}^\pm_{\phi,j}$ are both scalar, this proves that $\check{q}^\pm_{\phi,j} \in S^0_\infty(\langle \xi_j \rangle^{-1})$.

Observe now that

$$P_h = e^{-\phi/h} P_{\phi,h} e^{\phi/h} = e^{-\phi/h}\, \mathrm{Op}^l\big(\tfrac{1}{2}(\check{q}^+_\phi + \check{q}^-_\phi)\big) \cdot \Big(\frac{h}{i}\nabla_x\Big) e^{\phi/h} = e^{-\phi/h} \widetilde{Q} e^{\phi/h} \cdot d_{\phi,h}$$

with $\widetilde{Q}=\frac{1}{2}(\widetilde{Q}^{+}+\widetilde{Q}^{-})$ and $\widetilde{Q}^{\pm}=\mathrm{Op}^{l}(\check{q}^{\pm}_{\phi})$. Let $\widetilde{Q}^{\pm}_{\phi}=e^{-2\phi/h}\,\mathrm{Op}^{l}(\check{q}^{\pm}_{\phi})e^{2\phi/h}$; then $\widetilde{Q}^{\pm}_{\phi}=\widetilde{\mathrm{O}}\mathrm{p}(\tilde{q}^{\pm}_{\phi})$ with $\tilde{q}^{\pm}_{\phi}=(\tilde{q}^{\pm}_{\phi,1},\dots,\tilde{q}^{\pm}_{\phi,d})$ and

$$\begin{aligned}\tilde{q}^{\pm}_{\phi,j}(x,y,\xi)&=\check{q}^{\pm}_{\phi,j}(x,\xi-2i\theta(x,y))\\&=(2\pi h)^{-d}\int_{\mathbb{R}^{2d}}e^{ih^{-1}\xi'(x-z)}\int_0^1\partial_{\xi_j}p\big(\tfrac{1}{2}(x+z),\xi'+\gamma_j^{\pm}(s,\xi)-2i\gamma_j^{\pm}(s,\theta(x,y))+i\theta(x,z)\big)\,ds\,dz\,d\xi',\end{aligned}$$

and it follows from (3-6) that $\widetilde{Q}_{\phi}=\mathrm{Op}^{r}(\check{q}_{\phi})$ with $\check{q}_{\phi}=\check{q}^{+}_{\phi}+\check{q}^{-}_{\phi}$, $\check{q}^{\pm}_{\phi}=(\check{q}^{\pm}_{\phi,1},\dots,\check{q}^{\pm}_{\phi,d})$ and

$$\begin{aligned}&\check{q}^{\pm}_{\phi,j}(x,\xi)\\&=(2\pi h)^{-d}\int_{\mathbb{R}^{2d}}e^{ih^{-1}(\xi'-\xi)u}\tilde{q}^{\pm}_{\phi,j}(x+u,x,\xi')\,du\,d\xi'\\&=(2\pi h)^{-2d}\int_{\mathbb{R}^{4d}}\int_0^1 e^{ih^{-1}[(\xi'-\xi)u+(x+u-z)\eta]}\\&\qquad\times\partial_{\xi_j}p\big(\tfrac{1}{2}(x+u+z),\eta+\gamma_j^{\pm}(s,\xi')-2i\gamma_j^{\pm}(s,\theta(x+u,x))+i\theta(x+u,z)\big)\,ds\,dz\,du\,d\xi'\,d\eta.\end{aligned}$$

Make the change of variables $z=x+v$ and $\nu=\gamma_j^{\pm}(s,\xi')+\eta$; the above equation yields

$$\begin{aligned}\check{q}^{\pm}_{\phi,j}(x,\xi)=(2\pi h)^{-2d}\int_{\mathbb{R}^{4d}}\int_0^1 &e^{ih^{-1}[(\xi'-\xi)u+(u-v)(\nu-\gamma_j^{\pm}(s,\xi'))]}\\&\times\partial_{\xi_j}p\big(x+\tfrac{1}{2}(u+v),\nu+\psi_j^{\pm}(s,x,u,v)\big)\,ds\,du\,dv\,d\nu\,d\xi'\end{aligned}$$

with $\psi_j^{\pm}(s,x,u,v)=i\theta(x+u,x+v)-2i\gamma_j^{\pm}(s,\theta(x+u,x))$.

Define $\hat{p}^2(x,z)=\int e^{-iz\xi}p(x,\xi)\,d\xi$, the Fourier transform of $p$ with respect to the second variable, and observe that, since $\xi\mapsto p(x,\xi)$ is even, so is $z\mapsto\hat{p}^2(x,z)$. Using the above notations, we have

$$\partial_{\eta_j}p(x,\eta)=\frac{i}{(2\pi)^d}\int_{\mathbb{R}^d}e^{iz\eta}z_j\,\hat{p}^2(x,z)\,dz,$$

and we get

$$\begin{aligned}\check{q}^{\pm}_{\phi,j}(x,\xi)=\frac{i}{(2\pi)^d(2\pi h)^{2d}}\int_{\mathbb{R}^{4d}\times[0,1]\times\mathbb{R}^d}&z_je^{ih^{-1}[(u-v+hz)\nu+(\xi'-\xi)u-(u-v)\gamma_j^{\pm}(s,\xi')]}\\&\times\hat{p}^2\big(x+\tfrac{1}{2}(u+v),z\big)e^{iz\psi_j^{\pm}(s,x,u,v)}\,du\,dv\,d\xi'\,d\nu\,ds\,dz.\end{aligned}$$

Let $\mathscr{F}_{h,\nu\mapsto v}$ denote the semiclassical Fourier transform with respect to $\nu$, and $\overline{\mathscr{F}}_{h,u\mapsto\nu}$ its inverse. Writing

$$f_{s,x,v,z}(u)=z_j\,\hat{p}^2\big(x+\tfrac{1}{2}(u+v),z\big)e^{ih^{-1}[(\xi'-\xi)u-(u-v)\gamma_j^{\pm}(s,\xi')]}e^{iz\psi_j^{\pm}(s,x,u,v)},$$

we get

$$\begin{aligned}\breve{q}^{\pm}_{\phi,j}(x,\xi) &= \frac{i}{(2\pi)^d(2\pi h)^d}\int_{\mathbb{R}^{2d}\times[0,1]\times\mathbb{R}^d} \mathscr{F}_{h,v\mapsto v}\overline{\mathscr{F}}_{h,u\mapsto v}(f_{s,x,v,z})(v-hz)\,dv\,d\xi'\,ds\,dz\\ &= \frac{i}{(2\pi)^d(2\pi h)^d}\int_{\mathbb{R}^{2d}\times[0,1]\times\mathbb{R}^d} z_j e^{ih^{-1}[(\xi'-\xi)(v-hz)+h\gamma_j^{\pm}(s,z)\xi']}\\ &\qquad\qquad \times \hat{p}^2\big(x+v-\tfrac{1}{2}hz,z\big)e^{i\psi_j^{\pm}(s,x,v-hz,v)z}\,dv\,d\xi'\,ds\,dz,\end{aligned}$$

where we have used the fact that $\gamma_j^{\pm}(s,\xi')z = \gamma_j^{\pm}(s,z)\xi'$. Similarly, integrating with respect to $\xi'$ and $v$, we obtain

$$\breve{q}^{\pm}_{\phi,j}(x,\xi) = \frac{i}{(2\pi)^d}\int_{[0,1]\times\mathbb{R}^d} z_j e^{i\gamma_j^{\pm}(s,z)\xi}\hat{p}^2\big(x+h\big(\tfrac{1}{2}z-\gamma_j^{\pm}(s,z)\big),z\big)e^{\varphi_j^{\pm}(s,z)}\,ds\,dz$$

with $\varphi_j^{\pm}(s,z) = iz\psi_j^{\pm}\big(s,x,-h\gamma_j^{\pm}(s,z),h(z-\gamma_j^{\pm}(s,z))\big)$. From the definition of $\psi_j^{\pm}$, we get

$$\begin{aligned}\varphi_j^{\pm}(s,z) &= 2z\gamma_j^{\pm}\big(s,\theta(x-h\gamma_j^{\pm}(s,z),x)\big) - z\theta\big(x-h\gamma_j^{\pm}(s,z),x+h(z-\gamma_j^{\pm}(s,z))\big)\\ &= 2\gamma_j^{\pm}(s,z)\theta(x-h\gamma_j^{\pm}(s,z),x) - z\theta\big(x-h\gamma_j^{\pm}(s,z),x+h(z-\gamma_j^{\pm}(s,z))\big),\end{aligned}$$

and, since $\theta$ is defined by $\phi(x)-\phi(y)=(x-y)\theta(x,y)$, it follows easily that

$$\varphi_j^{\pm}(s,z) = \frac{1}{h}\big(2\phi(x)-\phi(x-h\gamma_j^{\pm}(s,z))-\phi\big(x+h(z-\gamma_j^{\pm}(s,z))\big)\big).$$

Let us write $\rho_j^{\pm}(x,s,z) = \hat{p}^2\big(x+h\big(\frac{1}{2}z-\gamma_j^{\pm}(s,z)\big),z\big)$; then

$$\breve{q}^{\pm}_{\phi,j}(x,0) = \frac{i}{(2\pi)^d}\int_{[0,1]\times\mathbb{R}^d} z_j\rho_j^{\pm}(x,s,z)e^{\varphi_j^{\pm}(s,z)}\,ds\,dz. \tag{3-8}$$

Observe now that we have the identities

$$\begin{aligned}\gamma_j^{\pm}(1-s,-z) &= -(z-\gamma_j^{\mp}(s,z)),\\ \tfrac{1}{2}z-\gamma_j^{\pm}(s,z) &= -\tfrac{1}{2}z-\gamma_j^{\mp}(1-s,-z)\end{aligned} \tag{3-9}$$

for all $s\in[0,1]$, $z\in\mathbb{R}^d$. In particular, since $\hat{p}^2$ is even with respect to the second variable, we get

$$\rho_j^{\pm}(x,1-s,-z)e^{\varphi_j^{\pm}(1-s,-z)} = \rho_j^{\mp}(x,s,z)e^{\varphi_j^{\mp}(s,z)}.$$

As a consequence, by the change of variables $(s,z)\mapsto(1-s,-z)$ in (3-8), we get $\breve{q}^{+}_{\phi,j}(x,0) = -\breve{q}^{-}_{\phi,j}(x,0)$, and hence $\breve{q}_\phi(x,0)=0$. Since $\breve{q}_{\phi,j}$ belongs to $S^0_\infty(\langle\xi_j\rangle^{-1})$ for all $j$, we get by the same trick as for the right factorization that there exists some symbol $q=(\bar{q}_{j,k})\in S^0_\infty(\mathscr{A})$ such that $\breve{q}_{\phi,j}(x,\xi)=\sum_{k=1}^d \xi_k\bar{q}_{j,k}(x,\xi)$. Since we use right quantization, it follows that, for all $u\in\mathscr{S}(\mathbb{R}^d,\mathbb{C}^d)$,

$$\operatorname{Op}^r(\breve{q}_\phi)u = \frac{h}{i}\operatorname{div}\operatorname{Op}^r(\bar{q})u = hD_x^*\operatorname{Op}^r(\bar{q})u,$$

where we have used the matrix-valued symbol $q=(q_{j,k})$. Consequently, for all $u\in\mathscr{S}(\mathbb{R}^d)$,

$$P_h u = e^{\phi/h}\operatorname{Op}^r(\breve{q}_\phi)e^{-\phi/h}d_{\phi,h}u = d^*_{\phi,h}e^{\phi/h}\operatorname{Op}^r(\bar{q})e^{-\phi/h}d_{\phi,h}u.$$

Using again the analyticity of $\bar{q}$, there exists $\hat{q} \in S^0_\infty(A)$ such that

$$\widehat{Q} := e^{\phi/h} \operatorname{Op}^r(\bar{q}) e^{-\phi/h} = \operatorname{Op}(\hat{q}),$$

and the factorization is proved. The fact that $\hat{q}$ admits an expansion in powers of $h$ follows easily from the above computations, since it is the case for $p$. □

Let us apply Lemma 3.2 to $P_h = \operatorname{Op}(p)$. Then, there exists a symbol $\hat{q} \in S^0_\infty(\mathscr{A})$ such that

$$P_h = d^*_{\phi,h} \widehat{Q} d_{\phi,h},$$

with $\widehat{Q} = \operatorname{Op}(\hat{q})$ and $\hat{q} = \hat{q}_0 + S^0(h)$. Now the strategy is the following: we will modify the operator $\widehat{Q}$ so that the new $\widehat{Q}$ is selfadjoint, nonnegative and $\widehat{Q}$ can be written as the square of a pseudodifferential operator, $\widehat{Q} = Q^* Q$.

First observe that, since $P_h$ is selfadjoint,

$$P_h = \tfrac{1}{2}(P_h + P_h^*) = d^*_{\phi,h} \tfrac{1}{2}(\widehat{Q} + \widehat{Q}^*) d_{\phi,h},$$

so that we can assume in the following that $\widehat{Q}$ is selfadjoint. This means that the partial operators $\widehat{Q}_{j,k} = \operatorname{Op}(\hat{q}_{j,k})$ satisfy $\widehat{Q}^*_{j,k} = \widehat{Q}_{k,j}$ (or, at the level of symbols, $\hat{q}_{k,j} = \overline{\hat{q}_{j,k}}$). For $k = 1, \dots, d$, let us write $d^k_{\phi,h} = h\partial_k + \partial_k \phi(x)$. Then

$$P_h = \sum_{j,k=1}^{d} (d^j_{\phi,h})^* \widehat{Q}_{j,k} d^k_{\phi,h}. \tag{3-10}$$

We would like to take the square root of $\widehat{Q}$ and show that it is still a pseudodifferential operator. The problem is that we don't even know if $\widehat{Q}$ is nonnegative. Nevertheless, we can use the nonuniqueness of operators $\widehat{Q}$ such that (3-10) holds to go to a situation where $\widehat{Q}$ is close to a diagonal operator with nonnegative partial operators on the diagonal. The starting point of this strategy is the commutation relation

$$[d^j_{\phi,h}, d^k_{\phi,h}] = 0 \quad \text{for all } j, k \in \{1, \dots, d\}, \tag{3-11}$$

which holds thanks to $d^j_{\phi,h} = e^{-\phi/h} h\partial_j e^{\phi/h}$ and Schwarz' theorem. Hence, for any bounded operator $B$, we have

$$P_h = d^*_{\phi,h} \widehat{Q}^{\mathrm{mod},\bullet} d_{\phi,h} = \sum_{j,k=1}^{d} (d^j_{\phi,h})^* \widehat{Q}^{\mathrm{mod},\bullet}_{j,k} d^k_{\phi,h}, \tag{3-12}$$

with $\widehat{Q}^{\mathrm{mod},\bullet} = \widehat{Q} + \mathscr{B}^\bullet$, $\bullet \in \{0, \infty\}$ for some $\mathscr{B}^\bullet$ having one of the two following forms:

- (*exchange between three coefficients*) For any $j_0, k_0, n \in \{1, \dots, d\}$, the operator $\mathscr{B}^\infty(j_0, k_0, n; B) = (\mathscr{B}^\infty_{j,k})_{j,k=1,\dots,d}$ is defined by

$$\begin{aligned} \mathscr{B}^\infty_{j,k} &= 0 \quad \text{if } (j,k) \notin \{(n,n), (j_0,k_0), (k_0,j_0)\}, \\ \mathscr{B}^\infty_{j_0,k_0} &= -(d^n_{\phi,h})^* B d^n_{\phi,h} \quad \text{and} \quad \mathscr{B}^\infty_{k_0,j_0} = (\mathscr{B}^\infty_{j_0,k_0})^*, \\ \mathscr{B}^\infty_{n,n} &= (d^{j_0}_{\phi,h})^* B d^{k_0}_{\phi,h} + (d^{k_0}_{\phi,h})^* B^* d^{j_0}_{\phi,h}. \end{aligned} \tag{3-13}$$

When $j_0 = k_0$, we use the convention that $\mathscr{B}^{\infty}_{j_0,j_0} = -(d^n_{\phi,h})^*(B + B^*)d^n_{\phi,h}$. Such modifications will be used away from the critical points.

- (*exchange between four coefficients*) For any $j_0, k_0, k_1 \in \{1, \dots, d\}$, the operator $\mathscr{B}^0(j_0, k_0, k_1; B) = (\mathscr{B}^0_{j,k})_{j,k=1,\dots,d}$ is defined by

$$\begin{aligned} \mathscr{B}^0_{j,k} &= 0 \quad \text{if } (j,k) \notin \{(j_0,k_0),(k_0,j_0),(j_0,k_1),(k_1,j_0)\}, \\ \mathscr{B}^0_{j_0,k_0} &= -Bd^{k_1}_{\phi,h} \quad \text{and} \quad \mathscr{B}^0_{k_0,j_0} = (\mathscr{B}^0_{j_0,k_0})^*, \\ \mathscr{B}^0_{j_0,k_1} &= Bd^{k_0}_{\phi,h} \quad \text{and} \quad \mathscr{B}^0_{k_1,j_0} = (\mathscr{B}^0_{j_0,k_1})^*. \end{aligned} \tag{3-14}$$

Such modifications will be used near the critical points.

Recall that the $d$-matrix weights $\mathscr{A}$ and $\Xi\mathscr{A}$ are given by $\mathscr{A}_{j,k} = \langle\xi_j\rangle^{-1}\langle\xi_k\rangle^{-1}$ and $(\Xi\mathscr{A})_{j,k} = \langle\xi_k\rangle^{-1}$. Using the preceding remark, we can prove the following:

**Lemma 3.4.** *Let $\widehat{Q} = \mathrm{Op}(\hat{q})$, where $\hat{q} \in S^0(\mathscr{A})$ is a Hermitian symbol with $\hat{q}(x,\xi;h) = \hat{q}_0(x,\xi) + S^0(h\mathscr{A})$. We let $P = d^*_{\phi,h}\widehat{Q}d_{\phi,h}$ and let $p(x,\xi;h) = p_0(x,\xi) + S^0(h) \in S^0(1)$ be its symbol. Assume that the following assumptions hold*:

(A1) *For all $\delta > 0$, there exists $\alpha > 0$ such that, for all $(x,\xi) \in T^*\mathbb{R}^d$, $|\xi|^2 + d(x,\mathscr{U})^2 \geq \delta$ implies $p_0(x,\xi) \geq \alpha$.*

(A2) *Near $(\boldsymbol{u}, 0)$, for any critical point $\boldsymbol{u} \in \mathscr{U}$, we have*

$$p_0(x,\xi) = |\xi|^2 + |\nabla\phi(x)|^2 + r(x,\xi) \tag{3-15}$$

*with $r(x,\xi) = \mathscr{O}(|(x-\boldsymbol{u},\xi)|^3)$.*

*Then, for $h$ small enough, there exists a symbol $q \in S^0(\Xi\mathscr{A})$ such that*

$$P_h = d^*_{\phi,h}Q^*Qd_{\phi,h},$$

*with $Q = \mathrm{Op}(q)$ and*

$$q(x,\xi;h) = \mathrm{Id} + \mathscr{O}(|(x-\boldsymbol{u},\xi)|) + S^0(h) \tag{3-16}$$

*near $(\boldsymbol{u}, 0)$ for any $\boldsymbol{u} \in \mathscr{U}$. Moreover, $Q = F\,\mathrm{Op}(\Xi^{-1})$ for some $F \in \Psi^0(1)$ that is invertible and selfadjoint with $F^{-1} \in \Psi^0(1)$.*

*If, additionally to the previous assumptions, we suppose*:

(A2′) *the remainder term in* (3-15) *satisfies $r(x,\xi) = \mathscr{O}(|(x-\boldsymbol{u},\xi)|^4)$;*

*then*

$$q(x,\xi;h) = \mathrm{Id} + \mathscr{O}(|(x-\boldsymbol{u},\xi)|^2) + S^0(h) \tag{3-17}$$

*near $(\boldsymbol{u}, 0)$.*

*Finally, if $\hat{q} \in S^0_{\mathrm{cl}}(\mathscr{A})$ then $q \in S^0_{\mathrm{cl}}(\Xi\mathscr{A})$.*

*Proof.* In the following, we assume that $\phi$ has a unique critical point $\boldsymbol{u}$ and that $\boldsymbol{u}=0$. Using some cutoff in space, we can always make this assumption without loss of generality. Given $\varepsilon>0$, let $w_0, w_1, \dots, w_d \in S^0(1)$ be nonnegative functions such that

$$w_0+w_1+\cdots+w_d=1 \tag{3-18}$$

whose support satisfies

$$\operatorname{supp}(w_0)\subset\{|\xi|^2+|\nabla\phi(x)|^2\leq 2\varepsilon\},$$

and, for all $\ell\geq 1$,

$$\operatorname{supp}(w_\ell)\subset\Big\{|\xi|^2+|\nabla\phi(x)|^2\geq\varepsilon \text{ and } |\xi_\ell|^2+|\partial_\ell\phi(x)|^2\geq\frac{1}{2d}(|\xi|^2+|\nabla\phi(x)|^2)\Big\}.$$

Let us decompose $\widehat{Q}$ according to these truncations:

$$\widehat{Q}=\sum_{\ell=0}^{d}\widehat{Q}^\ell \tag{3-19}$$

with $\widehat{Q}^\ell:=\operatorname{Op}(w_\ell\hat{q})$ for all $\ell\geq 0$. We will modify each of the operators $\widehat{Q}^\ell$ separately, using the following modifiers. For $j_0, k_0, n\in\{1,\dots,d\}$ and $\beta\in S^0(\langle\xi_{j_0}\rangle^{-1}\langle\xi_{k_0}\rangle^{-1}\langle\xi_n\rangle^{-2})$ we write for short

$$\mathcal{B}^\infty(j_0,k_0,n;\beta):=\mathcal{B}^\infty(j_0,k_0,n;\operatorname{Op}(\beta)),$$

where the right-hand side is defined by (3-13). In the same way, given $j_0,k_0,k_1\in\{1,\dots,d\}$ and $\beta\in S^0(\langle\xi_{j_0}\rangle^{-1}\langle\xi_{k_0}\rangle^{-1}\langle\xi_{k_1}\rangle^{-1})$ we write for short

$$\mathcal{B}^0(j_0,k_0,k_1;\beta):=\mathcal{B}^0(j_0,k_0,k_1;\operatorname{Op}(\beta)),$$

where the right-hand side is defined by (3-14). Observe that any operator of one of these two forms belongs to $\Psi^0(\mathscr{A})$. Let $\mathcal{M}(\mathscr{A})\subset\Psi^0(\mathscr{A})$ be the vector space of bounded operators on $L^2(\mathbb{R}^d)^d$ generated by these operators. Then, (3-12) says exactly that

$$P_h=d^*_{\phi,h}(\widehat{Q}+\mathcal{M})d_{\phi,h}\quad\text{for any } \mathcal{M}\in\mathcal{M}(\mathscr{A}). \tag{3-20}$$

**Step 1.** We first remove the terms of order 1 near the origin. More precisely, we show that there exists $\mathcal{M}^0\in\mathcal{M}(\mathscr{A})$ such that

$$\check{Q}^0:=\widehat{Q}^0+\mathcal{M}^0=\operatorname{Op}(\check{q}^0)+\Psi^0(h\mathscr{A}), \tag{3-21}$$

where $\check{q}^0\in S^0(\mathscr{A})$ satisfies, near $(0,0)\in T^*\mathbb{R}^d$,

$$\check{q}^0(x,\xi)=w_0(x,\xi)(\operatorname{Id}+\rho(x,\xi)) \tag{3-22}$$

with $\rho\in S(\mathscr{A})$ such that:

- $\rho(x,\xi)=\mathcal{O}(|(x,\xi)|)$ under the assumption (A2);
- $\rho(x,\xi)=\mathcal{O}(|(x,\xi)|^2)$ under the assumption (A2′).

From (3-10), we have

$$p_0(x,\xi) = \sum_{j,k=1}^{d} \hat{q}_{0;j,k}(x,\xi)(\xi_j + i\partial_j\phi(x))(\xi_k - i\partial_k\phi(x)),$$

where $\hat{q}_0 = (\hat{q}_{0;j,k})_{j,k}$ denotes the principal symbol of $\hat{q}$. Expanding $\hat{q}_0$ near the origin, we get

$$\hat{q}_0(x,\xi) = \hat{q}_0(0,0) + \nu(x,\xi)$$

with $\nu(x,\xi) = \mathcal{O}(|(x,\xi)|)$. Then, we deduce

$$p_0(x,\xi) = \sum_{j,k=1}^{d} (\hat{q}_{0;j,k}(0,0) + \nu_{j,k}(x,\xi))(\xi_j + i\partial_j\phi(x))(\xi_k - i\partial_k\phi(x)). \tag{3-23}$$

Identifying (3-15) and (3-23), we obtain $\hat{q}_{0;j,k}(0,0) = \delta_{j,k}$, which establishes (3-21)–(3-22) under the assumption (A2).

Suppose now that (A2′) is satisfied. Identifying (3-15) and (3-23) as before, we obtain

$$\sum_{j,k=1}^{d} \nu_{j,k}(x,\xi)(\xi_j + i\partial_j\phi(x))(\xi_k - i\partial_k\phi(x)) = \mathcal{O}(|(x,\xi)|^4). \tag{3-24}$$

Defining $A := \operatorname{Hess}(\phi)(0)$, we have $\partial_j\phi(x) = (Ax)_j + \mathcal{O}(x^2)$. Then, (3-24) becomes

$$\sum_{j,k=1}^{d} \nu_{j,k}(x,\xi)(\xi_j + i(Ax)_j)(\xi_k - i(Ax)_k) = \mathcal{O}(|(x,\xi)|^4). \tag{3-25}$$

Let us introduce the new variables $\eta = \xi + iAx$ and $\bar{\eta} = \xi - iAx$. Then, (3-25) reads

$$\sum_{j,k=1}^{d} \nu_{j,k}(x,\xi)\eta_j\bar{\eta}_k = \mathcal{O}(|(x,\xi)|^4) = \mathcal{O}(|(\eta,\bar{\eta})|^4). \tag{3-26}$$

On the other hand, since $A$ is invertible, there exist some complex numbers $\alpha_{j,k}^n, \tilde{\alpha}_{j,k}^n$ for $j,k,n = 1,\dots,d$ such that

$$\nu_{j,k}(x,\xi) = \sum_{n=1}^{d} (\alpha_{j,k}^n\bar{\eta}_n + \tilde{\alpha}_{j,k}^n\eta_n) + \mathcal{O}(|(\eta,\bar{\eta})|^2). \tag{3-27}$$

Combined with (3-26), this yields $\sum_{j,k,n=1}^{d}(\alpha_{j,k}^n\bar{\eta}_n + \tilde{\alpha}_{j,k}^n\eta_n)\eta_j\bar{\eta}_k = \mathcal{O}(|(\eta,\bar{\eta})|^4)$ and, since the left-hand side is a polynomial of degree 3 in $(\eta,\bar{\eta})$, it follows that

$$\sum_{j,k,n=1}^{d} (\alpha_{j,k}^n\bar{\eta}_n + \tilde{\alpha}_{j,k}^n\eta_n)\eta_j\bar{\eta}_k = 0 \tag{3-28}$$

for any $\eta \in \mathbb{C}^d$. Hence, uniqueness of coefficients of polynomials of $(\eta,\bar{\eta})$ implies

$$\alpha_{j,k}^n + \alpha_{j,n}^k = 0 \quad \text{for all } j,\ k,\ n \in \{1,\dots,d\}. \tag{3-29}$$

In particular, $\alpha^k_{j,k}=0$. On the other hand, $\tilde\alpha^n_{j,k}=\overline{\alpha^n_{k,j}}$ for all $j,k,n$ since $\widehat{Q}$ is selfadjoint.

Now, we define

$$\check{Q}^0:=\widehat{Q}^0+\mathcal{M}^0 \quad \text{with} \quad \mathcal{M}^0:=\sum_{j_0,k_0=1}^{d}\sum_{n=k_0+1}^{d}\alpha^n_{j_0,k_0}\mathcal{B}^0(j_0,k_0,n;w_0).$$

It follows from symbolic calculus that $\check{Q}^0=\mathrm{Op}(\check{q}^0)$, with $\check{q}^0\in S^0(\mathscr{A})$ given by

$$\begin{aligned}\check{q}^0_{j,k}=w_0\Big(\hat{q}_{0;j,k}-\sum_{n>k}\alpha^n_{j,k}(\xi_n-i\partial_n\phi(x))+\sum_{n<k}\alpha^k_{j,n}(\xi_n-i\partial_n\phi(x))\\ -\sum_{n>j}\overline{\alpha^n_{k,j}}(\xi_n+i\partial_n\phi(x))+\sum_{n<j}\overline{\alpha^j_{k,n}}(\xi_n+i\partial_n\phi(x))\Big)+S^0(h\mathscr{A})\end{aligned}$$

for any $j,k$. Moreover, from (3-29) and $\xi_n+i\partial_n\phi(x)=\eta_n+\mathcal{O}(|x|^2)$ near $(0,0)$, we get

$$\check{q}^0_{j,k}=w_0\Big(\hat{q}_{0;j,k}-\sum_{n=1}^{d}\alpha^n_{j,k}\bar\eta_n-\sum_{n=1}^{d}\overline{\alpha^n_{k,j}}\eta_n+\tilde\rho_{j,k}\Big)+S^0(h\mathscr{A})$$

with $\tilde\rho\in S^0(\mathscr{A})$ such that $\tilde\rho=\mathcal{O}(|(x,\xi)|^2)$ near the origin. Using the identity $\hat{q}_{0;j,k}=\delta_{j,k}+\nu_{j,k}$ together with (3-27), we get

$$\check{q}^0_{j,k}=w_0(\delta_{j,k}+\rho_{j,k})+S^0(h\mathscr{A})$$

with $\rho\in S^0(\mathscr{A})$ such that $\rho=\mathcal{O}(|(x,\xi)|^2)$ near the origin. This implies (3-21)–(3-22) under the assumption (A2′), and achieves the proof of Step 1.

**Step 2.** We now remove the antidiagonal terms away from the origin. More precisely, we show that there exist some $\mathcal{M}^\ell\in\mathcal{M}(\mathscr{A})$ and some diagonal symbols $\tilde{q}^\ell\in S^0(\mathscr{A})$ such that

$$\widetilde{Q}^\ell:=\widehat{Q}^\ell+\mathcal{M}^\ell=\mathrm{Op}(w_\ell\tilde{q}^\ell)+\Psi^0(h\mathscr{A}) \tag{3-30}$$

for any $\ell\in\{1,\dots,d\}$.

For $j_0,k_0,\ell\in\{1,\dots,d\}$ with $j_0\neq k_0$, let $\beta_{j_0,k_0,\ell}$ be defined by

$$\beta_{j_0,k_0,\ell}(x,\xi):=\frac{w_\ell(x,\xi)\hat{q}_{j_0,k_0}(x,\xi)}{|\xi_\ell|^2+|\partial_\ell\phi(x)|^2}.$$

By the support properties of $w_\ell$, we have $\beta_{j_0,k_0,\ell}\in S^0(\langle\xi_{j_0}\rangle^{-1}\langle\xi_{k_0}\rangle^{-1}\langle\xi_\ell\rangle^{-2})$, so $\mathcal{B}^\infty(j_0,k_0,\ell;\beta_{j_0,k_0,\ell})$ belongs to $\mathcal{M}(\mathscr{A})$. Defining

$$\mathcal{M}^\ell:=\sum_{j_0\neq k_0}\mathcal{B}^\infty(j_0,k_0,\ell;\beta_{j_0,k_0,\ell}),$$

the pseudodifferential calculus gives

$$(d^\ell_{\phi,h})^*\,\mathrm{Op}(\beta_{j_0,k_0,\ell})d^\ell_{\phi,h}=\mathrm{Op}(w_\ell\hat{q}_{j_0,k_0})+\Psi^0(h\langle\xi_{j_0}\rangle^{-1}\langle\xi_{k_0}\rangle^{-1}),$$

which implies

$$\widehat{Q}^\ell+\mathcal{M}^\ell=\mathrm{Op}(w_\ell\hat{q})+\mathcal{M}^\ell=\mathrm{Op}(w_\ell\tilde{q}^\ell)+\Psi^0(h\mathscr{A})$$

with $\tilde{q}^\ell \in S^0(\mathscr{A})$ diagonal. This proves (3-30).

**Step 3.** Let us now prove that we can modify each $\widetilde{Q}^\ell$ so that its diagonal coefficients are suitably bounded from below. More precisely, we claim that there exist $c > 0$ and $\widetilde{\mathscr{M}}^\ell \in \mathscr{M}(\mathscr{A})$ such that

$$\breve{Q}^\ell := \widetilde{Q}^\ell + \widetilde{\mathscr{M}}^\ell = \operatorname{Op}(\breve{q}^\ell) + \Psi^0(h\mathscr{A}) \tag{3-31}$$

with $\breve{q}^\ell$ diagonal and $\breve{q}^\ell_{i_0,i_0}(x,\xi) \geq c w_\ell(x,\xi)\langle \xi_{i_0}\rangle^{-2}$ for all $i_0 \in \{1,\dots,d\}$.

For $\ell$, $i_0 \in \{1,\dots,d\}$, let $\beta_{i_0,\ell}$ be defined by

$$\beta_{i_0,\ell}(x,\xi) := \frac{w_\ell(x,\xi)}{2(|\xi_\ell|^2 + |\partial_\ell\phi(x)|^2)}\biggl(\tilde{q}^\ell_{i_0,i_0}(x,\xi) - \frac{\gamma}{1+|\xi_{i_0}|^2 + |\partial_{i_0}\phi(x)|^2}\biggr),$$

where $\gamma > 0$ will be specified later. The symbol $\beta_{i_0,\ell}$ belongs to $S^0(\langle\xi_{i_0}\rangle^{-2}\langle\xi_\ell\rangle^{-2})$, so $\mathscr{B}^\infty(i_0,i_0,\ell;\beta_{i_0,\ell})$ is in $\mathscr{M}(\mathscr{A})$. Defining

$$\widetilde{\mathscr{M}}^\ell := \sum_{i_0\neq\ell} \mathscr{B}^\infty(i_0,i_0,\ell;\beta_{i_0,\ell}),$$

the symbolic calculus shows that $\widetilde{Q}^\ell + \widetilde{\mathscr{M}}^\ell = \operatorname{Op}(\breve{q}^\ell) + \Psi^0(h\mathscr{A})$ with $\breve{q}^\ell$ diagonal and

$$\breve{q}^\ell_{i_0,i_0}(x,\xi) = \frac{\gamma w_\ell(x,\xi)}{1+|\xi_{i_0}|^2 + |\partial_{i_0}\phi(x)|^2} \quad \text{for all } i_0 \neq \ell. \tag{3-32}$$

It remains to prove that we can choose $\gamma > 0$ above, so that $\breve{q}^\ell_{\ell,\ell}(x,\xi) \geq c w_\ell(x,\xi)\langle\xi_\ell\rangle^{-2}$. Thanks to assumption (A1), there exists $\alpha > 0$ such that

$$p_0(x,\xi) \geq \alpha \quad \text{for all } (x,\xi) \in \operatorname{supp}(w_\ell). \tag{3-33}$$

On the other hand, a simple commutator computation shows that $\operatorname{Op}(w_\ell)P_h = d^*_{\phi,h}\widehat{Q}^\ell d_{\phi,h} + \Psi^0(h)$. Combined with (3-20), (3-30) and the definition of $\breve{q}^\ell$, this yields

$$\operatorname{Op}(w_\ell)P_h = d^*_{\phi,h}\widetilde{Q}^\ell d_{\phi,h} + \Psi^0(h) = d^*_{\phi,h}\operatorname{Op}(\breve{q}^\ell)d_{\phi,h} + \Psi^0(h),$$

and then

$$(w_\ell p_0)(x,\xi) = \sum_{i_0=1}^d \breve{q}^\ell_{i_0,i_0}(x,\xi)(|\xi_{i_0}|^2 + |\partial_{i_0}\phi(x)|^2) + S^0(h).$$

Now, using (3-32), we get

$$(w_\ell p_0)(x,\xi) = \breve{q}^\ell_{\ell,\ell}(x,\xi)(|\xi_\ell|^2 + |\partial_\ell\phi(x)|^2) + \gamma(d-1)w_\ell(x,\xi) + S^0(h).$$

Combining this relation with (3-33) and choosing $\gamma = \alpha/(2d)$, we obtain

$$\breve{q}^\ell_{\ell,\ell}(x,\xi) \geq \frac{\alpha w_\ell(x,\xi)}{2(|\xi_\ell|^2 + |\partial_\ell\phi(x)|^2)} + S^0(h\langle\xi_\ell\rangle^{-2}). \tag{3-34}$$

Thus, $\breve{q}^\ell_{\ell,\ell}$ satisfies the required lower bound and (3-31) follows.

**Step 4.** Lastly, we take the square root of the modified operator. Let us define

$$\breve{Q} := \sum_{\ell=0}^{d} \breve{Q}^{\ell} \in \Psi^0(\mathscr{A}), \tag{3-35}$$

with $\breve{Q}^\ell$ defined above. Thanks to (3-20), we have $P_h = (d_{\phi,h})^* \widehat{Q} d_{\phi,h} = (d_{\phi,h})^* \breve{Q} d_{\phi,h}$ and it follows from the preceding constructions that the principal symbol $\breve{q}$ of $\breve{Q}$ satisfies

$$\breve{q}(x,\xi) \geq w_0(x,\xi)(\mathrm{Id} + \mathscr{O}(|(x,\xi)|)) + c \sum_{\ell \geq 1} w_\ell(x,\xi)\,\mathrm{diag}(\langle \xi_j \rangle^{-2}).$$

Shrinking $c > 0$ and the support of $w_0$ if necessary, it follows that

$$\breve{q}(x,\xi) \geq c\,\mathrm{diag}(\langle \xi_j \rangle^{-2}).$$

Letting $E = \mathrm{Op}(\Xi)\breve{Q}\,\mathrm{Op}(\Xi)$, and $e \in S^0(1)$ be the symbol of $E$, the pseudodifferential calculus gives $e(x,\xi;h) = e_0(x,\xi) + S^0(h)$ with

$$e_0(x,\xi) \geq c\,\mathrm{diag}(\langle \xi_j \rangle \langle \xi_j \rangle^{-2} \langle \xi_j \rangle) = c\,\mathrm{Id}, \tag{3-36}$$

so that, for $h > 0$ small enough, $e(x,\xi) \geq \frac{1}{2} c\,\mathrm{Id}$. Hence, we can adapt the proof of Theorem 4.8 of [Helffer and Nier 2005] to our semiclassical setting to get that $F := E^{1/2}$ belongs to $\Psi^0(1)$ and that $F^{-1} \in \Psi^0(1)$. Then, $\breve{Q} = Q^* Q$ with $Q := F\,\mathrm{Op}(\Xi^{-1})$ and, by construction, $Q \in \Psi^0(\Xi\mathscr{A})$.

In addition, as in Theorem 4.8 of [Helffer and Nier 2005], we can show that $F = \mathrm{Op}(e_0^{1/2}) + \Psi^0(h)$, so that $Q = \mathrm{Op}(q_0) + \Psi^0(h\Xi\mathscr{A})$ with $q_0 = e_0^{1/2}\Xi^{-1}$. If, moreover, $\hat{q}$ admits a classical expansion, then $\breve{q} \in S^0_{\mathrm{cl}}(\mathscr{A})$, and the same argument shows that both $e$ and $q$ admit classical expansions.

Let us now study $q_0$ near $(\boldsymbol{u}, 0)$. For $(x,\xi)$ close to $(\boldsymbol{u}, 0)$ we have $\Xi = \mathrm{Id} + \mathscr{O}(|\xi|^2)$ and $\breve{q}_0 = \mathrm{Id} + \rho(x,\xi)$, so

$$e_0(x,\xi) = \Xi \breve{q}_0 \Xi = \mathrm{Id} + \rho(x,\xi) + \mathscr{O}(|\xi|^2),$$

and we get easily $q_0 = e_0^{1/2}\Xi^{-1} = \mathrm{Id} + \mathscr{O}(|\xi|^2 + \rho(x,\xi))$, which proves (3-16) and (3-17). ☐

This completes the proof of Theorem 1.5. ☐

## 4. Quasimodes on $k$-forms and first exponential-type eigenvalue estimates

***Pseudodifferential Hodge–Witten Laplacian on the* 0*-forms.*** This part is devoted to the rough asymptotic of the small eigenvalues of $P_h$ and to the construction of associated quasimodes. From Theorem 1.5, this operator has the expression

$$P_h = a_h d^*_{\phi,h} G d_{\phi,h} a_h, \tag{4-1}$$

where $G$ is the matrix of pseudodifferential operators

$$G = (\mathrm{Op}(g_{j,k}))_{j,k} := Q^* Q = \mathrm{Op}(\Xi)^{-1} F^* F\,\mathrm{Op}(\Xi)^{-1}.$$

Using Corollary 3.1 and that $G$ is selfadjoint, we remark that $g_{j,k} \in S^0(\langle\xi_j\rangle^{-1}\langle\xi_k\rangle^{-1})$ and $g_{j,k} = \overline{g_{k,j}}$. Thus, $P_h$ can be viewed as a Hodge–Witten Laplacian on 0-forms (or a Laplace–Beltrami operator) with the pseudodifferential metric $G^{-1}$. In the following, we will then use the notation $P^{(0)} := P_h$.

Since $a_h(\boldsymbol{u}) = 1 + \mathcal{O}(h)$ and $g(\boldsymbol{u}, 0) = \beta_d \operatorname{Id} + \mathcal{O}(h)$ for all the critical points $\boldsymbol{u} \in \mathcal{U}$, it is natural to consider the operator with the coefficients $a_h$ and $G$ frozen at 1 and $\beta_d \operatorname{Id}$, respectively. For that, let $\Omega^p(\mathbb{R}^d)$, $p = 1, \dots, d$, be the space of $C^\infty$ $p$-forms on $\mathbb{R}^d$. We then define

$$P^W = d^*_{\phi,h} d_{\phi,h} + d_{\phi,h} d^*_{\phi,h}, \tag{4-2}$$

the semiclassical Witten Laplacian on the de Rham complex, and $P^{W,(p)}$, its restriction to the $p$-forms. This operator has been intensively studied (see, e.g., [Helffer and Sjöstrand 1985; Cycon et al. 2008; Bovier et al. 2004; 2005; Helffer et al. 2004]), and a lot is known concerning its spectral properties. In particular, from Lemma 1.6 and Proposition 1.7 of [Helffer and Sjöstrand 1985], we know that there are $n_0$ exponentially small (real nonnegative) eigenvalues, and that the others are above $h/C$.

From [Helffer et al. 2004; Hérau et al. 2011], we have good normalized quasimodes for $P^{W,(0)}$ associated to all minima of $\phi$. For $k \in \{1, \dots, n_0\}$, they are given by

$$f_k^{W,(0)}(x) = \chi_{k,\varepsilon}(x) b_k^{(0)}(h) e^{-(\phi(x)-\phi(\boldsymbol{m}_k))/h},$$

where $b_k^{(0)}(h) = (\pi h)^{-d/4} \det(\operatorname{Hess}\phi(\boldsymbol{m}_k))^{1/4}(1 + \mathcal{O}(h))$, and where the $\chi_{k,\varepsilon}$ are cutoff functions localized in sufficiently large areas containing $\boldsymbol{m}_k \in \mathcal{U}^{(0)}$. In fact, we need large support (associated to level sets of $\phi$) and properties for the cutoff functions $\chi_{k,\varepsilon}$, so that the refined analysis of the next section can be done. We postpone to the Appendix the construction of the cutoff functions, the definition of $\varepsilon > 0$, refined estimates on this family $(f_k^{W,(0)})_k$, and in particular the fact that it is a quasiorthonormal free family of functions, following closely [Helffer et al. 2004; Hérau et al. 2011].

We now define the quasimodes associated to $P^{(0)}$ in the following way:

$$f_k^{(0)}(x) := a_h(x)^{-1} f_k^{W,(0)}(x) = a_h(x)^{-1} b_k^{(0)}(h) \chi_{k,\varepsilon}(x) e^{-(\phi(x)-\phi(\boldsymbol{m}_k))/h} \tag{4-3}$$

for $1 \le k \le n_0$. We then have:

**Lemma 4.1.** *The system $(f_k^{(0)})_k$ is free, and there exists $\alpha > 0$ independent of $\varepsilon$ such that*

$$\langle f_k^{(0)}, f_{k'}^{(0)}\rangle = \delta_{k,k'} + \mathcal{O}(h) \quad \textit{and} \quad P^{(0)} f_k^{(0)} = \mathcal{O}(e^{-\alpha/h}).$$

**Remark 4.2.** For this result to be true, it would have been sufficient to take truncation functions with smaller support (say in a small neighborhood of each minimum $\boldsymbol{m}_k$). We emphasize again that the more complicated construction for the quasimodes is justified by their later use.

*Proof.* First, observe that

$$\langle f_k^{(0)}, f_{k'}^{(0)}\rangle = \langle a_h^{-2} f_k^{W,(0)}, f_{k'}^{W,(0)}\rangle = \delta_{kk'} + \langle (a_h^{-2} - 1) f_k^{W,(0)}, f_{k'}^{W,(0)}\rangle.$$

Moreover, near any minimum $\boldsymbol{m}_k$, $a_h^{-2} - 1 = \mathcal{O}(h + |x - \boldsymbol{m}_k|^2)$ and $\phi(x) - \phi(\boldsymbol{m}_k)$ is quadratic, so

$$\|(a_h^{-2} - 1) f_k^{W,(0)}\| = \mathcal{O}(h), \tag{4-4}$$

which proves the first statement. For the last statement, it is enough to notice that

$$P^{(0)} f_k^{(0)} = a_h^* d_{\phi,h}^* Q^* Q d_{\phi,h} f_k^{W,(0)}$$

and apply Lemma A.3. □

We now prove a first rough spectral result on $P^{(0)}$, using the preceding lemma.

**Proposition 4.3.** *The operator $P^{(0)}$ has exactly $n_0$ exponentially small (real nonnegative) eigenvalues, and the remaining part of its spectrum is in $[\varepsilon_0 h, +\infty[$ for some $\varepsilon_0 > 0$.*

Usually, this type of result is a consequence of an IMS formula. It is possible to do that here (with effort) but we prefer to give a simpler proof using what we know about $P^{W,(0)}$. The following proof is based on the spectral theorem and the maxi-min principle.

*Proof.* Thanks to Proposition 2.4, the spectrum of $P^{(0)}$ is discrete in $[0, \delta]$ and its $j$-th eigenvalue is given by

$$\sup_{\dim E = j-1} \inf_{u \in E^\perp,\, \|u\|=1} \langle P^{(0)} u, u \rangle. \tag{4-5}$$

Lemma 4.1 directly implies

$$\langle P^{(0)} f_k^{(0)}, f_{k'}^{(0)} \rangle \le \| P^{(0)} f_k^{(0)} \| \| f_{k'}^{(0)} \| = \mathcal{O}(e^{-\alpha/h})$$

for some $\alpha > 0$. Using the almost orthogonality of the $f_k^{(0)}$, (4-5) and $P^{(0)} \ge 0$, we deduce that $P^{(0)}$ has at least $n_0$ eigenvalues that are exponentially small.

We now want to prove that the remaining part of the spectrum of $P^{(0)}$ is above $\varepsilon_0 h$ for some $\varepsilon_0 > 0$ small enough. For this, we set

$$\mathscr{E} := \operatorname{Vect}\{ f_k^{W,(0)} : k = 1, \dots, n_0 \},$$

and we consider $u \in a_h^{-1} \mathscr{E}^\perp$ with $\|u\| = 1$. We have, again,

$$\langle P^{(0)} u, u \rangle = \| F \operatorname{Op}(\Xi^{-1}) d_{\phi,h} a_h u \|^2 \ge \varepsilon_0 \| \operatorname{Op}(\Xi^{-1}) d_{\phi,h} a_h u \|^2 \tag{4-6}$$

for some $\varepsilon_0 > 0$ independent of $h$, which may change from line to line. For the last inequality, we have used that $\|F^{-1}\|$ is uniformly bounded since $F^{-1} \in \Psi^0(1)$. On the other hand, using $0 \le P^{W,(1)} = -h^2 \Delta \otimes \operatorname{Id} + \mathcal{O}(1)$, we notice that

$$\operatorname{Op}(\Xi^{-1})^2 \ge (-h^2 \Delta + 1)^{-1} \otimes \operatorname{Id} \ge \varepsilon_0 (P^{W,(1)} + 1)^{-1}$$

for some (other) $\varepsilon_0 > 0$. Therefore, using the classical intertwining relations

$$(P^{W,(1)} + 1)^{-1/2} d_{\phi,h} = d_{\phi,h} (P^{W,(0)} + 1)^{-1/2},$$

and the fact that $P^{W,(0)} = d_{\phi,h}^* d_{\phi,h}$ on 0-forms, we get

$$\begin{aligned} \langle P^{(0)} u, u \rangle \ge \varepsilon_0 \| (P^{W,(1)} + 1)^{-1/2} d_{\phi,h} a_h u \|^2 &= \varepsilon_0 \| d_{\phi,h} (P^{W,(0)} + 1)^{-1/2} a_h u \|^2 \\ &= \varepsilon_0 \langle P^{W,(0)} (P^{W,(0)} + 1)^{-1} a_h u, a_h u \rangle. \end{aligned} \tag{4-7}$$

Now, let $\mathscr{F}$ be the eigenspace of $P^{W,(0)}$ associated to the $n_0$ exponentially small eigenvalues, and let $\Pi_{\mathscr{E}}$ (resp. $\Pi_{\mathscr{F}}$) be the orthogonal projectors onto $\mathscr{E}$ (resp. $\mathscr{F}$). Then, from Proposition 1.7 of [Helffer and Sjöstrand 1985] (see also Theorem 2.4 of [Helffer and Sjöstrand 1984]), we have $\|\Pi_{\mathscr{E}} - \Pi_{\mathscr{F}}\| = \mathcal{O}(e^{-\alpha/h})$. Moreover, since the $(n_0+1)$-st eigenvalue of $P^{W,(0)}$ is of order $h$, the spectral theorem gives

$$P^{W,(0)}(P^{W,(0)}+1)^{-1} \geq \varepsilon_0 h(1-\Pi_{\mathscr{F}}) + \mathcal{O}(e^{-\alpha/h}) \geq \varepsilon_0 h(1-\Pi_{\mathscr{E}}) + \mathcal{O}(e^{-\alpha/h}).$$

Then, using $a_h u \in \mathscr{E}^{\perp}$, $\|u\| = 1$ and Lemma 2.2, (4-7) becomes

$$\langle P^{(0)}u, u\rangle \geq \varepsilon_0 h\|a_h u\|^2 + \mathcal{O}(e^{-\alpha/h}) \geq \tfrac{1}{2}c_1\varepsilon_0 h.$$

Finally, this estimate and (4-5) imply that $P^{(0)}$ has at most $n_0$ eigenvalues below $\frac{1}{2}c_1\varepsilon_0 h$. Taking $\frac{1}{2}c_1\varepsilon_0$ as the new value of $\varepsilon_0$ gives the result. □

***Pseudodifferential Hodge–Witten Laplacian on the* 1*-forms.*** Since we want to follow a supersymmetric approach to prove the main theorem of this paper, we have to build an extension $P^{(1)}$ of $P^{(0)}$ defined on 1-forms which satisfies properties similar to those of $P^{W,(1)}$. To do this, we use the following coordinates for $\omega \in \Omega^1(\mathbb{R}^d)$ and $\sigma \in \Omega^2(\mathbb{R}^d)$:

$$\omega = \sum_{j=1}^{d} \omega_j(x)dx_j, \quad \sigma = \sum_{j<k} \sigma_{j,k}(x)dx_j \wedge dx_k,$$

and we extend the matrix $\sigma_{j,k}$ as a function with values in the space of antisymmetric matrices. Recall that the exterior derivative satisfies

$$(d^{(1)}\omega)_{j,k} = \partial_{x_j}\omega_k - \partial_{x_k}\omega_j \quad \text{and} \quad (d^{*(1)}\sigma)_j = -\sum_k \partial_{x_k}\sigma_{k,j}. \tag{4-8}$$

In the previous section, we saw that $P^{(0)}$ can be viewed as the Hodge–Witten Laplacian on 0-forms with a pseudodifferential metric $G^{-1}$. It is then natural to consider the corresponding Hodge–Witten Laplacian on 1-forms. Thus, mimicking the construction in the standard case, we define

$$P^{(1)} := Qd_{\phi,h}a_h^2 d_{\phi,h}^* Q^* + (Q^{-1})^* d_{\phi,h}^* M d_{\phi,h} Q^{-1}, \tag{4-9}$$

where $M$ is the linear operator acting on $\Omega^2(\mathbb{R}^d)$ with coefficients

$$M_{(j,k),(a,b)} := \tfrac{1}{2}\operatorname{Op}(a_h^2(g_{j,a}g_{k,b} - g_{k,a}g_{j,b})). \tag{4-10}$$

Note that $M$ is well-defined on $\Omega^2(\mathbb{R}^d)$ (i.e., $M\sigma$ is antisymmetric if $\sigma$ is antisymmetric) since $M_{(k,j),(a,b)} = M_{(j,k),(b,a)} = -M_{(j,k),(a,b)}$. Furthermore, we deduce from the properties of $g_{j,k}$ that

$$M_{(j,k),(a,b)} \in \Psi^0(\langle\xi_j\rangle^{-1}\langle\xi_k\rangle^{-1}\langle\xi_a\rangle^{-1}\langle\xi_b\rangle^{-1}). \tag{4-11}$$

**Remark 4.4.** When $G^{-1}$ is a true metric (and not a matrix of pseudodifferential operators), the operator $P^{(1)}$ defined in (4-9) is the usual Hodge–Witten Laplacian on 1-forms. Our construction is then an extension to the pseudodifferential case. Generalizing these structures to $p$-forms, it should be possible to define a Hodge–Witten Laplacian on the total de Rham complex. It could also be possible to define such an operator using only abstract geometric quantities (and not explicit formulas like (4-10)).

On the other hand, a precise choice for the operator $M$ is not relevant in the present paper. Indeed, for the study of the small eigenvalues of $P^{(0)}$, only the first part (in (4-9)) of $P^{(1)}$ is important (see Lemma 4.7 below). The second part is only used to make the operator $P^{(1)}$ elliptic. Thus, any $M$ satisfying (4-11) and $M_{(j,k),(a,b)} \geq \varepsilon \operatorname{Op}(\langle \xi_j \rangle^{-2} \langle \xi_k \rangle^{-2}) \otimes \operatorname{Id}$ should probably work.

We first show that $P^{(1)}$ acts diagonally (at the first order), as is the case for $P^{W,(1)}$.

**Lemma 4.5.** *The operator $P^{(1)} \in \Psi^0(1)$ is selfadjoint on $\Omega^1(\mathbb{R}^d)$. Moreover,*

$$P^{(1)} = P^{(0)} \otimes \operatorname{Id} + \Psi^0(h). \tag{4-12}$$

*Proof.* We begin by estimating the first part of $P^{(1)}$,

$$P_1^{(1)} := Q d_{\phi,h} a_h^2 d_{\phi,h}^* Q^*.$$

Let $q_{j,k} \in S^0(\langle \xi_k \rangle^{-1})$ denote the symbol of the coefficients of $Q$ and let $d_{\phi,h}^j = h\partial_j + (\partial_j \phi)$. Using the composition rules of matrices, a direct computation gives

$$(P_1^{(1)})_{j,k} = \sum_a \operatorname{Op}(q_{j,a}) d_{\phi,h}^a a_h^2 (d_{\phi,h}^* Q^*)_k = \sum_{a,b} \operatorname{Op}(q_{j,a}) d_{\phi,h}^a a_h^2 (d_{\phi,h}^b)^* \operatorname{Op}(\overline{q_{k,b}}). \tag{4-13}$$

We then deduce that $P_1^{(1)}$ is a selfadjoint operator on $\Omega^1(\mathbb{R}^d)$ with coefficients of class $\Psi^0(1)$. Moreover, this formula implies

$$(P_1^{(1)})_{j,k} = \sum_{a,b} \operatorname{Op}(a_h^2 q_{j,a} \overline{q_{k,b}}) d_{\phi,h}^a (d_{\phi,h}^b)^* + \Psi^0(h). \tag{4-14}$$

It remains to study

$$P_2^{(1)} := (Q^{-1})^* d_{\phi,h}^* M d_{\phi,h} Q^{-1}.$$

Let $q_{j,k}^{-1} \in S^0(\langle \xi_j \rangle)$ denote the symbol of the coefficients of $Q^{-1}$. The formulas of (4-8), the definition (4-10) and the composition rules of matrices imply

$$\begin{aligned}
(P_2^{(1)})_{j,k} &= \sum_\alpha \operatorname{Op}(\overline{q_{\alpha,j}^{-1}}) (d_{\phi,h}^* M d_{\phi,h} Q^{-1})_{\alpha,k} \\
&= -\sum_{a,\alpha} \operatorname{Op}(\overline{q_{\alpha,j}^{-1}}) (d_{\phi,h}^a)^* (M d_{\phi,h} Q^{-1})_{(a,\alpha),k} \\
&= -\sum_{a,b,\alpha,\beta} \operatorname{Op}(\overline{q_{\alpha,j}^{-1}}) (d_{\phi,h}^a)^* M_{(a,\alpha),(b,\beta)} (d_{\phi,h} Q^{-1})_{(b,\beta),k} \\
&= -\sum_{a,b,\alpha,\beta} \operatorname{Op}(\overline{q_{\alpha,j}^{-1}}) (d_{\phi,h}^a)^* M_{(a,\alpha),(b,\beta)} (d_{\phi,h}^b \operatorname{Op}(q_{\beta,k}^{-1}) - d_{\phi,h}^\beta \operatorname{Op}(q_{b,k}^{-1})) \\
&= -2 \sum_{a,b,\alpha,\beta} \operatorname{Op}(\overline{q_{\alpha,j}^{-1}}) (d_{\phi,h}^a)^* M_{(a,\alpha),(b,\beta)} d_{\phi,h}^b \operatorname{Op}(q_{\beta,k}^{-1}),
\end{aligned} \tag{4-15}$$

where we have used that $M_{(a,\alpha),(b,\beta)} = -M_{(a,\alpha),(\beta,b)}$. By (4-11), a typical term of these sums satisfies

$$\operatorname{Op}(\overline{q_{\alpha,j}^{-1}})(d_{\phi,h}^a)^* M_{(a,\alpha),(b,\beta)} d_{\phi,h}^b \operatorname{Op}(q_{\beta,k}^{-1}) \in \Psi^0(\langle \xi_\alpha\rangle \langle \xi_a\rangle \langle \xi_a\rangle^{-1} \langle \xi_\alpha\rangle^{-1} \langle \xi_b\rangle^{-1} \langle \xi_\beta\rangle^{-1} \langle \xi_b\rangle \langle \xi_\beta\rangle),$$

and then $P_2^{(1)} \in \Psi^0(1)$. On the other hand, using $g_{j,k} = \overline{g_{k,j}}$ and (4-10), we get

$$\begin{aligned}
(P_2^{(1)})_{j,k}^* &= -\sum_{a,b,\alpha,\beta} \operatorname{Op}(\overline{q_{\beta,k}^{-1}})(d_{\phi,h}^b)^* \operatorname{Op}(\overline{a_h^2(g_{a,b}g_{\alpha,\beta} - g_{\alpha,b}g_{a,\beta})}) d_{\phi,h}^a \operatorname{Op}(q_{\alpha,j}^{-1}) \\
&= -\sum_{a,b,\alpha,\beta} \operatorname{Op}(\overline{q_{\beta,k}^{-1}})(d_{\phi,h}^b)^* \operatorname{Op}(a_h^2(g_{b,a}g_{\beta,\alpha} - g_{b,\alpha}g_{\beta,a})) d_{\phi,h}^a \operatorname{Op}(q_{\alpha,j}^{-1}) \\
&= -\sum_{a,b,\alpha,\beta} \operatorname{Op}(\overline{q_{\alpha,k}^{-1}})(d_{\phi,h}^a)^* \operatorname{Op}(a_h^2(g_{a,b}g_{\alpha,\beta} - g_{a,\beta}g_{\alpha,b})) d_{\phi,h}^b \operatorname{Op}(q_{\beta,j}^{-1}) = (P_2^{(1)})_{k,j},
\end{aligned}$$

so that $P_2^{(1)}$ is selfadjoint on $\Omega^1(\mathbb{R}^d)$. Finally, (4-11) and (4-15) yield

$$\begin{aligned}
(P_2^{(1)})_{j,k} &= \sum_{a,b} \operatorname{Op}\biggl(a_h^2 \sum_{\alpha,\beta} \overline{q_{\alpha,j}^{-1}} q_{\beta,k}^{-1} (g_{a,b}g_{\alpha,\beta} - g_{a,\beta}g_{\alpha,b})\biggr)(d_{\phi,h}^a)^* d_{\phi,h}^b + \Psi^0(h) \\
&= \sum_{a,b} \operatorname{Op}(a_h^2 g_{a,b}\delta_{j,k} - a_h^2 q_{j,b}\overline{q_{k,a}})(d_{\phi,h}^a)^* d_{\phi,h}^b + \Psi^0(h), \qquad \text{(4-16)}
\end{aligned}$$

since

$$\sum_j g_{a,j} q_{j,b}^{-1} = \overline{q_{b,a}} + S^0(h\langle \xi_a\rangle^{-1}) \quad \text{and} \quad \sum_j q_{a,j} q_{j,b}^{-1} = \delta_{a,b} + S^0(h),$$

which follow from $GQ^{-1} = Q^*$ and $QQ^{-1} = \mathrm{Id}$.

Summing up the previous properties of $P_\bullet^{(1)}$, the operator $P^{(1)} = P_1^{(1)} + P_2^{(1)} \in \Psi^0(1)$ is selfadjoint on $\Omega^1(\mathbb{R}^d)$. Lastly, combining (4-14) and (4-16), we obtain

$$\begin{aligned}
P^{(1)} &= \sum_{a,b} (d_{\phi,h}^a)^* \operatorname{Op}(a_h^2 g_{a,b}) d_{\phi,h}^b \otimes \mathrm{Id} + \Psi^0(h) = a_h d_{\phi,h}^* G d_{\phi,h} a_h \otimes \mathrm{Id} + \Psi^0(h) \\
&= P^{(0)} \otimes \mathrm{Id} + \Psi^0(h), \qquad \text{(4-17)}
\end{aligned}$$

and the lemma follows. □

The next result compares $P^{(1)}$ and $P^{W,(1)}$.

**Lemma 4.6.** *There exist some pseudodifferential operators $(R_k)_{k=0,1,2}$ such that*

$$P^{(1)} = \beta_d P^{W,(1)} + R_0 + R_1 + R_2,$$

*where the remainder terms enjoy the following properties*:

(i) *$R_0$ is a $d \times d$ matrix whose coefficients are finite sums of terms of the form*

$$(d_{\phi,h}^a)^* (\operatorname{Op}(r_0) + \Psi^0(h)) d_{\phi,h}^b$$

*with $a, b \in \{1, \dots, d\}$ and $r_0 \in S^0(1)$ satisfying $r_0(x,\xi) = \mathcal{O}(|(x - \boldsymbol{u}, \xi)|^2)$ near $(\boldsymbol{u}, 0)$, $\boldsymbol{u} \in \mathcal{U}$;*

(ii) *$R_1$ is a matrix whose coefficients are finite sums of terms of the form $h\,\mathrm{Op}(r_1)d^a_{\phi,h}$ or $h(d^a_{\phi,h})^*\,\mathrm{Op}(r_1)$ with $a \in \{1,\dots,d\}$ and $r_1 \in S^0(1)$ satisfying $r_1(x,\xi) = \mathcal{O}(|(x-\boldsymbol{u},\xi)|)$ near $(\boldsymbol{u},0)$, $\boldsymbol{u} \in \mathcal{U}$;*

(iii) $R_2 \in \Psi^0(h^2)$.

*Proof.* As in the proof of Lemma 4.5, we use the decomposition $P^{(1)} = P^{(1)}_1 + P^{(1)}_2$. From Corollary 3.1 and Lemma 2.2, the coefficients appearing in these operators satisfy

$$\begin{aligned} a_h &= \tilde a + S^0(h) \in S^0(1),\\ q_{a,b} &= \tilde q_{a,b} + S^0(h\langle\xi_b\rangle^{-1}) \in S^0(\langle\xi_b\rangle^{-1}),\\ q^{-1}_{a,b} &= \tilde q^{-1}_{a,b} + S^0(h\langle\xi_a\rangle) \in S^0(\langle\xi_a\rangle),\\ M_{(j,k),(a,b)} &= \mathrm{Op}(\tilde m_{(j,k),(a,b)}) + \Psi^0(h\langle\xi_j\rangle^{-1}\langle\xi_k\rangle^{-1}\langle\xi_a\rangle^{-1}\langle\xi_b\rangle^{-1}) \end{aligned}$$

with $\tilde m_{(j,k),(a,b)} \in S^0(\langle\xi_j\rangle^{-1}\langle\xi_k\rangle^{-1}\langle\xi_a\rangle^{-1}\langle\xi_b\rangle^{-1})$ and

$$\tilde a = 1 + \mathcal{O}(|(x-\boldsymbol{u},\xi)|^2), \qquad \tilde m_{(j,k),(a,b)} = \tfrac12\beta_d^2(\delta_{j,a}\delta_{k,b} - \delta_{k,a}\delta_{j,b}) + \mathcal{O}(|(x-\boldsymbol{u},\xi)|^2),$$
$$\tilde q_{a,b} = \beta_d^{1/2}\delta_{a,b} + \mathcal{O}(|(x-\boldsymbol{u},\xi)|^2), \qquad \tilde q^{-1}_{a,b} = \beta_d^{-1/2}\delta_{a,b} + \mathcal{O}(|(x-\boldsymbol{u},\xi)|^2)$$

near $(\boldsymbol{u},0)$, $\boldsymbol{u} \in \mathcal{U}$. Then, making commutations in (4-13) and (4-15), we obtain the desired result. ☐

We now make the link between the eigenvalues of $P^{(0)}$ and $P^{(1)}$. For that, we will use the so-called intertwining relations, which are a fundamental tool in the supersymmetric approach. Recall that, thanks to Theorem 1.5, $P^{(0)}$ can be written as

$$P^{(0)} = L^*_\phi L_\phi \quad \text{with} \quad L_\phi = Q d_{\phi,h} a_h. \tag{4-18}$$

We obtain the following result:

**Lemma 4.7.** *On 0-forms, we have*

$$L_\phi P^{(0)} = P^{(1)} L_\phi = L_\phi L^*_\phi L_\phi. \tag{4-19}$$

*Moreover, for all $\lambda \in \mathbb{R}\setminus\{0\}$, the operator $L_\phi : \ker(P^{(0)}-\lambda) \to \ker(P^{(1)}-\lambda)$ is injective. Finally, $L_\phi(\ker(P^{(0)})) = \{0\}$.*

*Proof.* Let us first prove (4-19). Using (4-9), (4-18) and the usual cohomology rule (i.e., $d^2_{\phi,h} = 0$), we have

$$\begin{aligned} P^{(1)}L_\phi &= L_\phi L^*_\phi L_\phi + (Q^{-1})^* d^*_{\phi,h} M d_{\phi,h} Q^{-1} Q d_{\phi,h} a_h\\ &= L_\phi L^*_\phi L_\phi + (Q^{-1})^* d^*_{\phi,h} M d_{\phi,h} d_{\phi,h} a_h\\ &= L_\phi L^*_\phi L_\phi = L_\phi P^{(0)}. \end{aligned} \tag{4-20}$$

Now, let $u \neq 0$ be an eigenfunction of $P^{(0)}$ associated to $\lambda \in \mathbb{R}$. In particular, $\|L_\phi u\|^2 = \lambda\|u\|^2$ vanishes if and only if $\lambda = 0$. Moreover, (4-19) yields

$$P^{(1)} L_\phi u = L_\phi P^{(0)} u = \lambda L_\phi u.$$

This implies the second part of the lemma. ☐

We shall now study more precisely the small eigenvalues of $P^{(1)}$. Recall that $\boldsymbol{s}_j$, $j = 2, \dots, n_1 + 1$, denote the saddle points (of index 1) of $\phi$. Again, we will stick to the analysis already made for the Witten Laplacian on 1-forms $P^{W,(1)}$, for which we recall the following properties. From Lemma 1.6 and Proposition 1.7 of [Helffer and Sjöstrand 1985], the operator $P^{W,(1)}$ is selfadjoint, positive and has exactly $n_1$ exponentially small (nonzero) eigenvalues (counted with multiplicities). We next recall the construction of associated quasimodes made in Definition 4.3 of [Helffer et al. 2004]. Let $u_j$ denote a normalized fundamental state of $P^{W,(1)}$ restricted to an appropriate neighborhood of $\boldsymbol{s}_j$ with Dirichlet boundary conditions. The quasimodes $f_j^{W,(1)}$ are then defined by

$$f_j^{W,(1)} := \|\theta_j u_j\|^{-1} \theta_j(x) u_j(x), \tag{4-21}$$

where $\theta$ is a well-chosen $C_0^\infty$ localization function around $\boldsymbol{s}_j$. Since the $f_j^{W,(1)}$ have disjoint support, we immediately deduce

$$\langle f_j^{W,(1)}, f_{j'}^{W,(1)} \rangle = \delta_{j,j'}. \tag{4-22}$$

In particular, the family $\{f_j^{W,(1)} : j = 2, \dots, n_1 + 1\}$ is a free family of 1-forms. Furthermore, Theorem 1.4 of [Helffer and Sjöstrand 1985] implies that these quasimodes have a WKB expression,

$$f_j^{W,(1)}(x) = \theta_j(x) b_j^{(1)}(x,h) e^{-\phi_{+,j}(x)/h}, \tag{4-23}$$

where $b_j^{(1)}(x,h)$ is a normalization 1-form having a semiclassical asymptotic, and $\phi_{+,j}$ is the phase associated to the outgoing manifold of $\xi^2 + |\nabla_x \phi(x)|^2$ at $(\boldsymbol{s}_j, 0)$. Moreover, the phase function $\phi_{+,j}$ satisfies the eikonal equation $|\nabla_x \phi_{+,j}|^2 = |\nabla_x \phi|^2$ and $\phi_{+,j}(x) \sim |x - \boldsymbol{s}_j|^2$ near $\boldsymbol{s}_j$. For other properties of $\phi_{+,j}$, we refer to [Helffer and Sjöstrand 1985]. On the other hand, Lemma 1.6 and Proposition 1.7 of [Helffer and Sjöstrand 1985] imply that there exists $\alpha > 0$ independent of $\varepsilon$ such that

$$P^{W,(1)} f_j^{W,(1)} = \mathcal{O}(e^{-\alpha/h}). \tag{4-24}$$

Lastly, we deduce from Proposition 1.7 of [Helffer and Sjöstrand 1985] that there exists $\nu > 0$ such that

$$\langle P^{W,(1)} u, u \rangle \geq \nu h \|u\|^2 \tag{4-25}$$

for all $u \perp \operatorname{Vect}\{f_j^{W,(1)} : j = 2, \dots, n_1 + 1\}$.

Now, let us define the quasimodes associated to $P^{(1)}$ by

$$f_j^{(1)}(x) := \beta_d^{1/2} (Q^*)^{-1} f_j^{W,(1)} \tag{4-26}$$

for $2 \leq j \leq n_1 + 1$. Note that this is possible since $(Q^*)^{-1} \in \Psi^0(\langle \xi \rangle)$. Using that $(Q^*)^{-1}$ is close to $\beta_d^{-1/2} \operatorname{Id}$ microlocally near $(\boldsymbol{s}_j, 0)$, we will prove that they form a good, approximately normalized and orthogonal family of quasimodes for $P^{(1)}$.

**Lemma 4.8.** *The system $(f_j^{(1)})_j$ is free and, for all $j, j' = 2, \dots, n_1 + 1$, we have*

$$\|f_j^{(1)} - f_j^{W,(1)}\| = \mathcal{O}(h), \quad \langle f_j^{(1)}, f_{j'}^{(1)} \rangle = \delta_{j,j'} + \mathcal{O}(h) \quad \textit{and} \quad P^{(1)} f_j^{(1)} = \mathcal{O}(h^2).$$

*Proof.* From (4-26) and Corollary 3.1, we have

$$f_j^{(1)} - f_j^{W,(1)} = (\beta_d^{1/2}(Q^*)^{-1} - \mathrm{Id}) f_j^{W,(1)} = \mathrm{Op}(r) f_j^{W,(1)}$$

with $r \in S^0(\langle\xi\rangle^2)$ such that, modulo $S^0(h\langle\xi\rangle^2)$, $r(x,\xi) = \mathcal{O}(|(x-\boldsymbol{u},\xi)|^2)$ near $(\boldsymbol{u},0)$, $\boldsymbol{u} \in \mathcal{U}$. Moreover, using Taylor expansion and symbolic calculus, we can write

$$r(x,\xi) = \sum_{|\alpha+\beta|\in\{0,2\}} h^{1-|\alpha+\beta|/2} r_{\alpha,\beta}(x,\xi)(x-\boldsymbol{s}_j)^\alpha \xi^\beta$$

with $r_{\alpha,\beta} \in S^0(\langle\xi\rangle^2)$. Combined with the WKB form of the $f_j^{W,(1)}$ given in (4-23) (and, in particular, with $\phi_{+,j}(x) \sim |x-\boldsymbol{s}_j|^2$ near $\boldsymbol{s}_j$), it shows that

$$\mathrm{Op}(r) f_j^{W,(1)} = \mathcal{O}(h), \tag{4-27}$$

which proves the first statement.

The second statement is a direct consequence of the above estimate and (4-22).

For the last estimate, we follow the same strategy. Thanks to Lemma 4.6, we have

$$P^{(1)} f_j^{(1)} = \beta_d P^{W,(1)} f_j^{W,(1)} + \beta_d P^{W,(1)}(f_j^{(1)} - f_j^{W,(1)}) + R_0 f_j^{(1)} + R_1 f_j^{(1)} + R_2 f_j^{(1)}. \tag{4-28}$$

Proceeding as above, we write

$$P^{W,(1)}(f_j^{(1)} - f_j^{W,(1)}) = P^{W,(1)}(\beta_d^{1/2}(Q^*)^{-1} - \mathrm{Id}) f_j^{W,(1)},$$

where, using (4-2), Corollary 3.1 and the pseudodifferential calculus, the corresponding operator can be decomposed as

$$P^{W,(1)}(\beta_d^{1/2}(Q^*)^{-1} - \mathrm{Id}) = \mathrm{Op}\biggl(\sum_{|\alpha+\beta|\in\{0,2,4\}} h^{2-|\alpha+\beta|/2} \tilde{r}_{\alpha,\beta}(x,\xi;h)(x-\boldsymbol{s}_j)^\alpha \xi^\beta\biggr)$$

for some $\tilde{r}_{\alpha,\beta} \in S^0(\langle\xi\rangle^3)$. Thus, as in (4-27), we deduce

$$\beta_d P^{W,(1)}(f_j^{(1)} - f_j^{W,(1)}) = \mathcal{O}(h^2). \tag{4-29}$$

In the same way, we deduce from Lemma 4.6 that, for any $p = 0, 1, 2$,

$$R_p \beta_d^{1/2}(Q^*)^{-1} = \mathrm{Op}\biggl(\sum_{|\alpha+\beta|\in\{0,2,4\}} h^{2-|\alpha+\beta|/2} r^p_{\alpha,\beta}(x,\xi;h)(x-\boldsymbol{s}_j)^\alpha \xi^\beta\biggr)$$

with $r^p_{\alpha,\beta} \in S^0(\langle\xi\rangle^3)$. Thus,

$$R_p f_j^{(1)} = R_p \beta_d^{1/2}(Q^*)^{-1} f_j^{W,(1)} = \mathcal{O}(h^2). \tag{4-30}$$

Combining (4-28) with the estimates (4-24), (4-29) and (4-30), we obtain $P^{(1)} f_j^{(1)} = \mathcal{O}(h^2)$ and this concludes the proof of the lemma. □

The following proposition is the analogue of Proposition 4.3.

**Proposition 4.9.** *The operator $P^{(1)}$ has exactly $n_1$ $\mathcal{O}(h^2)$ (real) eigenvalues, and the remaining part of the spectrum is in $[\varepsilon_1 h, +\infty[$ for some $\varepsilon_1 > 0$.*

The idea of the proof is to consider separately the regions of the phase space close to the critical points $\mathcal{U}$ and away from this set. In the first one, we approximate $P^{(1)}$ by $P^{W,(1)}$ using that $Q \simeq \beta_d^{1/2}\,\mathrm{Id}$ microlocally near $(\boldsymbol{u}, 0)$, $\boldsymbol{u} \in \mathcal{U}$. In the second one, we use that (the symbol of) $P^{(1)}$ is elliptic by (4-12).

We start this strategy with a pseudodifferential IMS formula. For $\eta > 0$ fixed, let $\chi_0 \in C_0^\infty(\mathbb{R}^{2d}; [0, 1])$ be supported in a neighborhood of size $\eta$ of $\mathcal{U}$ and such that $\chi_0 = 1$ near $\mathcal{U}$ and $\chi_\infty := (1-\chi_0^2)^{1/2} \in C^\infty(\mathbb{R}^{2d})$. In particular,

$$\chi_0^2(x, \xi) + \chi_\infty^2(x, \xi) = 1 \quad \text{for all } (x, \xi) \in \mathbb{R}^{2d}. \tag{4-31}$$

In the sequel, the remainder terms may depend on $\eta$, but $C$ will denote a positive constant independent of $\eta$, which may change from line to line. Using Lemma 4.5 and the shorthand $\mathrm{Op}(a) = \mathrm{Op}(a) \otimes \mathrm{Id}$, the pseudodifferential calculus gives

$$\begin{aligned}
P^{(1)} &= \tfrac{1}{2}(\mathrm{Op}(\chi_0^2 + \chi_\infty^2) P^{(1)} + P^{(1)} \mathrm{Op}(\chi_0^2 + \chi_\infty^2)) \\
&= \tfrac{1}{2}(\mathrm{Op}(\chi_0)^2 P^{(1)} + P^{(1)} \mathrm{Op}(\chi_0)^2) + \tfrac{1}{2}(\mathrm{Op}(\chi_\infty)^2 P^{(1)} + P^{(1)} \mathrm{Op}(\chi_\infty)^2) + \Psi^0(h^2) \\
&= \mathrm{Op}(\chi_0) P^{(1)} \mathrm{Op}(\chi_0) + \mathrm{Op}(\chi_\infty) P^{(1)} \mathrm{Op}(\chi_\infty) \\
&\qquad + \tfrac{1}{2}[\mathrm{Op}(\chi_0), [\mathrm{Op}(\chi_0), P^{(1)}]] + \tfrac{1}{2}[\mathrm{Op}(\chi_\infty), [\mathrm{Op}(\chi_\infty), P^{(1)}]] + \mathcal{O}(h^2) \\
&= \mathrm{Op}(\chi_0) P^{(1)} \mathrm{Op}(\chi_0) + \mathrm{Op}(\chi_\infty) P^{(1)} \mathrm{Op}(\chi_\infty) + \mathcal{O}(h^2).
\end{aligned} \tag{4-32}$$

In the previous estimate, we have crucially used that $\mathrm{Op}(\chi_\bullet) \otimes \mathrm{Id}$ are matrices of pseudodifferential operators collinear to the identity.

**Lemma 4.10.** *There exists $\delta_\eta > 0$, which may depend on $\eta$, such that*

$$\mathrm{Op}(\chi_\infty) P^{(1)} \mathrm{Op}(\chi_\infty) \geq \delta_\eta \mathrm{Op}(\chi_\infty)^2 + \mathcal{O}(h^\infty). \tag{4-33}$$

*Moreover, there exists $C > 0$ such that, for all $\eta > 0$,*

$$\mathrm{Op}(\chi_0) P^{(1)} \mathrm{Op}(\chi_0) \geq (\beta_d - C\eta) \mathrm{Op}(\chi_0) P^{W,(1)} \mathrm{Op}(\chi_0) - (C\eta h + \mathcal{O}(h^2)). \tag{4-34}$$

*Proof.* We first estimate $P^{(1)}$ outside of the critical points $\mathcal{U}$. Since $\chi_\infty$ vanishes near $\mathcal{U}$, Proposition 2.3 yields that there exist $\delta_\eta > 0$ and $\tilde{p}_\eta \in S^0(1)$ (which may depend on $\eta$) such that $p = \tilde{p}_\eta$ in a vicinity of the support of $\chi_\infty$ and $\tilde{p}_\eta(x, \xi) \geq 2\delta_\eta$ for all $(x, \xi) \in \mathbb{R}^{2d}$. Then, Lemma 4.5 and the pseudodifferential calculus (in particular, the Gårding inequality) imply

$$\begin{aligned}
\mathrm{Op}(\chi_\infty) P^{(1)} \mathrm{Op}(\chi_\infty) &= \mathrm{Op}(\chi_\infty) P^{(0)} \mathrm{Op}(\chi_\infty) + \mathrm{Op}(\chi_\infty) \mathcal{O}(h) \mathrm{Op}(\chi_\infty) \\
&= \mathrm{Op}(\chi_\infty) \mathrm{Op}(\tilde{p}_\eta) \mathrm{Op}(\chi_\infty) + \mathrm{Op}(\chi_\infty) \mathcal{O}(h) \mathrm{Op}(\chi_\infty) + \mathcal{O}(h^\infty) \\
&\geq \mathrm{Op}(\chi_\infty)(2\delta_\eta + \mathcal{O}(h)) \mathrm{Op}(\chi_\infty) + \mathcal{O}(h^\infty),
\end{aligned}$$

which implies (4-33) for $h$ small enough. Here, we have identified as before $A$ with $A \otimes \mathrm{Id}$ for scalar operators $A$.

We now consider $\mathrm{Op}(\chi_0) P^{(1)} \mathrm{Op}(\chi_0)$. Thanks to Lemma 4.6, we can write

$$\mathrm{Op}(\chi_0) P^{(1)} \mathrm{Op}(\chi_0) = \beta_d \mathrm{Op}(\chi_0) P^{W,(1)} \mathrm{Op}(\chi_0) + \sum_{k=0}^{2} \mathrm{Op}(\chi_0) R_k \mathrm{Op}(\chi_0).$$

Let $\tilde{\chi}_0 \in C_0^\infty(\mathbb{R}^{2d}; [0,1])$ be supported in a neighborhood of size $\eta$ of $(\boldsymbol{u}, 0)$, $\boldsymbol{u} \in \mathcal{U}$, and such that $\tilde{\chi}_0 = 1$ near the support of $\chi_0$. Then, for $\omega \in \Omega^1(\mathbb{R}^d)$, $\langle R_0 \mathrm{Op}(\chi_0)\omega, \mathrm{Op}(\chi_0)\omega\rangle$ is a finite sum of terms of the form

$$\tilde{r}_0 = \langle (d^a_{\phi,h})^* (\mathrm{Op}(r_0) + \Psi^0(h)) d^b_{\phi,h} \mathrm{Op}(\chi_0)\omega_j, \mathrm{Op}(\chi_0)\omega_k\rangle. \tag{4-35}$$

Using functional analysis and pseudodifferential calculus, we get

$$\begin{aligned} |\tilde{r}_0| &= |\langle (\mathrm{Op}(r_0\tilde{\chi}_0) + \Psi^0(h)) d^b_{\phi,h} \mathrm{Op}(\chi_0)\omega_j, d^a_{\phi,h} \mathrm{Op}(\chi_0)\omega_k\rangle| + \mathcal{O}(h^\infty)\|\omega\|^2 \\ &\le (\|\mathrm{Op}(r_0\tilde{\chi}_0)\| + \mathcal{O}(h)) \|d^b_{\phi,h} \mathrm{Op}(\chi_0)\omega_j\| \|d^a_{\phi,h} \mathrm{Op}(\chi_0)\omega_k\| + \mathcal{O}(h^\infty)\|\omega\|^2 \\ &\le (\|\mathrm{Op}(r_0\tilde{\chi}_0)\| + \mathcal{O}(h)) \langle P^{W,(0)} \mathrm{Op}(\chi_0)\omega, \mathrm{Op}(\chi_0)\omega\rangle + \mathcal{O}(h^\infty)\|\omega\|^2. \end{aligned} \tag{4-36}$$

Recall now that, for $a \in S^0(1)$,

$$\|\mathrm{Op}(a)\|_{L^2(\mathbb{R}^d)\to L^2(\mathbb{R}^d)} = \|a\|_{L^\infty(\mathbb{R}^{2d})} + \mathcal{O}(h)$$

(see, e.g., [Zworski 2012, Theorem 13.13]). Thus, using that $\tilde{\chi}_0$ is supported in a neighborhood of size $\eta$ of $(\boldsymbol{u}, 0)$ at which $r_0$ vanishes yields $\|\mathrm{Op}(r_0\tilde{\chi}_0)\| \le C\eta$, and (4-36) implies

$$|\langle R_0 \mathrm{Op}(\chi_0)\omega, \mathrm{Op}(\chi_0)\omega\rangle| \le C\eta \langle P^{W,(0)} \mathrm{Op}(\chi_0)\omega, \mathrm{Op}(\chi_0)\omega\rangle + \mathcal{O}(h^\infty)\|\omega\|^2. \tag{4-37}$$

As before, $\langle R_1 \mathrm{Op}(\chi_0)\omega, \mathrm{Op}(\chi_0)\omega\rangle$ is a finite sum of terms of the form

$$\tilde{r}_1 = \langle \Psi^0(h) d^a_{\phi,h} \mathrm{Op}(\chi_0)\omega_j, \mathrm{Op}(\chi_0)\omega_k\rangle \tag{4-38}$$

or its complex conjugate. These terms can be estimated as

$$\begin{aligned} |\tilde{r}_1| &\le Ch \|d^a_{\phi,h} \mathrm{Op}(\chi_0)\omega_j\| \|\omega\| \\ &\le \eta \|d^a_{\phi,h} \mathrm{Op}(\chi_0)\omega_j\|^2 + \mathcal{O}(h^2)\|\omega\|^2 \\ &\le \eta \langle P^{W,(0)} \mathrm{Op}(\chi_0)\omega, \mathrm{Op}(\chi_0)\omega\rangle + \mathcal{O}(h^2)\|\omega\|^2, \end{aligned}$$

and then

$$|\langle R_1 \mathrm{Op}(\chi_0)\omega, \mathrm{Op}(\chi_0)\omega\rangle| \le C\eta \langle P^{W,(0)} \mathrm{Op}(\chi_0)\omega, \mathrm{Op}(\chi_0)\omega\rangle + \mathcal{O}(h^2)\|\omega\|^2. \tag{4-39}$$

Combining Lemma 4.6 with the estimates (4-37), (4-39) and $R_2 \in \Psi^0(h^2)$, we obtain

$$\mathrm{Op}(\chi_0) P^{(1)} \mathrm{Op}(\chi_0) \ge \beta_d \mathrm{Op}(\chi_0) P^{W,(1)} \mathrm{Op}(\chi_0) - C\eta \mathrm{Op}(\chi_0) P^{W,(0)} \mathrm{Op}(\chi_0) - \mathcal{O}(h^2).$$

Since $P^{W,(1)} = P^{W,(0)} \otimes \mathrm{Id} + \Psi^0(h)$ (see Equation (1.9) of [Helffer and Sjöstrand 1985], for example), this inequality gives (4-34). □

Let $\Pi$ denote the orthogonal projection onto $\mathrm{Vect}\{f_j^{(1)} : j = 2, \dots, n_1 + 1\}$. Using the previous lemma and its proof, we can describe the action of $P^{(1)}$ on $\Pi$:

**Lemma 4.11.** *The rank of* $\Pi$ *is* $n_1$ *for* $h$ *small enough. Moreover*,

$$P^{(1)}\Pi = \mathscr{O}(h^2) \quad \textit{and} \quad \Pi P^{(1)} = \mathscr{O}(h^2). \tag{4-40}$$

*Finally*, *there exists* $\varepsilon_1 > 0$ *such that*

$$(1-\Pi)P^{(1)}(1-\Pi) \geq \varepsilon_1 h(1-\Pi) \tag{4-41}$$

*for* $h$ *small enough.*

*Proof.* Since the functions $f_j^{(1)}$ are almost orthogonal (i.e., $\langle f_j^{(1)}, f_{j'}^{(1)}\rangle = \delta_{j,j'} + \mathscr{O}(h)$), the rank of $\Pi$ is $n_1$. Moreover, (4-40) is a direct consequence of Lemma 4.8.

We now give the lower bound for $P^{(1)}$ on the range of $1-\Pi$. Let $\mathscr{E}^{(1)}$ denote the space spanned by the $f_k^{W,(1)}$, $k = 2, \dots, n_1+1$ and $\mathscr{F}^{(1)}$ the eigenspace associated to the $n_1$ first eigenvalues of $P^{W,(1)}$. Let $\Pi_{\mathscr{E}^{(1)}}$, $\Pi_{\mathscr{F}^{(1)}}$ denote the corresponding orthogonal projectors. It follows from [Helffer and Sjöstrand 1985] that $\|\Pi_{\mathscr{E}^{(1)}} - \Pi_{\mathscr{F}^{(1)}}\| = \mathscr{O}(e^{-c/h})$ for some $c > 0$. On the other hand, it follows from the first estimate of Lemma 4.8 that $\|\Pi - \Pi_{\mathscr{E}^{(1)}}\| = \mathscr{O}(h)$. Combining these two estimates, we get

$$\|\Pi - \Pi_{\mathscr{F}^{(1)}}\| = \mathscr{O}(h).$$

Using this bound and the spectral properties of $P^{W,(1)}$, we get

$$P^{W,(1)} \geq \nu h - \nu h \Pi_{\mathscr{F}^{(1)}} \geq \nu h - \nu h \Pi + \mathscr{O}(h^2) \tag{4-42}$$

for some $\nu > 0$. From (4-23) and integration by parts, we also have $\mathrm{Op}(\chi_0)\Pi = \Pi + \mathscr{O}(h^\infty)$. Estimate (4-42) together with (4-31), (4-32), (4-33) and (4-34) give

$$\begin{aligned} P^{(1)} &= \mathrm{Op}(\chi_0)P^{(1)}\,\mathrm{Op}(\chi_0) + \mathrm{Op}(\chi_\infty)P^{(1)}\,\mathrm{Op}(\chi_\infty) + \mathscr{O}(h^2) \\ &\geq (\beta_d - C\eta)\,\mathrm{Op}(\chi_0)P^{W,(1)}\,\mathrm{Op}(\chi_0) + \delta_\eta\,\mathrm{Op}(\chi_\infty)^2 - (C\eta h + \mathscr{O}(h^2)) \\ &\geq \nu h(\beta_d - C\eta)\,\mathrm{Op}(\chi_0)^2 - \nu h(\beta_d - C\eta)\Pi + \delta_\eta\,\mathrm{Op}(\chi_\infty)^2 - (C\eta h + \mathscr{O}(h^2)) \\ &\geq \nu h(\beta_d - C\eta) - \nu h(\beta_d - C\eta)\Pi - (C\eta h + \mathscr{O}(h^2)). \end{aligned} \tag{4-43}$$

Thus, taking $\eta > 0$ small enough and applying $1-\Pi$, we finally obtain (4-41) for some $\varepsilon_1 > 0$. □

*Proof of Proposition 4.9.* From Proposition 2.4 and Lemma 4.5, the operator $P^{(1)}$ is bounded and its essential spectrum is above some positive constant independent of $h$. Next, the maxi-min principle together with (4-40) implies that $P^{(1)}$ has at least $\mathrm{rank}(\Pi) = n_1$ eigenvalues below $Ch^2$. In the same way, (4-41) yields that $P^{(1)}$ has at most $n_1$ eigenvalues below $\varepsilon_1 h$. Finally,

$$P^{(1)} = (1-\Pi)P^{(1)}(1-\Pi) + \Pi P^{(1)}(1-\Pi) + (1-\Pi)P^{(1)}\Pi + \Pi P^{(1)}\Pi \geq -Ch^2$$

proves that all the spectrum of $P^{(1)}$ is above $-Ch^2$. □

## 5. Eigenspace analysis and proof of the main theorem

Now we want to project the preceding quasimodes onto the generalized eigenspaces associated to exponentially small eigenvalues, and prove the main theorem. Recall that we have built in the preceding section quasimodes $f_k^{(0)}$, $k = 1, \dots, n_0$, for $P^{(0)}$ with good support properties. To each quasimode we will associate a function in $E^{(0)}$, the eigenspace associated to the $\mathbb{O}(h^2)$ eigenvalues. For this, we first define the spectral projector

$$\Pi^{(0)} = \frac{1}{2\pi i} \int_\gamma (z - P^{(0)})^{-1}\, dz, \tag{5-1}$$

where $\gamma = \partial B\big(0, \tfrac{1}{2}\varepsilon_0 h\big)$ and $\varepsilon_0 > 0$ is defined in Proposition 4.3. From the fact that $P^{(0)}$ is selfadjoint, we get that

$$\Pi^{(0)} = \mathbb{O}(1).$$

For the following, we denote the corresponding projection by

$$e_k^{(0)} = \Pi^{(0)}(f_k^{(0)}).$$

**Lemma 5.1.** *The system $(e_k^{(0)})_k$ is free and spans $E^{(0)}$. Further, there exists $\alpha > 0$ independent of $\varepsilon$ such that*

$$e_k^{(0)} = f_k^{(0)} + \mathbb{O}(e^{-\alpha/h}) \quad \textit{and} \quad \langle e_k^{(0)}, e_{k'}^{(0)} \rangle = \delta_{k,k'} + \mathbb{O}(h).$$

*Proof.* The proof follows [Helffer and Sjöstrand 1985] (see also [Dimassi and Sjöstrand 1999]). We sketch it for the sake of completeness and to give the necessary modifications. Using (5-1) and the Cauchy formula, we get

$$\begin{aligned} e_k^{(0)} - f_k^{(0)} = \Pi^{(0)} f_k^{(0)} - f_k^{(0)} &= \frac{1}{2\pi i} \int_\gamma (z - P^{(0)})^{-1} f_k^{(0)}\, dz - \frac{1}{2\pi i} \int_\gamma z^{-1} f_k^{(0)}\, dz \\ &= \frac{1}{2\pi i} \int_\gamma (z - P^{(0)})^{-1} z^{-1} P^{(0)} f_k^{(0)}\, dz. \end{aligned}$$

Since $P^{(0)}$ is selfadjoint and according to Proposition 4.3, we have

$$\|(z - P^{(0)})^{-1}\| = \mathbb{O}(h^{-1})$$

uniformly for $z \in \gamma$. Using also the second estimate in Lemma 4.1, this yields

$$\|(z - P^{(0)})^{-1} z^{-1} P^{(0)} f_k^{(0)}\| = \mathbb{O}(h^{-2} e^{-\alpha/h}),$$

and, after integration,

$$\|e_k^{(0)} - f_k^{(0)}\| = \mathbb{O}(h^{-1} e^{-\alpha/h}).$$

Decreasing $\alpha$, we obtain the first estimate of the lemma. In particular, this implies that the family $(e_k^{(0)})_k$ is free. Using that $E^{(0)}$ is of dimension $n_0$, the family $(e_k^{(0)})_k$ spans $E^{(0)}$.

For the last equality of the lemma, we just have to notice that

$$\langle e_k^{(0)}, e_{k'}^{(0)} \rangle = \langle f_k^{(0)}, f_{k'}^{(0)} \rangle + \mathbb{O}(e^{-\alpha/h}) = \delta_{k,k'} + \mathbb{O}(h) + \mathbb{O}(e^{-\alpha/h}) = \delta_{k,k'} + \mathbb{O}(h),$$

according to Lemma 4.1. The proof is complete. □

We can do a similar study for the analysis of $P^{(1)}$, for which we know that exactly $n_1$ (real) eigenvalues are $\mathcal{O}(h^2)$, and among them at least $n_0 - 1$ are exponentially small. Note that there is no particular reason for the remaining ones to also be exponentially small.

To the family of quasimodes $(f_j^{(1)})_j$, we now associate a family of functions in $E^{(1)}$, the eigenspace associated to the $\mathcal{O}(h^2)$ eigenvalues for $P^{(1)}$. By the spectral properties of the selfadjoint operator $P^{(1)}$, its spectral projector onto $E^{(1)}$ is given by

$$\Pi^{(1)} = \frac{1}{2\pi i} \int_\gamma (z - P^{(1)})^{-1}\, dz, \tag{5-2}$$

where $\gamma = \partial B\big(0, \frac{1}{2}\varepsilon_1 h\big)$, with $\varepsilon_1$ defined in Proposition 4.9. In the sequel, we write

$$e_j^{(1)} = \Pi^{(1)}(f_j^{(1)}).$$

Mimicking the proof of Lemma 5.1, one can show that the family $(e_j^{(1)})_j$ satisfies the following estimates:

**Lemma 5.2.** *The system $(e_j^{(1)})_j$ is free and spans $E^{(1)}$. Further, we have*

$$e_j^{(1)} = f_j^{(1)} + \mathcal{O}(h) \quad \text{and} \quad \langle e_j^{(1)}, e_{j'}^{(1)} \rangle = \delta_{j,j'} + \mathcal{O}(h).$$

Thanks to the preceding lemmas, the families $(e_k^{(0)})_k$ and $(e_j^{(1)})_j$ are orthonormal, apart from an $\mathcal{O}(h)$ factor. To accurately compute the eigenvalues of $P^{(0)}$ and prove the main theorem, we need more precise estimates of exponential type. For this, we will use the intertwining relation $L_\phi P^{(0)} = P^{(1)} L_\phi$.

More precisely, we denote by $L$ the $n_1 \times n_0$ matrix of this restriction of $L_\phi$ with respect to the bases $(e_j^{(1)})_j$ and $(e_k^{(0)})_k$:

$$L_{j,k} := \langle e_j^{(1)}, L_\phi e_k^{(0)} \rangle. \tag{5-3}$$

The classical way (e.g., [Helffer et al. 2004; Helffer and Sjöstrand 1985]) to compute the exponentially small eigenvalues of $P^{(0)}$ is to then accurately compute the singular values of $L$. For this, we first state a refined lemma about exponential estimates.

**Lemma 5.3.** *There exists $\alpha > 0$ independent of $\varepsilon$ such that*

$$L_\phi L_\phi^* f_j^{(1)} = \mathcal{O}(e^{-\alpha/h}), \tag{5-4}$$

*and also a smooth 1-form $r_j^{(1)}$ such that*

$$L_\phi^*(e_j^{(1)} - f_j^{(1)}) = L_\phi^* r_j^{(1)} \quad \text{and} \quad r_j^{(1)} = \mathcal{O}(e^{-\alpha/h}).$$

*Proof.* We first note that

$$\begin{aligned} L_\phi L_\phi^* f_j^{(1)} &= \beta_d^{1/2} L_\phi a_h d_{\phi,h}^* Q^* (Q^*)^{-1} f_j^{W,(1)} \\ &= \beta_d^{1/2} L_\phi a_h (d_{\phi,h}^* f_j^{W,(1)}). \end{aligned} \tag{5-5}$$

On the other hand, (4-2) and (4-24) give

$$\|d^*_{\phi,h} f_j^{W,(1)}\|^2 \le \|d^*_{\phi,h} f_j^{W,(1)}\|^2 + \|d_{\phi,h} f_j^{W,(1)}\|^2 = \langle P^{W,(1)} f_j^{W,(1)}, f_j^{W,(1)}\rangle = \mathbb{O}(e^{-\alpha/h})$$

for some $\alpha > 0$ independent of $\varepsilon$. Since $a_h$ and $L_\phi$ are uniformly bounded operators, (5-5) provides the required estimate.

Now we show the second and third equalities, following closely the proof of Lemma 5.1. Using (5-1), the intertwining relation (see Lemma 4.7) and the Cauchy formula, we have

$$\begin{aligned} L^*_\phi(e_j^{(1)} - f_j^{(1)}) &= L^*_\phi \Pi^{(1)} f_j^{(1)} - L^*_\phi f_j^{(1)} \\ &= \Pi^{(0)} L^*_\phi f_j^{(1)} - L^*_\phi f_j^{(1)} \\ &= \frac{1}{2\pi i}\int_\gamma (z - P^{(0)})^{-1} L^*_\phi f_j^{(1)}\,dz - \frac{1}{2\pi i}\int_\gamma z^{-1} L^*_\phi f_j^{(1)}\,dz \\ &= \frac{1}{2\pi i}\int_\gamma (z - P^{(0)})^{-1} z^{-1} P^{(0)} L^*_\phi f_j^{(1)}\,dz, \end{aligned} \tag{5-6}$$

where $\gamma = \partial B\big(0, \frac{1}{2}\min(\varepsilon_0, \varepsilon_1)h\big)$. Using again Lemma 4.7, this becomes

$$\begin{aligned} L^*_\phi(e_j^{(1)} - f_j^{(1)}) &= \frac{1}{2\pi i}\int_\gamma (z - P^{(0)})^{-1} z^{-1} L^*_\phi L_\phi L^*_\phi f_j^{(1)}\,dz \\ &= L^*_\phi \frac{1}{2\pi i}\int_\gamma (z - P^{(1)})^{-1} z^{-1} L_\phi L^*_\phi f_j^{(1)}\,dz. \end{aligned}$$

We then let

$$r_j^{(1)} = \bigg(\frac{1}{2\pi i}\int_\gamma (z - P^{(1)})^{-1} z^{-1}\,dz\bigg) L_\phi L^*_\phi f_j^{(1)}, \tag{5-7}$$

and the preceding equality reads

$$L^*_\phi(e_j^{(1)} - f_j^{(1)}) = L^*_\phi r_j^{(1)}. \tag{5-8}$$

Moreover, as in proof of Lemma 5.1, we have

$$\frac{1}{2\pi i}\int_\gamma (z - P^{(1)})^{-1} z^{-1}\,dz = \mathbb{O}(h^{-1}).$$

Combining with (5-4), this shows that $r_j^{(1)} = \mathbb{O}(e^{-\alpha/h})$ for some (new) $\alpha > 0$. □

We begin the study of the matrix $L$ with the following lemma:

**Lemma 5.4.** *There exists $\alpha' > 0$ such that, if $\varepsilon > 0$ is sufficiently small and fixed, we have, for all $2 \le j \le n_1 + 1$ and $2 \le k \le n_0$,*

$$L_{j,k} = \langle f_j^{(1)}, L_\phi f_k^{(0)}\rangle + \mathbb{O}(e^{-(S_k + \alpha')/h}).$$

*Moreover, $L_{j,1} = 0$ for all $2 \le j \le n_1 + 1$.*

*Proof.* We first treat the case $k = 1$. Since $f_1^{(0)}$ is collinear to $a_h^{-1}e^{-\phi/h}$, it belongs to $\ker(P^{(0)})$. Then, $e_1^{(0)} = \Pi^{(0)} f_1^{(0)} = f_1^{(0)}$ satisfies $L_\phi e_1^{(0)} = 0$ from Lemma 4.7. In particular, $L_{j,1} = 0$ for all $2 \le j \le n_1 + 1$.

We now assume $2 \le k \le n_0$. Using Lemma 4.7 and the definition of $e_\bullet^{(\star)}$, we can write

$$\begin{aligned} L_{j,k} &= \langle e_j^{(1)}, L_\phi e_k^{(0)}\rangle = \langle e_j^{(1)}, L_\phi \Pi^{(0)} f_k^{(0)}\rangle = \langle e_j^{(1)}, \Pi^{(1)} L_\phi f_k^{(0)}\rangle \\ &= \langle \Pi^{(1)} e_j^{(1)}, L_\phi f_k^{(0)}\rangle = \langle e_j^{(1)}, L_\phi f_k^{(0)}\rangle = \langle f_j^{(1)}, L_\phi f_k^{(0)}\rangle + \langle e_j^{(1)} - f_j^{(1)}, L_\phi f_k^{(0)}\rangle \\ &= \langle f_j^{(1)}, L_\phi f_k^{(0)}\rangle + \langle L_\phi^*(e_j^{(1)} - f_j^{(1)}), f_k^{(0)}\rangle. \end{aligned}$$

From Lemma 5.3, this becomes

$$\begin{aligned} L_{j,k} &= \langle f_j^{(1)}, L_\phi f_k^{(0)}\rangle + \langle L_\phi^* r_j^{(1)}, f_k^{(0)}\rangle \\ &= \langle f_j^{(1)}, L_\phi f_k^{(0)}\rangle + \langle r_j^{(1)}, L_\phi f_k^{(0)}\rangle. \end{aligned} \tag{5-9}$$

Now, since $Q$ is bounded and according to Lemma A.3, we have

$$L_\phi f_k^{(0)} = Q d_{\phi,h} f^{W,(0)} = \mathcal{O}(e^{-(S_k - C\varepsilon)/h}).$$

Using Lemma 5.3 again, this yields

$$\langle r_j^{(1)}, L_\phi f_k^{(0)}\rangle = \mathcal{O}(e^{-(S_k + \alpha - C\varepsilon)/h}) \tag{5-10}$$

with $\alpha > 0$ independent of $\varepsilon$. Taking $\varepsilon > 0$ small enough, the lemma follows from (5-9) and (5-10). □

Now we recall the explicit computation of the matrix $L$. This is just a consequence of the study of the corresponding Witten Laplacian.

**Lemma 5.5.** *For all* $2 \le j \le n_1 + 1$ *and* $2 \le k \le n_0$, *we have*

$$L_{k,k} = \left(\frac{h}{(2d+4)\pi}\right)^{1/2} \mu_k^{1/2} \left|\frac{\det \phi''(\boldsymbol{m}_k)}{\det \phi''(\boldsymbol{s}_k)}\right|^{1/4} e^{-S_k/h}(1 + \mathcal{O}(h)) =: h^{1/2}\ell_k(h) e^{-S_k/h}$$

*and*

$$L_{j,k} = \mathcal{O}(e^{-(S_k + \alpha')/h}) \quad \textit{for all } j \ne k,$$

*where* $S_k := \phi(\boldsymbol{s}_k) - \phi(\boldsymbol{m}_k)$ *and* $-\mu_k$ *denotes the unique negative eigenvalue of* $\phi''$ *at* $\boldsymbol{s}_k$.

*Proof.* First, we note that

$$\langle f_j^{(1)}, L_\phi f_k^{(0)}\rangle = \beta_d^{1/2} \langle f_j^{W,(1)}, d_{\phi,h} f_k^{W,(0)}\rangle,$$

by (4-3), (4-26) and $L_\phi = Q d_{\phi,h} a_h$. Thus, Lemma 5.4 implies

$$L_{j,k} = \beta_d^{1/2} \langle f_j^{W,(1)}, d_{\phi,h} f_k^{W,(0)}\rangle + \mathcal{O}(e^{-(S_k + \alpha')/h}).$$

The first term is exactly the approximate singular value of $d_{\phi,h}$ computed in [Helffer et al. 2004]. The result is then a direct consequence of Proposition 6.4 of [Helffer et al. 2004]. □

Now we are able to compute the singular values of $L$ (i.e., the eigenvalues of $(L^*L)^{1/2}$).

**Lemma 5.6.** *There exists $\alpha' > 0$ such that the singular values $\nu_k(L)$ of $L$, enumerated in a suitable order, satisfy*

$$\nu_k(L) = |L_{k,k}|(1 + \mathcal{O}(e^{-\alpha'/h})) \quad \textit{for all } 1 \le k \le n_0.$$

*Proof.* Since the first column of $L$ consists of zeros, we get $\nu_1 = 0$. Moreover, the other singular values of $L$ are those of the reduced matrix $L'$ with entries $L'_{j,k} = L_{j+1,k+1}$ for $1 \le j \le n_1$ and $1 \le k \le n_0 - 1$. We shall now use that the dominant term in each column of $L'$ lies on the diagonal. Define the $(n_0 - 1) \times (n_0 - 1)$ diagonal matrix $D$ by

$$D := \operatorname{diag}(L_{k+1,k+1} : k = 1, \dots, n_0 - 1).$$

Notice that $D$ is invertible, thanks to the ellipticity of $\ell_{k+1}(h)$, and that $\nu_k(D) = |L_{k+1,k+1}|$. We also define the $n_1 \times (n_0 - 1)$ characteristic matrix of $L'$

$$U = (\delta_{j,k})_{j,k}.$$

From Lemma 5.5, there is a constant $\alpha' > 0$ such that

$$L' = (U + \mathcal{O}(e^{-\alpha'/h}))D. \tag{5-11}$$

The Fan inequalities (see, for example, Theorem 1.6 of [Simon 1979]) therefore give

$$\nu_k(L') \le (1 + \mathcal{O}(e^{-\alpha'/h}))\nu_k(D). \tag{5-12}$$

To get the opposite estimate, we remark that $U^*U = \mathrm{Id}_{n_0-1}$. Then, (5-11) implies

$$D = (1 + \mathcal{O}(e^{-\alpha'/h}))U^*L',$$

and, as before,

$$\nu_k(D) \le (1 + \mathcal{O}(e^{-\alpha'/h}))\nu_k(L'). \tag{5-13}$$

The lemma follows from $\nu_{k+1}(L) = \nu_k(L')$, (5-12), (5-13) and $\nu_k(D) = |L_{k+1,k+1}|$. □

Now, Theorem 1.2 is a direct consequence of the explicit computations of Lemma 5.5 and of the following equivalent formulation:

**Lemma 5.7.** *The nonzero exponentially small eigenvalues of $P_h$ are of the form*

$$h(\ell_k^2(h) + \mathcal{O}(h))e^{-2S_k/h} \quad \textit{for } 2 \le k \le n_0.$$

*Proof.* According to Lemma 5.1 and Lemma 5.2, the bases $(e_k^{(0)})_k$ and $(e_j^{(1)})_j$ of $E^{(0)}$ and $E^{(1)}$ respectively are orthonormal up to $\mathcal{O}(h)$ small errors. Let $(\tilde{e}_k^{(0)})_k$ and $(\tilde{e}_j^{(1)})_j$ be the corresponding orthonormalizations (obtained by taking square roots of the Gramians), which differ from the original bases by $\mathcal{O}(h)$ small recombinations. Then, with respect to the new bases, the matrix of $L_\phi$ takes the form $\widetilde{L} = (1 + \mathcal{O}(h))L(1 + \mathcal{O}(h))$. Using the Fan inequalities, we see that the conclusion of Lemma 5.6 is also valid for $\widetilde{L}$ (note that there is no need to have exponentially small errors here). Since the matrix of the restriction of $P^{(0)}$ to $E^{(0)}$ with respect to the basis $(\tilde{e}_k^{(0)})_k$ is given by $\widetilde{L}^*\widetilde{L}$, the lemma follows. □

We end this section by showing that the main theorems stated in Section 1 imply the metastability of the system.

*Proof of Corollary 1.4.* We first prove (1-5) and (1-7). If $\phi$ has a unique minimum, Theorem 1.1 gives

$$\|(\boldsymbol{T}_h^\star)^n(d\nu_h) - d\nu_{h,\infty}\|_{\mathcal{H}_h} \leq (1-\delta h)^n \|d\nu_h\|_{\mathcal{H}_h} = e^{n\ln(1-\delta h)+|\ln h|} h \|d\nu_h\|_{\mathcal{H}_h}.$$

Using that $n\ln(1-\delta h) \sim -\delta h n$, this estimate yields

$$\|(\boldsymbol{T}_h^\star)^n(d\nu_h) - d\nu_{h,\infty}\|_{\mathcal{H}_h} \leq h\|d\nu_h\|_{\mathcal{H}_h}$$

for $n \gtrsim |\ln h| h^{-1}$. In the same way, if $\phi$ has several minima, Theorem 1.2 implies

$$\|(\boldsymbol{T}_h^\star)^n(d\nu_h) - d\nu_{h,\infty}\|_{\mathcal{H}_h} \leq (\lambda_2^\star(h))^n \|d\nu_h\|_{\mathcal{H}_h} = e^{n\ln(\lambda_2^\star(h))+|\ln h|} h \|d\nu_h\|_{\mathcal{H}_h}.$$

Using now that $n\ln(\lambda_2^\star(h)) \sim n(\lambda_2^\star(h)-1) \sim -Cnhe^{-S_2/h}$ for some $C>0$, this estimate yields

$$\|(\boldsymbol{T}_h^\star)^n(d\nu_h) - d\nu_{h,\infty}\|_{\mathcal{H}_h} \leq h\|d\nu_h\|_{\mathcal{H}_h}$$

for $n \gtrsim |\ln h| h^{-1} e^{S_2/h}$.

It remains to show (1-6). From Theorem 1.1, Theorem 1.2 and the proof of (1-5), we can write

$$(\boldsymbol{T}_h^\star)^n(d\nu_h) = \sum_{k=1}^{n_0} (\lambda_k^\star(h))^n \Pi_k d\nu_h + \mathcal{O}(h)\|d\nu_h\|_{\mathcal{H}_h},$$

for $n \gtrsim |\ln h| h^{-1}$. Here, $\Pi_k$ is the spectral projector of $\boldsymbol{T}_h^\star$ associated to the eigenvalue $\lambda_k^\star(h)$. If we assume in addition that $n \lesssim e^{2S_{n_0}/h}$, then $(\lambda_k^\star(h))^n = 1 + \mathcal{O}(h)$ for any $k = 1, \dots, n_0$. Thus, the previous equation becomes

$$(\boldsymbol{T}_h^\star)^n(d\nu_h) = \Pi^{(0)} d\nu_h + \mathcal{O}(h)\|d\nu_h\|_{\mathcal{H}_h}, \tag{5-14}$$

since $\Pi^{(0)} = \Pi_1 + \dots + \Pi_{n_0}$. Let

$$g_k(x) := \frac{\chi_k(x) e^{-(\phi(x)-\phi(\boldsymbol{m}_k))/h}}{\|\chi_k e^{-(\phi-\phi(\boldsymbol{m}_k))/h}\|}.$$

From (A-1), we immediately get $g_k = f_k^{W,(0)} + \mathcal{O}(h)$. Moreover, as in (4-4), we have

$$\|f_k^{(0)} - f_k^{W,(0)}\| = \|(a_h^{-1} - 1) f_k^{W,(0)}\| = \mathcal{O}(h).$$

Combining with Lemma 5.1, we deduce

$$g_k = e_k^{(0)} + \mathcal{O}(h). \tag{5-15}$$

Using Lemma 5.1 one more time, the bases $(e_k^{(0)})_k$ and $(g_k)_k$ of $\operatorname{Im}\Pi^{(0)}$ and $\operatorname{Im}\Pi$, respectively, are almost orthogonal, in the sense that

$$\langle e_k^{(0)}, e_{k'}^{(0)}\rangle = \delta_{k,k'} + \mathcal{O}(h) \quad \text{and} \quad \langle g_k, g_{k'}\rangle = \delta_{k,k'} + \mathcal{O}(h).$$

This then yields

$$\Pi = \Pi^{(0)} + \mathcal{O}(h), \tag{5-16}$$

and (1-6) follows from (5-14). □

## Appendix: Quasimodes, truncation procedure and labeling

In this appendix, we gather from [Helffer et al. 2004; Hérau et al. 2011] the refined construction of quasimodes on 0-forms for the Witten Laplacian, and the labeling procedure linking each minima with a saddle point of index 1. We recall briefly the construction proposed in [Hérau et al. 2011] (which was in the Fokker–Planck case there) but in a generic situation where all $\phi(\boldsymbol{s})-\phi(\boldsymbol{m})$ are distinct for $\boldsymbol{m}$ in the set of minima and $\boldsymbol{s}$ in the set of saddle points of $\phi$.

In the following, we will denote by $\mathscr{L}(\sigma)=\{x\in\mathbb{R}^n:\phi(x)<\sigma\}$ the sublevel set associated to the value $\sigma\in\mathbb{R}$. Let $\boldsymbol{s}$ be a saddle point of $\phi$ and $B(\boldsymbol{s},r)=\{x\in\mathbb{R}^n:|x-\boldsymbol{s}|<r\}$. Then, for $r>0$ small enough, the set

$$B(\boldsymbol{s},r)\cap\mathscr{L}(\phi(\boldsymbol{s}))=\{x\in B(\boldsymbol{s},r):\phi(x)<\phi(\boldsymbol{s})\}$$

has precisely 2 connected components, $C_j(\boldsymbol{s},r)$ with $j=1,2$.

**Definition A.1.** We say that $\boldsymbol{s}\in\mathbb{R}^n$ is a separating saddle point (ssp) if it is either $\infty$ or it is a usual saddle point such that $C_1(\boldsymbol{s},r)$ and $C_2(\boldsymbol{s},r)$ are contained in different connected components of the set $\{x\in\mathbb{R}^n:\phi(x)<\phi(\boldsymbol{s})\}$. We denote by SSP the set of ssps.

We also introduce the set of separating saddle values (ssv), $\mathrm{SSV}=\{\phi(\boldsymbol{s}):\boldsymbol{s}\in\mathrm{SSP}\}$ with the convention that $\phi(\infty)=+\infty$.

A connected component $E$ of the sublevel set $\mathscr{L}(\sigma)$ will be called a critical component if either $\partial E\cap\mathrm{SSP}\neq\varnothing$ or $E=\mathbb{R}^n$.

Let us now explain the way we label the critical points. We first order the saddle points in the following way. We recall from [Helffer et al. 2004] that $\sharp\mathrm{SSV}=n_0$ and then enumerate the ssvs in a decreasing way: $\infty=\sigma_1>\sigma_2>\cdots>\sigma_{n_0}$. To each ssv $\sigma_j$ we can associate a unique ssp: we define $\boldsymbol{s}_1=\infty$ and, for any $j=2,\dots,n_0$, we let $\boldsymbol{s}_j$ be the unique ssp such that $\phi(\boldsymbol{s}_j)=\sigma_j$ (note that this $\boldsymbol{s}_j$ is unique thanks to Hypothesis 2).

Then we can proceed to the labeling of minima. We denote by $\boldsymbol{m}_1$ the global minimum of $\phi$, $E_1=\mathbb{R}^d$ and by $S_1=\phi(\boldsymbol{s}_1)-\phi(\boldsymbol{m}_1)=+\infty$ the critical Arrhenius value.

Next we observe that the sublevel set $\mathscr{L}(\sigma_2)=\{x\in\mathbb{R}^n:\phi(x)<\sigma_2\}$ is the union of two critical components, with one containing $\boldsymbol{m}_1$. The remaining connected component of the sublevel set $\mathscr{L}(\sigma_2)$ will be denoted by $E_2$ and its minimum by $\boldsymbol{m}_2$. To the pair $(\boldsymbol{m}_2,\boldsymbol{s}_2)$ of critical points we associate the Arrhenius value $S_2=\phi(\boldsymbol{s}_2)-\phi(\boldsymbol{m}_2)$.

Continuing the labeling procedure, we decompose the sublevel set $\mathscr{L}(\sigma_3)$ into its connected components and perform the labeling as follows: we omit all those components that contain the already labeled minima $\boldsymbol{m}_1$ and $\boldsymbol{m}_2$. Some of these components may be noncritical. There is only one critical one remaining, and we denote it by $E_3$. We then let $\boldsymbol{m}_3$ be the point of global minimum of the restriction of $\phi$ to $E_3$ and $S_3=\phi(\boldsymbol{s}_3)-\phi(\boldsymbol{m}_3)$.

We go on with this procedure, proceeding in the order dictated by the elements of the set SSV, arranged in the decreasing order, until all $n_0$ local minima $\boldsymbol{m}$ have been enumerated. In this way we have associated each local minima to one ssp: to each local minimum $\boldsymbol{m}_k$, there is one critical component $E_k$ containing $\boldsymbol{m}_k$, and one ssp $\boldsymbol{s}_k$. We emphasize that in this procedure some of the saddle points (the noncritical ones) may not have been enumerated. For convenience, we enumerate these remaining saddle points from $n_0+1$ to $n_1+1$. Note that, with this labeling, $\mathcal{U}^{(1)} = \{\boldsymbol{s}_2, \dots, \boldsymbol{s}_{n_1+1}\}$. We then have

$$\text{minima} = \{\boldsymbol{m}_1, \dots, \boldsymbol{m}_{n_0}\}, \quad \text{SSP} = \{\boldsymbol{s}_1 = \infty, \boldsymbol{s}_2, \dots, \boldsymbol{s}_{n_0}\}.$$

We summarize the preceding discussion in the following proposition:

**Proposition A.2.** *The families of minima $\mathcal{U}^{(0)} = \{\boldsymbol{m}_k : k = 1, \dots, n_0\}$, separating saddle points $\{\boldsymbol{s}_k : k = 1, \dots, n_0\}$ and connected sets $\{E_k : k = 1, \dots, n_0\}$ satisfy*:

(i) $\boldsymbol{s}_1 = \infty$, $E_1 = \mathbb{R}^n$ *and $\boldsymbol{m}_1$ is the global minimum of $\phi$.*

(ii) *For every $k \geq 2$, $\overline{E}_k$ is compact, $E_k$ is the connected component containing $\boldsymbol{m}_k$ in*

$$\{x \in \mathbb{R}^n : \phi(x) < \phi(\boldsymbol{s}_k)\}$$

*and $\phi(\boldsymbol{m}_k) = \min_{E_k} \phi$.*

(iii) *If $\boldsymbol{s}_{k'} \in E_k$ for some $k, k' \in \{1, \dots, n_0\}$, then $k' > k$.*

To ensure that the eigenvalues $\lambda_k^\star$ are decreasing, if necessary we relabel the pairs of minima and critical saddle points so that the sequence $S_k$ is decreasing.

Using [Helffer et al. 2004; Hérau et al. 2011], we shall now introduce suitable refined quasimodes, adapted to the local minima of $\phi$ and the simplified labeling, described in Proposition A.2. Let $\varepsilon_0 > 0$ be such that the distance between critical points is larger than $10\varepsilon_0$ and such that, for every critical point $\boldsymbol{u}$ and $k \in \{1, \dots, n_0\}$, we have either $\boldsymbol{u} \in \overline{E}_k$ or $\operatorname{dist}(\boldsymbol{u}, \overline{E}_k) \geq 10\varepsilon_0$. Also let $C_0 > 1$, to be defined later, and note that $\varepsilon_0$ may also be taken smaller later. For $0 < \varepsilon < \varepsilon_0$ we build a family of functions $\chi_{k,\varepsilon}$, $k \in \{1, \dots, n_0\}$ as follows: for $k = 1$, we let $\chi_{1,\varepsilon} = 1$ and, for $k \geq 2$, we consider the open set $E_{k,\varepsilon} = E_k \setminus \overline{B(\boldsymbol{s}_k, \varepsilon)}$, and let $\chi_{k,\varepsilon}$ be a $C_0^\infty$-cutoff function supported in $E_{k,\varepsilon} + B(0, \varepsilon/C_0)$ and equal to 1 in $E_{k,\varepsilon} + B(0, \varepsilon/(2C_0))$. Then, we define the quasimodes for $1 \leq k \leq n_0$ by

$$f_k^{W,(0)} = b_k(h)\chi_{k,\varepsilon}(x)e^{-(\phi(x)-\phi(\boldsymbol{m}_k))/h}, \tag{A-1}$$

where $b_k$ is a normalization constant, given thanks to the stationary phase theorem by

$$b_k(h) = (\pi h)^{-d/4} \det(\operatorname{Hess}\phi(\boldsymbol{m}_k))^{1/4}(b_{k,0} + hb_{k,1} + \cdots), \quad b_{k,0} = 1.$$

Then, for $\varepsilon_0$ small enough and $C_0$ large enough, there exists $C > 0$ such that, for all $0 < \varepsilon < \varepsilon_0$, we have the following lemma:

**Lemma A.3.** *The system $(f_k^{W,(0)})$ is free and there exists $\alpha > 0$ uniform in $\varepsilon < \varepsilon_0$ such that*

$$\langle f_k^{W,(0)}, f_{k'}^{W,(0)} \rangle = \delta_{k,k'} + \mathcal{O}(e^{-\alpha/h}), \quad d_{\phi,h} f_k^{W,(0)} = \mathcal{O}(e^{-(S_k - C\varepsilon)/h}),$$

*and, in particular,*

$$P^{W,(0)} f_k^{W,(0)} = \mathcal{O}(e^{-\alpha/h}).$$

*Proof.* This is a direct consequence of the statement and proof of Proposition 5.3 in [Hérau et al. 2011]. □

## References

[Bovier et al. 2004] A. Bovier, M. Eckhoff, V. Gayrard, and M. Klein, "Metastability in reversible diffusion processes, I: Sharp asymptotics for capacities and exit times", *J. Eur. Math. Soc.* (*JEMS*) **6**:4 (2004), 399–424. MR 2006b:82112 Zbl 1105.82025

[Bovier et al. 2005] A. Bovier, V. Gayrard, and M. Klein, "Metastability in reversible diffusion processes, II: Precise asymptotics for small eigenvalues", *J. Eur. Math. Soc.* (*JEMS*) **7**:1 (2005), 69–99. MR 2006b:82113

[Cycon et al. 2008] H. L. Cycon, R. G. Froese, W. Kirsch, and B. Simon, *Schrödinger operators with application to quantum mechanics and global geometry*, Springer, Berlin, 2008. Corrected and extended 2nd printing. MR 88g:35003 Zbl 0619.47005

[Diaconis 2009] P. Diaconis, "The Markov chain Monte Carlo revolution", *Bull. Amer. Math. Soc.* (*N.S.*) **46**:2 (2009), 179–205. MR 2010b:60204 Zbl 1168.60032

[Diaconis and Lebeau 2009] P. Diaconis and G. Lebeau, "Micro-local analysis for the Metropolis algorithm", *Math. Z.* **262**:2 (2009), 411–447. MR 2010k:60255 Zbl 1178.60053

[Diaconis et al. 2011] P. Diaconis, G. Lebeau, and L. Michel, "Geometric analysis for the metropolis algorithm on Lipschitz domains", *Invent. Math.* **185**:2 (2011), 239–281. MR 2012f:47030 Zbl 1227.60093

[Dimassi and Sjöstrand 1999] M. Dimassi and J. Sjöstrand, *Spectral asymptotics in the semi-classical limit*, London Mathematical Society Lecture Note Series **268**, Cambridge University Press, 1999. MR 2001b:35237 Zbl 0926.35002

[Feller 1971] W. Feller, *An introduction to probability theory and its applications, II*, 2nd ed., John Wiley & Sons, New York, 1971. MR 42 #5292 Zbl 0219.60003

[Guillarmou and Michel 2011] C. Guillarmou and L. Michel, "Spectral analysis of random walk operators on Euclidean space", *Math. Res. Lett.* **18**:3 (2011), 405–424. MR 2012f:60159 Zbl 1256.47061

[Helffer and Nier 2005] B. Helffer and F. Nier, *Hypoelliptic estimates and spectral theory for Fokker–Planck operators and Witten Laplacians*, Lecture Notes in Mathematics **1862**, Springer, Berlin, 2005. MR 2006a:58039 Zbl 1072.35006

[Helffer and Sjöstrand 1984] B. Helffer and J. Sjöstrand, "Multiple wells in the semiclassical limit, I", *Comm. Partial Differential Equations* **9**:4 (1984), 337–408. MR 86c:35113 Zbl 0546.35053

[Helffer and Sjöstrand 1985] B. Helffer and J. Sjöstrand, "Puits multiples en mécanique semi-classique, IV: Étude du complexe de Witten", *Comm. Partial Differential Equations* **10**:3 (1985), 245–340. MR 87i:35162 Zbl 0597.35024

[Helffer et al. 2004] B. Helffer, M. Klein, and F. Nier, "Quantitative analysis of metastability in reversible diffusion processes via a Witten complex approach", *Mat. Contemp.* **26** (2004), 41–85. MR 2005i:58025 Zbl 1079.58025

[Hérau et al. 2011] F. Hérau, M. Hitrik, and J. Sjöstrand, "Tunnel effect and symmetries for Kramers–Fokker–Planck-type operators", *J. Inst. Math. Jussieu* **10**:3 (2011), 567–634. MR 2012h:35249 Zbl 1223.35246

[Hérau et al. 2013] F. Hérau, M. Hitrik, and J. Sjöstrand, "Supersymmetric structures for second order differential operators", *Algebra i Analiz* **25**:2 (2013), 125–154. In Russian; translated in *St. Petersbg. Math. J.* **25**:2 (2014), 241-263. MR 3114853 Zbl 1303.81086

[Holley and Stroock 1988] R. A. Holley and D. W. Stroock, "Simulated annealing via Sobolev inequalities", *Comm. Math. Phys.* **115**:4 (1988), 553–569. MR 90g:60091 Zbl 0643.60092

[Holley et al. 1989] R. A. Holley, S. Kusuoka, and D. W. Stroock, "Asymptotics of the spectral gap with applications to the theory of simulated annealing", *J. Funct. Anal.* **83**:2 (1989), 333–347. MR 92d:60081 Zbl 0706.58075

[Lebeau and Michel 2010] G. Lebeau and L. Michel, "Semi-classical analysis of a random walk on a manifold", *Ann. Probab.* **38**:1 (2010), 277–315. MR 2011f:58063 Zbl 1187.58033

[Lelièvre et al. 2010] T. Lelièvre, M. Rousset, and G. Stoltz, *Free energy computations: a mathematical perspective*, Imperial College Press, London, 2010. MR 2012c:82068 Zbl 1227.82002

[Martinez 2002] A. Martinez, *An introduction to semiclassical and microlocal analysis*, Springer, New York, 2002. MR 2003b:35010 Zbl 0994.35003

[Miclo 1992] L. Miclo, "Recuit simulé sur $\mathbf{R}^n$: étude de l'évolution de l'énergie libre", *Ann. Inst. H. Poincaré Probab. Statist.* **28**:2 (1992), 235–266. MR 93f:60118 Zbl 0747.60071

[Reed and Simon 1978] M. Reed and B. Simon, *Methods of modern mathematical physics, IV: Analysis of operators*, Academic Press, New York, 1978. MR 58 #12429c Zbl 0401.47001

[Simon 1979] B. Simon, *Trace ideals and their applications*, London Mathematical Society Lecture Note Series **35**, Cambridge University Press, 1979. MR 80k:47048 Zbl 0423.47001

[Zworski 2012] M. Zworski, *Semiclassical analysis*, Graduate Studies in Mathematics **138**, American Mathematical Society, Providence, RI, 2012. MR 2952218 Zbl 1252.58001

Received 22 Jan 2014. Revised 9 Jan 2015. Accepted 9 Feb 2015.

JEAN-FRANÇOIS BONY: bony@math.u-bordeaux1.fr
*Institut Mathématiques de Bordeaux, Université de Bordeaux, 351, cours de la Libération, 33405 Talence, France*

FRÉDÉRIC HÉRAU: frederic.herau@univ-nantes.fr
*Laboratoire de Mathématiques Jean Leray, Université de Nantes, 2, rue de la Houssinière, 44322 Nantes, France*

LAURENT MICHEL: lmichel@unice.fr
*Laboratoire J. A. Dieudonné, Université de Nice, Parc Valrose, 06000 Nice, France*

ANALYSIS AND PDE
Vol. 8, No. 2, 2015
dx.doi.org/10.2140/apde.2015.8.333

# TRAVELING WAVE SOLUTIONS IN A HALF-SPACE FOR BOUNDARY REACTIONS

XAVIER CABRÉ, NEUS CÓNSUL AND JOSÉ V. MANDÉ

We prove the existence and uniqueness of a traveling front and of its speed for the homogeneous heat equation in the half-plane with a Neumann boundary reaction term of unbalanced bistable type or of combustion type. We also establish the monotonicity of the front and, in the bistable case, its behavior at infinity. In contrast with the classical bistable interior reaction model, its behavior at the side of the invading state is of power type, while at the side of the invaded state its decay is exponential. These decay results rely on the construction of a family of explicit bistable traveling fronts. Our existence results are obtained via a variational method, while the uniqueness of the speed and of the front rely on a comparison principle and the sliding method.

## 1. Introduction

This paper concerns the problem

$$\begin{cases} v_t - \Delta v = 0 & \text{in } \mathbb{R}^2_+ \times (0,\infty), \\ \dfrac{\partial v}{\partial \nu} = f(v) & \text{on } \partial\mathbb{R}^2_+ \times (0,\infty) \end{cases} \tag{1-1}$$

for the homogeneous heat equation in a half-plane with a nonlinear Neumann boundary condition. To study the propagation of fronts given an initial condition, it is important to understand first the existence and properties of traveling fronts — or traveling waves — for (1-1). Taking $\mathbb{R}^2_+ = \{(x,y) \in \mathbb{R}^2 : x > 0\}$, these are solutions of the form $v(x,y,t) := u(x, y - ct)$ for some speed $c \in \mathbb{R}$. Thus, the pair $(c,u)$ must solve the elliptic problem

$$\begin{cases} \Delta u + c u_y = 0 & \text{in } \mathbb{R}^2_+ = \{(x,y) \in \mathbb{R}^2 : x > 0\}, \\ \dfrac{\partial u}{\partial \nu} = f(u) & \text{on } \partial\mathbb{R}^2_+, \end{cases} \tag{1-2}$$

where $\partial u/\partial\nu = -u_x$ is the exterior normal derivative of $u$ on $\partial\mathbb{R}^2_+ = \{x = 0\}$, $u$ is real valued, and $c \in \mathbb{R}$.

We look for solutions $u$ with $0 < u < 1$ and having the limits

$$\lim_{y\to-\infty} u(0,y) = 1 \quad \text{and} \quad \lim_{y\to+\infty} u(0,y) = 0. \tag{1-3}$$

Our results apply to nonlinearities $f$ of unbalanced bistable type or of combustion type, as defined next.

---

The authors were supported by grants MINECO MTM2011-27739-C04-01 and GENCAT 2009SGR-345.
*MSC2010:* 35J65, 35K57.
*Keywords:* traveling front, traveling wave, boundary reaction, homogeneous heat equation, bistable nonlinearity, combustion nonlinearity.

**Definition 1.1.** Let $f$ in $C^{1,\gamma}([0,1])$ for some $\gamma \in (0,1)$ satisfy

$$f(0) = f(1) = 0 \tag{1-4}$$

and, for some $\delta \in \left(0, \frac{1}{2}\right)$,

$$f' \leq 0 \quad \text{in } (0,\delta) \cup (1-\delta, 1). \tag{1-5}$$

(a) We say that $f$ is of positively balanced bistable type if it satisfies (1-4) and (1-5), it has a unique zero — named $\alpha$ — in (0, 1), and that it is "positively balanced" in the sense that

$$\int_0^1 f(s)\,ds > 0. \tag{1-6}$$

(b) We say that $f$ is of combustion type if it satisfies (1-4) and (1-5), and that there exists $0 < \beta < 1$ (called the ignition temperature) such that $f \equiv 0$ in $(0, \beta)$ and

$$f > 0 \quad \text{in } (\beta, 1). \tag{1-7}$$

In (1-2) one must find not only the solution $u$ but also the speed $c$, which is a priori unknown. We will establish that there is a unique speed $c \in \mathbb{R}$ for which (1-2) admits a solution $u$ satisfying the limits (1-3). For this speed $c$, using variational techniques we show the existence of a solution $u$ which is decreasing in $y$, with limits 1 and 0 at infinity on every vertical line. Moreover, we prove the uniqueness (up to translations in the $y$ variable) of a solution $u$ with limits 1 and 0 as $y \to \mp\infty$.

The speed $c$ of the front will be shown to be positive. Hence, since $c > 0$, we have that

$$v(x, y, t) = u(x, y - ct) \to 1 \quad \text{as } t \to +\infty.$$

That is, the state $u \equiv 1$ invades the state $u \equiv 0$.

For unbalanced bistable nonlinearities which satisfy in addition $f'(0) < 0$ and $f'(1) < 0$, we find the behaviors of the front at $y = \pm\infty$. In contrast with the classical bistable interior reaction model, its behavior at the side of the invading state $u = 1$, i.e., as $y \to -\infty$, is of power type, while its decay is exponential as $y \to +\infty$.

Our results are collected in the following theorem. Since $f$ is in $C^{1,\gamma}$, weak solutions to (1-2) can be shown to be classical, indeed $C^{2,\gamma}$ up to $\partial\mathbb{R}^2_+$. This is explained in the beginning of Section 4.

**Theorem 1.2.** *Let $f$ be of positively balanced bistable type or of combustion type as in* Definition 1.1. *We have*:

(i) *There exists a solution pair $(c, u)$ to* (1-2), *where $c > 0$, $0 < u < 1$, and $u$ has the limits* (1-3). *The solution $u$ lies in the weighted Sobolev space*

$$H^1_c(\mathbb{R}^2_+) := \left\{ w \in H^1_{loc}(\mathbb{R}^2_+) : \|w\|_c := \int_{\mathbb{R}^2_+} e^{cy}\{w^2 + |\nabla w|^2\}\,dx\,dy < \infty \right\}.$$

(ii) *Up to translations in the $y$ variable, $(c, u)$ is the unique solution pair to* (1-2) *among all constants $c \in \mathbb{R}$ and solutions $u$ satisfying* $0 \leq u \leq 1$ *and the limits* (1-3).

(iii) *For all* $x \geq 0$, $u$ *is decreasing in the* $y$ *variable and has limits* $u(x, -\infty) = 1$ *and* $u(x, +\infty) = 0$. *Furthermore*, $\lim_{x\to+\infty} u(x, y) = 0$ *for all* $y \in \mathbb{R}$. *If* $f$ *is of combustion type then we have*, *in addition*, $u_x \leq 0$ *in* $\mathbb{R}^2_+$.

(iv) *If* $f_1$ *is of positively balanced bistable type or of combustion type*, *if another nonlinearity* $f_2$ *is of the same type*, *with* $f_1 \geq f_2$ *and* $f_1 \not\equiv f_2$, *then their corresponding speeds satisfy* $c_1 > c_2$.

(v) *Assume that* $f$ *is of positively balanced bistable type and that*

$$f'(0) < 0 \quad \text{and} \quad f'(1) < 0. \tag{1-8}$$

*Then*, *there exists a constant* $b > 1$ *such that*:

$$\frac{1}{b}\frac{e^{-cy}}{y^{3/2}} \leq -u_y(0, y) \leq b\,\frac{e^{-cy}}{y^{3/2}} \qquad \text{for } y > 1, \tag{1-9}$$

$$\frac{1}{b}\frac{1}{(-y)^{3/2}} \leq -u_y(0, y) \leq b\,\frac{1}{(-y)^{3/2}} \qquad \text{for } y < -1, \tag{1-10}$$

$$\frac{1}{b}\frac{e^{-cy}}{y^{3/2}} \leq u(0, y) \leq b\,\frac{e^{-cy}}{y^{3/2}} \qquad \text{for } y > 1, \quad \text{and} \tag{1-11}$$

$$\frac{1}{b}\frac{1}{(-y)^{1/2}} \leq 1 - u(0, y) \leq b\,\frac{1}{(-y)^{1/2}} \qquad \text{for } y < -1. \tag{1-12}$$

*The lower bounds for* $-u_y$, $u$, *and* $1 - u$ *in* (1-9)–(1-12) *hold for any* $f$ *of positively balanced bistable type or of combustion type as in* Definition 1.1.

Our result on the existence of the traveling front will be proved using a variational method introduced by Steffen Heinze [2001]. It is explained later in this section. Heinze studied (1-2) in infinite cylinders of $\mathbb{R}^n$ instead of half-spaces. For these domains and for both bistable and combustion nonlinearities, he showed the existence of a traveling front. Using a rearrangement technique after making the change of variables $z = e^{ay}/a$, he also proved the monotonicity of the front. In addition, [Heinze 2001] found an interesting formula, (1-24), for the front speed in terms of the minimum value of the variational problem. The formula has interesting consequences, such as part (iv) of our theorem, the relation between the speeds for two comparable nonlinearities.

For our existence result, we will proceed as in [Heinze 2001]. The weak lower semicontinuity of the problem will be more delicate in our case due to the unbounded character of the problem in the $x$ variable — a feature not present in cylinders. The rearrangement technique will produce a monotone front. Its monotonicity will be crucial in order to establish that it has limits 1 and 0 as $y \to \mp\infty$. On the other hand, the front being in the weighted Sobolev space $H^1_c(\mathbb{R}^2_+)$ will lead easily to the fact that $u \to 0$ as $x \to +\infty$. While $u_x \leq 0$ in the case of combustion nonlinearities — as stated in part (iii) of the theorem — this property is not true for bistable nonlinearities since the normal derivative $-u_x = f(u)$ changes sign on $\{x = 0\}$.

The variational approach has another interesting feature. Obviously, the solutions that we produce in the half-plane are also traveling fronts for the same problem in a half-space $\mathbb{R}^n_+$ for $n \geq 3$: they only depend on two Euclidean variables. However, if the minimization problem is carried out directly in $\mathbb{R}^n_+$

for $n \geq 3$, then it produces a different type of solution that decays to 0 in all variables but one; see Remark 1.5 for more details.

In the case of combustion nonlinearities, (1-2) in a half-plane has been studied by Caffarelli, Mellet, and Sire [Caffarelli et al. 2012], a paper developed at the same time as most of our work. They establish the existence of a speed admitting a monotone front. As mentioned in [Caffarelli et al. 2012], our approaches towards the existence result are different. Their work does not use minimization methods, but instead approximation by truncated problems in bounded domains — as in [Berestycki and Nirenberg 1992]. They also rely in an interesting explicit formula for traveling fronts of a free boundary problem obtained as a singular limit of (1-2). In addition, [Caffarelli et al. 2012] establishes the following precise behavior of the combustion front at the side of the invaded state $u = 0$. For some constant $\mu_0 > 0$,

$$u(0, y) = \mu_0 \frac{e^{-cy}}{y^{1/2}} + \mathrm{O}\left(\frac{e^{-cy}}{y^{3/2}}\right) \quad \text{as } y \to +\infty \tag{1-13}$$

(here we follow our notation; [Caffarelli et al. 2012] reverses the states $u = 0$ and $u = 1$). The decay for combustion fronts is different than ours: $y^{1/2}$ in (1-13) is replaced by $y^{3/2}$ in the bistable case. Note however that the main order in the decays is $e^{-cy}$ and that the exponent $c$ depends in a highly nontrivial way on each nonlinearity $f$.

Uniqueness issues for the speed or for the front in (1-2) are treated for first time in the present paper. Our result on uniqueness of the speed and of the front relies heavily on the powerful sliding method of Berestycki and Nirenberg [1991]. We also use a comparison principle analogue to one in the paper by Cabré and Solà-Morales [2005], which studied (1-2) with $c = 0$. Among other things, [Cabré and Solà-Morales 2005] established the existence, uniqueness, and monotonicity of a front for (1-2) when $c = 0$ and $f$ is a balanced bistable nonlinearity. It was also shown there that, in the balanced bistable case, the front reaches its limits 1 and 0 at the power rate $1/|y|$. We point out that the variational method in the present paper requires $f$ to be unbalanced. It cannot be carried out in the balanced case.

Suppose now that $f$ satisfies the assumptions made above for bistable nonlinearities except for condition (1-6), and assume instead that

$$\int_0^1 f(s)\, ds \leq 0.$$

First, if the above integral is zero (i.e., $f$ is balanced), [Cabré and Solà-Morales 2005] established the existence of a monotone front for $f$ with speed $c = 0$. Suppose now that the above integral is negative. Then, the nonlinearity $\tilde{f}(s) := -f(1-s)$ has positive integral and is of bistable type. Thus, it produces a solution pair $(\tilde{c}, \tilde{u})$ for (1-2) with positive speed $\tilde{c}$. Then, if $u(x, y) := 1 - \tilde{u}(x, -y)$, $(-\tilde{c}, u)$ is a solution pair to (1-2) for the original $f$. The traveling speed $-\tilde{c}$ is now negative.

To prove our decay estimates as $y \to \pm\infty$, we use ideas from [Caffarelli et al. 2012; Cabré and Sire 2015]. The estimates rely on the construction of a family of explicit fronts for some bistable nonlinearities. Their formula and properties are stated in Theorem 1.3 below. To see how we construct these fronts, note that $u$ is a solution of (1-2) if and only if its trace $v(y) := u(0, y)$ solves the fractional diffusion equation

$$(-\partial_{yy} - c\partial_y)^{1/2} v = f(v) \quad \text{in } \mathbb{R} \qquad \text{for } v(y) := u(0, y). \tag{1-14}$$

This follows from two facts. First, if $u$ solves the first equation in (1-2), then so does $-u_x$. Second, we have $(-\partial_x)^2 = \partial_{xx} = -\partial_{yy} - c\partial_y$. Our main result states that there is a unique $c \in \mathbb{R}$ for which the fractional equation (1-14) admits a solution connecting 1 and 0.

As in the paper by Cabré and Sire [2015], which studied problem $(-\partial_{yy})^s v = f(v)$ in $\mathbb{R}$ for balanced bistable nonlinearities, the construction of explicit fronts will be based on the fundamental solution for the homogeneous heat equation associated to the fractional operator in (1-14), that is, equation

$$\partial_t v + (-\partial_{yy} - c\partial_y)^{1/2} v = 0.$$

The process to find such heat kernel uses an idea from [Caffarelli et al. 2012], and it is explained in Section 6 below.

Regarding the resulting decays (1-11) and (1-12) for bistable fronts, note that the exponential decay at the side of the invaded state $u = 0$ is much faster than the power decay (1-12) at the side of the invading state $u = 1$. This large difference of rates is clearly seen in the explicit fronts that we built. See Figure 1 for the plots of one such front, where the much steeper decay on the right is clearly seen.

These decays are also in contrast with the classical ones for the bistable equation $u_{yy} + cu_y + f(u) = 0$ in $\mathbb{R}$, which are both pure exponentials — with exponents that may be different at $+\infty$ and $-\infty$. Note however that the exponent $c$ in the exponential term at $+\infty$ for our problem will be different, in general, than the corresponding exponent in the classical case — taking the same nonlinearity $f$ for both problems.

The next theorem concerns the explicit bistable fronts that we construct. They will lead to the decay bounds of Theorem 1.2 for general bistable fronts. They involve the modified Bessel function of the second kind $K_1$ with index $\nu = 1$. We recall that $K_1(s)$ is a positive and decreasing function of $s > 0$ (see [Abramowitz and Stegun 1964]).

**Theorem 1.3.** *For every $c > 0$ and $t > 0$, let*

$$u^{t,c}(x, y) := u^t\left(\frac{c}{2}x, \frac{c}{2}y\right),$$

*where*

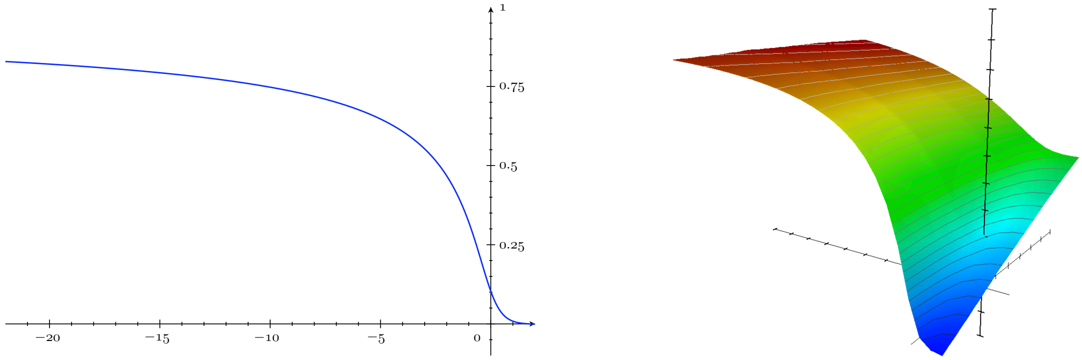
$$u^t(x, y) := \int_y^{+\infty} e^{-z} \frac{x+t}{\pi\sqrt{(x+t)^2 + z^2}} K_1(\sqrt{(x+t)^2 + z^2})\, dz$$

*and $K_1$ is the modified Bessel function of the second kind with index $\nu = 1$.*

*Then, there exists a nonlinearity $f^{t,c}$ of positively balanced bistable type for which $(c, u^{t,c})$ is the unique solution pair to* (1-2) *with $0 \le u^{t,c} \le 1$ satisfying the limits* (1-3). *In addition, we have*

$$(f^{t,c})'(0) = (f^{t,c})'(1) = -\frac{c}{2t}.$$

*and that $f^{t,c} = (c/2) f^t$ for a nonlinearity $f^t$ independent of $c$.*




$x = 0$ $\qquad$ $0 \le x \le 2$

**Figure 1.** The explicit bistable front $u^1$, for $y \in [-22, 2]$. Left: $x = 0$. Right: $x \in [0, 2]$.

*Furthermore, on* $\partial\mathbb{R}^2_+$ *the derivative of* $u^{t,c}$ *satisfies*

$$-u_y^{t,c}(0, y) = \frac{t}{(\pi c)^{1/2}} \frac{e^{-cy}}{y^{3/2}} + o\left(\frac{e^{-cy}}{y^{3/2}}\right) \quad \textit{as } y \to +\infty, \quad \textit{and}$$
$$-u_y^{t,c}(0, y) = \frac{t}{(\pi c)^{1/2}} \frac{1}{(-y)^{3/2}} + o\left(\frac{1}{(-y)^{3/2}}\right) \quad \textit{as } y \to -\infty.$$

Figure 1 shows plots of the explicit bistable front $u^1 = u^{1,2}$, a front with speed $c = 2$. In both of them we have $-22 \le y \le 2$. The much faster decay for positive values of $y$ than for negative values is clearly appreciated. In the three-dimensional plot, the steepest profile corresponds to $x = 0$ while the profile in the back of the picture is for $x = 2$.

Kyed [2008] also studies problem (1-2) in infinity cylinders of $\mathbb{R}^3$. It deals with nonlinearities $f$ that vanish only at 0 and that appear in models of boiling processes. Kyed also uses the variational principle of Heinze. In addition, [Kyed 2008] contains some exponential decay bounds. Landes [2009; 2012] studies problem (1-1) in finite cylinders, with special interest in bistable nonlinearities. These articles establish the presence of wavefront-type solutions for some initial conditions and give in addition bounds for their propagation speed. For this, appropriate sub- and supersolutions are constructed.

The variational method in [Heinze 2001] has also been used by Lucia, Muratov, and Novaga [Lucia et al. 2004] to study the classical interior reaction equation $u_{yy} + cu_y + f(u) = 0$ for monostable-type nonlinearities $f$. Their paper gives a very interesting characterization of the phenomenon of linear versus nonlinear selection for the front speed.

In relation to fractional diffusions — such as (1-14) — the existence of traveling fronts for

$$\partial_t v + (-\partial_{yy})^s v = f(v) \quad \text{in } \mathbb{R} \tag{1-15}$$

has been established in [Mellet et al. 2010] when $s \in \left(\frac{1}{2}, 1\right)$ and $f$ is a combustion nonlinearity. This article also shows that $v$ tends to 0 at $+\infty$ at the power rate $1/|y|^{2s-1}$. Note that the equation for traveling fronts of (1-15) is

$$\{(-\partial_{yy})^s - c\partial_y\}v = f(v) \quad \text{in } \mathbb{R},$$

which should be compared with (1-14). In the case of bistable nonlinearities, [Gui and Zhao 2014] establishes that (1-15) admits a unique traveling front and a unique speed for any $s \in (0, 1)$. In contrast with the decay in [Mellet et al. 2010] for combustion nonlinearities, in the bistable case [Gui and Zhao 2014] shows that the front reaches its two limiting values at the rate $1/|y|^{2s}$ — as in [Cabré and Solà-Morales 2005; Cabré and Sire 2015] for balanced bistable nonlinearities.

Next, let us describe the structure of the variational problem that will lead to our existence result. First, we enumerate the five concrete properties of the nonlinearity needed for all the results of the paper.

**Remark 1.4.** Even though we state our main result for bistable- and combustion-type nonlinearities (for the clarity of reading), all our proofs require only the following five conditions on $f$:

$$\text{(1-4), (1-5), (1-6), (1-7),} \quad \text{and} \quad \int_0^s f(\sigma)\,d\sigma \le 0 \quad \text{for all } s \in (0, \beta). \tag{1-16}$$

We claim that both positively balanced bistable nonlinearities and combustion nonlinearities as in Definition 1.1 satisfy the above five assumptions.

The claim is easily seen. For a combustion nonlinearity, (1-6) is obviously true, while the last condition in (1-16) holds (indeed with an equality) for the same $\beta$ as in part (b) of the definition. On the other hand, for $f$ of bistable type, since $f$ has a unique zero $\alpha$ in (0, 1) and (1-5) holds, it follows that $f < 0$ in $(0, \alpha)$ and $f > 0$ in $(\alpha, 1)$. Thus, by (1-6) there exists a unique $\beta \in (\alpha, 1)$ such that $\int_0^\beta f(s)\,ds = 0$. As a consequence, (1-7) and the last condition in (1-16) hold for such $\beta$.

To describe the potential energy of our problem, we first extend $f$ linearly to $(-\infty, 0)$ and to $(1, +\infty)$, keeping its $C^{1,\gamma}$ character. Consider now the potential $G \in C^2(\mathbb{R})$ defined by

$$G(s) := -\int_0^s f(\sigma)\,d\sigma \quad \text{for } s \in \mathbb{R}.$$

Note that $G' = -f$ in [0, 1]. Since $f'(0) \le 0$ and $f'(1) \le 0$ due to hypothesis (1-5), we have that

$$G(s) \ge \begin{cases} 0 = G(0) & \text{for } s \le 0, \\ G(1) & \text{for } s \ge 1. \end{cases} \tag{1-17}$$

Two other important properties of $G$ are the following. First, by (1-6), we have

$$G(1) < G(0) = 0 \quad \text{and} \quad G'(0) = -f(0) = 0. \tag{1-18}$$

Second, the last condition in (1-16) reads

$$G \ge 0 \quad \text{in } [0, \beta]. \tag{1-19}$$

On the other hand, since $G \in C^2(\mathbb{R})$ and $G(0) = G'(0) = 0$, we have that

$$-Cs^2 \le G(s) \le Cs^2 \quad \text{for all } s \in \mathbb{R}, \tag{1-20}$$

for some constant $C$.

Figure 2 shows the shape of the potential $G$ for a typical positively balanced bistable nonlinearity.

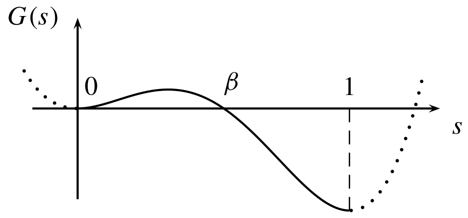



**Figure 2.** The potential $G$ for a positively balanced bistable $f$.

For $a > 0$, consider the weighted Sobolev space $(H^1_a(\mathbb{R}^2_+), \|\ \|_a)$ defined by

$$H^1_a(\mathbb{R}^2_+) = \{w \in H^1_{\mathrm{loc}}(\mathbb{R}^2_+) : \|w\|_a < \infty\},$$

where the norm $\|\ \|_a$ is defined by

$$\|w\|^2_a = \int_{\mathbb{R}^2_+} e^{ay}\{w^2 + |\nabla w|^2\}\, dx\, dy.$$

Notice that, by first truncating and then smoothing, the set $C^\infty_c(\overline{\mathbb{R}^2_+})$ — smooth functions with compact support in $\overline{\mathbb{R}^2_+}$ — is dense in $H^1_a(\mathbb{R}^2_+)$ with the $\|\ \|_a$ norm.

In both the bistable and combustion cases, the traveling front $u$ will be constructed from a minimizer $\underline{u}$ to the constraint problem

$$E_a(\underline{u}) = \inf_{w \in B_a} E_a(w) =: I_a \tag{1-21}$$

after scaling its independent variables $x$ and $y$. This is the method introduced by Heinze [2001] to study (1-2) in cylinders instead of half-spaces. The energy functional

$$E_a(w) = \frac{1}{2}\int_{\mathbb{R}^2_+} e^{ay}|\nabla w|^2\, dx\, dy + \int_{\partial\mathbb{R}^2_+} e^{ay} G(w(0, y))\, dy \tag{1-22}$$

will be minimized over the submanifold

$$B_a = \{w \in H^1_a(\mathbb{R}^2_+) : \Gamma_a(w) = 1\},$$

where

$$\Gamma_a(w) = \int_{\mathbb{R}^2_+} e^{ay}|\nabla w|^2\, dx\, dy.$$

To carry out this program, we will need to take a constant $a > 0$ small enough, depending only on $f$.

Note an important feature of the functionals $E_a$ and $\Gamma_a$. For $w \in H^1_a(\mathbb{R}^2_+)$ and $t \in \mathbb{R}$, define

$$w^t(x, y) := w(x, y + t)$$

(throughout the paper there is no risk of confusion with the same notation used for the explicit front $u^t$ of Theorem 1.3). We then have

$$E_a(w^t) = e^{-at} E_a(w) \quad \text{and} \quad \Gamma_a(w^t) = e^{-at}\Gamma_a(w). \tag{1-23}$$

The shape of the potential $G$ will lead to the existence of functions $u$ in $H^1_a(\mathbb{R}^2_+)$ with negative energy $E_a(u) < 0$. This will be essential in order to prove that our variational problem attains its infimum. In addition, the constraint will introduce a Lagrange multiplier and, through it, the a priori unknown speed $c$ of the traveling front.

As already noted by Heinze [2001], the above variational method produces an interesting formula for the speed $c$. One has

$$c = a(1 - 2I_a), \tag{1-24}$$

where $I_a$ is the minimum value (1-21) of the constraint variational problem; see Remark 2.8 below. This formula leads to the comparison result between the front speeds for two different comparable nonlinearities — part (iv) of Theorem 1.2. Note also that the value $a(1-2I_a)$ in (1-24) does not depend on which constant $a$ is chosen to carry out the minimization problem. The reason is that this value coincides with the speed $c$, and we prove uniqueness of the speed.

**Remark 1.5.** Obviously, the traveling front found in our paper on $\mathbb{R}^2_+$ is also a traveling front for problem (1-1) in the half-space $\mathbb{R}^n_+$, $n \geq 3$, traveling in any given unit direction $e$ of $\mathbb{R}^{n-1}$. That is, it is a solution of the problem

$$\begin{cases} \Delta u + c\partial_e u = 0 & \text{in } \mathbb{R}^n_+, \\ \dfrac{\partial u}{\partial \nu} = f(u) & \text{on } \partial\mathbb{R}^n_+, \end{cases}$$

where $\mathbb{R}^n_+ := \{(x, y) \in (0, +\infty) \times \mathbb{R}^{n-1}\}$. An interesting application of the constraint minimization method is that, when carried out in $\mathbb{R}^n_+$ and $n \geq 3$, it produces another type of traveling front. Their trace $u(0, \cdot)$ will not depend only on one Euclidean variable $e$ of $\mathbb{R}^{n-1}$ (as the solutions in the present paper do), but they will be monotone with limits 1 and 0 at infinity in the direction $e$ and will be even with limits 0 at $\pm\infty$ on the $y$ variables orthogonal to $e$. The reason is that here one minimizes the energy functional

$$E_a(w) = \frac{1}{2} \int_{\mathbb{R}^n_+} e^{a\cdot y} |\nabla w|^2 \, dx \, dy + \int_{\partial\mathbb{R}^n_+} e^{a\cdot y} G(w) \, dy$$

under the constraint on the Dirichlet energy, where $a \in \mathbb{R}^{n-1}$ is a nonunitary direction parallel to $e$. Now, note that the solutions built in this paper, which are constant in the $y$ variables orthogonal to $a$, do not belong to the corresponding energy space (since they have infinite Dirichlet energy).

The article is organized as follows. In Section 2 we study the variational structure of problem (1-2) and prove the existence of a solution pair. Section 3 uses the variational characterization and a monotone decreasing rearrangement of the minimizer to show that the front may be taken to be monotone in the $y$ direction. In Section 4 we establish the limits at infinity for the obtained solution. In Section 5 we prove a monotonicity and comparison result by means of a maximum principle and the sliding method. This result is the key ingredient to prove uniqueness of speed and of the front. Section 6 deals with the explicit fronts and supersolutions; here we give the proof of Theorem 1.3. Finally, in Section 7 we collect all results in the paper to establish Theorem 1.2.

## 2. The variational solution

***Two inequalities.*** We prove trace and Poincaré-type inequalities for functions in $H_a^1(\mathbb{R}_+^2)$. The existence of a lower bound for $E_a$ on $B_a$ will be a consequence of the following lemma:

**Lemma 2.1.** *Let $a > 0$. Then, for every $u \in H_a^1(\mathbb{R}_+^2)$, we have*

$$\int_{\mathbb{R}} e^{ay} u^2(0, y)\, dy \le \|u\|_a^2 = \int_{\mathbb{R}_+^2} e^{ay} \{u^2 + |\nabla u|^2\}\, dx\, dy$$

*and*

$$\int_{\mathbb{R}_+^2} e^{ay} u^2\, dx\, dy \le \frac{4}{a^2} \int_{\mathbb{R}_+^2} e^{ay} |\nabla u|^2\, dx\, dy.$$

*Proof.* By density, it suffices to establish both inequalities for $u \in C_c^\infty(\overline{\mathbb{R}_+^2})$. Since $u$ has compact support, we have

$$\int_{\mathbb{R}} e^{ay} u^2(0, y)\, dy = -\int_{\mathbb{R}_+^2} (e^{ay} u^2)_x\, dx\, dy = -2 \int_{\mathbb{R}_+^2} e^{ay} u u_x\, dx\, dy \le \int_{\mathbb{R}_+^2} e^{ay} (u^2 + u_x^2)\, dx\, dy.$$

This proves the first inequality.

Next, for every $x \ge 0$ we have that

$$\int_{\mathbb{R}} e^{ay} u^2(x, y)\, dy = -\frac{2}{a} \int_{\mathbb{R}} e^{ay} u(x, y) u_y(x, y)\, dy.$$

Thus, by Cauchy–Schwarz,

$$\int_{\mathbb{R}} e^{ay} u^2(x, y)\, dy \le \frac{4}{a^2} \int_{\mathbb{R}} e^{ay} u_y^2(x, y)\, dy.$$

Integrating in $x$ from 0 to $\infty$, we obtain the second inequality. □

***Construction of functions with negative energy.*** It is fundamental to show that $E_a$ takes a negative value somewhere on the constraint $B_a$. This will be accomplished by constructing test functions $u_0 \in H_a^1(\mathbb{R}_+^2)$ for which $E_a(u_0) < 0$ if $a$ is positive and small enough. We undertake this task next.

Let $u_0$ be defined by

$$u_0(x, y) := e^{-dx} h(y), \tag{2-1}$$

where $h$ is given by

$$h(y) := \begin{cases} 1 & \text{if } y \le 0, \\ e^{-amy} & \text{if } y > 0, \end{cases}$$

and the values of $d > 0$ and $m \ge 1$ are to be determined.

The following proposition applies to a class of nonlinearities which includes those of positively balanced bistable type and of combustion type.

**Proposition 2.2.** *Let $G$ satisfy* (1-18). *Let $a > 0$, $E_a$ be defined by* (1-22), *and $u_0$ by* (2-1). *Then, for small positive values of $a$ and $d$, and for large values of $m$ (all depending only on $f$), we have that $u_0 \in H^1_a(\mathbb{R}^2_+)$ and $E_a(u_0) < 0$. In addition,*

$$-\infty < \inf_{w \in B_a} E_a(w) < 0.$$

*Proof.* A simple calculation for the Dirichlet energy shows that, for $m \geq 1$,

$$\Gamma_a(u_0) = \frac{d}{2a}\left(1 + \frac{1}{2m-1}\right) + \frac{am^2}{2d(2m-1)}.$$

The potential energy can be computed as follows:

$$\begin{aligned}\int_{\mathbb{R}} e^{ay} G(u_0(0, y))\, dy &= \frac{G(1)}{a} + \int_0^{+\infty} e^{ay} G(e^{-amy})\, dy \\ &= \frac{1}{a}\left\{G(1) + \int_0^{+\infty} (e^{ay})' G(e^{-amy})\, dy\right\} \\ &= \frac{1}{a}\int_0^{+\infty} e^{ay} f(e^{-amy})(-am)e^{-amy}\, dy = -\frac{1}{a}\int_0^1 s^{-1/m} f(s)\, ds,\end{aligned}$$

where we have used property (1-20) of $G$ in order to integrate by parts. Note that (1-20) follows from assumption (1-18). Therefore,

$$aE_a(u_0) = \frac{d}{4}\left(1 + \frac{1}{2m-1}\right) + \frac{a^2m^2}{4d(2m-1)} - \int_0^1 s^{-1/m} f(s)\, ds.$$

Note that

$$\lim_{m \to +\infty} \int_0^1 s^{-1/m} f(s)\, ds = \int_0^1 f(s)\, ds > 0,$$

since $G(1) < G(0)$. It follows that $E_a(u_0) < 0$ if we first choose $m$ large enough, then choose $d$ small enough to make the first term above small, and finally $a$ also small to handle the second term.

It follows from property (1-23) that there exists a unique value $t$ such that $\Gamma_a(u_0^t) = e^{-at}\Gamma_a(u_0) = 1$. Since $u_0^t \in B_a$ and $E_a(u_0^t) = e^{-at} E_a(u_0) < 0$, we have shown that

$$\inf_{w \in B_a} E_a(w) < 0.$$

As a consequence of the two inequalities in Lemma 2.1 and of (1-20), we obtain that $E_a$ is bounded below on $B_a$. Therefore

$$-\infty < \inf_{w \in B_a} E_a(w) < 0,$$

as claimed. □

***A special minimizing sequence.*** To establish that our constraint variational problem (1-21) achieves its infimum it will be important to work with the following type of minimizing sequences.

**Lemma 2.3.** *Let $G$ satisfy* (1-17), (1-18), *and* (1-20). *Then, there exists a minimizing sequence $\{u_k\} \subset B_a$ of problem* (1-21) *such that, for all $k$, $u_k \in C_c(\overline{\mathbb{R}^2_+})$ has compact support in $\overline{\mathbb{R}^2_+}$ and $0 \leq u_k \leq 1$.*

*Proof.* By the previous proposition we know that

$$-\infty < I_a := \inf_{w \in B_a} E_a(w) < 0.$$

Let $\{w_k\} \subset B_a$ be any minimizing sequence. Approximating each $w_k$ by a function $v_k$ in $C_c^\infty(\overline{\mathbb{R}^2_+})$ (by first truncating and then smoothing), we may assume that $\lim_{k\to+\infty} E_a(v_k) = I_a$ and $\Gamma_a(v_k) = 1 + \tau_k$ with $\tau_k \to 0$. Consider now $t_k = \log\{(1+\tau_k)^{1/a}\}$. As a consequence of (1-23), we obtain $\Gamma_a(v_k^{t_k}) = 1$ — thus $\{v_k^{t_k}\} \subset B_a \cap C_c^\infty(\overline{\mathbb{R}^2_+})$ — and $\lim_{k\to+\infty} E_a(v_k^{t_k}) = I_a$.

Next, to show that we can restrict ourselves to minimizing sequences taking values in [0, 1], let us rename $\{v_k^{t_k}\}$ by $\{v_k\}$. We truncate $\{v_k\}$ and define $\tilde{v}_k$ by

$$\tilde{v}_k = \begin{cases} 0 & \text{if } v_k < 0, \\ v_k & \text{if } v_k \in [0,1], \\ 1 & \text{if } v_k > 1. \end{cases}$$

It is easy to see that $\tilde{v}_k \in C_c(\overline{\mathbb{R}^2_+}) \cap H^1_a(\mathbb{R}^2_+)$, $\Gamma_a(\tilde{v}_k) \le \Gamma_a(v_k)$ and, using (1-17), $E_a(\tilde{v}_k) \le E_a(v_k)$.

Next we claim that we may choose $s_k \le 0$ so that $u_k(x, y) := \tilde{v}_k(x, y + s_k)$ satisfies $\{u_k\} \subset B_a$. This claim follows from (1-23) and the fact that $0 < \Gamma_a(\tilde{v}_k) \le 1$ for $k$ large. To show this last assertion, note that $\Gamma_a(\tilde{v}_k) \le 1$ is a consequence of $\Gamma_a(\tilde{v}_k) \le \Gamma_a(v_k) = 1$. On the other hand, if $\Gamma_a(\tilde{v}_k) = 0$ then $\tilde{v}_k \equiv 0$ and thus $v_k \le 0$. From this and (1-17), we would get $E_a(v_k) \ge 0$. This is a contradiction if $k$ is large, since $I_a < 0$ and $\{v_k\}$ is a minimizing sequence.

Finally, since $s_k \le 0$ and $E_a(\tilde{v}_k) \le E_a(v_k) < 0$ for $k$ large, (1-23) gives $E_a(u_k) \le E_a(\tilde{v}_k) \le E_a(v_k)$. Therefore, $\{u_k\} \subset B_a$ is a minimizing sequence made of continuous functions with compact support and satisfying $0 \le u_k \le 1$. □

***Weak lower semicontinuity.*** Due to the unbounded character of $\mathbb{R}^2_+$, a delicate issue in this paper is to prove the weak lower semicontinuity (WLSC) of $E_a$ in $B_a$. The key point is to establish the WLSC result for the potential energy — the difficulty being that the potential $G(s)$ is negative near $s = 1$.

Note that for any sequence $\{u_k\} \in H^1_a(\mathbb{R}^2_+)$ converging weakly in $H^1_a(\mathbb{R}^2_+)$ to a function $\underline{u}$, $u_k \rightharpoonup \underline{u}$, we have $\Gamma_a(\underline{u}) \le \liminf_{k\to\infty} \Gamma_a(u_k)$. Also, if we split $G = G^+ - G^-$ into its positive and negative parts, Fatou's lemma gives

$$\int_{\partial\mathbb{R}^2_+} e^{ay} G^+(\underline{u})\, dy \le \liminf_{k\to\infty} \int_{\partial\mathbb{R}^2_+} e^{ay} G^+(u_k)\, dy.$$

Thus, we need to study the convergence of

$$\int_{\partial\mathbb{R}^2_+} e^{ay} G^-(u_k)\, dy.$$

To do this, the key observation (that already appears in [Heinze 2001]) is that any minimizing sequence cannot spend "too much time" (time meaning "positive $y$ variable") in $(\beta, 1)$, where $G$ may be negative, even if the state $u = 1$ invades $u = 0$. This key fact will be a consequence of the presence of the weight $e^{ay}$.

We will see that, for any given $R > 0$ and any minimizing sequence $\{u_k\}$, the Lebesgue measure of the sets $\{y > R : u_k(0, y) \geq \beta\}$ and $\{x > 0,\ y > R : u_k(x, y) \geq \beta\}$ both decrease to zero as $R \to +\infty$, uniformly in $k$.

To proceed from this, our analysis must be more delicate than in [Heinze 2001] — which deals with cylinders — due to the unbounded character of $\mathbb{R}^2_+$ in the $x$ variable. To handle this difficulty, we need to recall some facts about Riesz potentials; see [Gilbarg and Trudinger 1983, Section 7.8]. Let $\Omega$ be a bounded domain in $\mathbb{R}^2$. Consider the operator on $L^2(\Omega)$ defined by

$$Vw(z) := \int_\Omega \frac{w(\bar{z})}{|z - \bar{z}|}\, d\bar{z}.$$

It is well known (see [Gilbarg and Trudinger 1983, Section 7.8]) that

$$\|Vw\|_{L^2(\Omega)} \leq 2\sqrt{\pi}\,|\Omega|^{1/2}\,\|w\|_{L^2(\Omega)}. \tag{2-2}$$

Next, we use this inequality to prove a proposition that will be important to control the superlevel sets of $u_k$ mentioned above. The proposition will be applied to the functions $v = e^{ay/2}(u_k - \beta)^+$.

**Proposition 2.4.** *Given any constant $a > 0$, let $v \in C_c(\overline{\mathbb{R}^2_+}) \cap H^1_a(\mathbb{R}^2_+)$ have compact support with $\operatorname{supp}(v) \subset \overline{\Omega}$ for some bounded domain $\Omega \subset \mathbb{R}^2_+$. For any $R > 0$, define*

$$\Omega_R := \Omega \cap \{(x, y) \in \mathbb{R}^2_+ : y > R\} \quad \textit{and} \quad \partial^0\Omega_R := \overline{\Omega_R} \cap \partial\mathbb{R}^2_+.$$

*Then, we have*

$$\|v\|_{L^2(\Omega_R)} \leq \frac{4}{\sqrt{\pi}}|\Omega_R|^{1/2}\|\nabla v\|_{L^2(\Omega_R)} \quad \textit{and} \quad \left(\int_{\partial^0\Omega_R} v^2(0, y)\, dy\right)^{1/2} \leq \frac{\sqrt{8}}{\sqrt[4]{\pi}}\,|\Omega_R|^{1/4}\|\nabla v\|_{L^2(\Omega_R)}.$$

*Proof.* By density, it is enough to consider $v \in C^\infty_c(\overline{\mathbb{R}^2_+})$. Since $v$ has compact support, for $z \in \Omega_R$ and $\omega = (\omega_1, \omega_2) \in S^1$ with $\omega_1 > 0$ and $\omega_2 > 0$, we have

$$v(z) = -\int_0^\infty D_r v(z + r\omega)\, dr.$$

Integrating with respect to $\omega$ on the quarter of circle

$$S^1_+ := \{\omega \in S^1 : \omega_1 > 0 \text{ and } \omega_2 > 0\},$$

we obtain

$$v(z) = -\frac{2}{\pi}\int_0^\infty dr \int_{S^1_+} D_r v(z + r\omega)\, d\omega.$$

This leads to

$$|v(z)| \leq \frac{2}{\pi}\int_{\Omega_R} \frac{|\nabla v(\bar{z})|}{|z - \bar{z}|}\, d\bar{z}.$$

From this, the first inequality in Proposition 2.4 is now a consequence of (2-2).

As for the second inequality, we have

$$\int_{\partial^0\Omega_R} v^2(0,y)\,dy = -\int_{\Omega_R}(v^2)_x\,dx\,dy = -2\int_{\Omega_R} vv_x\,dx\,dy$$
$$\leq 2\|v\|_{L^2(\Omega_R)}\|\nabla v\|_{L^2(\Omega_R)} \leq \frac{8}{\sqrt{\pi}}|\Omega_R|^{1/2}\|\nabla v\|^2_{L^2(\Omega_R)},$$

as claimed. □

Next, note that, for any fixed $R > 0$, the embedding

$$H^1_a(\mathbb{R}^2_+\cap\{y<R\})\cap L^\infty(\mathbb{R}^2_+\cap\{y<R\}) \hookrightarrow L^2_a(\partial\mathbb{R}^2_+\cap\{y<R\})$$

is compact ($L^2_a$ is the $L^2$ space for the measure $e^{ay}dy$). Indeed, a bounded sequence in $H^1_a(\mathbb{R}^2_+\cap\{y<R\})$ is also bounded in $H^1((0,1)\times(-M,R))$ for all $M > 0$. This last space is compactly embedded in $L^2(\partial\mathbb{R}^2_+\cap\{-M<y<R\})$, and thus also in $L^2_a(\partial\mathbb{R}^2_+\cap\{-M<y<R\})$. In addition, since the sequence of functions is bounded in $L^\infty(\partial\mathbb{R}^2_+)$, their $L^2_a(\partial\mathbb{R}^2_+\cap\{-\infty<y<-M\})$ norms are as small as wished as $M\to\infty$ — since $e^{ay}\leq e^{-aM}$ in this set.

Thanks to the previous compact embedding, to achieve the desired WLSC result for $E_a$, it is enough to prove that

$$\int_R^{+\infty} e^{ay}G^-(u_k(0,y))\,dy$$

can be made — uniformly on $k$ — as small as we want, provided that $R$ is large enough. This is the content of the next proposition.

**Proposition 2.5.** *Let $G$ satisfy* (1-19), *i.e., $G\geq 0$ in $[0,\beta]$. Let $\{u_k\}\subset B_a\subset H^1_a(\mathbb{R}^2_+)$ be a minimizing sequence for $\inf\{E_a(w): w\in B_a\}$ such that $u_k\in C_c(\overline{\mathbb{R}^2_+})$ and $0\leq u_k\leq 1$ for all $k$.*

*Then, given $\varepsilon>0$, there exists $R>0$ such that we have*

$$\int_R^{+\infty} e^{ay}G^-(u_k(0,y))\,dy\leq\varepsilon$$

*for all $k$.*

*Proof.* For $R>0$, we define

$$A_k := \{(x,y)\in\mathbb{R}^2_+ : y>R,\ u_k(x,y)>\beta\}.$$

We can estimate the measures of $\partial^0A_k$ and $A_k$ respectively as follows (recall that the notation $\partial^0$ was introduced in Proposition 2.4). First,

$$e^{aR}|\partial^0A_k|\leq\int_{\partial^0A_k}e^{ay}\,dy\leq\frac{1}{\beta^2}\int_{\partial^0A_k}e^{ay}u_k^2(0,y)\,dy\leq\frac{C}{\beta^2}.$$

The last inequality is a consequence of $\Gamma_a(u_k)=1$ and the trace inequality in Lemma 2.1. In what follows, $C$ denotes different constants depending only on $a$ (and thus not on $k$). Therefore we have

$$|\partial^0A_k|\leq\frac{C}{\beta^2}e^{-aR}$$

and

$$\int_{\partial^0 A_k} e^{ay}\,dy \le \frac{C}{\beta^2}. \tag{2-3}$$

In an analogous way — integrating now on all of $A_k$ and not on its boundary, and using again Lemma 2.1 — we obtain that

$$|A_k| \le \frac{C}{\beta^2}\,e^{-aR}. \tag{2-4}$$

By (1-19), there exists a constant $C$ such that $G^-(s) \le C(s-\beta)^+$ for $s \in [0, 1]$. This and $0 \le u_k \le 1$ lead to

$$\begin{aligned}\int_R^{+\infty} e^{ay} G^-(u_k(0, y))\,dy &= \int_{\partial^0 A_k} e^{ay} G^-(u_k)\,dy \le C \int_{\partial^0 A_k} e^{ay}(u_k-\beta)\,dy \\ &\le C\biggl(\int_{\partial^0 A_k} e^{ay}(u_k-\beta)^2\,dy\biggr)^{1/2}\biggl(\int_{\partial^0 A_k} e^{ay}\,dy\biggr)^{1/2}.\end{aligned}$$

Because of (2-3) above, the last factor is bounded by $C/\beta$, a constant independent of $k$. Using the second inequality in Proposition 2.4 applied to the function $e^{ay/2}(u_k-\beta)^+$, we get

$$\int_R^{+\infty} e^{ay} G^-(u_k(0, y))\,dy \le \frac{C}{\beta}|A_k|^{1/4}\biggl(\int_{A_k} |\nabla(e^{ay/2}(u_k-\beta))|^2\,dx\,dy\biggr)^{1/2}. \tag{2-5}$$

Using Cauchy–Schwartz and the trace inequality of Lemma 2.1, we see that the integral on the right-hand side of (2-5) is bounded by a constant independent of $k$. In addition, as a consequence of inequality (2-4) we have that

$$\lim_{R\to\infty} |A_k| = 0.$$

Thus, the result follows from (2-5). □

We can now show that the infimum is achieved.

**Corollary 2.6.** *Let $f$ be of positively balanced bistable type or of combustion type as in* Definition 1.1. *Then, for every $a > 0$ small enough (depending only on $f$), there exists $\underline{u} \in B_a$ such that*

$$E_a(\underline{u}) = \inf_{w\in B_a} E_a(w).$$

*In addition,* $0 \le \underline{u} \le 1$, $|\{\underline{u}(0, y) : y \in \mathbb{R}\} \setminus [0, \beta]| > 0$, *and $\underline{u}$ is not identically constant.*

*Proof.* For $a > 0$ small enough, Proposition 2.2 shows that $-\infty < I_a < 0$, where

$$I_a := \inf_{w\in B_a} E_a(w). \tag{2-6}$$

By Lemma 2.3, there exists $\{u_k\} \subset B_a \subset H_a^1(\mathbb{R}^2_+)$ such that $E_a(u_k) \to I_a$, $u_k \in C_c(\overline{\mathbb{R}^2_+})$, and $0 \le u_k \le 1$. Since $\{u_k\} \subset B_a$, by Lemma 2.1 $\{u_k\}$ is bounded in $H_a^1(\mathbb{R}^2_+)$. Therefore, there exists a weakly convergent subsequence (still denoted by $\{u_k\}$) such that $u_k \rightharpoonup \underline{u}$ and $\underline{u} \in H_a^1(\mathbb{R}^2_+)$.

By the WLSC comments made on the beginning of this subsection on the kinetic energy and the potential energy corresponding to $G^+$, and by the compactness result of Proposition 2.5, we have that $E_a(\underline{u}) \le \liminf_k E_a(u_k)$. Thus, we will have that $\underline{u}$ is a minimizer if we show that

$$\underline{u} \in B_a.$$

To show this claim, recall that $\Gamma_a(\underline{u}) \le \liminf_k \Gamma_a(u_k) = 1$. If $\Gamma_a(\underline{u}) = 0$ then $\underline{u} \equiv 0$; thus, in that case we would have $0 = E_a(\underline{u}) \le \liminf_k E_a(u_k) = I_a < 0$, a contradiction. Hence, $\Gamma_a(\underline{u}) \in (0, 1]$.

Let us see now that $\Gamma_a(\underline{u}) \in (0, 1)$ is not possible either. Indeed, assume that $\Gamma_a(\underline{u}) < 1$. Then, for some $t < 0$, the function $\underline{u}^t$ defined by $\underline{u}^t(x, y) := \underline{u}(x, y + t)$ satisfies $\Gamma_a(\underline{u}^t) = e^{-at}\Gamma_a(\underline{u}) = 1$, and hence $\underline{u}^t \in B_a$. In addition, $E_a(\underline{u}^t) = e^{-at}E_a(\underline{u}) < E_a(\underline{u}) = I_a$, which is a contradiction. Therefore, we have shown our claim $\underline{u} \in B_a$.

To prove the last statements of the corollary, since $0 \le u_k \le 1$ the same holds for $\underline{u}$. Moreover, since $\underline{u} \in B_a$, $\underline{u}$ is not identically constant. Finally, if we had $|\{\underline{u}(0, y) : y \in \mathbb{R}\} \setminus [0, \beta]| = 0$, then $E_a(\underline{u}) \ge 0$ by (1-19) and this is a contradiction. □

***Solving the PDE.*** In this part we show that there exists a solution pair $(c, u)$ to (1-2) with $c > 0$ and $u$ not identically constant. The solution is constructed from a minimizer $\underline{u}$ of our variational problem, after scaling its independent variables $(x, y)$ to take care of a Lagrange multiplier $\lambda_a$. The speed turns out to be $c = a(1 - 2I_a) = a(1 - 2\lambda_a)$; see (2-11).

**Proposition 2.7.** *Let $f$ be of positively balanced bistable type or of combustion type as in* Definition 1.1. *Let $\underline{u}$ be a minimizer for problem* (1-21) *as given by* Corollary 2.6. *Then, there exists $c > 0$ and $\mu > 0$ such that, defining*

$$u(x, y) = \underline{u}(\mu x, \mu y),$$

*we have that $(c, u)$ is a solution pair for problem* (1-2), *$u$ is not identically constant*, $0 \le u \le 1$, *and* $u \in H^1_c(\mathbb{R}^2_+)$.

*Proof.* Let $\underline{u} \in B_a$ be a minimizer as in Corollary 2.6. We have that $D\Gamma_a(\underline{u}) \not\equiv 0$ because

$$D\Gamma_a(\underline{u}) \cdot \underline{u} = \int_{\mathbb{R}^2_+} 2e^{ay} |\nabla \underline{u}|^2 \, dx \, dy = 2.$$

Therefore, there exists a Lagrange multiplier $\lambda_a \in \mathbb{R}$ such that $DE_a(\underline{u}) \cdot \phi = \lambda_a D\Gamma_a(\underline{u}) \cdot \phi$ for all $\phi \in H^1_a(\mathbb{R}^2_+)$, that is,

$$(1 - 2\lambda_a) \int_{\mathbb{R}^2_+} e^{ay} \nabla \underline{u} \nabla \phi \, dx \, dy - \int_{\partial \mathbb{R}^2_+} e^{ay} f(\underline{u}(0, y)) \phi(0, y) \, dy = 0. \tag{2-7}$$

Let us see that $\lambda_a \ne \frac{1}{2}$. Indeed, otherwise, from (2-7) we deduce $f(\underline{u}(0, \cdot)) \equiv 0$ in $\mathbb{R}$. Thus, by assumption (1-7) on $f$, we would have that either $\underline{u}(0, \cdot) \equiv 1$ or that $0 \le \underline{u}(0, \cdot) \le \beta$. The first of these is not possible since $\underline{u} \equiv 1 \notin H^1_a(\partial \mathbb{R}^2_+)$. On the other hand, $0 \le \underline{u}(0, \cdot) \le \beta$ is ruled out by the last statement of Corollary 2.6.

Let us consider arbitrary functions $\varphi \in C_c^\infty(\mathbb{R}^2_+)$ vanishing on $\partial\mathbb{R}^2_+$, and also functions $\psi \in C_c^\infty(\overline{\mathbb{R}^2_+})$. From (2-7) and $\lambda_a \neq \frac{1}{2}$, we have

$$\int_{\mathbb{R}^2_+} e^{ay}\{\Delta\underline{u} + a\underline{u}_y\}\varphi\,dx\,dy = 0 \quad \text{and} \quad \int_{\partial\mathbb{R}^2_+} e^{ay}\{(1-2\lambda_a)\underline{u}_x + f(\underline{u})\}\psi\,dy = 0.$$

As a consequence, and since $\lambda_a \neq \frac{1}{2}$, the pair $(a, \underline{u})$ is a solution of

$$\begin{cases} \Delta\underline{u} + a\underline{u}_y = 0 & \text{in } \mathbb{R}^2_+, \\ \dfrac{\partial\underline{u}}{\partial\nu} = \dfrac{1}{1-2\lambda_a} f(\underline{u}) & \text{on } \partial\mathbb{R}^2_+. \end{cases}$$

Let us now show that $\lambda_a < \frac{1}{2}$. Consider the test function $(\underline{u} - \beta)^+ \in H^1_a(\mathbb{R}^2_+)$. Plugging it into (2-7), we get

$$(1-2\lambda_a)\int_{\{\underline{u}>\beta\}} e^{ay}|\nabla\underline{u}|^2\,dx\,dy - \int_{\{\underline{u}(0,\cdot)>\beta\}} e^{ay} f(\underline{u}(0,y))(\underline{u}(0,y)-\beta)\,dy = 0. \tag{2-8}$$

Recall that $|\{\underline{u}(0,y) : y \in \mathbb{R}\} \setminus [0,\beta]| > 0$, and thus

$$\int_{\{\underline{u}>\beta\}} e^{ay}|\nabla\underline{u}|^2\,dx\,dy > 0. \tag{2-9}$$

Since $f(\underline{u}(0,y))(\underline{u}(0,y)-\beta) > 0$ in $\{\underline{u} > \beta\}$ by (1-7), (2-8) and (2-9) lead to $1 - 2\lambda_a > 0$.

Let $\mu := 1 - 2\lambda_a > 0$ and define

$$u(x,y) := \underline{u}(\mu x, \mu y) \quad \text{and} \quad c := a(1-2\lambda_a) > 0. \tag{2-10}$$

We then have a solution pair $(c, u)$ for (1-2).

Note that $u \in H^1_c(\mathbb{R}^2_+)$ since

$$\int_{\mathbb{R}^2_+} e^{cy}\{|\nabla u|^2 + u^2\}\,dx\,dy = \int_{\mathbb{R}^2_+} e^{a\bar{y}}\{|\nabla\underline{u}|^2 + \mu^{-2}\underline{u}^2\}\,d\bar{x}\,d\bar{y} < \infty$$

and $\underline{u} \in H^1_a(\mathbb{R}^2_+)$.

Finally, since $f \in C^{1,\gamma}$, the weak solution that we have found can be shown to be classical, indeed $C^{2,\gamma}$ in all $\overline{\mathbb{R}^2_+}$. This is explained in the beginning of Section 4. □

**Remark 2.8.** It is interesting to note the following relation, already noted in [Heinze 2001], between the infimum value $I_a$ of our problem (2-6) and the speed $c$ of the traveling front. The formula, which is not strictly needed anywhere else in this paper, provides however with an alternative proof of part (iv) of Theorem 1.2 on the comparison of the front speeds for different nonlinearities.

We claim that

$$c = a(1-2I_a) = a(1-2\lambda_a), \tag{2-11}$$

where $a$ and $\lambda_a$ are the parameter and the multiplier in the proof of Proposition 2.7. To show this formula, we take a minimizing sequence $\{u_k\}$ made of $C^\infty$ functions with compact support, and we test (2-7) with

$\phi = \partial_y u_k \in H_a^1(\mathbb{R}^2_+)$. Integrating by parts in order to pass to the limit as $k \to \infty$, we obtain

$$\begin{aligned} 0 &= (1-2\lambda_a)\int_{\mathbb{R}^2_+} e^{ay}\partial_y \frac{|\nabla \underline{u}|^2}{2}\,dx\,dy + \int_{\partial\mathbb{R}^2_+} e^{ay}\partial_y G(\underline{u}(0,y))\,dy \\ &= -a\frac{1-2\lambda_a}{2} - a\int_{\partial\mathbb{R}^2_+} e^{ay} G(\underline{u}(0,y))\,dy, \end{aligned}$$

where in the last equality we have also integrated by parts. We deduce that

$$I_a = E_a(\underline{u}) = \tfrac{1}{2}\Gamma_a(\underline{u}) - \tfrac{1}{2}(1-2\lambda_a) = \tfrac{1}{2} - \tfrac{1}{2}(1-2\lambda_a) = \lambda_a,$$

which together with (2-10) shows the claim.

## 3. Monotonicity

In this section we show that the front $u$ built in the previous section can be taken to be nonincreasing in the $y$ variable. This fact will be crucial to show in the next section that such a nonincreasing front $u$ has limits 1 and 0 as $y \to \mp\infty$.

Note that it suffices to show the existence of a nonincreasing minimizer, since the scaling used in the proof of Proposition 2.7 does not change the monotonicity of the front. As in [Heinze 2001], the existence of a nonincreasing minimizer will be a consequence of an inequality for monotone decreasing rearrangements in a new variable $z$, defined by

$$z = \frac{e^{ay}}{a}.$$

**Proposition 3.1.** *The minimizer $\underline{u}$ of* Corollary 2.6 *can be taken to be nonincreasing in the $y$ variable.*

*Proof.* We follow ideas in [Heinze 2001] and perform the change of variables $(x, z) := (x, e^{ay}/a)$, which takes $\mathbb{R}^2_+$ into $(\mathbb{R}_+)^2 = \{(x,z) : x > 0,\ z > 0\}$, and the functionals $\Gamma_a$, $E_a$, into $\widetilde{\Gamma}_a$, $\widetilde{E}_a$, where

$$\widetilde{\Gamma}_a(v) := \iint_{(\mathbb{R}_+)^2} \{|\partial_x v|^2 + a^2 z^2 |\partial_z v|^2\}\,dx\,dz \tag{3-1}$$

and

$$\widetilde{E}_a(v) := \tfrac{1}{2}\widetilde{\Gamma}_a(v) + \int_0^{+\infty} G(v(0,z))\,dz.$$

Let $\{u_k\}$ be the minimizing sequence for problem (1-21) given by Lemma 2.3. The functions $\{u_k\}$ take values in $[0,1]$, are continuous, and have compact support in $\overline{\mathbb{R}^2_+}$. Let $v_k$ be defined by $v_k(x,z) := u_k(x,y)$. Since $v_k$ is nonnegative, continuous, and with compact support in $[0,+\infty)^2$, we may consider its monotone decreasing rearrangement in the $z$ variable, that we denote by $v_k^*$; see [Kawohl 1985]. That is, for each $x \geq 0$, we make the usual one-dimensional monotone decreasing rearrangement of the function $v_k(x,\cdot)$ of $z > 0$. Recall also that if we consider the even extension of $v_k$ across $\{z = 0\}$, then $v_k^*$ coincides with the Steiner symmetrization of $v_k$ with respect to $\{z = 0\}$.

As a consequence of equimeasurability, we have

$$\int_0^{+\infty} G(v_k^*(0,z))\,dz = \int_0^{+\infty} G(v_k(0,z))\,dz.$$

On the other hand, the inequality

$$\widetilde{\Gamma}_a(v_k^*) \le \widetilde{\Gamma}_a(v_k)$$

— and thus $\widetilde{E}_a(v_k^*) \le \widetilde{E}_a(v_k)$ — follows from a result of Landes [2007] for monotone decreasing rearrangements since the weight $w(x,z) = a^2 z^2$ in (3-1) is nonnegative and nondecreasing in $z \in (0,+\infty)$. It also follows from a previous result of Brock [1999] on Steiner symmetrization, which requires $w$ to be nonnegative and $w^{1/2}(x,z) = a|z|$ to be even and convex. These results require that the weight in front of $|\partial_x v|^2$ (which in our case is identically one) does not depend on $z$.

Finally, we pull back the sequence $v_k^*$ to the $(x, y)$ variables and name these functions $u_k^*$. We have that

$$\Gamma_a(u_k^*) = \widetilde{\Gamma}_a(v_k^*) \le \widetilde{\Gamma}_a(v_k) = 1$$

and

$$E_a(u_k^*) = \widetilde{E}_a(v_k^*) \le \widetilde{E}_a(v_k) = E_a(u_k).$$

Let $\underline{u}^*$ be a weak limit in $H_a^1(\mathbb{R}^2_+)$ of a subsequence of $\{u_k^*\}$. By the WLSC results of the previous section, it is easy to prove that we necessarily have $\underline{u}^* \in B_a$. This is done exactly as in the proof of Corollary 2.6. Thus, $\underline{u}^*$ is a minimizer which is nonincreasing in the $y$ variable. Note also that it still takes values in $[0, 1]$. □

## 4. Limits at infinity

In this section we prove that the front $u$ for (1-2) constructed in the previous sections satisfies

$$\lim_{y\to-\infty} u(x,y) = 1 \quad \text{and} \quad \lim_{y\to+\infty} u(x,y) = 0 \quad \text{for all } x \ge 0, \tag{4-1}$$

and

$$\lim_{x\to+\infty} u(x,y) = 0 \quad \text{for all } y \in \mathbb{R}. \tag{4-2}$$

To establish (4-1), it will be crucial to use that $u$ is nonincreasing in the $y$ variable.

In what follows, we will be using the following regularity fact. Assume that $u$ is a bounded $C^2$ function in $\mathbb{R}^2_+$, $C^1$ up to the boundary $\partial\mathbb{R}^2_+$, that satisfies our nonlinear problem

$$\begin{cases} \Delta u + c u_y = 0 & \text{in } \mathbb{R}^2_+, \\ \dfrac{\partial u}{\partial \nu} = f(u) & \text{on } \partial\mathbb{R}^2_+. \end{cases}$$

Since $f$ is $C^{1,\gamma}$ for some $\gamma \in (0,1)$, we have that, for every $R > 0$, $u \in C^{2,\gamma}(\overline{B_R^+})$ and

$$\|u\|_{C^{2,\gamma}(\overline{B_R^+})} \le C_R \tag{4-3}$$

for some constant $C_R$ depending only on $c$, $\gamma$, $R$, and on upper bounds for $\|u\|_{L^\infty(B^+_{4R})}$ and $\|f\|_{C^{1,\gamma}}$. Here $B^+_R = \{(x, y) \in \mathbb{R}^2 : x > 0,\ |(x, y)| < R\}$. This estimate is established by easily adapting the proof of [Cabré and Solà-Morales 2005, Lemma 2.3(a)]. As a consequence of the estimate, we also deduce that

$$|\nabla u| \in L^\infty(\mathbb{R}^2_+). \tag{4-4}$$

To establish (4-1), we first need the following easy result on limits as $|y| \to \infty$. It applies to any solution, not only to the variational one constructed in previous sections.

**Lemma 4.1.** *Assume that* $f(0) = f(1) = 0$ *and that* $0 \le u \le 1$ *is a solution of* (1-2) *satisfying*

$$\lim_{y\to-\infty} u(0, y) = 1 \quad \textit{and} \quad \lim_{y\to+\infty} u(0, y) = 0.$$

*Then, for all* $R > 0$, *we have*

$$\lim_{y\to-\infty} u(x, y) = 1, \quad \lim_{y\to+\infty} u(x, y) = 0, \quad \textit{and} \quad \lim_{|y|\to\infty} |\nabla u(x, y)| = 0 \tag{4-5}$$

*uniformly for* $x \in [0, R]$.

*Proof.* For $t \in \mathbb{R}$ let us define $u^t(x, y) := u(x, y + t)$, also a solution of (1-2). We claim that

$$\|u^t - 1\|_{L^\infty(B^+_R)} + \|\nabla u^t\|_{L^\infty(B^+_R)} \to 0 \quad \text{as } t \to -\infty.$$

Assume, to the contrary, that there exist $\varepsilon > 0$ and $\{t_k\} \subset \mathbb{R}$ with $t_k \to -\infty$ such that

$$\|u^{t_k} - 1\|_{L^\infty(B^+_R)} + \|\nabla u^{t_k}\|_{L^\infty(B^+_R)} \ge \varepsilon. \tag{4-6}$$

The estimates (4-3) lead to the existence of a subsequence $\{t_{k_j}\}$ for which $u^{t_{k_j}}$ converges in $C^2(\overline{B^+_R})$ to $u^\infty$. By the hypothesis of the lemma, we will have $0 \le u^\infty \le 1$ and

$$\begin{cases} \Delta u^\infty + c u^\infty_y = 0 & \text{in } \mathbb{R}^2_+, \\ u^\infty = 1 & \text{on } \partial\mathbb{R}^2_+, \\ \dfrac{\partial u^\infty}{\partial \nu} = f(u^\infty) = f(1) = 0 & \text{on } \partial\mathbb{R}^2_+. \end{cases}$$

From this and Hopf's boundary lemma, we deduce $u^\infty \equiv 1$ on $\mathbb{R}^2_+$, which contradicts (4-6).

In an analogous way we can show the limits as $y \to +\infty$. □

We can now prove the existence of limits as $y \to \pm\infty$ for the variational solution constructed in the last sections.

**Lemma 4.2.** *Let* $f$ *be of positively balanced bistable type or of combustion type as in* Definition 1.1. *Let* $u$ *be any front constructed as in* Proposition 2.7 *from the nonincreasing minimizer of* Proposition 3.1. *Then,*

$$\lim_{y\to-\infty} u(x, y) = 1 \quad \textit{and} \quad \lim_{y\to+\infty} u(x, y) = 0 \quad \textit{for all } x \ge 0.$$

*Proof.* By Corollary 2.6 we know that $0 \le u \le 1$, $u \in H^1_c(\mathbb{R}^2_+)$, and that the set $\{u(0, y) : y \in \mathbb{R}\}$ is not contained in $[0, \beta]$. We also know that $u_y(x, y) \le 0$. Therefore, for all $x \ge 0$, there exist $L^-(x) \in (\beta, 1]$ and $L^+(x) \in [0, 1]$ such that

$$\lim_{y\to-\infty} u(x, y) = L^-(x) \quad \text{and} \quad \lim_{y\to+\infty} u(x, y) = L^+(x).$$

Note that $L^+(x) \equiv 0$ is a consequence of the inequalities of Lemma 2.1.

To prove that $L^-(x) \equiv 1$, by Lemma 4.1 it is enough to show that $L^-(0) = 1$. To do this, we consider the sequence of solutions $\{u^k\}$ defined by $u^k(x, y) := u(x, y + k)$. Then, as in the proof of Lemma 4.1, as $k \to -\infty$ there exists a convergent subsequence to a solution $u^\infty$ of

$$\begin{cases} \Delta u^\infty + cu^\infty_y = 0 & \text{in } \mathbb{R}^2_+, \\ u^\infty = L^-(0) & \text{on } \partial\mathbb{R}^2_+, \\ \dfrac{\partial u^\infty}{\partial \nu} = f(L^-(0)) & \text{on } \partial\mathbb{R}^2_+. \end{cases} \tag{4-7}$$

Since $u^\infty(x, y) = L^-(x)$, we have that $\partial_y u^\infty(x, y) = \partial_{yy} u^\infty(x, y) \equiv 0$. Therefore, the first equation in (4-7) leads to $\partial_{xx} u^\infty(x, y) = 0$ for all $x > 0$ and $y \in \mathbb{R}$. Since $u^\infty$ is bounded, then it must be constant and equal to $L^-(0)$, its value at $x = 0$.

This and the last equation in (4-7) lead to $0 = -\partial_x u^\infty(0, y) = f(L^-(0))$. Since $L^-(0) \in (\beta, 1]$ and, on this interval, $f$ vanishes only at 1 by hypothesis (1-7), we conclude that $L^-(0) = 1$. □

It remains to prove (4-2) on the limits as $x \to +\infty$. This is a simple consequence of the Harnack inequality and the fact that the variational solution lies in $H^1_c(\mathbb{R}^2_+)$:

**Lemma 4.3.** *Let $f$ be of positively balanced bistable type or of combustion type as in* Definition 1.1. *Let $u$ be any front constructed as in* Proposition 2.7. *Then*,

$$\lim_{x\to+\infty} u(x, y) = 0 \quad \textit{for all } y \in \mathbb{R}.$$

*Proof.* Take any $y_0 \in \mathbb{R}$. Since $u \in H^1_c(\mathbb{R}^2_+)$ we have that

$$\lim_{x_0\to+\infty} \int_{y_0-1}^{y_0+1} dy \int_{x_0-1}^{+\infty} e^{cy} u^2 \, dx = 0. \tag{4-8}$$

Recall that $0 \le u \le 1$ satisfies $\Delta u + cu_y = 0$ in $\mathbb{R}^2_+$. Thus, by the Harnack inequality, for any $x_0 > 2$ we have

$$\sup_{B_1(x_0,y_0)} u \le C \inf_{B_1(x_0,y_0)} u \le C \int_{B_1(x_0,y_0)} u \, dx \, dy \le C \left( \int_{B_1(x_0,y_0)} u^2 \, dx \, dy \right)^{1/2}$$

for different constants $C$ independent of $x_0$. Using (4-8), it follows that

$$\lim_{x_0\to+\infty} u(x_0, y_0) = 0,$$

as claimed. □

## 5. Uniqueness of speed and of solution with limits

In the first part of this section we establish a useful comparison principle, Proposition 5.2, in the spirit of one in [Cabré and Solà-Morales 2005]. It will lead first to the asymptotic bounds on fronts stated in our main theorem (after building appropriate comparison barriers in next section). Then, it will be used in the second part of this section to establish a key result, Proposition 5.3 below.

Proposition 5.3 will have several important applications: first, the monotonicity in $y$ of every solution with limits; second, the uniqueness of a speed and of a front with limits; and third, the comparison result between speeds corresponding to different ordered nonlinearities. The proof of the proposition follows the powerful sliding method of Berestycki and Nirenberg [1991].

***A maximum principle.*** We start with the following easy lemma:

**Lemma 5.1.** *Let $w$ be a $C^2$ function in $\mathbb{R}^2_+$, bounded below, continuous up to $\partial\mathbb{R}^2_+$, and satisfying*

$$\Delta w + c w_y \le 0 \quad \textit{in } \mathbb{R}^2_+$$

*for some constant $c \in \mathbb{R}$. Assume also that $w \ge 0$ on $\partial\mathbb{R}^2_+$ and that, for every $R > 0$,*

$$\liminf_{|y|\to+\infty} w(x, y) \ge 0 \quad \textit{uniformly in } x \in [0, R]. \tag{5-1}$$

*Then, $w \ge 0$ in $\mathbb{R}^2_+$.*

*Proof.* Consider the new function

$$\overline{w} = \frac{w}{x+1} \quad \text{for } x \ge 0,\ y \in \mathbb{R}.$$

It satisfies

$$\Delta\overline{w} + \frac{2}{x+1}\overline{w}_x + c\overline{w}_y \le 0 \quad \text{in } \mathbb{R}^2_+.$$

Let $\varepsilon > 0$. Since $w$ is bounded below, if $R$ is sufficiently large we have that

$$\overline{w}(x, y) \ge -\varepsilon \tag{5-2}$$

for $x = R$. By assumption (5-1), we also have (5-2) for $x \in [0, R]$ and $|y| = S$, if $S$ is large enough (depending on $R$). Since, (5-2) also holds for $x = 0$, the maximum principle applied in $(0, R) \times (-S, S)$ gives that $\overline{w} \ge -\varepsilon$ in $(0, R) \times (-S, S)$.

Letting $S \to \infty$ we deduce that $\overline{w} \ge -\varepsilon$ in $(0, R) \times \mathbb{R}$. Now, letting $R \to \infty$ we conclude that $\overline{w} \ge -\varepsilon$ in $\mathbb{R}^2_+$ for any $\varepsilon > 0$. Thus $\overline{w} \ge 0$ in $\mathbb{R}^2_+$ and this finishes the proof. □

The following maximum principle (in the spirit of one in [Cabré and Solà-Morales 2005]) is a key ingredient in the remaining of this section. It will be applied to the difference of two solutions (and also of a supersolution and a solution) of our nonlinear problem.

**Proposition 5.2.** *Let $c \in \mathbb{R}$ and $v$ be a $C^2$ bounded function in $\overline{\mathbb{R}^2_+}$ satisfying*

$$\Delta v + c v_y \le 0 \quad \textit{in } \mathbb{R}^2_+$$

*and that, for all* $R > 0$,

$$\lim_{|y|\to\infty} v(x, y) = 0 \quad \textit{uniformly in } x \in [0, R]. \tag{5-3}$$

*Finally, assume that there exists a nonempty set* $H \subset \mathbb{R}$ *such that* $v(0, y) > 0$ *for* $y \in H$,

$$\frac{\partial v}{\partial \nu} + d(y)v \geq 0 \quad \textit{if } y \notin H \tag{5-4}$$

*and*

$$d(y) \geq 0 \quad \textit{if } y \notin H \tag{5-5}$$

*for some continuous function* $d$ *defined on* $\mathbb{R} \setminus H$.

*Then*, $v > 0$ *in* $\overline{\mathbb{R}^2_+}$.

*Proof.* We need to prove that $v \geq 0$ in $\overline{\mathbb{R}^2_+}$. It then follows that $v > 0$ in $\overline{\mathbb{R}^2_+}$: Indeed, since $H$ is nonempty, $v$ cannot be identically zero. If we assume that $v = 0$ at some point $(x_1, y_1) \in \overline{\mathbb{R}^2_+}$, we obtain a contradiction using the strong maximum principle (if $x_1 > 0$) and using the Hopf's boundary lemma and (5-4) (if $x_1 = 0$, since then $y_1 \notin H$ because $v(0, y_1) = 0$).

Let

$$A := \inf_{\partial\mathbb{R}^2_+} v.$$

By (5-3) used with $x = 0$, we have $A \leq 0$. Thus, we can apply Lemma 5.1 to $w := v - A$. We deduce that $v \geq A$ in all of $\mathbb{R}^2_+$.

It only remains to prove that $A \geq 0$. By contradiction, assume that $A < 0$. Then, by its definition and since $v(0, y) \to 0$ as $|y| \to \infty$, we have that the infimum $A$ of $v$ on $\partial\mathbb{R}^2_+$ is achieved at some point $(0, y_0)$. Since we have proved that $v \geq A$ in all of $\mathbb{R}^2_+$, then $(0, y_0)$ is also a minimum of $v$ in all $\overline{\mathbb{R}^2_+}$. Since $v(0, y_0) = A < 0$, $v$ is not identically constant, and thus the Hopf's boundary lemma gives that $-v_x(0, y_0) < 0$. This is a contradiction with (5-4) and (5-5) — since $y_0 \notin H$ because $v(0, y_0) < 0$. □

***Uniqueness.*** The goal of this section is to establish uniqueness of the traveling speed, as well as uniqueness — up to vertical translations — of solutions to (1-2) which have limits 1 and 0 as $y \to \mp\infty$ on $\partial\mathbb{R}^2_+$.

We also prove in this section that every solution with the above limits is necessarily decreasing in $y$.

All these three results will follow from the following proposition — an analogue of Lemma 5.2 in [Cabré and Solà-Morales 2005].

**Proposition 5.3.** *Assume that* $f$ *satisfies* (1-4) *and* (1-5), *and let* $c \in \mathbb{R}$. *Let* $u_1$ *and* $u_2$ *be, respectively, a supersolution and a solution of* (1-2) *such that*

$$0 \leq u_i \leq 1 \quad \textit{and} \quad u_i(0, 0) = \tfrac{1}{2}$$

*for* $i = 1, 2$. *Assume that, for* $i = 1, 2$ *and all* $R > 0$,

$$\lim_{y\to-\infty} u_i(x, y) = 1 \quad \textit{and} \quad \lim_{y\to+\infty} u_i(x, y) = 0 \quad \textit{uniformly in } x \in [0, R]. \tag{5-6}$$

*For* $t > 0$, *consider*

$$u_2^t(x, y) := u_2(x, y + t).$$

*Then,*

$$u_2^t \le u_1 \quad \textit{in } \overline{\mathbb{R}^2_+} \qquad \textit{for every } t > 0. \tag{5-7}$$

*As a consequence,* $u_2 \equiv u_1$ *in* $\overline{\mathbb{R}^2_+}$.

*In addition, for any solution* $u_2$ *satisfying* $0 \le u_2 \le 1$ *and* (5-6), *we have* $\partial_y u_2 < 0$ *in* $\overline{\mathbb{R}^2_+}$.

If we apply the proposition to $u_1 = u_2 = u$, where $u$ is a solution to (1-2) taking values in [0, 1] and with limits 1 and 0, the conclusion (5-7) establishes that $u$ is nonincreasing in $y$. From this, the strong maximum principle and Hopf's boundary lemma applied to the linearized problem satisfied by $u_y$, we deduce that $u_y < 0$, as claimed in the last statement of the proposition.

Second, by letting $t \to 0^+$ in (5-7) we deduce that $u_2 \le u_1$. But $u_2$ is a solution and $u_1$ a supersolution, with $u_2(0,0) = u_1(0,0)$. Again the strong maximum principle and the Hopf's boundary lemma give that $u_2 \equiv u_1$, as stated in the proposition.

The proposition also gives the uniqueness (up to vertical translations) of a solution with limits for a given speed $c$. For this, apply the proposition to two solutions after translating them in the $y$ direction.

In the proof of our main theorem in Section 7, we will give two other important applications of the proposition. First, the uniqueness of a speed admitting a solution with limits. This will follow from the fact that any front $u_1$ with limits 1 and 0 is necessarily decreasing, and hence the terms $c_1 \partial_y u_1$ and $c_2 \partial_y u_1$ will be comparable for two different speeds. Since one of the functions in the proposition may be taken to be only a supersolution, this will lead to the uniqueness of speed. A similar argument will show the comparison of speeds corresponding to two different ordered nonlinearities.

*Proof of Proposition 5.3.* As explained above, we only need to prove (5-7). The subsequent statements follow easily from this.

Note first that $u_i$ are not identically constant, by the assumption in (5-6) about their limits as $y \to \pm\infty$. Therefore, since $0 \le u_i \le 1$ and $f(0) = f(1) = 0$, the strong maximum principle leads to $0 < u_i < 1$ for $i = 1, 2$.

Let $\delta > 0$ be the constant in assumption (1-5) for $f$. By hypothesis (5-6), there exists a compact interval $[a, b]$ in $\mathbb{R}$ such that, for $i = 1, 2$,

$$\begin{aligned} &u_i(0, y) \in (1-\delta, 1) && \text{if } y \le a, \quad \text{and} \\ &u_i(0, y) \in (0, \delta) && \text{if } y \ge b. \end{aligned}$$

Note that $u_2^t$ is also a solution of (1-2), and hence

$$\begin{cases} \Delta(u_1 - u_2^t) + c(u_1 - u_2^t)_y \le 0 & \text{in } \mathbb{R}^2_+, \\ -(u_1 - u_2^t)_x \ge -d^t(y)(u_1 - u_2^t) & \text{on } \partial\mathbb{R}^2_+, \end{cases}$$

where

$$d^t(y) = -\frac{f(u_1) - f(u_2^t)}{u_1 - u_2^t}(0, y)$$

if $(u_1 - u_2^t)(0, y) \ne 0$, and $d^t(y) = -f'(u_1(0, y)) = -f'(u_2^t(0, y))$ otherwise. Note that $d^t$ is a continuous function since $f$ is $C^1$.

We also have, for all $R > 0$,

$$\lim_{|y|\to\infty} (u_1 - u_2^t)(x, y) = 0 \quad \text{uniformly in } x \in [0, R].$$

We finish the proof in three steps.

*Step 1.* We claim that $u_2^t < u_1$ in $\overline{\mathbb{R}^2_+}$ for $t > 0$ large enough.

To show this, we take $t > 0$ sufficiently large so that $u_2^t(0, y) < u_1(0, y)$ for $y \in [a, b]$. This is possible since $u_2(0, y+t) \to 0$ as $t \to +\infty$ and $u_1 > 0$. We apply Proposition 5.2 to $v = u_1 - u_2^t$ with

$$H = (a, b) \cup \{y \in \mathbb{R} : (u_1 - u_2^t)(0, y) > 0\}.$$

Clearly, $v(0, y) > 0$ in $H$.

To show that $d^t \ge 0$ in $\mathbb{R} \setminus H$, let $y \notin H$. There are two possibilities. First, if $y \ge b$ then $y + t \ge b$ also. Therefore, $u_1(0, y) \le \delta$ and $u_2^t(0, y) = u_2(0, y+t) \le \delta$. We conclude that $d^t(y) \ge 0$, since $f' \le 0$ in $(0, \delta)$ by (1-5).

The other possibility is that $y \le a$. In this case, we have $u_1(0, y) \ge 1 - \delta$, and since $y \notin H$ then $(u_1 - u_2^t)(0, y) \le 0$. Therefore $u_2^t(0, y) \ge u_1(0, y) \ge 1 - \delta$, and we conclude $d^t(y) \ge 0$, again by (1-5).

Proposition 5.2 gives that $u_1 - u_2^t > 0$ in $\overline{\mathbb{R}^2_+}$.

**Claim.** *If $u_2^t \le u_1$ for some $t > 0$, then $u_2^{t+\mu} \le u_1$ for every $\mu$ small enough* (*with $\mu$ either positive or negative*).

This statement will finish the proof of the proposition, since then $\{t > 0 : u_2^t \le u_1\}$ is a nonempty, closed and open set in $(0, \infty)$, and hence equal to this interval. We conclude $u_2^t \le u_1$ for all $t > 0$.

*Step 2.* To prove the claim, we will show that

$$\text{if } t > 0 \text{ and } u_2^t \le u_1, \text{ then } u_2^t \not\equiv u_1. \tag{5-8}$$

Once (5-8) is known, we can finish the proof of the claim as follows: First, by the strong maximum principle and Hopf's boundary lemma, $u_2^t < u_1$ in $\overline{\mathbb{R}^2_+}$. Let $K_t$ be a compact interval such that, on $\mathbb{R} \setminus K_t$, both $u_1$ and $u_2^t$ take values in $(0, \delta/2) \cup (1 - \delta/2, 1)$. Recall that $(u_1 - u_2^t)(0, \cdot) > 0$ in the compact set $K_t$. By continuity and the existence of limits at infinity, we have that, if $|\mu|$ is small enough, then $(u_1 - u_2^{t+\mu})(0, y) > 0$ for $y \in K_t$ and $u_2^{t+\mu}(0, y)$ takes values in $(0, \delta) \cup (1 - \delta, 1)$ for $y \notin K_t$. Hence, we can apply Proposition 5.2 to $v = u_1 - u_2^{t+\mu}$ with $H = K_t$, since $d^{t+\mu} \ge 0$ outside $K_t$. We conclude that $u_1 - u_2^{t+\mu} > 0$ in $\overline{\mathbb{R}^2_+}$.

*Step 3.* Here we establish (5-8), therefore completing the proof of the claim and of the proposition. We assume that $t > 0$ and $u_2^t \le u_1$, and we need to show that $u_2^t \not\equiv u_1$.

To prove this, consider first the case when both functions in the proposition are the same, that is, $u_1 \equiv u_2$. Assume that $t > 0$ and $u_2^t \equiv u_1 \equiv u_2$. Then, the function $u_2(0, y)$ is $t$-periodic. But this is a contradiction with the hypothesis (5-6) on limits. Therefore, in the case $u_1 \equiv u_2$, the two steps above can be carried out. We conclude that, for every solution $u_2$ as in the lemma, we have $u_2^t \le u_2$ for every $t > 0$. In particular, $\partial_y u_2 \le 0$ and, by the strong maximum principle and Hopf's boundary lemma, $\partial_y u_2 < 0$.

Finally, consider the general case of a supersolution $u_1$ and a solution $u_2$. Assume that $t > 0$ and $u_2^t \equiv u_1$. Then $\frac{1}{2} = u_1(0,0) = u_2^t(0,0) = u_2(0,t)$. Moreover, $u_2(0,0) = \frac{1}{2}$ by hypothesis. Hence, $u_2(0,0) = u_2(0,t)$. This is a contradiction, since in the previous paragraph we have established that $u_2$ is decreasing in $y$. □

## 6. Explicit traveling fronts

In this section we construct an explicit supersolution of the linearized problem for (1-2) in the case of positively balanced bistable nonlinearities satisfying

$$f'(0) < 0 \quad \text{and} \quad f'(1) < 0. \tag{6-1}$$

It will lead to our result on the asymptotic behavior of traveling fronts. In addition, we construct a family of explicit traveling fronts corresponding to some positively balanced bistable nonlinearities satisfying (6-1).

To simplify the notation in this section, by rescaling the independent variables we may assume that the speed of the front is

$$c = 2.$$

Recall that our nonlinear problem, when written for the trace $v = v(y)$ of functions on $x = 0$, becomes (1-14) with $c = 2$, i.e.,

$$(-\partial_{yy} - 2\partial_y)^{1/2} v = f(v) \quad \text{in } \mathbb{R}.$$

As in [Cabré and Sire 2015], the construction of explicit fronts will be based on the fundamental solution for the homogeneous heat equation associated to the previous fractional operator in $\mathbb{R}$, that is, equation

$$\partial_t v + (-\partial_{yy} - 2\partial_y)^{1/2} v = 0 \tag{6-2}$$

for functions $v = v(y,t)$. Taking one more derivative $\partial_t$ in (6-2), we see that the solution of this problem at time $t$ (given an initial condition $v_0$) coincides with the value of $w(x = t, \cdot)$ for the solution of

$$\begin{cases} L_2 w := \Delta w + 2w_y = 0 & \text{in } \mathbb{R}^2_+, \\ w(0,\cdot) = w_0 & \text{on } \mathbb{R}, \end{cases} \tag{6-3}$$

where the operator $L_2$ acts on functions $w = w(x,y)$. Thus, the heat kernel for (6-2) coincides with the Poisson kernel for (6-3).

To compute such a Poisson kernel, as in [Caffarelli et al. 2012] we start with the observation that, if

$$w = e^{-y}\phi,$$

then

$$L_2 w = \Delta w + 2w_y = 0 \quad \text{if and only if} \quad -\Delta\phi + \phi = 0. \tag{6-4}$$

The fundamental solution of Helmholtz's equation, the solution of

$$-\Delta\Phi + \Phi = \delta_0,$$

is given by

$$\Phi(r) = \frac{1}{2\pi} K_0(r),$$

where $r=\sqrt{x^2+y^2}$ and $K_0$ is the modified Bessel function of the second kind with index $\nu=0$ (see [Abramowitz and Stegun 1964]). The function $K_0=K_0(s)$ is a positive and decreasing function of $s>0$, whose asymptotic behavior at $s=0$ is given by

$$K_0(s)=-\log s+\mathrm{o}(|\log s|) \quad \text{as } s\to 0.$$

For $s\to+\infty$, all modified Bessel functions of the second kind $K_\nu$ have the same behavior

$$K_\nu(s)=\sqrt{\frac{\pi}{2}}\,s^{-1/2}e^{-s}+\mathrm{o}(s^{-1/2}e^{-s}) \quad \text{as } s\to+\infty. \tag{6-5}$$

By considering the fundamental solution $\Phi$ but now with pole at a point $(x_0, y_0)\in\mathbb{R}^2_+$, subtracting from it $\Phi$ with pole at the reflected point $(-x_0, y_0)$, and applying the divergence theorem, one sees that the Poisson kernel for the Helmholtz's equation $-\Delta\phi+\phi=0$ in the half-plane $\mathbb{R}^2_+$ is given by

$$-2\Phi_x=-\frac{1}{\pi}\,\frac{x}{r}\,K_0'(r).$$

Writing this convolution formula for $w=e^{-y}\phi$, we deduce that the Poisson kernel for (6-3) is given by

$$-2e^{-y}\Phi_x=-e^{-y}\frac{x}{\pi r}\,K_0'(r).$$

To avoid its singularity at the origin, given any constant $t>0$ we consider the Poisson kernel after "time" $x=t$ and define

$$\begin{aligned} P^t(x, y) &:= -2G^t_x=-e^{-y}\frac{x+t}{\pi\sqrt{(x+t)^2+y^2}}\,K_0'\big(\sqrt{(x+t)^2+y^2}\big) \\ &= e^{-y}\frac{x+t}{\pi\sqrt{(x+t)^2+y^2}}\,K_1\big(\sqrt{(x+t)^2+y^2}\big), \end{aligned} \tag{6-6}$$

where $K_1=-K_0'$ is the modified Bessel function of the second kind with index $\nu=1$, and where

$$G^t(x, y):=\frac{1}{2\pi}e^{-y}\,K_0\big(\sqrt{(x+t)^2+y^2}\big).$$

By (6-4), $G^t$ is a solution of the homogeneous equation $L_2w=0$ in $\mathbb{R}^2_+$. Thus, so is $P^t=-2G^t_x$.

Finally, the explicit traveling front will be given by

$$u^t(x, y):=\int_y^{+\infty}P^t(x, z)\,dz=\int_y^{+\infty}e^{-z}\frac{x+t}{\pi\sqrt{(x+t)^2+z^2}}\,K_1\big(\sqrt{(x+t)^2+z^2}\big)\,dz.$$

Next, let us check all the properties of $P^t$ for $t>0$ that will be needed in order to use it as a supersolution of the linearized problem for (1-2). We know that $L_2P^t=0$ in $\mathbb{R}^2_+$. Using (6-5) we see that $P_t$ is bounded in all $\mathbb{R}^2_+$. We also have that $P^t>0$ in $\overline{\mathbb{R}^2_+}$ since $K_0$ is radially decreasing. Next, we have that, for every $R>0$, $P^t(x, y)\to 0$ as $|y|\to\infty$ uniformly in $x\in[0, R]$. This follows from the last equality in (6-6) and from (6-5). Also from the last equality in (6-6), we see that

$$\frac{-\partial_x P^t}{P^t}(0, y)=\frac{-1}{t}\left\{1-\frac{t^2}{t^2+y^2}+\frac{t^2}{\sqrt{t^2+y^2}}\,\frac{K_1'(\sqrt{t^2+y^2})}{K_1(\sqrt{t^2+y^2})}\right\}. \tag{6-7}$$

Now, using that $K_1' = -\frac{1}{2}(K_0 + K_2)$ and the asymptotic behavior (6-5), we deduce that

$$\lim_{|y|\to\infty} \frac{-\partial_x P^t}{P^t}(0, y) = \frac{-1}{t}. \tag{6-8}$$

Finally, since

$$P^t(0, y) = e^{-y} \frac{t}{\pi\sqrt{t^2+y^2}} K_1\bigl(\sqrt{t^2+y^2}\bigr)$$

and $K_1$ has the asymptotic behavior (6-5), we deduce

$$P^t(0, y) = \frac{t}{\sqrt{2\pi}} y^{-3/2} e^{-2y} + \mathrm{o}(y^{-3/2} e^{-2y}) \quad \text{as } y \to +\infty, \text{ and} \tag{6-9}$$

$$P^t(0, y) = \frac{t}{\sqrt{2\pi}} (-y)^{-3/2} + \mathrm{o}((-y)^{-3/2}) \quad \text{as } y \to -\infty. \tag{6-10}$$

We can now establish our result on explicit traveling fronts. We need to verify that each $u^t$ is a traveling front for some nonlinearity $f^t$ of positively balanced bistable type satisfying $(f^t)'(0) < 0$ and $(f^t)'(1) < 0$.

*Proof of Theorem 1.3.* The statements for $u^{t,c}$ follow from the corresponding ones for $u^{t,2} = u^t$. To prove them for $u^t$, note first that the solution of (6-3) when $w_0 \equiv 1$ is $w \equiv 1$. We deduce that its Poisson kernel satisfies

$$\int_{-\infty}^{+\infty} P^t(x, z)\, dz = 1$$

for all $x > 0$. It follows that $0 < u^t < 1$ and that $\lim_{y\to-\infty} u^t(x, y) = 1$ for all $x \geq 0$. Clearly we also have $\lim_{y\to+\infty} u^t(x, y) = 0$. In addition, $\partial_y u^t = -P^t < 0$ in $\overline{\mathbb{R}^2_+}$.

Next, let us see that we have $L_2 u^t = 0$ in $\mathbb{R}^2_+$. Indeed, $\partial_y L_2 u^t = L_2 u^t_y = -L_2 P^t = 0$. Thus, $L_2 u^t$ is a function of $x$ alone. But

$$\begin{aligned} L_2 u^t = &\int_y^{+\infty} \partial_{xx} \left\{ e^{-z} \frac{x+t}{\pi\sqrt{(x+t)^2+z^2}} K_1\bigl(\sqrt{(x+t)^2+z^2}\bigr) \right\} dz \\ &- \partial_y \left\{ e^{-y} \frac{x+t}{\pi\sqrt{(x+t)^2+y^2}} K_1\bigl(\sqrt{(x+t)^2+y^2}\bigr) \right\} - 2e^{-y} \frac{x+t}{\pi\sqrt{(x+t)^2+y^2}} K_1\bigl(\sqrt{(x+t)^2+y^2}\bigr). \end{aligned}$$

Since $L_2 u^t$ does not depend on $y$, we may let $y \to +\infty$ in this expression. From this, and since $K_j' = -\frac{1}{2}(K_{j-1} + K_{j+1})$ for all $j$ and all functions $K_\nu(s)$ have the asymptotic behavior (6-5), we deduce that $L_2 u^t \equiv 0$ in $\mathbb{R}^2_+$.

The asymptotic behaviors for $-u^t_y = P^t$ in the statement of the theorem follow from (6-9) and (6-10).

Next, we find the expression for the nonlinearity $f^t$. Since $u^t(0, \cdot)$ is decreasing from 1 to 0, $f^t$ is implicitly well defined in $[0, 1]$ by

$$\begin{aligned} f^t(u^t(0, y)) := -u^t_x(0, y) &= \int_y^{+\infty} 2G^t_{xx}(0, z)\, dz = \int_y^{+\infty} 2(-G^t_{yy} - 2G^t_y)(0, z)\, dz \\ &= 2(G^t_y + 2G^t)(0, y) \\ &= \frac{1}{\pi} e^{-y} \left\{ K_0\bigl(\sqrt{t^2+y^2}\bigr) - \frac{y}{\sqrt{t^2+y^2}} K_1\bigl(\sqrt{t^2+y^2}\bigr) \right\}. \end{aligned} \tag{6-11}$$

From this, we clearly see that $f^t(0) = f^t(1) = 0$, again by (6-5).

The remaining properties of $f^t$ will be deduced from the following implicit formula for its derivative. We have $(f^t)'(u^t(0, y))u^t_y(0, y) = -u^t_{xy}(0, y)$ and thus, by (6-7),

$$
\begin{aligned}
(f^t)'(u^t(0, y)) &= \frac{-u^t_{xy}}{u^t_y}(0, y) \\
&= \frac{-\partial_x P^t}{P^t}(0, y) = \frac{-1}{t}\left\{1 - \frac{t^2}{t^2+y^2} + \frac{t^2}{\sqrt{t^2+y^2}}\,\frac{K_1'(\sqrt{t^2+y^2})}{K_1(\sqrt{t^2+y^2})}\right\} \\
&= \frac{-1}{t}\left\{1 - \frac{t^2}{t^2+y^2} - \frac{t^2}{\sqrt{t^2+y^2}}\,\frac{K_0(\sqrt{t^2+y^2}) + K_2(\sqrt{t^2+y^2})}{2K_1(\sqrt{t^2+y^2})}\right\} \\
&= \frac{t}{\sqrt{t^2+y^2}}\left\{-\frac{\sqrt{t^2+y^2}}{t^2} + \frac{1}{\sqrt{t^2+y^2}} + \frac{K_0(\sqrt{t^2+y^2}) + K_2(\sqrt{t^2+y^2})}{2K_1(\sqrt{t^2+y^2})}\right\} \\
&=: \frac{t}{\sqrt{t^2+y^2}}\, h^t(y).
\end{aligned}
$$

It turns out that the function $\{K_0(s) + K_2(s)\}/(2K_1(s))$ is a decreasing function of $s \in (0, +\infty)$ which behaves as $1/s$ as $s \to 0^+$ and as $1 + 1/(2s)$ as $s \to +\infty$. Therefore,

$$
(f^t)'(0) = (f^t)'(1) = -\frac{1}{t} < 0.
$$

It also follows that $h^t$ is a decreasing function of $y \in [0, +\infty)$ with $h^t(0) = (K_0(t)+K_2(t))/(2K_1(t)) > 0$ and $\lim_{y\to+\infty} h^t(y) = -\infty$. Therefore, since $h^t$ is an even function of $y$, there exists a $y^t > 0$ such that $h^t$ is negative in $(-\infty, -y^t)$, positive in $(-y^t, y^t)$, and negative in $(y^t, +\infty)$. As a consequence, for some $0 < \gamma_1 < \gamma_2 < 1$, we have that $(f^t)'$ is negative in $(0, \gamma_1)$, positive in $(\gamma_1, \gamma_2)$, and negative in $(\gamma_2, 1)$. This gives that $f^t$ has a unique zero in $(0, 1)$ and that $f^t$ is of bistable type.

We finally check the positively balanced character of $f^t$ (after the end of the proof we give an alternative, more synthetic argument for this). Using formula (6-11) for $f^t$ in terms of $G^t$ and that $G^t_{xx} = -\partial_y(G^t_y + 2G^t)$, we have

$$
\begin{aligned}
\int_0^1 f^t(s)ds &= \int_{\mathbb{R}} f(u^t(0, y))(-u^t_y)(0, y)\,dy = -4\int_{\mathbb{R}} \{(G^t_y + 2G^t)G^t_x\}(0, y)\,dy \\
&= 4\int_{\mathbb{R}^2_+} \{(G^t_y + 2G^t)G^t_{xx} + (G^t_{xy} + 2G^t_x)G^t_x\}\,dx\,dy \\
&= 4\int_{\mathbb{R}^2_+} \left\{-\tfrac{1}{2}\partial_y(G^t_y + 2G^t)^2 + \tfrac{1}{2}\partial_y(G^t_x)^2 + 2(G^t_x)^2\right\} dx\,dy \\
&= 8\int_{\mathbb{R}^2_+} (G^t_x)^2\,dx\,dy > 0,
\end{aligned}
$$

which finishes the proof. □

The following is an alternative way to prove that the integral of $f^t$ is positive. It is more synthetic but it relies on deeper results. Assume by contradiction that $\int_0^1 f^t(s)\,ds \le 0$. Then, by the remarks made after

Theorem 1.2, the existence part of Theorem 1.2, and the results of [Cabré and Solà-Morales 2005], there exists a solution of (1-2) for some $c \le 0$ which satisfies the limits (1-3). But $u^t$ is also a solution of (1-2), now with $c = 2$, and satisfying the limits (1-3). By the uniqueness of the speed proved in Theorem 1.2, we arrive at a contradiction.

## 7. Proof of Theorem 1.2

In this section we use all the previous results to establish our main theorem.

*Proof of Theorem 1.2.* Let $f$ be of positively balanced bistable type or of combustion type as in Definition 1.1.

*Part (i).* This first part has been established in Proposition 2.7, together with Lemma 4.2, where we proved the existence of limits as $y \to \pm\infty$. That $0 < u < 1$ follows from the strong maximum principle and the Hopf's boundary lemma, since we know that $0 \le u \le 1$ and $f(0) = f(1) = 0$.

*Part (ii).* Let $(c_1, u_1)$ and $(c_2, u_2)$ be two solution pairs with $u_i$ taking values in [0, 1] and having limits 1 and 0 as $y \to \mp\infty$. By Lemma 4.1, both $u_i$ satisfy the uniform limits assumption (5-6). Translate each one in the $y$ variable so that both satisfy $u_i(0, 0) = \frac{1}{2}$. Assume that $c_1 \le c_2$. Proposition 5.3 applied with $c = c_1$ gives that the solution $u_1$ is decreasing in $y$. Thus

$$0 = \Delta u_1 + c_1 \partial_y u_1 \ge \Delta u_1 + c_2 \partial_y u_1,$$

and hence $u_1$ is a supersolution for the problem with $c = c_2$. Proposition 5.3 applied with $c = c_2$ leads to $u_1 \equiv u_2$. As a consequence, since $\partial_y u_1 < 0$, we deduce from the equations that $c_1 = c_2$.

*Part (iii).* The monotonicity in $y$ of a variational solution as in part (i) was established in Proposition 3.1. From $u_y \le 0$ we deduce $u_y < 0$ using the strong maximum principle and the Hopf's boundary lemma for the linearized problem satisfied by $u_y$. The existence of its vertical and horizontal limits has been proved in Lemmas 4.2 and 4.3. Let us now show that $u_x \le 0$ in the case that $f$ is of combustion type — this fact is not true for bistable nonlinearities since the normal derivative $-u_x = f(u)$ changes sign on $\{x = 0\}$. Note that $u_x$ is a solution of $(\Delta + c\partial_y)u_x = 0$ in $\mathbb{R}^2_+$, it is bounded by (4-4), and has limits 0 as $|y| \to \infty$ uniformly on compact sets of $x$ by (4-5). In addition, $u_x = -f(u) \le 0$ on $x = 0$. Lemma 5.1 leads to $u_x \le 0$ in $\mathbb{R}^2_+$.

*Part (iv).* We can give two different proofs of this part. Let $f_1$ and $f_2$ be not identically equal with $f_1 \ge f_2$. Let $(c_i, u_i)$ be the unique solution pair for the nonlinearity $f_i$ with $u_i$ taking values in [0, 1] and having limits as $y \to \pm\infty$.

The first proof is variational and uses formula (2-11) for the speed. Take $a > 0$ small enough so that both problems (1-21), for $f_1$ and for $f_2$, can be minimized in $B_a$. Since $G_1 \le G_2$, the minimum values satisfy $I_{1,a} \le I_{2,a}$, and the inequality is in fact strict since $G_1 \not\equiv G_2$ and the minimizers $\underline{u}_1$ and $\underline{u}_2$ take all values in (0, 1). Thus, from (2-11) we deduce $c_1 > c_2$.

The second proof of (iv) is nonvariational. Recall that, by Lemma 4.1, both $u_i$ satisfy the uniform limits assumption (5-6). Translate each front in the $y$ variable so that both satisfy $u_i(0, 0) = \frac{1}{2}$. Assume, arguing by contradiction, that $c_1 \le c_2$. Since $u_1$ is decreasing in $y$, we have

$$0 = \Delta u_1 + c_1 \partial_y u_1 \ge \Delta u_1 + c_2 \partial_y u_1.$$

In addition,

$$\frac{\partial u_1}{\partial \nu} = f_1(u_1) \ge f_2(u_1) \quad \text{on } \partial \mathbb{R}^2_+.$$

Hence, $u_1$ is a supersolution for the problem with $c = c_2$ and $f = f_2$. Proposition 5.3 leads to $u_1 \equiv u_2$. As a consequence, we obtain $f_1 \equiv f_2$ — since the image of $u_1$ is the whole of (0, 1). This is a contradiction.

*Part (v).* To establish this part, it suffices to show the bounds for $-u_y$. From them, the bounds for $u$ and $1-u$ follow by integration. Defining $\tilde{u}$ by $u(x, y) = \tilde{u}(cx/2, cy/2)$, we see that $\tilde{u}$ is a front with speed 2 for (1-2) with nonlinearity $(2/c)f$. Since the constants on the bounds of part (v) do not reflect the dependence on $f$, we may rename $\tilde{u}$ by $u$ and $(2/c)f$ by $f$, and assume that $u$ is a front for the nonlinearity $f$ with speed $c = 2$. Note however that the factor $e^{-2y}$ in (1-9) will change to $e^{-cy}$ after the scaling.

As stated in the theorem, the lower bounds for $-u_y$ hold for any $f$ of positively balanced bistable type or of combustion type. To prove them, we take $t > 0$ small enough that

$$-\frac{1}{2t} \le \min_{[0,1]} f'.$$

For such $t$, consider the Poisson kernel $P^t$ defined by (6-6) and, for any positive constant $C > 0$, the function

$$v := C(-u_y) - P^t.$$

Note that $\Delta v + 2v_y = 0$ in $\mathbb{R}^2_+$. By using (6-8) and that $-u_y$ and $P^t$ are positive, it follows that $\{-\partial_x v + (2t)^{-1} v\}(0, y) \ge 0$ for $|y|$ large enough, say for $y$ in the complement of a compact interval $H$. Next, take the constant $C > 0$ large enough so that $v > 0$ in the compact set $H$. By (4-4), the limits of $u_y$ established in (4-5), and the properties of $P^t$ checked in Section 6, we can apply Proposition 5.2 with $c = 2$ and $d(y) = (2t)^{-1}$ to deduce that $v > 0$ in $\mathbb{R}^2_+$. By using the asymptotic behaviors (6-9) and (6-10) of $P^t$ at $\pm\infty$, we conclude the two lower bounds for $-u_y$.

To prove the upper bounds for $-u_y$ we need to assume that $f'(0) < 0$ and $f'(1) < 0$. We proceed in the same way as for the lower bounds, but replacing the roles of $-u_y$ and $P^t$. We now take $t > 0$ large enough that

$$\max\{f'(0), f'(1)\} < -\frac{2}{t}.$$

Using (6-8), for any $C > 0$, $\tilde{v} := CP^t - (-u_y)$ satisfies $\{-\partial_x \tilde{v} + (2/t)\tilde{v}\}(0, y) \ge 0$ for $|y|$ large enough, say for $y$ in the complement of a compact interval $H$. One proceeds exactly as before to obtain $\tilde{v} > 0$ in $\mathbb{R}^2_+$ for $C$ large enough. This gives the desired upper bounds for $-u_y$. □

## References

[Abramowitz and Stegun 1964] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, National Bureau of Standards Applied Mathematics Series **55**, U.S. Government Printing Office, Washington, DC, 1964. Reprinted by Dover, New York, 1974. MR 29 #4914 Zbl 0171.38503

[Berestycki and Nirenberg 1991] H. Berestycki and L. Nirenberg, "On the method of moving planes and the sliding method", *Bol. Soc. Brasil. Mat.* (*N.S.*) **22**:1 (1991), 1–37. MR 93a:35048 Zbl 0784.35025

[Berestycki and Nirenberg 1992] H. Berestycki and L. Nirenberg, "Travelling fronts in cylinders", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **9**:5 (1992), 497–572. MR 93k:35019 Zbl 0799.35073

[Brock 1999] F. Brock, "Weighted Dirichlet-type inequalities for Steiner symmetrization", *Calc. Var. Partial Differential Equations* **8**:1 (1999), 15–25. MR 2000e:35038 Zbl 0947.35056

[Cabré and Sire 2015] X. Cabré and Y. Sire, "Nonlinear equations for fractional Laplacians, II: Existence, uniqueness, and qualitative properties of solutions", *Trans. Amer. Math. Soc.* **367**:2 (2015), 911–941. MR 3280032

[Cabré and Solà-Morales 2005] X. Cabré and J. Solà-Morales, "Layer solutions in a half-space for boundary reactions", *Comm. Pure Appl. Math.* **58**:12 (2005), 1678–1732. MR 2006i:35116 Zbl 1102.35034

[Caffarelli et al. 2012] L. Caffarelli, A. Mellet, and Y. Sire, "Traveling waves for a boundary reaction-diffusion equation", *Adv. Math.* **230**:2 (2012), 433–457. MR 2914954 Zbl 1255.35072

[Gilbarg and Trudinger 1983] D. Gilbarg and N. S. Trudinger, *Elliptic partial differential equations of second order*, 2nd ed., Grundlehren der Mathematischen Wissenschaften **224**, Springer, Berlin, 1983. MR 86c:35035 Zbl 0562.35001

[Gui and Zhao 2014] C. Gui and M. Zhao, "Traveling wave solutions of Allen–Cahn equation with a fractional Laplacian", *Ann. Inst. H. Poincaré Anal. Non Linéaire* (online publication May 2014).

[Heinze 2001] S. Heinze, "A variational approach to traveling waves", preprint 85, Max-Planck-Institut für Mathematik in den Naturwissenschaften, Leipzig, 2001.

[Kawohl 1985] B. Kawohl, *Rearrangements and convexity of level sets in PDE*, Lecture Notes in Math. **1150**, Springer, Berlin, 1985. MR 87a:35001 Zbl 0593.35002

[Kyed 2008] M. Kyed, "Existence of travelling wave solutions for the heat equation in infinite cylinders with a nonlinear boundary condition", *Math. Nachr.* **281**:2 (2008), 253–271. MR 2009d:35196 Zbl 1139.35062

[Landes 2007] R. Landes, "Some remarks on rearrangements and functionals with non-constant density", *Math. Nachr.* **280**:5-6 (2007), 560–570. MR 2008c:35006 Zbl 1127.26008

[Landes 2009] R. Landes, "Wavefront solution in the theory of boiling liquids", *Analysis* (*Munich*) **29**:3 (2009), 283–298. MR 2011b:35195 Zbl 1183.35172

[Landes 2012] R. Landes, "Stable and unstable initial configuration in the theory wave fronts", *Discrete Contin. Dyn. Syst.* (*S*) **5**:4 (2012), 797–808. MR 2012j:35203 Zbl 1248.35104

[Lucia et al. 2004] M. Lucia, C. B. Muratov, and M. Novaga, "Linear vs. nonlinear selection for the propagation speed of the solutions of scalar reaction-diffusion equations invading an unstable equilibrium", *Comm. Pure Appl. Math.* **57**:5 (2004), 616–636. MR 2005a:35162 Zbl 1053.35065

[Mellet et al. 2010] A. Mellet, J.-M. Roquejoffre, and Y. Sire, "Generalized fronts for one-dimensional reaction-diffusion equations", *Discrete Contin. Dyn. Syst.* (*A*) **26**:1 (2010), 303–312. MR 2011b:35238 Zbl 1180.35294

Received 23 Apr 2014. Accepted 26 Nov 2014.

XAVIER CABRÉ: xavier.cabre@upc.edu
*Icrea and Universitat Politécnica de Catalunya, Departament de Matemática Aplicada I, Diagonal 647, 08028 Barcelona, Spain*

NEUS CÓNSUL: neus.consul@upc.edu
*Universitat Politécnica de Catalunya, Departament de Matemática Aplicada I, Diagonal 647, 08028 Barcelona, Spain*

JOSÉ V. MANDÉ: josemande@gmail.com
*Universitat Politécnica de Catalunya, Departament de Matemática Aplicada I, Diagonal 647, 08028 Barcelona, Spain*

**mathematical sciences publishers** msp

ANALYSIS AND PDE
Vol. 8, No. 2, 2015
dx.doi.org/10.2140/apde.2015.8.365

# LOCALLY CONFORMALLY FLAT ANCIENT RICCI FLOWS

GIOVANNI CATINO, CARLO MANTEGAZZA AND LORENZO MAZZIERI

We show that any locally conformally flat ancient solution to the Ricci flow must be rotationally symmetric. As a by-product, we prove that any locally conformally flat Ricci soliton is a gradient soliton in the shrinking and steady cases as well as in the expanding case, provided the soliton has nonnegative curvature.

## 1. Introduction

In this paper, we study ancient solutions to the Ricci flow. We recall that a time-dependent metric $g(t)$ on a Riemannian manifold $M$ is a solution to the Ricci flow if it evolves by the equation

$$\frac{\partial}{\partial t} g(t) = -2\,\mathrm{Ric}_{g(t)}\,.$$

A solution is called *ancient* if it is defined for every negative time. Ancient solutions typically arise as the limit of a sequence of suitable blow-ups as the time approaches a singular time for the Ricci flow. In dimension two there exists a compact, rotationally symmetric, ancient solution due to King [1990], Rosenau [1995] and Fateev, Onofri and Zamolodchikov [Fateev et al. 1993]. In dimension three, Perelman [2003] constructed a compact, rotationally symmetric, ancient solution on the three-sphere. In the nonrotationally symmetric case, the first construction is due to Fateev [1996] in dimension three. Motivated by this construction, Bakas, Kong and Ni [Bakas et al. 2012] produced high-dimensional, compact, ancient solutions to the Ricci flow which are not rotationally symmetric.

In dimension two, Daskalopoulos, Hamilton and Sesum [Daskalopoulos et al. 2012] have obtained a complete classification of all compact ancient solutions to the Ricci flow. Ni [2010] showed that any compact ancient solution to the Ricci flow which is of type I is $k$-noncollapsed and has positive curvature operator has constant sectional curvature. Brendle, Huisken and Sinestrari [Brendle et al. 2011] proved that any compact ancient solution which satisfies a suitable pinching condition must have constant sectional curvature.

In this article, we show that any complete ancient solution to the Ricci flow in dimension $n \geq 4$ which is locally conformally flat along the flow must be rotationally symmetric.

**Theorem 1.1.** *Let $(M^n, g(t))$, $n \geq 4$, be a complete ancient solution to the Ricci flow which is locally conformally flat at every time. Then $(M^n, g(t))$ is rotationally symmetric.*

---

*MSC2010:* 53C24, 53C25, 53C44.
*Keywords:* Ricci solitons, Ricci flow.

The nonrotationally symmetric examples of Bakas, Kong and Ni show that the locally conformally flatness assumption cannot be removed. The proof of Theorem 1.1 relies on previous work of Catino and Mantegazza [2011] about the behavior of the Weyl tensor under the Ricci flow, combined with a more recent result [Catino et al. 2014] concerning the classification of Riemannian manifolds admitting a Codazzi tensor with exactly two distinct eigenvalues.

As a consequence of Theorem 1.1, we classify locally conformally flat Ricci solitons. We recall that a complete Riemannian manifold $(M^n, g)$ is a *Ricci soliton* if there exists a vector field $X$ on $M^n$ such that

$$\mathrm{Ric} + \tfrac{1}{2}\mathscr{L}_X g = \lambda g$$

for some constant $\lambda$. The Ricci soliton is called shrinking if $\lambda > 0$, steady if $\lambda = 0$, and expanding if $\lambda < 0$. If $X = \nabla f$ for some smooth function $f$, then the soliton is called a *gradient Ricci soliton*. It follows from the work of Perelman [2002] (see [Eminenti et al. 2008], for instance, for a direct proof) that any compact Ricci soliton is actually a gradient soliton. Moreover, Naber [2010] has shown that any shrinking Ricci soliton with bounded curvature has a gradient soliton structure. On the other hand, steady and expanding Ricci solitons which do not support a gradient structure were found by [Lauret 2001; Baird and Danielo 2007; Lott 2007; Baird 2009].

In this article we prove the following result:

**Theorem 1.2.** *Let $(M^n, g, X)$, $n \geq 4$, be a complete, locally conformally flat, shrinking or steady Ricci soliton. Then it is a gradient Ricci soliton. The conclusion still holds in the expanding case provided the soliton has nonnegative curvature operator.*

In particular, from the classification results of locally conformally flat gradient Ricci solitons in the shrinking case [Cao et al. 2011; Eminenti et al. 2008; Ni and Wallach 2008; Petersen and Wylie 2010; Zhang 2009a], in the steady case [Cao and Chen 2012; Catino and Mantegazza 2011], as well as in the expanding case [Catino and Mantegazza 2011], we obtain the following corollaries:

**Corollary 1.3.** *Let $(M^n, g, X)$, $n \geq 4$, be a complete, locally conformally flat, shrinking Ricci soliton. Then it is isometric to a quotient of $\mathbb{S}^n$, $\mathbb{R} \times \mathbb{S}^{n-1}$ or $\mathbb{R}^n$.*

**Corollary 1.4.** *Let $(M^n, g, X)$, $n \geq 4$, be a complete, locally conformally flat, steady Ricci soliton. Then it is isometric to a quotient of $\mathbb{R}^n$ or the Bryant soliton.*

**Corollary 1.5.** *Let $(M^n, g, X)$, $n \geq 4$, be a complete, locally conformally flat, expanding Ricci soliton with nonnegative curvature operator. Then it is a rotationally symmetric gradient expanding Ricci soliton.*

We note that rotationally symmetric gradient expanding Ricci solitons were constructed in [Bryant 2005; Cao 1997; Feldman et al. 2003].

## 2. Notations and preliminaries

The Riemann curvature operator of a Riemannian manifold $(M^n, g)$ is defined as in [Gallot et al. 1990] by

$$\mathrm{Riem}(X, Y)Z = \nabla_Y \nabla_X Z - \nabla_X \nabla_Y Z + \nabla_{[X,Y]} Z \,.$$

In a local coordinate system the components of the $(3,1)$-Riemann curvature tensor are given by $\mathrm{R}^l_{ijk}\partial/\partial x^l = \mathrm{Riem}(\partial/\partial x^i, \partial/\partial x^j)\partial/\partial x^k$ and we denote by $\mathrm{R}_{ijkl} = g_{lm}\mathrm{R}^m_{ijk}$ its $(4,0)$-version. With this choice, we have that the round sphere $\mathbb{S}^n$ has positive curvature, meaning that

$$\mathrm{Riem}(v, w, v, w) = \mathrm{R}_{ijkl}v^i w^j v^k w^l > 0$$

for every couple $u$ and $v$ of nonparallel vector fields. The Ricci tensor is obtained by the contraction $\mathrm{R}_{ik} = g^{jl}\mathrm{R}_{ijkl}$, and $\mathrm{R} = g^{ik}\mathrm{R}_{ik}$ will denote the scalar curvature. The so-called Weyl tensor is then defined by the following decomposition formula (see [Gallot et al. 1990, Chapter 3, Section K]) in dimension $n \geq 3$:

$$\mathrm{W}_{ijkl} = \mathrm{R}_{ijkl} + \frac{\mathrm{R}}{(n-1)(n-2)}(g_{ik}g_{jl} - g_{il}g_{jk}) - \frac{1}{n-2}(\mathrm{R}_{ik}g_{jl} - \mathrm{R}_{il}g_{jk} + \mathrm{R}_{jl}g_{ik} - \mathrm{R}_{jk}g_{il}). \tag{2-1}$$

The Weyl tensor shares the symmetries of the curvature tensor. Moreover, as can be easily seen by the formula above, all of its contractions with the metric are zero.

In dimension three, W is identically zero on every Riemannian manifold, whereas, when $n \geq 4$, the vanishing of the Weyl tensor is a relevant condition, since it is equivalent to the *locally conformally flatness* of $(M^n, g)$. We recall that this latter condition means that around every point $p \in M^n$ there exists a smooth function $f$ defined in a open neighborhood $U_p$ of $p$ such that the conformal deformation $\tilde{g}$ of the original metric $g$ defined by $\tilde{g}_{ij} = e^f g_{ij}$ is flat. In particular, the Riemann tensor associated to $\tilde{g}$ is zero in $U_p$.

We also recall that, in dimension $n = 3$, *locally conformally flatness* is equivalent to the vanishing of the Cotton tensor

$$\mathrm{C}_{ijk} = \nabla_k \mathrm{R}_{ij} - \nabla_j \mathrm{R}_{ik} - \frac{1}{2(n-1)}(\nabla_k \mathrm{R} g_{ij} - \nabla_j \mathrm{R} g_{ik}).$$

By direct computation, we can see that the tensor $\mathrm{C}_{ijk}$ satisfies the symmetries

$$\mathrm{C}_{ijk} = -\mathrm{C}_{ikj}, \quad \mathrm{C}_{ijk} + \mathrm{C}_{jki} + \mathrm{C}_{kij} = 0; \tag{2-2}$$

moreover, it is trace-free in any two indices:

$$g^{ij}\mathrm{C}_{ijk} = g^{ik}\mathrm{C}_{ijk} = g^{jk}\mathrm{C}_{ijk} = 0 \tag{2-3}$$

by its skew-symmetry and the Schur lemma. We note that, for $n \geq 4$,

$$\nabla^l \mathrm{W}_{ijkl} = -\frac{n-3}{n-2}\mathrm{C}_{ijk}, \tag{2-4}$$

and we refer the reader to [Besse 1988] for the detailed computation. It follows from this formula that, in every dimension $n \geq 3$, the vanishing of the Cotton tensor is a necessary condition for a Riemannian manifold $(M^n, g)$ to be *locally conformally flat*. We also note that the vanishing of the Cotton tensor can be rephrased in terms of the so-called Schouten tensor

$$\mathrm{S}_{ij} = \mathrm{R}_{ij} - \frac{1}{2(n-1)}\mathrm{R} g_{ij}$$

by saying that S must satisfy the Codazzi equation

$$(\nabla_X \mathrm{S})Y = (\nabla_Y \mathrm{S})X, \quad X,\ Y \in TM.$$

Any symmetric two-tensor satisfying this condition is called a Codazzi tensor (see [Besse 1988, Chapter 16] for a general overview on Codazzi tensors). Hence, if $(M^n, g)$, $n \geq 3$, is a *locally conformally flat* manifold, then the Schouten tensor is a Codazzi tensor.

## 3. Proof of Theorem 1.1

Let $(M^n, g(t))$, $n \geq 4$, be a complete ancient solution to the Ricci flow. We assume that, along the flow, the Weyl tensor remains identically zero. As was observed in [Catino and Mantegazza 2011], this condition implies a strong rigidity on the eigenvalues of the Ricci tensor. More precisely, one has the following result:

**Lemma 3.1** [Catino and Mantegazza 2011, Corollary 1.2]. *Let $(M^n, g)$, $n \geq 4$, be a solution to the Ricci flow such that the Weyl tensor remains identically zero at every time. Then, at every point, either the Ricci tensor is proportional to the metric or it has an eigenvalue of multiplicity* $n-1$ *and another of multiplicity* 1.

By the results in [Chen 2009; Zhang 2009a], which generalize the well-known Hamilton–Ivey curvature estimate, we know that every complete ancient solution $g(t)$ to the Ricci flow whose Weyl tensor is identically zero for all times is forced to have nonnegative curvature operator for every time $t$. Moreover, by Hamilton's strong maximum principle for systems in [Hamilton 1986], we have that either the metric has strictly positive curvature operator or it splits a line. By Theorem 1.167 in [Besse 1988], a Riemannian product $(\mathbb{R} \times N^{n-1}, ds \times h)$ is locally conformally flat if and only if the manifold $(N^{n-1}, h)$ has constant curvature; hence, one of the following possibilities holds: $(M^n, g(t))$ is flat, or it is a quotient of a rescaling of $\mathbb{R} \times \mathbb{S}^{n-1}$, or it has positive curvature operator. Since the first two cases satisfy the conclusion of the theorem, from now on we assume that $(M^n, g(t))$ is a complete, locally conformally flat, ancient solution to the Ricci flow with positive curvature operator.

As we have seen in the previous section, the relation (2-4) implies that the Cotton tensor is identically zero, hence the Schouten tensor

$$\mathrm{S}_{ij} = \mathrm{R}_{ij} - \frac{1}{2(n-1)}\mathrm{R} g_{ij}$$

is a Codazzi tensor. Moreover, from Lemma 3.1, we know that, at every point, either the metric is Einstein or the Ricci tensor (and so the Schouten tensor) has two distinct eigenvalues of multiplicity 1 and $n-1$, respectively. Now, it was proved by Bando [1987] that solutions to the Ricci flow are real analytic. To be precise, Bando showed that any Ricci flow solutions is real analytic if $M^n$ is compact or if it is complete with uniformly bounded curvature. This result was recently improved by Kotschwar [2013], who showed a local version of Bando's result. It follows that, if the metric is Einstein in some open subset of $M^n$, then it is Einstein everywhere and, by conformally flatness and positivity of the curvature, the manifold $(M^n, g)$ must be isometric to a quotient of $\mathbb{S}^n$. Thus, either $(M^n, g(t))$ has constant positive sectional curvature or the Schouten tensor has an eigenvalue of multiplicity 1 and a different one of multiplicity $n-1$ at every point of some open dense subset $U$ of $M^n$. In the latter case, we apply to such an open set $U$ the following classification result of Riemannian manifolds admitting a Codazzi tensor with two distinct eigenvalues:

**Lemma 3.2** [Catino et al. 2014, Theorem 2.1]. *Let* T *be a Codazzi tensor on the Riemannian manifold* $(U, g)$ *with* $n \geq 3$. *Suppose that, at every point of* $U$, *the tensor* T *has exactly two distinct eigenvalues* $\rho$ *and* $\sigma$, *of multiplicity* 1 *and* $n-1$, *respectively. Finally, we let* $W = \{p \in U \mid d\sigma(p) \neq 0\}$. *Then we have that*:

(1) *The closed set* $\overline{W} = W \cup \partial W$ *with the metric* $g|_{\overline{W}}$ *is locally isometric to the warped product of some* $(n-1)$*-dimensional Riemannian manifold on an interval of* $\mathbb{R}$ *and* $\sigma$ *is constant along the "leaves" of the warped product.*

(2) *The boundary of* $W$, *if present*, *is given by the disjoint union of connected*, *totally geodesic hypersurfaces where* $\sigma$ *is constant.*

(3) *Each connected component of the complement of* $\overline{W}$ *in* $U$, *if present*, *has* $\sigma$ *constant and it is foliated by totally geodesic hypersurfaces.*

We are going to show that, under our assumptions, case (3) cannot occur and $W = U$. In fact, if a connected component of the complement of $\overline{W}$ in $U$ is present, the proof of this lemma (see also [Besse 1988, Proposition 16.11]) shows that the totally geodesic foliation of $(U, g(t))$ is *integrable*. Then, since the manifold $(U, g(t))$ has nonnegative sectional curvature, it follows from [Abe 1973, Corollary 2] that such a component must split a flat factor. This is clearly in contradiction with the positivity of the curvature, hence it must be that $U = W$ in this lemma and $(U, g(t))$ is locally a warped product of some $(n-1)$-dimensional manifold on an interval of $\mathbb{R}$. Since $(U, g(t))$ is locally conformally flat with positive curvature operator, we have that the $(n-1)$-dimensional fibers of the warped product are isometric to $\mathbb{S}^{n-1}$ and the metric is rotationally symmetric. By the density of $U$ in $M^n$, this conclusion clearly holds for the whole $(M^n, g)$. This concludes the proof of Theorem 1.1.

**Remark 3.3.** We would like to note that the same argument shows that the conclusion of Theorem 1.1 still holds if one consider a Ricci flow solution $(M^n, g(t))$, $n \geq 4$, defined on some time interval $I \subseteq \mathbb{R}$, which is locally conformally flat with nonnegative curvature operator for every $t \in I$.

## 4. Proof of Theorem 1.2

Now we turn our attention to the classification of locally conformally flat Ricci solitons. Let $(M^n, g, X)$ be a complete, locally conformally flat shrinking or steady Ricci soliton. In particular, it generates a self-similar ancient solution $g(t)$ to Ricci flow (see [Zhang 2009b]) which is locally conformally flat at every time $t$. Hence, Theorem 1.1 implies that the manifold is rotationally symmetric with nonnegative curvature operator. As we observed in Remark 3.3, the conclusion still holds if we consider an expanding Ricci soliton with nonnegative curvature operator. To prove Theorem 1.2 we then apply the following result:

**Lemma 4.1** [Catino and Mantegazza 2011, Proposition 2.6]. *Let $(M^n, g, X)$ be a complete, locally warped, locally conformally flat Ricci soliton with nonnegative Ricci tensor; then it is a gradient Ricci soliton with a potential function $f : M^n \to \mathbb{R}$ (hence, $X = \nabla f$) depending only on the $r$ variable of the warping interval.*

## Acknowledgments

This work was partially supported by the Italian projects FIRB–IDEAS "Analysis and beyond" and GNAMPA projects "Flussi geometrici e soluzioni autosimilari" and "Equazioni di evoluzione geometriche e strutture di tipo Einstein". Mantegazza and Mazzieri were partially supported by the SNS project "Geometric flows and related topics". Mazzieri was also partially supported by the Italian project FIRB 2012 "Geometria Differenziale e Teoria Geometrica delle Funzioni".

## References

[Abe 1973] K. Abe, "Applications of a Riccati type differential equation to Riemannian manifolds with totally geodesic distributions", *Tôhoku Math. J.* (2) **25** (1973), 425–444. MR 50 #3163 Zbl 0283.53045

[Baird 2009] P. Baird, "A class of three-dimensional Ricci solitons", *Geom. Topol.* **13**:2 (2009), 979–1015. MR 2010a:53131 Zbl 1157.35465

[Baird and Danielo 2007] P. Baird and L. Danielo, "Three-dimensional Ricci solitons which project to surfaces", *J. Reine Angew. Math.* **608** (2007), 65–91. MR 2008f:53086 Zbl 1128.53020

[Bakas et al. 2012] I. Bakas, S. Kong, and L. Ni, "Ancient solutions of Ricci flow on spheres and generalized Hopf fibrations", *J. Reine Angew. Math.* **663** (2012), 209–248. MR 2889711 Zbl 1269.53065

[Bando 1987] S. Bando, "Real analyticity of solutions of Hamilton's equation", *Math. Z.* **195**:1 (1987), 93–97. MR 88i:53073 Zbl 0606.58051

[Besse 1988] A. L. Besse, *Einstein manifolds*, Ergebnisse der Math. und ihrer Grenzgebiete **10**, Springer, Berlin, 1988. MR 2008k:53084 Zbl 1147.53001

[Brendle et al. 2011] S. Brendle, G. Huisken, and C. Sinestrari, "Ancient solutions to the Ricci flow with pinched curvature", *Duke Math. J.* **158**:3 (2011), 537–551. MR 2012f:53133 Zbl 1219.53062

[Bryant 2005] R. L. Bryant, "Ricci flow solitons in dimension three with SO(3)-symmetries", preprint, 2005, available at http://www.math.duke.edu/~bryant/3DRotSymRicciSolitons.pdf.

[Cao 1997] H.-D. Cao, "Limits of solutions to the Kähler–Ricci flow", *J. Differential Geom.* **45** (1997), 257–272. MR 99g:53042 Zbl 0889.58067

[Cao and Chen 2012] H.-D. Cao and Q. Chen, "On locally conformally flat gradient steady Ricci solitons", *Trans. Amer. Math. Soc.* **364**:5 (2012), 2377–2391. MR 2888210 Zbl 1245.53038

[Cao et al. 2011] X. Cao, B. Wang, and Z. Zhang, "On locally conformally flat gradient shrinking Ricci solitons", *Commun. Contemp. Math.* **13**:2 (2011), 269–282. MR 2012e:53074 Zbl 1215.53061

[Catino and Mantegazza 2011] G. Catino and C. Mantegazza, "The evolution of the Weyl tensor under the Ricci flow", *Ann. Inst. Fourier* (*Grenoble*) **61**:4 (2011), 1407–1435. MR 2951497 Zbl 1255.53034

[Catino et al. 2014] G. Catino, C. Mantegazza, and L. Mazzieri, "A note on Codazzi tensors", *Math. Ann.* (online publication November 2014).

[Chen 2009] B.-L. Chen, "Strong uniqueness of the Ricci flow", *J. Differential Geom.* **82**:2 (2009), 363–382. MR 2010h:53095 Zbl 1177.53036

[Daskalopoulos et al. 2012] P. Daskalopoulos, R. Hamilton, and N. Sesum, "Classification of ancient compact solutions to the Ricci flow on surfaces", *J. Differential Geom.* **91**:2 (2012), 171–214. MR 2971286 Zbl 1257.53095

[Eminenti et al. 2008] M. Eminenti, G. La Nave, and C. Mantegazza, "Ricci solitons: the equation point of view", *Manuscripta Math.* **127**:3 (2008), 345–367. MR 2009m:53088 Zbl 1160.53031

[Fateev 1996] V. A. Fateev, "The sigma model (dual) representation for a two-parameter family of integrable quantum field theories", *Nuclear Phys. B* **473**:3 (1996), 509–538. MR 97j:81310 Zbl 0925.81297

[Fateev et al. 1993] V. A. Fateev, E. Onofri, and A. B. Zamolodchikov, "Integrable deformations of the O(3) sigma model: the sausage model", *Nuclear Phys. B* **406**:3 (1993), 521–565. MR 95d:81111 Zbl 0990.81683

[Feldman et al. 2003] M. Feldman, T. Ilmanen, and D. Knopf, "Rotationally symmetric shrinking and expanding gradient Kähler–Ricci solitons", *J. Differential Geom.* **65**:2 (2003), 169–209. MR 2005e:53102 Zbl 1069.53036

[Gallot et al. 1990] S. Gallot, D. Hulin, and J. Lafontaine, *Riemannian geometry*, 2nd ed., Springer, Berlin, 1990. MR 91j:53001 Zbl 0716.53001

[Hamilton 1986] R. S. Hamilton, "Four-manifolds with positive curvature operator", *J. Differential Geom.* **24**:2 (1986), 153–179. MR 87m:53055 Zbl 0628.53042

[King 1990] J. R. King, "Exact similarity solutions to some nonlinear diffusion equations", *J. Phys. A* **23**:16 (1990), 3681–3697. MR 91j:35062 Zbl 0728.35052

[Kotschwar 2013] B. L. Kotschwar, "A local version of Bando's theorem on the real-analyticity of solutions to the Ricci flow", *Bull. Lond. Math. Soc.* **45**:1 (2013), 153–158. MR 3033963 Zbl 1259.53065

[Lauret 2001] J. Lauret, "Ricci soliton homogeneous nilmanifolds", *Math. Ann.* **319**:4 (2001), 715–733. MR 2002k:53083 Zbl 0987.53019

[Lott 2007] J. Lott, "On the long-time behavior of type-III Ricci flow solutions", *Math. Ann.* **339**:3 (2007), 627–666. MR 2008i:53093 Zbl 1135.53046

[Naber 2010] A. Naber, "Noncompact shrinking four solitons with nonnegative curvature", *J. Reine Angew. Math.* **645** (2010), 125–153. MR 2012k:53081 Zbl 1196.53041

[Ni 2010] L. Ni, "Closed type I ancient solutions to Ricci flow", pp. 147–150 in *Recent advances in geometric analysis*, edited by Y.-I. Lee et al., Adv. Lect. Math. (ALM) **11**, International Press, 2010. MR 2011d:53161 Zbl 1198.53074

[Ni and Wallach 2008] L. Ni and N. Wallach, "On a classification of gradient shrinking solitons", *Math. Res. Lett.* **15**:5 (2008), 941–955. MR 2010i:53127 Zbl 1158.53052

[Perelman 2002] G. Perelman, "The entropy formula for the Ricci flow and its geometric applications", preprint, 2002. Zbl 1130.53001 arXiv 0211159

[Perelman 2003] G. Perelman, "Ricci flow with surgery on three–manifolds", preprint, 2003. Zbl 1130.53002 arXiv 0303109

[Petersen and Wylie 2010] P. Petersen and W. Wylie, "On the classification of gradient Ricci solitons", *Geom. Topol.* **14**:4 (2010), 2277–2300. MR 2012a:53060 Zbl 1202.53049

[Rosenau 1995] P. Rosenau, "Fast and superfast diffusion processes", *Phys. Rev. Lett.* **74** (1995), 1056–1059.

[Zhang 2009a] Z.-H. Zhang, "Gradient shrinking solitons with vanishing Weyl tensor", *Pacific J. Math.* **242**:1 (2009), 189–200. MR 2010f:53116 Zbl 1171.53332

[Zhang 2009b] Z.-H. Zhang, "On the completeness of gradient Ricci solitons", *Proc. Amer. Math. Soc.* **137**:8 (2009), 2755–2759. MR 2010a:53057 Zbl 1176.53046

Received 8 May 2014. Accepted 9 Feb 2015.

GIOVANNI CATINO: giovanni.catino@polimi.it
*Dipartimento di Matematica, Politecnico di Milano, Piazza Leonardo da Vinci 32, I-20133 Milano, Italy*

CARLO MANTEGAZZA: c.mantegazza@sns.it
*Scuola Normale Superiore di Pisa, Piazza dei Cavalieri 7, I-56126 Pisa, Italy*

LORENZO MAZZIERI: l.mazzieri@sns.it
*Scuola Normale Superiore di Pisa, Piazza dei Cavalieri 7, I-56126 Pisa, Italy*

ANALYSIS AND PDE
Vol. 8, No. 2, 2015
dx.doi.org/10.2140/apde.2015.8.373

# MOTION OF THREE-DIMENSIONAL ELASTIC FILMS BY ANISOTROPIC SURFACE DIFFUSION WITH CURVATURE REGULARIZATION

IRENE FONSECA, NICOLA FUSCO, GIOVANNI LEONI AND MASSIMILIANO MORINI

Short time existence for a surface diffusion evolution equation with curvature regularization is proved in the context of epitaxially strained three-dimensional films. This is achieved by implementing a minimizing movement scheme, which is hinged on the $H^{-1}$-gradient flow structure underpinning the evolution law. Long-time behavior and Liapunov stability in the case of initial data close to a flat configuration are also addressed.



## 1. Introduction

In this paper we study the morphologic evolution of anisotropic, epitaxially strained films, driven by stress and surface mass transport in three dimensions. This can be viewed as the evolutionary counterpart of the static theory developed in [Bonnetier and Chambolle 2002; Fonseca et al. 2007; 2011; Fusco and Morini 2012; Bonacini 2013a; Capriani et al. 2013] in the two-dimensional case and in [Bonacini 2013b] in three dimensions. The two-dimensional formulation of the same evolution problem has been addressed in [Fonseca et al. 2012] (see also [Piovano 2014] for the case of motion by evaporation–condensation).

The physical setting behind the evolution equation is the following. The free interface is allowed to evolve via *surface diffusion* under the influence of a chemical potential $\mu$. Assuming that mass transport in the bulk occurs at a much faster time scale, and thus can be neglected (see [Mullins 1963]), we have, according to the Einstein–Nernst relation, that the evolution is governed by the *volume-preserving* equation

$$V = C\Delta_\Gamma \mu, \tag{1-1}$$

where $C > 0$, $V$ denotes the normal velocity of the evolving interface $\Gamma$, $\Delta_\Gamma$ stands for the tangential laplacian, and the chemical potential $\mu$ is given by the first variation of the underlying free energy functional.

*MSC2010:* primary 35K25, 53C44, 74K35, 35Q74; secondary 37B25.
*Keywords:* minimizing movements, surface diffusion, gradient flows, higher order geometric flows, elastically stressed epitaxial films, volume-preserving evolution, long-time behavior, Liapunov stability.

In our case, the free energy functional associated with the physical system is given by

$$\int_{\Omega_h} W(E(u))\,dz + \int_{\Gamma_h} \psi(\nu)\,d\mathcal{H}^2, \tag{1-2}$$

where $h$ is the function whose graph $\Gamma_h$ describes the evolving profile of the film, $\Omega_h$ is the region occupied by the film, $u$ is displacement of the material, which is assumed to be in (quasistatic) elastic equilibrium at each time, $E(u)$ is the symmetric part of $Du$, $W$ is a positive definite quadratic form, and $\mathcal{H}^2$ denotes the two-dimensional Hausdorff measure. Finally, $\psi$ is an anisotropic surface energy density, evaluated at the unit normal $\nu$ to $\Gamma_h$. The first variation of (1-2) can be written as the sum of three contributions: a constant Lagrange multiplier related to mass conservation, the (anisotropic) curvature of the surface, and the elastic energy density evaluated at the displacement of the solid on the profile of the film. Hence, (1-1) takes the form (assuming $C = 1$)

$$V = \Delta_\Gamma\big[\mathrm{Div}_\Gamma(D\psi(\nu)) + W(E(u))\big], \tag{1-3}$$

where $\mathrm{Div}_\Gamma$ stands for the tangential divergence along $\Gamma_{h(\cdot,t)}$, and $u(\cdot,t)$ is the elastic equilibrium in $\Omega_{h(\cdot,t)}$, i.e., the minimizer of the elastic energy under the prescribed periodicity and boundary conditions (see (1-6) below).

In the physically relevant case of a highly anisotropic nonconvex interfacial energy, there may exist certain directions $\nu$ at which the ellipticity condition

$$D^2\psi(\nu)[\tau,\tau] > 0 \quad \text{for all } \tau \perp \nu,\ \tau \neq 0$$

fails; see for instance [Di Carlo et al. 1992; Siegel et al. 2004]. Correspondingly, the above evolution equation becomes *backward parabolic* and thus ill-posed. To overcome this ill-posedness, and following the work of Herring [1951], an additive curvature regularization to surface energy has been proposed; see [Di Carlo et al. 1992; Gurtin and Jabbour 2002]. Here we consider the regularized surface energy

$$\int_{\Gamma_h} \Big(\psi(\nu) + \frac{\varepsilon}{p}|H|^p\Big)\,d\mathcal{H}^2,$$

where $p > 2$, $H$ stands for the sum $\kappa_1 + \kappa_2$ of the principal curvatures of $\Gamma_h$, and $\varepsilon$ is a (small) positive constant. The restriction on the range of exponents $p > 2$ is of technical nature and it is motivated by the fact that, in two dimensions, the Sobolev space $W^{2,p}$ embeds into $C^{1,(p-2)/p}$ if $p > 2$. The extension of our analysis to the case $p = 2$ seems to require different ideas.

The regularized free energy functional then reads

$$\int_{\Omega_h} W(E(u))\,dz + \int_{\Gamma_h} \Big(\psi(\nu) + \frac{\varepsilon}{p}|H|^p\Big)\,d\mathcal{H}^2, \tag{1-4}$$

and (1-1) becomes

$$V = \Delta_\Gamma\Big[\mathrm{Div}_\Gamma(D\psi(\nu)) + W(E(u)) - \varepsilon\Big(\Delta_\Gamma(|H|^{p-2}H) - |H|^{p-2}H\Big(\kappa_1^2 + \kappa_2^2 - \frac{1}{p}H^2\Big)\Big)\Big]. \tag{1-5}$$

Sixth-order evolution equations of this type have already been considered in [Gurtin and Jabbour 2002] for the case without elasticity. Its two-dimensional version was studied numerically in [Siegel et al. 2004] for the evolution of voids in elastically stressed materials, and analytically in [Fonseca et al. 2012] in the context of evolving one-dimensional graphs. We also refer to [Rätz et al. 2006; Burger et al. 2007] and references therein for some numerical results in the three-dimensional case. However, to the best of our knowledge no analytical results were available in the literature prior to ours.

As in [Fonseca et al. 2012], in this paper we focus on evolving graphs, and to be precise on the case where (1-5) models the evolution toward equilibrium of epitaxially strained elastic films deposited over a rigid substrate. Given $Q := (0, b)^2$, $b > 0$, we look for a spatially $Q$-periodic solution to the Cauchy problem

$$\begin{cases} \dfrac{1}{J}\dfrac{\partial h}{\partial t} = \Delta_\Gamma\big[\mathrm{Div}_\Gamma(D\psi(\nu)) + W(E(u)) - \varepsilon\big(\Delta_\Gamma(|H|^{p-2}H) - |H|^{p-2}H\big(\kappa_1^2+\kappa_2^2-\frac{1}{p}H^2\big)\big)\big] \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{in } \mathbb{R}^2\times(0,T_0), \\ \mathrm{Div}\,\mathbb{C}E(u) = 0 \quad \text{in } \Omega_h, \\ \mathbb{C}E(u)[\nu] = 0 \quad \text{on } \Gamma_h, \quad u(x,0,t) = (e_0^1 x_1, e_0^2 x_2, 0), \\ h(\cdot,t) \text{ and } Du(\cdot,t) \quad \text{are } Q\text{-periodic}, \\ h(\cdot,0) = h_0, \end{cases} \tag{1-6}$$

where, we recall, $h : \mathbb{R}^2 \times [0, T_0] \to (0, +\infty)$ denotes the function describing the two-dimensional profile $\Gamma_h$ of the film;

$$J := \sqrt{1 + |D_x h|^2};$$

$W(A) := \frac{1}{2}\mathbb{C}A : A$ for all $A \in \mathbb{M}^{2\times 2}_{\mathrm{sym}}$ with $\mathbb{C}$ a positive definite fourth-order tensor; $e_0 := (e_0^1, e_0^2)$, with $e_0^1, e_0^2 > 0$, is a vector that embodies the mismatch between the crystalline lattices of the film and the substrate; and $h_0 \in H^2_{\mathrm{loc}}(\mathbb{R}^2)$ is a $Q$-periodic function. Note that, in (1-6), the sixth-order (geometric) parabolic equation for the film profile is coupled with the elliptic system of elastic equilibrium equations in the bulk.

It was observed by Cahn and Taylor [1994] that the surface diffusion equation can be regarded as a gradient flow of the free energy functional with respect to a suitable $H^{-1}$-Riemannian structure. To formally illustrate this point, consider the manifold of subsets of $Q \times (0, +\infty)$ of fixed volume $d$, which are subgraphs of a $Q$-periodic function, that is,

$$\mathcal{M} := \left\{\Omega_h : h \;\; Q\text{-periodic},\; h \in H^2(Q),\; \int_Q h\,dx = d\right\},$$

where $\Omega_h := \{(x, y) : x \in Q,\ 0 < y < h(x)\}$. The tangent space $T_{\Omega_h}\mathcal{M}$ at an element $\Omega_h$ is described by the kinematically admissible normal velocities:

$$T_{\Omega_h}\mathcal{M} := \left\{V : \Gamma_h \to \mathbb{R} : V \text{ is } Q\text{-periodic},\; V \in L^2(\Gamma_h; \mathcal{H}^2),\; \int_{\Gamma_h} V\,d\mathcal{H}^2 = 0\right\},$$

where $\Gamma_h$ is the graph of $h$ over the periodicity cell $Q$; it is endowed with the $H^{-1}$ metric tensor

$$g_{\Omega_h}(V_1, V_2) := \int_{\Gamma_h} \nabla_{\Gamma_h} w_1 \nabla_{\Gamma_h} w_2\, d\mathcal{H}^2 \quad \text{for all } V_1,\ V_2 \in T_{\Omega_h}\mathcal{M},$$

where $w_i$, $i = 1, 2$, is the solution to

$$\begin{cases} -\Delta_{\Gamma_h} w_i = V_i \ \text{ on } \ \Gamma_h, \\ w_i \ \text{ is } Q\text{-periodic}, \\ \int_{\Gamma_h} w_i \, d\mathcal{H}^2 = 0. \end{cases}$$

Consider now the *reduced free energy functional*

$$G(\Omega_h) := \int_{\Omega_h} W(E(u_h))\, dz + \int_{\Gamma_h} \Big( \psi(\nu) + \frac{\varepsilon}{p}|H|^p \Big)\, d\mathcal{H}^2,$$

where $u_h$ is the minimizer of the elastic energy in $\Omega_h$ under the boundary and periodicity conditions described above. Then, the evolution described by (1-6) is such that at each time the normal velocity $V$ of the evolving profile $h(t)$ is the element of the tangent space $T_{\Omega_{h(t)}}\mathcal{M}$ corresponding to the steepest descent of $G$, i.e., (1-6) may be formally rewritten as

$$g_{\Omega_{h(t)}}(V, \widetilde{V}) = -\partial G(\Omega_{h(t)})[\widetilde{V}] \quad \text{for all } \ \widetilde{V} \in T_{\Omega_{h(t)}}\mathcal{M},$$

where $\partial G(h(t))[\widetilde{V}]$ stands for the first variation of $G$ at $\Omega_{h(t)}$ in the direction $\widetilde{V}$.

In order to solve (1-6), we take advantage of this gradient flow structure and we implement a *minimizing movements scheme* (see [Ambrosio 1995]), which consists in constructing discrete time evolutions by solving iteratively suitable minimum incremental problems.

It is interesting to observe that the gradient flow of the free energy functional $G$ with respect to an $L^2$-Riemannian structure (instead of $H^{-1}$) leads to a fourth-order evolution equation, which describes motion by evaporation–condensation (see [Cahn and Taylor 1994; Gurtin and Jabbour 2002] and [Piovano 2014], where the one-dimensional case was studied analytically).

This paper is organized as follows. In Section 2 we set up the problem and introduce the discrete time evolutions. In Section 3 we prove our main local-in-time existence result for (1-6), by showing that (up to subsequences) the discrete time evolutions converge to a weak solution of (1-6) in $[0, T_0]$ for some $T_0 > 0$ (see Theorem 3.16). By a *Q-periodic weak solution* we mean a function $h \in H^1(0, T_0; H^{-1}_\#(Q)) \cap L^\infty(0, T_0; H^2_\#(Q))$ such that $(h, u_h)$ satisfies the system (1-6) in the distributional sense (see Definition 3.1). To the best of our knowledge, Theorem 3.16 is the first (short time) existence result for a surface diffusion-type geometric evolution equation in the presence of elasticity in three dimensions. Moreover, the use of minimizing movements also appears to be new in the context of higher-order geometric flows (the only other paper we are aware of in which a similar approach is adopted, but in two dimensions, is [Fonseca et al. 2012]).

Compared to mean curvature flows, where the minimizing movements algorithm is nowadays classical after the pioneering work of [Almgren et al. 1993] (see also [Chambolle 2004; Bellettini et al. 2006; Caselles and Chambolle 2006]), a major technical difference lies in the fact that no comparison principle is available in this higher-order framework. The convergence analysis is instead based on subtle interpolation and regularity estimates. It is worth mentioning that, for geometric surface diffusion equation without elasticity and without curvature regularization,

$$V = \Delta_\Gamma H$$

(corresponding to the case $W = 0$, $\psi = 1$, and $\varepsilon = 0$), short time existence of a smooth solution was proved in [Escher et al. 1998] using semigroup techniques. See also [Bellettini et al. 2007; Mantegazza 2002]. It is still an open question whether the solution constructed via the minimizing movement scheme is unique, and thus independent of the subsequence.

In Section 4 we address the Liapunov stability of the flat configuration, corresponding to a horizontal (flat) profile. Roughly speaking, we show that if the surface energy density is strictly convex and the second variation of the functional (1-2) at a given flat configuration is positive definite, then such a configuration is asymptotically stable, that is, for all initial data $h_0$ sufficiently close to it the corresponding evolutions constructed via minimizing movements exist for all times, and converge asymptotically to the flat configuration as $t \to +\infty$ (see Theorem 4.8). We remark that Theorem 4.8 may be regarded as an evolutionary counterpart of the static stability analysis of the flat configuration performed in [Fusco and Morini 2012; Bonacini 2013a; 2013b]. In Theorem 4.7 we address also the case of a nonconvex anisotropy, and we show that, if the corresponding Wulff shape contains a horizontal facet, then the Asaro–Grinfeld–Tiller instability does not occur and the flat configuration is *always* Liapunov stable (see [Bonacini 2013a; 2013b] for the corresponding result in the static case). Both results are completely new even in the two-dimensional case, to which they obviously apply (see Section 4C). We remark that our treatment is purely variational and it is hinged on the fact that (1-4) is a Liapunov functional for the evolution.

Finally, in the Appendix, we collect several auxiliary results that are used throughout the paper.

## 2. Setting of the problem

Let $Q := (0, b)^2 \subset \mathbb{R}^2$, $b > 0$, $p > 2$, and let $h_0 \in W^{2,p}_\#(Q)$ be a positive function, describing the initial profile of the film. We recall that $W^{2,p}_\#(Q)$ stands for the subspace of $W^{2,p}(Q)$ of all functions whose $Q$-periodic extension belong to $W^{2,p}_{\mathrm{loc}}(\mathbb{R}^2)$. Given $h \in W^{2,p}_\#(Q)$, with $h \geq 0$, we set

$$\Omega_h := \{(x, y) \in Q \times \mathbb{R} : 0 < y < h(x)\}$$

and we denote by $\Gamma_h$ the graph of $h$ over $Q$. We will identify a function $h \in W^{2,p}_\#(Q)$ with its periodic extension to $\mathbb{R}^2$, and denote by $\Omega^\#_h$ and $\Gamma^\#_h$ the open subgraph and the graph of this extension, respectively. Note that $\Omega^\#_h$ is the periodic extension of $\Omega_h$. Set

$$LD_\#(\Omega_h; \mathbb{R}^3) := \big\{u \in L^2_{\mathrm{loc}}(\Omega^\#_h; \mathbb{R}^3) : u(x, y) = u(x+bk, y) \text{ for } (x, y) \in \Omega^\#_h \text{ and } k \in \mathbb{Z}^2,\ E(u)|_{\Omega_h} \in L^2(\Omega_h; \mathbb{R}^3)\big\},$$

where $E(u) := \frac{1}{2}(Du + D^T u)$, with $Du$ the distributional gradient of $u$ and $D^T u$ its transpose, is the strain of the displacement $u$. We prescribe the Dirichlet boundary condition $u(x, 0) = w_0(x, 0)$ for $x \in Q$, with $w_0 \in H^1(U \times (0, +\infty))$ for every bounded open subset $U \subset \mathbb{R}^2$ and such that $Dw_0(\cdot\,, y)$ is $Q$-periodic for a.e. $y > 0$. A typical choice is given by $w_0(x, y) := (e^1_0 x_1, e^2_0 x_2, 0)$, where the vector

$e_0 := (e_0^1, e_0^2)$, with $e_0^1$, $e_0^2 > 0$, embodies the mismatch between the crystalline lattices of film and substrate. Define

$$X := \{(h,u) : h \in W^{2,p}_{\#}(Q),\ h \geq 0,\ u : \Omega_h^{\#} \to \mathbb{R}^3,\ u - w_0 \in LD_{\#}(\Omega_h; \mathbb{R}^3),\ u(x,0) = w_0 \text{ for all } x \in \mathbb{R}^2\}.$$

The elastic energy density $W : \mathbb{M}^{3\times 3}_{\mathrm{sym}} \to [0, +\infty)$ takes the form

$$W(A) := \tfrac{1}{2}\mathbb{C}A : A,$$

with $\mathbb{C}$ a positive definite fourth-order tensor, so that $W(A) > 0$ for all $A \in \mathbb{M}^{3\times 3}_{\mathrm{sym}} \setminus \{0\}$. Given $h \in W^{2,p}_{\#}(Q)$, $h \geq 0$, we denote by $u_h$ the corresponding elastic equilibrium in $\Omega_h$, i.e.,

$$u_h := \operatorname{argmin}\left\{\int_{\Omega_h} W(E(u))\,dz : u \in w_0 + LD_{\#}(\Omega_h; \mathbb{R}^3),\ u(x,0) = w_0(x,0)\right\}.$$

Let $\psi : \mathbb{R}^3 \to [0, +\infty)$ be a positively one-homogeneous function of class $C^2$ away from the origin. Note that, in particular,

$$\frac{1}{c}|\xi| \leq \psi(\xi) \leq c|\xi| \quad \text{for all } \xi \in \mathbb{R}^3 \tag{2-1}$$

for some $c > 0$.

We now introduce the energy functional

$$F(h,u) := \int_{\Omega_h} W(E(u))\,dz + \int_{\Gamma_h} \left(\psi(\nu) + \frac{\varepsilon}{p}|H|^p\right) d\mathcal{H}^2, \tag{2-2}$$

defined for all $(h,u) \in X$, where $\nu$ is the outer unit normal to $\Omega_h$, $H = \mathrm{Div}_{\Gamma_h} \nu$ denotes the sum of the principal curvatures of $\Gamma_h$, and $\varepsilon$ is a positive constant. In the sequel we will often use the fact that

$$-\mathrm{Div}\left(\frac{Dh}{\sqrt{1+|Dh|^2}}\right) = H \quad \text{in } Q, \tag{2-3}$$

which, in turn, implies

$$\int_Q H\,dx = 0. \tag{2-4}$$

**Remark 2.1** (notation). In the sequel we denote by $z$ a generic point in $Q \times \mathbb{R}$ and we write $z = (x,y)$ with $x \in Q$ and $y \in \mathbb{R}$. Moreover, given $g : \Gamma_h \to \mathbb{R}$, where $\Gamma_h$ is the graph of some function $h$ defined in $Q$, we denote by the same symbol $g$ the function from $Q$ to $\mathbb{R}$ given by $x \mapsto g(x, h(x))$. Consistently, $Dg$ will stand for the gradient of the function from $Q$ to $\mathbb{R}$ just defined.

**2A.** ***The incremental minimum problem.*** In this subsection we introduce the incremental minimum problems that will be used to define the discrete time evolutions. As a standing assumption throughout this paper, we start from an initial configuration $(h_0, u_0) \in X$ such that

$$h_0 \in W^{2,p}_{\#}(Q), \quad h_0 > 0, \tag{2-5}$$

and $u_0$ minimizes the elastic energy in $\Omega_{h_0}$ among all $u$ with $(h_0, u) \in X$.

Fix a sequence $\tau_n \searrow 0$ representing the discrete time increments. For $i \in \mathbb{N}$ we define inductively $(h_{i,n}, u_{i,n})$ as a solution of the minimum problem

$$\min\left\{ F(h,u) + \frac{1}{2\tau_n}\int_{\Gamma_{i-1,n}} |D_{\Gamma_{i-1,n}} v_h|^2 \, d\mathcal{H}^2 : (h,u) \in X,\ \|Dh\|_{L^\infty(Q)} \le \Lambda_0,\ \int_Q h\,dx = \int_Q h_0\,dx \right\}, \tag{2-6}$$

where $\Gamma_{i-1,n}$ stands for $\Gamma_{h_{i-1,n}}$, $\Lambda_0$ is a positive constant such that

$$\Lambda_0 > \|h_0\|_{C^1_\#(Q)}, \tag{2-7}$$

and $v_h$ is the unique solution in $H^1_\#(\Gamma_{h_{i-1,n}})$ to the following problem:

$$\begin{cases} \Delta_{\Gamma_{i-1,n}} v_h = \dfrac{(h - h_{i-1,n})}{\sqrt{1+|Dh_{i-1,n}|^2}} \circ \pi, \\ \int_{\Gamma_{h_{i-1,n}}} v_h \, d\mathcal{H}^2 = 0, \end{cases} \tag{2-8}$$

where $\pi$ is the canonical projection $\pi(x,y) = x$. We note that the formulation of the problem in (2-6) with the upper bound $\Lambda_0$ is usually adopted in the literature in order to ensure existence of solutions of the minimal surface equation (see Chapter 12 in [Giusti 1984]).

For $x \in Q$ and $(i-1)\tau_n \le t \le i\tau_n$, $i \in \mathbb{N}$, we define the linear interpolation

$$h_n(x,t) := h_{i-1,n}(x) + \frac{1}{\tau_n}(t - (i-1)\tau_n)(h_{i,n}(x) - h_{i-1,n}(x)), \tag{2-9}$$

and we let $u_n(\cdot,t)$ be the *elastic equilibrium corresponding to* $h_n(\cdot,t)$, i.e.,

$$F(h_n(\cdot,t), u_n(\cdot,t)) = \min_{(h_n(\cdot,t),u) \in X} F(h_n(\cdot,t), u)\,. \tag{2-10}$$

The remainder of this subsection is devoted to the proof of the existence of a minimizer for the minimum incremental problem (2-6).

**Theorem 2.2.** *The minimum problem* (2-6) *admits a solution* $(h_{i,n}, u_{i,n}) \in X$.

*Proof.* Let $\{(h_k, u_k)\} \subset X$ be a minimizing sequence for (2-6). Let $H_k$ denote the sum of principal curvatures of $\Gamma_{h_k}$. Since the sequence $\{H_k\}$ is bounded in $L^p(Q)$ and $\|Dh_k\|_{L^\infty_\#(Q)} \le \Lambda_0$, it follows from (2-3) and Lemma A.3 that $\|h_k\|_{W^{2,p}_\#(Q)} \le C$. Then, up to a subsequence (not relabelled), we may assume that $h_k \rightharpoonup h$ weakly in $W^{2,p}_\#(Q)$, and thus strongly in $C^{1,\alpha}_\#(Q)$ for some $\alpha > 0$. As a consequence, $H_k \rightharpoonup H$ in $L^p(Q)$, where $H$ is the sum of the principal curvatures of $\Gamma_h$. In turn, the $L^p$-weak convergence of $\{H_k\}$ and the $C^1$-convergence of $\{h_k\}$ imply by lower semicontinuity that

$$\int_{\Gamma_h} \left(\psi(\nu) + \frac{\varepsilon}{p}|H|^p\right) d\mathcal{H}^2 \le \liminf_k \int_{\Gamma_{h_k}} \left(\psi(\nu) + \frac{\varepsilon}{p}|H_k|^p\right) d\mathcal{H}^2. \tag{2-11}$$

Moreover, we also have that $v_{h_k} \to v_h$ strongly in $H^1(\Gamma_{i-1,n})$, and thus

$$\lim_k \frac{1}{2\tau_n}\int_{\Gamma_{i-1,n}} |D_{\Gamma_{i-1,n}} v_{h_k}|^2 \, d\mathcal{H}^2 = \frac{1}{2\tau_n}\int_{\Gamma_{i-1,n}} |D_{\Gamma_{i-1,n}} v_h|^2 \, d\mathcal{H}^2. \tag{2-12}$$

Finally, since $\sup_k \int_{\Omega_{h_k}} |Eu_k|^2\,dz < +\infty$, reasoning as in [Fonseca et al. 2007, Proposition 2.2], from the uniform convergence of $\{h_k\}$ to $h$ and Korn's inequality we conclude that there exists $u \in H^1_{\mathrm{loc}}(\Omega^\#_h;\mathbb{R}^3)$ such that $(h,u) \in X$ and, up to a subsequence, $u_k \rightharpoonup u$ weakly in $H^1_{\mathrm{loc}}(\Omega^\#_h;\mathbb{R}^3)$. Therefore, we have that

$$\int_{\Omega_h} W(E(u))\,dz \le \liminf_k \int_{\Omega_{h_k}} W(E(u_k))\,dz,$$

which, together with (2-11) and (2-12), allows us to conclude that $(h,u)$ is a minimizer. □

## 3. Existence of the evolution

In this section we prove short time existence of a solution of the geometric evolution equation

$$V = \Delta_\Gamma\Big[\mathrm{Div}_\Gamma(D\psi(\nu)) + W(E(u)) - \varepsilon\Big(\Delta_\Gamma(|H|^{p-2}H) - \frac{1}{p}|H|^pH + |H|^{p-2}H|B|^2\Big)\Big], \tag{3-1}$$

where $V$ denotes the outer normal velocity of $\Gamma_{h(\cdot,t)}$, $|B|^2$ is the sum of the squares of the principal curvatures of $\Gamma_{h(\cdot,t)}$, $u(\cdot,t)$ is the elastic equilibrium in $\Omega_{h(\cdot,t)}$, and $W(E(u))$ is the trace of $W\big(E(u(\cdot,t))\big)$ on $\Gamma_{h(\cdot,t)}$. In the sequel, we denote by $H^{-1}_\#(Q)$ the dual space of $H^1_\#(Q)$. Note that, if $f \in H^1_\#(Q)$, then $\Delta f$ can be identified with the element of $H^{-1}_\#(Q)$ defined by

$$\langle \Delta f, g\rangle := -\int_Q Df\,Dg\,dx \quad \text{for all } g \in H^1_\#(Q).$$

Moreover, a function $f \in L^2(Q)$ can be identified with the element of $H^{-1}_\#(Q)$ defined by

$$\langle f, g\rangle := \int_Q fg\,dx \quad \text{for all } g \in H^1_\#(Q).$$

**Definition 3.1.** Let $T_0 > 0$. We say that $h \in L^\infty(0,T_0;W^{2,p}_\#(Q)) \cap H^1(0,T_0;H^{-1}_\#(Q))$ is a *solution of* (3-1) in $[0,T_0]$ if:

(i) $\mathrm{Div}_\Gamma(D\psi(\nu)) + W(E(u)) - \varepsilon\Big(\Delta_\Gamma(|H|^{p-2}H) - \frac{1}{p}|H|^pH + |H|^{p-2}H|B|^2\Big) \in L^2(0,T_0;H^1_\#(Q))$,

(ii) for a.e. $t \in (0,T_0)$, in $H^{-1}_\#(Q)$ we have

$$\frac{1}{J}\frac{\partial h}{\partial t} = \Delta_\Gamma\Big[\mathrm{Div}_\Gamma(D\psi(\nu)) + W(E(u)) - \varepsilon\Big(\Delta_\Gamma(|H|^{p-2}H) - \frac{1}{p}|H|^pH + |H|^{p-2}H|B|^2\Big)\Big],$$

where $J := \sqrt{1+|Dh|^2}$, $u(\cdot,t)$ is the elastic equilibrium in $\Omega_{h(\cdot,t)}$, and where we wrote $\Gamma$ in place of $\Gamma_{h(\cdot,t)}$.

**Remark 3.2.** An immediate consequence of the above definition is that the evolution is *volume-preserving*, that is, $\int_Q h(x,t)\,dx = \int_Q h_0(x)\,dx$ for all $t \in [0,T_0]$. Indeed, for all $t_1, t_2 \in [0,T_0]$ and for $\varphi \in H^1_\#(Q)$, we have

$$\begin{aligned}
&\int_Q [h(x,t_2)-h(x,t_1)]\varphi\,dx \\
&\quad = \int_{t_1}^{t_2} \Big\langle \frac{\partial h}{\partial t}(\,\cdot\,,t),\varphi \Big\rangle\,dt \\
&\quad = \int_{t_1}^{t_2} \Big\langle J\Delta_\Gamma\Big[\mathrm{Div}_\Gamma(D\psi(\nu)) + W(E(u)) - \varepsilon\Big(\Delta_\Gamma(|H|^{p-2}H) - \frac{1}{p}|H|^p H + |H|^{p-2}H|B|^2\Big)\Big],\varphi\Big\rangle\,dt \\
&\quad = -\int_{t_1}^{t_2}\int_\Gamma D_\Gamma\Big[\mathrm{Div}_\Gamma(D\psi(\nu)) + W(E(u)) \\
&\qquad\qquad - \varepsilon\Big(\Delta_\Gamma(|H|^{p-2}H) - \frac{1}{p}|H|^p H + |H|^{p-2}H|B|^2\Big)\Big] D_\Gamma(\varphi\circ\pi)\,d\mathcal{H}^2\,dt.
\end{aligned}$$

Choosing $\varphi = 1$, we conclude that

$$\int_Q [h(x,t_2)-h(x,t_1)]\,dx = 0.$$

**Remark 3.3.** In the sequel, we consider the following equivalent norm on $H^{-1}_{\#}(Q)$. Given $\mu \in H^{-1}_{\#}(Q)$, we set

$$\|\mu\|_{H^{-1}_{\#}(Q)} := \sup\Big\{\langle\mu,g\rangle : g\in H^1_{\#}(Q),\ \Big|\int_Q g\,dx\Big| + \|Dg\|_{L^2(Q)} \le 1\Big\}.$$

Note that, if $f\in L^2(Q)$ with $\int_Q f\,dx = 0$, we have

$$\|f\|_{H^{-1}_{\#}(Q)} = \|Dw\|_{L^2(Q)},$$

where $w\in H^1_{\#}(Q)$ is the unique periodic solution to the problem

$$\begin{cases} \Delta w = f \ \text{ in } \ Q, \\ \int_Q w\,dx = 0. \end{cases} \tag{3-2}$$

To see this, first observe that, since $\int_Q f\,dx = 0$, we have

$$\|f\|_{H^{-1}_{\#}(Q)} = \sup\Big\{\int_Q fg\,dx : g\in H^1_{\#}(Q),\ \int_Q g\,dx = 0 \ \text{ and } \ \|Dg\|_{L^2(Q)}\le 1\Big\}.$$

Thus, since by (3-2)

$$\int_Q fg\,dx = -\int_Q DwDg\,dx \le \|Dw\|_{L^2(Q)},$$

we have $\|f\|_{H^{-1}_{\#}(Q)} \le \|Dw\|_{L^2(Q)}$. The opposite inequality follows by taking $g = -w/\|Dw\|_{L^2(Q)}$.

**Theorem 3.4.** *For all $n, i \in \mathbb{N}$, we have*

$$\int_0^{+\infty} \left\| \frac{\partial h_n}{\partial t} \right\|^2_{H^{-1}_\#(Q)} dt \le CF(h_0, u_0), \tag{3-3}$$

$$F(h_{i,n}, u_{i,n}) \le F(h_{i-1,n}, u_{i-1,n}) \le F(h_0, u_0), \tag{3-4}$$

$$\text{and} \quad \sup_{t\in[0,+\infty)} \|h_n(\,\cdot\,, t)\|_{W^{2,p}_\#(Q)} < +\infty \tag{3-5}$$

*for some $C = C(\Lambda_0) > 0$. Moreover, up to a subsequence,*

$$h_n \to h \text{ in } C^{0,\alpha}([0, T]; L^2(Q)) \text{ for all } \alpha \in \left(0, \tfrac14\right), \quad h_n \rightharpoonup h \text{ weakly in } H^1(0, T; H^{-1}_\#(Q)) \tag{3-6}$$

*for all $T > 0$ and for some function $h$ such that $h(\,\cdot\,, t) \in W^{2,p}_\#(Q)$ for every $t \in [0, +\infty)$ and*

$$F(h(\,\cdot\,, t), u_{h(\cdot,t)}) \le F(h_0, u_0) \quad \text{for all } t \in [0, +\infty). \tag{3-7}$$

*Proof.* By the minimality of $(h_{i,n}, u_{i,n})$ (see (2-6)) we have that

$$F(h_{i,n}, u_{i,n}) + \frac{1}{2\tau_n} \int_{\Gamma_{i-1,n}} |D_{\Gamma_{i-1,n}} v_{h_{i,n}}|^2 \, d\mathcal{H}^2 \le F(h_{i-1,n}, u_{i-1,n}) \tag{3-8}$$

for all $i \in \mathbb{N}$, which yields in particular (3-4). Hence,

$$\frac{1}{2\tau_n} \int_{\Gamma_{i-1,n}} |D_{\Gamma_{i-1,n}} v_{h_{i,n}}|^2 \, d\mathcal{H}^2 \le F(h_{i-1,n}, u_{i-1,n}) - F(h_{i,n}, u_{i,n})$$

and, summing over $i$, we obtain

$$\sum_{i=1}^{\infty} \frac{1}{2\tau_n} \int_{\Gamma_{i-1,n}} |D_{\Gamma_{i-1,n}} v_{h_{i,n}}|^2 \, d\mathcal{H}^2 \le F(h_0, u_0). \tag{3-9}$$

Let $w_{h_{i,n}} \in H^1_\#(Q)$ denote the unique periodic solution to the problem

$$\begin{cases} \Delta w_{h_{i,n}} = h_{i,n} - h_{i-1,n} \text{ in } Q, \\ \int_Q w_{h_{i,n}} \, dx = 0. \end{cases}$$

Note that

$$\begin{aligned}
\int_Q |Dw_{h_{i,n}}|^2 \, dx &= \int_Q \Delta w_{h_{i,n}} w_{h_{i,n}} \, dx = \int_{\Gamma_{i-1,n}} \frac{h_{i,n} - h_{i-1,n}}{\sqrt{1 + |Dh_{i-1,n}|^2}} \circ \pi \, w_{h_{i,n}} \, d\mathcal{H}^2 \\
&= \int_{\Gamma_{i-1,n}} \Delta_{\Gamma_{i-1,n}} v_{h_{i,n}} w_{h_{i,n}} \, d\mathcal{H}^2 = -\int_{\Gamma_{i-1,n}} D_{\Gamma_{i-1,n}} v_{h_{i,n}} D_{\Gamma_{i-1,n}} w_{h_{i,n}} \, d\mathcal{H}^2 \\
&\le \|D_{\Gamma_{i-1,n}} v_{h_{i,n}}\|_{L^2(\Gamma_{i-1,n})} \|D_{\Gamma_{i-1,n}} w_{h_{i,n}}\|_{L^2(\Gamma_{i-1,n})} \\
&\le C(\Lambda_0) \|D_{\Gamma_{i-1,n}} v_{h_{i,n}}\|_{L^2(\Gamma_{i-1,n})} \|Dw_{h_{i,n}}\|_{L^2(Q)}.
\end{aligned}$$

Combining this inequality with (3-9) and recalling (2-9) and Remark 3.3, we get (3-3).

Note from (3-4) it follows that

$$\sup_{i,n} \int_{\Gamma_{i,n}} |H|^p \, d\mathcal{H}^2 < +\infty.$$

Hence, (3-5) follows immediately by Lemma A.3, taking into account that $\|Dh_{i,n}\|_{L^\infty(Q)} \le \Lambda_0$. Using a diagonalizing argument, it can be shown that there exist $h$ such that $h_n \rightharpoonup h$ weakly in $H^1(0,T;H^{-1}_\#(Q))$ for all $T > 0$. Note also that, by (3-3) and using Hölder's inequality, we have for $t_2 > t_1$ that

$$\|h_n(\cdot,t_2) - h_n(\cdot,t_1)\|_{H^{-1}(Q)} \le \int_{t_1}^{t_2} \left\| \frac{\partial h_n(\cdot,t)}{\partial t} \right\|_{H^{-1}(Q)} dt \le C(t_2-t_1)^{\frac{1}{2}}. \tag{3-10}$$

Therefore, applying Theorem A.4 to the solution $w \in H^1_\#(Q)$ of the problem

$$\begin{cases} \Delta w = h_n(\cdot,t_2) - h_n(\cdot,t_1) \text{ in } Q, \\ \int_Q w\,dx = 0, \end{cases}$$

we get

$$\begin{aligned} \|h_n(\cdot,t_2) - h_n(\cdot,t_1)\|_{L^2(Q)} &= \|\Delta w\|_{L^2(Q)} \le C\|D^3 w\|^{\frac{1}{2}}_{L^2(Q)} \|Dw\|^{\frac{1}{2}}_{L^2(Q)} \\ &\le C\|Dh(\cdot,t_2) - Dh(\cdot,t_1)\|^{\frac{1}{2}}_{L^2(Q)} \|h(\cdot,t_2) - h(\cdot,t_1)\|^{\frac{1}{2}}_{H^{-1}(Q)} \\ &\le C(\Lambda_0)(t_2-t_1)^{\frac{1}{4}}, \end{aligned} \tag{3-11}$$

where the last inequality follows from (3-10). By the Ascoli–Arzelà theorem (see, e.g., [Ambrosio et al. 2008, Proposition 3.3.1]), we get (3-6). Finally, inequality (3-7) follows from (3-4) by lower semicontinuity, using (3-6) and (3-5). □

In what follows, $\{h_n\}$ and $h$ are the subsequence and the function found in Theorem 3.4, respectively. The next result shows that the convergence of $\{h_n\}$ to $h$ can be significantly improved for short time.

**Theorem 3.5.** *There exist $T_0 > 0$ and $C > 0$ depending only $(h_0, u_0)$ such that*:

(i) $$h_n \to h \quad \textit{in } C^{0,\beta}([0,T_0];C^{1,\alpha}_\#(Q))$$

*for every $\alpha \in (0, p-2/p)$ and $\beta \in (0,(p-2-\alpha p)(p+2)/(16p^2))$;*

(ii) $$\sup_{t\in[0,T_0]} \|Du_n(\cdot,t)\|_{C^{0,\frac{p-2}{p}}(\overline{\Omega}_{h_n(\cdot,t)})} \le C;$$

(iii) $$E(u_n(\cdot,h_n)) \to E(u(\cdot,h)) \quad \textit{in } C^{0,\beta}([0,T_0];C^{0,\alpha}_\#(Q))$$

*for every $\alpha \in (0,(p-2)/p)$ and $0 \le \beta < (p-2-\alpha p)(p+2)/(16p^2)$, where $u(\cdot,t)$ is the elastic equilibrium in $\Omega_{h(\cdot,t)}$.*

*Moreover, $h(\cdot,t) \to h_0$ in $C^{1,\alpha}_\#(Q)$ as $t \to 0^+$, $h_n, h \ge C_0 > 0$ for some positive constant $C_0$, and*

$$\sup_{t\in[0,T_0]} \|Dh_n(\cdot,t)\|_{L^\infty(Q)} < \Lambda_0 \tag{3-12}$$

*for all $n$.*

*Proof.* To prove assertion (i), we start by observing that, by Theorem A.6, (3-5), Theorem A.6 again, and (3-11) we have

$$\begin{aligned}\|Dh_n(\cdot,t_2)-Dh_n(\cdot,t_1)\|_{L^\infty} &\le C\|D^2h_n(\cdot,t_2)-D^2h_n(\cdot,t_1)\|_{L^p}^{\frac{p+2}{2p}}\|h_n(\cdot,t_2)-h_n(\cdot,t_1)\|_{L^p}^{\frac{p-2}{2p}}\\ &\le C\|h_n(\cdot,t_2)-h_n(\cdot,t_1)\|_{L^p}^{\frac{p-2}{2p}}\\ &\le C\Big(\|D^2h_n(\cdot,t_2)-D^2h_n(\cdot,t_1)\|_{L^2}^{\frac{p-2}{2p}}\|h_n(\cdot,t_2)-h_n(\cdot,t_1)\|_{L^2}^{\frac{p+2}{2p}}\Big)^{\frac{p-2}{2p}}\\ &\le C|t_2-t_1|^{\frac{p^2-4}{16p^2}} \qquad (3\text{-}13)\end{aligned}$$

for all $t_1$, $t_2\in[0,T_0]$. Notice that from (3-5) we have

$$\sup_{n,t\in[0,T_0]}\|h_n(\cdot,t)\|_{C_\#^{1,\frac{p-2}{p}}(Q)}<+\infty. \qquad (3\text{-}14)$$

Take $\alpha\in(0,(p-2)/p)$ and observe that

$$[Dh_n(\cdot,t_2)-Dh_n(\cdot,t_1)]_\alpha\le[Dh_n(\cdot,t_2)-Dh_n(\cdot,t_1)]_{\frac{p-2}{p}}^{\frac{\alpha p}{p-2}}\big[\operatorname{osc}_{[0,b]}(Dh_n(\cdot,t_2)-Dh_n(\cdot,t_1))\big]^{\frac{p-2-\alpha p}{p-2}},$$

where $[\cdot]_\beta$ denotes the $\beta$-Hölder seminorm. From this inequality, (3-13), (3-14), and the Ascoli–Arzelà theorem [Ambrosio et al. 2008, Proposition 3.3.1], assertion (i) follows.

Standard elliptic estimates ensure that, if $h_n(\cdot,t)\in C_\#^{1,\alpha}(Q)$ for some $\alpha\in(0,1)$, then $Du_n(\cdot,t)$ can be estimated in $C^{0,\alpha}(\overline{\Omega}_{h_n(\cdot,t)})$ with a constant depending only on the $C^{1,\alpha}$-norm of $h_n(\cdot,t)$; see for instance [Fusco and Morini 2012, Proposition 8.9], where this property is proved in two dimensions but an entirely similar argument works in all dimensions. Hence, assertion (ii) follows from (3-14). Assertion (iii) is an immediate consequence of (i) and Lemma A.1. Finally, (3-12) follows from (2-7) and (i). ☐

**Remark 3.6.** Note that in the previous theorem we can take

$$T_0:=\sup\{t>0:\|Dh_n(\cdot,s)\|_{L^\infty(Q)}<\Lambda_0\ \text{ for all }\ s\in[0,t)\}.$$

In Theorem 3.16 we will show that $h$ is a solution to (3-1) in $[0,T_0)$, in the sense of Definition 3.1.

We begin with some auxiliary results.

**Proposition 3.7.** *Let $h\in W_\#^{3,q}(Q)$ for some $q>2$ and let $\Gamma$ be its graph. Let $\Phi:Q\times\mathbb{R}\times(-1,1)\to Q\times\mathbb{R}$ be the flow*

$$\frac{\partial\Phi}{\partial t}=X(\Phi),\quad \Phi(\cdot,0)=\mathrm{Id},$$

*where $X$ is a smooth vector field $Q$-periodic in the first two variables. Set $\Gamma_t:=\Phi(\cdot,t)(\Gamma)$, denote by $\nu_t$ the normal to $\Gamma_t$, let $H_t$ be the sum of principal curvatures of $\Gamma_t$, and let $|B_t|^2$ be the sum of squares of* *the principal curvatures of* $\Gamma_t$*. Then*

$$\frac{d}{dt}\frac{1}{p}\int_{\Gamma_t}|H_t|^p\,d\mathcal{H}^2 \\ = \int_{\Gamma_t} D_{\Gamma_t}(|H_t|^{p-2}H_t)D_{\Gamma_t}(X\cdot\nu_t)\,d\mathcal{H}^2 - \int_{\Gamma_t}|H_t|^{p-2}H_t\Big(|B_t|^2-\frac{1}{p}H_t^2\Big)(X\cdot\nu_t)\,d\mathcal{H}^2. \quad (3\text{-}15)$$

*Proof.* Set $\Phi_t(\cdot) := \Phi(\cdot\,,t)$. We can extend $\nu_t$ to a tubular neighborhood of $\Gamma_t$ as the gradient of the signed distance from $\Gamma_t$. We have

$$\frac{d}{dt}\frac{1}{p}\int_{\Gamma_t}|H_t|^p\,d\mathcal{H}^2 = \frac{d}{ds}\bigg(\frac{1}{p}\int_{\Gamma_{t+s}}|H_{t+s}|^p\,d\mathcal{H}^2\bigg)\bigg|_{s=0} = \frac{d}{ds}\bigg(\frac{1}{p}\int_{\Gamma_t}|H_{t+s}\circ\Phi_s|^p J_2\Phi_s\,d\mathcal{H}^2\bigg)\bigg|_{s=0},$$

where $J_2$ denotes the two-dimensional Jacobian of $\Phi_s$ on $\Gamma_t$. Then we have

$$\frac{d}{dt}\frac{1}{p}\int_{\Gamma_t}|H_t|^p\,d\mathcal{H}^2 = \frac{1}{p}\int_{\Gamma_t}|H_t|^p\,\mathrm{Div}_{\Gamma_t}X\,d\mathcal{H}^2 + \int_{\Gamma_t}|H_t|^{p-2}H_t\frac{d}{ds}(H_{t+s}\circ\Phi_s)\Big|_{s=0}\,d\mathcal{H}^2.$$

Concerning the last integral, we observe that

$$\frac{d}{ds}(H_{t+s}\circ\Phi_s)\Big|_{s=0} = \frac{d}{ds}(\mathrm{Div}_{\Gamma_{t+s}}\nu_{t+s})\Big|_{s=0} + DH_t\cdot X.$$

Set

$$\dot\nu_t := \frac{d}{ds}\nu_{t+s}\Big|_{s=0}.$$

By differentiating with respect to $s$ the identity $D\nu_{t+s}[\nu_{t+s}]=0$, we get

$$D\dot\nu_t[\nu_t] + D\nu_t[\dot\nu_t] = 0.$$

Multiplying this identity by $\nu_t$ and recalling that $D\nu$ is a symmetric matrix, we get

$$D\dot\nu_t[\nu_t]\cdot\nu_t = -D\nu_t[\nu_t]\cdot\dot\nu_t = 0.$$

This implies that $\mathrm{Div}_{\Gamma_t}\dot\nu_t = \mathrm{Div}\,\dot\nu_t$, and so

$$\frac{d}{ds}(\mathrm{Div}_{\Gamma_{t+s}}\nu_{t+s})\Big|_{s=0} = \mathrm{Div}_{\Gamma_t}\dot\nu_t.$$

In turn — see [Cagnetti et al. 2008, Lemma 3.8(f)] —

$$\dot\nu_t = -(D_{\Gamma_t}X)^T[\nu_t] - D_{\Gamma_t}\nu_t[X] = -D_{\Gamma_t}(X\cdot\nu_t).$$

Collecting the above identities, integrating by parts, and using the identity $\partial_{\nu_t} H_t = -\operatorname{trace}((D\nu_t)^2) = -|B_t|^2$ proved in [Cagnetti et al. 2008, Lemma 3.8(d)], we have

$$\begin{aligned}
&\frac{d}{dt}\frac{1}{p}\int_{\Gamma_t}|H_t|^p\,d\mathcal{H}^2\\
&\quad=\frac{1}{p}\int_{\Gamma_t}|H_t|^p\operatorname{Div}_{\Gamma_t}X\,d\mathcal{H}^2+\int_{\Gamma_t}|H_t|^{p-2}H_t(-\Delta_{\Gamma_t}(X\cdot\nu_t)+DH_t\cdot X)\,d\mathcal{H}^2\\
&\quad=-\int_{\Gamma_t}|H_t|^{p-2}H_tD_{\Gamma_t}H_t\cdot X\,d\mathcal{H}^2+\frac{1}{p}\int_{\Gamma_t}|H_t|^pH_t(X\cdot\nu_t)\,d\mathcal{H}^2\\
&\qquad\qquad+\int_{\Gamma_t}|H_t|^{p-2}H_t(-\Delta_{\Gamma_t}(X\cdot\nu_t)+DH_t\cdot X)\,d\mathcal{H}^2\\
&\quad=\int_{\Gamma_t}|H_t|^{p-2}H_t\Big(-\Delta_{\Gamma_t}(X\cdot\nu_t)+\partial_{\nu_t}H_t(X\cdot\nu_t)+\frac{1}{p}H_t^2(X\cdot\nu_t)\Big)\,d\mathcal{H}^2\\
&\quad=\int_{\Gamma_t}D_{\Gamma_t}(|H_t|^{p-2}H_t)D_{\Gamma_t}(X\cdot\nu_t)\,d\mathcal{H}^2-\int_{\Gamma_t}|H_t|^{p-2}H_t\Big\{\Big(|B_t|^2-\frac{1}{p}H_t^2\Big)(X\cdot\nu_t)\Big\}\,d\mathcal{H}^2. \qquad (3\text{-}16)
\end{aligned}$$

Thus (3-15) follows. □

Proposition 3.7 motivates the following definition:

**Definition 3.8.** We say that $(h, u_h) \in X$ is a *critical pair* for the functional $F$ defined in (2-2) if $|H|^{p-2}H \in H^1(\Gamma_h)$ and

$$\begin{aligned}
\varepsilon\int_{\Gamma_h}D_{\Gamma_h}(|H|^{p-2}H)D_{\Gamma_h}\phi\,d\mathcal{H}^2+\varepsilon\int_{\Gamma_h}\Big(\frac{1}{p}|H|^pH-|H|^{p-2}H|B|^2\Big)\phi\,d\mathcal{H}^2&\\
+\int_{\Gamma_h}\big[\operatorname{Div}_{\Gamma_h}(D\psi(\nu))+W(E(u_h))\big]\phi\,d\mathcal{H}^2&=0
\end{aligned}$$

for all $\phi \in H^1_\#(\Gamma_h)$ with $\int_{\Gamma_h}\phi\,d\mathcal{H}^2 = 0$. We will also say that $h$ is a *critical profile* if $(h, u_h)$ is a critical pair.

**Lemma 3.9.** *Let $h \in W^{2,p}_\#(Q)$ be such that $|H|^{p-2}H \in W^{1,q}_\#(Q)$ for some $q > 2$. Then there exists a sequence $\{h_j\} \subset W^{3,q}_\#(Q)$ such that $h_j \to h$ in $W^{2,p}_\#(Q)$ and $|H_j|^{p-2}H_j \to |H|^{p-2}H$ in $W^{1,q}_\#(Q)$, where $H_j$ stands for the sum of the principal curvatures of $\Gamma_{h_j}$.*

*Proof.* We may assume without loss of generality that $H \neq 0$, otherwise $h$ would have already the required regularity (see (2-3)). By the Sobolev embedding theorem it follows that $|H|^{p-2}H \in C^{0,1-2/q}_\#(Q)$ and, in turn, using the $1/(p-1)$ Hölder continuity of the function $t \mapsto t^{1/(p-1)}$, $H \in C^{0,\alpha}_\#(Q)$ for $\alpha := (q-2)/(q(p-1))$. Standard Schauder estimates yield $h \in C^{2,\alpha}_\#(Q)$.

For $\delta > 0$ set

$$H_\delta := \begin{cases} H-\delta & \text{if } H \geq \delta,\\ H+\delta' & \text{if } H \leq -\delta',\\ 0 & \text{otherwise,}\end{cases}$$

where $\delta'$ is chosen in such a way that $\int_Q H_\delta\,dx = 0$. Observe that this choice of $\delta'$ is always possible, although not necessarily unique. Indeed, by (2-4) and the fact that $H \neq 0$, if $\delta$ is sufficiently small

$$\int_{\{H>\delta\}} (H-\delta)\,dx + \int_{\{H<0\}} H\,dx < 0 \quad \text{and} \quad \int_{\{H>\delta\}} (H-\delta)\,dx > 0.$$

By continuity it is then clear that we may find $\delta' > 0$ such that

$$\int_{\{H>\delta\}} (H-\delta)\,dx + \int_{\{H<-\delta'\}} (H+\delta')\,dx = 0. \tag{3-17}$$

We now show that, independently of the choice of $\delta'$ satisfying (3-17), $\delta' \to 0$ as $\delta \to 0$. Indeed, if not, there would exist a sequence $\delta_n \to 0$ and a corresponding sequence $\delta'_n \to \delta' > 0$ such that (3-17) holds with $\delta$ and $\delta'$ replaced by $\delta_n$ and $\delta'_n$, respectively. But then, passing to the limit as $n \to \infty$, we would get

$$\int_{\{H>0\}} H\,dx + \int_{\{H<-\delta'\}} (H+\delta')\,dx = 0,$$

which contradicts (2-4).

Note that $H_\delta \to H$ in $C^{0,\beta}_\#(Q)$ for all $\beta < \alpha$ as $\delta \to 0$. Moreover, we claim that $|H_\delta|^{p-2}H_\delta \to |H|^{p-2}H$ in $W^{1,q}_\#(Q)$. Indeed, observe that $H \in W^{1,q}(A_\delta)$ where $A_\delta := \{H > \delta\} \cup \{H < -\delta'\}$ for all $\delta > 0$. Hence,

$$D(|H|^{p-2}H) = \begin{cases} (p-1)|H|^{p-2}DH & \text{if } H \neq 0, \\ 0 & \text{elsewhere,} \end{cases}$$

and

$$D(|H_\delta|^{p-2}H_\delta) = \begin{cases} (p-1)|H_\delta|^{p-2}DH & \text{in } A_\delta, \\ 0 & \text{elsewhere.} \end{cases}$$

The claim follows by observing that $D(|H_\delta|^{p-2}H_\delta) \to D(|H|^{p-2}H)$ a.e. and that $|D(|H_\delta|^{p-2}H_\delta)| \leq |D(|H|^{p-2}H)|$. Note now that $H \in W^{1,q}(A_\delta)$ implies $H_\delta \in W^{1,q}_\#(Q)$. In order to conclude the proof it is enough to show that for $\delta$ sufficiently small there exists a unique periodic solution $h_\delta$ to the problem

$$\begin{cases} -\operatorname{Div}(Dh_\delta/\sqrt{1+|Dh_\delta|^2}) = H_\delta \\ \int_Q h_\delta\,dx = \int_Q h\,dx. \end{cases} \tag{3-18}$$

This follows from Lemma 3.10 below. □

**Lemma 3.10.** *Let $h \in C^{2,\alpha}_\#(Q)$ and let $H$ be the sum of the principal curvatures of $\Gamma_h$. Then there exist $\sigma, C > 0$ with the following property: for all $K \in C^{0,\alpha}_\#(Q)$ with $\int_Q K\,dx = 0$ and $\|K-H\|_{C^{0,\alpha}_\#(Q)} \leq \sigma$, there exists a unique periodic solution $k \in C^{2,\alpha}_\#(Q)$ to*

$$\begin{cases} -\operatorname{Div}(Dk/\sqrt{1+|Dk|^2}) = K \\ \int_Q k\,dx = \int_Q h\,dx, \end{cases}$$

*and*

$$\|k-h\|_{C^{2,\alpha}_\#(Q)} \leq C\|K-H\|_{C^{0,\alpha}_\#(Q)}. \tag{3-19}$$

*Proof.* Without loss of generality we may assume that $\int_Q h\,dx = 0$.

Set $X := \{k \in C^{2,\alpha}_\#(Q) : \int_Q k\, dx = 0\}$ and $Y := \{K \in C^{0,\alpha}_\#(Q) : \int_Q K\, dx = 0\}$, and consider the operator $T\colon X \to Y$ defined by

$$T(k) := -\operatorname{Div}\biggl(\frac{Dk}{\sqrt{1+|Dk|^2}}\biggr).$$

By assumption we have that $T(h) = H$. We now use the inverse function theorem (see, e.g., [Ambrosetti and Prodi 1993, Chapter 2, Theorem 1.2]) to prove that $T$ is invertible in a $C^{2,\alpha}$-neighborhood of $h$ with a $C^1$-inverse. To see this, note that for any $k \in X$ we have that $T'(k) : X \to Y$ is the continuous linear operator defined by

$$T'(h)[\varphi] := -\operatorname{Div}\biggl[\frac{1}{\sqrt{1+|Dh|^2}}\biggl(I - \frac{Dh\otimes Dh}{1+|Dh|^2}\biggr)D\varphi\biggr].$$

It is easily checked that $T'$ is a continuous map from $X$ to the space $\mathscr{L}(X, Y)$ of linear bounded operators from $X$ to $Y$, so that $T \in C^1(X, Y)$. Finally, standard existence arguments for elliptic equations imply that for any $k \in X$ the operator $T'(k)$ is invertible. Thus we may apply the inverse function theorem to conclude that there exist $\sigma > 0$ such that, for all $K \in C^{0,\alpha}_\#(Q)$ with $\int_Q K\, dx = 0$ and $\|K - H\|_{C^{0,\alpha}_\#(Q)} \le \sigma$, there exists a unique periodic function $k = T^{-1}K \in C^{2,\alpha}_\#(Q)$. Moreover, the continuity of $T^{-1}$, together with standard Schauder estimates, implies that (3-19) holds for $\sigma$ sufficiently small. ☐

In what follows, $J_{i,n}$ stands for

$$J_{i,n} := \sqrt{1+|Dh_{i,n}|^2},$$

$H_{i,n}$ is the sum of the principal curvatures of $\Gamma_{i,n}$, $|B_{i,n}|^2$ denotes the sum of the squares of the principal curvatures of $\Gamma_{i,n}$, and $\widetilde{H}_n\colon Q\times[0, T_0] \to \mathbb{R}$ is the function defined as

$$\widetilde{H}_n(x,t) := H_{i,n}(x, h_{i,n}(x), t) \quad \text{if } t \in [(i-1)\tau_n, i\tau_n). \tag{3-20}$$

**Theorem 3.11.** *Let $T_0$ be as in Theorem 3.5 and let $\widetilde{H}_n$ be given in* (3-20). *Then there exists $C > 0$ such that*

$$\int_0^{T_0}\int_Q |D^2(|\widetilde{H}_n|^{p-2}\widetilde{H}_n)|^2\, dx\, dt \le C \tag{3-21}$$

*for $n \in \mathbb{N}$.*

*Proof.* **Step 1.** We claim that $|H_{i,n}|^{p-2}H_{i,n} \in W^{1,q}_\#(\Gamma_{i,n})$ for all $q \ge 1$ and that $h_{i,n} \in C^{2,\sigma}_\#(Q)$ for all $\sigma \in (0, 1/(p-1))$.

We recall that $h_{i,n}$ is the solution to the incremental minimum problem (2-6). We are going to show that $h_{i,n} \in W^{2,q}_\#(Q)$ for all $q \ge 2$. Fix a function $\varphi \in C^2_\#(Q)$ such that $\int_Q \varphi\, dx = 0$. Then, by minimality and by (3-12), we have

$$\frac{d}{ds}\biggl(F(h_{i,n}+s\varphi, u_{i,n}) + \frac{1}{2\tau_n}\int_{\Gamma_{i-1,n}} |D_{\Gamma_{i-1,n}} v_{h_{i,n}+s\varphi}|^2\, d\mathscr{H}^2\biggr)\bigg|_{s=0} = 0,$$

where, we recall, $v_{h_{i,n}+s\varphi}$ solves (2-8) with $h$ replaced by $h_{i,n}+s\varphi$. It can be shown that

$$\begin{aligned}\int_Q W\big(E\big(u_{i,n}(x,h_{i,n}(x))\big)\big)\varphi\,dx+\int_Q D\psi(-Dh_{i,n},1)\cdot(-D\varphi,0)\,dx+\frac{\varepsilon}{p}\int_Q|H_{i,n}|^p\frac{Dh_{i,n}\cdot D\varphi}{J_{i,n}}\,dx\\ -\varepsilon\int_Q|H_{i,n}|^{p-2}H_{i,n}\bigg[\Delta\varphi-\frac{D^2\varphi[Dh_{i,n},Dh_{i,n}]}{J_{i,n}^2}-\frac{\Delta h_{i,n}Dh_{i,n}\cdot D\varphi}{J_{i,n}^2}\\ -2\frac{D^2h_{i,n}[Dh_{i,n},D\varphi]}{J_{i,n}^2}+3\frac{D^2h_{i,n}[Dh_{i,n},Dh_{i,n}]Dh_{i,n}\cdot D\varphi}{J_{i,n}^4}\bigg]\,dx\\ -\frac{1}{\tau_n}\int_Q v_{h_{i,n}}\varphi\,dx=0,\end{aligned}\tag{3-22}$$

where the last integral is obtained by observing that $v_{h_{i,n}+s\varphi}=v_{h_{i,n}}+sv_\varphi$, with $v_\varphi$ solving

$$\begin{cases}\Delta_{\Gamma_{i-1,n}}v_\varphi=\dfrac{\varphi}{\sqrt{1+|Dh_{i-1,n}|^2}}\circ\pi,\\ \int_{\Gamma_{h_{i-1,n}}}v_\varphi\,d\mathcal{H}^2=0.\end{cases}$$

Setting $w:=|H_{i,n}|^{p-2}H_{i,n}$,

$$\begin{aligned}A&:=\varepsilon\bigg(I-\frac{Dh_{i,n}\otimes Dh_{i,n}}{J_{i,n}^2}\bigg),\\ b&:=\pi(D\psi(-Dh_{i,n},1))-\frac{\varepsilon}{p}|H_{i,n}|^p\frac{Dh_{i,n}}{J_{i,n}}\\ &\quad+\varepsilon w\bigg[-\frac{\Delta h_{i,n}Dh_{i,n}}{J_{i,n}^2}-2\frac{D^2h_{i,n}[Dh_{i,n}]}{J_{i,n}^2}+3\frac{D^2h_{i,n}[Dh_{i,n},Dh_{i,n}]Dh_{i,n}}{J_{i,n}^4}\bigg],\\ c&:=-W\big(E\big(u(x,h_{i,n}(x))\big)\big)+\frac{1}{\tau_n}v_{h_{i,n}},\end{aligned}\tag{3-23}$$

we have by (3-5) and Theorem 3.5 that $A\in W^{1,p}_\#(Q;\mathbb{M}^{2\times2}_{\mathrm{sym}})$, $b\in L^1(Q;\mathbb{R}^2)$, $c\in C^{0,\alpha}_\#(Q)$ for some $\alpha$, and we may rewrite (3-22) as

$$\int_Q w\,AD^2\varphi\,dx+\int_Q b\cdot D\varphi+\int_Q c\varphi\,dx=0\quad\text{for all }\varphi\in C^\infty_\#(Q)\text{ with }\int_Q\varphi\,dx=0.\tag{3-24}$$

By Lemma A.2 we get that $w\in L^q(Q)$ for $q\in(p/(p-1),2)$. Therefore, for any such $q$ we have $H_{i,n}\in L^{q(p-1)}(Q)$ and thus, by Lemma A.3, $h_{i,n}\in W^{2,q(p-1)}_\#(Q)$. In turn, using Hölder's inequality, this implies that $b$, $w\,\mathrm{Div}\,A\in L^{r_0}(Q;\mathbb{R}^2)$, where $r_0:=q(p-1)/p$. Observe that $r_0\in(1,2)$. By applying Lemma A.2 again, we deduce that $w\in W^{1,r_0}_\#(Q)$ and thus $w\in L^{2r_0/(2-r_0)}(Q)$. In turn, arguing as before, this implies that $b$, $w\,\mathrm{Div}\,A\in L^{r_1}(Q;\mathbb{R}^2)$, where $r_1:=2r_0(p-1)/((2-r_0)p)>r_0$. If $r_1\ge2$, then using again Lemma A.2 we conclude that $w\in W^{1,r_1}_\#(Q)$, which implies the claim, since $D^2h_{i,n}\in L^q(Q;\mathbb{M}^{2\times2}_{\mathrm{sym}})$ and, in turn, $b$, $w\,\mathrm{Div}\,A\in L^q(Q;\mathbb{R}^2)$ for all $q$. Then the conclusion follows by Lemma A.2. Otherwise, we proceed by induction. Assume that $w\in W^{1,r_{i-1}}_\#(Q)$. If $r_{i-1}\ge2$ then the claim follows. If not, a further application of Lemma A.2 implies that $w\in W^{1,r_i}_\#(Q)$ with $r_i:=2r_{i-1}(p-1)/((2-r_{i-1})p)$.

Since $r_{i-1} < 2$, we have $r_i > r_{i-1}$. We claim that there exists $j$ such that $r_j > 2$. Indeed, if not, the increasing sequence $\{r_i\}$ would converge to some $\ell \in (1, 2]$ satisfying

$$\ell = \frac{2\ell(p-1)}{(2-\ell)p}.$$

However, this is impossible since the above identity is equivalent to $\ell = 2/p < 1$.

Finally, observe that, since $|H_{i,n}|^{p-2}H_{i,n} \in W^{1,q}_{\#}(Q)$ for all $q \geq 1$, we have $|H_{i,n}|^{p-1} \in C^{0,\alpha}_{\#}(Q)$ for every $\alpha \in (0, 1)$. Hence $H_{i,n} \in C^{0,\sigma}_{\#}(Q)$ for all $\sigma \in (0, 1/(p-1))$ and so, by standard Schauder estimates, $h_{i,n} \in C^{2,\sigma}_{\#}(Q)$ for all $\sigma \in (0, 1/(p-1))$.

**Step 2.** By Step 1 we may now write the Euler–Lagrange equation for $h_{i,n}$ in intrinsic form. To be precise, we claim that, for all $\varphi \in C^2_{\#}(Q)$ with $\int_Q \varphi\, dx = 0$, we have

$$\varepsilon \int_{\Gamma_{i,n}} D_{\Gamma_{i,n}}(|H_{i,n}|^{p-2}H_{i,n}) D_{\Gamma_{i,n}}\phi \, d\mathcal{H}^2 - \varepsilon \int_{\Gamma_{i,n}} |H_{i,n}|^{p-2}H_{i,n}\Big(|B_{i,n}|^2 - \frac{1}{p}H^2_{i,n}\Big)\phi \, d\mathcal{H}^2 + \int_{\Gamma_{i,n}} \big[\mathrm{Div}_{\Gamma_{i,n}}(D\psi(\nu_{i,n})) + W(E(u_{i,n}))\big]\phi \, d\mathcal{H}^2 - \frac{1}{\tau_n}\int_{\Gamma_{i,n}} v_{h_{i,n}}\phi \, d\mathcal{H}^2 = 0, \quad (3\text{-}25)$$

where $\phi := (\varphi/J_{i,n}) \circ \pi$. To see this, fix $h \in W^{3,q}_{\#}(Q)$ for some $q > 2$, denote by $\Gamma$ and $\Gamma_t$ the graphs of $h$ and $h + t\varphi$, respectively, and by $H$ and $H_t$ the corresponding sums of the principal curvatures. Then, by Proposition 3.7 and arguing as in the proof of (3-22), we have

$$\begin{aligned}
&\int_\Gamma D_\Gamma(|H|^{p-2}H) D_\Gamma \phi \, d\mathcal{H}^2 - \int_\Gamma |H|^{p-2}H\Big(|B|^2 - \frac{1}{p}H^2\Big)\phi \, d\mathcal{H}^2 \\
&= \frac{1}{p}\int_Q |H|^p \frac{Dh \cdot D\varphi}{J}\, dx \\
&\quad - \int_Q |H|^{p-2}H\Big[\Delta\varphi - \frac{D^2\varphi[Dh, Dh]}{J^2} - \frac{\Delta h\, Dh \cdot D\varphi}{J^2} - 2\frac{D^2h[Dh, D\varphi]}{J^2} + 3\frac{D^2h[Dh, Dh]Dh \cdot D\varphi}{J^4}\Big]\, dx,
\end{aligned}$$

where $\phi$ stands for $(\varphi/J) \circ \pi$ and $J := \sqrt{1 + |Dh|^2}$. By the approximation Lemma 3.9, this identity still holds if $h \in C^{2,\alpha}_{\#}(Q)$ and thus (3-25) follows from (3-22), recalling that, by Step 1, $h_{i,n} \in C^{2,\sigma}_{\#}(Q)$ for some $\sigma > 0$.

In order to show (3-21), observe that Lemma A.3, together with the bound $\|Dh_{i,n}\|_{L^\infty} \leq \Lambda_0$, implies that

$$\|D^2h_{i,n}\|_{L^q(Q)} \leq C(q, \Lambda_0)\|H_{i,n}\|_{L^q(Q)}. \quad (3\text{-}26)$$

Moreover, since $\Gamma_{i,n}$ is of class $C^{2,\sigma}$, (3-25) yields that $|H_{i,n}|^{p-2}H_{i,n} \in H^2(\Gamma_{i,n})$, and in turn that $|H_{i,n}|^{p-2}H_{i,n} \in H^2(Q)$ (see Remark 2.1).

As before, setting $w := |H_{i,n}|^{p-2}H_{i,n}$, by approximation we may rewrite (3-25) as

$$\int_Q A(x) Dw D\Big(\frac{\varphi}{J_{i,n}}\Big) J_{i,n}\, dx - \varepsilon \int_Q w\varphi\Big(|B_{i,n}|^2 - \frac{1}{p}H^2_{i,n}\Big)\, dx + \int_Q \big[\mathrm{Div}_{\Gamma_{i,n}}(D\psi(\nu_{i,n})) + W(E(u_{i,n}))\big]\varphi\, dx - \frac{1}{\tau_n}\int_Q v_{h_{i,n}}\varphi\, dx = 0 \quad (3\text{-}27)$$

for all $\varphi \in H^1_\#(Q)$ with $\int_Q \varphi\, dx = 0$, where $A$, defined as in (3-23), is an elliptic matrix with ellipticity constants depending only on $\Lambda_0$. Recall that $w \in H^2(Q)$. We now choose $\varphi = D_s\eta$ with $\eta \in H^2_\#(Q)$, and observe that integrating by parts twice yields

$$\begin{aligned}\int_Q & ADwD\Big(\frac{D_s\eta}{J_{i,n}}\Big)J_{i,n}\,dx\\ &= -\int_Q AD(D_sw)D\Big(\frac{\eta}{J_{i,n}}\Big)J_{i,n}\,dx - \int_Q D_s(AJ_{i,n})DwD\Big(\frac{\eta}{J_{i,n}}\Big)\,dx + \int_Q ADwD\Big(\frac{\eta D_sJ_{i,n}}{J^2_{i,n}}\Big)J_{i,n}\,dx\\ &= -\int_Q AD(D_sw)D\Big(\frac{\eta}{J_{i,n}}\Big)J_{i,n}\,dx - \int_Q D_s(AJ_{i,n})DwD\Big(\frac{\eta}{J_{i,n}}\Big)\,dx\\ &\qquad\qquad - \int_Q AD^2w\frac{\eta D_sJ_{i,n}}{J_{i,n}}\,dx - \int_Q D(AJ_{i,n})Dw\frac{\eta D_sJ_{i,n}}{J^2_{i,n}}\,dx.\end{aligned}$$

Therefore, recalling (3-27) and by density, we may conclude that, for every $\eta \in H^1_\#(Q)$,

$$\begin{aligned}\int_Q & AD(D_sw)D\Big(\frac{\eta}{J_{i,n}}\Big)J_{i,n}\,dx\\ &= -\int_Q D_s(AJ_{i,n})DwD\Big(\frac{\eta}{J_{i,n}}\Big)\,dx - \int_Q AD^2w\frac{\eta D_sJ_{i,n}}{J_{i,n}}\,dx\\ &\quad - \int_Q D(AJ_{i,n})Dw\frac{\eta D_sJ_{i,n}}{J^2_{i,n}}\,dx - \varepsilon\int_Q wD_s\eta\Big(|B_{i,n}|^2 - \frac{1}{p}H^2_{i,n}\Big)\,dx\\ &\quad + \int_Q\big[\mathrm{Div}_{\Gamma_{i,n}}(D\psi(\nu_{i,n})) + W(E(u_{i,n}))\big]D_s\eta\,dx - \frac{1}{\tau_n}\int_Q v_{h_{i,n}}D_s\eta\,dx.\end{aligned}$$

Finally, choosing $\eta = D_swJ_{i,n}$, we obtain

$$\begin{aligned}\int_Q & AD(D_sw)D(D_sw)J_{i,n}\,dx\\ &= -\int_Q D_s(AJ_{i,n})DwD(D_sw)\,dx - \int_Q AD^2wD_swD_sJ_{i,n}\,dx - \int_Q D(AJ_{i,n})Dw\frac{D_swD_sJ_{i,n}}{J_{i,n}}\,dx\\ &\quad - \varepsilon\int_Q wD_s(D_swJ_{i,n})\Big(|B_{i,n}|^2 - \frac{1}{p}H^2_{i,n}\Big)\,dx\\ &\quad + \int_Q\big[\mathrm{Div}_{\Gamma_{i,n}}(D\psi(\nu_{i,n})) + W(E(u_{i,n}))\big]D_s(D_swJ_{i,n})\,dx - \frac{1}{\tau_n}\int_Q v_{h_{i,n}}D_s(D_swJ_{i,n})\,dx.\end{aligned}$$

Summing the resulting equations for $s = 1, 2$, estimating $D(AJ_{i,n})$ by $D^2h_{i,n}$, and using Young's inequality several times, we deduce

$$\begin{aligned}\int_Q & |D^2w|^2\,dx\\ &\le C\int_Q\Big(|Dw|^2|D^2h_{i,n}|^2\,dx + |H_{i,n}|^{2p+2} + |H_{i,n}|^{2p-2}|D^2h_{i,n}|^4 + \frac{v^2_{i,n}}{(\tau_n)^2} + 1\Big)\,dx \qquad (3\text{-}28)\end{aligned}$$

for some constant $C$ depending only on $\Lambda_0$, $D^2\psi$, and on the $C^{1,\alpha}$ bounds on $u_{i,n}$ provided by Theorem 3.5. Note that, by Young's inequality and (3-26), we have

$$\int_Q |H_{i,n}|^{2p-2}|D^2h_{i,n}|^4\,dx \le C\int_Q (|H_{i,n}|^{2p+2}+|D^2h_{i,n}|^{2p+2})\,dx \le C\int_Q |H_{i,n}|^{2p+2}\,dx.$$

Combining the last estimate with (3-28), we therefore have

$$\int_Q |D^2w|^2\,dx \le C_0\int_Q\Big(|D^2h_{i,n}|^2|Dw|^2+|w|^{\frac{2(p+1)}{p-1}}+\frac{v_{i,n}^2}{(\tau_n)^2}+1\Big)\,dx. \tag{3-29}$$

To deal with the first term on the right-hand side, we use Hölder's inequality, (3-26) and Theorem A.6 twice to get

$$\begin{aligned} C_0\int_Q |D^2h_{i,n}|^2|Dw|^2\,dx &\le C_0\Big(\int_Q |D^2h|^{2(p-1)}\,dx\Big)^{\frac{1}{p-1}}\Big(\int_Q |Dw|^{\frac{2(p-1)}{p-2}}\,dx\Big)^{\frac{p-2}{p-1}}\\ &\le C\|w\|_2^{\frac{2}{p-1}}\|Dw\|^2_{\frac{2(p-1)}{p-2}} \le C\|w\|_2^{\frac{2}{p-1}}\Big(\|D^2w\|_2^{\frac{p}{2(p-1)}}\|w\|_2^{\frac{p-2}{2(p-1)}}\Big)^2\\ &= C\|D^2w\|_2^{\frac{p}{p-1}}\|w\|_2^{\frac{p}{p-1}} \le C\|D^2w\|_2^{\frac{p}{p-1}}\Big(\|D^2w\|_{\frac{p}{p-1}}^{\frac{p-2}{2p}}\|w\|_{\frac{p}{p-1}}^{\frac{p+2}{2p}}\Big)^{\frac{p}{p-1}}\\ &\le C\|D^2w\|_2^{\frac{3p-2}{2(p-1)}}\|w\|_{\frac{p}{p-1}}^{\frac{p+2}{2(p-1)}} \le \tfrac14\|D^2w\|_2^2+C, \end{aligned}$$

where in the last inequality we used the fact that $(3p-2)/(2(p-1))<2$ and that $\|w\|_{\frac{p}{p-1}}=\|H_{i,n}\|_p^{p-1}$ is uniformly bounded with respect to $i$, $n$. Using Theorem A.6 again, we also have

$$C_0\int_Q |w|^{\frac{2(p+1)}{p-1}}\,dx \le C\|D^2w\|_{\frac{p}{p-1}}^{\frac{p+2}{p}}\|w\|_{\frac{p}{p-1}}^{\frac{p^2+p+2}{p(p-1)}} \le \tfrac14\|D^2w\|_2^2+C,$$

where, as before, we used the fact that $(p+2)/p<2$ and $\|w\|_{\frac{p}{p-1}}$ is uniformly bounded. Inserting the two estimates above in (3-29), we then get

$$\int_Q |D^2w|^2\,dx \le C\int_Q\Big(1+\frac{v_{i,n}^2}{(\tau_n)^2}\Big)\,dx. \tag{3-30}$$

Integrating the last inequality with respect to time and using (3-9) we conclude the proof of the theorem. □

**Remark 3.12.** The same argument used in Step 1 of the proof of Theorem 3.11 and in the proof of (3-25) shows that, if $(h,u_h)\in X$ satisfies

$$\begin{aligned} \int_Q W\big(E\big(u_h(x,h(x))\big)\big)\varphi\,dx + \int_Q D\psi(-Dh,1)\cdot(-D\varphi,0)\,dx + \frac{\varepsilon}{p}\int_Q |H|^p\frac{Dh\cdot D\varphi}{J}\\ -\varepsilon\int_Q |H|^{p-2}H\Big[\Delta\varphi-\frac{D^2\varphi[Dh,Dh]}{J^2}-\frac{\Delta h\,Dh\cdot D\varphi}{J^2}-2\frac{D^2h[Dh,D\varphi]}{J^2}\\ +3\frac{D^2h[Dh,Dh]Dh\cdot D\varphi}{J^4}\Big]\,dx=0 \end{aligned}$$

for all $\varphi\in C^2_\#(Q)$ such that $\int_Q\varphi\,dx=0$, then $(h,u_h)$ is a critical pair for the functional $F$.

**Lemma 3.13.** *With $T_0$ and $\widetilde{H}_n$ as in Theorem 3.11, $|\widetilde{H}_n|^p$ is a Cauchy sequence in $L^1(0,T_0;L^1(Q))$. Moreover, $|\widetilde{H}_n|^{p-2}\widetilde{H}_n$ is a Cauchy sequence in $L^1(0,T_0;L^2(Q))$.*

For the proof of the lemma we need the following inequality:

**Lemma 3.14.** *Let $p>1$. There exists $c_p>0$ such that*

$$\frac{1}{c_p}(x^{p-1}+y^{p-1})\le\frac{|x^p-y^p|}{|x-y|}\le c_p(x^{p-1}+y^{p-1}).$$

*Proof.* By homogeneity it is enough to assume $y=1$ and $x>1$ and to observe that

$$\lim_{x\to+\infty}\frac{x^p-1}{(x-1)(x^{p-1}+1)}=1\quad\text{and}\quad\lim_{x\to 1}\frac{x^p-1}{(x-1)(x^{p-1}+1)}=\frac{p}{2}.\qquad\square$$

*Proof of Lemma 3.13.* We split the proof into two steps.

**Step 1.** We start by showing that $|\widetilde{H}_n|^p$ is a Cauchy sequence in $L^1(0,T_0;L^1(Q))$. Set $k:=[p]$, where $[\cdot]$ denotes the integer part. Note that $k\ge 2$ since $p>2$. From Lemma 3.14 we get

$$\begin{aligned}
&\int_0^{T_0}\int_Q\big||\widetilde{H}_n|^p-|\widetilde{H}_m|^p\big|\,dx\,dt\\
&\qquad=\int_0^{T_0}\int_Q\big||\widetilde{H}_n^k|^{\frac{p}{k}}-|\widetilde{H}_m^k|^{\frac{p}{k}}\big|\,dx\,dt\\
&\qquad\le c\int_0^{T_0}\int_Q\big||\widetilde{H}_n^k|-|\widetilde{H}_m^k|\big|(|\widetilde{H}_n|^k+|\widetilde{H}_m|^k)^{\frac{p}{k}-1}\,dx\,dt\\
&\qquad\le c\int_0^{T_0}\Big(\int_Q|\widetilde{H}_n^k-\widetilde{H}_m^k|^2\,dx\Big)^{\frac12}(\|\widetilde{H}_n\|_\infty+\|\widetilde{H}_m\|_\infty)^{p-k}\,dt\\
&\qquad\le c\int_0^{T_0}(\|\widetilde{H}_n^k-\widetilde{H}_m^k-M_{m,n}\|_2+|M_{m,n}|)(\|\widetilde{H}_n\|_\infty+\|\widetilde{H}_m\|_\infty)^{p-k}\,dt,
\end{aligned}\tag{3-31}$$

where $M_{m,n}:=\int_Q(\widetilde{H}_n^k-\widetilde{H}_m^k)\,dx$. Set

$$w_n:=|\widetilde{H}_n|^{p-2}\widetilde{H}_n\tag{3-32}$$

and observe that $\widetilde{H}_n^k=(w_n^+)^{\frac{k}{p-1}}+(-1)^k(w_n^-)^{\frac{k}{p-1}}$. Thus,

$$|D\widetilde{H}_n^k|\le|D(w_n^+)^{\frac{k}{p-1}}|+|D(w_n^-)^{\frac{k}{p-1}}|\le c|Dw_n||w_n|^{\frac{k}{p-1}-1}=c|Dw_n||\widetilde{H}_n|^{k-p+1}.\tag{3-33}$$

From Lemma A.7 and inequalities (3-31), (3-33) we get

$$\begin{aligned}
&\int_0^{T_0}\int_Q\big||\widetilde{H}_n|^p-|\widetilde{H}_m|^p\big|\,dx\,dt\\
&\quad\le c\int_0^{T_0}\big(\|\widetilde{H}_n^k-\widetilde{H}_m^k-M_{m,n}\|_{H^{-1}}^{\frac12}\|D\widetilde{H}_n^k-D\widetilde{H}_m^k\|_2^{\frac12}+|M_{m,n}|\big)(\|\widetilde{H}_n\|_\infty+\|\widetilde{H}_m\|_\infty)^{p-k}\,dt
\end{aligned}$$

$$\leq c\int_0^{T_0} \|\widetilde{H}_n^k-\widetilde{H}_m^k-M_{m,n}\|_{H^{-1}}^{\frac{1}{2}}(\|Dw_n\|_2+\|Dw_m\|_2)^{\frac{1}{2}}(\|\widetilde{H}_n\|_\infty+\|\widetilde{H}_m\|_\infty)^{\frac{p-k+1}{2}}\,dt \\ +\int_0^{T_0}|M_{m,n}|(\|\widetilde{H}_n\|_\infty+\|\widetilde{H}_m\|_\infty)^{p-k}\,dt. \tag{3-34}$$

Fix $n, m\in\mathbb{N}$. We now estimate the $H^{-1}$-norm of $\widetilde{H}_n^k-\widetilde{H}_m^k-M_{m,n}$. Recall that, in view of Remark 3.3,

$$\|\widetilde{H}_n^k-\widetilde{H}_m^k-M_{m,n}\|_{H^{-1}}=\|Du\|_2, \tag{3-35}$$

where $u$ is the unique $Q$-periodic solution of

$$\begin{cases}-\Delta u=\widetilde{H}_n^k-\widetilde{H}_m^k-M_{m,n} \text{ in } Q,\\ \int_Q u\,dx=0.\end{cases} \tag{3-36}$$

Thus,

$$\int_Q |Du|^2\,dx=\int_Q u(\widetilde{H}_n^k-\widetilde{H}_m^k-M_{m,n})\,dx=\int_Q u(\widetilde{H}_n-\widetilde{H}_m)\sum_{i=0}^{k-1}\widetilde{H}_n^{k-1-i}\widetilde{H}_m^i\,dx, \tag{3-37}$$

where we also used that $\int_Q u\,dx=0$. Fix $\delta\in(0,1)$ (to be chosen) and let $T^\delta(t):=(t\vee-\delta)\wedge\delta$. Then

$$\widetilde{H}_n=[(\widetilde{H}_n-\delta)^+ +\delta]+T^\delta(\widetilde{H}_n)-[(-\widetilde{H}_n-\delta)^+ +\delta] \tag{3-38}$$

and (see (3-32))

$$(\widetilde{H}_n-\delta)^+ +\delta=\begin{cases}w_n^{\frac{1}{p-1}} & \text{if } w_n\geq\delta^{p-1},\\ \delta & \text{otherwise.}\end{cases}$$

Hence,

$$|D[(\widetilde{H}_n-\delta)^+ +\delta]|\leq c\frac{|Dw_n|}{\delta^{p-2}}, \tag{3-39}$$

and a similar estimate holds for $D[(-\widetilde{H}_n-\delta)^+ +\delta]$. We now set

$$V_{n,\delta}:=[(\widetilde{H}_n-\delta)^+ +\delta]-[(-\widetilde{H}_n-\delta)^+ +\delta]. \tag{3-40}$$

From (3-37) we have

$$\begin{aligned}\int_Q |Du|^2\,dx &\\ &=\int_Q u(\widetilde{H}_n-\widetilde{H}_m)\sum_{i=0}^{k-1}\sum_{r=0}^{k-1-i}\sum_{s=0}^{i}\binom{k-1-i}{r}\binom{i}{s}V_{n,\delta}^{k-1-i-r}V_{m,\delta}^{i-s}[T^\delta(\widetilde{H}_n)]^r[T^\delta(\widetilde{H}_m)]^s\,dx\\ &=\int_Q u(\widetilde{H}_n-\widetilde{H}_m)\sum_{i=0}^{k-1}V_{n,\delta}^{k-1-i}V_{m,\delta}^i\,dx\\ &\quad+\int_Q u(\widetilde{H}_n-\widetilde{H}_m)\sum_{i=0}^{k-1}\sum_{(r,s)\neq(0,0)}\binom{k-1-i}{r}\binom{i}{s}V_{n,\delta}^{k-1-i-r}V_{m,\delta}^{i-s}[T^\delta(\widetilde{H}_n)]^r[T^\delta(\widetilde{H}_m)]^s\,dx\\ &=: L+M.\end{aligned} \tag{3-41}$$

We start by estimating the last term in the previous chain of equalities:

$$
\begin{aligned}
|M| &\le c\int_Q |u||\widetilde H_n-\widetilde H_m|\sum_{i=0}^{k-1}\sum_{(r,s)\neq(0,0)}\delta^{r+s}V_{n,\delta}^{k-1-i-r}V_{m,\delta}^{i-s}\,dx\\
&\le c\int_Q |u|(|\widetilde H_n|+|\widetilde H_m|)\sum_{\ell=1}^{k-1}\delta^{\ell}[V_{n,\delta}^{k-1-\ell}+V_{m,\delta}^{k-1-\ell}]\,dx\\
&\le c\delta\int_Q |u|(|\widetilde H_n|+|\widetilde H_m|)(1+V_{n,\delta}^{k-2}+V_{m,\delta}^{k-2})\,dx\\
&\le c\delta\Big(\int_Q u^2\,dx\Big)^{\frac12}(1+\|\widetilde H_n\|_\infty+\|\widetilde H_m\|_\infty)^{k-1}\\
&\le \frac16\int_Q |Du|^2\,dx+c\delta^2(1+\|\widetilde H_n\|_\infty+\|\widetilde H_m\|_\infty)^{2(k-1)}, \qquad (3\text{-}42)
\end{aligned}
$$

where we used (3-40) and the Poincaré and Young inequalities. To deal with $L$, we integrate by parts and use (2-3) and the periodicity of $u$, $\tilde h_n$, and $\tilde h_m$ to get

$$
L=\int_Q\Big(\frac{D\tilde h_n}{\widetilde J_n}-\frac{D\tilde h_m}{\widetilde J_m}\Big)Du\sum_{i=0}^{k-1}V_{n,\delta}^{k-1-i}V_{m,\delta}^{i}\,dx+\int_Q\Big(\frac{D\tilde h_n}{\widetilde J_n}-\frac{D\tilde h_m}{\widetilde J_m}\Big)u\sum_{i=0}^{k-1}D(V_{n,\delta}^{k-1-i}V_{m,\delta}^{i})\,dx,
$$

where

$$
\tilde h_n(x,t):=h_{i,n}(x)\quad \text{if } t\in[(i-1)\tau_n,i\tau_n) \text{ and } \widetilde J_n(x.t):=\sqrt{1+|D\tilde h_n(x,t)|^2}. \qquad (3\text{-}43)
$$

From the equality above, recalling (3-32), (3-39), and (3-40), and setting

$$
\varepsilon_{n,m}:=\sup_{t\in[0,T_0]}\Big\|\frac{D\tilde h_n}{\widetilde J_n}(\cdot,t)-\frac{D\tilde h_m}{\widetilde J_m}(\cdot,t)\Big\|_\infty,
$$

we may estimate

$$
\begin{aligned}
|L| &\le c\varepsilon_{n,m}\int_Q |Du|(1+|\widetilde H_n|^{k-1}+|\widetilde H_m|^{k-1})\,dx\\
&\qquad + c\varepsilon_{n,m}\int_Q |u|\sum_{i=0}^{k-1}[|DV_{n,\delta}^{k-1-i}|V_{m,\delta}^{i}+|DV_{m,\delta}^{i}|V_{n,\delta}^{k-1-i}]\,dx\\
&\le \frac16\int_Q |Du|^2\,dx+c\varepsilon_{n,m}^2(1+\|\widetilde H_n\|_\infty+\|\widetilde H_m\|_\infty)^{2(k-1)}\\
&\qquad + c\varepsilon_{n,m}\int_Q |u|\frac{|Dw_n|}{\delta^{p-2}}\sum_{i=0}^{k-2}V_{n,\delta}^{k-2-i}V_{m,\delta}^{i}\,dx+c\varepsilon_{n,m}\int_Q |u|\frac{|Dw_m|}{\delta^{p-2}}\sum_{i=0}^{k-2}V_{m,\delta}^{i-1}V_{n,\delta}^{k-i-1}\,dx\\
&\le \frac16\int_Q |Du|^2\,dx+c\varepsilon_{n,m}^2(1+\|\widetilde H_n\|_\infty+\|\widetilde H_m\|_\infty)^{2(k-1)}\\
&\qquad + c\frac{\varepsilon_{n,m}}{\delta^{p-2}}\int_Q |u|(|Dw_n|+|Dw_m|)(1+\|\widetilde H_n\|_\infty+\|\widetilde H_m\|_\infty)^{k-2}\,dx
\end{aligned}
$$

$$\le \frac{1}{3}\int_Q |Du|^2\,dx + c\varepsilon_{n,m}^2(1+\|\widetilde{H}_n\|_\infty+\|\widetilde{H}_m\|_\infty)^{2(k-1)} + c\frac{\varepsilon_{n,m}^2}{\delta^{2(p-2)}}\int_Q (|Dw_n|+|Dw_m|)^2(1+\|\widetilde{H}_n\|_\infty+\|\widetilde{H}_m\|_\infty)^{2(k-2)}\,dx.$$

From this estimate, (3-35), (3-36), (3-41), and (3-42), choosing $\delta^{2(p-2)}=\varepsilon_{n,m}$, with $n$, $m$ so large that $\varepsilon_{n,m}<1$ (see Theorem 3.5(i)), we obtain

$$\|\widetilde{H}_n^k-\widetilde{H}_m^k-M_{m,n}\|_{H^{-1}}^2 \le c\varepsilon_{n,m}^\alpha\big[(1+\|\widetilde{H}_n\|_\infty+\|\widetilde{H}_m\|_\infty)^{2(k-1)} + (\|Dw_n\|_2+\|Dw_m\|_2)^2(1+\|\widetilde{H}_n\|_\infty+\|\widetilde{H}_m\|_\infty)^{2(k-2)}\big], \tag{3-44}$$

where $\alpha := \min\{1, 1/(p-2)\}$.

We now estimate $M_{m,n}$. Since

$$M_{m,n}=\int_Q(\widetilde{H}_n^k-\widetilde{H}_m^k)\,dx=\int_Q(\widetilde{H}_n-\widetilde{H}_m)\sum_{i=0}^{k-1}\widetilde{H}_n^{k-1-i}\widetilde{H}_m^i\,dx,$$

using the same argument with $u\equiv 1$ (see (3-44)) gives

$$|M_{m,n}|\le c(\varepsilon_{n,m})^{\frac{\alpha}{2}}\big[(1+\|\widetilde{H}_n\|_\infty+\|\widetilde{H}_m\|_\infty)^{k-1}+(\|Dw_n\|_2+\|Dw_m\|_2)(1+\|\widetilde{H}_n\|_\infty+\|\widetilde{H}_m\|_\infty)^{k-2}\big].$$

From this inequality, recalling (3-32), (3-34), and (3-44), we deduce

$$\begin{aligned}\int_0^{T_0}\!\!\int_Q\big||\widetilde{H}_n|^p-|\widetilde{H}_m|^p\big|\,dx\,dt &\le c(\varepsilon_{n,m})^{\frac{\alpha}{4}}\int_0^{T_0}(\|Dw_n\|_2+\|Dw_m\|_2)^{\frac12}(1+\|w_n\|_\infty+\|w_m\|_\infty)^{\frac{p}{2(p-1)}}\,dt\\ &\quad + c(\varepsilon_{n,m})^{\frac{\alpha}{4}}\int_0^{T_0}(\|Dw_n\|_2+\|Dw_m\|_2)(1+\|w_n\|_\infty+\|w_m\|_\infty)^{\frac12}\,dt\\ &\quad + c(\varepsilon_{n,m})^{\frac{\alpha}{2}}\int_0^{T_0}(1+\|w_n\|_\infty+\|w_m\|_\infty)\,dt\\ &\quad + c(\varepsilon_{n,m})^{\frac{\alpha}{2}}\int_0^{T_0}(\|Dw_n\|_2+\|Dw_m\|_2)(\|w_n\|_\infty+\|w_m\|_\infty)^{\frac{p-2}{p-1}}\,dt.\end{aligned}$$

Observe now that, by (3-5) and (3-20), there exists $C>0$ such that $\int_Q|w_n|\,dx\le\|\widetilde{H}_n\|_{p-1}^{p-1}\le C$ for all $n$ and thus, using the embedding of $H^2(Q)$ into $C(\overline{Q})$ and Poincaré's inequality,

$$\|Dw_n\|_2+\|w_n\|_\infty\le C(1+\|D^2w_n\|_2). \tag{3-45}$$

Therefore, from the above inequalities and also using the fact that $\frac12+p/(2(p-1))<2$ and that $1+(p-2)/(p-1)<2$, we conclude that

$$\int_0^{T_0}\!\!\int_Q\big||\widetilde{H}_n|^p-|\widetilde{H}_m|^p\big|\,dx\,dt\le c(\varepsilon_{n,m})^{\frac{\alpha}{4}}\int_0^{T_0}(1+\|D^2w_n\|_2+\|D^2w_m\|_2)^2\,dt\le c(\varepsilon_{n,m})^{\frac{\alpha}{4}},$$

where the last inequality follows from (3-21). This proves that the sequence $|\widetilde{H}_n|^p$ is a Cauchy sequence in $L^1(0, T_0; L^1(Q))$. Note also that using Lemma 3.14 we have

$$\begin{aligned}\int_0^{T_0}\int_Q \big||\widetilde{H}_n| - |\widetilde{H}_m|\big|^p \, dx\, dt &\le c\int_0^{T_0}\int_Q \big||\widetilde{H}_n| - |\widetilde{H}_m|\big|(|\widetilde{H}_n| + |\widetilde{H}_m|)^{p-1}\, dx\, dt \\ &\le c\int_0^{T_0}\int_Q \big||\widetilde{H}_n|^p - |\widetilde{H}_m|^p\big|\, dx\, dt. \end{aligned} \tag{3-46}$$

**Step 2.** We now conclude the proof by showing that $w_n$ is a Cauchy sequence in $L^1(0, T_0; L^2(Q))$. To this purpose, we use Lemma A.7 to obtain

$$\begin{aligned}\int_0^{T_0} \|w_n - w_m\|_2\, dt &\le \int_0^{T_0} \|w_n - w_m - N_{m,n}\|_2\, dt + \int_0^{T_0} |N_{m,n}|\, dt \\ &\le c\int_0^{T_0} \|w_n - w_m - N_{m,n}\|_{H^{-1}}^{\frac{2}{3}} \|D^2 w_n - D^2 w_m\|_2^{\frac{1}{3}}\, dt + \int_0^{T_0} |N_{m,n}|\, dt, \end{aligned} \tag{3-47}$$

where $N_{m,n} := \fint_Q (w_n - w_m)\, dx$. As observed in (3-35) and (3-36) , $\|w_n - w_m - N_{m,n}\|_{H^{-1}} = \|Dv\|_2$, where $v$ is the unique $Q$-periodic solution of

$$\begin{cases} -\Delta v = w_n - w_m - N_{m,n} & \text{in } Q, \\ \int_Q v\, dx = 0. \end{cases}$$

As in (3-37), using the fact that $\int_Q v\, dx = 0$, we have

$$\begin{aligned}\int_Q |Dv|^2\, dx &= \int_Q (w_n - w_m - N_{m,n})v = \int_Q (|\widetilde{H}_n|^{p-2}\widetilde{H}_n - |\widetilde{H}_m|^{p-2}\widetilde{H}_m)v\, dx \\ &= \int_Q (|\widetilde{H}_n|^{p-2} - |\widetilde{H}_m|^{p-2})\widetilde{H}_n v\, dx + \int_Q (\widetilde{H}_n - \widetilde{H}_m)|\widetilde{H}_m|^{p-2} v\, dx \\ &=: \widetilde{L} + \widetilde{M}. \end{aligned} \tag{3-48}$$

Now, by Hölder's inequality twice and the Sobolev embedding theorem,

$$\begin{aligned}|\widetilde{L}| &\le \int_Q \Big|(|\widetilde{H}_n|^p)^{\frac{p-2}{p}} - (|\widetilde{H}_m|^p)^{\frac{p-2}{p}}\Big||\widetilde{H}_n||v|\, dx \\ &\le \int_Q \big||\widetilde{H}_n|^p - |\widetilde{H}_m|^p\big|^{\frac{p-2}{p}}|\widetilde{H}_n||v|\, dx \\ &\le \|v\|_p \|\widetilde{H}_n\|_\infty \bigg(\int_Q \big||\widetilde{H}_n|^p - |\widetilde{H}_m|^p\big|^{\frac{p-2}{p-1}}\, dx\bigg)^{\frac{p-1}{p}} \\ &\le c\|Dv\|_2 \|\widetilde{H}_n\|_\infty \||\widetilde{H}_n|^p - |\widetilde{H}_m|^p\|_1^{\frac{p-2}{p}} \\ &\le \frac{1}{6}\int_Q |Dv|^2\, dx + c\|\widetilde{H}_n\|_\infty^2 \||\widetilde{H}_n|^p - |\widetilde{H}_m|^p\|_1^{\frac{2(p-2)}{p}}. \end{aligned} \tag{3-49}$$

To estimate $\widetilde{M}$, arguing as in the previous step (see (3-38)) and observing that $(-|\widetilde{H}_m|^{p-2}-\delta)^+=0$, we write

$$\begin{aligned}\widetilde{M} &= \int_Q (\widetilde{H}_n-\widetilde{H}_m)\big[(|\widetilde{H}_m|^{p-2}-\delta)^+ +\delta\big]v\,dx + \int_Q (\widetilde{H}_n-\widetilde{H}_m)\big[T^\delta(|\widetilde{H}_m|^{p-2})-\delta\big]v\,dx\\ &= \int_Q\Big(\frac{D\tilde{h}_n}{\widetilde{J}_n}-\frac{D\tilde{h}_m}{\widetilde{J}_m}\Big)Dv\big[(|\widetilde{H}_m|^{p-2}-\delta)^+ +\delta\big]\,dx\\ &\qquad+\int_Q\Big(\frac{D\tilde{h}_n}{\widetilde{J}_n}-\frac{D\tilde{h}_m}{\widetilde{J}_m}\Big)vD\big[(|\widetilde{H}_m|^{p-2}-\delta)^+ +\delta\big]\,dx+\int_Q (\widetilde{H}_n-\widetilde{H}_m)\big[T^\delta(|\widetilde{H}_m|^{p-2})-\delta\big]v\,dx.\end{aligned}$$

Similarly to what we proved in (3-39), we have

$$\big|D\big[(|\widetilde{H}_m|^{p-2}-\delta)^+ +\delta\big]\big|\le c\frac{|Dw_m|}{\delta^{\frac{1}{p-2}}}.$$

Therefore, arguing as in the previous step, we obtain

$$\begin{aligned}|\widetilde{M}| &\le \frac{1}{6}\int_Q |Dv|^2\,dx + c\varepsilon_{n,m}^2(1+\|\widetilde{H}_m\|_\infty)^{2(p-2)} + c\varepsilon_{n,m}\int_Q |v|\frac{|Dw_m|}{\delta^{\frac{1}{p-2}}}\,dx\\ &\qquad\qquad + c\delta\int_Q |v|(\|\widetilde{H}_n\|_\infty+\|\widetilde{H}_m\|_\infty)\,dx\\ &\le \frac{1}{3}\int_Q |Dv|^2\,dx + c\varepsilon_{n,m}^2(1+\|\widetilde{H}_m\|_\infty)^{2(p-2)} + c\frac{\varepsilon_{n,m}^2}{\delta^{\frac{2}{p-2}}}\|Dw_m\|_2^2 + c\delta^2(\|\widetilde{H}_n\|_\infty+\|\widetilde{H}_m\|_\infty)^2,\end{aligned}$$

where in the last line we used the Young and Poincaré inequalities. Choosing $\delta$ so that $\delta^{2/(p-2)}=\varepsilon_{n,m}$ and recalling (3-48) and (3-49), we conclude that

$$\begin{aligned}&\|w_n-w_m-N_{m,n}\|_{H^{-1}}\\ &\le c\|\widetilde{H}_n\|_\infty\||\widetilde{H}_n|^p-|\widetilde{H}_m|^p\|_1^{\frac{p-2}{p}} + c(\varepsilon_{n,m})^{\frac{\beta}{2}}(1+\|\widetilde{H}_n\|_\infty+\|\widetilde{H}_m\|_\infty+\|\widetilde{H}_m\|_\infty^{p-2}+\|Dw_m\|_2),\end{aligned}\tag{3-50}$$

where $\beta=\min\{1,\ p-2\}$.

Since, by (3-32),

$$N_{m,n}=\int_Q (w_n-w_m)\,dx = \int_Q (|\widetilde{H}_n|^{p-2}-|\widetilde{H}_m|^{p-2})\widetilde{H}_n\,dx + \int_Q (\widetilde{H}_n-\widetilde{H}_m)|\widetilde{H}_m|^{p-2}\,dx,$$

the same argument used to estimate the last two integrals in (3-48) (with $v\equiv 1$) gives

$$|N_{m,n}|\le c\|\widetilde{H}_n\|_\infty\||\widetilde{H}_n|^p-|\widetilde{H}_m|^p\|_1^{\frac{p-2}{p}} + c(\varepsilon_{n,m})^{\frac{\beta}{2}}(\|\widetilde{H}_n\|_\infty+\|\widetilde{H}_m\|_\infty+\|\widetilde{H}_m\|_\infty^{p-2}+\|Dw_m\|_2).$$

From this estimate, recalling (3-32), (3-47) and (3-50), we have that

$$\begin{aligned}
&\int_0^{T_0} \|w_n - w_m\|_2\,dt \\
&\le c\int_0^{T_0} \||\widetilde{H}_n|^p - |\widetilde{H}_m|^p\|_1^{\frac{2(p-2)}{3p}} \|w_n\|_\infty^{\frac{2}{3(p-1)}} (\|D^2 w_n\|_2 + \|D^2 w_m\|_2)^{\frac{1}{3}}\,dt \\
&\quad + c(\varepsilon_{n,m})^{\frac{\beta}{3}} \int_0^{T_0} \big(1 + \|w_n\|_\infty^{\frac{1}{p-1}} + \|w_m\|_\infty^{\frac{1}{p-1}} + \|w_m\|_\infty^{\frac{p-2}{p-1}} + \|Dw_m\|_2\big)^{\frac{2}{3}} (\|D^2 w_n\|_2 + \|D^2 w_m\|_2)^{\frac{1}{3}}\,dt \\
&\quad + c\int_0^{T_0} \|w_n\|_\infty^{\frac{1}{p-1}} \||\widetilde{H}_n|^p - |\widetilde{H}_m|^p\|_1^{\frac{p-2}{p}}\,dt \\
&\quad + c(\varepsilon_{n,m})^{\frac{\beta}{2}} \int_0^{T_0} (\|w_n\|_\infty^{\frac{1}{p-1}} + \|w_m\|_\infty^{\frac{1}{p-1}} + \|w_m\|_\infty^{\frac{p-2}{p-1}} + \|Dw_m\|_2)\,dt.
\end{aligned}$$

Using (3-45) and Hölder's inequality, we can bound the right-hand side of this inequality by

$$\begin{aligned}
&c\int_0^{T_0} \||\widetilde{H}_n|^p - |\widetilde{H}_m|^p\|_1^{\frac{2(p-2)}{3p}} (1 + \|D^2 w_n\|_2 + \|D^2 w_m\|_2)^{\frac{1}{3} + \frac{2}{3(p-1)}}\,dt \\
&\quad + c(\varepsilon_{n,m})^{\frac{\beta}{3}} \int_0^{T_0} \big(1 + \|D^2 w_n\|_2 + \|D^2 w_m\|_2\big)^{\frac{2}{3}} (\|D^2 w_n\|_2 + \|D^2 w_m\|_2)^{\frac{1}{3}}\,dt \\
&\quad + c\int_0^{T_0} \big(1 + \|D^2 w_n\|_2\big)^{\frac{1}{p-1}} \||\widetilde{H}_n|^p - |\widetilde{H}_m|^p\|_1^{\frac{p-2}{p}}\,dt + c(\varepsilon_{n,m})^{\frac{\beta}{2}} \int_0^{T_0} (1 + \|D^2 w_n\|_2 + \|D^2 w_m\|_2)\,dt \\
&\le c\Big(\int_0^{T_0}\int_Q \big||\widetilde{H}_n|^p - |\widetilde{H}_m|^p\big|\,dx\,dt\Big)^{\frac{2(p-2)}{3p}} \Big[\int_0^{T_0} \big(\|D^2 w_n\|_2 + \|D^2 w_m\|_2\big)^{\frac{p(p+1)}{(p-1)(p+4)}}\Big]^{\frac{p+4}{3p}} \\
&\quad + c\Big(\int_0^{T_0}\int_Q \big||\widetilde{H}_n|^p - |\widetilde{H}_m|^p\big|\,dx\,dt\Big)^{\frac{p-2}{2}} \Big[\int_0^{T_0} \big(1 + \|D^2 w_n\|_2\big)^{\frac{p}{2(p-1)}}\Big]^{\frac{2}{p}} \\
&\quad + c(\varepsilon_{n,m})^{\frac{\beta}{3}} \int_0^{T_0} (1 + \|D^2 w_n\|_2 + \|D^2 w_m\|_2)\,dt.
\end{aligned}$$

Since $p(p+1)/((p-1)(p+4)) < 2$ and $p/(2(p-1)) < 2$, recalling (3-21), we finally have

$$\begin{aligned}
&\int_0^{T_0} \|w_n - w_m\|_2\,dt \\
&\le c\Big(\int_0^{T_0}\int_Q \big||\widetilde{H}_n|^p - |\widetilde{H}_m|^p\big|\,dx\,dt\Big)^{\frac{2(p-2)}{3p}} + c\Big(\int_0^{T_0}\int_Q \big||\widetilde{H}_n|^p - |\widetilde{H}_m|^p\big|\,dx\,dt\Big)^{\frac{p-2}{2}} + c(\varepsilon_{n,m})^{\frac{\beta}{3}}.
\end{aligned}$$

The conclusion follows from Step 1. □

**Corollary 3.15.** *Let* $\widetilde{H}_n$ *be the functions defined in* (3-20), *let* $h$ *be the limiting function provided by Theorem 3.5*, *and set*

$$H := -\operatorname{Div}\Big(\frac{Dh}{1 + |Dh|^2}\Big).$$

*Then,*

$$|\widetilde{H}_n|^p \to |H|^p \ \text{ in } L^1(0,T_0;L^1(Q)) \quad \text{and} \quad |\widetilde{H}_n|^{p-2}\widetilde{H}_n \to |H|^{p-2}H \ \text{ in } L^1(0,T_0;L^2(Q)). \tag{3-51}$$

*Proof.* Let $\tilde{h}_n$ and $\widetilde{J}_n$ be as in the proof of Lemma 3.13. From Theorem 3.5(i) we get that, for all $t \in (0, T_0)$ and for all $\varphi \in C^1_\#(Q)$, we have

$$\int_Q \widetilde{H}_n \varphi\, dx = \int_Q \frac{D\tilde{h}_n}{\widetilde{J}_n} \cdot D\varphi\, dx \longrightarrow \int_Q \frac{Dh}{J} \cdot D\varphi\, dx = \int_Q \widetilde{H}\varphi\, dx,$$

where $J = \sqrt{1+|Dh|^2}$. Since, for every $t$, $\widetilde{H}_n(\cdot,t)$ is bounded in $L^p(Q)$, we deduce that, for all $t \in (0, T_0)$,

$$\widetilde{H}_n(\cdot,t) \rightharpoonup H(\cdot,t) \quad \text{weakly in } L^p(Q). \tag{3-52}$$

On the other hand, from Lemma 3.13 we know that there exist a subsequence $n_j$ and two functions $z$, $w$ such that, for a.e. $t$,

$$|\widetilde{H}_{n_j}(\cdot,t)|^p \longrightarrow z(\cdot,t) \quad \text{in } L^1(Q) \qquad \text{and} \qquad (|\widetilde{H}_{n_j}|^{p-2}\widetilde{H}_{n_j})(\cdot,t) \longrightarrow w(\cdot,t) \quad \text{in } L^2(Q). \tag{3-53}$$

Moreover, for any such $t$ there exists a further subsequence (depending on $t$), not relabelled, such that $|\widetilde{H}_{n_j}(x,t)|^p$, $|\widetilde{H}_{n_j}(x,t)|^{p-2}\widetilde{H}_{n_j}(x,t)$, and thus $\widetilde{H}_{n_j}(x,t)$ converge for a.e. $x$. Then, by (3-52), $\widetilde{H}_{n_j}(x,t) \to H(x,t)$ for a.e. $x$. Thus, we conclude that $z = |H|^p$ and $w = |H|^{p-2}H$. ☐

We now prove short time existence for (3-1).

**Theorem 3.16.** *Let $h_0 \in W^{2,p}_\#(Q)$, let $h$ be the function given in Theorem 3.4, and let $T_0 > 0$ be as in Theorem 3.5. Then $h$ is a solution of* (3-1) *in* $[0, T_0]$ *in the sense of Definition 3.1 with initial datum $h_0$. Moreover, there exists a nonincreasing $g$ such that*

$$F(h(\cdot,t), u_h(\cdot,t)) = g(t) \quad \text{for } t \in [0,T_0] \setminus Z_0, \tag{3-54}$$

*where $Z_0$ is a set of zero measure, and*

$$F(h(\cdot,t), u_h(\cdot,t)) \le g(t+) \quad \text{for } t \in Z_0. \tag{3-55}$$

This result motivates the following definition:

**Definition 3.17.** We say that a solution to (3-1) is *variational* if it is the limit of a subsequence of the minimizing movements scheme as in Theorem 3.5(i).

*Proof of Theorem 3.16.* Let $\widetilde{H}_n$, $\tilde{h}_n$, $\widetilde{J}_n$ be the functions given in (3-20), and (3-43). Set $\widetilde{W}_n(x,t) := W\big(E(u_{i,n})(x, h_{i,n}(x))\big)$ and $\tilde{v}_n(x,t) := v_{h_{i,n}}(x)$ for $t \in [(i-1)\tau_n, i\tau_n)$. Moreover, define $\hat{v}_n := \tilde{v}_n/\tau_n$. Note that, for all $t$, $\hat{v}_n(\cdot,t)$ is the unique $Q$-periodic solution to

$$\begin{cases} \Delta_{\Gamma_{\tilde{h}_n(\cdot,t-\tau_n)}} w = \dfrac{1}{\widetilde{J}_n(\cdot,t-\tau_n)} \dfrac{\partial h_n(\cdot,t)}{\partial t} \\ \int_{\Gamma_{\tilde{h}_n(\cdot,t-\tau_n)}} w \, d\mathcal{H}^2 = 0. \end{cases} \tag{3-56}$$

Fix $t \in (0, T_0)$ and a sequence $(i_k, n_k)$ such that $t_k := i_k \tau_{n_k} \to t$. Summing (3-22) from $i = 1$ to $i = i_k$, we get

$$\int_0^{t_k}\int_Q \widetilde{W}_{n_k}\varphi\,dx\,dt + \int_0^{t_k}\int_Q D\psi(-D\tilde{h}_{n_k},1)\cdot(-D\varphi,0)\,dx\,dt + \frac{\varepsilon}{p}\int_0^{t_k}\int_Q |\widetilde{H}_{n_k}|^p \frac{D\tilde{h}_{n_k}\cdot D\varphi}{\widetilde{J}_{n_k}}\,dx\,dt$$
$$-\varepsilon\int_0^{t_k}\int_Q |\widetilde{H}_{n_k}|^{p-2}\widetilde{H}_{n_k}\Big[\Delta\varphi - \frac{D^2\varphi[D\tilde{h}_{n_k},D\tilde{h}_{n_k}]}{\widetilde{J}^2_{n_k}} - \frac{\Delta\tilde{h}_{n_k}D\tilde{h}_{n_k}\cdot D\varphi}{\widetilde{J}^2_{n_k}} - 2\frac{D^2\tilde{h}_{n_k}[D\tilde{h}_{n_k},D\varphi]}{\widetilde{J}^2_{n_k}}$$
$$+3\frac{D^2\tilde{h}_{n_k}[D\tilde{h}_{n_k},D\tilde{h}_{n_k}]D\tilde{h}_{n_k}\cdot D\varphi}{\widetilde{J}^4_{n_k}}\Big]\,dx\,dt$$
$$-\int_0^{t_k}\int_Q \hat{v}_{n_k}\varphi\,dx\,dt = 0. \tag{3-57}$$

We claim that we can pass to the limit in the above equation to get

$$\int_0^t\int_Q W\big(E\big(u(x,h(x,s),s)\big)\big)\varphi\,dx\,ds + \int_0^t\int_Q D\psi(-Dh,1)\cdot(-D\varphi,0)\,dx\,ds$$
$$+\frac{\varepsilon}{p}\int_0^t\int_Q |H|^p\frac{Dh\cdot D\varphi}{J}\,dx\,ds$$
$$-\varepsilon\int_0^t\int_Q |H|^{p-2}H\Big[\Delta\varphi - \frac{D^2\varphi[Dh,Dh]}{J^2} - \frac{\Delta h Dh\cdot D\varphi}{J^2} - 2\frac{D^2h[Dh,D\varphi]}{J^2}$$
$$+3\frac{D^2h[Dh,Dh]Dh\cdot D\varphi}{J^4}\Big]\,dx\,ds - \int_0^t\int_Q \hat{v}\varphi\,dx\,ds = 0, \tag{3-58}$$

where $\hat{v}(\cdot,t)$ is the unique periodic solution in $H^1_\#(\Gamma(t))$ to

$$\begin{cases} \Delta_{\Gamma_{h(\cdot,t)}} w = \dfrac{1}{J(\cdot,t)}\dfrac{\partial h(\cdot,t)}{\partial t}, \\ \int_{\Gamma_{h(\cdot,t)}} w\,d\mathcal{H}^2 = 0 \end{cases} \tag{3-59}$$

for a.e. $t \in (0, T_0)$. To prove the claim, observe that the convergence of the first two integrals in (3-57) immediately follows from (i) and (iii) of Theorem 3.5. The convergence of the third integral in (3-57) follows from (3-51) and Theorem 3.5(i). Similarly, (3-51) and of Theorem 3.5(i) imply that

$$\int_0^{t_k}\int_Q |\widetilde{H}_{n_k}|^{p-2}\widetilde{H}_{n_k}\Big[\Delta\varphi - \frac{D^2\varphi[D\tilde{h}_{n_k},D\tilde{h}_{n_k}]}{\widetilde{J}^2_{n_k}}\Big]dx\,dt \to \int_0^t\int_Q |H|^{p-2}H\Big[\Delta\varphi - \frac{D^2\varphi[Dh,Dh]}{J^2}\Big]dx\,ds.$$

Next we show the convergence of

$$\int_0^{t_k}\int_Q |\widetilde{H}_{n_k}|^{p-2}\widetilde{H}_{n_k}\Big[-\frac{\Delta\tilde{h}_{n_k}D\tilde{h}_{n_k}\cdot D\varphi}{\widetilde{J}^2_{n_k}} - 2\frac{D^2\tilde{h}_{n_k}[D\tilde{h}_{n_k},D\varphi]}{\widetilde{J}^2_{n_k}}$$
$$+3\frac{D^2\tilde{h}_{n_k}[D\tilde{h}_{n_k},D\tilde{h}_{n_k}]D\tilde{h}_{n_k}\cdot D\varphi}{\widetilde{J}^4_{n_k}}\Big]\,dx\,dt$$

to the corresponding term in (3-58). To this purpose, we only show that

$$\int_0^{t_k}\int_Q |\widetilde{H}_{n_k}|^{p-2}\widetilde{H}_{n_k}\frac{\Delta\tilde{h}_{n_k}D\tilde{h}_{n_k}\cdot D\varphi}{\widetilde{J}^2_{n_k}}\,dx\,dt \longrightarrow \int_0^t\int_Q |H|^{p-2}H\frac{\Delta h Dh\cdot D\varphi}{J^2}\,dx\,ds, \tag{3-60}$$

since the convergence of the other terms can be shown in a similar way. To prove (3-60), we first observe that, by (3-5) and Theorem 3.5(i), we have $\Delta\tilde{h}_{n_k}(\cdot,t)\rightharpoonup\Delta h(\cdot,t)$ in $L^p(Q)$ for all $t\in(0,T_0)$. On the other hand, (3-51) yields that for a.e. $t\in(0,T_0)$ we have $(\widetilde{H}_{n_k}|^{p-2}\widetilde{H}_{n_k})(\cdot,t)\to(|H|^{p-2}H)(\cdot,t)$ in $L^2(Q)$. Therefore, for a.e. $t\in(0,T_0)$,

$$\int_Q |\widetilde{H}_{n_k}|^{p-2}\widetilde{H}_{n_k}\frac{\Delta\tilde{h}_{n_k}D\tilde{h}_{n_k}\cdot D\varphi}{\widetilde{J}^2_{n_k}}\,dx \longrightarrow \int_Q |H|^{p-2}H\frac{\Delta h Dh\cdot D\varphi}{J^2}\,dx.$$

The conclusion then follows by applying the Lebesgue dominated convergence theorem after observing that, by (2-9) and (3-5),

$$\begin{aligned}\left|\int_Q |\widetilde{H}_{n_k}|^{p-2}\widetilde{H}_{n_k}\frac{\Delta\tilde{h}_{n_k}D\tilde{h}_{n_k}\cdot D\varphi}{\widetilde{J}^2_{n_k}}\,dx\right| &\le C\|\Delta\tilde{h}_{n_k}\|_{L^2(Q)}\||\widetilde{H}_{n_k}|^{p-2}\widetilde{H}_{n_k}\|_{L^2(Q)}\\ &\le C\||\widetilde{H}_{n_k}|^{p-2}\widetilde{H}_{n_k}\|_{L^2(Q)}\end{aligned}$$

and that $\||\widetilde{H}_{n_k}|^{p-2}\widetilde{H}_{n_k}\|_{L^2(Q)}$ converges in $L^1(0,T_0)$ thanks to (3-51).

Note (3-51) implies that for a.e. $t\in(0,T_0)$ we have $\|\widetilde{H}_{n_k}(\cdot,t)\|_{L^p(Q)}\to\|H(\cdot,t)\|_{L^p(Q)}$. Since $\widetilde{H}_{n_k}(\cdot,t)\rightharpoonup H(\cdot,t)$ in $L^p(Q)$ (see (3-52)), we may conclude that $\widetilde{H}_{n_k}(\cdot,t)\to H(\cdot,t)$ in $L^p(Q)$ for a.e. $t\in(0,T_0)$. Therefore, by (2-3) and [Acerbi et al. 2013, Lemma 7.2], we also have $\tilde{h}_{n_k}(\cdot,t)\to h(\cdot,t)$ in $W^{2,p}_\#(Q)$ for a.e. $t\in(0,T_0)$. Thus, by (2-9) and (3-5) and the Lebesgue dominated convergence theorem we infer that

$$\int_0^{T_0}\int_Q |D^2\tilde{h}_{n_k}-D^2h|^p\,dx\,dt\longrightarrow 0. \tag{3-61}$$

This, together with the fact that $h_n\rightharpoonup h$ weakly in $H^1(0,T_0;H^{-1}_\#(Q))$ (see (3-6)), implies that

$$\frac{1}{\widetilde{J}_{n_k}(\cdot,\cdot-\tau_{n_k})}\frac{\partial h_{n_k}}{\partial t}\rightharpoonup\frac{1}{J}\frac{\partial h}{\partial t}\quad\text{in } L^2(0,T_0;H^{-1}_\#(Q)). \tag{3-62}$$

Indeed, for any $\varphi\in L^2(0,T_0;H^1_\#(Q))$,

$$\begin{aligned}&\left|\int_0^{T_0}\int_Q\Big(\frac{1}{\widetilde{J}_{n_k}(\cdot,\cdot-\tau_{n_k})}\frac{\partial h_{n_k}}{\partial t}-\frac{1}{J}\frac{\partial h}{\partial t}\Big)\varphi\,dx\,dt\right|\\ &\quad\le\left|\int_0^{T_0}\int_Q\Big(\frac{1}{\widetilde{J}_{n_k}(\cdot,\cdot-\tau_{n_k})}-\frac{1}{J}\Big)\frac{\partial h_{n_k}}{\partial t}\varphi\,dx\,dt\right|+\left|\int_0^{T_0}\int_Q\Big(\frac{\partial h_{n_k}}{\partial t}-\frac{\partial h}{\partial t}\Big)\frac{\varphi}{J}\,dx\,dt\right|\\ &\quad\le\int_0^{T_0}\int_Q\Big\|\frac{\partial h_{n_k}}{\partial t}\Big\|_{H^{-1}}\Big\|\frac{\varphi}{\widetilde{J}_{n_k}(\cdot,\cdot-\tau_{n_k})}-\frac{\varphi}{J}\Big\|_{H^1}\,dx\,dt+\left|\int_0^{T_0}\int_Q\Big(\frac{\partial h_{n_k}}{\partial t}-\frac{\partial h}{\partial t}\Big)\frac{\varphi}{J}\,dx\,dt\right|.\end{aligned} \tag{3-63}$$

Since $H^1_\#(Q)$ is embedded in $L^q(Q)$ for all $q\ge1$, we deduce from (3-61) that $\varphi/\widetilde{J}_{n_k}(\cdot,\cdot-\tau_{n_k})\to\varphi/J$ in $L^2(0,T_0;H^1_\#(Q))$. This convergence together with (3-3) shows that the second-to-last integral in

(3-63) vanishes in the limit. On the other hand, the last integral in (3-63) also vanishes in the limit, since $h_{n_k} \rightharpoonup h$ weakly in $H^1(0, T_0; H^{-1}_\#(Q))$. Thus, (3-62) follows.

Arguing as in the proof of Theorem 3.11 and integrating with respect to $t$, we have, from (3-56),

$$\int_0^t \int_Q A_{n_k} D\hat{v}_n \cdot D\varphi \, dx \, ds = \int_0^t \int_Q \frac{1}{\widetilde{J}_{n_k}(\cdot, \cdot - \tau_{n_k})} \frac{\partial h_{n_k}}{\partial t} \varphi \, dx \, ds \tag{3-64}$$

for all $\varphi \in L^2(0, T_0; H^1_\#(Q))$, where

$$A_{n_k}(x,t) := \left( I - \frac{D\tilde{h}_{n_k}(\cdot, \cdot - \tau_{n_k}) \otimes D\tilde{h}_{n_k}(\cdot, \cdot - \tau_{n_k})}{\widetilde{J}_{n_k}(\cdot, \cdot - \tau_{n_k})^2} \right) \widetilde{J}_{n_k}(\cdot, \cdot - \tau_{n_k}).$$

Note that (3-12) implies that $A_{n_k}(x,t)$ is an elliptic matrix with ellipticity constants depending only on $\Lambda_0$ for all $(x,t)$. Therefore, (3-64) immediately implies that

$$\int_0^{T_0} \int_Q |D\hat{v}_{n_k}|^2 \, dx \, dt \le c \int_0^{T_0} \left\| \frac{\partial h_{n_k}}{\partial t} \right\|^2_{H^{-1}} dt \le c$$

thanks to (3-3). Since $A_{n_k} \to A := (I - (Dh \otimes Dh)/J^2)J$ in $L^\infty(0, T_0; L^\infty(Q))$ by Theorem 3.5(i), from the estimate above and recalling (3-62) and (3-64) we conclude that

$$\hat{v}_{n_k} \rightharpoonup \hat{v} \quad \text{weakly in } L^2(0, T_0; H^1_\#(Q)),$$

where $\hat{v}$ satisfies

$$\int_0^t \int_Q A D\hat{v} \cdot D\varphi \, dx \, ds = \int_0^t \int_Q \frac{1}{J} \frac{\partial h}{\partial t} \varphi \, dx \, ds$$

for all $\varphi \in L^2(0, T_0; H^1_\#(Q))$ and for all $t \in (0, T_0)$. In turn, letting $\varphi$ vary in a countable dense subset of $H^1_\#(Q)$ and differentiating the above equation with respect to $t$, we conclude that, for a.e. $t \in (0, T_0)$, $\hat{v}(\cdot, t)$ is the unique solution in $H^1_\#(\Gamma_{h(\cdot,t)})$ to (3-59) for a.e. $t \in (0, T_0)$. This shows that the last integral in (3-57) converges and thus (3-58) holds. Again letting $\varphi$ vary in a countable dense subset of $H^1_\#(Q)$ and differentiating (3-58) with respect to $t$, we obtain

$$\begin{aligned} \int_Q W\big(E(u(x, h(x,t), t))\big)\varphi \, dx + \int_Q D\psi(-Dh, 1) \cdot (-D\varphi, 0) \, dx + \frac{\varepsilon}{p} \int_Q |H|^p \frac{Dh \cdot D\varphi}{J} \, dx \\ -\varepsilon \int_Q |H|^{p-2} H \left[ \Delta\varphi - \frac{D^2\varphi[Dh, Dh]}{J^2} - \frac{\Delta h \, Dh \cdot D\varphi}{J^2} - 2\frac{D^2 h[Dh, D\varphi]}{J^2} + 3\frac{D^2 h[Dh, Dh] Dh \cdot D\varphi}{J^4} \right] dx \\ - \int_Q \hat{v}\varphi \, dx = 0 \end{aligned} \tag{3-65}$$

for all $\varphi \in H^1_\#(Q)$. Since, by (3-21), $|H|^{p-2}H \in L^2(0, T_0; H^2_\#(Q))$, arguing as in Step 2 of the proof of Theorem 3.11 we have that the above equation is equivalent to

$$\begin{aligned} \varepsilon \int_{\Gamma_h} D_{\Gamma_h}(|H|^{p-2}H) D_{\Gamma_h}\phi \, d\mathcal{H}^2 - \varepsilon \int_{\Gamma_h} |H|^{p-2} H \Big( |B|^2 - \frac{1}{p} H^2 \Big) \phi \, d\mathcal{H}^2 \\ + \int_{\Gamma_h} \big[ \mathrm{Div}_{\Gamma_h}(D\psi(\nu)) + W(E(u)) \big] \phi \, d\mathcal{H}^2 - \int_{\Gamma_h} \hat{v}\phi \, d\mathcal{H}^2 = 0 \end{aligned}$$

for a.e. $t \in (0, T_0)$, where $\phi := \varphi / J$. This equation, together with (3-59), implies that $h$ is a solution to (3-1) in the sense of Definition 3.1.

Next, to show that the energy decreases during the evolution, we observe first that, for every $n$, the map $t \mapsto F(\tilde{h}_n(\cdot, t), \tilde{u}_n(\cdot, t))$ is nonincreasing, as shown in (3-4). Note also that thanks to (3-51) we may assume up to extracting a further subsequence that, for a.e. $t$, $\widetilde{H}_n \to H$ in $L^p(Q)$. This fact, together with (i) and (iii) of Theorem 3.5, implies that for all such $t$, $F(\tilde{h}_n(\cdot, t), \tilde{u}_n(\cdot, t)) \to F(h(\cdot, t), u(\cdot, t))$. Thus also (3-54) follows. Let $t \in Z_0$ and choose $t_n \to t+$ with $t_n \notin Z_0$ for every $n$. Finally, since $h(\cdot, t_n) \rightharpoonup h(\cdot, t)$ weakly in $W^{2,p}_\#(Q)$ by (3-5), by lower semicontinuity we get that

$$F(h(\cdot, t), u(\cdot, t)) \le \liminf_n F(h(\cdot, t_n), u(\cdot, t_n)) = \lim_n g(t_n) = g(t+). \qquad \square$$

## 4. Liapunov stability of the flat configuration

In this section we are going to study the Liapunov stability of an admissible flat configuration. Take $h(x) \equiv d > 0$ and let $u_d$ denote the corresponding elastic equilibrium. Throughout this section we assume that the Dirichlet datum $w_0$ is affine, i.e., of the form $w_0(x, y) = (A[x], 0)$ for some $A \in \mathbb{M}^{2\times 2}$. As already mentioned, a typical choice is given by $w_0(x, y) := (e_0^1 x_1, e_0^2 x_2, 0)$, where the vector $e_0 := (e_0^1, e_0^2)$ with $e_0^1, e_0^2 > 0$ embodies the mismatch between the crystalline lattices of film and substrate.

A detailed analysis of the so-called Asaro–Tiller–Grinfeld morphological stability/instability was undertaken in [Bonacini 2013b; Fusco and Morini 2012]. It was shown that, if $d$ is sufficiently small, then the flat configuration $(d, u_d)$ is a volume constrained local minimizer for the functional

$$G(h, u) := \int_{\Omega_h} W(E(u))\, dz + \int_{\Gamma_h} \psi(\nu)\, d\mathcal{H}^2. \tag{4-1}$$

To be precise, it was proved that, if $d$ is small enough, then the second variation $\partial^2 G(d, u_d)$ is positive definite and that, in turn, this implies the local minimality property. In order to state the results of this section, we need to introduce some preliminary notation. In the following, given $h \in C^2_\#(Q)$, $h \ge 0$, $\nu$ will denote the unit vector field coinciding with the gradient of the signed distance from $\Omega^\#_h$, which is well defined in a sufficiently small tubular neighborhood of $\Gamma^\#_h$. Moreover, for every $x \in \Gamma_h$ we set

$$\mathbb{B}(x) := D\nu(x). \tag{4-2}$$

Note that the bilinear form associated with $\mathbb{B}(x)$ is symmetric and, when restricted to $T_x\Gamma_h \times T_x\Gamma_h$, it coincides with the *second fundamental form of* $\Gamma_h$ *at* $x$. Here $T_x\Gamma_h$ denotes the tangent space to $\Gamma_h$ at $x$. For $x \in \Gamma_h$ we also set $H(x) := \operatorname{Div} \nu(x) = \operatorname{trace} \mathbb{B}(x)$, which is the *sum of the principal curvatures* of $\Gamma_h$ at $x$. Given a (sufficiently) smooth and positively one-homogeneous function $\omega : \mathbb{R}^N \setminus \{0\} \to \mathbb{R}$, we consider the *anisotropic second fundamental form* defined as

$$\mathbb{B}^\omega := D(D\omega \circ \nu),$$

and we set

$$H^\omega := \operatorname{trace} \mathbb{B}^\omega = \operatorname{Div}(D\omega \circ \nu). \tag{4-3}$$

We also introduce the space of periodic displacements

$$A(\Omega_h) := \{u \in LD_\#(\Omega_h; \mathbb{R}^3) : u(x,0) = 0\}. \tag{4-4}$$

Given a regular configuration $(h, u_h) \in X$ with $h \in C^2_\#(Q)$ and $\varphi \in \widetilde{H}^1_\#(Q)$, where

$$\widetilde{H}^1_\#(Q) := \left\{\varphi \in H^1_\#(Q) : \int_Q \varphi\, dx = 0\right\}, \tag{4-5}$$

we recall that the second variation of $G$ at $(h, u_h)$ with respect to the direction $\varphi$ is

$$\frac{d^2}{dt^2} G(h + t\varphi, u_{h+t\varphi})\Big|_{t=0},$$

where, as usual, $u_{h+t\varphi}$ denotes the elastic equilibrium in $\Omega_{h+t\varphi}$. It turns out (see [Bonacini 2013b, Theorem 4.1]) that

$$\frac{d^2}{dt^2} G(h + t\varphi, u_{h+t\varphi})\Big|_{t=0} = \partial^2 G(h, u_h)[\varphi] - \int_{\Gamma_h} \big(W(E(u_h)) + H^\psi\big) \operatorname{Div}_{\Gamma_h}\left[\left(\frac{(Dh, |Dh|^2)}{\sqrt{1 + |Dh|^2}} \circ \pi\right)\phi^2\right] d\mathcal{H}^2, \tag{4-6}$$

where $\partial^2 G(h, u_h)[\varphi]$ is the (nonlocal) quadratic form defined as

$$\begin{aligned} \partial^2 G(h, u_h)[\varphi] := -2\int_{\Omega_h} W(E(v_\phi))\, dz + \int_{\Gamma_h} D^2\psi(\nu)[D_{\Gamma_h}\phi, D_{\Gamma_h}\phi]\, d\mathcal{H}^2 \\ + \int_{\Gamma_h} \big(\partial_\nu\big[W(E(u_h))\big] - \operatorname{trace}(\mathbb{B}^\psi \mathbb{B})\big)\phi^2\, d\mathcal{H}^2, \end{aligned} \tag{4-7}$$

$$\phi := \frac{\varphi}{\sqrt{1 + |Dh|^2}} \circ \pi,$$

and $v_\phi$ the unique solution in $A(\Omega_h)$ to

$$\int_{\Omega_h} \mathbb{C}E(v_\phi) : E(w)\, dz = \int_{\Gamma_h} \operatorname{Div}_{\Gamma_h}(\phi\, \mathbb{C}E(u_h)) \cdot w\, d\mathcal{H}^2 \quad \text{for all } w \in A(\Omega_h). \tag{4-8}$$

Note that, if $(h, u_h)$ is a critical pair of $G$ (see Definition 3.8 with $\varepsilon = 0$), then the integral in (4-6) vanishes, so that

$$\frac{d^2}{dt^2} G(h + t\varphi, u_{h+t\varphi})\Big|_{t=0} = \partial^2 G(h, u_h)[\varphi].$$

Throughout this section $\alpha$ will denote a fixed number in the interval $(0, 1 - 2/p)$. The next result is a simple consequence of [Bonacini 2013b, Theorem 6.6].

**Theorem 4.1.** *Assume that the surface density $\psi$ is of class $C^3$ away from the origin, it satisfies* (2-1), *and the following convexity condition holds: for every $\xi \in S^2$,*

$$D^2\psi(\xi)[w, w] > 0 \quad \textit{for all } w \perp \xi,\ w \neq 0. \tag{4-9}$$

*If*

$$\partial^2 G(d, u_d)[\varphi] > 0 \quad \textit{for all } \varphi \in \widetilde{H}^1_\#(Q) \setminus \{0\}, \tag{4-10}$$

*then there exists* $\delta > 0$ *such that*

$$G(d, u_d) < G(k, v)$$

*for all* $(k, v) \in X$ *with* $|\Omega_k| = |\Omega_d|$, $0 < \|k - d\|_{C^{1,\alpha}_\#(Q)} \leq \delta$.

*Proof.* By condition (4-10) and [Bonacini 2013b, Theorem 6.6] there exists $\delta_0 > 0$ such that, if $0 < \|k - d\|_{C^1_\#(Q)} \leq \delta_0$ and $\|D\eta\|_\infty \leq 1 + \|Du_d\|_\infty$ with $(k, \eta) \in X$, then

$$G(d, u_d) < G(k, \eta). \tag{4-11}$$

Note that we may choose $0 < \delta < \delta_0$ such that, if $\|k - d\|_{C^{1,\alpha}_\#(Q)} \leq \delta$ and $u_k$ is the elastic equilibrium corresponding to $k$, by elliptic regularity (see also Lemma A.1) we have that $\|Du_k\|_\infty \leq 1 + \|Du_d\|_\infty$. Therefore, using (4-11) with $\eta := u_k$, we may conclude that

$$G(d, u_d) < G(k, u_k) \leq G(k, v),$$

where in the last inequality we used the minimality of $u_k$, and the result follows. □

**Remark 4.2.** It can be shown that Theorem 4.1 continues to hold if (4-9) is replaced by the weaker condition

$$D^2\psi(e_3)[w, w] > 0 \quad \text{for all } w \perp e_3,\ w \neq 0. \tag{4-12}$$

Indeed, (4-12) implies that (4-9) holds for all $\xi$ belonging to a suitable neighborhood $U \subset S^2$ of $e_3$. In turn, by choosing $\delta$ sufficiently small we can ensure that the outer unit normals to $\Gamma_k$ lie in $U$ provided $\|k - d\|_{C^{1,\alpha}_\#(Q)} < \delta$. A careful inspection of the proof of [Bonacini 2013b, Theorem 6.6] shows that, under these circumstances, condition (4-9) is only required to hold at vectors $\xi \in U$.

**Remark 4.3.** Under assumption (4-9), it can be shown that (4-10) is equivalent to having

$$\inf\{\partial^2 G(d, u_d)[\varphi] : \varphi \in \widetilde{H}^1_\#(Q),\ \|\varphi\|_{H^1_\#(Q)} = 1\} =: m_0 > 0 \tag{4-13}$$

(see [Bonacini 2013b, Corollary 4.8]), i.e.,

$$\partial^2 G(d, u_d)[\varphi] \geq m_0 \|\varphi\|^2_{H^1_\#(Q)} \quad \text{for all } \varphi \in \widetilde{H}^1_\#(Q).$$

**Remark 4.4.** Note that, if the profile $h \equiv d$ is flat, then the corresponding elastic equilibrium $u_d$ is affine. It immediately follows that $(d, u_d)$ is a critical pair in the sense of Definition 3.8.

We now consider the case of a nonconvex surface energy density $\psi$, and introduce the "relaxed" functional defined for all $(h, u) \in X$ as

$$\overline{G}(h, u) := \int_{\Omega_h} W(E(u))\, dz + \int_{\Gamma_h} \psi^{**}(\nu)\, d\mathcal{H}^2, \tag{4-14}$$

where $\psi^{**}$ is the convex envelope of $\psi$. It turns out that, if the boundary of the Wulff shape $W_\psi$ associated with the nonconvex density $\psi$ contains a flat horizontal facet, then the flat configuration is always an isolated volume-constrained local minimizer, irrespective of the value of $d$. We recall that the Wulff shape $W_\psi$ is given by

$$W_\psi := \{z \in \mathbb{R}^3 : z \cdot \nu < \psi(\nu) \text{ for all } \nu \in S^2\}$$

(see [Fonseca 1991, Definition 3.1]). The following result can be easily obtained from [Bonacini 2013b, Theorem 7.5 and Remark 7.6] arguing as in the last part of the proof of Theorem 4.1.

**Theorem 4.5.** *Let $\psi : \mathbb{R}^3 \to [0, +\infty)$ be a Lipschitz positively one-homogeneous function satisfying* (2-1), *and let $\{(x, y) \in \mathbb{R}^3 : |x| \le \alpha, y = \beta\} \subset \partial W_\psi$ for some $\alpha, \beta > 0$. Then there exists $\delta > 0$ such that*

$$\overline{G}(d, u_d) < \overline{G}(k, v)$$

*for all $(k, v) \in X$ with $|\Omega_k| = |\Omega_d|$, $0 < \|k - d\|_{C^{1,\alpha}_\#(Q)} \le \delta$.*

In the next two subsections we use the previous theorems to study the Liapunov stability of the flat configuration both in the convex and nonconvex case.

**Definition 4.6.** We say that the flat configuration $(d, u_d)$ is *Liapunov stable* if, for every $\sigma > 0$, there exists $\delta(\sigma) > 0$ such that, if $(h_0, u_0) \in X$ with $|\Omega_{h_0}| = |\Omega_d|$ and $\|h_0 - d\|_{W^{2,p}_\#(Q)} \le \delta(\sigma)$, then every variational solution $h$ to (3-1) according to Definition 3.17, with initial datum $h_0$, exists for all times, and $\|h(\cdot, t) - d\|_{W^{2,p}_\#(Q)} \le \sigma$ for all $t > 0$.

**4A.** ***The case of a nonconvex surface density.*** In this subsection will show that, if the boundary of the Wulff shape $W_\psi$ associated with $\psi$ contains a flat horizontal facet, then the flat configuration is always Liapunov stable.

**Theorem 4.7.** *Let $\psi : \mathbb{R}^3 \to [0, +\infty)$ be a positively one-homogeneous function of class $C^2$ away from the origin such that* (2-1) *holds, and let $\{(x, y) \in \mathbb{R}^3 : |x| \le \alpha, y = \beta\} \subset \partial W_\psi$ for some $\alpha, \beta > 0$. Then for every $d > 0$ the flat configuration $(d, u_d)$ is Liapunov stable* (*according to Definition 4.6*).

*Proof.* We start by observing that, from the assumptions on $\psi$, $e_3$ is normal to boundary $\partial W_\psi$ of the Wulff shape $W_\psi$ associated with $\psi$. Thus, by [Fonseca 1991, Proposition 3.5(iv)], it follows that $\psi(e_3) = \psi^{**}(e_3)$. In turn, by Theorem 4.5, we may find $\delta > 0$ such that

$$F(d, u_d) = \overline{G}(d, u_d) < \overline{G}(k, v) \le F(k, v) \tag{4-15}$$

for all $(k, v) \in X$ with $|\Omega_k| = |\Omega_d|$ and $0 < \|k - d\|_{C^{1,\alpha}_\#(Q)} \le \delta$. Fix $\sigma > 0$ and choose $\delta_0 \in \big(0, \min\{\delta, \frac{1}{2}\sigma\}\big)$ so small that

$$\|h - d\|_{C^{1,\alpha}_\#(Q)} \le \delta_0 \implies \|Dh\|_\infty < \Lambda_0, \tag{4-16}$$

where $\Lambda_0$ is as in (2-6). For every $\tau > 0$, set

$$\omega(\tau) := \sup\{\|k - d\|_{C^{1,\alpha}_\#(Q)}\},$$

where the supremum is taken over all $(k, v) \in X$ such that

$$|\Omega_k| = |\Omega_d|, \quad \|k - d\|_{C^{1,\alpha}_\#(Q)} \le \delta, \quad \text{and} \quad F(k, v) - F(d, u_d) < \tau.$$

Clearly, $\omega(\tau) > 0$ for $\tau > 0$. We claim that $\omega(\tau) \to 0$ as $\tau \to 0^+$. Indeed, to see this we assume by contradiction that there exists a sequence $(k_n, v_n) \in X$ with $|\Omega_{k_n}| = |\Omega_d|$ such that

$$\liminf_n F(k_n, v_n) \le F(d, u_d) \quad \text{and} \quad 0 < c_0 \le \|k_n - d\|_{C^{1,\alpha}_\#(Q)} \le \delta \tag{4-17}$$

for some $c_0 > 0$. By Lemma A.3, up to a subsequence we may assume that $k_n \rightharpoonup k$ in $W^{2,p}_\#(Q)$ and that $v_n \rightharpoonup v$ in $H^1_{\text{loc}}(\Omega_k; \mathbb{R}^3)$ for some $(k, v) \in X$ satisfying $\delta \ge \|k - d\|_{C^{1,\alpha}_\#(Q)} \ge c_0$, since $W^{2,p}_\#(Q)$ is compactly embedded in $C^{1,\alpha}_\#(Q)$. By lower semicontinuity we also have that

$$F(k, v) \le \liminf_n F(k_n, v_n) \le F(d, u_d),$$

which contradicts (4-15).

Choose $\delta(\sigma)$ so small that, if $\|h_0 - d\|_{W^{2,p}_\#(Q)} \le \delta(\sigma)$, then

$$\|h_0 - d\|_{C^{1,\alpha}_\#(Q)} < \delta_0 \quad \text{and} \quad F(h_0, u_0) - F(d, u_d) \le \omega^{-1}\big(\tfrac{1}{2}\delta_0\big),$$

where $\omega^{-1}$ is the generalized inverse of $\omega$ defined as $\omega^{-1}(s) := \sup\{\tau > 0 : \omega(\tau) \le s\}$ for all $s > 0$. Note that, since $\omega(\tau) > 0$ for $\tau > 0$ and $\omega(\tau) \to 0$ as $\tau \to 0+$, we have that $\omega^{-1}(s) \to 0$ as $s \to 0+$. Let $h$ be a variational solution as in Theorem 3.4 (see Definition 3.17). Let

$$T_1 := \sup\{t > 0 : \|h(\cdot, s) - d\|_{C^{1,\alpha}_\#(Q)} \le \delta_0 \ \text{ for all } \ s \in (0, t)\}.$$

Note that, by Theorem 3.5, $T_1 > 0$. We claim that $T_1 = +\infty$. Indeed, if $T_1$ were finite, then, recalling (3-7), we would get, for all $s \in [0, T_1]$,

$$F(h(\cdot, T_1), u_{h(\cdot, T_1)}) - F(d, u_d) \le F(h_0, u_0) - F(d, u_d) \le \omega^{-1}\big(\tfrac{1}{2}\delta_0\big), \tag{4-18}$$

which implies $\|h(\cdot, T_1) - d\|_{C^{1,\alpha}_\#(Q)} \le \frac{1}{2}\delta_0$ by the definition of $\omega$. Then, (4-16), Remark 3.6, and Theorem 3.5 would imply that there exists $T > T_1$ such that $\|h(\cdot, t) - d\|_{C^{1,\alpha}_\#(Q)} \le \delta_0$ for all $t \in (T_1, T)$, thus giving a contradiction. We conclude that $T_1 = +\infty$ and that $\|h(\cdot, t) - d\|_{C^{1,\alpha}_\#(Q)} \le \delta_0$ for all $t > 0$. Therefore, (4-16) implies that $\|Dh(\cdot, t)\|_\infty < \Lambda_0$ for all times, which, together with Remark 3.6, gives that $h$ is a solution to (3-1) for all times. Moreover, by (4-18) we have also shown that $F(h(\cdot, t), u_{h(\cdot, t)}) - F(d, u_d) \le \omega^{-1}\big(\frac{1}{2}\delta_0\big)$ for all $t > 0$, which by (4-15) implies that

$$\varepsilon \int_{\Gamma_{h(\cdot, t)}} |H|^p \, d\mathcal{H}^2 \le \omega^{-1}\big(\tfrac{1}{2}\delta_0\big).$$

Using elliptic regularity (see (2-3)), this inequality and the fact that $\|h(\cdot, t) - d\|_\infty \le \frac{1}{2}\sigma$ for all $t > 0$ imply that $\|h(\cdot, t) - d\|_{W^{2,p}_\#(Q)} \le \sigma$ provided that $\delta_0$ and, in turn, $\delta(\sigma)$ are chosen sufficiently small. □

**4B.** ***The case of a convex surface density.*** In this section we will show that, under the convexity assumption (4-9), the condition $\partial^2 G(d, u_d) > 0$ implies that $(d, u_d)$ is asymptotically stable for the regularized evolution equation (3-1) (see Theorem 4.14 below). We start by addressing the Liapunov stability (see Definition 4.6).

**Theorem 4.8.** *Assume that the surface density $\psi$ satisfies the assumptions of Theorem 4.1 and that the flat configuration $(d, u_d)$ satisfies* (4-10)*. Then $(d, u_d)$ is Liapunov stable.*

*Proof.* Since (4-15) still holds with $\bar{G}$ replaced by $G$ in view of Theorem 4.1, we can conclude as in the proof of Theorem 4.7. □

**Remark 4.9** (stability of the flat configuration for small volumes). If the surface density $\psi$ satisfies the assumptions of Theorem 4.1, then there exists $d_0 > 0$ (depending only on the Dirichlet boundary datum $w_0$) such that (4-10) holds for all $d \in (0, d_0)$ (see [Bonacini 2013b, Proposition 7.3]).

**Definition 4.10.** We say that a flat configuration $(d, u_d)$ is *asymptotically stable* if there exists $\delta > 0$ such that, if $(h_0, u_0) \in X$ with $|\Omega_{h_0}| = |\Omega_d|$ and $\|h_0 - d\|_{W^{2,p}_{\#}(Q)} \le \delta$, then every variational solution $h$ to (3-1) according to Definition 3.17, with initial datum $h_0$, exists for all times and $\|h(\cdot\,, t) - d\|_{W^{2,p}_{\#}(Q)} \to 0$ as $t \to +\infty$.

We start by showing that, if a variational solution to (3-1) exists for all times, then there exists a sequence $\{t_n\} \subset (0, +\infty)$, with $t_n \to \infty$, such that $h(\cdot\,, t_n)$ converges to a critical profile (see Definition 3.8).

**Proposition 4.11.** *Assume that for a certain initial datum $h_0 \in W^{2,p}_{\#}(Q)$ there exists a global-in-time variational solution $h$. Then there exists a sequence $\{t_n\} \subset (0, +\infty) \setminus Z_0$, where $Z_0$ is the set in* (3-54)*, and a critical profile $\bar{h}$ for $F$ such that $t_n \to \infty$ and $h(\cdot\,, t_n) \to \bar{h}$ strongly in $W^{2,p}_{\#}(Q)$.*

*Proof.* From (3-3), by lower semicontinuity we have that

$$\int_0^{\infty} \left\| \frac{\partial h}{\partial t} \right\|^2_{H^{-1}(Q)} dt \le C F(h_0, u_0).$$

Since the set $Z_0$ has measure zero, we may find a sequence $\{t_n\} \subset (0, +\infty) \setminus Z_0$, $t_n \to \infty$, such that $\|\partial h(\cdot\,, t_n)/\partial t\|_{H^{-1}(Q)} \to 0$. Since $h \in L^{\infty}(0, \infty; W^{2,p}_{\#}(Q)) \cap H^1(0, \infty; H^{-1}_{\#}(Q))$, setting $h^n = h(\cdot\,, t_n)$ we may also assume that there exists $\bar{h} \in W^{2,p}_{\#}(Q)$ such that $h^n \rightharpoonup \bar{h}$ weakly in $W^{2,p}_{\#}(Q)$. In turn, denoting by $u_{h^n}$ the corresponding elastic equilibria, by elliptic regularity (see also Lemma A.1 ) we have that $u_{h^n}(\cdot\,, h^n(\cdot)) \to u_{\bar{h}}(\cdot\,, \bar{h}(\cdot))$ in $C^{1,\alpha}_{\#}(Q; \mathbb{R}^3)$. Let $\hat{v}^n$ be the unique $Q$-periodic solution to (3-59) with $t = t_n$ and note that $\hat{v}^n \to 0$ in $H^1_{\#}(Q)$, since $\|\partial h(\cdot\,, t_n)/\partial t\|_{H^{-1}(Q)} \to 0$. Writing the equation satisfied by $h^n$ as in (3-22), we have, for all $\varphi \in C^2_{\#}(Q)$ with $\int_Q \varphi\, dx = 0$,

$$\begin{aligned}\int_Q W\big(E\big(u_{h^n}(x, h^n(x))\big)\big)\varphi\, dx &+ \int_Q D\psi(-Dh^n, 1)\cdot(-D\varphi, 0)\, dx + \frac{\varepsilon}{p}\int_Q |H^n|^p \frac{Dh^n \cdot D\varphi}{J^n} \\ &- \varepsilon \int_Q |H^n|^{p-2} H^n \bigg[\Delta\varphi - \frac{D^2\varphi[Dh^n, Dh^n]}{(J^n)^2} - \frac{\Delta h^n\, Dh^n \cdot D\varphi}{(J^n)^2} \\ &\qquad - 2\frac{D^2 h^n[Dh^n, D\varphi]}{(J^n)^2} + 3\frac{D^2 h^n[Dh^n, Dh^n] Dh^n \cdot D\varphi}{(J^n)^4}\bigg]\, dx \\ &\qquad - \int_Q \hat{v}^n \varphi\, dx = 0, \end{aligned} \tag{4-19}$$

where $H^n$ stands for the sum of the principal curvatures of $h^n$ and $J^n = \sqrt{1+|Dh^n|^2}$. Arguing exactly as in the proof of Theorem 3.11 (see (3-30)), we deduce that

$$\int_Q |D^2(|H^n|^{p-2}H^n)|^2\,dx \le C\int_Q (1+(\hat{v}^n)^2)\,dx \tag{4-20}$$

for some constant $C$ independent of $n$. Thus, passing to a subsequence, if necessary, we may also assume that there exists $w \in H^2_\#(Q)$ such that $|H^n|^{p-2}H^n \rightharpoonup w$ weakly in $H^2_\#(Q)$ and $|H^n|^{p-2}H^n \to w$ strongly in $H^1_\#(Q)$. Since $H^1_\#(Q)$ is continuously embedded in $L^q(Q)$ for every $1 \le q < \infty$ by the Sobolev embedding theorem, there exists $z \in L^1(Q)$ such that $|H^n|^p \to z$ in $L^1(Q)$. The same argument used at the end of the proof of Corollary 3.15 shows that $z = |\overline{H}|^p$ and $w = |\overline{H}|^{p-2}\overline{H}$, where $\overline{H}$ is the sum of the principal curvatures of $\bar{h}$.

Using all the convergences proved above, and arguing as in the proof of Theorem 3.16, we may pass to the limit in (4-19), thus getting that $\bar{h}$ is a critical profile by Remark 3.12. □

**Lemma 4.12.** *Assume that* (4-9) *and* (4-10) *hold. Then there exist* $\sigma > 0$ *and* $c_0 > 0$ *such that*

$$\partial^2 G(h,u_h)[\varphi] \ge c_0\|\varphi\|^2_{H^1_\#(Q)} \quad \textit{for all } \varphi \in \widetilde{H}^1_\#(Q)$$

*provided* $\|h-d\|_{C^{2,\alpha}_\#(Q)} \le \sigma$, *where* $\widetilde{H}^1_\#(Q)$ *is defined in* (4-5).

*Proof.* Throughout this proof, with a slight abuse of notation, we denote by $\mathbb{C}$ the tensor acting on a generic $3\times 3$ matrix $M$ as $\mathbb{C}M := \mathbb{C}(M+M^T)/2$. Let $m_0$ be the positive constant defined in (4-13). We claim that there exists $\sigma > 0$ such that

$$\inf\{\partial^2 G(h,u_h)[\varphi] : \varphi \in \widetilde{H}^1_\#(Q),\ \|\varphi\|_{H^1_\#(Q)} = 1\} \ge \tfrac{1}{2}m_0$$

whenever $\|h-d\|_{C^{2,\alpha}_\#(Q)} \le \sigma$. Indeed, if not, then there exist two sequences $\{h_n\} \subset C^{2,\alpha}_\#(Q)$ with $h_n \to d$ in $C^{2,\alpha}_\#(Q)$ and $\{\varphi_n\} \subset \widetilde{H}^1_\#(Q)$ with $\|\varphi_n\|_{H^1_\#(Q)} = 1$ such that

$$\partial^2 G(h_n,u_{h_n})[\varphi_n] < \tfrac{1}{2}m_0. \tag{4-21}$$

Set

$$\phi_n := \frac{\varphi_n}{\sqrt{1+|Dh_n|^2}} \circ \pi, \tag{4-22}$$

where we recall that $\pi(x,y) = x$. Let $v_{\phi_n}$ be the unique solution in $A(\Omega_{h_n})$ — see (4-4) — to

$$\int_{\Omega_{h_n}} \mathbb{C}E(v_{\phi_n}) : E(w)\,dz = \int_{\Gamma_{h_n}} \mathrm{Div}_{\Gamma_{h_n}}(\phi_n \mathbb{C}E(u_{h_n}))\cdot w\,d\mathcal{H}^2 \quad \text{for all } w \in A(\Omega_{h_n}) \tag{4-23}$$

and let $v_{\varphi_n}$ be the unique solution in $A(\Omega_d)$ to

$$\int_{\Omega_d} \mathbb{C}E(v_{\varphi_n}) : E(w)\,dz = \int_{\Gamma_d} \mathrm{Div}_{\Gamma_d}(\varphi_n \mathbb{C}E(u_d))\cdot w\,d\mathcal{H}^2 \quad \text{for all } w \in A(\Omega_d). \tag{4-24}$$

Observe that (see, e.g., Lemma A.1)

$$\|\mathrm{Div}_{\Gamma_{h_n}}(\phi_n\,\mathbb{C}E(u_{h_n}))\|_{L^2(\Gamma_{h_n})} \le C\|\varphi_n\|_{H^1_\#(Q)}$$

for some constant $C > 0$ depending only on

$$\sup_n \big( \|\mathbb{C}E(u_{h_n})\|_{C^1(\Gamma_{h_n})} + \|h_n\|_{C^2_\#(Q)} \big)$$

and thus independent of $n$. Therefore, choosing $w = v_{\phi_n}$ in (4-23), and using Korn's inequality, we deduce that

$$\sup_n \|v_{\phi_n}\|_{H^1(\Omega_{h_n})} < +\infty. \tag{4-25}$$

The same bound holds for the sequence $\{v_{\varphi_n}\}$.

Next we show that

$$\int_{\Omega_{h_n}} W(E(v_{\phi_n}))\,dz - \int_{\Omega_d} W(E(v_{\varphi_n}))\,dz \longrightarrow 0 \tag{4-26}$$

as $n \to \infty$. Consider a sequence $\{\Phi_n\}$ of diffeomorphisms $\Phi_n : \Omega_d \to \Omega_{h_n}$ such that $\Phi_n - \mathrm{Id}$ is $Q$-periodic with respect to $x$, $\Phi_n(x, y) = (x, y + d - h_n(x))$ in a neighborhood of $\Gamma_d$, and $\|\Phi_n - \mathrm{Id}\,\|_{C^{2,\alpha}(\overline{\Omega}_d;\mathbb{R}^3)} \le C\|h_n - d\|_{C^{2,\alpha}_\#(Q)} \to 0$. Set $w_n := v_{\phi_n} \circ \Phi_n$. Changing variables, we get that $w_n \in A(\Omega_d)$ satisfies

$$\int_{\Omega_d} A_n Dw_n : Dw\,dz = \int_{\Gamma_d} \big(\mathrm{Div}_{\Gamma_{h_n}}(\phi_n \mathbb{C}E(u_{h_n})) \circ \Phi_n\big) \cdot w J_{\Phi_n}\,d\mathcal{H}^2 \tag{4-27}$$

for every $w \in A(\Omega_d)$, where $J_{\Phi_n}$ stands for the $(N-1)$-Jacobian of $\Phi_n$ and the fourth-order tensor-valued functions $A_n$ satisfy $A_n \to \mathbb{C}$ in $C^{1,\alpha}(\overline{\Omega}_d)$. We claim that

$$\int_{\Omega_d} W(E(w_n - v_{\varphi_n}))\,dz \longrightarrow 0 \tag{4-28}$$

as $n \to \infty$. Note that this would immediately imply $\int_{\Omega_d} W(E(w_n))\,dz - \int_{\Omega_d} W(E(v_{\varphi_n}))\,dz \to 0$ and, in turn, taking also into account that $A_n \to \mathbb{C}$ uniformly and that $\frac{1}{2}\int_{\Omega_d} A_n Dw_n : Dw_n\,dz = \int_{\Omega_{h_n}} W(E(v_{\phi_n}))\,dz$, claim (4-26) would follow. In order to prove (4-28), we write

$$\begin{aligned}
&\int_{\Omega_d} \mathbb{C}D(v_{\varphi_n} - w_n) : D(v_{\varphi_n} - w_n)\,dz \\
&\quad = \int_{\Omega_d} \mathbb{C}Dv_{\varphi_n} : D(v_{\varphi_n} - w_n)\,dz - \int_{\Omega_d} (\mathbb{C} - A_n)Dw_n : D(v_{\varphi_n} - w_n)\,dz - \int_{\Omega_d} A_n Dw_n : D(v_{\varphi_n} - w_n)\,dz \\
&\quad = \int_{\Gamma_d} \mathrm{Div}_{\Gamma_d}(\varphi_n \mathbb{C}E(u_d)) \cdot (v_{\varphi_n} - w_n)\,d\mathcal{H}^2 - \int_{\Omega_d} (\mathbb{C} - A_n)Dw_n : D(v_{\varphi_n} - w_n)\,dz \\
&\qquad\qquad - \int_{\Gamma_d} \big(\mathrm{Div}_{\Gamma_{h_n}}(\phi_n \mathbb{C}E(u_{h_n})) \circ \Phi_n\big) \cdot (v_{\varphi_n} - w_n) J_{\Phi_n}\,d\mathcal{H}^2 \\
&\quad =: I_1 - I_2 - I_3,
\end{aligned}$$

where we used (4-24) and (4-27). From (4-25), the analogous bound for the sequence $\{v_{\varphi_n}\}$, and the uniform convergence of $A_n$ to $\mathbb{C}$ we deduce that $I_2$ tends to 0.

Fix $\eta=(\eta_1,\eta_2,\eta_3)\in C^1_\#(\Gamma_d;\mathbb{R}^3)\simeq C^1_\#(Q;\mathbb{R}^3)$. Using the fact that $\Phi_n^{-1}(x,y)=(x,y-h_n(x)+d)$ in a neighborhood of $\Gamma_{h_n}$, we have

$$D_{\Gamma_{h_n}}(\eta_j\circ\Phi_n^{-1})=(I-\nu_{h_n}\otimes\nu_{h_n})D_{\Gamma_d}\eta_j\circ\Phi_n^{-1},$$

where we set $\nu_{h_n}:=(-Dh_n,1)/\sqrt{1+|Dh_n|^2}$. Using this fact, we then have, by repeated integrations by parts and changes of variables,

$$\begin{aligned}\int_{\Gamma_d}\big(\mathrm{Div}_{\Gamma_{h_n}}(\phi_n\mathbb{C}E(u_{h_n}))\circ\Phi_n\big)\cdot\eta\, J_{\Phi_n}\,d\mathcal{H}^2 &= \int_{\Gamma_{h_n}}\mathrm{Div}_{\Gamma_{h_n}}(\phi_n\mathbb{C}E(u_{h_n}))\cdot\eta\circ\Phi_n^{-1}\,d\mathcal{H}^2\\ &=-\int_{\Gamma_{h_n}}\phi_n\mathbb{C}E(u_{h_n}):D_{\Gamma_{h_n}}(\eta\circ\Phi_n^{-1})\,d\mathcal{H}^2\\ &=-\int_{\Gamma_{h_n}}(I-\nu_{h_n}\otimes\nu_{h_n})\phi_n\mathbb{C}E(u_{h_n}):D_{\Gamma_d}\eta\circ\Phi_n^{-1}\;d\mathcal{H}^2\\ &=-\int_{\Gamma_d}[(I-\nu_{h_n}\otimes\nu_{h_n})\phi_n\mathbb{C}E(u_{h_n})]\circ\Phi_n:D_{\Gamma_d}\eta\,J_{\Phi_n}\,d\mathcal{H}^2\\ &=\int_{\Gamma_d}\mathrm{Div}_{\Gamma_d}\big[[(I-\nu_{h_n}\otimes\nu_{h_n})\phi_n\mathbb{C}E(u_{h_n})]\circ\Phi_n J_{\Phi_n}\big]\cdot\eta\,d\mathcal{H}^2.\end{aligned}$$

Hence, we may rewrite

$$I_1-I_3=\int_{\Gamma_d}\mathrm{Div}_{\Gamma_d}\,g_n\cdot(v_{\varphi_n}-w_n)\,d\mathcal{H}^2, \tag{4-29}$$

where, by (4-22),

$$\begin{aligned}g_n&:=\varphi_n\mathbb{C}E(u_d)-[(I-\nu_{h_n}\otimes\nu_{h_n})\phi_n\mathbb{C}E(u_{h_n})]\circ\Phi_n J_{\Phi_n}\\ &=\varphi_n\bigg[\mathbb{C}E(u_d)-[(I-\nu_{h_n}\otimes\nu_{h_n})\mathbb{C}E(u_{h_n})]\circ\Phi_n\frac{J_{\Phi_n}}{\sqrt{1+|Dh_n|^2}}\bigg].\end{aligned}$$

Since $h_n\to d$ in $C^{2,\alpha}_\#(Q)$, by standard Schauder estimates for the elastic displacements $u_{h_n}$ we get

$$\mathbb{C}E(u_d)-[(I-\nu_{h_n}\otimes\nu_{h_n})\mathbb{C}E(u_{h_n})]\circ\Phi_n\frac{J_{\Phi_n}}{\sqrt{1+|Dh_n|^2}}\longrightarrow 0\quad\text{in } C^{1,\alpha}(\Gamma_d).$$

Therefore, by (4-29) and the equiboundedness of $\{v_{\phi_n}\}$ and $\{w_n\}$, we have that $I_1-I_3\to 0$. This concludes the proof of (4-28) and, in turn, of (4-26).

Finally, again from the $C^{2,\alpha}$-convergence of $\{h_n\}$ to $d$ and the fact that

$$\partial_\nu\big[W(E(u_{h_n})\big]\circ\Phi_n\longrightarrow\partial_\nu\big[W(E(u_d))\big]\quad\text{in } C^{0,\alpha}_\#(\Gamma_d)$$

by standard Schauder elliptic estimates, recalling (4-7) we easily infer that

$$\bigg(\partial^2G(h_n,u_{h_n})[\varphi_n]+2\int_{\Omega_{h_n}}W(E(v_{\phi_n}))\,dz\bigg)-\bigg(\partial^2G(d,u_d)[\varphi_n]+2\int_{\Omega_d}W(E(v_{\varphi_n}))\,dz\bigg)\longrightarrow 0 \tag{4-30}$$

as $n\to\infty$. Thus, recalling (4-26), we also have

$$\partial^2G(h_n,u_{h_n})[\varphi_n]-\partial^2G(d,u_d)[\varphi_n]\longrightarrow 0$$

and, in turn, by (4-21)

$$\limsup \partial^2 G(d, u_d)[\varphi_n] \le \tfrac{1}{2} m_0,$$

which is a contradiction to (4-13). This concludes the proof of the lemma. □

Next we prove that $(d, u_d)$ is an isolated critical pair.

**Proposition 4.13.** *Assume that* (4-9) *and* (4-10) *hold. Then there exists* $\sigma > 0$ *such that, if* $(h, u_h) \in X$ *with* $|\Omega_h| = |\Omega_d|$ *and* $0 < \|h - d\|_{W^{2,p}_\#(Q)} \le \sigma$, *then* $(h, u_h)$ *is not a critical pair.*

*Proof.* Assume by contradiction that there exists a sequence $h_n \to d$ in $W^{2,p}_\#(Q)$ with $h_n \ne d$ and $|\Omega_{h_n}| = |\Omega_d|$ such that $(h_n, u_{h_n})$ is a critical pair. Using the Euler–Lagrange equation and arguing as in the proof of Theorem 3.11, one can show that

$$\int_Q \big|D^2(|H_n|^{p-2}H_n)\big|^2\,dx \le C \int_Q \big(|D^2 h_n|^2 |D(|H_n|^{p-2}H_n)|^2 + |H_n|^{2(p+1)} + 1\big)\,dx.$$

Indeed, this can obtained in the same way as (3-29), taking into account that there is no contribution from the time derivative. From this inequality, arguing exactly as in the final part of the proof of Theorem 3.11 we deduce that

$$\int_Q \big|D^2(|H_n|^{p-2}H_n)\big|^2\,dx \le C$$

for some $C$ independent of $n$. In particular, by the Sobolev embedding theorem, $\{|H_n|^{p-2}H_n\}$ is bounded in $C^{0,\beta}_\#(Q)$ for every $\beta \in (0,1)$. Hence, $\{H_n\}$ is bounded in $C^{0,\beta}_\#(Q)$ for all $\beta \in (0, 1/(p-1))$. In turn, by (2-3) and standard elliptic regularity this implies that $\{h_n\}$ is bounded in $C^{2,\beta}_\#(Q)$ for all $\beta \in (0, 1/(p-1))$ and thus $h_n \to d$ in $C^{2,\beta}(Q)$ for all such $\beta$. Since $(d, u_d)$ is a critical pair (see Remark 4.4),

$$\frac{d}{ds} F(d + s(h_n - d), u_{d+s(h_n-d)})\Big|_{s=0} = 0,$$

and so by (4-6) to reach a contradiction it is enough to show that, for $n$ large,

$$\begin{aligned}
&\frac{d^2}{ds^2} F(d + s(h_n - d), u_{d+s(h_n-d)})\Big|_{s=t} \\
&\quad = \partial^2 G(h_{n,t}, u_{h_{n,t}})[h_n - d] \\
&\qquad - \int_{\Gamma_{h_{n,t}}} \big(W(E(u_{h_{n,t}})) + H^{\psi}_{h_{n,t}}\big)\,\mathrm{Div}_{\Gamma_{h_{n,t}}}\left(\frac{(Dh_{n,t}, |Dh_{n,t}|^2)(h_{n,t} - d)^2}{(1 + |Dh_{n,t}|)^{\frac{3}{2}}} \circ \pi\right) d\mathcal{H}^2 \\
&\qquad\quad + \varepsilon \frac{d^2}{ds^2} \mathcal{W}_p(d + s(h_n - d))\Big|_{s=t} > 0
\end{aligned}$$

for all $t \in (0,1)$, where $h_{n,t} := d + t(h_n - d)$, $H^{\psi}_{h_{n,t}}$ is defined as in (4-3) with $h$ replaced by $h_{n,t}$, and

$$\mathcal{W}_p(h) := \int_{\Gamma_h} |H|^p\,d\mathcal{H}^2.$$

To this purpose note that, since $h_n \to d$ in $C^{2,\beta}$, by Lemma A.1 we have

$$\sup_{t\in(0,1)} \|W(E(u_{h_{n,t}})) + H^{\psi} - W_d\|_{L^\infty(\Gamma_{h_{n,t}})} \to 0$$

as $n \to \infty$, where $W_d$ is the constant value of $W(E(u_d))$ on $\Gamma_d$ (see Remark 4.4). Therefore, also by Lemma 4.12, we deduce that

$$\begin{aligned}
&\partial^2 G(h_{n,t}, u_{h_{n,t}})[h_n - d] - \int_{\Gamma_{h_{n,t}}} \big(W(E(u_{h_{n,t}})) + H^{\psi}_{h_{n,t}}\big) \operatorname{Div}_{\Gamma_{h_{n,t}}} \left( \frac{(Dh_{n,t}, |Dh_{n,t}|^2)(h_{n,t}-d)^2}{(1+|Dh_{n,t}|)^{\frac{3}{2}}} \circ \pi \right) d\mathcal{H}^2 \\
&\quad = \partial^2 G(h_{n,t}, u_{h_{n,t}})[h_n - d] \\
&\qquad - \int_{\Gamma_{h_{n,t}}} \big(W(E(u_{h_{n,t}})) + H^{\psi}_{h_{n,t}} - W_d\big) \operatorname{Div}_{\Gamma_{h_{n,t}}} \left( \frac{(Dh_{n,t}, |Dh_{n,t}|^2)(h_{n,t}-d)^2}{(1+|Dh_{n,t}|)^{\frac{3}{2}}} \circ \pi \right) d\mathcal{H}^2 \\
&\quad \geq c_0 \|h_n - d\|^2_{H^1_\#(Q)} - C \|W(E(u_{h_{n,t}})) + H^{\psi}_{h_{n,t}} - W_d\|_{L^\infty(\Gamma_{h_{n,t}})} \|h_n - d\|^2_{H^1_\#(Q)} \geq \tfrac{1}{2} c_0 \|h_n - d\|^2_{H^1_\#(Q)}
\end{aligned}$$

for $n$ large and for some constant $c_0 > 0$ independent of $n$, where we used the facts that

$$\int_{\Gamma_{h_{n,t}}} \left\| \operatorname{Div}_{\Gamma_{h_{n,t}}} \left( \frac{(Dh_{n,t}, |Dh_{n,t}|^2)(h_{n,t}-d)^2}{(1+|Dh_{n,t}|)^{\frac{3}{2}}} \circ \pi \right) \right\| d\mathcal{H}^2 \leq C \|h_n\|_{C^2_\#(Q)} \|h_n - d\|^2_{H^1_\#(Q)}$$

and that $h_n \to d$ in $C^{2,\beta}(Q)$.

Since

$$\mathcal{W}_p(d + t(h_n - d)) = t^p \int_Q \left| \operatorname{Div} \frac{Dh_n}{\sqrt{1 + t^2|Dh_n|^2}} \right|^p dx =: f_n(t),$$

in order to conclude it is enough to show that $f_n''(t) \geq 0$ for all $t \in (0,1)$. Set

$$g_n(x,t) := \left| \operatorname{Div} \frac{Dh_n(x)}{\sqrt{1 + t^2|Dh_n(x)|^2}} \right|^2$$

so that

$$f_n'' = \int_Q \Big[ p(p-1) t^{p-2} g_n^{\frac{p}{2}} + p^2 t^{p-1} g_n^{\frac{p-2}{2}} \partial_t g_n + \tfrac{p}{2} t^p \Big( \big(\tfrac{p}{2} - 1\big) g_n^{\frac{p-4}{2}} (\partial_t g_n)^2 + g_n^{\frac{p-2}{2}} \partial_{tt} g_n \Big) \Big] dx. \quad (4\text{-}31)$$

On the other hand, observe that

$$g_n = \frac{|\Delta h_n|^2}{1 + t^2|Dh_n|^2} + t^4 \frac{|D^2 h_n[Dh_n, Dh_n]|^2}{(1 + t^2|Dh_n|^2)^3} - 2t^2 \frac{D^2 h_n[Dh_n, Dh_n]\Delta h_n}{(1 + t^2|Dh_n|^2)^2}$$

so that, for $n$ large,

$$g_n \geq \tfrac{1}{2}|\Delta h_n|^2 - C|D^2 h_n|^2 |Dh_n|^2 \quad \text{and} \quad |\partial_t g_n| + |\partial_{tt} g_n| \leq C|D^2 h_n|^2 |Dh_n|^2.$$

We then deduce from (4-31) that there exist $C_0, C_1 > 0$ independent of $n$ and $t \in (0,1)$ such that

$$f_n''(t) \geq C_0 \int_Q |\Delta h_n|^p \, dx - C_1 \|Dh_n\|^p_\infty \int_Q |D^2 h_n|^p \, dx.$$

Since $\|Dh_n\|_\infty \to 0$, by Lemma A.3 we conclude that the right-hand side in the above inequality is nonnegative for $n$ large, thus concluding the proof of the proposition. □

Finally, we prove the main result of this section, namely, the asymptotic stability of the flat configuration (see Definition 4.10).

**Theorem 4.14.** *Under the assumptions of Theorem 4.8, $(d, u_d)$ is asymptotically stable.*

*Proof.* By Proposition 4.13 there exists $\sigma > 0$ such that, if $h$ is a critical profile with $|\Omega_h| = |\Omega_d|$ and $\|h-d\|_{W^{2,p}_\#(Q)} \leq \sigma$, then $h = d$. In view of Theorem 4.1 we may take $\sigma$ so small that

$$F(d,u_d) < F(k,u_k) \quad \text{for all } (k,u_k) \in X \text{ with } 0 < \|k-d\|_{W^{2,p}_\#(Q)} \leq \sigma. \tag{4-32}$$

Since $(d, u_d)$ is Liapunov stable by Theorem 4.8, for every fixed $(h_0, u_0) \in X$ with $|\Omega_{h_0}| = |\Omega_d|$ and $\|h_0 - d\|_{W^{2,p}_\#(Q)} \leq \delta(\sigma)$, we have

$$\|h(\cdot\,,t) - d\|_{W^{2,p}_\#(Q)} \leq \sigma \quad \text{for all } t > 0. \tag{4-33}$$

Here $\delta(\sigma)$ is the number given in Definition 4.6. We claim that

$$F(h(\cdot\,,t), u_h(\cdot\,,t)) \longrightarrow F(d,u_d) \quad \text{as } t \to +\infty. \tag{4-34}$$

By Proposition 4.11 there exists a sequence $\{t_n\} \subset (0,+\infty) \setminus Z_0$ such that $t_n \to +\infty$ and $\{h(\cdot\,,t_n)\}$ converges to a critical profile in $W^{2,p}_\#(Q)$, where $Z_0$ is the set in (3-54). In view of the choice of $\sigma$ and by (4-33), we conclude that $h(\cdot\,,t_n) \to d$ in $W^{2,p}_\#(Q)$.

In particular, $F(h(\cdot\,,t_n), u_{h(\cdot,t_n)}) \to F(d,u_d)$. Then, by (3-54), $F(h(\cdot\,,t), u_h(\cdot\,,t)) \to F(d,u_d)$ as $t \to +\infty$, $t \notin Z_0$. On the other hand, for $t \in Z_0$ we have that $F(h(\cdot\,,t), u_h(\cdot\,,t)) \leq F(h(\cdot\,,\tau), u_h(\cdot\,,\tau))$ for all $\tau < t$, $\tau \notin Z_0$, by (3-55). Therefore,

$$\limsup_{t\to+\infty, t\in Z_0} F(h(\cdot\,,t), u_h(\cdot\,,t)) \leq F(d,u_d).$$

Recalling (4-32), we finally obtain (4-34). In turn, reasoning as in the proof of Theorem 4.7 (see (4-17)), it follows from (4-32) and (4-33) that, for every sequence $\{s_n\} \subset (0,+\infty)$ with $s_n \to +\infty$, there exists a subsequence such that $\{h(\cdot\,,s_n)\}$ converges to $d$ in $W^{2,p}_\#(Q)$. This implies that $h(\cdot\,,t) \to d$ in $W^{2,p}_\#(Q)$ as $t \to +\infty$ and concludes the proof. □

**4C.** ***The two-dimensional case.*** As remarked in the introduction, the arguments presented in the previous subsections apply to the two-dimensional version of (3-1), with $p = 2$, studied in [Fonseca et al. 2012], with

$$V = \big((g_{\theta\theta} + g)k + W(E(u)) - \varepsilon\big(k_{\sigma\sigma} + \tfrac{1}{2}k^3\big)\big)_{\sigma\sigma}. \tag{4-35}$$

Here $V$ denotes the outer normal velocity of $\Gamma_{h(\cdot,t)}$, $k$ is its curvature, $W(E(u))$ is the trace of $W\big(E(u(\cdot\,,t))\big)$ on $\Gamma_{h(\cdot,t)}$, with $u(\cdot\,,t)$ the elastic equilibrium in $\Omega_{h(\cdot,t)}$ under the conditions that $Du(\cdot\,,y)$ is $b$-periodic and $u(x,0) = e_0(x,0)$ for some $e_0 > 0$; and $(\cdot)_\sigma$ stands for tangential differentiation along $\Gamma_{h(\cdot,t)}$. The constant $e_0 > 0$ measures the lattice mismatch between the elastic film and the (rigid) substrate. Moreover, $g : [0, 2\pi] \to (0, +\infty)$ is defined as

$$g(\theta) = \psi(\cos\theta, \sin\theta) \tag{4-36}$$

and is evaluated at $\arg(\nu(\cdot, t))$, where $\nu(\cdot, t)$ is the outer normal to $\Gamma_{h(\cdot,t)}$. The underlying energy functional is then given by

$$F(h, u) := \int_{\Omega_h} W(E(u))\, dz + \int_{\Gamma_h} \bigl(\psi(\nu) + \tfrac{1}{2}\varepsilon k^2\bigr)\, d\mathcal{H}^1.$$

In the two-dimensional framework, given $b > 0$ we search for $b$-periodic solutions to (4-35). A local-in-time *b-periodic weak solution* to (4-35) is a function $h \in H^1(0, T_0; H^{-1}_{\#}(0, b)) \cap L^\infty(0, T_0; H^2_{\#}(0, b))$ such that:

(i) $(g_{\theta\theta} + g)k + W(E(u)) - \varepsilon\bigl(k_{\sigma\sigma} + \tfrac{1}{2}k^3\bigr) \in L^2(0, T_0; H^1_{\#}(0, b))$,

(ii) for almost every $t \in [0, T_0]$,

$$\frac{\partial h}{\partial t} = J\bigl((g_{\theta\theta} + g)k + Q(E(u)) - \varepsilon\bigl(k_{\sigma\sigma} + \tfrac{1}{2}k^3\bigr)\bigr)_{\sigma\sigma} \quad \text{in } H^{-1}_{\#}(0, b).$$

Given $(h_0, u_0)$ with $h_0 \in H^2_{\#}(0, b)$, $h_0 > 0$, and $u_0$ the corresponding elastic equilibrium, local-in-time existence of a *unique* weak solution with initial datum $(h_0, u_0)$ has been established in [Fonseca et al. 2012]. The Liapunov and asymptotic stability analysis of the flat configuration established in Sections 4A and 4B extends to the two-dimensional case, where, in addition, the range of those $d$ under which (4-10) holds can be analytically determined for isotropic elastic energies of the form

$$W(\xi) := \mu|\xi|^2 + \tfrac{1}{2}\lambda(\operatorname{trace}\xi)^2.$$

In the above formula, the *Lamé coefficients* $\mu$ and $\lambda$ are chosen to satisfy the ellipticity conditions $\mu > 0$ and $\mu + \lambda > 0$; see [Fusco and Morini 2012; Bonacini 2013a]. The stability range of the flat configuration depends on $\mu$, $\lambda$, and the mismatch constant $e_0$ appearing in the Dirichlet condition $u(x, 0) = e_0(x, 0)$. For the reader's convenience, we recall the results. Consider the *Grinfeld function* $K$ defined by

$$K(y) := \max_{n \in \mathbb{N}} \frac{1}{n} J(ny), \qquad y \ge 0, \tag{4-37}$$

where

$$J(y) := \frac{y + (3 - 4\nu_p)\sinh y \cosh y}{4(1 - \nu_p)^2 + y^2 + (3 - 4\nu_p)\sinh^2 y},$$

and $\nu_p$ is the *Poisson modulus* of the elastic material, i.e.,

$$\nu_p := \frac{\lambda}{2(\lambda + \mu)}. \tag{4-38}$$

It turns out that $K$ is strictly increasing and continuous, $K(y) \le Cy$, and $\lim_{y \to +\infty} K(y) = 1$ for some positive constant $C$. We also set, as in the previous subsections,

$$G(h, u) := \int_{\Omega_h} W(E(u))\, dz + \int_{\Gamma_h} \psi(\nu)\, d\mathcal{H}^1.$$

Combining [Fusco and Morini 2012, Theorem 2.9] and [Bonacini 2013a, Theorem 2.8] with the results of the previous subsection, we obtain the two-dimensional asymptotic stability of the flat configuration.

**Theorem 4.15.** *Assume $\partial^2_{11}\psi(0,1) > 0$ and define*

$$B = \frac{\pi}{4}\frac{(2\mu+\lambda)\partial^2_{11}\psi(0,1)}{e_0^2\mu(\mu+\lambda)}.$$

*Let $d_{\mathrm{loc}} : (0,+\infty) \to (0,+\infty]$ be defined as $d_{\mathrm{loc}}(b) := +\infty$ if $0 < b \le B$, and as the solution to*

$$K\left(\frac{2\pi d_{\mathrm{loc}}(b)}{b}\right) = \frac{B}{b} \tag{4-39}$$

*otherwise. Then the second variation of $G$ at $(d, u_d)$ is positive definite, i.e.,*

$$\partial^2 G(d,u_d)[\varphi] > 0 \quad \text{for all } \varphi \in H^1_\#(0,b)\setminus\{0\} \text{ with } \int_0^b \varphi\,dx = 0,$$

*if and only if $0 < d < d_{\mathrm{loc}}(b)$. In particular, for all $d \in (0, d_{\mathrm{loc}}(b))$ the flat configuration $(d, u_d)$ is asymptotically stable.*

## Appendix

**A1. *Regularity results.*** In this subsection we collect a few regularity results that have been used in the previous sections. We start with the following elliptic estimate, whose proof is essentially contained in [Fonseca et al. 2012, Lemma 6.10].

**Lemma A.1.** *Let $M > 0$, $c_0 > 0$. Let $h_1, h_2 \in C^{1,\alpha}_\#(Q)$ for some $\alpha \in (0,1)$, with $\|h_i\|_{C^{1,\alpha}_\#(Q)} \le M$ and $h_i \ge c_0$, $= i = 1, 2$, and let $u_1$ and $u_2$ be the corresponding elastic equilibria in $\Omega_{h_1}$ and $\Omega_{h_2}$, respectively. Then,*

$$\big\| E\big(u_1(\cdot, h_1(\cdot))\big) - E\big(u_2(\cdot, h_2(\cdot))\big)\big\|_{C^{1,\alpha}_\#(Q)} \le C\|h_1 - h_2\|_{C^{1,\alpha}_\#(Q)} \tag{A-1}$$

*for some constant $C > 0$ depending only on $M$, $c_0$, and $\alpha$.*

The following lemma is probably well known to the experts, however for the reader's convenience we provide a proof.

**Lemma A.2.** *Let $p > 2$, $u \in L^{\frac{p}{p-1}}(Q)$ such that*

$$\int_Q u\,AD^2\varphi\,dx + \int_Q b\cdot D\varphi + \int_Q c\varphi\,dx = 0 \quad \text{for all } \varphi \in C^\infty_\#(Q) \text{ with } \int_Q \varphi\,dx = 0,$$

*where $A \in W^{1,p}_\#(Q;\mathbb{M}^{2\times2}_{\mathrm{sym}})$ satisfies standard uniform ellipticity conditions (see (A-6)), $b \in L^1(Q;\mathbb{R}^2)$, and $c \in L^1(Q)$. Then $u \in L^q(Q)$ for all $q \in (1,2)$. Moreover, if $b$, $u \operatorname{Div} A \in L^r(Q;\mathbb{R}^2)$ and $c \in L^r(Q)$ for some $r > 1$, then $u \in W^{1,r}_\#(Q)$.*

*Proof.* We only prove the first assertion, since the other one can be proven using similar arguments. Denote by $A_\varepsilon$, $u_\varepsilon$, $b_\varepsilon$, and $c_\varepsilon$ the standard mollifications of $A$, $u$, $b$, and $c$, and let $v_\varepsilon \in C^\infty_\#(Q)$ be the unique solution to the problem

$$\begin{cases} \int_Q (A_\varepsilon D v_\varepsilon + u_\varepsilon \operatorname{Div} A_\varepsilon - b_\varepsilon)\cdot D\varphi\,dx - \int_Q c_\varepsilon \varphi\,dx = 0 \quad \text{for all } \varphi \in C^1_\#(Q),\ \int_Q \varphi\,dx = 0,\\ \int_Q v_\varepsilon\,dx = \int_Q u\,dx. \end{cases}$$

Denoting by $G_\varepsilon$ the Green's function associated with the elliptic operator

$$-\operatorname{Div}(A_\varepsilon Du)$$

it is known [Dong and Kim 2009, Equation (3.66); Grüter and Widman 1982, Equation (1.6)] that for all $q \in [1,2)$ and for all $x \in Q$ we have

$$\|D_y G_\varepsilon(x,\cdot)\|_{L^q(Q)} \le C,$$

with $C$ depending only on the ellipticity constants and $q$ and not on $\varepsilon$. Since

$$v_\varepsilon(x) = \int_Q G_\varepsilon(x,y)[-\operatorname{Div}(u_\varepsilon \operatorname{Div} A_\varepsilon - b_\varepsilon) + c_\varepsilon]\,dy = \int_Q [(u_\varepsilon \operatorname{Div} A_\varepsilon - b_\varepsilon)\cdot D_y G_\varepsilon(x,y) + G_\varepsilon(x,y)c_\varepsilon]\,dy,$$

it follows by standard properties of convolution that for all $q \in (1,2)$ there exists $C > 0$, depending only on $q$ and the $L^1$-norms of $u_\varepsilon \operatorname{Div} A_\varepsilon$, $b_\varepsilon$, $c_\varepsilon$, and hence on the $L^1$-norms of $b$, $c$, the $L^{p/(p-1)}$ norm of $u$, and the $W^{1,p}$ norm of $A$, such that $\|v_\varepsilon\|_{L^q(Q)} \le C$ for $\varepsilon$ sufficiently small. Thus, we may assume (up to subsequences) that $v_\varepsilon \rightharpoonup v$ weakly in $L^q(Q)$, where $v$ solves

$$\int_Q vAD^2\varphi\,dx + \int_Q (v \operatorname{Div} A - u \operatorname{Div} A + b)\cdot D\varphi\,dx + \int_Q c\varphi\,dx = 0 \tag{A-2}$$

for all $\varphi \in C^2_\#(Q)$ with $\int_Q \varphi\,dx = 0$, and satisfies

$$\int_Q v\,dx = \int_Q u\,dx. \tag{A-3}$$

Since by assumption $u$ solves the problem (A-2)–(A-3), it is enough to show that the problem admits a unique solution. Let $v_1$ and $v_2$ be two solutions and set $w := v_2 - v_1$. Then, we have

$$\int_Q wAD^2\varphi\,dx + \int_Q w \operatorname{Div} A\cdot D\varphi\,dx = 0 \tag{A-4}$$

for all $\varphi \in C^2_\#(Q)$ with $\int_Q \varphi\,dx = 0$. Let $g \in C^1_\#(Q)$ with $\int_Q g\,dx = 0$ and denote by $\varphi_g$ the unique solution in $W^{1,2}_\#(Q)$ to the equation $\operatorname{Div}(A[D\varphi_g]) = g$ such that $\int_Q \varphi_g\,dx = 0$. By a standard elliptic regularity argument and using the fact that $A \in W^{1,p}_\#(Q;\mathbb{M}^{2\times 2}_{\mathrm{sym}})$ for $p > 2$ it follows that $\varphi_g \in W^{2,2}_\#(Q)$. Therefore, setting $f := g - \operatorname{Div} A\cdot D\varphi_g$, we have that $AD^2\varphi_g = f$ and that $f \in L^s(Q)$ for all $s \in (1,p)$. Thus, we may apply Lemma A.3 to get that $\varphi_g \in W^{2,s}_\#(Q)$ for all $s \in (1,p)$. In turn, this implies that $f \in L^p(Q)$ and Lemma A.3 again yields that $\varphi_g \in W^{2,p}_\#(Q)$. Therefore $\varphi_g$ is an admissible test function for equation (A-4) and thus we deduce that $\int_Q wg\,dx = 0$ for all $g \in C^1_\#(Q)$ with $\int_Q g\,dx = 0$. This implies that $w$ is constant and, in turn, $w \equiv 0$ since $\int_Q w\,dx = 0$. ☐

In the next lemma we denote by $Lu$ an elliptic operator of the form

$$Lu := \sum_{ij} a_{ij}(x) D_{ij} u + \sum_i b_i(x) D_i u, \tag{A-5}$$

where all the coefficients are $Q$-periodic functions, the $a_{ij}$ are continuous, and the $b_i$ are bounded. Moreover, there exist $\lambda$, $\Lambda > 0$ such that

$$\Lambda |\xi|^2 \geq \sum_{ij} a_{ij}(x) \xi_i \xi_j \geq \lambda |\xi|^2 \quad \text{for all } \xi \in \mathbb{R}^2, \ \sum_i |b_i| \leq \Lambda. \tag{A-6}$$

**Lemma A.3.** *Let $p \geq 2$. Then, there exists $C > 0$ such that for all $u \in W^{2,p}_{\#}(Q)$ we have*

$$\|D^2 u\|_{L^p(Q)} \leq C \|Lu\|_{L^p(Q)},$$

*where $L$ is the differential operator defined in* (A-5). *The constant $C$ depends only on $p$, $\lambda$, $\Lambda$ and the moduli of continuity of the coefficients $a_{ij}$.*

*Proof.* We argue by contradiction, assuming that there exists a sequence $\{u_h\} \subset W^{2,p}_{\#}(Q)$, a modulus of continuity $\omega$, and a sequence of operators $\{L_h\}$ as in (A-5), with periodic coefficients $a^h_{ij}$, $b^h_i$ satisfying (A-6) and

$$|a^h_{ij}(x_1) - a^h_{ij}(x_2)| \leq \omega(|x_1 - x_2|)$$

for all $x_1, x_2 \in Q$, such that

$$\|D^2 u_h\|_{L^p(Q)} \geq h \|L_h u_h\|_{L^p(Q)}.$$

By homogeneity we may assume that

$$\|D^2 u_h\|_{L^p(Q)} = 1 \quad \text{for all } h \in \mathbb{N}. \tag{A-7}$$

Recall that, by periodicity,

$$\int_Q D u_h \, dx = 0.$$

Moreover, by adding a constant if needed, we may also assume that $\int_Q u_h \, dx = 0$. Therefore, by Poincaré's inequality and up to a subsequence, $u_h \rightharpoonup u$ weakly in $W^{2,p}_{\#}(Q)$. Moreover, we may also assume that there exist $a_{ij}$ and $b_i$ satisfying (A-6) such that

$$a^h_{ij} \to a_{ij} \quad \text{uniformly in } Q \qquad \text{and} \qquad b^h_i \overset{*}{\rightharpoonup} b_i \quad \text{weakly* in } L^\infty(Q).$$

Since $\|L_h u_h\|_{L^p(Q)} \to 0$, we have that $u$ is a periodic function satisfying $Lu = 0$, where $L$ is the operator associated with the coefficients $a_{ij}$ and $b_i$. Thus, by the maximum principle [Gilbarg and Trudinger 1983, Theorem 9.6] $u$ is constant, and thus $u = 0$. On the other hand, by elliptic regularity (see [ibid., Theorem 9.11]) there exists a constant $C > 0$ depending on $p$, $\lambda$, $\Lambda$, and $\omega$ such that

$$\|D^2 u_h\|_{L^p(Q)} \leq C(\|u_h\|_{W^{1,p}(Q)} + \|L_h u_h\|_{L^p(Q)}).$$

Since the right-hand side vanishes, we reach a contradiction to (A-7). □

**A2.** ***Interpolation results.***

**Theorem A.4.** *Let $\Omega \subset \mathbb{R}^n$ be a bounded open set satisfying the cone condition. Let $1 \le p \le \infty$ and $j, m$ be two integers such that $0 \le j \le m$ and $m \ge 1$. Then there exists $C > 0$ such that*

$$\|D^j f\|_{L^p(\Omega)} \le C(\|D^m f\|_{L^p(\Omega)}^{\frac{j}{m}} \|f\|_{L^p(\Omega)}^{\frac{m-j}{m}} + \|f\|_{L^p(\Omega)}) \tag{A-8}$$

*for all $f \in W^{m,p}(\Omega)$. Moreover, if $\Omega$ is a cube, $f \in W^{m,p}_{\#}(\Omega)$ and, if either $f$ vanishes at the boundary or $\int_\Omega f\,dx = 0$, then* (A-8) *holds in the stronger form*

$$\|D^j f\|_{L^p(\Omega)} \le C\|D^m f\|_{L^p(\Omega)}^{\frac{j}{m}} \|f\|_{L^p(\Omega)}^{\frac{m-j}{m}}. \tag{A-9}$$

*Proof.* Inequality (A-8) follows by combining inequalities (1) and (3) in [Adams and Fournier 2003, Theorem 5.2]. If $\Omega$ is a cube, $f$ is periodic and, if either $f$ vanishes at the boundary or $\int_\Omega f\,dx = 0$, then inequality (A-9) follows by observing that

$$\|f\|_{W^{m,p}(\Omega)} \le C\|D^m f\|_{L^p(\Omega)},$$

as a straightforward application of the Poincaré inequality. □

The next interpolation result is obtained by combining [Adams and Fournier 2003, Theorem 5.8] with (A-8).

**Theorem A.5.** *Let $\Omega \subset \mathbb{R}^n$ be a bounded open set satisfying the cone condition. If $mp > n$, let $1 \le p \le q \le \infty$; if $mp = n$, let $1 \le p \le q < \infty$; if $mp < n$, let $1 \le p \le q \le np/(n - mp)$. Then there exists $C > 0$ such that*

$$\|f\|_{L^q(\Omega)} \le C(\|D^m f\|_{L^p(\Omega)}^{\theta} \|f\|_{L^p(\Omega)}^{1-\theta} + \|f\|_{L^p(\Omega)}) \tag{A-10}$$

*for all $f \in W^{m,p}(\Omega)$, where $\theta := n/(mp) - n/(mq)$. Moreover, if $\Omega$ is a cube, $f \in W^{m,p}_{\#}(\Omega)$ and, if either $f$ vanishes at the boundary or $\int_\Omega f\,dx = 0$, then* (A-10) *holds in the stronger form*

$$\|f\|_{L^q(\Omega)} \le C\|D^m f\|_{L^p(\Omega)}^{\theta} \|f\|_{L^p(\Omega)}^{1-\theta}. \tag{A-11}$$

Combining Theorems A.4 and A.5, and arguing as in the proof of [Fonseca et al. 2012, Theorem 6.4], we have the following theorem:

**Theorem A.6.** *Let $\Omega \subset \mathbb{R}^n$ be a bounded open set satisfying the cone condition. Let $s$, $j$, and $m$ be integers such that $0 \le s \le j \le m$. Let $1 \le p \le q < \infty$ if $(m - j)p \ge n$, and let $1 \le p \le q \le \infty$ if $(m - j)p > n$. Then, there exists $C > 0$ such that*

$$\|D^j f\|_{L^q(\Omega)} \le C(\|D^m f\|_{L^p(\Omega)}^{\theta} \|D^s f\|_{L^p(\Omega)}^{1-\theta} + \|D^s f\|_{L^p(\Omega)}) \tag{A-12}$$

*for all $f \in W^{m,p}(\Omega)$, where*

$$\theta := \frac{1}{m-s}\Big(\frac{n}{p} - \frac{n}{q} + j - s\Big).$$

*Moreover, if $\Omega$ is a cube, $f \in W^{m,p}_{\#}(\Omega)$ and, if either $f$ vanishes at the boundary or $\int_\Omega f\,dx = 0$, then* (A-12) *holds in the stronger form*

$$\|D^j f\|_{L^q(\Omega)} \le C\|D^m f\|^{\theta}_{L^p(\Omega)}\|D^s f\|^{1-\theta}_{L^p(\Omega)}. \tag{A-13}$$

Finally, we conclude with an interpolation estimate involving the $H^{-1}$-norm; see Remark 3.3.

**Lemma A.7.** *There exists $C > 0$ such that, for all $f \in H^1_{\#}(Q)$ with $\int_Q f\,dx = 0$, we have*

$$\|f\|_{L^2(Q)} \le C\|Df\|^{\frac{1}{2}}_{L^2(Q)}\|f\|^{\frac{1}{2}}_{H^{-1}_{\#}(Q)}.$$

*Similarly, there exists $C > 0$ such that, for all $f \in H^2_{\#}(Q)$ with $\int_Q f\,dx = 0$, we have*

$$\|f\|_{L^2(Q)} \le C\|D^2 f\|^{\frac{1}{3}}_{L^2(Q)}\|f\|^{\frac{2}{3}}_{H^{-1}_{\#}(Q)}.$$

*Proof.* Let $w$ be the unique $Q$-periodic solution to

$$\begin{cases} -\Delta w = f \text{ in } Q, \\ \int_Q w\,dx = 0. \end{cases}$$

Combining Lemma A.3 with (A-9) we obtain

$$\begin{aligned} \|f\|_{L^2(Q)} = \|\Delta w\|_{L^2(Q)} &\le C\|D^2 w\|_{L^2(Q)} \le C\|D^3 w\|^{\frac{1}{2}}_{L^2(Q)}\|Dw\|^{\frac{1}{2}}_{L^2(Q)} \\ &\le C\|\Delta(Dw)\|^{\frac{1}{2}}_{L^2(Q)}\|Dw\|^{\frac{1}{2}}_{L^2(Q)} = C\|Df\|^{\frac{1}{2}}_{L^2(Q)}\|f\|^{\frac{1}{2}}_{H^{-1}_{\#}(Q)}. \end{aligned}$$

The second inequality of the statement is proven similarly. □

## Acknowledgements

The authors warmly thank the Center for Nonlinear Analysis (NSF Grants No. DMS-0405343 and DMS-0635983), where part of this research was carried out. The research of I. Fonseca was partially funded by the National Science Foundation under Grant No. DMS-0905778 and that of G. Leoni under Grant No. DMS-1007989. Fonseca and Leoni also acknowledge support of the National Science Foundation under the PIRE Grant No. OISE-0967140. The work of N. Fusco was supported by ERC under FP7 Advanced Grant No. 226234 and was partially carried on at the University of Jyväskylä under the FiDiPro No. 2100002028.

## References

[Acerbi et al. 2013] E. Acerbi, N. Fusco, and M. Morini, "Minimality via second variation for a nonlocal isoperimetric problem", *Comm. Math. Phys.* **322**:2 (2013), 515–557. MR 3077924 Zbl 1270.49043

[Adams and Fournier 2003] R. A. Adams and J. J. F. Fournier, *Sobolev spaces*, 2nd ed., Pure and Applied Mathematics (Amsterdam) **140**, Elsevier/Academic Press, Amsterdam, 2003. MR 2009e:46025 Zbl 1098.46001

[Almgren et al. 1993] F. Almgren, J. E. Taylor, and L. Wang, "Curvature-driven flows: a variational approach", *SIAM J. Control Optim.* **31**:2 (1993), 387–438. MR 94h:58067 Zbl 0783.35002

[Ambrosetti and Prodi 1993] A. Ambrosetti and G. Prodi, *A primer of nonlinear analysis*, Cambridge Studies in Advanced Mathematics **34**, Cambridge University Press, 1993. MR 94f:58016 Zbl 0781.47046

[Ambrosio 1995] L. Ambrosio, "Minimizing movements", *Rend. Accad. Naz. Sci. XL Mem. Mat. Appl.* (5) **19** (1995), 191–246. MR 97c:49044 Zbl 0957.49029

[Ambrosio et al. 2008] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, 2nd ed., Birkhäuser, Basel, 2008. MR 2009h:49002 Zbl 1145.35001

[Bellettini et al. 2006] G. Bellettini, V. Caselles, A. Chambolle, and M. Novaga, "Crystalline mean curvature flow of convex sets", *Arch. Ration. Mech. Anal.* **179**:1 (2006), 109–152. MR 2007a:53126 Zbl 1148.53049

[Bellettini et al. 2007] G. Bellettini, C. Mantegazza, and M. Novaga, "Singular perturbations of mean curvature flow", *J. Differential Geom.* **75**:3 (2007), 403–431. MR 2008b:53087 Zbl 1119.53043

[Bonacini 2013a] M. Bonacini, "Epitaxially strained elastic films: the case of anisotropic surface energies", *ESAIM Control Optim. Calc. Var.* **19**:1 (2013), 167–189. MR 3023065 Zbl 06144571

[Bonacini 2013b] M. Bonacini, "Stability of equilibrium configurations for elastic films in two and three dimensions", preprint, 2013. arXiv 1308.1537

[Bonnetier and Chambolle 2002] E. Bonnetier and A. Chambolle, "Computing the equilibrium configuration of epitaxially strained crystalline films", *SIAM J. Appl. Math.* **62**:4 (2002), 1093–1121. MR 2003c:74070 Zbl 1001.49017

[Burger et al. 2007] M. Burger, F. Haußer, C. Stöcker, and A. Voigt, "A level set approach to anisotropic flows with curvature regularization", *J. Comput. Phys.* **225**:1 (2007), 183–205. MR 2008h:82098 Zbl 1201.82049

[Cagnetti et al. 2008] F. Cagnetti, M. G. Mora, and M. Morini, "A second order minimality condition for the Mumford–Shah functional", *Calc. Var. Partial Differential Equations* **33**:1 (2008), 37–74. MR 2010b:49074 Zbl 1191.49018

[Cahn and Taylor 1994] J. W. Cahn and J. E. Taylor, "Overview no. 113: Surface motion by surface diffusion", *Acta Metallurgica et Materialia* **42**:4 (1994), 1045–1063.

[Capriani et al. 2013] G. M. Capriani, V. Julin, and G. Pisante, "A quantitative second order minimality criterion for cavities in elastic bodies", *SIAM J. Math. Anal.* **45**:3 (2013), 1952–1991. MR 3070533 Zbl 06200955

[Caselles and Chambolle 2006] V. Caselles and A. Chambolle, "Anisotropic curvature-driven flow of convex sets", *Nonlinear Anal.* **65**:8 (2006), 1547–1577. MR 2007d:35143 Zbl 1107.35069

[Chambolle 2004] A. Chambolle, "An algorithm for mean curvature motion", *Interfaces Free Bound.* **6**:2 (2004), 195–218. MR 2005f:53106 Zbl 1061.35147

[Di Carlo et al. 1992] A. Di Carlo, M. E. Gurtin, and P. Podio-Guidugli, "A regularized equation for anisotropic motion-by-curvature", *SIAM J. Appl. Math.* **52**:4 (1992), 1111–1119. MR 93g:73005 Zbl 0800.73021

[Dong and Kim 2009] H. Dong and S. Kim, "Green's matrices of second order elliptic systems with measurable coefficients in two dimensional domains", *Trans. Amer. Math. Soc.* **361**:6 (2009), 3303–3323. MR 2010h:35005 Zbl 1173.35050

[Escher et al. 1998] J. Escher, U. F. Mayer, and G. Simonett, "The surface diffusion flow for immersed hypersurfaces", *SIAM J. Math. Anal.* **29**:6 (1998), 1419–1433. MR 99f:58042 Zbl 0912.35161

[Fonseca 1991] I. Fonseca, "The Wulff theorem revisited", *Proc. Roy. Soc. London Ser. A* **432**:1884 (1991), 125–145. MR 92e:49053 Zbl 0725.49017

[Fonseca et al. 2007] I. Fonseca, N. Fusco, G. Leoni, and M. Morini, "Equilibrium configurations of epitaxially strained crystalline films: existence and regularity results", *Arch. Ration. Mech. Anal.* **186**:3 (2007), 477–537. MR 2009a:74088 Zbl 1126.74029

[Fonseca et al. 2011] I. Fonseca, N. Fusco, G. Leoni, and V. Millot, "Material voids in elastic solids with anisotropic surface energies", *J. Math. Pures Appl.* (9) **96**:6 (2011), 591–639. MR 2851683 Zbl 1285.74003

[Fonseca et al. 2012] I. Fonseca, N. Fusco, G. Leoni, and M. Morini, "Motion of elastic thin films by anisotropic surface diffusion with curvature regularization", *Arch. Ration. Mech. Anal.* **205**:2 (2012), 425–466. MR 2947537 Zbl 1270.74127

[Fusco and Morini 2012] N. Fusco and M. Morini, "Equilibrium configurations of epitaxially strained elastic films: second order minimality conditions and qualitative properties of solutions", *Arch. Ration. Mech. Anal.* **203**:1 (2012), 247–327. MR 2864412 Zbl 1281.74024

[Gilbarg and Trudinger 1983] D. Gilbarg and N. S. Trudinger, *Elliptic partial differential equations of second order*, 2nd ed., Grundlehren der Math. Wissenschaften **224**, Springer, Berlin, 1983. MR 86c:35035 Zbl 0562.35001

[Giusti 1984] E. Giusti, *Minimal surfaces and functions of bounded variation*, Monographs in Mathematics **80**, Birkhäuser, Basel, 1984. MR 87a:58041 Zbl 0545.49018

[Grüter and Widman 1982] M. Grüter and K.-O. Widman, "The Green function for uniformly elliptic equations", *Manuscripta Math.* **37**:3 (1982), 303–342. MR 83h:35033 Zbl 0485.35031

[Gurtin and Jabbour 2002] M. E. Gurtin and M. E. Jabbour, "Interface evolution in three dimensions with curvature-dependent energy and surface diffusion: interface-controlled evolution, phase transitions, epitaxial growth of elastic films", *Arch. Ration. Mech. Anal.* **163**:3 (2002), 171–208. MR 2003d:74073 Zbl 1053.74005

[Herring 1951] C. Herring, "Some theorems on the free energies of crystal surfaces", *Phys. Rev.* **82** (1951), 87–93. Zbl 0042.23201

[Mantegazza 2002] C. Mantegazza, "Smooth geometric evolutions of hypersurfaces", *Geom. Funct. Anal.* **12**:1 (2002), 138–182. MR 2003f:53118 Zbl 1011.58010

[Mullins 1963] W. W. Mullins, "Solid surface morphologies governed by capillarity", in *Metal surfaces: structure, energetics and kinetics*, American Society for Metals, Metals Park, OH, 1963.

[Piovano 2014] P. Piovano, "Evolution of elastic thin films with curvature regularization via minimizing movements", *Calc. Var. Partial Differential Equations* **49**:1–2 (2014), 337–367. MR 3148120 Zbl 1288.35282

[Rätz et al. 2006] A. Rätz, A. Ribalta, and A. Voigt, "Surface evolution of elastically stressed films under deposition by a diffuse interface model", *J. Comput. Phys.* **214**:1 (2006), 187–208. MR 2006k:74084 Zbl 1088.74035

[Siegel et al. 2004] M. Siegel, M. J. Miksis, and P. W. Voorhees, "Evolution of material voids for highly anisotropic surface energy", *J. Mech. Phys. Solids* **52**:6 (2004), 1319–1353. MR 2004m:74013 Zbl 1079.74513

Received 9 May 2014. Accepted 22 Jan 2015.

IRENE FONSECA: fonseca@andrew.cmu.edu
*Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA 15213, United States*

NICOLA FUSCO: n.fusco@unina.it
*Dipartimento di Matematica e Applicazioni "R. Caccioppoli", Università degli Studi di Napoli "Federico II", I-80126 Napoli, Italy*

GIOVANNI LEONI: giovanni@andrew.cmu.edu
*Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA 15213, United States*

MASSIMILIANO MORINI: massimiliano.morini@unipr.it
*Dipartimento di Matematica, Università degli Studi di Parma, I-43121 Parma, Italy*

ANALYSIS AND PDE
Vol. 8, No. 2, 2015
dx.doi.org/10.2140/apde.2015.8.425

# EXPONENTIAL CONVERGENCE TO EQUILIBRIUM IN A COUPLED GRADIENT FLOW SYSTEM MODELING CHEMOTAXIS

JONATHAN ZINSL AND DANIEL MATTHES

We study a system of two coupled nonlinear parabolic equations. It constitutes a variant of the Keller–Segel model for chemotaxis; i.e., it models the behavior of a population of bacteria that interact by means of a signaling substance. We assume an external confinement for the bacteria and a nonlinear dependency of the chemotactic drift on the signaling substance concentration.

We perform an analysis of existence and long-time behavior of solutions based on the underlying gradient flow structure of the system. The result is that, for a wide class of initial conditions, weak solutions exist globally in time and converge exponentially fast to the unique stationary state under suitable assumptions on the convexity of the confinement and the strength of the coupling.

## 1. Introduction

**1A.** ***The equations and their variational structure.*** This paper is concerned with existence and long-time behavior of weak nonnegative solutions to the initial value problem

$$\partial_t u(t,x) = \operatorname{div}\big(u(t,x)\mathrm{D}[u(t,x)+W(x)+\varepsilon\phi(v(t,x))]\big), \tag{1}$$

$$\partial_t v(t,x) = \Delta v(t,x) - \kappa v(t,x) - \varepsilon u(t,x)\phi'(v(t,x)), \tag{2}$$

$$u(0,x) = u_0(x) \geq 0, \qquad v(0,x) = v_0(x) \geq 0, \tag{3}$$

where the sought functions $u$ and $v$ are defined for $(t,x) \in [0,\infty) \times \mathbb{R}^3$. Below, we comment in detail on the origin of (1)–(2) from mathematical biology. In brief, $u$ is the spatial density of bacteria that interact with each other by means of a signaling substance of local concentration $v$.

In (1)–(2), $\varepsilon$ and $\kappa$ are given positive constants; we are mainly concerned with the case where the *coupling strength* $\varepsilon$ is sufficiently small. Strict positivity of $\kappa$ is essential for our approach, as explained below. The *response function* $\phi \in C^2([0,\infty))$ is assumed to be convex and strictly decreasing, with

$$0 < -\phi'(w) \leq -\phi'(0) < \infty, \quad 0 \leq \phi''(w) \leq \overline{\phi''} < \infty \quad \text{for all } w \geq 0 \tag{4}$$

This research was supported by the German Research Foundation (DFG) Collaborative Research Center SFB-TR 109.
*MSC2010:* primary 35K45; secondary 35A15, 35B40, 35D30, 35Q92.
*Keywords:* gradient flow, Wasserstein metric, chemotaxis.

for an appropriate constant $\overline{\phi''} \geq 0$, the paradigmatic examples being

$$\phi(w) = -w \qquad \text{(classical Keller–Segel model)}, \tag{5}$$

$$\phi(w) = -\log(1+w) \qquad \text{(weak saturation effect)}, \tag{6}$$

$$\phi(w) = \frac{1}{1+w} \qquad \text{(strong saturation effect)}. \tag{7}$$

The *external potential* $W \in C^2(\mathbb{R}^3)$ is assumed to grow quadratically: it has globally bounded second-order partial derivatives and is uniformly convex with a constant $\lambda_0 > 0$, that is,

$$\mathrm{D}^2 W(x) \geq \lambda_0 \mathbb{1} \quad \text{for all } x \in \mathbb{R}^3 \text{ in the sense of symmetric matrices.} \tag{8}$$

Without loss of generality, we may assume that $W \geq 0$.

Equations (1)–(2) possess a variational structure. Formally, they can be written as a gradient flow of the entropy functional

$$\mathcal{H}(u, v) := \int_{\mathbb{R}^3} \bigl(\tfrac{1}{2}u^2 + uW + \tfrac{1}{2}|\mathrm{D}v|^2 + \tfrac{1}{2}\kappa v^2 + \varepsilon u \phi(v)\bigr)\,\mathrm{d}x$$

with respect to a metric **dist**, defined on the space $\boldsymbol{X} := \mathcal{P}_2(\mathbb{R}^3) \times L^2_+(\mathbb{R}^3)$ by

$$\mathbf{dist}((u_1, v_1), (u_2, v_2)) := \sqrt{\boldsymbol{W}_2^2(u_1, u_2) + \|v_1 - v_2\|^2_{L^2(\mathbb{R}^3)}} \quad \text{for } (u_1, v_1), (u_2, v_2) \in \boldsymbol{X}. \tag{9}$$

Here $\boldsymbol{W}_2$ is the *$L^2$-Wasserstein metric* on the space $\mathcal{P}_2(\mathbb{R}^3)$ of probability measures on $\mathbb{R}^3$ with finite second moment; see Section 2A for the definition. This gradient flow structure is at the basis of our proof for global existence of weak solutions to (1)–(3), and it is also the key element for our analysis of long-time behavior. We remark that, even with this variational structure at hand, the analysis is far from trivial since $\mathcal{H}$ is *not* convex along geodesics. Therefore, the established general theory on $\lambda$-contractive gradient flows in metric spaces [Ambrosio et al. 2008] is not directly applicable.

**1B. *Statement of the main results.*** In the first part of this work, we show that a weak solution to (1)–(2) can be obtained by means of the time-discrete implicit Euler approximation (also known as *minimizing movement* or the *JKO scheme*). More precisely, for each sufficiently small time step $\tau > 0$, let $(u_\tau^0, v_\tau^0) := (u_0, v_0)$, and then define inductively for each $n \in \mathbb{N}$

$$(u_\tau^n, v_\tau^n) \in \operatorname*{argmin}_{(u,v)\in\mathcal{P}_2(\mathbb{R}^3)\times L^2(\mathbb{R}^3)} \left(\frac{1}{2\tau}\,\mathbf{dist}\bigl((u, v), (u_\tau^{n-1}, v_\tau^{n-1})\bigr)^2 + \mathcal{H}(u, v)\right). \tag{10}$$

We will prove in Section 4A that this construction is well defined, i.e., that a minimizer exists for every $n \in \mathbb{N}$. Further, introduce the piecewise-constant interpolation $(u_\tau, v_\tau) : \mathbb{R}_+ \to \mathcal{P}_2(\mathbb{R}^3) \times L^2(\mathbb{R}^3)$ by

$$u_\tau(t) = u_\tau^n, \quad v_\tau(t) = v_\tau^n \quad \text{for all } t \in ((n-1)\tau, n\tau]. \tag{11}$$

Our existence result — which does not require a small coupling strength $\varepsilon$ — reads as follows.

**Theorem 1.1** (existence of weak solutions to (1)–(2)). *Let $\kappa > 0$ and $\varepsilon > 0$ be given, and assume that the response function $\phi$ satisfies* (4) *and that the convex confinement potential $W$ grows quadratically.*

*Let further initial conditions* $u_0 \in \mathcal{P}_2(\mathbb{R}^3) \cap L^2(\mathbb{R}^3)$ *and* $v_0 \in W^{1,2}(\mathbb{R}^3)$ *be given, with* $v_0 \geq 0$, *and define for each* $\tau > 0$ *a function* $(u_\tau, v_\tau)$ *by means of the scheme* (10) *and* (11). *Then there is a sequence* $(\tau_k)_{k\in\mathbb{N}}$ *with* $\tau_k \downarrow 0$ *such that* $(u_{\tau_k}, v_{\tau_k})$ *converges to a weak solution* $(u, v) : [0, \infty) \times \mathbb{R}^3 \to [0, \infty]^2$ *of* (1)–(3) *in the following sense*:

$$
\begin{aligned}
&u_{\tau_k}(t) \to u(t) \quad \textit{narrowly in } \mathcal{P}(\mathbb{R}^3), \textit{pointwise with respect to } t \in [0, T],\\
&v_{\tau_k}(t) \to v(t) \quad \textit{in } L^2(\mathbb{R}^3), \textit{uniformly with respect to } t \in [0, T],\\
&\quad u \in C^{1/2}([0, T], \mathcal{P}_2(\mathbb{R}^3)) \cap L^\infty([0, T], L^2(\mathbb{R}^3)) \cap L^2([0, T], W^{1,2}(\mathbb{R}^3)),\\
&\quad v \in C^{1/2}([0, T], L^2(\mathbb{R}^3)) \cap L^\infty([0, T], W^{1,2}(\mathbb{R}^3)) \cap L^2([0, T], W^{2,2}(\mathbb{R}^3)) \cap W^{1,2}([0, T], L^2(\mathbb{R}^3))
\end{aligned}
$$

*for all* $T > 0$, *and* $(u, v)$ *satisfies*

$$\partial_t u = \operatorname{div}(u\mathrm{D}[u + W + \varepsilon\phi(v)]) \quad \textit{in the sense of distributions}, \tag{12}$$

$$\partial_t v = \Delta v - \kappa v - \varepsilon u\phi'(v) \qquad \textit{a.e. in } (0, +\infty) \times \mathbb{R}^3, \tag{13}$$

$$u(0) = u_0, \qquad v(0) = v_0. \tag{14}$$

The convergence of $(u_{\tau_k}, v_{\tau_k})$ is actually much stronger; see Proposition 4.7 for details.

The key a priori estimate yielding sufficient compactness of $(u_\tau, v_\tau)$ follows from a dissipation estimate, which formally amounts to

$$
\begin{aligned}
-\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}^3}\bigl(u\log u + \tfrac{1}{2}|\mathrm{D}v|^2 + \tfrac{1}{2}\kappa v^2\bigr)\,\mathrm{d}x \geq \frac{1}{2}\int_{\mathbb{R}^3}&(|\mathrm{D}u|^2 + (\Delta v - \kappa v)^2)\,\mathrm{d}x\\
&- C\bigl(\|u\|^2_{L^2(\mathbb{R}^3)} + \|v\|^2_{W^{1,2}(\mathbb{R}^3)} + \|\Delta W\|_{L^\infty(\mathbb{R}^3)}\bigr).
\end{aligned}
$$

Related existence results have been proved recently for similar systems of equations, using essentially the same technique, in [Laurençot and Matioc 2013; Blanchet and Laurençot 2013; Zinsl 2014; Blanchet et al. 2014]. Therefore, we keep the technical details to a minimum. Note that our method of proof yields neither contractivity of the flow nor uniqueness of weak solutions due to the lack of convexity of the entropy functional.

Our main result is the following on the long-time behavior of solutions:

**Theorem 1.2** (exponential convergence to equilibrium). *Let* $\kappa$, $\phi$ *and* $W$ *be as in Theorem 1.1 above. Then there are constants* $\bar{\varepsilon} > 0$, $L > 0$ *and* $C > 0$ *such that, for every* $\varepsilon \in (0, \bar{\varepsilon})$ *and with* $\Lambda_\varepsilon := \min(\kappa, \lambda_0) - L\varepsilon$, *the following is true.*

*Let initial conditions* $u_0 \in \mathcal{P}_2(\mathbb{R}^3) \cap L^2(\mathbb{R}^3)$ *and* $v_0 \in W^{1,2}(\mathbb{R}^3)$ *be given, with* $v_0 \geq 0$, *and assume in addition that* $v_0 \in L^{6/5}(\mathbb{R}^3)$. *Let further* $(u, v)$ *be a weak solution to* (1)–(3) *obtained as a limit of the scheme* (10) *and* (11). *Then* $(u, v)$ *converges to the unique nonnegative stationary solution* $(u_\infty, v_\infty) \in (\mathcal{P}_2 \cap L^2)(\mathbb{R}^3) \times W^{1,2}(\mathbb{R}^3)$ *of* (1)–(2) *exponentially fast with rate* $\Lambda_\varepsilon$ *in the sense*

$$
\begin{aligned}
&\boldsymbol{W}_2(u(t,\cdot), u_\infty) + \|u(t,\cdot) - u_\infty\|_{L^2(\mathbb{R}^3)} + \|v(t,\cdot) - v_\infty\|_{W^{1,2}(\mathbb{R}^3)}\\
&\qquad\leq C\bigl(1 + \|v_0\|_{L^{6/5}(\mathbb{R}^3)}\bigr)(\mathcal{H}(u_0, v_0) - \mathcal{H}(u_\infty, v_\infty) + 1)e^{-\Lambda_\varepsilon t} \quad \textit{for all } t \geq 0.
\end{aligned} \tag{15}
$$

We give a brief and formal indication of the main idea for the proof of Theorem 1.2. First, we decompose the entropy in the form

$$\mathcal{H}(u,v)-\mathcal{H}(u_\infty,v_\infty)=\mathcal{L}_u(u)+\mathcal{L}_v(v)+\varepsilon\mathcal{L}_*(u,v), \tag{16}$$

where

$$\begin{aligned}\mathcal{L}_u(u)&:=\int_{\mathbb{R}^3}\bigl(\tfrac{1}{2}(u^2-u_\infty^2)+[W+\varepsilon\phi(v_\infty)](u-u_\infty)\bigr)\,\mathrm{d}x,\\ \mathcal{L}_v(v)&:=\int_{\mathbb{R}^3}\tfrac{1}{2}(|\mathrm{D}(v-v_\infty)|^2+\kappa(v-v_\infty)^2)\,\mathrm{d}x,\\ \mathcal{L}_*(u,v)&:=\int_{\mathbb{R}^3}\bigl(u[\phi(v)-\phi(v_\infty)]-u_\infty\phi'(v_\infty)[v-v_\infty]\bigr)\,\mathrm{d}x.\end{aligned}$$

There, $\mathcal{L}_u$ and $\mathcal{L}_v$ are $\lambda_\varepsilon$-convex and $\kappa$-convex functionals — with $\lambda_\varepsilon=\lambda_0-C\varepsilon>0$ — in $(\mathcal{P}_2,\boldsymbol{W}_2)$ and in $L^2$, respectively, which are minimized by the stationary solution $(u_\infty,v_\infty)$; the functional $\mathcal{L}_*$ has no useful convexity properties. On a very formal level — pretending that $\mathcal{L}_u$, $\mathcal{L}_v$ and $\mathcal{L}_*$ are smooth functionals on Euclidean spaces and denoting their "gradients" by $\nabla_u$ and $\nabla_v$ — the dissipation of the *principal entropy* $\mathcal{L}_u+\mathcal{L}_v$ amounts to

$$\begin{aligned}-\frac{\mathrm{d}}{\mathrm{d}t}(\mathcal{L}_u+\mathcal{L}_v)&=\nabla_u\mathcal{L}_u\cdot\nabla_u\mathcal{H}+\nabla_v\mathcal{L}_v\cdot\nabla_v\mathcal{H}\\ &=\|\nabla_u\mathcal{L}_u\|^2+\|\nabla_v\mathcal{L}_v\|^2+\varepsilon\nabla_u\mathcal{L}_u\cdot\nabla_u\mathcal{L}_*+\varepsilon\nabla_v\mathcal{L}_v\cdot\nabla_v\mathcal{L}_*\\ &\geq(1-\varepsilon)\|\nabla_u\mathcal{L}_u\|^2+(1-\varepsilon)\|\nabla_v\mathcal{L}_v\|^2-\tfrac{1}{2}\varepsilon(\|\nabla_u\mathcal{L}_*\|^2+\|\nabla_v\mathcal{L}_*\|^2).\end{aligned} \tag{17}$$

By convexity of $\mathcal{L}_u$ and $\mathcal{L}_v$, one has the inequalities

$$\|\nabla_u\mathcal{L}_u\|^2\geq 2\lambda_\varepsilon\mathcal{L}_u,\qquad \|\nabla_v\mathcal{L}_v\|^2\geq 2\kappa\mathcal{L}_v,$$

and so we are *almost* in the situation to apply the Gronwall estimate to (17) and conclude convergence to equilibrium with an exponential rate of $\min(\lambda_\varepsilon,\kappa)>0$. However, it remains to estimate the terms involving the "gradients" of $\mathcal{L}_*$. This is relatively straightforward if the entropy $\mathcal{H}(u,v)$ is sufficiently close to its minimal value $\mathcal{H}(u_\infty,v_\infty)$ but is rather difficult for $(u,v)$ far from equilibrium. Moreover, rigorous estimates have to be carried out on the time-discrete level (with subsequent passage to continuous time) since our notion of solution is too weak to carry out the respective estimates in continuous time.

In the language of gradient flows, our results can be interpreted as follows. For $\varepsilon=0$, the functional $\mathcal{H}$ is $\Lambda_0$-convex along geodesics in $(\boldsymbol{X},\mathbf{dist})$ with $\Lambda_0=\min(\lambda_0,\kappa)>0$. Consequently, there is an associated $\Lambda_0$-contractive gradient flow defined on all of $\boldsymbol{X}$ that satisfies (1)–(2), and in particular, all solutions converge with the exponential rate $\Lambda_0$ to the unique equilibrium. For every $\varepsilon>0$, the convexity of $\mathcal{H}$ is lost; see [Zinsl 2014] for a discussion of (non)convexity in a similar situation. By Theorem 1.1, Equations (1)–(2) still define a continuous flow on the proper domain of $\mathcal{H}$, which is $\boldsymbol{X}\cap(L^2(\mathbb{R}^3)\times W^{1,2}(\mathbb{R}^3))$. Further, we show that, on the (almost exhaustive) subset of those $(u,v)$ with $v\in L^{6/5}(\mathbb{R}^3)$, this flow still converges to an equilibrium with an exponential rate $\Lambda_\varepsilon\geq\Lambda_0-L\varepsilon>0$ (Theorem 1.2) for all $\varepsilon>0$ sufficiently small.

From this point of view, our result is perturbative: the uncoupled system ($\varepsilon = 0$) exhibiting a strictly contractive flow is perturbed in such a way that the perturbed system ($\varepsilon > 0$) still yields exponential convergence towards the unique equilibrium — with a slightly slower convergence rate than in the unperturbed case. For this approach to work, we obviously need to require $\kappa > 0$ and $\lambda_0 > 0$. On the other hand, this theorem is stronger than a usual perturbation result: the crucial point is that we do *not* require the initial condition $(u_0, v_0)$ to be close to equilibrium, apart from the rather harmless additional hypothesis that $v_0 \in L^{6/5}(\mathbb{R}^3)$, which could be weakened further with additional technical effort.

Our result on global existence of weak solutions (Theorem 1.1), however, can be generalized to the case of $\kappa = 0$ and no convexity assumption on the confinement potential (see, e.g., [Zinsl 2014]). Further generalization of Theorem 1.1 to the case of nonlinear, but nonquadratic, diffusion can be achieved with similar techniques as in [Zinsl 2014]. However, in our analysis of the long-time behavior, the right entropy dissipation estimates are not at hand to the best of our knowledge when dealing with nonquadratic diffusion. To keep technicalities to a minimum, we consider the quadratic case throughout this work.

We expect that similar results can be proved for system (1)–(2) on a bounded domain $\Omega \subset \mathbb{R}^3$, even with vanishing confinement $W \equiv 0$. The role of the confinement will then be played by Poincaré's inequality. Our setup with a convex confinement on $\mathbb{R}^3$ fits much more naturally into the variational framework.

**1C.** ***Modeling background.*** The system of equations (1)–(2) is a variant of the so-called *Keller–Segel model for chemotaxis* describing the time-dependent distribution of biological cells or microorganisms in response to gradients of chemical substances (*chemotaxis*). The original model — corresponding to the linear response function from (5) — has been developed by Keller and Segel [1970] to describe slime mold aggregation. However, chemotactic processes occur in many (and highly different) biological systems; for the biological details, we refer to the book by Eisenbach [Eisenbach 2004]. For example, many bacteria like *Escherichia coli* possess *flagella* driven by small motors that respond to gradients of signaling molecules in the environment. Chemotaxis also plays an important role in embryonal development, e.g., in the development of blood vessels (*angiogenesis*), which is also a crucial step in tumor growth. Starting from the basic Keller–Segel model, many different model extensions are conceivable. A broad range of those is summarized in the review articles by Hillen and Painter [2009] and Horstmann [2003]. Details on the modeling aspects can be found, e.g., in the books by Murray [2003] and Perthame [2007].

In the model (1)–(2) under consideration here, $u$ is the time-dependent spatial density of the cells and $v$ is the time-dependent concentration of the signaling substance. Equation (1) describes the temporal change in cell density due to the directed drift of cells towards regions with a higher concentration of the substance and due to undirected diffusion. Equation (2) models the degradation of the signaling substance as well as its production by the cells. Two special aspects are included in this particular model: nonlinear diffusion, i.e., the use of a nonconstant, $u$-dependent mobility coefficient for the diffusive motion of the bacteria, and signal-dependent chemotactic sensitivity, i.e., the use of the — in general nonlinear — response $\phi(v)$ instead of the concentration $v$ itself. For the first, we refer to [Hillen and Painter 2009] and the references therein for biological motivation. The second is motivated by the fact that the conversion of an external signal into a reaction of the considered microorganism (*signal transduction*) often occurs by binding and dissociation of molecules to certain receptors. The movement of the cell is then caused by gradients in the number of receptors occupied by signaling molecules rather than by concentration gradients of signaling molecules themselves. For growing concentrations, the number of bound receptors can exhibit a saturation such that the gradient vanishes. In [Hillen and Painter 2009; Segel 1977; Lapidus and Schiller 1976], this was incorporated into the model by the *chemotactic sensitivity function*

$$\phi'(v) = -\frac{1}{(1+v)^2},$$

which fits into our model with the response function $\phi$ defined in (7). Finally, an external background potential $W$ is included in order to generate a spatial confinement of the bacterial population.

For the dynamics of the signaling substance, we assume linear diffusion according to Fick's laws and degradation with a constant, exponential rate $\kappa$. The nonnegative term $-\varepsilon u\phi'(v)$ models the production of signaling substance by the microorganisms; here it is taken into account that the cells might be less active in producing additional substance the higher its local concentration already is. This is consistent with the models presented in [Horstmann 2003, §6]. Often, the two processes of chemotactic response and production of the chemoattractant are modeled with different response functions. Here we require them to be equal (or scalar multiples of each other) to ensure that the system has a gradient flow structure.

By definition, $u$ and $v$ are density/concentration functions and thus should be nonnegative. Note that it is part of our results that, for given nonnegative initial data (of sufficient regularity), there exists a weak solution that is nonnegative for all times $t > 0$. On a formal level, nonnegativity is an easy consequence of the particular structure of the system (1)–(2).

**1D.** ***Relation to the existing literature.*** The rapidly growing mathematical literature about the Keller–Segel model and its manifold variants is devoted primarily to the dichotomy *global existence versus finite-time blow-up* of (weak, possibly measure-valued) solutions, but the long-time behavior of global solutions has been intensively investigated as well.

Global existence and blow-up in the classical parabolic-parabolic Keller–Segel model, which is (1)–(2) with $\phi(v) = -v$, $W \equiv 0$ and *linear* diffusion, has been thoroughly studied by Calvez and Corrias [2008] in space dimension $d = 2$ and by Corrias and Perthame [2008] in higher space dimensions $d > 2$; see also [Biler et al. 2011; Kozono and Sugiyama 2009; Mizoguchi 2013; Nagai et al. 2003; Senba and Suzuki 2006; Sugiyama and Kunii 2006; Yamada 2011]. Recently, in [Carrapatoso and Mischler 2014], uniqueness and long-time behavior of solutions of the parabolic-parabolic Keller–Segel system was studied by means of a perturbation of the parabolic-elliptic framework.

Variants with nonlinear diffusion and drift have been studied for instance by Sugiyama [2006; 2007]. The results from [Sugiyama 2006] already indicate that, in the model (1)–(2) under consideration, blow-up *never* occurs, in accordance with Theorem 1.1.

In the aforementioned works [Corrias and Perthame 2008; Nagai et al. 2003], the intermediate asymptotics of global solutions have been studied as well: it is proved that the cell density converges to the self-similar solution of the heat equation at an algebraic rate, i.e., in a properly scaled frame, the density approaches a Gaussian. See also [Di Francesco and Rosado 2008] for an extension of this result to a model with size-exclusion. Similar asymptotic behavior has been proved in models with nonlinear, homogeneous diffusion, e.g., by Luckhaus and Sugiyama [2006; 2007]. There, the intermediate asymptotics are that of a porous medium equation with the respective homogeneous nonlinearity; i.e., the rescaled bacterial density converges to a Barenblatt profile. These intermediate asymptotics are — at least morally — related to Theorem 1.2: recall that algebraic convergence to self-similarity for the unconfined porous medium equation is comparable to exponential convergence to an equilibrium for the equation with $\lambda$-convex confinement.

The fully parabolic model (1)–(2) with a nonlinear response $\phi$ has not been rigorously analyzed so far, with the following exception: in her thesis, Post [1999] proves existence and uniqueness of solutions to a similar system with linear diffusion and vanishing confinement on a bounded domain by nonvariational methods and obtains convergence to the (spatially homogeneous) stationary solution from compactness arguments. Variants of the classical parabolic-parabolic or parabolic-elliptic Keller–Segel models with a nonlinear chemotactic sensitivity coefficient have also been studied, e.g., in [Nagai and Senba 1998; Winkler 2010].

Despite the fact that energy/entropy methods are one of the key tools for the analysis of Keller–Segel-type systems, the use of genuine variational methods is relatively recent in that context. The variational machinery of gradient flows in transportation metrics, originally developed by Jordan, Kinderlehrer and Otto [Jordan et al. 1998] for the linear Fokker–Planck equation, has been applied to a variety of dynamical systems: mainly to nonlinear diffusion [Carrillo and Toscani 2000; Otto 2001; Carrillo et al. 2006a; Agueh 2008] but also to aggregation [Carrillo et al. 2003; 2006b; 2011] and fourth-order equations [Giacomelli and Otto 2001; Gianazza et al. 2009; Matthes et al. 2009].

For the parabolic-elliptic Keller–Segel model, which can be reduced to a single nonlocal scalar equation, the variational framework was established by Blanchet, Calvez and Carrillo [Blanchet et al. 2008], who represented the evolution as a gradient flow of an appropriate potential with respect to the Wasserstein distance and constructed a numerical scheme on these grounds. Later, the gradient flow structure has been used for a detailed analysis of the basin of attraction in the critical mass case by Blanchet, Carlen and Carrillo [Blanchet et al. 2012] (see also, e.g., [Blanchet et al. 2009; Calvez and Carrillo 2012; López-Gómez et al. 2013]).

The parabolic-parabolic Keller–Segel model was somewhat harder to fit into the framework since the two equations are (formally) gradient flows with respect to *different* metrics: Wasserstein and $L^2$. The first rigorous analytical result on grounds of this structure was given by Blanchet and Laurençot [2013], where they constructed weak solutions for the system with critical exponents of nonlinear diffusion. Later their result was generalized to other, noncritical parameter situations in [Zinsl 2014]. In the recent work by Blanchet et al. [2014], a similar strategy was used to reprove the result in [Calvez and Corrias 2008] about the global existence of weak solutions to the classical Keller–Segel system in two spatial dimensions. To the best of our knowledge, our approach taken here to prove long-time asymptotics by gradient flow techniques in a combined Wasserstein-$L^2$-metric is novel.

**1E.** ***Plan of the paper.*** First, we summarize common facts and definitions on gradient flows in metric spaces in Section 2. After that, various properties of the entropy functional are derived in Section 3. On grounds of these properties, we construct a weak solution by means of the minimizing movement scheme in Section 4, proving Theorem 1.1. Existence, uniqueness and regularity of stationary solutions are studied in Section 5, and the proof of Theorem 1.2 is completed in Section 6.

## 2. Preliminaries

In this section, we recall the relevant definitions and properties related to gradient flows in metric spaces $(X, \boldsymbol{d})$, following [Ambrosio et al. 2008]. The two metric spaces of interest here are $L^2(\mathbb{R}^d)$ with the metric induced by the norm and the space $\mathcal{P}_2(\mathbb{R}^d)$ of probability measures, endowed with the $L^2$-Wasserstein distance $\boldsymbol{W}_2$. We also discuss the compound metric **dist** from (9).

**2A.** ***Spaces of probability measures and the Wasserstein distance.*** We denote by $\mathcal{P}(\mathbb{R}^d)$ the space of probability measures on $\mathbb{R}^d$. By abuse of notation, we will frequently identify absolutely continuous measures $\mu \in \mathcal{P}(\mathbb{R}^d)$ with their respective (Lebesgue) density functions $u = \mathrm{d}\mu/\mathrm{d}x \in L^1_+(\mathbb{R}^d)$, where $L^p_+(\mathbb{R}^d)$ for $p \geq 1$ denotes the subspace of those $L^p(\mathbb{R}^d)$ functions with nonnegative values.

A sequence $(\mu_n)_{n\in\mathbb{N}}$ in $\mathcal{P}(\mathbb{R}^d)$ is called *narrowly convergent* to its limit $\mu \in \mathcal{P}(\mathbb{R}^d)$ if

$$\lim_{n\to\infty} \int_{\mathbb{R}^d} \varphi(x)\,\mathrm{d}\mu_n(x) = \int_{\mathbb{R}^d} \varphi(x)\,\mathrm{d}\mu(x)$$

for every bounded, continuous function $\varphi : \mathbb{R}^d \to \mathbb{R}$. By $\mathcal{P}_2(\mathbb{R}^d)$, we denote the subspace of those $\mu \in \mathcal{P}(\mathbb{R}^d)$ with finite second moment

$$\boldsymbol{m}_2(\mu) := \int_{\mathbb{R}^d} |x|^2\,\mathrm{d}\mu(x).$$

$\mathcal{P}_2(\mathbb{R}^d)$ turns into a complete metric space when endowed with the $L^2$-Wasserstein distance $\boldsymbol{W}_2$. We do not recall the general definition of $\boldsymbol{W}_2$ here. Instead, since we are concerned with absolutely continuous measures in $\mathcal{P}_2(\mathbb{R}^d)$ only, we remark that, for probability density functions $u_1, u_2 \in L^1_+(\mathbb{R}^d)$, the Wasserstein distance is given by the infimum

$$\boldsymbol{W}_2^2(u_1, u_2) = \inf\left\{ \int_{\mathbb{R}^d} |t(x) - x|^2 u_1(x)\,\mathrm{d}x \;\middle|\; t : \mathbb{R}^d \to \mathbb{R}^d \text{ Borel-measurable and } t \# u_1 = u_2 \right\},$$

where $t \# u$ denotes the *push-forward* with respect to the map $t$. In this case, the infimum above is attained by an optimal transport map [Villani 2003, Theorem 2.32]. Convergence in the metric space $(\mathcal{P}_2(\mathbb{R}^d), \boldsymbol{W}_2)$ is equivalent to narrow convergence and convergence of the second moment. Further, $\boldsymbol{W}_2$ is lower semicontinuous in both components with respect to narrow convergence.

**2B.** ***Geodesic convexity and gradient flows in metric spaces.*** A functional $\mathfrak{A} : X \to \mathbb{R} \cup \{\infty\}$ defined on the metric space $(X, \boldsymbol{d})$ is called *geodesically* $\lambda$*-convex* for some $\lambda \in \mathbb{R}$ if, for every $w_0, w_1 \in X$ and $s \in [0, 1]$, one has

$$\mathfrak{A}(w_s) \leq (1-s)\mathfrak{A}(w_0) + s\mathfrak{A}(w_1) - \tfrac{1}{2}\lambda s(1-s)\boldsymbol{d}^2(w_0, w_1),$$

where $w_s : [0, 1] \to X,\ s \mapsto w_s$ is a *geodesic connecting* $w_0$ and $w_1$.

On $L^2(\mathbb{R}^d)$, the (unique up to rescaling) geodesic from $w_0$ to $w_1$ is given by linear interpolation, i.e., $w_s = (1-s)w_0 + sw_1$. Hence, a functional $\mathfrak{F} : L^2(\mathbb{R}^d) \to \mathbb{R}\cup\{\infty\}$ of the form

$$\mathfrak{F}(w) = \int_{\mathbb{R}^d} f(w(x), \mathrm{D}w(x), \mathrm{D}^2 w(x))\,\mathrm{d}x$$

with a given continuous function $f : \mathbb{R}\times\mathbb{R}^d\times\mathbb{R}^{d\times d} \to \mathbb{R}$ is $\lambda$-convex if and only if $(z, p, Q) \mapsto f(z, p, Q) - \frac{1}{2}\lambda z^2$ is (jointly) convex.

In the metric space $(\mathcal{P}_2(\mathbb{R}^d), \boldsymbol{W}_2)$, geodesic $\lambda$-convexity is a much more complicated concept. We recall two important classes of $\lambda$-convex functionals (see, e.g., [Ambrosio et al. 2008, Chapter 9.3; Villani 2003, Theorem 5.15]).

**Theorem 2.1** (criteria for geodesic convexity in $(\mathcal{P}_2(\mathbb{R}^d), \boldsymbol{W}_2)$). *The following statements are true*:

(a) *Let a function $h \in C^0([0,\infty))$ be given, and define the functional $\mathfrak{H}$ on $\mathcal{P}_2(\mathbb{R}^d)$ by $\mathfrak{H}(u) := \int_{\mathbb{R}^d} h(u(x))\,\mathrm{d}x$ for $u \in (\mathcal{P}_2 \cap L^1)(\mathbb{R}^d)$. If $h(0) = 0$ and $r \mapsto r^d h(r^{-d})$ is convex and nonincreasing on $(0,\infty)$, then $\mathfrak{H}$ is 0-geodesically convex and lower semicontinuous in $(\mathcal{P}_2(\mathbb{R}^d), \boldsymbol{W}_2)$.*

(b) *Let a function $W \in C^0(\mathbb{R}^d)$ be given, and define the functional $\mathfrak{H}(\mu) := \int_{\mathbb{R}^d} W(x)\,\mathrm{d}\mu(x)$ for all $\mu \in \mathcal{P}_2(\mathbb{R}^d)$. If $W$ is $\lambda$-convex (as a functional on the metric space $\mathbb{R}^d$ with the Euclidean distance) for some $\lambda \in \mathbb{R}$, then $\mathfrak{H}$ is $\lambda$-geodesically convex in $(\mathcal{P}_2(\mathbb{R}^d), \boldsymbol{W}_2)$.*

Next, we introduce a notion of *gradient flow*. There are various possible characterizations. For our purposes here, we need the following very strong one:

**Definition 2.2.** Let $\mathfrak{A} : X \to \mathbb{R}\cup\{\infty\}$ be a lower semicontinuous functional on the metric space $(X, \boldsymbol{d})$. A continuous semigroup $\mathsf{S}^{\mathfrak{A}}_{(\cdot)}$ on $(X, \boldsymbol{d})$ is called *$\kappa$-flow* for some $\kappa \in \mathbb{R}$ if the *evolution variational inequality*

$$\frac{1}{2}\frac{\mathrm{d}^+}{\mathrm{d}t}\boldsymbol{d}^2(\mathsf{S}^{\mathfrak{A}}_t(w), \widetilde{w}) + \frac{\kappa}{2}\boldsymbol{d}^2(\mathsf{S}^{\mathfrak{A}}_t(w), \widetilde{w}) + \mathfrak{A}(\mathsf{S}^{\mathfrak{A}}_t(w)) \le \mathfrak{A}(\widetilde{w}) \tag{18}$$

holds for arbitrary $w$ and $\widetilde{w}$ in the domain of $\mathfrak{A}$ and for all $t \ge 0$.

If $\mathsf{S}^{\mathfrak{A}}_{(\cdot)}$ is a $\kappa$-flow for the $\lambda$-convex functional $\mathfrak{A}$, then $\mathsf{S}^{\mathfrak{A}}_{(\cdot)}$ is also a *gradient flow* for $\mathfrak{A}$ in essentially all possible interpretations of that notion. For the metric spaces $(\mathcal{P}_2(\mathbb{R}^d), \boldsymbol{W}_2)$ and $L^2(\mathbb{R}^d)$, it can be proved that every lower semicontinuous and geodesically $\lambda$-convex functional possesses a unique $\kappa$-flow, with $\kappa := \lambda$ (see [Ambrosio et al. 2008, Theorem 11.1.4, Corollary 4.3.3], respectively).

In these metric spaces, $\lambda$-geodesic convexity with $\lambda > 0$ implies existence and uniqueness of a minimizer $w_{\min}$ of $\mathfrak{A}$, for which the following holds (see, e.g., [Ambrosio et al. 2008, Lemma 2.4.8, Theorem 4.0.4]):

$$\frac{\lambda}{2}\boldsymbol{d}^2(w, w_{\min}) \le \mathfrak{A}(w) - \mathfrak{A}(w_{\min}) \le \frac{1}{2\lambda}\lim_{h\downarrow 0}\frac{\mathfrak{A}(w) - \mathfrak{A}(\mathsf{S}^{\mathfrak{A}}_h(w))}{h}. \tag{19}$$

**Remark 2.3** (formal calculation of evolution equations associated to gradient flows). In the metric spaces of interest here, one can explicitly write an evolution equation for the flow $\mathsf{S}^{\mathfrak{A}}_{(\cdot)}$ of a sufficiently regular functional $\mathfrak{A}$; see, e.g., [Villani 2003, §8.2]. On $(\mathscr{P}_2(\mathbb{R}^d), \boldsymbol{W}_2)$, one has

$$\partial_t \mathsf{S}_t^{\mathfrak{A}}(w) = \mathrm{div}\biggl(\mathsf{S}_t^{\mathfrak{A}}(w)\mathrm{D}\biggl(\frac{\delta \mathfrak{A}}{\delta w}(\mathsf{S}_t^{\mathfrak{A}}(w))\biggr)\biggr),$$

and on $L^2(\mathbb{R}^d)$, one has

$$\partial_t \mathsf{S}_t^{\mathfrak{A}}(w) = -\frac{\delta \mathfrak{A}}{\delta w}(\mathsf{S}_t^{\mathfrak{A}}(w)).$$

Here, $\delta\mathfrak{A}/\delta w$ stands for the usual first variation of the functional $\mathfrak{A}$ on $L^2$.

**2C. *The metric* dist.** It is easily verified that $\boldsymbol{X} := \mathscr{P}_2(\mathbb{R}^3) \times L^2_+(\mathbb{R}^3)$ becomes a complete metric space when endowed with the compound metric **dist** defined in (9). The topology on $\boldsymbol{X}$ induced by **dist** is that of the cartesian product. Moreover:

**Lemma 2.4.** *The distance* **dist** *is weakly lower semicontinuous on $\boldsymbol{X}$ in the following sense: if $(u_n, v_n)_{n\in\mathbb{N}}$ is a sequence in $\boldsymbol{X}$ such that $u_n$ converges to $u \in \mathscr{P}_2(\mathbb{R}^3)$ narrowly and $v_n$ converges to $v \in L^2(\mathbb{R}^3)$ weakly in $L^2(\mathbb{R}^3)$, then*

$$\mathbf{dist}((u, v), (\tilde{u}, \tilde{v})) \leq \liminf_{n\to\infty} \mathbf{dist}((u_n, v_n), (\tilde{u}, \tilde{v}))$$

*holds, for every $(\tilde{u}, \tilde{v}) \in \boldsymbol{X}$.*

For our purposes, it suffices to discuss convexity and gradient flows for functionals $\Phi : \boldsymbol{X} \to \mathbb{R} \cup \{\infty\}$ of the separable form $\Phi(u, v) = \Phi_1(u) + \Phi_2(v)$. One immediately verifies:

**Lemma 2.5.** *Assume that $\Phi_1$ and $\Phi_2$ are $\lambda$-convex and lower semicontinuous functionals on the respective spaces $(\mathscr{P}_2(\mathbb{R}^3), \boldsymbol{W}_2)$ and $L^2(\mathbb{R}^3)$, and denote their respective gradient flows by $\mathsf{S}^1_{(\cdot)}$ and $\mathsf{S}^2_{(\cdot)}$. Then $\Phi : \boldsymbol{X} \to \mathbb{R} \cup \{\infty\}$ with $\Phi(u, v) = \Phi_1(u) + \Phi_2(v)$ is a $\lambda$-convex and lower semicontinuous functional on $(\boldsymbol{X}, \mathbf{dist})$, and the semigroup $\mathsf{S}^{\Phi}_{(\cdot)}$ given by $\mathsf{S}^{\Phi}_t(u, v) = (\mathsf{S}^1_t(u), \mathsf{S}^2_t(v))$ is a $\lambda$-flow for $\Phi$.*

## 3. Properties of the entropy functional

Recall the definition of the metric space $(\boldsymbol{X}, \mathbf{dist})$. We define the *entropy functional* $\mathscr{H} : \boldsymbol{X} \to \mathbb{R} \cup \{\infty\}$ as follows. For all $(u, v) \in \boldsymbol{X} \cap (L^2(\mathbb{R}^3) \times W^{1,2}(\mathbb{R}^3))$, set

$$\mathscr{H}(u, v) := \int_{\mathbb{R}^3} \bigl(\tfrac{1}{2}u^2 + uW + \tfrac{1}{2}|\mathrm{D}v|^2 + \tfrac{1}{2}\kappa v^2 + \varepsilon u\phi(v)\bigr)\,\mathrm{d}x, \tag{20}$$

which is a finite value by our assumptions on $\phi$ and $W$. For all other $(u, v) \in \mathscr{P}_2(\mathbb{R}^3) \times L^2(\mathbb{R}^3)$, we set $\mathscr{H}(u, v) = +\infty$.

**Proposition 3.1** (properties of the entropy functional $\mathscr{H}$). *The functional $\mathscr{H}$ defined in* (20) *has the following properties*:

(a) *There exist $C_0, C_1 > 0$ such that*

$$\mathscr{H}(u, v) \geq C_0\bigl[\|u\|^2_{L^2(\mathbb{R}^3)} + \boldsymbol{m}_2(u) + \|v\|^2_{W^{1,2}(\mathbb{R}^3)} - C_1\bigr]. \tag{21}$$

*In particular, $\mathscr{H}$ is bounded from below.*

(b) *$\mathcal{H}$ is weakly lower semicontinuous in the following sense: for every sequence $(u_n, v_n)_{n\in\mathbb{N}}$ in $\boldsymbol{X}$, where $(u_n)_{n\in\mathbb{N}}$ converges narrowly to some $u \in \mathcal{P}_2(\mathbb{R}^3)$ and where $(v_n)_{n\in\mathbb{N}}$ converges weakly in $L^2(\mathbb{R}^3)$ to some $v \in L^2(\mathbb{R}^3)$, one has*

$$\mathcal{H}(u, v) \le \liminf_{n\to\infty} \mathcal{H}(u_n, v_n).$$

(c) *For sufficiently small $\varepsilon > 0$, $\mathcal{H}$ is $\lambda'$-geodesically convex for some $\lambda' > 0$ with respect to the distance induced by the norm $\|(\tilde{u}, \tilde{v})\|_{L^2\times L^2} := \sqrt{\|\tilde{u}\|^2_{L^2(\mathbb{R}^3)} + \|\tilde{v}\|^2_{L^2(\mathbb{R}^3)}}$.*

*Proof.* For part (a), we observe that, due to $\lambda_0$-convexity of $W$, one has $W(x) \ge \frac{1}{4}\lambda_0|x|^2 - \frac{1}{2}\lambda_0|x_{\min}|^2$, where $x_{\min} \in \mathbb{R}^3$ is the unique minimizer of $W$. Moreover, with convexity of $\phi$, we deduce

$$\int_{\mathbb{R}^3} u\phi(v)\,\mathrm{d}x \ge \phi(0) + \phi'(0)\|uv\|_{L^1(\mathbb{R}^3)} \ge \phi(0) + C\phi'(0)\|\mathrm{D}v\|_{L^2(\mathbb{R}^3)}\|u\|^{1/3}_{L^2(\mathbb{R}^3)},$$

using that $\|u\|_{L^1(\mathbb{R}^3)} = 1$ and the chain of inequalities

$$\|uv\|_{L^1(\mathbb{R}^3)} \le \|u\|_{L^{6/5}(\mathbb{R}^3)}\|v\|_{L^6(\mathbb{R}^3)} \le C\|\mathrm{D}v\|_{L^2(\mathbb{R}^3)}\|u\|^{2/3}_{L^1(\mathbb{R}^3)}\|u\|^{1/3}_{L^2(\mathbb{R}^3)}. \tag{22}$$

All in all, we arrive at

$$\mathcal{H}(u, v) \ge \tfrac{1}{2}\|u\|^2_{L^2(\mathbb{R}^3)} + \tfrac{1}{4}\lambda_0\boldsymbol{m}_2(u) - \lambda_0|x_{\min}|^2 + \tfrac{1}{2}\|\mathrm{D}v\|^2_{L^2(\mathbb{R}^3)} + \tfrac{1}{2}\kappa\|v\|^2_{L^2(\mathbb{R}^3)} - \varepsilon|\phi(0)| - \varepsilon C|\phi'(0)|\|\mathrm{D}v\|_{L^2(\mathbb{R}^3)}\|u\|^{1/3}_{L^2(\mathbb{R}^3)}.$$

From this, the desired estimate follows by Young's inequality.

In (b), the claimed lower semicontinuity of the integral with $\varepsilon = 0$ follows from joint convexity of the map

$$\mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}^3 \ni (r, z, p) \mapsto \tfrac{1}{2}r^2 + W(x)r + \tfrac{1}{2}|p|^2 + \tfrac{1}{2}\kappa z^2,$$

for every $x \in \mathbb{R}^3$. It thus remains to prove semicontinuity of the integral of $u\phi(v)$. Let a sequence $(u_n, v_n)_{n\in\mathbb{N}}$ with the mentioned properties be given, and assume — without loss of generality — that $\mathcal{H}(u_n, v_n) \to H < \infty$. With these prerequisites at hand, we are even able to prove the *continuity* of the integral of $u\phi(v)$: it follows by (21) that $(u_n)_{n\in\mathbb{N}}$ and $(v_n)_{n\in\mathbb{N}}$ are bounded sequences in $L^2(\mathbb{R}^3)$ and in $W^{1,2}(\mathbb{R}^3)$, respectively. Moreover, the sequence of second moments $(\boldsymbol{m}_2(u_n))_{n\in\mathbb{N}}$ is bounded. Hence, $(u_n)_{n\in\mathbb{N}}$ converges to $u$ weakly in $L^2(\mathbb{R}^3)$, and $(v_n)_{n\in\mathbb{N}}$ converges to $v$ weakly in $W^{1,2}(\mathbb{R}^3)$ and strongly in $L^2(B_R(0))$, for every ball $B_R(0) \subset \mathbb{R}^3$. Recalling our assumptions (4) on $\phi$, we conclude that

$$|\phi(v_n) - \phi(v)|^2 \le \phi'(0)^2|v_n - v|^2,$$

and thus, $(\phi(v_n))_{n\in\mathbb{N}}$ converges to $\phi(v)$ strongly in $L^2(B_R(0))$. We proceed by a truncation argument. Let therefore $R > 0$, and choose $\beta_R \in C^\infty(\mathbb{R}^3)$ with

$$0 \le \beta_R \le 1, \qquad \beta_R \equiv 1 \quad \text{on } B_R(0), \qquad \beta_R \equiv 0 \quad \text{on } \mathbb{R}^3 \setminus B_{2R}(0).$$

Using the triangle inequality, we see

$$\left|\int_{\mathbb{R}^3}(u_n\phi(v_n)-u\phi(v))\,\mathrm{d}x\right| \le \left|\int_{\mathbb{R}^3}\phi(v)(u_n-u)\,\mathrm{d}x\right|+\left|\int_{\mathbb{R}^3}\beta_R u_n(\phi(v_n)-\phi(v))\,\mathrm{d}x\right|+\left|\int_{\mathbb{R}^3}(1-\beta_R)u_n(\phi(v_n)-\phi(v))\,\mathrm{d}x\right|. \quad (23)$$

Since $u_n \rightharpoonup u$ weakly on $L^2(\mathbb{R}^3)$ and $\phi(v)\in L^2(\mathbb{R}^3)$, the first term in (23) converges to zero. The same holds for the second one due to strong convergence of $\phi(v_n)$ to $\phi(v)$ on $L^2(B_{2R}(0))$ and boundedness of $\|u_n\|_{L^2(\mathbb{R}^3)}$. The third term in (23) can be estimated using (22):

$$\left|\int_{\mathbb{R}^3}(1-\beta_R)u_n(\phi(v_n)-\phi(v))\,\mathrm{d}x\right| \le \|\phi(v_n)-\phi(v)\|_{L^6(\mathbb{R}^3)}\|u_n\|_{L^{6/5}(\mathbb{R}^3\setminus B_R(0))}.$$

Consequently,

$$\begin{aligned}\left|\int_{\mathbb{R}^3}(1-\beta_R)u_n(\phi(v_n)-\phi(v))\,\mathrm{d}x\right| &\le C\|\phi(v_n)-\phi(v)\|_{W^{1,2}(\mathbb{R}^3)}\|u_n\|_{L^2(\mathbb{R}^3)}^{1/3}\left(\int_{\mathbb{R}^3\setminus B_R(0)}\frac{|x|^2}{R^2}u_n(x)\,\mathrm{d}x\right)^{2/3}\\ &\le CR^{-4/3}(\|\phi(v_n)\|_{W^{1,2}(\mathbb{R}^3)}+\|\phi(v)\|_{W^{1,2}(\mathbb{R}^3)})\|u_n\|_{L^2(\mathbb{R}^3)}^{1/3}(\boldsymbol{m}_2(u_n))^{2/3}\\ &\le 2\widetilde{C}R^{-4/3}.\end{aligned}$$

Hence, for all $R>0$,

$$\limsup_{n\to\infty}\left|\int_{\mathbb{R}^3}(u_n\phi(v_n)-u\phi(v))\,\mathrm{d}x\right| \le 2\widetilde{C}R^{-4/3},$$

proving the claim.

Finally, to prove (c), consider a geodesic $w_s=(u_s,v_s)$ with respect to the flat metric induced by $\|\cdot\|_{L^2\times L^2}$; that is, $u_s=(1-s)u_0+su_1$ and $v_s=(1-s)v_0+sv_1$ for given $u_0,u_1\in(\mathscr{P}_2\cap L^2)(\mathbb{R}^3)$ and $v_0,v_1\in W^{1,2}(\mathbb{R}^3)$. It then follows that

$$\begin{aligned}\frac{\mathrm{d}^2}{\mathrm{d}s^2}\mathscr{H}(u_s,v_s) &= \int_{\mathbb{R}^3}\big((u_1-u_0)^2+|\mathrm{D}(v_1-v_0)|^2+\kappa(v_1-v_0)^2\\ &\qquad +2\varepsilon\phi'(v_s)(u_1-u_0)(v_1-v_0)+\varepsilon u_s\phi''(v_s)(v_1-v_0)^2\big)\,\mathrm{d}x\\ &\ge \int_{\mathbb{R}^3}\begin{pmatrix}u_1-u_0\\ v_1-v_0\end{pmatrix}^{\mathrm{T}}A_s\begin{pmatrix}u_1-u_0\\ v_1-v_0\end{pmatrix}\mathrm{d}x \quad \text{with } A_s:=\begin{pmatrix}1 & \varepsilon\phi'(v_s)\\ \varepsilon\phi'(v_s) & \kappa\end{pmatrix},\end{aligned}$$

where we have used that $\phi$ is convex. Thus, $\mathscr{H}$ is $\lambda'$-convex with respect to the flat distance above if $A_s\ge\lambda'\mathbb{1}$ for all $s\in[0,1]$. Recalling that $0<-\phi'(v_s)\le-\phi'(0)$ by hypothesis (4), it follows from elementary linear algebra that $\varepsilon^2\phi'(0)^2<\kappa$ is sufficient to find a suitable $\lambda'>0$ with $A_s\ge\lambda'\mathbb{1}$. □

## 4. Existence of weak solutions

In this section, we prove Theorem 1.1 by construction of a weak solution using the minimizing movement scheme.

**4A.** ***Time discretization.*** Recall the discretization scheme from (10). We introduce the step size $\tau > 0$ and define the associated Yosida penalization $\mathcal{H}_\tau$ of the entropy by

$$\mathcal{H}_\tau(u, v \mid \tilde{u}, \tilde{v}) := \frac{1}{2\tau}\,\mathbf{dist}^2((u, v), (\tilde{u}, \tilde{v})) + \mathcal{H}(u, v) \tag{24}$$

for all $(u, v), (\tilde{u}, \tilde{v}) \in \boldsymbol{X}$. Set $(u_\tau^0, v_\tau^0) := (u_0, v_0)$, and define the sequence $(u_\tau^n, v_\tau^n)_{n\in\mathbb{N}}$ inductively by choosing

$$(u_\tau^n, v_\tau^n) \in \underset{(u,v)\in\boldsymbol{X}}{\operatorname{argmin}}\, \mathcal{H}_\tau(u, v \mid u_\tau^{n-1}, v_\tau^{n-1}). \tag{25}$$

**Lemma 4.1.** *For every $(\tilde{u}, \tilde{v}) \in \boldsymbol{X}$, there exists at least one minimizer $(u, v) \in \boldsymbol{X}$ of $\mathcal{H}_\tau(\,\cdot \mid \tilde{u}, \tilde{v})$ that satisfies $u \in L^2(\mathbb{R}^3)$ and $v \in W^{1,2}(\mathbb{R}^3)$.*

*Proof.* The proof is an application of the direct methods from the calculus of variations to the functional $\mathcal{H}_\tau(\,\cdot \mid \tilde{u}, \tilde{v})$.

First, observe that, on any given sublevel $S$ of $\mathcal{H}_\tau(\,\cdot \mid \tilde{u}, \tilde{v})$, both $\boldsymbol{W}_2(u, \tilde{u})$ and $\|v\|_{L^2(\mathbb{R}^3)}$ are uniformly bounded. The first bound implies that also the second moment $\boldsymbol{m}_2(u)$ is uniformly bounded, and thus, the $u$-components in $S$ belong to a subset of $\mathcal{P}_2(\mathbb{R}^3)$ that is relatively compact in the narrow topology by Prokhorov's theorem. The other bound implies via Alaoglu's theorem that the $v$-components belong to a weakly relatively compact subset of $L^2(\mathbb{R}^3)$.

Next, recall the properties of $\mathcal{H}$ and of **dist** given in Proposition 3.1 and Lemma 2.4. From these, it follows that $\mathcal{H}_\tau(\,\cdot \mid \tilde{u}, \tilde{v})$ is lower semicontinuous with respect to narrow convergence in the first and $L^2$-weak convergence in the second components.

Combining these properties with the fact that $\mathcal{H}_\tau(\,\cdot \mid \tilde{u}, \tilde{v})$ is bounded from below (e.g., by zero), the existence of a minimizer follows. The additional regularity is a consequence of the fact that the proper domain of $\mathcal{H}$ is a subset of $L^2(\mathbb{R}^3) \times W^{1,2}(\mathbb{R}^3)$. □

Given the sequence $(u_\tau^n, v_\tau^n)_{n\in\mathbb{N}}$, define the *discrete solution* $(u_\tau, v_\tau) : [0, \infty) \to \boldsymbol{X}$ as in (11) by piecewise constant interpolation:

$$(u_\tau, v_\tau)(t) := (u_\tau^n, v_\tau^n) \quad \text{for } t \in ((n-1)\tau, n\tau] \text{ and } n \geq 1. \tag{26}$$

We start be recalling a collection of estimates on $(u_\tau, v_\tau)$ that follows immediately from the construction by minimizing movements.

**Proposition 4.2** (classical estimates). *The following hold for $T > 0$:*

$$\mathcal{H}(u_\tau^n, v_\tau^n) \leq \mathcal{H}(u_0, v_0) < \infty \qquad \textit{for all } n \geq 0, \tag{27}$$

$$\sum_{n=1}^{\infty}\bigl(\boldsymbol{W}_2^2(u_\tau^n, u_\tau^{n-1}) + \|v_\tau^n - v_\tau^{n-1}\|_{L^2(\mathbb{R}^3)}^2\bigr) \leq 2\tau(\mathcal{H}(u_0, v_0) - \inf\mathcal{H}), \tag{28}$$

$$\boldsymbol{W}_2(u_\tau(s), u_\tau(t)) + \|v_\tau(s) - v_\tau(t)\|_{L^2(\mathbb{R}^3)} \leq 2[2(\mathcal{H}(u_0, v_0) - \inf\mathcal{H})\max(\tau, |t-s|)]^{1/2} \qquad \textit{for all } 0 \leq s, t \leq T, \tag{29}$$

*the infimum* $\inf\mathcal{H}$ *of* $\mathcal{H}$ *on* $\boldsymbol{X}$ *being finite.*

By the well known *JKO method* [Jordan et al. 1998], we derive an approximate weak formulation satisfied by $(u_\tau, v_\tau)$. The idea is to choose test functions $\eta, \gamma \in C_c^\infty(\mathbb{R}^3)$ and perturb the minimizer $(u_\tau^n, v_\tau^n)$ of the functional $\mathcal{H}_\tau(\,\cdot \mid u_\tau^{n-1}, v_\tau^{n-1})$ over an auxiliary time $s \geq 0$ as follows:

$$u_\tau^n \rightsquigarrow \mathsf{S}_s^{\mathrm{D}\eta} \# u_\tau^n, \qquad v_\tau^n \rightsquigarrow v + s\gamma.$$

Here $\mathsf{S}_{(\cdot)}^{\mathrm{D}\eta}$ is the flow on $\mathbb{R}^3$ generated by the gradient vector field $\mathrm{D}\eta$. Since the calculations are very similar to the ones performed in [Zinsl 2014], we skip the details and directly state the result.

**Lemma 4.3.** *For all $n \in \mathbb{N}$ and all test functions $\eta, \gamma \in C_c^\infty(\mathbb{R}^3)$ and $\psi \in C_c^\infty((0,\infty)) \cap C([0,\infty))$, the following* discrete weak formulation *holds*:

$$\begin{aligned}
0 = \int_0^\infty \int_{\mathbb{R}^3} [u_\tau(t,x)\eta(x) - v_\tau(t,x)\gamma(x)] \frac{\psi(\lfloor t/\tau \rfloor \tau) - \psi(\lfloor t/\tau \rfloor \tau + \tau)}{\tau} \,\mathrm{d}x\,\mathrm{d}t + O(\tau) \\
+ \int_0^\infty \int_{\mathbb{R}^3} \psi(\lfloor t/\tau \rfloor \tau)\big(-\tfrac{1}{2} u_\tau(t,x)^2 \Delta\eta(x) + u_\tau(t,x)\mathrm{D}W(x)\cdot \mathrm{D}\eta(x) \\
+ \mathrm{D}v_\tau(t,x)\cdot \mathrm{D}\gamma(x) + \kappa v_\tau(t,x)\gamma(x) \\
+ \varepsilon u_\tau(t,x)\phi'(v_\tau(t,x))[\gamma(x) + \mathrm{D}v_\tau(t,x)\cdot \mathrm{D}\eta(x)]\big)\,\mathrm{d}x\,\mathrm{d}t. \qquad (30)
\end{aligned}$$

Our goal for the rest of this section is to pass to the limit $\tau \downarrow 0$ in (30) and obtain the (time-continuous) weak formulation (12)–(13).

**4B.** ***Regularity of the discrete solution.*** Since the discrete weak formulation (30) contains nonlinear terms with respect to $u_\tau$ and $v_\tau$, further compactness estimates are needed to pass to the continuous time limit $\tau \to 0$. As a preparation, we state:

**Lemma 4.4** (flow interchange lemma [Matthes et al. 2009, Theorem 3.2]). *Let $\mathfrak{A}$ be a proper, lower semicontinuous and $\lambda$-geodesically convex functional on $(\boldsymbol{X}, \mathbf{dist})$, which is defined on $\boldsymbol{X} \cap L^2(\mathbb{R}^3) \times W^{1,2}(\mathbb{R}^3)$ at least. Further, assume that $\mathsf{S}_{(\cdot)}^{\mathfrak{A}}$ is a $\lambda$-flow for $\mathfrak{A}$. Then, the following holds for every $n \in \mathbb{N}$:*

$$\mathfrak{A}(u_\tau^n, v_\tau^n) + \tau \mathrm{D}^{\mathfrak{A}}\mathcal{H}(u_\tau^n, v_\tau^n) + \tfrac{1}{2}\lambda\,\mathbf{dist}^2((u_\tau^n, v_\tau^n), (u_\tau^{n-1}, v_\tau^{n-1})) \leq \mathfrak{A}(u_\tau^{n-1}, v_\tau^{n-1}).$$

*There*, $\mathrm{D}^{\mathfrak{A}}\mathcal{H}(w)$ *denotes the* dissipation *of the entropy $\mathcal{H}$ along $\mathsf{S}_{(\cdot)}^{\mathfrak{A}}$, i.e.,*

$$\mathrm{D}^{\mathfrak{A}}\mathcal{H}(w) := \limsup_{h \downarrow 0} \frac{\mathcal{H}(w) - \mathcal{H}(\mathsf{S}_h^{\mathfrak{A}}(w))}{h}.$$

The necessary additional regularity is provided by the following estimate on the minimizers of $\mathcal{H}_\tau$:

**Proposition 4.5** (additional regularity). *Let $(u,v), (\tilde{u}, \tilde{v}) \in \boldsymbol{X} \cap (L^2(\mathbb{R}^3) \times W^{1,2}(\mathbb{R}^3))$ with $(u,v) \in$* $\operatorname{argmin} \mathcal{H}_\tau(\,\cdot \mid \tilde{u}, \tilde{v})$. *Denoting $\mathscr{E}(u) := \int_{\mathbb{R}^3} u \log(u)\,\mathrm{d}x$ and $\mathscr{F}(v) := \int_{\mathbb{R}^3} (\frac{1}{2}|\mathrm{D}v|^2 + \frac{1}{2}\kappa v^2)\,\mathrm{d}x$, the following estimate holds for some constant $K > 0$:*

$$\begin{aligned}
&\|\mathrm{D}u\|_{L^2(\mathbb{R}^3)}^2 + \|\Delta v - \kappa v\|_{L^2(\mathbb{R}^3)}^2 \\
&\qquad \leq K\Big(\|u\|_{L^2(\mathbb{R}^3)}^2 + \|v\|_{W^{1,2}(\mathbb{R}^3)}^2 + \|\Delta W\|_{L^\infty(\mathbb{R}^3)} + \frac{1}{\tau}(\mathscr{E}(\tilde{u}) - \mathscr{E}(u) + \mathscr{F}(\tilde{v}) - \mathscr{F}(v))\Big). \qquad (31)
\end{aligned}$$

*Proof.* The method of proof used here is based on the flow interchange lemma (Lemma 4.4). The idea is to calculate the dissipation of $\mathscr{H}$ along the gradient flow of an auxiliary functional, namely the heat flow and the heat flow with decay, respectively.

Therefore, we recall that the functional $\mathscr{E}(u) := \int_{\mathbb{R}^3} u \log(u)\,\mathrm{d}x$ is 0-geodesically convex on $\mathscr{P}_2(\mathbb{R}^3)$ and its gradient flow $\mathsf{S}^{\mathscr{E}}_{(\cdot)}$ is the heat flow satisfying

$$\partial_s \mathsf{S}^{\mathscr{E}}_s(u) = \Delta \mathsf{S}^{\mathscr{E}}_s(u).$$

Moreover, with the evolution variational inequality (18), we deduce as in [Blanchet and Laurençot 2013; Zinsl 2014] by integration over time using that $\mathscr{E}$ is a Lyapunov functional along $\mathsf{S}^{\mathscr{E}}_{(\cdot)}$

$$\tfrac{1}{2}(\boldsymbol{W}_2^2(\mathsf{S}^{\mathscr{E}}_s(u),\tilde{u}) - \boldsymbol{W}_2^2(u,\tilde{u})) \le \int_0^s (\mathscr{E}(\tilde{u}) - \mathscr{E}(\mathsf{S}^{\mathscr{E}}_\sigma(u)))\,\mathrm{d}\sigma \le s[\mathscr{E}(\tilde{u}) - \mathscr{E}(\mathsf{S}^{\mathscr{E}}_s(u))]. \tag{32}$$

Analogous to that, $\mathscr{F}(v) := \int_{\mathbb{R}^3} (\frac{1}{2}|\mathrm{D}v|^2 + \frac{1}{2}\kappa v^2)\,\mathrm{d}x$ is $\kappa$-geodesically convex on $L^2(\mathbb{R}^3)$ and its gradient flow $\mathsf{S}^{\mathscr{F}}_{(\cdot)}$ is given by

$$\partial_s \mathsf{S}^{\mathscr{F}}_s(v) = \Delta \mathsf{S}^{\mathscr{F}}_s(v) - \kappa \mathsf{S}^{\mathscr{F}}_s(v).$$

The application of the evolution variational inequality (18) then shows

$$\tfrac{1}{2}\big(\|\mathsf{S}^{\mathscr{F}}_s(v) - \tilde{v}\|^2_{L^2(\mathbb{R}^3)} - \|v - \tilde{v}\|^2_{L^2(\mathbb{R}^3)}\big) \le \int_0^s (\mathscr{F}(\tilde{v}) - \mathscr{F}(\mathsf{S}^{\mathscr{F}}_\sigma(v)))\,\mathrm{d}\sigma \le s[\mathscr{F}(\tilde{v}) - \mathscr{F}(\mathsf{S}^{\mathscr{F}}_s(v))]. \tag{33}$$

Well known results of parabolic theory ensure that $(\mathsf{S}^{\mathscr{E}}_s(u), \mathsf{S}^{\mathscr{F}}_s(v)) \in \boldsymbol{X} \cap (L^2(\mathbb{R}^3) \times W^{1,2}(\mathbb{R}^3))$ if $(u, v) \in \boldsymbol{X} \cap (L^2(\mathbb{R}^3) \times W^{1,2}(\mathbb{R}^3))$. For the sake of clarity, we introduce the notation $(\mathscr{U}_s, \mathscr{V}_s) := (\mathsf{S}^{\mathscr{E}}_s(u), \mathsf{S}^{\mathscr{F}}_s(v))$ and calculate for $s > 0$

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}s}\mathscr{H}(\mathscr{U}_s, \mathscr{V}_s) \\
&= \int_{\mathbb{R}^3} \big([\mathscr{U}_s + W + \varepsilon\phi(\mathscr{V}_s)]\Delta\mathscr{U}_s + [-\Delta\mathscr{V}_s + \kappa\mathscr{V}_s + \varepsilon\mathscr{U}_s\phi'(\mathscr{V}_s)][\Delta\mathscr{V}_s - \kappa\mathscr{V}_s]\big)\,\mathrm{d}x \\
&= \int_{\mathbb{R}^3} \big(-|\mathrm{D}\mathscr{U}_s|^2 - \mathscr{U}_s\Delta W - (\Delta\mathscr{V}_s - \kappa\mathscr{V}_s)^2 - \varepsilon\phi'(\mathscr{V}_s)\mathrm{D}\mathscr{V}_s\cdot\mathrm{D}\mathscr{U}_s + \varepsilon\mathscr{U}_s\phi'(\mathscr{V}_s)[\Delta\mathscr{V}_s - \kappa\mathscr{V}_s]\big)\,\mathrm{d}x,
\end{aligned} \tag{34}$$

where the last line follows by integration by parts. An application of Young's inequality yields

$$\begin{aligned}
&\int_{\mathbb{R}^3} \big(-|\mathrm{D}\mathscr{U}_s|^2 - \mathscr{U}_s\Delta W - (\Delta\mathscr{V}_s - \kappa\mathscr{V}_s)^2 - \varepsilon\phi'(\mathscr{V}_s)\mathrm{D}\mathscr{V}_s\cdot\mathrm{D}\mathscr{U}_s + \varepsilon\mathscr{U}_s\phi'(\mathscr{V}_s)[\Delta\mathscr{V}_s - \kappa\mathscr{V}_s]\big)\,\mathrm{d}x \\
&\qquad \le \int_{\mathbb{R}^3} \big(-\tfrac{1}{2}|\mathrm{D}\mathscr{U}_s|^2 - \mathscr{U}_s\Delta W - \tfrac{1}{2}(\Delta\mathscr{V}_s - \kappa\mathscr{V}_s)^2 + \tfrac{1}{2}\varepsilon^2\phi'(0)^2(|\mathrm{D}\mathscr{V}_s|^2 + \mathscr{U}_s^2)\big)\,\mathrm{d}x.
\end{aligned}$$

Exploiting the monotonicity of the $L^2$ norm along $\mathsf{S}^{\mathscr{E}}_{(\cdot)}$ and of the $W^{1,2}$ norm along $\mathsf{S}^{\mathscr{F}}_{(\cdot)}$, one gets

$$\begin{aligned}
&\int_{\mathbb{R}^3} \big(-\tfrac{1}{2}|\mathrm{D}\mathscr{U}_s|^2 - \mathscr{U}_s\Delta W - \tfrac{1}{2}(\Delta\mathscr{V}_s - \kappa\mathscr{V}_s)^2 + \tfrac{1}{2}\varepsilon^2\phi'(0)^2(|\mathrm{D}\mathscr{V}_s|^2 + \mathscr{U}_s^2)\big)\,\mathrm{d}x \\
&\quad \le -\tfrac{1}{2}\|\mathrm{D}\mathscr{U}_s\|^2_{L^2(\mathbb{R}^3)} - \tfrac{1}{2}\|\Delta\mathscr{V}_s - \kappa\mathscr{V}_s\|^2_{L^2(\mathbb{R}^3)} + \|\Delta W\|_{L^\infty(\mathbb{R}^3)} + \tfrac{1}{2}\varepsilon^2\phi'(0)^2\big(\|\mathrm{D}v\|^2_{L^2(\mathbb{R}^3)} + \|u\|^2_{L^2(\mathbb{R}^3)}\big).
\end{aligned}$$

All in all, we have estimated the dissipation of $\mathscr{H}$ along $\mathsf{S}^{\mathscr{E}}_{(\cdot)}$ and $\mathsf{S}^{\mathscr{F}}_{(\cdot)}$:

$$\frac{\mathrm{d}}{\mathrm{d}s}\mathscr{H}(\mathscr{U}_s, \mathscr{V}_s) \le -\tfrac{1}{2}(\|\mathrm{D}\mathscr{U}_s\|_2^2 + \|\Delta\mathscr{V}_s - \kappa\mathscr{V}_s\|_2^2) + C\|u\|_2^2 + C\|v\|^2_{W^{1,2}(\mathbb{R}^3)} + \|\Delta W\|_\infty. \tag{35}$$

As a final step of the proof of Proposition 4.5, we use the minimizing property of $(u, v)$. Clearly,

$$0 \le \mathscr{H}_\tau(\mathscr{U}_s, \mathscr{V}_s \mid \tilde{u}, \tilde{v}) - \mathscr{H}_\tau(u, v \mid \tilde{u}, \tilde{v}).$$

We insert (32), (33) and (35) and obtain

$$\begin{aligned}\frac{1}{s}\int_0^s &\big(\|\mathrm{D}\mathscr{U}_\sigma\|^2_{L^2(\mathbb{R}^3)} + \|\Delta\mathscr{V}_\sigma - \kappa\mathscr{V}_\sigma\|^2_{L^2(\mathbb{R}^3)}\big)\,\mathrm{d}\sigma \\ &\le K\Big(\|u\|^2_{L^2(\mathbb{R}^3)} + \|v\|^2_{W^{1,2}(\mathbb{R}^3)} + \|\Delta W\|_{L^\infty(\mathbb{R}^3)} + \frac{1}{\tau}(\mathscr{E}(\tilde{u}) - \mathscr{E}(\mathscr{U}_s) + \mathscr{F}(\tilde{v}) - \mathscr{F}(\mathscr{V}_s))\Big),\end{aligned}$$

for some constant $K > 0$. Similar to [Blanchet and Laurençot 2013; Zinsl 2014; Blanchet 2013], passing to the lim inf as $s \to 0$ yields (31) by lower semicontinuity of norms and continuity of the entropies $\mathscr{E}$ and $\mathscr{F}$ along their respective gradient flows. □

**4C. *Compactness estimates and passage to continuous time.*** The following compactness estimates in addition to the results of Proposition 4.2 are needed to pass to the limit $\tau \to 0$ in the nonlinear terms of the discrete weak formulation (30) afterwards. The method of proof is essentially the same as in [Zinsl 2014, §7]. For the sake of brevity, the details are omitted here.

**Proposition 4.6** (additional a priori estimates). *Let $(u_\tau, v_\tau)$ be the discrete solution obtained by the minimizing movement scheme* (25). *Then the following hold for $T > 0$:*

$$\boldsymbol{m}_2(u_\tau^n) \le C_1 < \infty \quad \text{for all } n \le \lfloor T/\tau \rfloor, \tag{36}$$
$$\|u_\tau^n\|_{L^2(\mathbb{R}^3)} \le C_3 < \infty \quad \text{for all } n \ge 0, \tag{37}$$
$$\|v_\tau^n\|_{W^{1,2}(\mathbb{R}^3)} \le C_5 < \infty \quad \text{for all } n \ge 0, \tag{38}$$
$$\int_0^T \|u_\tau(t)\|^2_{W^{1,2}(\mathbb{R}^3)}\,\mathrm{d}t \le C_6 < \infty, \tag{39}$$
$$\int_0^T \|v_\tau(t)\|^2_{W^{2,2}(\mathbb{R}^3)}\,\mathrm{d}t \le C_7 < \infty, \tag{40}$$

*with constants $C_j > 0$ only depending on $T$ and the initial condition $(u_0, v_0)$.*

The estimates (36)–(38) are a consequence of those in Proposition 4.2. Also employing Proposition 4.5 yields (39)–(40).

The estimates of Propositions 4.2 and 4.6 enable us to prove the existence of the continuous-time limit of the discrete solution.

**Proposition 4.7** (continuous-time limit). *Let $(\tau_k)_{k\ge0}$ be a vanishing sequence of step sizes, i.e., $\tau_k \to 0$ as $k \to \infty$, and let $(u_{\tau_k}, v_{\tau_k})_{k\ge0}$ be the corresponding sequence of discrete solutions obtained by the minimizing movement scheme.*

*Then there exists a subsequence (nonrelabeled) such that, for fixed* $t \in [0, T]$, $u_{\tau_k}(t)$ *converges to a limit* $u(t)$ *narrowly in* $\mathcal{P}(\mathbb{R}^3)$ *and* $v_{\tau_k}(t)$ *converges to a limit* $v(t)$ *strongly in* $L^2(\mathbb{R}^3)$. *The second convergence is uniform with respect to* $t \in [0, T]$. *The limit curves satisfy* $u \in C^{1/2}([0, T], \mathcal{P}_2(\mathbb{R}^3))$ *and* $v \in C^{1/2}([0, T], L^2_+(\mathbb{R}^3))$. *Furthermore, the following additional convergence properties hold for* $k \to \infty$:

(a) $u_{\tau_k} \rightharpoonup u$ *weakly in* $L^2([0, T], W^{1,2}(\mathbb{R}^3))$,

(b) $v_{\tau_k} \rightharpoonup v$ *weakly in* $L^2([0, T], W^{2,2}(\mathbb{R}^3))$,

(c) $u_{\tau_k} \to u$ *strongly in* $L^2([0, T], L^2(\Omega))$ *for all bounded domains* $\Omega \subset \mathbb{R}^3$,

(d) $v_{\tau_k} \to v$ *strongly in* $L^2([0, T], W^{1,2}(\Omega))$ *for all bounded domains* $\Omega \subset \mathbb{R}^3$.

Now, to complete the proof of Theorem 1.1, one needs to verify that the obtained limit curve $(u, v)$ indeed satisfies the weak formulation (12)–(13). This will be omitted here for the sake of brevity.

## 5. The stationary solution

In this section, we provide the characterization of a stationary state of system (1)–(2) and prove some relevant properties.

**5A. *Existence and uniqueness.*** At first, we show existence and uniqueness of the stationary solution to system (1)–(2).

**Proposition 5.1.** *For each sufficiently small* $\varepsilon > 0$, *there exists a unique minimizer* $(u_\infty, v_\infty) \in \boldsymbol{X} \cap (W^{1,2}(\mathbb{R}^3) \times W^{2,2}(\mathbb{R}^3))$ *of* $\mathcal{H}$, *for which the following holds*: $(u_\infty, v_\infty)$ *is a stationary solution to* (1)–(2) *and to the Euler–Lagrange system*

$$\Delta v_\infty - \kappa v_\infty = \varepsilon u_\infty \phi'(v_\infty), \tag{41}$$

$$u_\infty = [U_\varepsilon - W - \varepsilon\phi(v_\infty)]_+, \tag{42}$$

*where* $U_\varepsilon \in \mathbb{R}$ *is chosen such that* $\|u_\infty\|_{L^1(\mathbb{R}^3)} = 1$ *and* $[\,\cdot\,]_+$ *denotes the positive part.*

*Moreover,* $v_\infty \in C^0(\mathbb{R}^3)$ *and there exists* $V > 0$ *independent of* $\varepsilon > 0$ *such that* $\|v_\infty\|_{L^\infty(\mathbb{R}^3)} \le V$.

*Proof.* We prove that $\mathcal{H}$ possesses a unique minimizer $(u_\infty, v_\infty)$. Let a minimizing sequence $(u_n, v_n)_{n\in\mathbb{N}}$ be given such that $\lim_{n\to\infty} \mathcal{H}(u_n, v_n) = \inf \mathcal{H} > -\infty$. As the sequence $(\mathcal{H}(u_n, v_n))_{n\in\mathbb{N}}$ is bounded, we can, by the same argument as in the proof of Proposition 3.1, extract a (nonrelabeled) subsequence, on which $(u_n)_{n\in\mathbb{N}}$ converges weakly in $L^2(\mathbb{R}^3)$ to some $u_\infty \in L^2_+(\mathbb{R}^3)$ and $(v_n)_{n\in\mathbb{N}}$ converges weakly in $W^{1,2}(\mathbb{R}^3)$ to some $v_\infty \in L^2_+(\mathbb{R}^3) \cap W^{1,2}(\mathbb{R}^3)$ as $n \to \infty$. By the same argument as in the proof of Proposition 3.1(b), $(u_\infty, v_\infty)$ is indeed a minimizer of $\mathcal{H}$ and hence an element of $\boldsymbol{X} \cap (L^2(\mathbb{R}^3) \times W^{1,2}(\mathbb{R}^3))$.

Since $(u_\infty, v_\infty) \in \operatorname{argmin} \mathcal{H}_\tau(\,\cdot \mid u_\infty, v_\infty)$ for arbitrary $\tau > 0$, Proposition 4.5 immediately yields

$$\|u_\infty\|^2_{W^{1,2}(\mathbb{R}^3)} + \|v_\infty\|^2_{W^{2,2}(\mathbb{R}^3)} \le V_0(\mathcal{H}(u_\infty, v_\infty) + \|\Delta W\|_{L^\infty(\mathbb{R}^3)} + V_1)$$

for some constants $V_0, V_1 > 0$. Because of the continuous embedding of $W^{2,2}(\mathbb{R}^3)$ into $C^0(\mathbb{R}^3)$, it follows that $\|v_\infty\|_{L^\infty(\mathbb{R}^3)} \le V$ for some $V > 0$.

Uniqueness of the minimizer is, by [Villani 2003, Theorem 5.32], a consequence of $\lambda'$-geodesic convexity of $\mathcal{H}$ with respect to the distance induced by $\|\cdot\|_{L^2\times L^2}$ for some $\lambda'>0$ as proved in Proposition 3.1(c).

It remains to show that there is a set of Euler–Lagrange equations characterizing $(u_\infty, v_\infty)$.

The following variational inequality holds thanks to the minimizing property of $(u_\infty, v_\infty)$:

$$0 \le \frac{\mathrm{d}^+}{\mathrm{d}s}\bigg|_{s=0} \mathcal{H}(u_\infty + s\tilde{u}, v_\infty + s\tilde{v}) \\ = \int_{\mathbb{R}^3} (u_\infty + W + \varepsilon\phi(v_\infty))\tilde{u}\,\mathrm{d}x + \int_{\mathbb{R}^3} (-\Delta v_\infty + \kappa v_\infty + \varepsilon u_\infty \phi'(v_\infty))\tilde{v}\,\mathrm{d}x \tag{43}$$

for arbitrary maps $\tilde{u}$ and $\tilde{v}$ such that $u_\infty + \tilde{u} \ge 0$ on $\mathbb{R}^3$ and $\int_{\mathbb{R}^3} \tilde{u}\,\mathrm{d}x = 0$.

First, we consider the second component and thus set $\tilde{u} = 0$ in (43). As there are no constraints on $v_\infty$, it is allowed to replace $\tilde{v}$ by $-\tilde{v}$ in (43), yielding equality and hence (41).

Second, we consider the first component and set $\tilde{v} = 0$ in (43). For arbitrary $\psi$ such that $\int_{\mathbb{R}^3} \psi\,\mathrm{d}x \le 1$ and $\psi + u_\infty \ge 0$ on $\mathbb{R}^3$, we put

$$\tilde{u}_\psi := \tfrac{1}{2}\psi - \tfrac{1}{2}u_\infty \int_{\mathbb{R}^3} \psi\,\mathrm{d}x$$

and observe that $u_\infty + u_\psi \ge 0$ on $\mathbb{R}^3$ and $\int_{\mathbb{R}^3} \tilde{u}_\psi\,\mathrm{d}x = 0$ since $u_\infty$ has mass equal to 1. By straightforward calculation, we obtain

$$0 \le \int_{\mathbb{R}^3} (u_\infty + W + \varepsilon\phi(v_\infty) - U_\varepsilon)\psi\,\mathrm{d}x \tag{44}$$

for all $\psi$ as above and the constant

$$U_\varepsilon := \int_{\mathbb{R}^3} (u_\infty^2 + W u_\infty + \varepsilon u_\infty \phi(v_\infty))\,\mathrm{d}x \in \mathbb{R}.$$

Fix $x \in \mathbb{R}^3$. If $u_\infty(x) > 0$, choosing $\psi$ supported on a small neighborhood of $x$ and replacing by $-\psi$ in (44) eventually yields

$$u_\infty(x) = U_\varepsilon - W(x) - \varepsilon\phi(v_\infty(x)).$$

If $u_\infty(x) = 0$, we obtain

$$U_\varepsilon - W(x) - \varepsilon\phi(v_\infty(x)) \le 0.$$

Hence, for all $x \in \mathbb{R}^3$,

$$u_\infty(x) = [U_\varepsilon - W(x) - \varepsilon\phi(v_\infty(x))]_+.$$ □

**5B.** ***Properties.*** As a preparation to prove some crucial regularity estimates on the stationary solution $(u_\infty, v_\infty)$, several properties of solutions to the elliptic partial differential equation $-\Delta h + \kappa h = f$ are needed.

Therefore, we introduce for $\kappa > 0$ the *Yukawa potential* (also called *screened Coulomb* or *Bessel potential*) $\boldsymbol{G}_\kappa$ by

$$\boldsymbol{G}_\kappa(x) := \frac{1}{4\pi|x|}\exp(-\sqrt{\kappa}|x|) \quad \text{for all } x \in \mathbb{R}^3 \setminus \{0\}. \tag{45}$$

Additionally, we define for $\sigma > 0$ the kernel $\boldsymbol{Y}_\sigma$ by

$$\boldsymbol{Y}_\sigma := \frac{1}{\sigma}\boldsymbol{G}_{1/\sigma}.$$

In subsequent parts of this work, we will need the *iterates* $\boldsymbol{Y}_\sigma^k$ for $k \in \mathbb{N}$ defined inductively by

$$\boldsymbol{Y}_\sigma^1 := \boldsymbol{Y}_\sigma, \qquad \boldsymbol{Y}_\sigma^{k+1} := \boldsymbol{Y}_\sigma * \boldsymbol{Y}_\sigma^k.$$

The relevant properties of $\boldsymbol{G}_\kappa$ and $\boldsymbol{Y}_\sigma$ are summarized in Lemma 5.2 below. For the proof, we refer to Appendix A.

**Lemma 5.2** (Yukawa potential). *The following statements hold for all $\kappa > 0$, $\sigma > 0$ and $k \in \mathbb{N}$:*

(a) *$\boldsymbol{G}_\kappa$ and $\boldsymbol{Y}_\sigma$ are the fundamental solutions to $-\Delta h + \kappa h = f$ and $-\sigma\Delta h + h = f$ on $\mathbb{R}^3$, respectively.*

(b) *Let $p > 1$. If $f \in L^p(\mathbb{R}^3)$, then $\boldsymbol{G}_\kappa * f \in W^{2,p}(\mathbb{R}^3)$ and*

$$\kappa\|\boldsymbol{G}_\kappa * f\|_{L^p(\mathbb{R}^3)} + \sqrt{\kappa}\|\mathrm{D}(\boldsymbol{G}_\kappa * f)\|_{L^p(\mathbb{R}^3)} + \|\mathrm{D}^2(\boldsymbol{G}_\kappa * f)\|_{L^p(\mathbb{R}^3)} \le C_p\|f\|_{L^p(\mathbb{R}^3)} \tag{46}$$

*for some $p$-dependent constant $C_p > 0$. (Note that this fact is not obvious as $\mathrm{D}^2(\boldsymbol{G}_\kappa) \notin L^1(\mathbb{R}^3, \mathbb{R}^{3\times 3})$.)*

(c) *For all $x \in \mathbb{R}^3 \setminus \{0\}$,*

$$\boldsymbol{Y}_\sigma(x) = \int_0^\infty \boldsymbol{H}_{\sigma t}(x)e^{-t}\,\mathrm{d}t,$$

*where $\boldsymbol{H}_t$ is the* heat kernel *on $\mathbb{R}^3$ at time $t > 0$, i.e.,*

$$\boldsymbol{H}_t(\xi) = t^{-3/2}\boldsymbol{H}_1(t^{-1/2}\xi) \quad \text{with } \boldsymbol{H}_1(\zeta) = (4\pi)^{-3/2}\exp\bigl(-\tfrac{1}{4}|\zeta|^2\bigr).$$

*Additionally, one has*

$$\boldsymbol{Y}_\sigma^k = \int_0^\infty \boldsymbol{H}_{\sigma r}\frac{r^{k-1}e^{-r}}{\Gamma(k)}\,\mathrm{d}r. \tag{47}$$

*Moreover, $\boldsymbol{Y}_\sigma^k \in W^{1,q}(\mathbb{R}^3)$ for each $q \in [1, \frac{3}{2})$, and there are universal constants $Y_q$ such that*

$$\|\mathrm{D}\boldsymbol{Y}_\sigma^k\|_{L^q(\mathbb{R}^3)} \le Y_q(\sigma k)^{-Q}, \quad \text{where } Q := 2 - \frac{3}{2q} \in \bigl[\tfrac{1}{2}, 1\bigr). \tag{48}$$

Now, we are in position to prove several estimates on the stationary solution.

**Proposition 5.3** (estimates on the stationary solution). *The following uniform estimates hold for all $x \in \mathbb{R}^3$:*

(a) *$u_\infty(x) \le U_0 - \varepsilon V\phi'(0)$, where $U_0 \in \mathbb{R}$ is chosen in such a way that $\int_{\mathbb{R}^3}[U_0 - W]_+\,\mathrm{d}x = 1$ and $V > 0$ is the constant from Proposition 5.1.*

(b) *$|\mathrm{D}v_\infty(x)| \le C\varepsilon$ for some constant $C > 0$.*

(c) *$-C'\varepsilon\mathbb{1} \le \mathrm{D}^2 v_\infty(x) \le C'\varepsilon\mathbb{1}$ in the sense of symmetric matrices, for some constant $C' > 0$.*

*Proof.* (a) We first prove that $U_\varepsilon \le U_0 + \varepsilon\phi(0)$, which in turn follows if

$$\int_{\mathbb{R}^3}[U_0 + \varepsilon\phi(0) - W - \varepsilon\phi(v_\infty)]_+\,\mathrm{d}x \ge 1.$$

One has

$$\int_{\mathbb{R}^3} [U_0 + \varepsilon\phi(0) - W - \varepsilon\phi(v_\infty)]_+ \,\mathrm{d}x = \int_{\{U_0 - W \geq 0\}} [U_0 - W + \varepsilon(\phi(0) - \phi(v_\infty))] \,\mathrm{d}x \\ + \int_{\{0 > U_0 - W \geq \varepsilon(\phi(v_\infty) - \phi(0))\}} [U_0 - W + \varepsilon(\phi(0) - \phi(v_\infty))] \,\mathrm{d}x. \quad (49)$$

From $\phi(0) - \phi(v_\infty) \geq 0$ and the definition of $U_0$, we deduce that the first term on the right-hand side of (49) is larger than or equal to 1. The second term on the right-hand side of (49) is nonnegative because the integrand is nonnegative on the domain of integration.

Now, if $u_\infty(x) > 0$ for some $x \in \mathbb{R}^3$, we also have due to convexity of $\phi$

$$u_\infty(x) \leq U_\varepsilon - W(x) - \varepsilon\phi(0) - \varepsilon v_\infty(x)\phi'(0) \leq U_0 + \varepsilon\phi(0) - \varepsilon\phi(0) - \varepsilon V\phi'(0),$$

from which the desired estimate follows.

(b) Define

$$f_v : \mathbb{R}^3 \to \mathbb{R}, \qquad f_v(x) := \varepsilon[U_\varepsilon - W(x) - \varepsilon\phi(v(x))]_+ \phi'(v(x)).$$

Then, $f_v \in L^\infty(\mathbb{R}^3)$ with compact support $\operatorname{supp}(f_v) \subset B_R(0)$ where $R > 0$ can be chosen independently of $\varepsilon \in (0, 1)$. Moreover, by Lemma 5.2(a), $(u_\infty, v_\infty)$ is the solution to the integral equation

$$v = -(\boldsymbol{G}_\kappa * f_v)$$

with the Yukawa potential $\boldsymbol{G}_\kappa$ defined in (45). Since $W^{2,4}(\mathbb{R}^3)$ is continuously embedded in $C^1(\mathbb{R}^3)$ [Zeidler 1990, Appendix, §(45) et seq.] and $f_v \in L^4(\mathbb{R}^3)$, we deduce from Lemma 5.2(b) that

$$\|v\|_{C^1(\mathbb{R}^3)} \leq \widetilde{C} \|f_v\|_{L^4(\mathbb{R}^3)},$$

for some constant $\widetilde{C} > 0$. Hence, we obtain (b) by using (a):

$$\|\mathrm{D}v_\infty\|_{L^\infty(\mathbb{R}^3)} \leq \widetilde{C} \|f_{v_\infty}\|_{L^4(\mathbb{R}^3)} \leq \widetilde{C}\varepsilon(U_0 - \varepsilon V\phi'(0))|\phi'(0)||B_R(0)|^{1/4} =: C\varepsilon.$$

(c) First, consider $x \in \mathbb{R}^3 \setminus B_{R+1}(0)$, where $R > 0$ is such that $\operatorname{supp}(f_{v_\infty}) \subset B_R(0)$. Smoothness of $\boldsymbol{G}_\kappa$ on $\mathbb{R}^3 \setminus \{0\}$ yields for all $i, j \in \{1, 2, 3\}$

$$|\partial_i \partial_j v_\infty(x)| = \left| \int_{B_R(0)} (\partial_{x_i} \partial_{x_j} \boldsymbol{G}_\kappa(x - y)) f_{v_\infty}(y) \,\mathrm{d}y \right| \\ = \left| \int_{x + B_R(0)} (\partial_i \partial_j \boldsymbol{G}_\kappa(z)) f_{v_\infty}(x - z) \,\mathrm{d}z \right|,$$

where the last equality follows by the transformation $z := x - y$. Obviously, we obtain the estimate

$$|\partial_i \partial_j v_\infty(x)| \leq \|f_{v_\infty}\|_{L^\infty(\mathbb{R}^3)} \int_{\mathbb{R}^3 \setminus B_1(0)} |\partial_i \partial_j \boldsymbol{G}_\kappa(z)| \,\mathrm{d}z.$$

Since, for $|z| \geq 1$, one has (see Appendix B for the derivatives of $\boldsymbol{G}_\kappa$)

$$|\partial_i \partial_j \boldsymbol{G}_\kappa(z)| \leq \frac{C(\kappa)\exp(-\sqrt{\kappa} z)}{4\pi |z|},$$

we arrive at

$$|\partial_i \partial_j v_\infty(x)| \le C(\kappa) \| f_{v_\infty} \|_{L^\infty(\mathbb{R}^3)} \int_1^\infty \exp(-\sqrt{\kappa} r) r \, \mathrm{d}r,$$

the last integral obviously being finite.

Consider now the case $|x| \le R+1$, and set $y := (R+2)e_1 \neq x$. By the triangular inequality, we have for $\alpha \in (0, 1)$ that

$$|\partial_i \partial_j v_\infty(x)| \le |\partial_i \partial_j v_\infty(y)| + \frac{|\partial_i \partial_j v_\infty(x) - \partial_i \partial_j v_\infty(y)|}{|x-y|^\alpha} |x-y|^\alpha.$$

By the arguments above, $f_{v_\infty}$ is $\alpha$-Hölder-continuous for some $\alpha \in (0, 1)$ since $u_\infty$ is Lipschitz-continuous and of compact support. By Lemma B.1 in Appendix B, we know that there exists $C > 0$ such that

$$[\partial_i \partial_j v_\infty]_{C^{0,\alpha}(\mathbb{R}^3)} \le C [f_{v_\infty}]_{C^{0,\alpha}(\mathbb{R}^3)}.$$

Hence, since $|x-y| \le 2R+3$, one has

$$|\partial_i \partial_j v_\infty(x)| \le |\partial_i \partial_j v_\infty(y)| + C(2R+3)^\alpha [f_{v_\infty}]_{C^{0,\alpha}(\mathbb{R}^3)}.$$

Combining both cases yields

$$\begin{aligned} |\partial_i \partial_j v_\infty(x)| &\le |\partial_i \partial_j v_\infty((R+2)e_1)| + C(2R+3)^\alpha [f_{v_\infty}]_{C^{0,\alpha}(\mathbb{R}^3)} \\ &\le C_0 \| f_{v_\infty} \|_{L^\infty(\mathbb{R}^3)} + C_1 \| f_{v_\infty} \|_{W^{1,\infty}(\mathbb{R}^3)} \end{aligned}$$

for some $C_0, C_1 > 0$ and *all* $x \in \mathbb{R}^3$. Using (a) and (b), it is straightforward to conclude that there exists $C_2 > 0$ with

$$(\| \mathrm{D} f \|_{C^0(\mathbb{R}^3)} + \| f \|_{L_\infty(\mathbb{R}^3)}) \le C_2 \varepsilon.$$

All in all, we proved the existence of $C_3 > 0$ such that for all $x \in \mathbb{R}^3$ and all $i, j \in \{1, 2, 3\}$

$$|\partial_i \partial_j v_\infty(x)| \le C_3 \varepsilon.$$

Obviously, this estimate yields the assertion (for a different constant $C' > 0$). □

## 6. Convergence to equilibrium

In this section, we prove Theorem 1.2. The strategy of proof is as follows. We first show that the entropy $\mathcal{H}(u, v) - \mathcal{H}_\infty$ can indeed be decomposed as in (16). Furthermore, the second component $v_\tau^n$ of the discrete solution admits a control estimate enabling us to prove boundedness of the auxiliary entropy $\mathcal{L}_u(u) + \mathcal{L}_v(v)$ in (16) for large times. From that, we can deduce an explicit temporal bound such that exponential decay to zero of this entropy occurs for sufficiently large times. The previous two steps essentially comprise a rigorous version of (17) from the introduction. Finally, these estimates are converted into the desired estimate for the continuous weak solution, completing the proof of Theorem 1.2.

Since our claim only concerns the solutions $(u, v)$ to (1)–(3) that are constructed as in the proof of Theorem 1.1, i.e., by the minimizing movement scheme, we assume in the following that we are given a family of time-discrete approximations $(u_\tau^n, v_\tau^n)_{n \in \mathbb{N}}$ that converge to the weak solution $(u, v)$ in the sense discussed in Section 4 as $\tau \downarrow 0$. Therefore, we may assume without loss of generality that $\tau > 0$ is sufficiently small.

Throughout this section, we shall use the abbreviation $[a]_\tau := (1/\tau)\log(1+a\tau)$, where $a > 0$. Note that, for every $\tau > 0$ and an index $m_\tau \in \mathbb{N}$ given such that $m_\tau \tau \geq T$ with a fixed $T \geq 0$,

$$(1+a\tau)^{-m_\tau} \leq e^{-[a]_\tau T} \downarrow e^{-aT} \quad \text{as } \tau \downarrow 0. \tag{50}$$

In order to keep track of the dependencies of certain quantities on $\varepsilon$, we are going to define several positive numbers $\varepsilon_j$ such that the estimates in a certain proof are uniform with respect to $\varepsilon \in (0, \varepsilon_j)$. When we want to emphasize that a quantity is independent of $\varepsilon \in (0, \varepsilon_j)$ — and also of $\tau$ and the initial condition $(u_0, v_0)$ — we call it a *system constant*. System constants are (in principle) expressible as a function of $\lambda_0$, $\kappa$ and $\phi$ and truly universal constants. Finally, we write $\mathcal{H}_\infty := \mathcal{H}(u_\infty, v_\infty)$.

**6A.** ***Decomposition of the entropy.*** The key element in the proof of Theorem 1.2 is the decomposition of the entropy functional as announced in (16). Introduce the *perturbed potential* $W_\varepsilon$ by

$$W_\varepsilon(x) := W(x) + \varepsilon\phi(v_\infty(x)). \tag{51}$$

Recall that $(u_\infty, v_\infty)$ is the minimizer of $\mathcal{H}$ on $\boldsymbol{X}$, and define

$$\begin{aligned} \mathcal{L}_u(u) &:= \int_{\mathbb{R}^3} \big(\tfrac{1}{2}(u^2 - u_\infty^2) + W_\varepsilon(u - u_\infty)\big)\,\mathrm{d}x, \\ \mathcal{L}_v(v) &:= \int_{\mathbb{R}^3} \tfrac{1}{2}\big(|\mathrm{D}(v - v_\infty)|^2 + \kappa(v - v_\infty)^2\big)\,\mathrm{d}x, \\ \mathcal{L}_*(u, v) &:= \int_{\mathbb{R}^3} \big(u[\phi(v) - \phi(v_\infty)] - u_\infty\phi'(v_\infty)[v - v_\infty]\big)\,\mathrm{d}x. \end{aligned}$$

Finally, let $\mathcal{L}(u, v) := \mathcal{L}_u(u) + \mathcal{L}_v(v)$ denote the *auxiliary entropy*.

**Lemma 6.1.** *The decomposition* (16) *holds*:

$$\mathcal{H}(u, v) - \mathcal{H}_\infty = \mathcal{L}(u, v) + \varepsilon\mathcal{L}_*(u, v).$$

*Proof.* By the properties of $\phi$ and the fact that $u_\infty$ has compact support, $\mathcal{L}_*$ is well defined on all of $\boldsymbol{X}$ while $\mathcal{L}_u$ and $\mathcal{L}_v$ are finite precisely on $(\mathcal{P}_2 \cap L^2)(\mathbb{R}^3)$ and $W^{1,2}(\mathbb{R}^3)$, respectively. Thus, both sides in (16) are finite on the same subset of $\boldsymbol{X}$. Now, for every such pair $(u, v)$, we have on the one hand that

$$\mathcal{L}_u(u) = \int_{\mathbb{R}^3} \big(\tfrac{1}{2}u^2 + uW + \varepsilon u\phi(v_\infty)\big)\,\mathrm{d}x - \int_{\mathbb{R}^3} \big(\tfrac{1}{2}u_\infty^2 + u_\infty W + \varepsilon u_\infty\phi(v_\infty)\big)\,\mathrm{d}x \tag{52}$$

and on the other hand that

$$\mathcal{L}_v(v) = \int_{\mathbb{R}^3} \big(\tfrac{1}{2}|\mathrm{D}v|^2 + \tfrac{1}{2}\kappa v^2\big)\,\mathrm{d}x + \int_{\mathbb{R}^3} \big(\tfrac{1}{2}|\mathrm{D}v_\infty|^2 + \tfrac{1}{2}\kappa v_\infty^2\big)\,\mathrm{d}x - \int_{\mathbb{R}^3} (\mathrm{D}v \cdot \mathrm{D}v_\infty + \kappa v v_\infty)\,\mathrm{d}x.$$

Integration by parts in the last integral yields, recalling the defining equation (41) for $v_\infty$, that

$$-\int_{\mathbb{R}^3} (\mathrm{D}v \cdot \mathrm{D}v_\infty + \kappa v v_\infty)\,\mathrm{d}x = \int_{\mathbb{R}^3} (\Delta v_\infty - \kappa v_\infty)v\,\mathrm{d}x = \varepsilon\int_{\mathbb{R}^3} u_\infty\phi'(v_\infty)v\,\mathrm{d}x.$$

Similarly, integration by parts in the middle integral leads to

$$\int_{\mathbb{R}^3}\big(\tfrac{1}{2}|\mathrm{D}v_\infty|^2+\tfrac{1}{2}\kappa v_\infty^2\big)\,\mathrm{d}x=-\frac{\varepsilon}{2}\int_{\mathbb{R}^3}u_\infty\phi'(v_\infty)v_\infty\,\mathrm{d}x.$$

And so,

$$\mathcal{L}_v(v)=\int_{\mathbb{R}^3}\big(\tfrac{1}{2}|\mathrm{D}v|^2+\tfrac{1}{2}\kappa v^2\big)\,\mathrm{d}x-\int_{\mathbb{R}^3}\big(\tfrac{1}{2}|\mathrm{D}v_\infty|^2+\tfrac{1}{2}\kappa v_\infty^2\big)\,\mathrm{d}x+\varepsilon\int_{\mathbb{R}^3}u_\infty\phi'(v_\infty)(v-v_\infty)\,\mathrm{d}x. \tag{53}$$

Combining (52) and (53) with the definition of $\mathcal{L}_*$ yields (16). □

We summarize some useful properties of the auxiliary entropy $\mathcal{L}$ in the following:

**Proposition 6.2** (properties of $\mathcal{L}$). *There are constants $K, L>0$ and some $\varepsilon_0>0$ such that the following are true for every $\varepsilon\in(0,\varepsilon_0)$:*

(a) *$W_\varepsilon\in C^2(\mathbb{R}^3)$ is $\lambda_\varepsilon$-convex with $\lambda_\varepsilon:=\lambda_0-L\varepsilon>0$.*

(b) *$\mathcal{L}_u$ is $\lambda_\varepsilon$-convex in $(\mathcal{P}_2(\mathbb{R}^3),\boldsymbol{W}_2)$, and for every $u\in(\mathcal{P}_2\cap W^{1,2})(\mathbb{R}^3)$, one has*

$$\tfrac{1}{2}\|u-u_\infty\|^2_{L^2(\mathbb{R}^3)}\le\mathcal{L}_u(u)\le\frac{1}{2\lambda_\varepsilon}\int_{\mathbb{R}^3}u|\mathrm{D}(u+W_\varepsilon)|^2\,\mathrm{d}x. \tag{54}$$

(c) *$\mathcal{L}_v$ is $\kappa$-convex in $L^2(\mathbb{R}^3)$, and for every $v\in W^{2,2}(\mathbb{R}^3)$, one has*

$$\tfrac{1}{2}\kappa\|v-v_\infty\|^2_{L^2(\mathbb{R}^3)}\le\mathcal{L}_v(v)\le\frac{1}{2\kappa}\int_{\mathbb{R}^3}(\Delta(v-v_\infty)-\kappa(v-v_\infty))^2\,\mathrm{d}x. \tag{55}$$

(d) *For every $(u,v)\in\boldsymbol{X}$,*

$$\mathcal{L}(u,v)\le(1+K\varepsilon)(\mathcal{H}(u,v)-\mathcal{H}_\infty). \tag{56}$$

*Proof.* (a) Since $W_\varepsilon=W+\varepsilon\phi(v_\infty)$, the chain rule yields

$$\mathrm{D}^2W_\varepsilon=\mathrm{D}^2W+\varepsilon\phi''(v_\infty)\mathrm{D}v_\infty\otimes\mathrm{D}v_\infty+\varepsilon\phi'(v_\infty)\mathrm{D}^2v_\infty.$$

Using our assumptions on $\phi$ and by Proposition 5.3, there are some $L>0$ and some $\varepsilon_0$ such that

$$\phi''(v_\infty)\mathrm{D}v_\infty\otimes\mathrm{D}v_\infty+\phi'(v_\infty)\mathrm{D}^2v_\infty\ge-L\mathbb{1}$$

holds uniformly with respect to $\varepsilon\in(0,\varepsilon_0)$. And thus also $\mathrm{D}^2W_\varepsilon\ge\lambda_\varepsilon\mathbb{1}$ with the indicated definition of $\lambda_\varepsilon$.

(b) Since $W_\varepsilon$ is $\lambda_\varepsilon$-convex, $\mathcal{L}_u$ is also $\lambda_\varepsilon$-geodesically convex in $\boldsymbol{W}_2$ because it is the sum of a 0-geodesically convex functional and a $\lambda_\varepsilon$-geodesically convex functional; see Theorem 2.1.

The Wasserstein subdifferential of $\mathcal{L}_u$ has been calculated in [Ambrosio et al. 2008, Lemma 10.4.1]. Together with (19), this shows the second inequality in (54). For the first inequality, observe that

$$\mathcal{L}_u(u)=\frac{1}{2}\int_{\mathbb{R}^3}(u-u_\infty)^2\,\mathrm{d}x+\int_{\mathbb{R}^3}(W_\varepsilon+u_\infty)(u-u_\infty)\,\mathrm{d}x.$$

It thus suffices to prove nonnegativity of the second integral term for all $u\in\mathcal{P}_2(\mathbb{R}^3)$. First, as $u$ and $u_\infty$ have equal mass, and by the definition of $u_\infty$,

$$0=\int_{\mathbb{R}^3}(u_\infty-u)\,\mathrm{d}x=\int_{\{U_\varepsilon-W_\varepsilon>0\}}u_\infty\,\mathrm{d}x-\int_{\mathbb{R}^3}u\,\mathrm{d}x,$$

and consequently,

$$\int_{\{U_\varepsilon - W_\varepsilon > 0\}} (u-u_\infty)\,\mathrm{d}x = -\int_{\{U_\varepsilon - W_\varepsilon \le 0\}} u\,\mathrm{d}x. \tag{57}$$

Also, by the definition of $u_\infty$,

$$\int_{\mathbb{R}^3} (W_\varepsilon + u_\infty)(u-u_\infty)\,\mathrm{d}x = \int_{\{U_\varepsilon - W_\varepsilon > 0\}} U_\varepsilon(u-u_\infty)\,\mathrm{d}x + \int_{\{U_\varepsilon - W_\varepsilon \le 0\}} W_\varepsilon u\,\mathrm{d}x.$$

Combining this with (57) yields

$$\int_{\{U_\varepsilon - W_\varepsilon > 0\}} U_\varepsilon(u-u_\infty)\,\mathrm{d}x + \int_{\{U_\varepsilon - W_\varepsilon \le 0\}} W_\varepsilon u\,\mathrm{d}x = \int_{\{U_\varepsilon - W_\varepsilon \le 0\}} (W_\varepsilon - U_\varepsilon)u\,\mathrm{d}x \ge 0$$

as the integrand is nonnegative on the domain of integration.

(c) This is an immediate consequence of (19) for the $L^2$ subdifferential of $\mathscr{L}_v$.

(d) Since $\phi$ is convex, we have

$$\phi(v) - \phi(v_\infty) - \phi'(v_\infty)[v-v_\infty] \ge 0,$$

and so we can estimate $\mathscr{L}_*$ from below by

$$\begin{aligned}\mathscr{L}_*(u,v) &= \int_{\mathbb{R}^3} (u-u_\infty)[\phi(v)-\phi(v_\infty)]\,\mathrm{d}x + \int_{\mathbb{R}^3} \big(\phi(v) - \phi(v_\infty) - \phi'(v_\infty)[v-v_\infty]\big)\,\mathrm{d}x \\ &\ge -\frac{1}{2}\int_{\mathbb{R}^3}(u-u_\infty)^2\,\mathrm{d}x - \frac{\phi'(0)^2}{2}\int_{\mathbb{R}^3}(v-v_\infty)^2\,\mathrm{d}x \\ &\ge -\mathscr{L}_u(u) - \frac{\phi'(0)^2}{\kappa}\mathscr{L}_v(v),\end{aligned}$$

using the properties (b) and (c) above. By (16), we conclude

$$(1-K'\varepsilon)\mathscr{L}(u,v) = \mathscr{H}(u,v) - \mathscr{H}_\infty \quad \text{with } K' := \max\left(1, \frac{\phi'(0)^2}{\kappa}\right),$$

which clearly implies (56) for all $\varepsilon \in (0, \varepsilon_0)$, possibly after diminishing $\varepsilon_0$. □

**6B.** ***Dissipation.*** We can now formulate the main a priori estimate for the time-discrete solution.

**Proposition 6.3.** *Given $(\tilde u, \tilde v) \in \boldsymbol{X}$ with $\mathscr{H}(\tilde u, \tilde v) < \infty$, let $(u, v) \in \boldsymbol{X}$ be a minimizer of the functional $\mathscr{H}_\tau(\,\cdot \mid \tilde u, \tilde v)$ introduced in* (24)*. Then*

$$\mathscr{L}_u(u) + \tau\mathscr{D}_u(u,v) \le \mathscr{L}_u(\tilde u), \qquad \mathscr{L}_v(v) + \tau\mathscr{D}_v(u,v) \le \mathscr{L}_v(\tilde v), \tag{58}$$

*where the dissipation terms are given by*

$$\mathscr{D}_u(u,v) = \left(1-\frac{\varepsilon}{2}\right)\int_{\mathbb{R}^3} u|\mathrm{D}(u+W_\varepsilon)|^2\,\mathrm{d}x - \frac{\varepsilon}{2}\int_{\mathbb{R}^3} u|\mathrm{D}(\phi(v)-\phi(v_\infty))|^2\,\mathrm{d}x, \tag{59}$$

$$\mathscr{D}_v(u,v) = \left(1-\frac{\varepsilon}{2}\right)\int_{\mathbb{R}^3} (\Delta(v-v_\infty) - \kappa(v-v_\infty))^2\,\mathrm{d}x - \frac{\varepsilon}{2}\int_{\mathbb{R}^3} (u\phi'(v) - u_\infty\phi'(v_\infty))^2\,\mathrm{d}x. \tag{60}$$

*Proof.* Naturally, these estimates are derived by means of the flow interchange lemma (Lemma 4.4).

For given $\nu > 0$, introduce the regularized functional $\mathcal{L}_u^\nu = \mathcal{L}_u + \nu\mathcal{E}$, where

$$\mathcal{E}(u) = \int_{\mathbb{R}^3} u \log u \, \mathrm{d}x.$$

Note that $\mathcal{E}$ is finite on $(\mathcal{P}_2 \cap L^2)(\mathbb{R}^3)$; see, e.g., [Zinsl 2014, Lemma 5.3]. Moreover, $\mathcal{L}_u^\nu$ is $\lambda_\varepsilon$-convex in $\boldsymbol{W}_2$ by Theorem 2.1. We claim that the $\lambda_\varepsilon$-flow associated to $\mathcal{L}_u^\nu$ satisfies the evolution equation

$$\partial_s \mathcal{U} = \nu \Delta \mathcal{U} + \tfrac{1}{2}\Delta \mathcal{U}^2 + \operatorname{div}(\mathcal{U} \mathrm{D} W_\varepsilon). \tag{61}$$

Since $\nu > 0$, this equation is strictly parabolic. Therefore, for every initial condition $\mathcal{U}_0 \in (\mathcal{P}_2 \cap L^2)(\mathbb{R}^3)$, there exists a smooth and positive solution $\mathcal{U} : \mathbb{R}_+ \times \mathbb{R}^3 \to \mathbb{R}$ such that $\mathcal{U}(s, \cdot) \to \mathcal{U}_0$ both in $\boldsymbol{W}_2$ and in $L^2(\mathbb{R}^3)$ as $s \downarrow 0$. By [Ambrosio et al. 2008, Theorem 11.2.8], the solution operator to (61) can be identified with the $\lambda_\varepsilon$-flow of $\mathcal{L}_u^\nu$.

Now, let $\mathcal{U}$ be the smooth solution to (61) with initial condition $\mathcal{U}_0 = u$. By smoothness of $\mathcal{U}$, the equation (61) is satisfied in the classical sense at every time $s > 0$, and the following integration by parts is justified:

$$\begin{aligned}
-\frac{\mathrm{d}}{\mathrm{d}s}\mathcal{H}(\mathcal{U}, v) &= -\int_{\mathbb{R}^3} [\mathcal{U} + W_\varepsilon + \varepsilon(\phi(v) - \phi(v_\infty))] \operatorname{div}[\mathcal{U}\mathrm{D}(\mathcal{U} + W_\varepsilon) + \nu \mathrm{D}\mathcal{U}] \, \mathrm{d}x \\
&= \int_{\mathbb{R}^3} \mathcal{U}|\mathrm{D}(\mathcal{U} + W_\varepsilon)|^2 \, \mathrm{d}x + \varepsilon \int_{\mathbb{R}^3} \mathcal{U}\mathrm{D}(\phi(v) - \phi(v_\infty)) \cdot \mathrm{D}(\mathcal{U} + W_\varepsilon) \, \mathrm{d}x \\
&\qquad + \nu \int_{\mathbb{R}^3} \mathrm{D}[\mathcal{U} + W + \varepsilon\phi(v)] \cdot \mathrm{D}\mathcal{U} \, \mathrm{d}x.
\end{aligned}$$

The very last integral has already been estimated in the proof of Proposition 4.5 (see (34) and following). Rewriting the middle integral by means of the elementary inequality

$$2ab \le a^2 + b^2, \tag{62}$$

we arrive at

$$\begin{aligned}
-\frac{\mathrm{d}}{\mathrm{d}s}\mathcal{H}(\mathcal{U}, v) \ge \Big(1 - \frac{\varepsilon}{2}\Big) \int_{\mathbb{R}^3} \mathcal{U}|\mathrm{D}(\mathcal{U} + W_\varepsilon)|^2 \, \mathrm{d}x - \frac{\varepsilon}{2} \int_{\mathbb{R}^3} \mathcal{U}|\mathrm{D}(\phi(v) - \phi(v_\infty))|^2 \, \mathrm{d}x \\
- \nu K \big(\|\mathcal{U}\|^2_{L^2(\mathbb{R}^3)} + \|v\|^2_{W^{1,2}(\mathbb{R}^3)}\big).
\end{aligned}$$

We pass to the limit $s \downarrow 0$. Recall that $\mathcal{U}$ converges (strongly) to its initial datum $\mathcal{U}_0 = u$ in $L^2(\mathbb{R}^3)$, and observe that the expressions on the right-hand side are lower semicontinuous with respect to that convergence. In fact, this is clear except perhaps for the first integral, which however can be rewritten, using integration by parts, in the form

$$\int_{\mathbb{R}^3} \mathcal{U}|\mathrm{D}(\mathcal{U} + W_\varepsilon)|^2 \, \mathrm{d}x = \frac{4}{9} \int_{\mathbb{R}^3} |\mathrm{D}\mathcal{U}^{3/2}|^2 \, \mathrm{d}x - \int_{\mathbb{R}^3} \mathcal{U}^2 \Delta W_\varepsilon \, \mathrm{d}x + \int_{\mathbb{R}^3} \mathcal{U}|\nabla W_\varepsilon|^2 \, \mathrm{d}x,$$

in which the lower semicontinuity is obvious since $\Delta W_\varepsilon \in L^\infty(\mathbb{R}^3)$. Applying now Lemma 4.4, we arrive at

$$\mathcal{L}_u^\nu(u) + (1-\varepsilon)\int_{\mathbb{R}^3} u|\mathrm{D}(u+W_\varepsilon)|^2\,\mathrm{d}x - \frac{\varepsilon}{2}\int_{\mathbb{R}^3} u|\mathrm{D}(\phi(v)-\phi(v_\infty))|^2\,\mathrm{d}x \\ \le \mathcal{L}_u^\nu(\tilde{u}) + \nu K\big(\|u\|^2_{L^2(\mathbb{R}^3)} + \|u\|^2_{W^{1,2}(\mathbb{R}^3)}\big).$$

Finally, passage to the limit $\nu \downarrow 0$ yields the dissipation (59).

The dissipation (60) is easier to obtain. It is immediate that the $\kappa$-flow in $L^2(\mathbb{R}^3)$ of $\mathcal{L}_v$ satisfies the linear parabolic evolution equation

$$\partial_s \mathcal{V} = \Delta(\mathcal{V} - v_\infty) - \kappa(\mathcal{V} - v_\infty). \tag{63}$$

Solutions $\mathcal{V}$ to (63) exist for arbitrary initial conditions $\mathcal{V}_0 \in L^2(\mathbb{R}^3)$, and they have at least the spatial regularity of $v_\infty$. Hence, with $\mathcal{V}_0 := v$, we have, also recalling the defining equation (41) for $v_\infty$,

$$-\frac{\mathrm{d}}{\mathrm{d}s}\mathcal{H}(u,\mathcal{V}) = \int_{\mathbb{R}^3}\big[\Delta(\mathcal{V}-v_\infty)-\kappa(\mathcal{V}-v_\infty)-\varepsilon(u\phi'(\mathcal{V})-u_\infty\phi'(v_\infty))\big]\cdot\big[\Delta(\mathcal{V}-v_\infty)-\kappa(\mathcal{V}-v_\infty)\big]\,\mathrm{d}x.$$

Another application of the elementary inequality (62) yields

$$-\frac{\mathrm{d}}{\mathrm{d}s}\mathcal{H}(u,\mathcal{V}) \ge \Big(1-\frac{\varepsilon}{2}\Big)\int_{\mathbb{R}^3}[\Delta(\mathcal{V}-v_\infty)-\kappa(\mathcal{V}-v_\infty)]^2\,\mathrm{d}x - \frac{\varepsilon}{2}\int_{\mathbb{R}^3}(u\phi'(\mathcal{V})-u_\infty\phi'(v_\infty))^2\,\mathrm{d}x.$$

We pass to the limit $s \downarrow 0$ so that $\mathcal{V}$ converges to $v$ in $L^2(\mathbb{R}^3)$. The first integral is obviously lower semicontinuous. Concerning the second integral, note that the integrand converges pointwise a.e. on $\mathbb{R}^3$ on a subsequence and that it is pointwise a.e. bounded by the integrable function $2\phi'(0)^2(u^2+u_\infty^2)$. Hence, we can pass to the limit using the dominated convergence theorem. Now another application of Lemma 4.4 yields the desired result. □

We will need below two further estimates for the dissipation terms from (59)–(60).

**Lemma 6.4.** *There is a constant $\theta > 0$ such that, for every $\varepsilon \in (0, \varepsilon_0)$ and every $u \in (\mathcal{P}_2 \cap W^{1,2})(\mathbb{R}^3)$,*

$$\|u\|^4_{L^3(\mathbb{R}^3)} \le \theta\Big(1 + \int_{\mathbb{R}^3} u|\mathrm{D}(u+W_\varepsilon)|^2\,\mathrm{d}x\Big). \tag{64}$$

*Proof.* Integrating by parts, it is easily seen that

$$\frac{4}{9}\int_{\mathbb{R}^3}\big|\mathrm{D}u^{3/2}\big|^2\,\mathrm{d}x + \int_{\mathbb{R}^3} u|\mathrm{D}W_\varepsilon|^2\,\mathrm{d}x = \int_{\mathbb{R}^3} u|\mathrm{D}(u+W_\varepsilon)|^2\,\mathrm{d}x + \int_{\mathbb{R}^3} u^2\Delta W_\varepsilon\,\mathrm{d}x.$$

By Proposition 5.3 on the regularity of $u_\infty$ and $v_\infty$, there exists a constant $C$ such that

$$\Delta W_\varepsilon = \Delta W + \varepsilon\phi'(v_\infty)\Delta v_\infty + \varepsilon\phi''(v_\infty)|\mathrm{D}v_\infty|^2 \le C \quad \text{on } \mathbb{R}^3$$

for all $\varepsilon \in (0, \varepsilon_1)$. Moreover,

$$\frac{1}{2}\int_{\mathbb{R}^3} u^2\,\mathrm{d}x \le \int_{\mathbb{R}^3} u_\infty^2\,\mathrm{d}x + \frac{1}{\lambda_\varepsilon}\int_{\mathbb{R}^3} u|\mathrm{D}(u+W_\varepsilon)|^2\,\mathrm{d}x$$

by (54). Invoking again Proposition 5.3, it follows that there exists an $\varepsilon$-uniform constant $C'$ such that

$$\|\mathrm{D}u^{3/2}\|_{L^2(\mathbb{R}^3)}^2 \leq C'\left(1 + \int_{\mathbb{R}^3} u|\mathrm{D}(u+W_\varepsilon)|^2\,\mathrm{d}x\right)$$

holds for all $u \in \mathscr{P}_2(\mathbb{R}^3)$. On the other hand, Hölder's and Sobolev's inequalities provide

$$\|u\|_{L^3(\mathbb{R}^3)} \leq \|u^{3/2}\|_{L^6(\mathbb{R}^3)}^{1/2}\|u\|_{L^1(\mathbb{R}^3)}^{1/4} \leq C''\|\mathrm{D}u^{3/2}\|_{L^2(\mathbb{R}^3)}^{1/2},$$

where we have used that $u$ is of unit mass. Together, this yields (64). □

**Lemma 6.5.** *For every $v \in W^{2,2}(\mathbb{R}^3)$,*

$$\min(1, 2\kappa, \kappa^2)\|v - v_\infty\|_{W^{2,2}(\mathbb{R}^3)}^2 \leq \int_{\mathbb{R}^3} (\Delta(v - v_\infty) - \kappa(v - v_\infty))^2\,\mathrm{d}x. \tag{65}$$

*Proof.* Set $\hat{v} := v - v_\infty$ for brevity. Integration by parts yields

$$\begin{aligned}\int_{\mathbb{R}^3} (\Delta\hat{v} - \kappa\hat{v})^2\,\mathrm{d}x &= \int_{\mathbb{R}^3} (\Delta\hat{v})^2\,\mathrm{d}x - 2\kappa\int_{\mathbb{R}^3} \hat{v}\Delta\hat{v}\,\mathrm{d}x + \kappa^2\int_{\mathbb{R}^3}\hat{v}^2\,\mathrm{d}x\\ &= \int_{\mathbb{R}^3} \|\mathrm{D}^2\hat{v}\|^2\,\mathrm{d}x + 2\kappa\int_{\mathbb{R}^3} |\mathrm{D}\hat{v}|^2\,\mathrm{d}x + \kappa^2\int_{\mathbb{R}^3}\hat{v}^2\,\mathrm{d}x,\end{aligned}$$

which clearly implies (65). □

**6C. *Control of the v component.*** For our estimates below, we need some preliminaries concerning solutions to the time-discrete heat equation. Here, we use the iterates $\boldsymbol{Y}_\sigma^k$ defined in (47) to write a semiexplicit representation of the components $v_\tau^n$ for a particular choice of $\sigma$.

**Lemma 6.6.** *For every $n \in \mathbb{N}$,*

$$v_\tau^n = (1+\kappa\tau)^{-n}\boldsymbol{Y}_\sigma^n * v_0 + \tau\sum_{m=1}^{n}(1+\kappa\tau)^{-m}\boldsymbol{Y}_\sigma^m * f_\tau^{n+1-m}, \tag{66}$$

*where we have set*

$$f_\tau^k := -\varepsilon u_\tau^k \phi'(v_\tau^k), \qquad \sigma := \frac{\tau}{1+\kappa\tau}.$$

*Proof.* We proceed by induction on $n$. By the flow interchange lemma (Lemma 4.4), using the auxiliary functional $\mathfrak{A}(u, v) := \int_{\mathbb{R}^3} \gamma v\,\mathrm{d}x$ for an arbitrary test function $\gamma \in C_c^\infty(\mathbb{R}^3)$, one sees by analogous (but easier) arguments as in the proof of (60) that $v_\tau^n$ is the — unique in $L^2(\mathbb{R}^3)$ — distributional solution to

$$v_\tau^n - \sigma\Delta v_\tau^n = (1+\kappa\tau)^{-1}v_\tau^{n-1} + \tau(1+\kappa\tau)^{-1}f_\tau^n.$$

Hence, it can be written as

$$v_\tau^n = (1+\kappa\tau)^{-1}\boldsymbol{Y}_\sigma * v_\tau^{n-1} + \tau(1+\kappa\tau)^{-1}\boldsymbol{Y}_\sigma * f_\tau^n.$$

For $n = 1$, this is (66) because $v_\tau^0 = v_0$. Now, if $n > 1$ and (66) holds with $n-1$ in place of $n$, then

$$v_\tau^n = (1+\kappa\tau)^{-n}\boldsymbol{Y}_\sigma * (\boldsymbol{Y}_\sigma^{n-1} * v_0) + \tau\sum_{m=1}^{n-1}(1+\kappa\tau)^{-(m+1)}\boldsymbol{Y}_\sigma * (\boldsymbol{Y}_\sigma^m * f_\tau^{n-m}) + \tau(1+\kappa\tau)^{-1}\boldsymbol{Y}_\sigma * f_\tau^n.$$

Using that $\boldsymbol{Y}_\sigma * (\boldsymbol{Y}_\sigma^k * f) = \boldsymbol{Y}_\sigma^{k+1} * f$, we obtain (66). □

We are now able to prove the main result of this section.

**Proposition 6.7.** *Provided that* $v_0 \in L^{6/5}(\mathbb{R}^3)$, *then* $\mathrm{D}v_\tau^n \in L^{6/5}(\mathbb{R}^3)$ *for every* $n \in \mathbb{N}$, *and the following estimate holds*:

$$\|\mathrm{D}v_\tau^n\|_{L^{6/5}(\mathbb{R}^3)} \le a\|v_0\|_{L^{6/5}(\mathbb{R}^3)} e^{-[\kappa]_\tau n\tau}(n\tau)^{-1/2} + \varepsilon M_1 \tag{67}$$

*with the system constants*

$$a := (1+\kappa)Y_1, \qquad M_1 := |\phi'(0)| Y_{6/5}(1+\kappa)^{3/4}\int_0^\infty (1+\kappa)^{-s}s^{-3/4}\,\mathrm{d}s. \tag{68}$$

*Proof.* From the representation formula (66), it follows that

$$\|\mathrm{D}v_\tau^n\|_{L^{6/5}(\mathbb{R}^3)} \le (1+\kappa\tau)^{-n}\|\mathrm{D}\boldsymbol{Y}_\sigma^n\|_{L^1(\mathbb{R}^3)}\|v_0\|_{L^{6/5}(\mathbb{R}^3)} + \tau\sum_{m=1}^{n}(1+\kappa\tau)^{-m}\|\mathrm{D}\boldsymbol{Y}_\sigma^m\|_{L^{6/5}(\mathbb{R}^3)}\|f_\tau^{n+1-m}\|_{L^1(\mathbb{R}^3)}.$$

Now apply estimate (48), once with $q := 1$ and $Q := \frac{1}{2}$ to the first term and once with $q := \frac{6}{5}$ and $Q := \frac{3}{4}$ to the second term on the right-hand side. Further, since $u_\tau^n$ is of unit mass, one has

$$\|f_\tau^k\|_{L^1(\mathbb{R}^3)} = \varepsilon\|u_\tau^k\phi'(v_\tau^k)\|_{L^1(\mathbb{R}^3)} \le \varepsilon|\phi'(0)|.$$

This yields

$$\|\mathrm{D}v_\tau^n\|_{L^{6/5}(\mathbb{R}^3)} \le Y_1\|v_0\|_{L^{6/5}(\mathbb{R}^3)}(1+\kappa\tau)^{-n}(\sigma n)^{-1/2} + \varepsilon|\phi'(0)|Y_{6/5}\,\tau\sum_{m=1}^{n}(1+\kappa\tau)^{-m}(\sigma m)^{-3/4}. \tag{69}$$

The sum in (69) is bounded uniformly in $n$ and $\tau$ because

$$\tau\sum_{m=1}^{\infty}(1+\kappa\tau)^{-m}(\sigma m)^{-3/4} \le (1+\kappa\tau)^{3/4}\int_0^\infty e^{-[\kappa]_\tau t}t^{-3/4}\,\mathrm{d}t.$$

Without loss of generality, we assume that $\tau \le 1$. By the monotone convergence $e^{-[\kappa]_\tau t} \downarrow e^{-\kappa t}$ as $\tau \downarrow 0$, we can estimate the sum in (69) as

$$\tau\sum_{m=1}^{\infty}(1+\kappa\tau)^{-m}(\sigma m)^{-3/4} \le (1+\tau)^{3/4}\int_0^\infty (1+\kappa)^{-t}t^{-3/4}\,\mathrm{d}t,$$

and the right-hand side is finite. Thus, (69) implies (67) with the given constants. □

In view of (50), we can draw the following conclusion from (67) with $\varepsilon_1 := \min(\varepsilon_0, \frac{1}{2})$, where $\varepsilon_0 > 0$ was implicitly characterized in Proposition 6.2:

**Corollary 6.8.** *Assume that* $v_0 \in L^{6/5}(\mathbb{R}^3)$, *and define*

$$T_1 := \max\Big(0, \frac{1}{\kappa}\log\frac{a\|v_0\|_{L^{6/5}(\mathbb{R}^3)}}{M_1}\Big) \tag{70}$$

*with the system constants* $a$ *and* $M_1$ *from* (68). *Then for every* $\varepsilon \in (0, \varepsilon_1)$, *for every sufficiently small* $\tau$ *and for every* $n$ *such that* $n\tau \ge T_1$, *one has*

$$\|\mathrm{D}v_\tau^n\|_{L^{6/5}(\mathbb{R}^3)} \le 2M_1. \tag{71}$$

**6D.** ***Bounds on the auxiliary entropy.*** We are now in position to prove the main estimate leading towards exponential decay and boundedness of the auxiliary entropy $\mathscr{L}$ along the discrete solution.

**Lemma 6.9.** *There are system constants $L'$ and $M'$ and an $\varepsilon_2 \in (0, \varepsilon_1)$ such that, for every $\varepsilon \in (0, \varepsilon_2)$, for every sufficiently small $\tau > 0$ and for every $n$ with $n\tau > T_1$, we have that*

$$(1 + 2\Lambda'_\varepsilon \tau)\mathscr{L}(u^n_\tau, v^n_\tau) \le \mathscr{L}(u^{n-1}_\tau, v^{n-1}_\tau) + \tau\varepsilon M' \tag{72}$$

*with $\Lambda'_\varepsilon := \min(\kappa, \lambda_0) - L'\varepsilon$.*

*Proof.* For brevity, we simply write $u$ and $v$ in place of $u^n_\tau$ and $v^n_\tau$, respectively, and we introduce $\hat{v} := v - v_\infty$. Since $n\tau > T_1$ by hypothesis, Corollary 6.8 implies that

$$\|\mathrm{D}\hat{v}\|_{L^{6/5}(\mathbb{R}^3)} \le \|\mathrm{D}v\|_{L^{6/5}(\mathbb{R}^3)} + \|\mathrm{D}v_\infty\|_{L^{6/5}(\mathbb{R}^3)} \le Z := 2M_1 + \sup_{0<\varepsilon<\varepsilon_1} \|\mathrm{D}v_\infty\|_{L^{6/5}(\mathbb{R}^3)} < \infty.$$

Now, since

$$|\mathrm{D}(\phi(v) - \phi(v_\infty))|^2 \le 2\phi'(v)^2|\mathrm{D}\hat{v}|^2 + 2(\phi'(v) - \phi'(v_\infty))^2|\mathrm{D}v_\infty|^2 \le \alpha|\mathrm{D}\hat{v}|^2 + \beta\hat{v}^2,$$

with the system constants

$$\alpha := 2\phi'(0)^2, \qquad \beta := 2\overline{\phi''}^2 \sup_{0<\varepsilon<\varepsilon_1} \|\mathrm{D}v_\infty\|^2_{L^\infty(\mathbb{R}^3)}, \tag{73}$$

we conclude that

$$\begin{aligned}
\int_{\mathbb{R}^3} u|\mathrm{D}(\phi(v) - \phi(v_\infty))|^2\,\mathrm{d}x &\le \alpha\int_{\mathbb{R}^3} u|\mathrm{D}\hat{v}|^2\,\mathrm{d}x + \beta\int_{\mathbb{R}^3} u\hat{v}^2\,\mathrm{d}x \\
&\le \alpha\|u\|_{L^3(\mathbb{R}^3)}\|\mathrm{D}\hat{v}\|^2_{L^3(\mathbb{R}^3)} + \beta\|u\|_{L^1(\mathbb{R}^3)}\|\hat{v}\|^2_{L^\infty(\mathbb{R}^3)} \\
&\le \|u\|^4_{L^3(\mathbb{R}^3)} + \alpha^{4/3}\|\mathrm{D}\hat{v}\|^{8/3}_{L^3(\mathbb{R}^3)} + \beta\|\hat{v}\|^2_{L^\infty(\mathbb{R}^3)} \\
&\le \|u\|^4_{L^3(\mathbb{R}^3)} + \alpha^{4/3}\big(S_1\|\hat{v}\|^{3/4}_{W^{2,2}(\mathbb{R}^3)}\|\mathrm{D}\hat{v}\|^{1/4}_{L^{6/5}(\mathbb{R}^3)}\big)^{8/3} + \beta S_2\|\hat{v}\|^2_{W^{2,2}(\mathbb{R}^3)} \\
&\le \theta\left(1 + \int_{\mathbb{R}^3} u|\mathrm{D}(u + W_\varepsilon)|^2\,\mathrm{d}x\right) + \frac{\alpha^{4/3}S_1^{8/3}Z^{2/3} + \beta S_2}{\min(1, 2\kappa, \kappa^2)}\int_{\mathbb{R}^3}(\Delta\hat{v} - \kappa\hat{v})^2\,\mathrm{d}x,
\end{aligned} \tag{74}$$

where $\theta$ is the constant from (64) and $S_1$ and $S_2$ are Sobolev constants. Next, observe that

$$(u\phi'(v) - u_\infty\phi'(v_\infty))^2 \le 2(u - u_\infty)^2\phi'(v)^2 + 2u_\infty^2(\phi'(v) - \phi'(v_\infty))^2 \le \alpha(u - u_\infty)^2 + \beta\|u_\infty\|^2_{L^\infty(\mathbb{R}^3)}\hat{v}^2$$

with the same constants as in (73). Therefore, using (54), (55) and Proposition 5.3(a),

$$\begin{aligned}
\int_{\mathbb{R}^3}(u\phi'(v) - u_\infty\phi'(v_\infty))^2\,\mathrm{d}x &\le \alpha\|u - u_\infty\|^2_{L^2(\mathbb{R}^3)} + \beta(U_0 - \varepsilon V\phi'(0))^2\|\hat{v}\|^2_{L^2(\mathbb{R}^3)} \\
&\le 2\alpha\mathscr{L}_u(u) + \frac{2\beta}{\kappa}(U_0 - \varepsilon V\phi'(0))^2\mathscr{L}_v(v).
\end{aligned}$$

Altogether, we have shown that there is a system constant $M'$ such that (recall the dissipation terms $\mathfrak{D}_u(u,v)$ and $\mathfrak{D}_v(u,v)$ from (59)–(60))

$$\mathfrak{D}_u(u,v)+\mathfrak{D}_v(u,v) \geq (1-M'\varepsilon)\int_{\mathbb{R}^3} u|\mathrm{D}(u+W_\varepsilon)|^2\,\mathrm{d}x + (1-M'\varepsilon)\int_{\mathbb{R}^3}(\Delta\hat{v}-\kappa\hat{v})^2\,\mathrm{d}x \\ - M'\varepsilon\mathcal{L}_u(u) - M'\varepsilon\mathcal{L}_v(v) - M'\varepsilon \quad (75)$$

for all $\varepsilon\in(0,\varepsilon_1)$. Provided that $M'\varepsilon<1$, we can apply (54) and (55) to estimate further:

$$\mathfrak{D}_u(u,v)+\mathfrak{D}_v(u,v) \geq \big(2\lambda_\varepsilon(1-M'\varepsilon)-M'\varepsilon\big)\mathcal{L}_u(u) + \big(2\kappa(1-M'\varepsilon)-M'\varepsilon\big)\mathcal{L}_v(v) - M'\varepsilon.$$

Finally, we can choose $\varepsilon_2\in(0,\varepsilon_1)$ so small that the coefficients of $\mathcal{L}_u$ and $\mathcal{L}_v$ above are nonnegative for every $\varepsilon\in(0,\varepsilon_2)$, and thus, we arrive at the final estimate

$$\mathfrak{D}_u(u,v)+\mathfrak{D}_v(u,v) \geq 2(\min(\kappa,\lambda_\varepsilon)-L'\varepsilon)\mathcal{L}(u,v) - \varepsilon M'$$

with a suitable choice of $L'$. Now estimate (58) implies (72) with $\Lambda'_\varepsilon$ given as above. □

Diminishing $\varepsilon_2$ such that the constant $1+K\varepsilon_2$ in (56) is less than or equal to 2, we derive the following explicit estimate:

**Proposition 6.10.** *Assume that $v_0\in L^{6/5}(\mathbb{R}^3)$, and let $T_1$ be defined as in* (70). *Then, for every $\varepsilon\in(0,\varepsilon_2)$, for every sufficiently small $\tau$ and for every $n$ with $n\tau>T_1$, the following estimate holds*:

$$\mathcal{L}(u_\tau^n, v_\tau^n) \leq 2(\mathcal{H}(u_0,v_0)-\mathcal{H}_\infty)e^{-2[\Lambda'_\varepsilon]_\tau(n\tau-T_1)} + \varepsilon M_2 \quad (76)$$

*with the system constant*

$$M_2 := \frac{M'}{2\inf_{0<\varepsilon<\varepsilon_2}\Lambda'_\varepsilon}.$$

*Proof.* We prove a slightly refined estimate: given $\bar{n}\in\mathbb{N}$ with $\bar{n}\tau\geq T_1$, we conclude by induction on $n\geq\bar{n}$ that

$$\mathcal{L}(u_\tau^n, v_\tau^n) \leq 2(\mathcal{H}(u_0,v_0)-\mathcal{H}_\infty)(1+2\Lambda'_\varepsilon\tau)^{-(n-\bar{n})} + \frac{M'\varepsilon}{2\Lambda'_\varepsilon}(1-(1+2\Lambda'_\varepsilon\tau)^{-(n-\bar{n})}), \quad (77)$$

which clearly implies (76). For $n=\bar{n}$, (77) is a consequence of (56) and the energy estimate (27). Now assume (77) for some $n\geq\bar{n}$, and apply the iterative estimate (72):

$$\begin{aligned}\mathcal{L}(u_\tau^{n+1}, v_\tau^{n+1}) &\leq (1-2\Lambda'_\varepsilon\tau)^{-1}\mathcal{L}(u_\tau^n,v_\tau^n) + (1+2\Lambda'_\varepsilon\tau)^{-1}\tau M'\varepsilon \\ &\leq 2(\mathcal{H}(u_0,v_0)-\mathcal{H}_\infty)(1+2\Lambda'_\varepsilon\tau)^{-((n+1)-\bar{n})} \\ &\quad + \frac{M'\varepsilon}{2\Lambda'_\varepsilon}\big((1+2\Lambda'_\varepsilon\tau)^{-1} - (1+2\Lambda'_\varepsilon\tau)^{-((n+1)-\bar{n})}\big) + (1+2\Lambda'_\varepsilon\tau)^{-1}\tau M'\varepsilon.\end{aligned}$$

Elementary calculations show that the last expression above equals the right-hand side of (77) with $n+1$ in place of $n$. □

Invoking again (50), we obtain the following analog to Corollary 6.8:

**Corollary 6.11.** *Assume that* $v_0 \in L^{6/5}(\mathbb{R}^3)$, *and define*

$$T_2 := T_1 + \max\left(0, \frac{1}{2\Lambda'_\varepsilon} \log \frac{2(\mathcal{H}(u_0, v_0) - \mathcal{H}_\infty)}{M_2}\right). \tag{78}$$

*Then for every* $\varepsilon \in (0, \varepsilon_2)$, *for every sufficiently small* $\tau$ *and for every* $n$ *such that* $n\tau \geq T_2$, *one has*

$$\mathcal{L}(u_\tau^n, v_\tau^n) \leq 2M_2. \tag{79}$$

We have thus proved that, for $t \geq T_2$, the auxiliary entropy $\mathcal{L}$ is bounded by a system constant. Next, we prove that $\mathcal{L}$ is not only bounded but actually convergent to zero exponentially fast.

### 6E. *Exponential decay for large times.*

**Lemma 6.12.** *There is a constant* $L'' > L'$ *and some* $\varepsilon_3 \in (0, \varepsilon_2)$ *such that, for every* $\varepsilon \in (0, \varepsilon_3)$, *for every sufficiently small* $\tau > 0$ *and for every* $n$ *such that* $n\tau > T_2$, *we have*

$$(1 + 2\Lambda''_\varepsilon \tau)\mathcal{L}(u_\tau^n, v_\tau^n) \leq \mathcal{L}(u_\tau^{n-1}, v_\tau^{n-1}) \tag{80}$$

*with* $\Lambda''_\varepsilon := \min(\lambda_0, \kappa) - L''\varepsilon$.

*Proof.* We proceed like in the proof of Lemma 6.9 with the following modification. By Corollary 6.11, we know that

$$\mathcal{L}_u(u_\tau^n) \leq \mathcal{L}(u_\tau^n, v_\tau^n) \leq 2M_2.$$

Using the first inequality in (54), we can estimate the $L^2$-norm of $u_\tau^n$ by a system constant:

$$\|u\|_{L^2(\mathbb{R}^3)} \leq \|u_\infty\|_{L^2(\mathbb{R}^3)} + \|u - u_\infty\|_{L^2(\mathbb{R}^3)} \leq Z := \sup_{0<\varepsilon<\varepsilon_2} \|u_\infty\|_{L^2(\mathbb{R}^3)} + 2\sqrt{M_2}.$$

This allows us to replace the chain of estimates (74) by a simpler one:

$$\int_{\mathbb{R}^3} u|\mathrm{D}(\phi(v) - \phi(v_\infty))|^2 \,\mathrm{d}x \leq \|u\|_{L^2(\mathbb{R}^3)}\big(\alpha\|\mathrm{D}\hat{v}\|^2_{L^4(\mathbb{R}^3)} + \beta\|\hat{v}\|^2_{L^4(\mathbb{R}^3)}\big)$$

with the constants from (73). Using the Sobolev inequalities

$$\|\mathrm{D}\hat{v}\|_{L^4(\mathbb{R}^3)} \leq S\|\hat{v}\|_{W^{2,2}(\mathbb{R}^3)}, \qquad \|\hat{v}\|_{L^4(\mathbb{R}^3)} \leq S\|\hat{v}\|_{W^{1,2}(\mathbb{R}^3)}$$

in combination with (65) and (55), respectively, we arrive at

$$\int_{\mathbb{R}^3} u|\mathrm{D}(\phi(v) - \phi(v_\infty))|^2 \,\mathrm{d}x \leq \frac{\alpha Z S^2}{\min(1, 2\kappa, \kappa^2)} \int_{\mathbb{R}^3} (\Delta\hat{v} - \kappa\hat{v})^2 \,\mathrm{d}x + \frac{2\beta Z S^2}{\min(1, \kappa)} \mathcal{L}_v(v).$$

This eventually leads to the dissipation estimate (75) again, with a different constant $M'$, but *without the constant term* $-\varepsilon M'$. By means of (58), this implies (80) for appropriate choices of $L''$ and $\varepsilon_3$. ☐

By iteration of (80), starting from (79), one immediately obtains

**Proposition 6.13.** *For all sufficiently small* $\tau$ *and every* $n$ *such that* $n\tau \geq T_2$, *we have*

$$\mathcal{L}(u_\tau^n, v_\tau^n) \leq 2M_2 e^{-2[\Lambda''_\varepsilon]_\tau (n\tau - T_2)}. \tag{81}$$

**6F.** ***Passage to continuous time.*** To complete the proof of Theorem 1.2, we consider the limit $\tau \downarrow 0$ of the estimates obtained above. Here $\tau \downarrow 0$ means that we consider a vanishing sequence $(\tau_k)_{k\in\mathbb{N}}$ such that the corresponding sequence of discrete solutions $(u_{\tau_k}, v_{\tau_k})_{k\in\mathbb{N}}$ converges in the sense of Section 4 to a weak solution $(u, v)$ to (1)–(3). Since the convergence of $(u_{\tau_k}, v_{\tau_k})_{k\in\mathbb{N}}$ in $\boldsymbol{X}$ is locally uniform on each compact time interval, the lower semicontinuity of $\mathscr{L}$ in $\boldsymbol{X}$ allows one to conclude that

$$\mathscr{L}(t) := \mathscr{L}(u(t), v(t)) \le \liminf_{\tau\downarrow 0} \mathscr{L}(u_\tau(t), v_\tau(t)) \quad \text{for every } t \ge 0.$$

We prove that

$$\mathscr{L}(t) \le C(1 + \|v_0\|_{L^{6/5}(\mathbb{R}^3)})^2(1 + \mathscr{H}(u_0, v_0) - \mathscr{H}_\infty)^2 e^{-2\Lambda''_\varepsilon t} \quad \text{for all } t \ge 0. \tag{82}$$

From this, claim (15) in Theorem 1.2 follows with $\Lambda_\varepsilon := \Lambda''_\varepsilon$.

Recalling (50), we conclude from (81) that

$$\mathscr{L}(t) \le 2M_2 e^{-2\Lambda''_\varepsilon (t-T_2)} \quad \text{for all } t \ge T_2. \tag{83}$$

Moreover, from (56) and the energy estimate (27), we obtain

$$\mathscr{L}(t) \le 2(\mathscr{H}(u_0, v_0) - \mathscr{H}_\infty) \quad \text{for all } t \ge 0. \tag{84}$$

We distinguish:

**Case 1** ($\mathscr{H}(u_0, v_0) - \mathscr{H}_\infty \le \frac{1}{2}M_2$). Then, from the definition of $T_2$ in (78), one has $T_2 = T_1$, and in consequence of (83),

$$\mathscr{L}(t) \le 2M_2 e^{-2\Lambda''_\varepsilon (t-T_1)} \quad \text{for all } t \ge T_2 = T_1.$$

Since further $\mathscr{L}(t) \le M_2$ for all $t \ge 0$ by (84), the first inequality extends to all times $t \ge 0$:

$$\mathscr{L}(t) \le 2M_2 e^{-2\Lambda''_\varepsilon (t-T_1)} \quad \text{for all } t \ge 0. \tag{85}$$

**Case 2** ($\mathscr{H}(u_0, v_0) - \mathscr{H}_\infty > \frac{1}{2}M_2$). Substitute

$$T_2 = T_1 + \frac{1}{2\Lambda'_\varepsilon} \log\left(\frac{2(\mathscr{H}(u_0, v_0) - \mathscr{H}_\infty)}{M_2}\right)$$

in (83) to find

$$\mathscr{L}(t) \le 2M_2 e^{-2\Lambda''_\varepsilon (t-T_1)} \left(\frac{2(\mathscr{H}(u_0, v_0) - \mathscr{H}_\infty)}{M_2}\right)^{\Lambda''_\varepsilon/\Lambda'_\varepsilon} \quad \text{for all } t \ge T_2 > T_1. \tag{86}$$

Using $\Lambda''_\varepsilon \le \Lambda'_\varepsilon$, we conclude from (86)

$$\mathscr{L}(t) \le 4(\mathscr{H}(u_0, v_0) - \mathscr{H}_\infty) e^{-2\Lambda''_\varepsilon (t-T_1)} \quad \text{for all } t \ge T_2 > T_1. \tag{87}$$

Define $A := 4(\mathscr{H}(u_0, v_0) - \mathscr{H}_\infty) \max(1, (\mathscr{H}(u_0, v_0) - \mathscr{H}_\infty)/M_2)$. Then, from (87) and the fact that $Ae^{-2\Lambda''_\varepsilon (T_2-T_1)} \ge 2(\mathscr{H}(u_0, v_0) - \mathscr{H}_\infty)$, we deduce

$$\mathscr{L}(t) \le Ae^{-2\Lambda''_\varepsilon (t-T_1)} \quad \text{for all } t \ge 0. \tag{88}$$

Together, (85) and (88) yield

$$\begin{aligned}\mathcal{L}(t) &\le \max(2M_2, 4(\mathcal{H}(u_0, v_0) - \mathcal{H}_\infty)) \max\Big(1, \frac{1}{M_2}(\mathcal{H}(u_0, v_0) - \mathcal{H}_\infty)\Big)e^{-2\Lambda''_\varepsilon(t-T_1)}\\ &\le \frac{2}{M_2}e^{2\Lambda''_\varepsilon T_1}\max(M_2, 2(\mathcal{H}(u_0, v_0) - \mathcal{H}_\infty))^2 e^{-2\Lambda''_\varepsilon t} \quad \text{for all } t \ge 0.\end{aligned}$$

Since $\kappa \ge \Lambda''_\varepsilon$, we have

$$e^{2\Lambda''_\varepsilon T_1} \le e^{2\kappa T_1} \le \max\bigg(1, \bigg[\frac{a}{M_1}\|v_0\|_{L^{6/5}(\mathbb{R}^3)}\bigg]^2\bigg)$$

and consequently (82):

$$\mathcal{L}(t) \le C(1 + \|v_0\|_{L^{6/5}(\mathbb{R}^3)})^2(1 + \mathcal{H}(u_0, v_0) - \mathcal{H}_\infty)^2 e^{-2\Lambda''_\varepsilon t} \quad \text{for all } t \ge 0.$$

## Appendix A: Proof of Lemma 5.2

(a) The proof of the first assertion can be found in [Lieb and Loss 2001, Theorem 6.23]. From that, the second one follows by elementary calculations.

(b) According to [Stein 1970, Chapter V, §3.3, Theorem 3], one has for $p > 1$

$$\|\boldsymbol{G}_1 * f\|_{W^{2,p}(\mathbb{R}^3)} \le C_p\|f\|_{L^p(\mathbb{R}^3)}. \tag{89}$$

To prove assertion (b), we use a rescaling of the equation $-\Delta h + \kappa h = f$ by $\tilde{x} := \sqrt{\kappa}x$. Consequently, $h(\tilde{x}) = (\boldsymbol{G}_\kappa * f)(\tilde{x}/\sqrt{\kappa})$ is a solution to $-\Delta_{\tilde{x}}h + h = f/\kappa$, i.e., $h(\tilde{x}) = (\boldsymbol{G}_1 * (f/\kappa))(\tilde{x})$. By the transformation theorem, we obtain

$$\begin{aligned}\bigg(\int_{\mathbb{R}^3}\bigg|\frac{f(\tilde{x})}{\kappa}\bigg|^p \mathrm{d}\tilde{x}\bigg)^{1/p} &= \kappa^{3/2-1}\|f\|_{L^p(\mathbb{R}^3)},\\ \bigg(\int_{\mathbb{R}^3}\bigg|\bigg(\boldsymbol{G}_1 * \frac{f}{\kappa}\bigg)(\tilde{x})\bigg|^p \mathrm{d}\tilde{x}\bigg)^{1/p} &= \kappa^{3/2}\|\boldsymbol{G}_\kappa * f\|_{L^p(\mathbb{R}^3)},\\ \bigg(\int_{\mathbb{R}^3}\bigg|\mathrm{D}_{\tilde{x}}\bigg(\boldsymbol{G}_1 * \frac{f}{\kappa}\bigg)(\tilde{x})\bigg|^p \mathrm{d}\tilde{x}\bigg)^{1/p} &= \kappa^{3/2-1/2}\|\mathrm{D}_x(\boldsymbol{G}_\kappa * f)\|_{L^p(\mathbb{R}^3)},\\ \bigg(\int_{\mathbb{R}^3}\bigg|\mathrm{D}^2_{\tilde{x}}\bigg(\boldsymbol{G}_1 * \frac{f}{\kappa}\bigg)(\tilde{x})\bigg|^p \mathrm{d}\tilde{x}\bigg)^{1/p} &= \kappa^{3/2-1}\|\mathrm{D}^2_x(\boldsymbol{G}_\kappa * f)\|_{L^p(\mathbb{R}^3)},\end{aligned}$$

which yields (46) after insertion into (89) and simplification.

(c) The first statement is a straightforward consequence of the integral-type representation of $\boldsymbol{G}_\kappa$ in [Lieb and Loss 2001, Theorem 6.23]. To prove the first claim of the second statement, we proceed by induction. For $k = 1$, Equation (47) is just the definition of $\boldsymbol{Y}_\sigma$. Now assume that (47) holds for some $k \in \mathbb{N}$. Using the semigroup property $\boldsymbol{H}_{t_1+t_2} = \boldsymbol{H}_{t_1} * \boldsymbol{H}_{t_2}$ of the heat kernel, we find that

$$\begin{aligned}\boldsymbol{Y}_\sigma^{k+1} &= \int_0^\infty \int_0^\infty \boldsymbol{H}_{\sigma r_1} * \boldsymbol{H}_{\sigma r_2} e^{-r_1} r_2^{k-1} e^{-r_2} \,\frac{\mathrm{d}r_1\,\mathrm{d}r_2}{\Gamma(k)} \\ &= \int_0^\infty \int_0^\infty \boldsymbol{H}_{\sigma(r_1+r_2)} e^{-(r_1+r_2)} r_2^{k-1} \,\frac{\mathrm{d}r_1\,\mathrm{d}r_2}{\Gamma(k)}.\end{aligned}$$

Now perform a change of variables

$$r := r_1 + r_2, \qquad s := r_2,$$

which is of determinant 1 and leads to

$$\boldsymbol{Y}_\sigma^{k+1} = \int_0^\infty \boldsymbol{H}_{\sigma r} e^{-r} \left( \int_0^r s^{k-1}\,\mathrm{d}s \right) \frac{\mathrm{d}r}{\Gamma(k)} = \int_0^\infty \boldsymbol{H}_{\sigma r} \frac{e^{-r} r^k\,\mathrm{d}r}{k\Gamma(k)},$$

which is (47) with $k+1$ in place of $k$, using that $k\Gamma(k) = \Gamma(k+1)$.

For (48), first observe that $r \mapsto r^{k-1} e^{-r} / \Gamma(k)$ defines a probability density on $\mathbb{R}_+$. We can thus apply Jensen's inequality to obtain

$$\|\mathrm{D}\boldsymbol{Y}_\sigma^k\|_{L^q(\mathbb{R}^3)} \le \int_0^\infty \|\mathrm{D}\boldsymbol{H}_{\sigma r}\|_{L^q(\mathbb{R}^3)} \frac{r^{k-1} e^{-r}\,\mathrm{d}r}{\Gamma(k)}. \tag{90}$$

The $L^q$-norm of $\mathrm{D}\boldsymbol{H}_{\sigma r}$ is easily evaluated using its definition,

$$\begin{aligned}\|\mathrm{D}\boldsymbol{H}_{\sigma r}\|_{L^q(\mathbb{R}^3)} &= (\sigma r)^{-3/2} \left( \int_{\mathbb{R}^3} |\mathrm{D}_\xi \boldsymbol{H}_1((\sigma r)^{-1/2}\xi)|^q \,\mathrm{d}\xi \right)^{1/q} \\ &= (\sigma r)^{-3/2} \left( \int_{\mathbb{R}^3} |(\sigma r)^{-1/2} \mathrm{D}_\zeta \boldsymbol{H}_1(\zeta)|^q \,(\sigma r)^{3/2}\,\mathrm{d}\zeta \right)^{1/q} = (\sigma r)^{-Q} \|\mathrm{D}\boldsymbol{H}_1\|_{L^q(\mathbb{R}^3)}.\end{aligned}$$

By definition of the gamma function, we thus obtain from (90) that

$$\|\mathrm{D}\boldsymbol{Y}_\sigma^k\|_{L^q(\mathbb{R}^3)} \le \|\mathrm{D}\boldsymbol{H}_1\|_{L^q(\mathbb{R}^3)} \frac{\Gamma(k-Q)}{\Gamma(k)} \sigma^{-Q}.$$

For further estimation, observe that the sequence $(a_k)_{k\in\mathbb{N}}$ with $a_k = k^Q \Gamma(k-Q)/\Gamma(k)$ is monotonically decreasing (to zero). Indeed,

$$\frac{a_{k+1}}{a_k} = \frac{(k+1)k^Q\,(k-Q)\Gamma(k-Q)\Gamma(k)}{k^Q\,k\Gamma(k)\Gamma(k-Q)} = \left(1 + \frac{1}{k}\right)^Q \left(1 - \frac{Q}{k}\right)$$

is always smaller than 1 since $\xi \mapsto (1+\xi)^{-Q}$ is convex. Therefore, $a_k \le a_1$ for all $k \in \mathbb{N}$, and so (48) follows with $Y_q := \Gamma(1-Q)\|\mathrm{D}\boldsymbol{H}_1\|_{L^q(\mathbb{R}^3)}$. □

## Appendix B: Hölder estimate for the kernel $\boldsymbol{G}_\kappa$

As a preparation, we calculate the derivatives of $\boldsymbol{G}_\kappa$ in $\mathbb{R}^3 \setminus \{0\}$. For all $i, j, k \in \{1, 2, 3\}$, one has

$$\partial_i \boldsymbol{G}_\kappa(x) = -\frac{1}{4\pi}\frac{\exp(-\sqrt{\kappa}|x|)}{|x|^3}(\sqrt{\kappa}|x|+1)x_i,$$

$$\partial_i\partial_j \boldsymbol{G}_\kappa(x) = -\frac{1}{4\pi}\exp(-\sqrt{\kappa}|x|)\bigg[\Big(\frac{\kappa}{|x|^3}+\frac{3\sqrt{\kappa}}{|x|^4}+\frac{3}{|x|^5}\Big)x_i x_j - \Big(\frac{\sqrt{\kappa}}{|x|^2}+\frac{1}{|x|^3}\Big)\delta_{ij}\bigg],$$

$$\begin{aligned}\partial_i\partial_j\partial_k \boldsymbol{G}_\kappa(x) = -\frac{1}{4\pi}\exp(-\sqrt{\kappa}|x|)\frac{-\sqrt{\kappa}x_k}{|x|}\bigg[\Big(\frac{\kappa}{|x|^3}+\frac{3\sqrt{\kappa}}{|x|^4}+\frac{3}{|x|^5}\Big)x_i x_j - \Big(\frac{\sqrt{\kappa}}{|x|^2}+\frac{1}{|x|^3}\Big)\delta_{ij}\bigg]\\ -\frac{1}{4\pi}\exp(-\sqrt{\kappa}|x|)\bigg[\Big(-\frac{3\kappa}{|x|^4}-\frac{12\sqrt{\kappa}}{|x|^5}-\frac{15}{|x|^6}\Big)\frac{x_i x_j x_k}{|x|}\\ +\delta_{ij}\Big(\frac{2\sqrt{\kappa}}{|x|^3}+\frac{3}{|x|^4}\Big)\frac{x_k}{|x|}+\Big(\frac{\kappa}{|x|^3}+\frac{3\sqrt{\kappa}}{|x|^4}+\frac{3}{|x|^5}\Big)(\delta_{ik}x_j+\delta_{jk}x_i)\bigg],\end{aligned}$$

where $\delta_{ij}$ denotes Kronecker's delta.

We prove the following:

**Lemma B.1** (Hölder estimate for second derivative). *Let* $f \in C^{0,\alpha}(\mathbb{R}^3)$ *for some* $\alpha \in (0, 1)$, *and assume that it is of compact support. Then, there exists* $C > 0$ *such that for all* $i, j \in \{1, 2, 3\}$ *the following estimate holds*:

$$[\partial_i\partial_j(\boldsymbol{G}_\kappa * f)]_{C^{0,\alpha}(\mathbb{R}^3)} \le C[f]_{C^{0,\alpha}(\mathbb{R}^3)}.$$

*Here*,

$$[g]_{C^{0,\alpha}(\mathbb{R}^3)} := \sup_{x,y\in\mathbb{R}^3,\ x\neq y}\frac{|g(x)-g(y)|}{|x-y|}$$

*denotes the* Hölder seminorm *of* $g : \mathbb{R}^3 \to \mathbb{R}$.

*Proof.* This result is an extension of the respective result for Poisson's equation (corresponding to $\kappa = 0$) proved by Lieb and Loss [2001, Theorem 10.3]. Their method of proof is adapted here. In the following, $C$ and $\widetilde{C}$ denote generic nonnegative constants.

The following holds for arbitrary test functions $\psi \in C_c^\infty(\mathbb{R}^3)$:

$$-\int_{\mathbb{R}^3}(\partial_j\psi)(x)(\partial_i v)(x)\,\mathrm{d}x = \int_{\mathbb{R}^3} f(y)\int_{\mathbb{R}^3}(\partial_j\psi)(x)\partial_{x_i}\boldsymbol{G}_\kappa(x-y)\,\mathrm{d}x\,\mathrm{d}y,$$

which can be rewritten by the dominated convergence theorem and integration by parts as

$$\begin{aligned}&\int_{\mathbb{R}^3} f(y)\int_{\mathbb{R}^3}(\partial_j\psi)(x)\partial_{x_i}\boldsymbol{G}_\kappa(x-y)\,\mathrm{d}x\,\mathrm{d}y\\ &= \lim_{\delta\to 0}\int_{\mathbb{R}^3} f(y)\int_{\mathbb{R}^3\setminus B_\delta(y)}(\partial_j\psi)(x)\partial_{x_i}\boldsymbol{G}_\kappa(x-y)\,\mathrm{d}x\,\mathrm{d}y\\ &= \lim_{\delta\to 0}\int_{\mathbb{R}^3} f(y)\bigg[-\int_{\partial B_\delta(y)}\psi(x)\partial_{x_i}\boldsymbol{G}_\kappa(x-y)e_j\cdot\nu_{y,\delta}(x)\,\mathrm{d}S(x)-\int_{\mathbb{R}^3\setminus B_\delta(y)}\psi(x)\partial_{x_i}\partial_{x_j}\boldsymbol{G}_\kappa(x-y)\,\mathrm{d}x\bigg]\mathrm{d}y,\end{aligned}$$

where $e_j$ is the $j$-th unit vector and $\nu_{y,\delta}(x) = (x-y)/\delta$ is the unit outward normal vector in $x$ on the sphere $\partial B_\delta(y)$.

The first part can be simplified explicitly by the transformation $z := (x-y)/\delta$:

$$-\int_{\partial B_\delta(y)} \psi(x)\partial_{x_i}\boldsymbol{G}_\kappa(x-y)e_j\cdot\nu_{y,\delta}(x)\,\mathrm{d}S(x) = \frac{1}{4\pi}\int_{\partial B_1(0)} \psi(\delta z+y)\exp(-\sqrt{\kappa}\delta)(\sqrt{\kappa}\delta+1)z_iz_j\,\mathrm{d}S(z),$$

which converges as $\delta\to 0$ to $\psi(y)\delta_{ij}/3$.

For the second part, we split the domain of integration $\mathbb{R}^3\setminus B_\delta(y)$ into two parts:

$$\begin{aligned}\int_{\mathbb{R}^3\setminus B_\delta(y)} &\psi(x)\partial_{x_i}\partial_{x_j}\boldsymbol{G}_\kappa(x-y)\,\mathrm{d}x\\ &= \int_{\mathbb{R}^3\setminus B_\delta(1)} \psi(x)\partial_{x_i}\partial_{x_j}\boldsymbol{G}_\kappa(x-y)\,\mathrm{d}x + \int_{\{1\ge|x-y|\ge\delta\}} \psi(x)\partial_{x_i}\partial_{x_j}\boldsymbol{G}_\kappa(x-y)\,\mathrm{d}x.\end{aligned}$$

We use integration by parts to insert convenient additional terms:

$$\begin{aligned}\int_{\mathbb{R}^3\setminus B_\delta(1)} &\psi(x)\partial_{x_i}\partial_{x_j}\boldsymbol{G}_\kappa(x-y)\,\mathrm{d}x + \int_{\{1\ge|x-y|\ge\delta\}} \psi(x)\partial_{x_i}\partial_{x_j}\boldsymbol{G}_\kappa(x-y)\,\mathrm{d}x\\ &= \int_{\mathbb{R}^3\setminus B_\delta(1)} \psi(x)\partial_{x_i}\partial_{x_j}\boldsymbol{G}_\kappa(x-y)\,\mathrm{d}x + \int_{\{1\ge|x-y|\ge\delta\}} \psi(x)\partial_{x_i}\partial_{x_j}\boldsymbol{G}_\kappa(x-y)\,\mathrm{d}x\\ &\quad - \int_{\{1\ge|x-y|\ge\delta\}} \psi(y)\partial_{x_i}\partial_{x_j}\boldsymbol{G}_\kappa(x-y)\,\mathrm{d}x\\ &\quad + \int_{\partial B_1(y)} \psi(y)\partial_{x_j}\boldsymbol{G}_\kappa(x-y)e_i\cdot\nu_{y,1}(x)\,\mathrm{d}S(x) - \int_{\partial B_\delta(y)} \psi(y)\partial_{x_j}\boldsymbol{G}_\kappa(x-y)e_i\cdot\nu_{y,\delta}(x)\,\mathrm{d}S(x).\end{aligned}$$

Now we calculate again explicitly and obtain in the limit $\delta\to 0$

$$\begin{aligned}\int_{\partial B_1(y)} \psi(y)\partial_{x_j}\boldsymbol{G}_\kappa(x-y)e_i\cdot\nu_{y,1}(x)\,\mathrm{d}S(x) - \int_{\partial B_\delta(y)} \psi(y)\partial_{x_j}\boldsymbol{G}_\kappa(x-y)e_i\cdot\nu_{y,\delta}(x)\,\mathrm{d}S(x)\\ \to -\tfrac{1}{3}\delta_{ij}[\exp(-\sqrt{\kappa})(\sqrt{\kappa}+1)-1]\psi(y).\end{aligned}$$

In summary, one gets

$$\begin{aligned}-\int_{\mathbb{R}^3}(\partial_j\psi)(x)(\partial_i v)(x)\,\mathrm{d}x = \int_{\mathbb{R}^3}\psi(x)\Big[\tfrac{1}{3}\delta_{ij}f(x)\exp(-\sqrt{\kappa})(\sqrt{\kappa}+1) + \int_{\mathbb{R}^3\setminus B_1(x)} f(y)\partial_{x_i}\partial_{x_j}\boldsymbol{G}_\kappa(x-y)\,\mathrm{d}y\\ + \lim_{\delta\to 0}\int_{\{1\ge|x-y|\ge\delta\}}(f(x)-f(y))\partial_{x_i}\partial_{x_j}\boldsymbol{G}_\kappa(x-y)\,\mathrm{d}y\Big]\,\mathrm{d}x.\end{aligned}$$

From $\alpha$-Hölder continuity of $f$, we conclude that, independent of $\delta$,

$$\mathbb{1}_{\{1\ge|x-y|\ge\delta\}}(y)\big|[f(x)-f(y)]\partial_{x_i}\partial_{x_j}\boldsymbol{G}_\kappa(x-y)\big| \le C|x-y|^{\alpha-3},$$

which is integrable as $\alpha - 3 + 2 > -1$. So using again the dominated convergence theorem, we have, with [Lieb and Loss 2001, Theorem 6.10],

$$(\partial_i \partial_j v)(x) = \tfrac{1}{3}\delta_{ij} \exp(-\sqrt{\kappa})(\sqrt{\kappa}+1) \\ + \int_{\mathbb{R}^3 \setminus B_1(x)} f(y) \partial_{x_i} \partial_{x_j} \boldsymbol{G}_\kappa(x-y)\,\mathrm{d}y + \int_{B_1(x)} [f(x) - f(y)] \partial_{x_i} \partial_{x_j} \boldsymbol{G}_\kappa(x-y)\,\mathrm{d}y. \quad (91)$$

Obviously, the first term in (91) is Hölder-continuous. For the second term in (91), we obtain for all $x, z \in \mathbb{R}^3$, $x \neq z$,

$$\left| \int_{\mathbb{R}^3 \setminus B_1(x)} f(y) \partial_{x_i} \partial_{x_j} \boldsymbol{G}_\kappa(x-y)\,\mathrm{d}y - \int_{\mathbb{R}^3 \setminus B_1(z)} f(y) \partial_{z_i} \partial_{z_j} \boldsymbol{G}_\kappa(z-y)\,\mathrm{d}y \right| \\ = \left| \int_{B_1(0)} [f(z-a) - f(x-a)] \partial_{a_i} \partial_{a_j} \boldsymbol{G}_\kappa(a)\,\mathrm{d}a \right|$$

by the transformations $a := x - y$ in the first and $a := z - y$ in the second integrals. From $\alpha$-Hölder continuity of $f$, we get the estimate

$$\left| \int_{B_1(0)} [f(z-a) - f(x-a)] \partial_{a_i} \partial_{a_j} \boldsymbol{G}_\kappa(a)\,\mathrm{d}a \right| \le C|x-z|^\alpha \int_{\mathbb{R}^3 \setminus B_1(x)} |\partial_{a_i} \partial_{a_j} \boldsymbol{G}_\kappa(a)\,\mathrm{d}a|,$$

where the integral on the right-hand side is finite because $\partial_{a_i} \partial_{a_j} \boldsymbol{G}_\kappa(a)$ behaves as $r^{-1}\exp(-r)$ for $r \to \infty$, which is integrable.

The same integral transformation yields for the third term in (91)

$$\left| \int_{B_1(x)} [f(x) - f(y)] \partial_{x_i} \partial_{x_j} \boldsymbol{G}_\kappa(x-y)\,\mathrm{d}y - \int_{B_1(z)} [f(z) - f(y)] \partial_{z_i} \partial_{z_j} \boldsymbol{G}_\kappa(z-y)\,\mathrm{d}y \right| \\ = \left| \int_{B_1(0)} [f(z) - f(z-a) - f(x) + f(x-a)] \partial_{a_i} \partial_{a_j} \boldsymbol{G}_\kappa(a)\,\mathrm{d}a \right|.$$

We now proceed as in [Lieb and Loss 2001] and write $B_1(0) = A \cup B$ with

$$A := \{a : 0 \le |a| < 4|x-z|\}, \\ B := \{a : 4|x-z| < |a| < 1\},$$

where $B = \varnothing$ for $|x-z| \ge \frac{1}{4}$, and calculate, using that $|\partial_{a_i} \partial_{a_j} \boldsymbol{G}_\kappa(a)| \le C|a|^{-3}$,

$$\left| \int_A [f(z) - f(z-a) - f(x) + f(x-a)] \partial_{a_i} \partial_{a_j} \boldsymbol{G}_\kappa(a)\,\mathrm{d}a \right| \le \int_A 2C|a|^{\alpha-3}\,\mathrm{d}a = \widetilde{C}|x-z|^\alpha.$$

It remains to consider the case $|x-z| < \frac{1}{4}$. One has

$$\left| \int_B [f(z) - f(x)] \partial_{a_i} \partial_{a_j} \boldsymbol{G}_\kappa(a)\,\mathrm{d}a \right| = \left| \int_{\partial B} [f(z) - f(x)] \partial_{a_j} \boldsymbol{G}_\kappa(a) e_i \cdot \nu(a)\,\mathrm{d}S(a) \right|,$$

and by similar arguments as above,

$$\left|\int_{\partial B}[f(z)-f(x)]\partial_{a_j}\boldsymbol{G}_\kappa(a)e_i\cdot\nu(a)\,\mathrm{d}S(a)\right|$$
$$=\tfrac{1}{3}\delta_{ij}|f(z)-f(x)|\left|\exp(-\sqrt{\kappa})(\sqrt{\kappa}+1)-\exp(-4\sqrt{\kappa}|x-z|)(4\sqrt{\kappa}|x-z|+1)\right|.$$

Note that the real-valued map $[0,\infty)\ni r\mapsto\exp(-\sqrt{\kappa}r)(\sqrt{\kappa}r+1)$ is monotonically decreasing. This yields

$$\left|\int_B[f(z)-f(x)]\partial_{a_i}\partial_{a_j}\boldsymbol{G}_\kappa(a)\,\mathrm{d}a\right|\le C|z-x|^\alpha.$$

By the transformations $b:=x-a-z$ and $b:=-a$, we get

$$\left|\int_B[f(x-a)-f(z-a)]\partial_{a_i}\partial_{a_j}\boldsymbol{G}_\kappa(a)\,\mathrm{d}a\right|$$
$$=\left|\int_B f(z+b)\partial_{b_i}\partial_{b_j}\boldsymbol{G}_\kappa(b)\,\mathrm{d}b-\int_D f(b+z)\partial_{b_i}\partial_{b_j}\boldsymbol{G}_\kappa(b-x+z)\,\mathrm{d}b\right| \tag{92}$$

with $D:=\{b:4|x-z|<|b-x+z|<1\}$.

Note that

$$\int_B\partial_{b_i}\partial_{b_j}\boldsymbol{G}_\kappa(b)\,\mathrm{d}b=\int_D\partial_{b_i}\partial_{b_j}\boldsymbol{G}_\kappa(b-x+z)\,\mathrm{d}b.$$

This enables us to rewrite (92) as

$$\left|\int_B f(z+b)\partial_{b_i}\partial_{b_j}\boldsymbol{G}_\kappa(b)\,\mathrm{d}b-\int_D f(b+z)\partial_{b_i}\partial_{b_j}\boldsymbol{G}_\kappa(b-x+z)\,\mathrm{d}b\right|$$
$$=\left|\int_B[f(z+b)-f(z)]\partial_{b_i}\partial_{b_j}\boldsymbol{G}_\kappa(b)\,\mathrm{d}b-\int_D[f(z+b)-f(z)]\partial_{b_i}\partial_{b_j}\boldsymbol{G}_\kappa(b-x+z)\,\mathrm{d}b\right|. \tag{93}$$

We consider (93) separately on the sets $B\cap D$, $B\setminus D$ and $D\setminus B$.

Note that, by the triangular inequality, $B\cap D\subset\{b:3|x-z|<|b|<1+|x-z|\}$ and by Taylor's theorem

$$(\partial_{b_i}\partial_{b_j}\boldsymbol{G}_\kappa)(b)-(\partial_{b_i}\partial_{b_j}\boldsymbol{G}_\kappa)(b-x+z)=\sum_{k=1}^3(\partial_k\partial_i\partial_j\boldsymbol{G}_\kappa)(b^*)(x_k-z_k)$$

for some $b^*=b-\beta(x-z)$ with $\beta\in(0,1)$. Therefore, one has, by the triangular inequality, $|b^*|\ge|b|-\beta|x-z|\ge(1-\frac{1}{3}\beta)|b|\ge\frac{2}{3}|b|$ on $B\cap D$ and consequently

$$|(\partial_{b_i}\partial_{b_j}\boldsymbol{G}_\kappa)(b)-(\partial_{b_i}\partial_{b_j}\boldsymbol{G}_\kappa)(b-x+z)|\le C|b^*|^{-4}|x-z|\le\widetilde{C}|b|^{-4}|x-z|.$$

This allows us to estimate

$$\left|\int_{B\cap D}[f(z+b)-f(z)][\partial_{b_i}\partial_{b_j}\boldsymbol{G}_\kappa(b)-\partial_{b_i}\partial_{b_j}\boldsymbol{G}_\kappa(b-x+z)]\,\mathrm{d}b\right|$$

$$\le C|x-z|\int_{3|x-z|}^{1+|x-z|} r^{-4+\alpha+2}\,\mathrm{d}r$$

$$\le \frac{C|x-z|}{1-\alpha}\big[(3|x-z|)^{\alpha-1}-(1+|x-z|)^{\alpha-1}\big]\frac{\widetilde{C}}{1-\alpha}|x-z|^\alpha.$$

For the remaining terms, we split up as in [Lieb and Loss 2001]:

$$B\setminus D\subset E\cup G,$$
$$D\setminus B\subset E'\cup G',$$

where

$$E:=\{b:4|x-z|<|b|\le 5|x-z|\},$$
$$G:=\{b:1-|x-z|\le|b|<1\},$$
$$E':=\{b:4|x-z|<|b-x+z|\le 5|x-z|\},$$
$$G':=\{b:1-|x-z|\le|b-x+z|<1\}.$$

Consider at first the real-valued map $[0,\frac{1}{4}]\ni s\mapsto(1-s)^\beta$ for arbitrary $\beta>0$. Obviously, it is continuously differentiable and therefore $\alpha$-Hölder continuous because its domain of definition is compact. Hence, the following holds for all $0\le s\le\frac{1}{4}$:

$$1-(1-s)^\beta=(1-0)^\beta-(1-s)^\beta\le Cs^\alpha. \tag{94}$$

Now, we estimate the integral on $B\setminus D$, where we use again the estimate $|\partial_{a_i}\partial_{a_j}\boldsymbol{G}_\kappa(a)|\le C|a|^{-3}$:

$$\left|\int_{B\setminus D}[f(z+b)-f(z)]\partial_{b_i}\partial_{b_j}\boldsymbol{G}_\kappa(b)\,\mathrm{d}b\right|\le C\left(\int_{4|x-z|}^{5|x-z|} r^{\alpha-3+2}\,\mathrm{d}r+\int_{1-|x-z|}^{1} r^{\alpha-3+2}\,\mathrm{d}r\right)$$
$$=\frac{C}{\alpha}[(5|x-z|)^\alpha-(4|x-z|)^\alpha+1-(1-|x-z|)^\alpha]\le\frac{C}{\alpha}(5^\alpha+\widetilde{C})|x-z|^\alpha,$$

where we have used (94) for $\beta:=\alpha$ in the last step.

For the remaining integral on $D\setminus B$, we consider the domains $E'$ and $G'$ separately and note at first that, using the triangular inequality, $E'\subset\{0<|b|\le 6|x-z|\}$. Subsequently, this yields that $|b-x+z|^{-3}<(4|x-z|)^{-3}\le C|b|^{-3}$ on $E'$. Hence, by the estimate $|\partial_{a_i}\partial_{a_j}\boldsymbol{G}_\kappa(a)|\le C|a|^{-3}$,

$$\int_{E'}\big|[f(z+b)-f(z)]\partial_{b_i}\partial_{b_j}\boldsymbol{G}_\kappa(b-x+z)\big|\,\mathrm{d}b\le C\int_0^{6|x-z|} r^{\alpha-3+2}\,\mathrm{d}r=\widetilde{C}|x-z|^\alpha.$$

On $G'$, one has $|b-x+z|\ge 1-|x-z|>\frac{3}{4}$. Consequently, it holds that

$$\int_{G'}\big|[f(z+b)-f(z)]\partial_{b_i}\partial_{b_j}\boldsymbol{G}_\kappa(b-x+z)\big|\,\mathrm{d}b\le C\big(\tfrac{3}{4}\big)^{-3}\int_{1-|x-z|}^{1} r^{\alpha+2}\,\mathrm{d}r$$
$$=\widetilde{C}[1-(1-|x-z|)^{3+\alpha}]\le\widetilde{C}|x-z|^\alpha,$$

where we have used (94) for $\beta := 3+\alpha$ in the last step. Together,

$$\left|\int_{D\setminus B}[f(z+b)-f(z)]\partial_{b_i}\partial_{b_j}\boldsymbol{G}_\kappa(b-x+z)\,\mathrm{d}b\right| \le C|x-z|^\alpha,$$

and the assertion is proved. □

## References


[Agueh 2008] M. Agueh, "Rates of decay to equilibria for $p$-Laplacian type equations", *Nonlinear Anal.* **68**:7 (2008), 1909–1927. MR 2009e:35117 Zbl 1185.35017

[Ambrosio et al. 2008] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, 2nd ed., Birkhäuser, Basel, 2008. MR 2009h:49002 Zbl 1145.35001

[Biler et al. 2011] P. Biler, L. Corrias, and J. Dolbeault, "Large mass self-similar solutions of the parabolic-parabolic Keller–Segel model of chemotaxis", *J. Math. Biol.* **63**:1 (2011), 1–32. MR 2012c:35431 Zbl 1230.92011

[Blanchet 2013] A. Blanchet, "A gradient flow approach to the Keller–Segel systems", preprint, 2013, Available at http://www.tse-fr.eu/sites/default/files/medias/doc/by/blanchet/dec_2010/b4.pdf.

[Blanchet and Laurençot 2013] A. Blanchet and P. Laurençot, "The parabolic-parabolic Keller–Segel system with critical diffusion as a gradient flow in $\mathbb{R}^d$, $d \ge 3$", *Comm. Partial Differential Equations* **38**:4 (2013), 658–686. MR 3040679 Zbl 1282.35202

[Blanchet et al. 2008] A. Blanchet, V. Calvez, and J. A. Carrillo, "Convergence of the mass-transport steepest descent scheme for the subcritical Patlak–Keller–Segel model", *SIAM J. Numer. Anal.* **46**:2 (2008), 691–721. MR 2009a:35113 Zbl 1205.65332

[Blanchet et al. 2009] A. Blanchet, J. A. Carrillo, and P. Laurençot, "Critical mass for a Patlak–Keller–Segel model with degenerate diffusion in higher dimensions", *Calc. Var. Partial Differential Equations* **35**:2 (2009), 133–168. MR 2010e:35149 Zbl 1172.35035

[Blanchet et al. 2012] A. Blanchet, E. A. Carlen, and J. A. Carrillo, "Functional inequalities, thick tails and asymptotics for the critical mass Patlak–Keller–Segel model", *J. Funct. Anal.* **262**:5 (2012), 2142–2230. MR 2876403 Zbl 1237.35155

[Blanchet et al. 2014] A. Blanchet, J. A. Carrillo, D. Kinderlehrer, M. Kowalczyk, P. Laurençot, and S. Lisini, "A hybrid variational principle for the Keller–Segel system in $\mathbb{R}^2$", preprint, 2014. arXiv 1407.5562v1

[Calvez and Carrillo 2012] V. Calvez and J. A. Carrillo, "Refined asymptotics for the subcritical Keller–Segel system and related functional inequalities", *Proc. Amer. Math. Soc.* **140**:10 (2012), 3515–3530. MR 2929020 Zbl 1277.35051

[Calvez and Corrias 2008] V. Calvez and L. Corrias, "The parabolic-parabolic Keller–Segel model in $\mathbb{R}^2$", *Commun. Math. Sci.* **6**:2 (2008), 417–447. MR 2009m:35518 Zbl 1149.35360

[Carrapatoso and Mischler 2014] K. Carrapatoso and S. Mischler, "Uniqueness and long time asymptotic for the parabolic-parabolic Keller–Segel equation", preprint, 2014. arXiv 1406.6006v2

[Carrillo and Toscani 2000] J. A. Carrillo and G. Toscani, "Asymptotic $L^1$-decay of solutions of the porous medium equation to self-similarity", *Indiana Univ. Math. J.* **49**:1 (2000), 113–142. MR 2001j:35155 Zbl 0963.35098

[Carrillo et al. 2003] J. A. Carrillo, R. J. McCann, and C. Villani, "Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates", *Rev. Mat. Iberoamericana* **19**:3 (2003), 971–1018. MR 2005a:35126 Zbl 1073.35127

[Carrillo et al. 2006a] J. A. Carrillo, M. Di Francesco, and G. Toscani, "Intermediate asymptotics beyond homogeneity and self-similarity: long time behavior for $u_t = \Delta\phi(u)$", *Arch. Ration. Mech. Anal.* **180**:1 (2006), 127–149. MR 2006k:35131 Zbl 1096.35015

[Carrillo et al. 2006b] J. A. Carrillo, R. J. McCann, and C. Villani, "Contractions in the 2-Wasserstein length space and thermalization of granular media", *Arch. Ration. Mech. Anal.* **179**:2 (2006), 217–263. MR 2006j:76121 Zbl 1082.76105

[Carrillo et al. 2011] J. A. Carrillo, M. Di Francesco, A. Figalli, T. Laurent, and D. Slepčev, "Global-in-time weak measure solutions and finite-time aggregation for nonlocal interaction equations", *Duke Math. J.* **156**:2 (2011), 229–271. MR 2012c:35447 Zbl 1215.35045

[Corrias and Perthame 2008] L. Corrias and B. Perthame, "Asymptotic decay for the solutions of the parabolic-parabolic Keller–Segel chemotaxis system in critical spaces", *Math. Comput. Modelling* **47**:7–8 (2008), 755–764. MR 2009c:35200 Zbl 1134.92006

[Di Francesco and Rosado 2008] M. Di Francesco and J. Rosado, "Fully parabolic Keller–Segel model for chemotaxis with prevention of overcrowding", *Nonlinearity* **21**:11 (2008), 2715–2730. MR 2009m:35196 Zbl 1157.35398

[Eisenbach 2004] M. Eisenbach, *Chemotaxis*, Imperial College, London, 2004.

[Giacomelli and Otto 2001] L. Giacomelli and F. Otto, "Variational formulation for the lubrication approximation of the Hele–Shaw flow", *Calc. Var. Partial Differential Equations* **13**:3 (2001), 377–403. MR 2003a:76046 Zbl 1086.35004

[Gianazza et al. 2009] U. Gianazza, G. Savaré, and G. Toscani, "The Wasserstein gradient flow of the Fisher information and the quantum drift-diffusion equation", *Arch. Ration. Mech. Anal.* **194**:1 (2009), 133–220. MR 2010i:35371 Zbl 1223.35264

[Hillen and Painter 2009] T. Hillen and K. J. Painter, "A user's guide to PDE models for chemotaxis", *J. Math. Biol.* **58**:1–2 (2009), 183–217. MR 2009m:92017 Zbl 1161.92003

[Horstmann 2003] D. Horstmann, "From 1970 until present: the Keller–Segel model in chemotaxis and its consequences, I", *Jahresber. Deutsch. Math.-Verein.* **105**:3 (2003), 103–165. MR 2005f:35163 Zbl 1071.35001

[Jordan et al. 1998] R. Jordan, D. Kinderlehrer, and F. Otto, "The variational formulation of the Fokker–Planck equation", *SIAM J. Math. Anal.* **29**:1 (1998), 1–17. MR 2000b:35258 Zbl 0915.35120

[Keller and Segel 1970] E. F. Keller and L. A. Segel, "Initiation of slime mold aggregation viewed as an instability", *J. Theor. Biol.* **26**:3 (1970), 399–415. Zbl 1170.92306

[Kozono and Sugiyama 2009] H. Kozono and Y. Sugiyama, "Global strong solution to the semi-linear Keller–Segel system of parabolic-parabolic type with small data in scale invariant spaces", *J. Differential Equations* **247**:1 (2009), 1–32. MR 2010i:35147 Zbl 1207.35178

[Lapidus and Schiller 1976] I. R. Lapidus and R. Schiller, "Model for the chemotactic response of a bacterial population", *Biophys. J.* **16**:7 (1976), 779–789.

[Laurençot and Matioc 2013] P. Laurençot and B.-V. Matioc, "A gradient flow approach to a thin film approximation of the Muskat problem", *Calc. Var. Partial Differential Equations* **47**:1–2 (2013), 319–341. MR 3044141 Zbl 1264.35129

[Lieb and Loss 2001] E. H. Lieb and M. Loss, *Analysis*, 2nd ed., Graduate Studies in Mathematics **14**, Amer. Math. Soc., Providence, RI, 2001. MR 2001i:00001 Zbl 0966.26002

[López-Gómez et al. 2013] J. López-Gómez, T. Nagai, and T. Yamada, "The basin of attraction of the steady-states for a chemotaxis model in $\mathbb{R}^2$ with critical mass", *Arch. Ration. Mech. Anal.* **207**:1 (2013), 159–184. MR 3004770 Zbl 06148937

[Luckhaus and Sugiyama 2006] S. Luckhaus and Y. Sugiyama, "Large time behavior of solutions in super-critical cases to degenerate Keller–Segel systems", *M2AN Math. Model. Numer. Anal.* **40**:3 (2006), 597–621. MR 2007e:35127 Zbl 1113.35028

[Luckhaus and Sugiyama 2007] S. Luckhaus and Y. Sugiyama, "Asymptotic profile with the optimal convergence rate for a parabolic equation of chemotaxis in super-critical cases", *Indiana Univ. Math. J.* **56**:3 (2007), 1279–1297. MR 2008k:35272 Zbl 1118.35006

[Matthes et al. 2009] D. Matthes, R. J. McCann, and G. Savaré, "A family of nonlinear fourth order equations of gradient flow type", *Comm. Partial Differential Equations* **34**:10–12 (2009), 1352–1397. MR 2011e:35149 Zbl 1187.35131

[Mizoguchi 2013] N. Mizoguchi, "Global existence for the Cauchy problem of the parabolic-parabolic Keller–Segel system on the plane", *Calc. Var. Partial Differential Equations* **48**:3–4 (2013), 491–505. MR 3116019 Zbl 1288.35286

[Murray 2003] J. D. Murray, *Mathematical biology, II: Spatial models and biomedical applications*, 3rd ed., Interdisciplinary Applied Mathematics **18**, Springer, New York, 2003. MR 2004b:92001 Zbl 1006.92002

[Nagai and Senba 1998] T. Nagai and T. Senba, "Global existence and blow-up of radial solutions to a parabolic-elliptic system of chemotaxis", *Adv. Math. Sci. Appl.* **8**:1 (1998), 145–156. MR 2000a:35241 Zbl 0902.35010

[Nagai et al. 2003] T. Nagai, R. Syukuinn, and M. Umesako, "Decay properties and asymptotic profiles of bounded solutions to a parabolic system of chemotaxis in $\mathbb{R}^n$", *Funkcial. Ekvac.* **46**:3 (2003), 383–407. MR 2005h:35144 Zbl 02112808

[Otto 2001] F. Otto, "The geometry of dissipative evolution equations: the porous medium equation", *Comm. Partial Differential Equations* **26**:1–2 (2001), 101–174. MR 2002j:35180 Zbl 0984.35089

[Perthame 2007] B. Perthame, *Transport equations in biology*, Birkhäuser, Basel, 2007. MR 2007j:35004 Zbl 1185.92006

[Post 1999] K. Post, *A non-linear parabolic system modeling chemotaxis with sensitivity functions*, Ph.D. thesis, Humboldt-Universität zu Berlin, 1999, Available at http://edoc.hu-berlin.de/docviews/abstract.php?id=26314.

[Segel 1977] L. Segel, "A theoretical study of receptor mechanisms in bacterial chemotaxis", *SIAM J. Appl. Math.* **32**:3 (1977), 653–665. Zbl 0356.92009

[Senba and Suzuki 2006] T. Senba and T. Suzuki, "A quasi-linear parabolic system of chemotaxis", *Abstr. Appl. Anal.* **2006** (2006), 23061. MR 2006j:35121 Zbl 1134.35059

[Stein 1970] E. M. Stein, *Singular integrals and differentiability properties of functions*, Princeton Mathematical Series **30**, Princeton University, 1970. MR 44 #7280 Zbl 0207.13501

[Sugiyama 2006] Y. Sugiyama, "Global existence in sub-critical cases and finite time blow-up in super-critical cases to degenerate Keller–Segel systems", *Differential Integral Equations* **19**:8 (2006), 841–876. MR 2008h:35134 Zbl 1212.35240

[Sugiyama 2007] Y. Sugiyama, "Time global existence and asymptotic behavior of solutions to degenerate quasi-linear parabolic systems of chemotaxis", *Differential Integral Equations* **20**:2 (2007), 133–180. MR 2008a:35169 Zbl 1212.35241

[Sugiyama and Kunii 2006] Y. Sugiyama and H. Kunii, "Global existence and decay properties for a degenerate Keller–Segel model with a power factor in drift term", *J. Differential Equations* **227**:1 (2006), 333–364. MR 2008j:35103 Zbl 1102.35046

[Villani 2003] C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics **58**, Amer. Math. Soc., Providence, RI, 2003. MR 2004e:90003 Zbl 1106.90001

[Winkler 2010] M. Winkler, "Absence of collapse in a parabolic chemotaxis system with signal-dependent sensitivity", *Math. Nachr.* **283**:11 (2010), 1664–1673. MR 2011k:35090 Zbl 1205.35037

[Yamada 2011] T. Yamada, "Improvement of convergence rates for a parabolic system of chemotaxis in the whole space", *Math. Methods Appl. Sci.* **34**:17 (2011), 2103–2124. MR 2012k:35217 Zbl 1242.35054

[Zeidler 1990] E. Zeidler, *Nonlinear functional analysis and its applications, II/B: Nonlinear monotone operators*, Springer, New York, 1990. MR 91b:47002 Zbl 0684.47029

[Zinsl 2014] J. Zinsl, "Existence of solutions for a nonlinear system of parabolic equations with gradient flow structure", *Monats. Math.* **174**:4 (2014), 653–679. MR 3233116 Zbl 1302.35202

Received 12 May 2014. Revised 25 Nov 2014. Accepted 9 Jan 2015.

JONATHAN ZINSL: zinsl@ma.tum.de
*Zentrum für Mathematik, Technische Universität München, Boltzmannstraße 3, 85747 Garching, Germany*

DANIEL MATTHES: matthes@ma.tum.de
*Zentrum für Mathematik, Technische Universität München, Boltzmannstraße 3, 85747 Garching, Germany*

ANALYSIS AND PDE
Vol. 8, No. 2, 2015
dx.doi.org/10.2140/apde.2015.8.467

# SCATTERING FOR THE RADIAL 3D CUBIC WAVE EQUATION

BENJAMIN DODSON AND ANDREW LAWRIE

Consider the Cauchy problem for the radial cubic wave equation in $1+3$ dimensions with either the focusing or defocusing sign. This problem is critical in $\dot{H}^{1/2} \times \dot{H}^{-1/2}(\mathbb{R}^3)$ and subcritical with respect to the conserved energy. Here we prove that if the critical norm of a solution remains bounded on the maximal time interval of existence, then the solution must in fact be global in time and must scatter to free waves as $t \to \pm\infty$.

## 1. Introduction

Consider the Cauchy problem for the cubic semilinear wave equation in $\mathbb{R}^{1+3}$, namely,

$$\begin{aligned} u_{tt} - \Delta u + \mu u^3 = 0, \\ \vec{u}(0) = (u_0, u_1), \end{aligned} \tag{1-1}$$

restricted to the radial setting and with $\mu \in \{\pm 1\}$. The case $\mu = 1$ yields what is referred to as the defocusing problem, since here the conserved energy,

$$E(\vec{u})(t) := \int_{\mathbb{R}^3} \left( \frac{1}{2}\big(|u_t(t)|^2 + |\nabla u(t)|^2\big) + \frac{1}{4}|u(t)|^4 \right) dx = \text{constant}, \tag{1-2}$$

is positive for sufficiently regular nonzero solutions, and the $\dot{H}^1 \times L^2(\mathbb{R}^3)$ norm of a solution,

$$\vec{u}(t) := (u(t), u_t(t)),$$

is bounded by its energy.

The case $\mu = -1$ gives the focusing problem, and the conserved energy for sufficiently regular solutions to (1-1) is given by

$$E(\vec{u})(t) := \int_{\mathbb{R}^3} \left( \frac{1}{2}\big(|u_t(t)|^2 + |\nabla u(t)|^2\big) - \frac{1}{4}|u(t)|^4 \right) dx = \text{constant}. \tag{1-3}$$

As we will only be considering radial solutions to (1-1), we will often slightly abuse notation by writing $u(t, x) = u(t, r)$, where $(r, \omega)$, with $r = |x|$, $x = r\omega$, $\omega \in \mathbb{S}^2$, are polar coordinates on $\mathbb{R}^3$. In this setting

Support of the National Science Foundation, DMS-1103914 for Dodson and DMS-1302782 for Lawrie, is gratefully acknowledged.
*MSC2010:* 35L05.
*Keywords:* scattering, concentration compactness, double Duhamel.

we can rewrite the Cauchy problem (1-1) as

$$
\begin{aligned}
u_{tt} - u_{rr} - \frac{2}{r}u_r \pm u^3 = 0, \\
\vec{u}(0) = (u_0, u_1),
\end{aligned}
\tag{1-4}
$$

and the conserved energy (up to a constant multiple) as

$$
E(\vec{u})(t) = \int_0^\infty \left( \frac{1}{2}\big(u_t^2(t) + u_r^2(t)\big) \pm \frac{1}{4}u^4(t) \right) r^2 \, dr. \tag{1-5}
$$

The Cauchy problem (1-4) is invariant under the scaling

$$
\vec{u}(t, r) \mapsto \vec{u}_\lambda(t, r) := \left( \lambda^{-1} u\left(\frac{t}{\lambda}, \frac{r}{\lambda}\right), \lambda^{-2} u_t\left(\frac{t}{\lambda}, \frac{r}{\lambda}\right) \right). \tag{1-6}
$$

One can also check that this scaling leaves unchanged the $\dot{H}^{1/2} \times \dot{H}^{-1/2}$-norm of the solution. It is for this reason that (1-4) is called *energy-subcritical*. It is natural to consider the Cauchy problem with initial data $(u_0, u_1) \in \dot{H}^{1/2} \times \dot{H}^{-1/2}$. We remark that (1-4) is also invariant under conformal inversion:

$$
u(t, r) \mapsto \frac{1}{t^2 - r^2} u\left( \frac{t}{t^2 - r^2}, \frac{r}{t^2 - r^2} \right). \tag{1-7}
$$

A standard argument based on Strichartz estimates shows that both the defocusing and focusing problems are locally well-posed in $\dot{H}^{1/2} \times \dot{H}^{-1/2}(\mathbb{R}^3)$. This means that for all initial data $\vec{u}(0) = (u_0, u_1) \in \dot{H}^{1/2} \times \dot{H}^{-1/2}$, there is a unique solution $\vec{u}(t)$ defined on a maximal interval of existence $I_{\max}$ with $\vec{u}(t) \in C(I_{\max}; \dot{H}^{1/2} \times \dot{H}^{-1/2})$. Moreover, for every compact time interval $J \subset I_{\max}$, we have $u \in S(J) := L_t^4(J; L_x^4)$. The Strichartz norm $S(J)$ determines criteria for both scattering and finite-time blow-up, and we make these statements precise in Proposition 2.4. Here we note that in particular, one can show that if the initial data $\vec{u}(0)$ has sufficiently small $\dot{H}^{1/2} \times \dot{H}^{-1/2}$-norm, then the corresponding solution $\vec{u}(t)$ has finite $S(\mathbb{R})$-norm and hence scatters to free waves as $t \to \pm\infty$.

The theory for solutions to (1-4) with initial data that is small in $\dot{H}^{1/2} \times \dot{H}^{-1/2}$ is thus very well understood: all solutions are global in time and scatter to free waves as $t \to \pm\infty$. However, much less is known regarding the *asymptotic dynamics* of solutions to either the defocusing or focusing problems once one leaves the perturbative regime.

It is well known that there are solutions to the focusing problem that blow up in finite time. To give an example,

$$
\phi_T(t, r) = \frac{\sqrt{2}}{T - t} \tag{1-8}
$$

solves the ODE $\phi_{tt} = \phi^3$. Using finite speed of propagation, one can construct from $\phi_T$ a compactly supported (in space) self-similar blow-up solution to (1-4). Indeed, define $\vec{u}_T(t)$ to be the solution to (1-4) with initial data $\vec{u}_T(0, x) = \chi_{2T}(x)\sqrt{2}$, where $\chi_{2T} \in C_0^\infty(\mathbb{R}^3)$ satisfies $\chi_{2T}(x) = 1$ if $|x| \le 2T$. Then $\vec{u}_T(t)$ equals $\phi_T(t)$ for all $r \le T$ and $0 \le t < T$, and blows up at time $t = T$. However, such a self-similar solution must have its critical $\dot{H}^{1/2} \times \dot{H}^{-1/2}$-norm tending to $\infty$ as $t \to T$:

$$\lim_{t\to T} \|\vec{u}_T(t)\|_{\dot{H}^{1/2}\times\dot{H}^{-1/2}} = \infty.$$

Indeed, one can show by a direct computation that the $L^3(\mathbb{R}^3)$-norm of $\vec{u}_T(t, x)$ tends to $\infty$ as $t \to T_+$. Since $\dot{H}^{1/2} \subset L^3$, this means that the $\dot{H}^{1/2} \times \dot{H}^{-1/2}$-norm must blow up as well. Such behavior is typically referred to as type I, or ODE, blow-up.

One the other hand, type II solutions, $\vec{u}(t)$, are those whose critical norm remains bounded on their maximal interval of existence, $I_{\max}$:

$$\sup_{t\in I_{\max}} \|\vec{u}(t)\|_{\dot{H}^{1/2}\times\dot{H}^{-1/2}} < \infty. \tag{1-9}$$

In this paper we restrict our attention to type II solutions, i.e., those which satisfy (1-9). We prove that if a solution $\vec{u}(t)$ to (1-4) satisfies (1-9), then $\vec{u}(t)$ must in fact exist globally in time and scatter to free waves in both time directions. To be precise, we establish the following result.

**Theorem 1.1.** *Let $\vec{u}(t) \in \dot{H}^{1/2} \times \dot{H}^{-1/2}(\mathbb{R}^3)$ be a radial solution to* (1-4) *defined on its maximal interval of existence $I_{\max} = (T_-, T_+)$. Suppose in addition that*

$$\sup_{t\in I_{\max}} \|\vec{u}(t)\|_{\dot{H}^{1/2}\times\dot{H}^{-1/2}(\mathbb{R}^3)} < \infty. \tag{1-10}$$

*Then $I_{\max} = \mathbb{R}$, i.e., $\vec{u}(t)$ is defined globally in time. Moreover*,

$$\|u\|_{L^4_{t,x}(\mathbb{R}^{1+3})} < \infty, \tag{1-11}$$

*which means that $\vec{u}(t)$ scatters to a free wave in each time direction*, *i.e.*, *there exist radial solutions $\vec{u}_L^{\pm}(t) \in \dot{H}^{1/2} \times \dot{H}^{-1/2}(\mathbb{R}^3)$ to the free wave equation*, $\Box u_L^{\pm} = 0$, *such that*

$$\|\vec{u}(t) - \vec{u}_L^{\pm}(t)\|_{\dot{H}^{1/2}\times\dot{H}^{-1/2}(\mathbb{R}^3)} \to 0 \quad \text{as} \quad t \to +\infty. \tag{1-12}$$

**Remark 1.** Theorem 1.1 is a conditional result. Other than the requirement that the initial data be small in $\dot{H}^{1/2} \times \dot{H}^{-1/2}$, there is no known general criterion that ensures that (1-10) is satisfied by the evolution for either the defocusing or the focusing equation. While the methods in this paper apply equally well to both the focusing and defocusing equations, one should expect drastically different behavior from generic initial data in these two cases.

**Remark 2.** The proof of Theorem 1.1 readily generalizes to all subcritical powers $p \le 3$ for which there is a satisfactory small data/local well-posedness theory. In particular, the methods presented here allow one to deduce the exact analog of Theorem 1.1 for radial equations (1-13) for all powers $p$ with $1 + \sqrt{2} < p \le 3$; here $1 + \sqrt{2}$ is the F. John exponent [1979; Schaeffer 1985]. We have chosen to present the details for only the cubic equation to keep the exposition as simple as possible. We also remark that only the material in Section 4 relies on the assumption of radiality.

**1A.** ***History of the problem.*** The cubic wave equation on $\mathbb{R}^{1+3}$ has been extensively studied and we certainly cannot give a complete account of the vast body of literature devoted to this problem.

For the defocusing equation, the positivity of the conserved energy can be used to extend a local existence result to a global one if one begins with initial data that is sufficiently regular. Jörgens [1961] showed global existence for the defocusing equation for smooth compactly supported data. There has been a good deal of recent work extending the local existence result of Lindblad and Sogge [1995] in $H^s \times H^{s-1}$ for $s > \frac{1}{2}$ to an unconditional global well-posedness result, and we refer the reader to [Kenig et al. 2000; Gallagher and Planchon 2003; Bahouri and Chemin 2006; Roy 2009] and the references therein for details. However, since these works are not carried out in the scaling critical space, the issue of global dynamics, and in particular scattering, is not addressed.

For the focusing equation, type II finite-time blow-up has recently been ruled out for initial data that lies in $\dot{H}^1 \times L^2$ in the work of Killip, Stovall, and Vişan [Killip et al. 2012]. There are several works that open up interesting lines of inquiry related to the question of asymptotic dynamics. In two remarkable works, Merle and Zaag [2003; 2005] determined that all blow-up solutions must blow up *at the self-similar rate*. In the radial case, an infinite family of smooth self-similar solutions is constructed by Bizoń et al. [2010]. Bizoń and Zenginoğlu [2009] give numerical evidence to support a conjecture that a two-parameter family of solutions, obtained via time translation and conformal inversion of a self-similar solution, serves as a global attractor for a large set of initial data. In fact, Donninger and Schörkhuber [2012] showed that the blow-up profile (1-8) is stable under small perturbations in the energy topology.

Equations of the form

$$\Box u = \pm |u|^{p-1} u \tag{1-13}$$

for different values of $p$ and for different dimensions have also been extensively studied. For $d = 3$, the energy-critical power, $p = 5$, exhibits quite different phenomena than both the subcritical and supercritical equations. Global existence and scattering for all finite-energy data was proved by Struwe [1988] for the radial defocusing equation and by Grillakis [1990] in the nonradial, defocusing case.

For the focusing energy-critical equation, type II blow-up can occur, as explicitly demonstrated by Krieger, Schlag, and Tataru [Krieger et al. 2009] via an energy concentration scenario resulting in the bubbling off of the ground state solution, $W$, for the underlying elliptic equation; see also [Krieger and Schlag 2014a; Donninger et al. 2014; Donninger and Krieger 2013].

Kenig and Merle [2008] initiated a powerful program of attack for semilinear equations (1-13) with the concentration compactness/rigidity method, giving a characterization of possible dynamics for solutions with energy below the threshold energy of the ground state elliptic solution. The subsequent work of Duyckaerts, Kenig, and Merle [Duyckaerts et al. 2011; 2012a; 2012b; 2013] resulted in a classification of possible dynamics for large energies. In particular, all type II radial solutions asymptotically resolve into a sum of rescaled solitons plus a radiation term at their maximal time of existence. Dynamics at the threshold energy of $W$ have been examined by Duyckaerts and Merle [2008] and above the threshold by Krieger, Nakanishi, and Schlag [Krieger et al. 2013a; 2013b; 2014].

Analogs of Theorem 1.1 have been established for radial equations with different powers in 3 dimensions.

Shen [2012] proved the exact analog of Theorem 1.1 for subcritical powers $3 < p < 5$; and Kenig and Merle [2011], and then Duyckaerts, Kenig, and Merle [Duyckaerts et al. 2014], established the analog of Theorem 1.1 for all supercritical powers $p > 5$. Here we address type II behavior in the remainder of the subcritical range for the radial equation, $1 + \sqrt{2} < p \le 3$. While we focus on the cubic equation, our proof readily generalizes to other subcritical powers. The extra regularity for critical elements proved in Section 4 gives an extension and simplification of the argument in [Shen 2012] which allows us to treat the cubic and lower-power equations.

Leaving the setting of type II solutions, Krieger and Schlag [2014b] have recently constructed a family of solutions to the supercritical equation, $p > 5$, which are smooth, global in time, and stable under small perturbations, and have infinite critical norm.

**1B.** ***Outline of the proof of Theorem 1.1.*** The proof of Theorem 1.1 follows the concentration compactness/rigidity method developed in [Kenig and Merle 2006; 2008]. The proof follows a contradiction argument: if Theorem 1.1 were not true, the linear and nonlinear profile decompositions of Bahouri and Gérard would allow one to construct a minimal solution to (1-4), called the critical element, which does not scatter (here the minimality refers to the size of the norm in (1-10)). This construction, which is by now standard in the field and is outlined in Section 3, yields a critical element whose trajectory in the space $\dot{H}^{1/2} \times \dot{H}^{-1/2}$ is precompact up to modulation. The goal is then to prove that this compactness property is too rigid a property for a nonzero solution and thus the critical element cannot exist.

A significant hurdle in the way of ruling out a critical element $\vec{u}_c(t)$ for the cubic equation (or any subcritical equation) is that $\vec{u}_c(t)$ is constructed in the space $\dot{H}^{1/2} \times \dot{H}^{-1/2}$, and thus useful global monotone quantities that require more regularity, such as the conserved energy and virial type identities, are not, a priori, well defined. In general, solutions to the cubic wave equation are only as regular as their initial data, as evidenced by the presence of the free propagator $S(t)$ in the Duhamel representation for the solution

$$\vec{u}_c(t_0) = S(t_0 - t)\vec{u}_c(t) + \int_t^{t_0} S(t_0 - s)(0, \pm u^3)\, ds. \tag{1-14}$$

The critical element is rescued by the fact that the precompactness of its trajectory is at odds with the dispersive properties of the free part, $S(t_0 - t)\vec{u}(t)$, and thus the first term on the right of (1-14) is forced to vanish weakly as $t \to \sup I_{\max}$ and as $t \to \inf I_{\max}$. The second term on the right thus encodes the regularity of the critical element, and a gain can be expected due to the presence of the cubic term. The additional regularity is extracted by way of the "double Duhamel trick", which refers to the consideration of the pairing

$$\left\langle \int_{T_1}^{t_0} S(t_0 - s)(0, \pm u^3)\, ds,\ \int_{t_0}^{T_2} S(t_0 - \tau)(0, \pm u^3)\, d\tau \right\rangle_{\dot{H}^1 \times L^2},$$

where $T_1 < t_0$ and $T_2 > t_0$. This technique was developed by Tao [2007] and utilized in the Kenig–Merle framework for nonlinear Schrödinger problems by Killip and Vişan [2010a; 2010b; 2013], and for semilinear wave equations in [Killip and Vişan 2011; Bulut 2012a; 2012b]. This method is also closely related to the in/out decomposition used by Killip, Tao, and Vişan in [Killip et al. 2009, Section 6]. For more details on how to exploit the different time directions above, we refer the reader to Section 4, and in particular to the proof of Theorem 4.1.

Indeed, we bound the critical element in $\dot{H}^1 \times L^2$. We then use the conserved energy to rule out a critical element which fails to be compact by a low frequency concentration, as such a solution would have vanishing energy; see Section 5A. One is then left with a critical element that is global in time and evolves at a fixed scale. In Section 6, we prove that such a solution cannot exist by way of a virial identity. We note that this virial-based rigidity argument works for precompact solutions to (1-13) with powers $p \le 3$, but fails to produce useful estimates for powers $3 < p < 5$. However, in this range one can use the "channels of energy" method pioneered in [Duyckaerts et al. 2013; 2014]; see [Shen 2012]. For more on this, see Remark 12.

## 2. Preliminaries

**2A. *Harmonic analysis.*** In what follows we will denote by $P_k$ the usual Littlewood–Paley projections onto frequencies of size $|\xi| \simeq 2^k$ and by $P_{\le k}$ the projection onto frequencies $|\xi| \lesssim 2^k$. These projections satisfy Bernstein's inequalities.

**Lemma 2.1** (Bernstein's inequalities [Tao 2006, Appendix A]). *Let $1 \le p \le q \le \infty$ and $s \ge 0$. Let $f : \mathbb{R}^d \to \mathbb{R}$. Then*

$$\begin{aligned} \|P_{\ge N} f\|_{L^p} &\lesssim N^{-s} \||\nabla|^s P_{\ge N} f\|_{L^p}, \\ \|P_{\le N} |\nabla|^s f\|_{L^p} &\lesssim N^s \|P_{\le N} f\|_{L^p}, \qquad \|P_N |\nabla|^{\pm s} f\|_{L^p} \simeq N^{\pm s} \|P_N f\|_{L^p}, \\ \|P_{\le N} f\|_{L^q} &\lesssim N^{d/p-d/q} \|P_{\le N} f\|_{L^p}, \qquad \|P_N f\|_{L^q} \lesssim N^{d/p-d/q} \|P_N f\|_{L^p}. \end{aligned} \tag{2-1}$$

Next, we define the notion of a frequency envelope.

**Definition 3** [Tao 2001, Definition 1]. We define a *frequency envelope* to be a sequence $\beta = \{\beta_k\}$ of positive real numbers with $\beta \in \ell^2$. Moreover, we require the local constancy condition

$$2^{-\sigma|j-k|}\beta_k \lesssim \beta_j \lesssim 2^{\sigma|j-k|}\beta_k,$$

where $\sigma > 0$ is a small fixed constant; in what follows we will use $\sigma = \frac{1}{8}$. If $\beta$ is a frequency envelope and $(f, g) \in \dot{H}^s \times \dot{H}^{s-1}$, then we say that $(f, g)$ *lies underneath* $\beta$ if

$$\|(P_k f, P_k g)\|_{\dot{H}^s \times \dot{H}^{s-1}} \le \beta_k \quad \text{for all } k \in \mathbb{Z},$$

and we note that if $(f, g)$ lies underneath $\beta$, then we have

$$\|(f, g)\|_{\dot{H}^s \times \dot{H}^{s-1}} \lesssim \|\beta\|_{\ell^2}.$$

We will require the following refinement of the Sobolev embedding for radial functions, which is a consequence of the Hardy–Littlewood–Sobolev inequality.

**Lemma 2.2** (radial Sobolev embedding [Tao et al. 2007, Corollary A.3]). *Let* $0 < s < 3$ *and suppose* $f \in \dot{H}^s(\mathbb{R}^3)$ *is a radial function. Suppose that*

$$\beta > -\frac{3}{q}, \quad \frac{1}{2} - s \le \frac{1}{q} \le \frac{1}{2}, \quad \frac{1}{q} = \frac{1}{2} - \frac{\beta + s}{3},$$

*and at most one of the equalities* $q = 1$, $q = \infty$, $\frac{1}{q} + s = \frac{1}{2}$ *holds. Then*

$$\|r^\beta f\|_{L^q} \le C \|f\|_{\dot{H}^s}. \tag{2-2}$$

**2B.** ***Strichartz estimates.*** An essential ingredient for the small data theory are Strichartz estimates for the linear wave equation in $\mathbb{R}^{1+3}$,

$$\begin{aligned} v_{tt} - \Delta v &= F, \\ \vec{v}(0) &= (v_0, v_1). \end{aligned} \tag{2-3}$$

A free wave will mean a solution to (2-3) with $F = 0$ and will be denoted by $\vec{u}(t) = S(t)\vec{u}(0)$. In what follows we will say that a pair $(p, q)$ is admissible if

$$p, q \ge 2, \quad \frac{1}{p} + \frac{1}{q} \le \frac{1}{2}. \tag{2-4}$$

The Strichartz estimates we state below are standard and we refer the reader to [Keel and Tao 1998; Lindblad and Sogge 1995] or the book [Sogge 2008] and the references therein for more details.

**Remark 4.** We note that since we will only consider the waves with radial initial data and with $F$ radial, we can allow the endpoint $(p, q) = (2, \infty)$ as an admissible pair. The admissibility of $(2, \infty)$ in the radial setting was established in [Klainerman and Machedon 1993]. This endpoint is of course forbidden for nonradial data in dimension $d = 3$.

**Proposition 2.3** [Keel and Tao 1998; Klainerman and Machedon 1993; Lindblad and Sogge 1995; Sogge 2008]. *Let* $\vec{v}(t)$ *be a solution to* (2-3) *with initial data* $\vec{v}(0) \in \dot{H}^s \times \dot{H}^{s-1}(\mathbb{R}^3)$ *for* $s > 0$. *Let* $(p, q)$ *and* $(a, b)$ *be admissible pairs satisfying the gap condition*

$$\frac{1}{p} + \frac{3}{q} = \frac{1}{a'} + \frac{3}{b'} - 2 = \frac{3}{2} - s, \tag{2-5}$$

*where* $(a', b')$ *are the conjugate exponents of* $(a, b)$. *Then, for any time interval* $I \ni 0$, *we have the estimates*

$$\|v\|_{L^p_t(I; L^q_x)} \lesssim \|(v_0, v_1)\|_{\dot{H}^s \times \dot{H}^{s-1}} + \|F\|_{L^{a'}_t(I; L^{b'}_x)}. \tag{2-6}$$

**2C.** ***Small data theory — global existence, scattering, perturbative theory.*** A standard argument based on Proposition 2.3 with $s = \frac{1}{2}$, $(p, q) = (4, 4)$, and $(a', b') = \left(\frac{4}{3}, \frac{4}{3}\right)$ yields the following small data result.

**Proposition 2.4** (small data theory). *Let* $\vec{u}(0) = (u_0, u_1) \in \dot{H}^{1/2} \times \dot{H}^{-1/2}(\mathbb{R}^3)$. *Then there is a unique solution* $\vec{u}(t) \in \dot{H}^{1/2} \times \dot{H}^{-1/2}(\mathbb{R}^3)$ *defined on a maximal interval of existence* $I_{\max}(\vec{u}) = (T_-(\vec{u}), T_+(\vec{u}))$. *Moreover, for any compact interval* $J \subset I_{\max}$, *we have*

$$\|u\|_{L^4_t(J; L^4_x)} < \infty.$$

*A globally defined solution* $\vec{u}(t)$ *for* $t \in [0, \infty)$ *scatters as* $t \to \infty$ *to a free wave, i.e., a solution* $\vec{u}_L(t)$ *of* $\Box u_L = 0$, *if and only if* $\|u\|_{L_t^4([0,\infty),L_x^4)} < \infty$. *In particular, there exists a constant* $\delta > 0$ *such that*

$$\|\vec{u}(0)\|_{\dot{H}^{1/2}\times\dot{H}^{-1/2}} < \delta \implies \|u\|_{L_t^4(\mathbb{R};L_x^4)} \lesssim \|\vec{u}(0)\|_{\dot{H}^{1/2}\times\dot{H}^{-1/2}} \lesssim \delta, \tag{2-7}$$

*and hence* $\vec{u}(t)$ *scatters to free waves as* $t \to \pm\infty$. *Finally, we have the standard finite-time blow-up criterion*:

$$T_+(\vec{u}) < \infty \implies \|u\|_{L_t^4([0,T_+(\vec{u}));L_x^4)} = +\infty. \tag{2-8}$$

*A similar statement holds if* $-\infty < T_-(\vec{u})$.

For the concentration compactness procedure in Section 3 one requires the following perturbation theory for approximate solutions to (1-4).

**Lemma 2.5** (perturbation lemma). *There are continuous functions* $\varepsilon_0, C_0 : (0,\infty) \to (0,\infty)$ *such that the following holds*: *Let* $I \subset \mathbb{R}$ *be an open interval* (*possibly unbounded*); *let* $\vec{u}, \vec{v} \in C(I; \mathcal{H})$ *satisfy, for some* $A > 0$,

$$\|\vec{v}\|_{L^\infty(I;\dot{H}^{1/2}\times\dot{H}^{-1/2})} + \|v\|_{L_t^4(I;L_x^4)} \le A,$$
$$\|\mathrm{eq}(u)\|_{L_t^{4/3}(I;L_x^{4/3})} + \|\mathrm{eq}(v)\|_{L_t^{4/3}(I;L_x^{4/3})} + \|w_0\|_{L_t^4(I;L_x^4)} \le \varepsilon \le \varepsilon_0(A),$$

*where* $\mathrm{eq}(u) := \Box u \pm u^3$ *in the sense of distributions, and where* $\vec{w}_0(t) := S(t-t_0)(\vec{u}-\vec{v})(t_0)$, *with* $t_0 \in I$ *arbitrary but fixed. Then*

$$\|\vec{u} - \vec{v} - \vec{w}_0\|_{L^\infty(I;\dot{H}^{1/2}\times\dot{H}^{-1/2})} + \|u - v\|_{L_t^4(I;L_x^4)} \le C_0(A)\varepsilon.$$

*In particular*, $\|u\|_{L_t^4(I;L_x^4)} < \infty$.

**2D.** ***Blow-up for nonpositive energies.*** Finally, we recall that in the case of the focusing equation, any nontrivial solution with negative energy must blow up in both time directions. This result was proved in [Killip et al. 2012] for solutions to (1-4).

**Proposition 2.6** [Killip et al. 2012, Theorem 3.1]. *Let* $\vec{u}(t)$ *be a solution to* (1-4) *with the focusing sign and with maximal interval of existence* $I_{\max} = (T_-, T_+)$. *If* $E(\vec{u}) \le 0$, *then* $\vec{u}(t)$ *is either identically zero or blows up in finite time in both time directions, i.e.*, $T_+ < +\infty$ *and* $T_- > -\infty$.

## 3. Concentration compactness

In this section we begin the proof of Theorem 1.1. We will follow the concentration compactness/rigidity method introduced by Kenig and Merle [2006; 2008]. The concentration compactness part of the argument, which is based on the profile decompositions of Bahouri and Gérard [1999], is by now standard, and we will essentially follow the scheme from [Kenig and Merle 2010], which is a refinement of the methods from [Kenig and Merle 2006; 2008]. Indeed, the main conclusion of this section is that in the event that Theorem 1.1 fails, there exists a minimal, nontrivial, nonscattering solution to (1-4), which we will call the critical element.

We begin with some notation, following [Kenig and Merle 2010] for convenience. For initial data $(u_0, u_1) \in \dot{H}^{1/2} \times \dot{H}^{-1/2}$, we let $\vec{u}(t) \in \dot{H}^{1/2} \times \dot{H}^{-1/2}$ be the unique solution to (1-4) with initial data $\vec{u}(0) = (u_0, u_1)$ defined on its maximal interval of existence $I_{\max}(\vec{u}) := (T_-(\vec{u}), T_+(\vec{u}))$. For $A > 0$, define

$$\mathscr{B}(A) := \left\{(u_0, u_1) \in \dot{H}^{1/2} \times \dot{H}^{-1/2} : \|\vec{u}(t)\|_{L_t^\infty(I_{\max}(\vec{u}); \dot{H}^{1/2} \times \dot{H}^{-1/2})} \le A\right\}. \tag{3-1}$$

**Definition 5.** We say that $\mathcal{SC}(A)$ holds if for all $\vec{u} = (u_0, u_1) \in \mathscr{B}(A)$, we have $I_{\max}(\vec{u}) = \mathbb{R}$ and $\|u\|_{L_t^4 L_x^4} < \infty$. We also say that $\mathcal{SC}(A; \vec{u})$ holds if $\vec{u} \in \mathscr{B}(A)$, $I_{\max}(\vec{u}) = \mathbb{R}$, and $\|u\|_{L_t^4 L_x^4} < \infty$.

**Remark 6.** Recall from Proposition 2.4 that $\|u\|_{L_t^4; L_x^4} < \infty$ if and only if $\vec{u}$ scatters to a free wave as $t \to \pm\infty$. Therefore Theorem 1.1 is equivalent to the statement that $\mathcal{SC}(A)$ holds for all $A > 0$.

Now suppose that Theorem 1.1 is *false*. By Proposition 2.4, there is an $A_0 > 0$ small enough that $\mathcal{SC}(A_0)$ holds. Given that we are assuming that Theorem 1.1 fails, we can find a threshold, or critical value $A_C$ such that for $A < A_C$, $\mathcal{SC}(A)$ holds, and for $A > A_C$, $\mathcal{SC}(A)$ fails. Note that $0 < A_0 < A_C$. The standard conclusion of this assumed failure of Theorem 1.1 is that there is a minimal nonscattering solution $\vec{u}(t)$ to (1-4) such that $\mathcal{SC}(A_C, \vec{u})$ fails, which enjoys certain compactness properties.

We will state a refined version of this result below, and we refer the reader to [Kenig and Merle 2010; Shen 2012; Tao et al. 2007; 2008] for the details of the argument. As usual, the main ingredients are the linear and nonlinear Bahouri–Gérard type profile decompositions [1999] used in conjunction with the perturbation theory in Lemma 2.5.

**Proposition 3.1.** *Suppose that Theorem 1.1 is false. Then there exists a solution $\vec{u}(t)$ such that $\mathcal{SC}(A_C; \vec{u})$ fails. Moreover, we can assume that $\vec{u}(t)$ does not scatter in either time direction*:

$$\|u\|_{L^4((T_-(\vec{u}), 0]; L_x^4)} = \|u\|_{L^4([0, T_+(\vec{u})); L_x^4)} = \infty. \tag{3-2}$$

*In addition, there exists a continuous function $N : I_{\max}(\vec{u}) \to (0, \infty)$ such that the set*

$$K := \left\{ \left( \frac{1}{N(t)} u\left(t, \frac{\cdot}{N(t)}\right), \frac{1}{N^2(t)} u_t\left(t, \frac{\cdot}{N(t)}\right)\right) \,\Big|\, t \in I_{\max} \right\} \tag{3-3}$$

*is precompact in $\dot{H}^{1/2} \times \dot{H}^{-1/2}(\mathbb{R}^3)$.*

**Remark 7.** After passing to subsequences, scaling considerations, and possibly time reversal, we can assume, without loss of generality, that $T_+(\vec{u}) = +\infty$, and $N(t) \le 1$ on $[0, \infty)$. We can further reduce this to two separate cases: either we have

- $N(t) \equiv 1$ for all $t \ge 0$, or
- $\limsup_{t \to \infty} N(t) = 0$.

These reductions follow from general arguments and are now standard. See, for example, [Kenig and Merle 2010; Killip et al. 2009; Shen 2012] for more details.

In what follows it will be convenient to give a name to the compactness property (3-3) satisfied by the critical element.

**Definition 8.** Let $I \ni 0$ be a time interval and suppose $\vec{u}(t)$ is a solution to (1-4) on $I$. We will say $\vec{u}(t)$ has the *compactness property on $I$* if there exists a continuous function $N : I \to (0, \infty)$ such that the set

$$K := \left\{ \left( \frac{1}{N(t)} u\left(t, \frac{\cdot}{N(t)}\right), \frac{1}{N^2(t)} u_t\left(t, \frac{\cdot}{N(t)}\right) \right) \Bigm| t \in I_{\max} \right\}$$

is precompact in $\dot{H}^{1/2} \times \dot{H}^{-1/2}(\mathbb{R}^3)$.

**Remark 9.** A consequence of a critical element having the compactness property on an interval $I$ is that, after modulation, we can control the $\dot{H}^{1/2} \times \dot{H}^{-1/2}$ tails uniformly in $t \in I$. Indeed, by the Arzelà–Ascoli theorem, for any $\eta > 0$ there exists $C(\eta) < \infty$ such that

$$\begin{aligned} \int_{|x| \geq C(\eta)/N(t)} \left| |\nabla|^{1/2} u(t,x) \right|^2 dx + \int_{|\xi| \geq C(\eta) N(t)} |\xi| |\hat{u}(t,\xi)|^2 \, d\xi \leq \eta, \\ \int_{|x| \geq C(\eta)/N(t)} \left| |\nabla|^{-1/2} u_t(t,x) \right|^2 dx + \int_{|\xi| \geq C(\eta) N(t)} |\xi|^{-1} |\hat{u_t}(t,\xi)|^2 \, d\xi \leq \eta, \end{aligned} \tag{3-4}$$

for all $t \in I$.

Another standard fact about solutions to (1-4) with the compactness property is that any Strichartz norm of the linear part of the evolution vanishes as $t \to T_-$ and as $t \to T_+$. A concentration compactness argument then implies that the linear part of the evolution must vanish weakly in $\dot{H}^{1/2} \times \dot{H}^{-1/2}$, i.e., for any $t_0 \in I$,

$$S(t_0 - t) u(t) \rightharpoonup 0 \tag{3-5}$$

weakly in $\dot{H}^{1/2} \times \dot{H}^{-1/2}$ as $t \nearrow \sup I$, $t \searrow \inf I$; see [Tao et al. 2008, Section 6; Shen 2012, Proposition 3.6]. This implies the following lemma, which will be crucial in the proof of Theorem 1.1.

**Lemma 3.2** [Tao et al. 2008, Section 6; Shen 2012, Proposition 3.6]. *Let $\vec{u}(t)$ be a solution to* (1-4) *with the compactness property on an interval $I = (T_-, T_+)$. Then for any $t_0 \in I$, we can write*

$$\begin{aligned} \int_{t_0}^{T} S(t_0 - s)(0, \pm u^3) \, ds \rightharpoonup \vec{u}(t_0) \quad \text{as } T \nearrow T_+ \quad \text{weakly in } \dot{H}^{1/2} \times \dot{H}^{-1/2}, \\ -\int_{T}^{t_0} S(t_0 - s)(0, \pm u^3) \, ds \rightharpoonup \vec{u}(t_0) \quad \text{as } T \searrow T_- \quad \text{weakly in } \dot{H}^{1/2} \times \dot{H}^{-1/2}. \end{aligned} \tag{3-6}$$

## 4. Additional regularity for critical elements

In this section we show that the critical element $\vec{u}(t)$ from Section 3 has additional regularity for $t \in I$. In particular, we prove the following result.

**Theorem 4.1.** *Let $\vec{u}(t)$ be a solution to* (1-4) *defined on a time interval $I = (T_-, \infty)$ with $T_- < 0$ and suppose that $\vec{u}(t)$ has the compactness property on $I$ with $N(t) \leq 1$ for all $t \in [0, \infty)$. Then for each $t \in I$ we have $\vec{u}(t) \in \dot{H}^1 \times L^2$, and the estimate*

$$\|\vec{u}(t)\|_{\dot{H}^1 \times L^2(\mathbb{R}^3)} \lesssim N(t)^{1/2} \tag{4-1}$$

*holds with a constant that is uniform for $t \in I$.*

**Remark 10.** We note that all constants this section implicit in the symbol $\lesssim$ will be allowed to depend on the $L_t^\infty(I;\dot H^{1/2}\times\dot H^{-1/2})$-norm of $\vec u$, which is fixed.

We will prove Theorem 4.1 using a bootstrap procedure with two steps. In particular, we will first show that if $\vec u(t)$ has the compactness property on an interval $I$ as in Theorem 4.1, then $\vec u(t)$ must lie in $\dot H^{2/3}\times\dot H^{-1/3}$. We then use this result to attain Theorem 4.1.

**Proposition 4.2.** *Let $\vec u(t)$ be as in Theorem 4.1. Then for any $t_0\in I$, we have*

$$\|\vec u(t_0)\|_{\dot H^{2/3}\times\dot H^{-1/3}(\mathbb{R}^3)}\lesssim N(t_0)^{1/6}. \tag{4-2}$$

We momentarily postpone the proof of Proposition 4.2 and use it to deduce Theorem 4.1.

**4A.** ***Proof of Theorem 4.1 assuming Proposition 4.2.*** The first step is to establish refined Strichartz estimates.

**Lemma 4.3.** *Let $\vec u(t)$ satisfy the conclusions of Proposition 4.2. Then there exists $\delta>0$ sufficiently small that for any $t_0\in I$,*

$$\|u\|_{L_t^3L_x^6([t_0-\delta/N(t_0),t_0+\delta/N(t_0)]\times\mathbb{R}^3)}\lesssim N(t_0)^{1/6}. \tag{4-3}$$

*Proof.* To simplify notation, let $J=[t_0-\delta/N(t_0),\,t_0+\delta/N(t_0)]$. We also let $Y=L_t^\infty L_x^{18/5}\cap L_t^3L_x^6$ with the natural norm. Using Strichartz estimates, we have

$$\begin{aligned}\|u\|_{Y(J\times\mathbb{R}^3)}&\lesssim\|\vec u(t_0)\|_{\dot H^{2/3}\times\dot H^{-1/3}(\mathbb{R}^3)}+\|u^3\|_{L_t^{6/5}L_x^{3/2}(J\times\mathbb{R}^3)}\\&\lesssim N(t_0)^{1/6}+(\delta/N(t_0))^{1/3}\|u\|^{3/2}_{L_t^3L_x^6(J\times\mathbb{R}^3)}\|u\|^{3/2}_{L_t^\infty L_x^{18/5}(J\times\mathbb{R}^3)}\\&\lesssim N(t_0)^{1/6}+\delta^{1/3}\big(N(t_0)^{-1/6}\|u\|_{Y(J\times\mathbb{R}^3)}\big)^2\|u\|_{Y(J\times\mathbb{R}^3)}.\end{aligned}\tag{4-4}$$

Choosing $\delta=\delta\big(N(t_0)^{-1/6}\|u(t_0)\|_{\dot H^{2/3}\times\dot H^{-1/3}}\big)>0$ small enough, the lemma follows by a standard bootstrapping argument. We remark that here it is important that the constant in (4-2) is uniform in $t_0\in I$. □

An immediate consequence of Lemma 4.3 is the following estimate.

**Corollary 4.4.** *There exists $\delta>0$ such that for each $t_0\in I$, we have*

$$\|u^3\|_{L_t^1L_x^2([t_0-\delta/N(t_0),t_0+\delta/N(t_0)]\times\mathbb{R}^3)}\lesssim N(t_0)^{1/2}. \tag{4-5}$$

We are now ready to begin the proof of Theorem 4.1 assuming Proposition 4.2.

*Proof that Proposition 4.2 implies Theorem 4.1.* Fix $t_0\in I$. By translating in time, we can, without loss of generality, assume that $t_0=0$. Let

$$v=u+\frac{i}{\sqrt{-\Delta}}u_t. \tag{4-6}$$

Then we have

$$\|v(t)\|_{\dot H^1}\simeq\|\vec u(t)\|_{\dot H^1\times L^2}. \tag{4-7}$$

And if $\vec{u}(t)$ solves (1-4), then $v(t)$ is a solution to

$$v_t = -i\sqrt{-\Delta}v \pm \frac{i}{\sqrt{-\Delta}}u^3, \tag{4-8}$$

where $+$ corresponds to the focusing equation and $-$ to the defocusing equation. By Duhamel's principle, for any $T$ such that $T_- < T < 0$,

$$v(0) = e^{iT\sqrt{-\Delta}}v(T) \pm \frac{i}{\sqrt{-\Delta}} \int_T^0 e^{i\tau\sqrt{-\Delta}}u^3(\tau)\,d\tau. \tag{4-9}$$

Next, we define an approximate identity $\{\psi_M\}_{M>0}$. Indeed, let $\psi \in C_0^\infty(\mathbb{R}^3)$ be a radial function, normalized in $L^1(\mathbb{R}^3)$ so that $\|\psi\|_{L^1} = 1$. Set $\psi_M(x) := M^3\psi(Mx)$. We then define the operator $Q_M$ given by convolution with $\psi_M$:

$$Q_M f(x) := \int_{\mathbb{R}^3} \psi_M(x-y) f(y)\,dy. \tag{4-10}$$

Of course $Q_M$ is also a Fourier multiplier operator, given by multiplication on the Fourier side by $\widehat{\psi_M}$, where $\widehat{\psi_M}(\xi) = \hat{\psi}\left(\frac{\xi}{M}\right)$. Since $\psi \in C_0^\infty$, we have $\hat{\psi} \in \mathcal{S}(\mathbb{R}^3)$.

With this setup, it clearly suffices to prove that there exists an $M_0 > 0$ such that

$$\|Q_M v(0)\|_{\dot{H}^1} \lesssim N(0)^{1/2} \tag{4-11}$$

for all $M \geq M_0 > 0$ with a constant that is independent of $M$.

To begin, let $T_- < T_1 < 0 < T_2 < \infty$ and let $M$ be a large number to be determined below. By the Duhamel formula, we have

$$\begin{aligned}\langle Q_M v(0), Q_M v(0)\rangle_{\dot{H}^1} = \Big\langle Q_M\Big(e^{iT_2\sqrt{-\Delta}}v(T_2) \mp \frac{i}{\sqrt{-\Delta}}\int_0^{T_2} e^{it\sqrt{-\Delta}}u^3\,dt\Big),\\ Q_M\Big(e^{iT_1\sqrt{-\Delta}}v(T_1) \pm \frac{i}{\sqrt{-\Delta}}\int_{T_1}^0 e^{i\tau\sqrt{-\Delta}}u^3\,d\tau\Big)\Big\rangle_{\dot{H}^1},\end{aligned} \tag{4-12}$$

where the bracket $\langle\cdot,\cdot\rangle_{\dot{H}^1}$ is the $\dot{H}^1$ inner product, namely

$$\langle f, g\rangle_{\dot{H}^1} = \operatorname{Re}\int_{\mathbb{R}^3} \sqrt{-\Delta}f \cdot \overline{\sqrt{-\Delta}g}.$$

We start by estimating the term that contains both Duhamel terms:

$$\begin{aligned}\Big|\Big\langle Q_M\Big(\frac{i}{\sqrt{-\Delta}}\int_0^{T_2} e^{it\sqrt{-\Delta}}u^3(t)\,dt\Big), Q_M\Big(\frac{i}{\sqrt{-\Delta}}\int_0^{T_1} e^{i\tau\sqrt{-\Delta}}u^3(\tau)\,d\tau\Big)\Big\rangle_{\dot{H}^1}\Big|\\ = \Big|\Big\langle Q_M\Big(\int_0^{T_2} e^{it\sqrt{-\Delta}}u^3(t)\,dt\Big), Q_M\Big(\int_{T_1}^0 e^{i\tau\sqrt{-\Delta}}(u^3(\tau))\,d\tau\Big)\Big\rangle_{L^2}\Big|.\end{aligned} \tag{4-13}$$

With $\delta > 0$ as in Corollary 4.4, we use (4-5) to deduce that

$$\int_{-\delta/N(0)}^{\delta/N(0)} \|Q_M(u^3(t))\|_{L_x^2(\mathbb{R}^3)}\,dt \lesssim N(0)^{1/2}. \tag{4-14}$$

Next, define a decreasing, smooth, radial function $\chi \in C_0^\infty(\mathbb{R}^3)$, with $\chi(x) \equiv 1$ for all $|x| \le 1$ and $\chi(x) = 0$ if $|x| \ge 2$. Also, let $c > 0$ be a small constant, say $c = \frac{1}{4}$. We have

$$\left\| Q_M\left(\int_{\delta/N(0)}^\infty e^{it\sqrt{-\Delta}}\left(1-\chi\left(\frac{x}{c|t|}\right)\right)u^3(t)\,dt\right)\right\|_{L_x^2(\mathbb{R}^3)} \lesssim \int_{\delta/N(0)}^\infty \left\|(1-\chi)\left(\frac{x}{c|t|}\right)u^3(t)\right\|_{L_x^2(\mathbb{R}^3)} dt. \quad (4\text{-}15)$$

By the radial Sobolev embedding (i.e., Lemma 2.2), we note that

$$\||x|^{3/2}u^3\|_{L_x^2} \lesssim \||x|^{1/2}u\|_{L_x^6}^3 \lesssim \|u\|_{\dot H^{1/2}}. \quad (4\text{-}16)$$

Therefore,

$$\left\|(1-\chi)\left(\frac{x}{c|t|}\right)u^3\right\|_{L_x^2(\mathbb{R}^3)} \lesssim \frac{1}{|t|^{3/2}}\|u\|_{\dot H^{1/2}}^3. \quad (4\text{-}17)$$

Thus,

$$\int_{\delta/N(0)}^\infty \left\|Q_M\left((1-\chi)\left(\frac{x}{c|t|}\right)u^3(t)\right)\right\|_{L_x^2(\mathbb{R}^3)} dt \lesssim \delta^{-1/2}N(0)^{1/2}. \quad (4\text{-}18)$$

The same is also true in the negative time direction. With these estimates in hand, we write (4-13) as a pairing

$$\langle A+B,\, A'+B'\rangle = \langle A+B,\, A'\rangle + \langle A,\, A'+B'\rangle + \langle B,\, B'\rangle - \langle A,\, A'\rangle, \quad (4\text{-}19)$$

where

$$\begin{aligned} A &:= Q_M\left(\int_0^{\delta/N(0)} e^{it\sqrt{-\Delta}}u^3\,dt + \int_{\delta/N(0)}^{T_2} e^{it\sqrt{-\Delta}}(1-\chi)\left(\frac{x}{c|t|}\right)u^3\,dt\right),\\ B &:= Q_M\left(\int_{\delta/N(0)}^{T_2} e^{it\sqrt{-\Delta}}\chi\left(\frac{x}{c|t|}\right)u^3(t)\,dt\right), \end{aligned} \quad (4\text{-}20)$$

and $A'$, $B'$ are the corresponding integrals in the negative time direction. We start by estimating the term $\langle A, A'\rangle$. By (4-14) and (4-18),

$$\begin{aligned} &\left\langle Q_M\left(\int_0^{\delta/N(0)} e^{it\sqrt{-\Delta}}u^3\,dt + \int_{\delta/N(0)}^{T_2} e^{it\sqrt{-\Delta}}(1-\chi)\left(\frac{x}{c|t|}\right)u^3\,dt\right),\right.\\ &\qquad \left. Q_M\left(\int_{-\delta/N(0)}^0 e^{i\tau\sqrt{-\Delta}}u^3\,d\tau + \int_{T_1}^{-\delta/N(0)} e^{i\tau\sqrt{-\Delta}}(1-\chi)\left(\frac{x}{c|\tau|}\right)u^3\,d\tau\right)\right\rangle_{L^2} \lesssim N(0). \end{aligned} \quad (4\text{-}21)$$

Next, we examine the term $\langle B, B'\rangle$, which is given by

$$\begin{aligned} &\int_{T_1}^{-\delta/N(0)}\int_{\delta/N(0)}^{T_2}\left\langle Q_M\left(e^{it\sqrt{-\Delta}}\chi\left(\frac{x}{c|t|}\right)u^3(t)\right),\, Q_M\left(e^{i\tau\sqrt{-\Delta}}\chi\left(\frac{x}{c|\tau|}\right)u^3(\tau)\right)\right\rangle_{L^2} dt\,d\tau\\ &\quad = \int_{T_1}^{-\delta/N(0)}\int_{\delta/N(0)}^{T_2}\left\langle Q_M\left(\chi\left(\frac{x}{c|t|}\right)u^3(t)\right),\, Q_M\left(e^{i(\tau-t)\sqrt{-\Delta}}\chi\left(\frac{x}{c|\tau|}\right)u^3(\tau)\right)\right\rangle_{L^2} dt\,d\tau. \end{aligned}$$

To estimate the above, we begin by noting that the sharp Huygens principle implies that

$$\left(e^{i(t-\tau)\sqrt{-\Delta}}\chi\left(\frac{\cdot}{c|\tau|}\right)u^3(\tau)\right)(x)$$

is supported on the set $|x| > \frac{3}{4}|t-\tau|$ for $c = \frac{1}{4}$. Also, we note that since $t > 0$ and $\tau < 0$, we have $|t-\tau| > |t|, |\tau|$. Next, recall that the kernel of $Q_M$ is given by the function $\psi_M(x) = M^3\psi(Mx)$, where $\psi \in C_0^\infty$. This implies that for $M \gg N(0)^{-1}$ large enough, we have

$$\operatorname{supp}\left(\int_{\mathbb{R}^3}\psi_M(x-z)\chi\left(\frac{z}{c|t|}\right)u^3(t,z)\,dz\right) \subset \left\{|x| < \frac{1}{2}|t|\right\},$$
$$\operatorname{supp}\left(\int_{\mathbb{R}^3}\psi_M(x-y)\left(e^{i(\tau-t)\sqrt{-\Delta}}\chi\left(\frac{\cdot}{c|\tau|}\right)u^3(\tau)\right)(y)\,dy\right) \subset \left\{|x| > \frac{1}{2}|t-\tau|\right\}.$$

Therefore, as long as $M$ is chosen large enough, say for $M \geq M_0 \gg N(0)^{-1}$, and since $|t| < |t-\tau|$ for $t > \delta/N(0)$ and $\tau < -\delta/N(0)$, we have

$$\left\langle \int \psi_M(x-z)\chi\left(\frac{z}{c|t|}\right)u^3(t,z)\,dz, \int \psi_M(x-y)\left(e^{i(\tau-t)\sqrt{-\Delta}}\chi\left(\frac{\cdot}{c|\tau|}\right)u^3(\tau)\right)(y)\,dy\right\rangle_{L^2} = 0. \quad (4\text{-}22)$$

It remains to estimate the terms $\langle A, A'+B'\rangle$ and $\langle A+B, A'\rangle$, which are given by

$$\left\langle Q_M\left(\int_0^{\delta/N(0)} e^{it\sqrt{-\Delta}}u^3(t)\,dt + \int_{\delta/N(0)}^{T_2} e^{it\sqrt{-\Delta}}(1-\chi)\left(\frac{x}{c|t|}\right)u^3(t)\,dt\right),\right.$$
$$\left. Q_M\left(\int_{T_1}^0 e^{i\tau\sqrt{-\Delta}}u^3(\tau)\,d\tau\right)\right\rangle_{L^2} \quad (4\text{-}23)$$

and

$$\left\langle Q_M\left(\int_0^{T_2} e^{it\sqrt{-\Delta}}u^3(t)\,dt\right),\right.$$
$$\left. Q_M\left(\int_{T_1}^{-\delta/N(0)} e^{i\tau\sqrt{-\Delta}}(1-\chi)\left(\frac{\cdot}{c|\tau|}\right)u^3(\tau)\,d\tau + \int_{-\delta/N(0)}^0 e^{i\tau\sqrt{-\Delta}}u^3(\tau)\,d\tau\right)\right\rangle_{L^2}. \quad (4\text{-}24)$$

We provide the details for how to handle (4-23), as the estimates for (4-24) are similar. First recall that by the Duhamel principle (4-9), we can write

$$Q_M\int_{T_1}^0 e^{i\tau\sqrt{-\Delta}}u^3(\tau)\,d\tau = \mp i\sqrt{-\Delta}Q_M v(0) \pm i\sqrt{-\Delta}e^{iT_1\sqrt{-\Delta}}Q_M v(T_1). \quad (4\text{-}25)$$

Using again (4-14) and (4-18), we have

$$\left|\left\langle \sqrt{-\Delta}Q_M v(0), Q_M\left(\int_0^{\delta/N(0)} e^{it\sqrt{-\Delta}}u^3\,dt + \int_{\delta/N(0)}^{T_2} e^{it\sqrt{-\Delta}}(1-\chi)\left(\frac{x}{c|t|}\right)u^3\,dt\right)\right\rangle_{L^2}\right|$$
$$\lesssim N(0)^{1/2}\|Q_M v(0)\|_{\dot H^1(\mathbb{R}^3)}. \quad (4\text{-}26)$$

We remark that all of the estimates established so far have been uniform in $T_- < T_1 < 0 < T_2 < T_+$. This is important as we will now take limits, $T_1 \searrow T_-$ and $T_2 \nearrow T_+$. Indeed, using the weak convergence result in Lemma 3.2, we claim that for any fixed $T_2 \in (0, T_+)$, we have

$$\lim_{T_1 \searrow T_-} \Big\langle i\sqrt{-\Delta} e^{iT_1\sqrt{-\Delta}} Q_M v(T_1), \\ Q_M \Big( \int_0^{\delta/N(0)} e^{it\sqrt{-\Delta}} u^3 \, dt + \int_{\delta/N(0)}^{T_2} e^{it\sqrt{-\Delta}} (1-\chi)\Big(\frac{x}{c|t|}\Big) u^3 \, dt \Big) \Big\rangle_{L^2} = 0. \tag{4-27}$$

In fact, (4-14) and (4-18) imply that letting $T_2 \nearrow T_+$, for $M$ fixed,

$$(-\Delta)^{1/4} Q_M \Big( \int_0^{\delta/N(0)} e^{it\sqrt{-\Delta}} u^3 \, dt + \int_{\delta/N(0)}^{T_2} e^{it\sqrt{-\Delta}} (1-\chi)\Big(\frac{x}{c|t|}\Big) u^3 \, dt \Big)$$

converges in $L^2(\mathbb{R}^3)$ to

$$(-\Delta)^{1/4} Q_M \Big( \int_0^{\delta/N(0)} e^{it\sqrt{-\Delta}} u^3 \, dt + \int_{\delta/N(0)}^{T_+} e^{it\sqrt{-\Delta}} (1-\chi)\Big(\frac{x}{c|t|}\Big) u^3 \, dt \Big) \in L^2(\mathbb{R}^3).$$

Therefore, since Lemma 3.2 says that $e^{iT_1\sqrt{-\Delta}} v(T_1) \rightharpoonup 0$ weakly in $\dot H^{1/2}(\mathbb{R}^3)$ as $T_1 \searrow T_-$, we have

$$\lim_{T_1 \searrow T_-} \lim_{T_2 \nearrow T_+} \Big\langle \sqrt{-\Delta} e^{iT_1\sqrt{-\Delta}} Q_M v(T_1), \\ Q_M \Big( \int_0^{\delta/N(0)} e^{it\sqrt{-\Delta}} u^3 \, dt + \int_{\delta/N(0)}^{T_2} e^{it\sqrt{-\Delta}} (1-\chi)\Big(\frac{x}{c|t|}\Big) u^3 \, dt \Big) \Big\rangle_{L^2} = 0. \tag{4-28}$$

Thus we have proved that

$$\Big| \lim_{T_1 \searrow T_-} \lim_{T_2 \nearrow T_+} \langle A, A' + B' \rangle \Big| \lesssim N(0)^{1/2} \| Q_M v(0) \|_{\dot H^1}. \tag{4-29}$$

Using an identical argument, we can similarly prove that

$$\Big| \lim_{T_1 \searrow T_-} \lim_{T_2 \nearrow T_+} \langle A + B, A' \rangle \Big| \lesssim N(0)^{1/2} \| Q_M v(0) \|_{\dot H^1}, \tag{4-30}$$

where $A$, $B$, $A'$, $B'$ are defined as in (4-20). Therefore, combining (4-21), (4-26), (4-29), and (4-30), we have proved that

$$\Big| \lim_{T_1 \searrow T_-} \lim_{T_2 \nearrow T_+} \int_0^{T_2} \int_{T_1}^0 \big\langle e^{it\sqrt{-\Delta}} Q_M(u^3), e^{i\tau\sqrt{-\Delta}} Q_M(u^3) \big\rangle_{L^2} \, dt \, d\tau \Big| \lesssim \|v(0)\|_{\dot H^1(\mathbb{R}^3)} N(0)^{1/2} + N(0). \tag{4-31}$$

We are left to examine the terms in (4-12) (once expanded) that contain at most one Duhamel integral. Here we will rely heavily on the $\dot H^{1/2}$-weak convergence in Lemma 3.2.

Indeed, for a fixed $T_1$ and fixed $M$, we see that $\sqrt{-\Delta} Q_M v(T_1) \in \dot H^{1/2}(\mathbb{R}^3)$. Therefore, by Lemma 3.2, we have

$$\lim_{T_2 \nearrow T_+} \big\langle e^{iT_1\sqrt{-\Delta}} Q_M v(T_1),\ e^{iT_2\sqrt{-\Delta}} Q_M v(T_2) \big\rangle_{\dot H^1} = 0. \tag{4-32}$$

Next, for fixed $T_1 > T_-$, Corollary 4.4 and the bound (4-5) imply that

$$\|u^3\|_{L_t^1 L_x^2([T_1,0]\times\mathbb{R}^3)} < \infty,$$

which in turn implies that $\int_{T_1}^0 Q_M e^{it\sqrt{-\Delta}} u^3\, dt \in \dot{H}^{1/2}(\mathbb{R}^3)$, where again we are using that the multiplier $\widehat{\psi_M}$ is in $\mathcal{S}(\mathbb{R}^3)$. Therefore, Lemma 3.2 implies

$$\lim_{T_2 \nearrow T_+} \left\langle Q_M\left(\int_{T_1}^0 e^{it\sqrt{-\Delta}} u^3\, dt\right), e^{iT_2\sqrt{-\Delta}} Q_M v(T_2)\right\rangle_{\dot{H}^{1/2}} = 0. \tag{4-33}$$

Finally, we claim that

$$\lim_{T_1 \searrow T_-} \lim_{T_2 \nearrow T_+} \left\langle Q_M\big(e^{iT_1\sqrt{-\Delta}} v(T_1)\big), Q_M\left(\int_0^{T_2} e^{i\tau\sqrt{-\Delta}} u^3\, d\tau\right)\right\rangle_{\dot{H}^{1/2}} = 0. \tag{4-34}$$

To see this, we use (4-25). Indeed, using Lemma 3.2 again, we have

$$\begin{aligned} \lim_{T_1 \searrow T_-} \lim_{T_2 \nearrow T_+} &\left\langle Q_M\big(e^{iT_1\sqrt{-\Delta}} v(T_1)\big), \sqrt{-\Delta}\, Q_M\big(v(0) - e^{iT_2\sqrt{-\Delta}} v(T_2)\big)\right\rangle_{\dot{H}^{1/2}} \\ &= \lim_{T_1 \searrow T_-} \left\langle Q_M\big(e^{iT_1\sqrt{-\Delta}} v(T_1)\big), Q_M\big(\sqrt{-\Delta}\, v(0)\big)\right\rangle_{\dot{H}^{1/2}} = 0. \end{aligned} \tag{4-35}$$

Therefore, (4-12) together with (4-31)–(4-35) imply that

$$\| Q_M v(0)\|^2_{\dot{H}^1(\mathbb{R}^3)} \lesssim \| Q_M v(0)\|_{\dot{H}^1(\mathbb{R}^3)} N(0)^{1/2} + N(0), \tag{4-36}$$

for all $M \geq M_0$ and with a uniform-in-$M$ constant. We can then conclude that

$$\| Q_M v(0)\|_{\dot{H}^1(\mathbb{R}^3)} \lesssim N(0)^{1/2} \tag{4-37}$$

uniformly in $M \geq M_0$. Therefore, $\|v(0)\|_{\dot{H}^1(\mathbb{R}^3)} \lesssim N(0)^{1/2}$. This proves Theorem 4.1, assuming the conclusions of Proposition 4.2. □

**4B. *Proof of Proposition 4.2.*** To complete the proof of Theorem 4.1 we prove Proposition 4.2. We begin with another refined Strichartz-type estimate.

**Lemma 4.5.** *Let $\eta > 0$. There exists $\delta = \delta(\eta) > 0$ such that for all $t_0 \in I$ we have*

$$\|u\|_{L^4_{t,x}([t_0 - \delta/N(t_0), t_0 + \delta/N(t_0)] \times \mathbb{R}^3)} \lesssim \eta. \tag{4-38}$$

*Proof.* Again, without loss of generality, suppose that $t_0 = 0$. Then define the interval $J = \left[-\frac{\delta}{N(0)}, \frac{\delta}{N(0)}\right]$. Using the Duhamel formula, we have

$$\|u\|_{L^4_{t,x}(J\times\mathbb{R}^3)} \leq \|S(t)\vec{u}(0)\|_{L^4_{t,x}(J\times\mathbb{R}^3)} + \left\| \int_0^t S(t-s)(0, \pm u^3)\, ds \right\|_{L^4_{t,x}(J\times\mathbb{R}^3)}. \tag{4-39}$$

We estimate the first term on the right side of (4-39) as follows. First choose $C(\eta)$ as in Remark 9, (3-4), so that

$$\| P_{\geq C(\eta)N(0)} \vec{u}(0)\|_{\dot{H}^{1/2}\times \dot{H}^{-1/2}(\mathbb{R}^3)} \leq \eta. \tag{4-40}$$

Note that by compactness, $C(\eta)$ above can be chosen uniformly in $t \in I$, which is why it suffices to only consider $t_0 = 0$ in this argument. Next, we have

$$\|S(t)\vec{u}(0)\|_{L^4_{t,x}(J\times\mathbb{R}^3)} \lesssim \|S(t) P_{\geq C(\eta)N(0)}\vec{u}(0)\|_{L^4_{t,x}(J\times\mathbb{R}^3)} + \|S(t) P_{\leq C(\eta)N(0)}\vec{u}(0)\|_{L^4_{t,x}(J\times\mathbb{R}^3)}. \tag{4-41}$$

We use (4-40) together with Strichartz estimates to handle the first term on the right side above:

$$\|S(t)P_{\geq C(\eta)N(0)}\vec{u}(0)\|_{L^4_{t,x}(J\times\mathbb{R}^3)} \lesssim \|P_{\geq C(\eta)N(0)}\vec{u}(0)\|_{\dot{H}^{1/2}\times\dot{H}^{-1/2}(\mathbb{R}^3)} \lesssim \eta. \tag{4-42}$$

To control the second term we use Bernstein's inequalities, (2-1), and Sobolev embedding,

$$\|P_{\leq C(\eta)N(0)}S(t)\vec{u}(0)\|_{L^4_x(\mathbb{R}^3)} \lesssim C(\eta)^{1/4}N(0)^{1/4}\|u(0)\|_{\dot{H}^{1/2}(\mathbb{R}^3)}. \tag{4-43}$$

Taking the $L^4_t(J)$-norm of both sides above gives

$$\|S(t)P_{\leq C(\eta)N(0)}\vec{u}(0)\|_{L^4_{t,x}(J\times\mathbb{R}^3)} \lesssim C(\eta)^{1/4}\delta^{1/4}. \tag{4-44}$$

Next we use Strichartz estimates on the second term on the right side of (4-39).

$$\left\|\int_0^t S(t-s)(0,\pm u^3)\,ds\right\|_{L^4_{t,x}(J\times\mathbb{R}^3)} \lesssim \|u^3\|_{L^{4/3}_{t,x}(J\times\mathbb{R}^3)} \lesssim \|u\|^3_{L^4_{t,x}(J\times\mathbb{R}^3)}. \tag{4-45}$$

Combining all of the above, we obtain

$$\|u\|_{L^4_{t,x}(J\times\mathbb{R}^3)} \lesssim \eta + C(\eta)^{1/4}\delta^{1/4} + \|u\|^3_{L^4_{t,x}(J\times\mathbb{R}^3)}. \tag{4-46}$$

The proof is concluded using the usual continuity argument after taking $\delta$ small enough. □

*Proof of Proposition 4.2.* We can again, without loss of generality, just consider the case $t_0 = 0$. We will prove Proposition 4.2 by finding a frequency envelope $\alpha_k(0)$ such that

$$\|(P_k u(0), P_k u_t(0))\|_{\dot{H}^{2/3}\times\dot{H}^{-1/3}} \lesssim 2^{k/6}\alpha_k(0), \quad \|\{2^{k/6}\alpha_k(0)\}_{k\in\mathbb{Z}}\|_{\ell^2} \lesssim N(0)^{1/6}. \tag{4-47}$$

Once we find $\alpha_k(0)$ satisfying (4-47), Proposition 4.2 follows from Definition 3. With this in mind we first establish the following claim:

**Claim 4.6.** *There exists a number $\eta_0 > 0$ such that the following holds. Let $0 < \eta < \eta_0$ and let $J := [-\delta/N(0), \delta/N(0)]$, where $\delta = \delta(\eta) > 0$ is chosen as in Lemma 4.5. Define*

$$\begin{aligned} a_k &:= 2^{k/2}\|P_k u\|_{L^\infty_t L^2_x(J)} + 2^{-k/2}\|P_k u_t\|_{L^\infty_t L^2_x(J)} + 2^{k/4}\|P_k u\|_{L^8_t L^{8/3}_x(J)}, \\ a_k(0) &:= 2^{k/2}\|P_k u(0)\|_{L^2_x(\mathbb{R}^3)} + 2^{-k/2}\|P_k u_t(0)\|_{L^2_x(\mathbb{R}^3)}. \end{aligned} \tag{4-48}$$

*Next define frequency envelopes $\alpha_k$ and $\alpha_k(0)$ by*

$$\alpha_k := \sum_j 2^{-|j-k|/8}a_j, \quad \alpha_k(0) := \sum_j 2^{-|j-k|/8}a_j(0). \tag{4-49}$$

*Then, as long as $\eta_0$ is chosen small enough, we have*

$$a_k \lesssim a_k(0) + \eta^2 \sum_{j\geq k-3} 2^{(k-j)/4}a_j \tag{4-50}$$

*and*

$$\alpha_k \lesssim \alpha_k(0). \tag{4-51}$$

*Proof of Claim 4.6.* To prove (4-50), we note that Strichartz estimates, together with Lemma 4.5, imply that

$$\begin{aligned} a_k &= 2^{k/2}\|P_k u\|_{L_t^\infty L_x^2(J)} + 2^{k/4}\|P_k u\|_{L_t^8 L_x^{8/3}(J)} \\ &\lesssim 2^{k/2}\|P_k u(0)\|_{L_x^2(\mathbb{R}^3)} + 2^{-k/2}\|P_k u_t(0)\|_{L_x^2(\mathbb{R}^3)} + 2^{k/4}\|P_k(u^3)\|_{L_t^{8/5} L_x^{8/7}(J)} \\ &\lesssim a_k(0) + \eta^2 \sum_{j\ge k-3} 2^{(k-j)/4} a_j. \end{aligned} \tag{4-52}$$

To prove the last line above we note that it will suffice, by Hölder's inequality in time and Lemma 4.5, to show that

$$\|P_k(u^3)\|_{L_x^{8/7}} \lesssim \|u\|_{L^4}^2 \sum_{j\ge k-3} \|P_j u\|_{L^{8/3}}. \tag{4-53}$$

First, since $P_k((P_{\le k-4}u)^3) = 0$, we have

$$\begin{aligned} \|P_k u^3\|_{L_x^{8/7}} &\lesssim \big\|P_k[(P_{\le k-4}u)^2 P_{\ge k-3}u]\big\|_{L^{8/7}} + \big\|P_k[P_{\le k-4}u(P_{\ge k-3}u)^2]\big\|_{L^{8/7}} + \|P_k[P_{\ge k-3}u]^3\|_{L^{8/7}} \\ &\lesssim \|u\|_{L^4}^2 \|P_{\ge k-3}u\|_{L^{8/3}}, \end{aligned}$$

where the last inequality follows from the boundedness of $P_k$ on $L^p$ and by Holder's inequality. This proves (4-53), and thus we have established (4-50).

To prove (4-51), we use (4-50) to obtain

$$\sum_j 2^{-|j-k|/8} a_j \lesssim \sum_j a_j(0) 2^{-|j-k|/8} + \eta^2 \sum_j 2^{-|j-k|/8} \sum_{j_1\ge j-3} 2^{(j-j_1)/4} a_{j_1}. \tag{4-54}$$

Reversing the order of summation in the second term above gives

$$\begin{aligned} \sum_{j_1\le k}\sum_{j\le j_1+3} 2^{(j-j_1)/4} 2^{(j-k)/8} a_{j_1} &\lesssim \sum_{j_1\le k} 2^{(j_1-k)/8} a_{j_1} \lesssim \alpha_k, \\ \sum_{j_1> k}\sum_{j\le j_1+3} 2^{(j-j_1)/4} 2^{-|j-k|/8} a_{j_1} &\lesssim \sum_{j_1> k} \big(2^{-(k-j_1)/4} + 2^{-(k-j_1)/8}\big) a_{j_1} \lesssim \alpha_k. \end{aligned} \tag{4-55}$$

Therefore, (4-54) implies that

$$\alpha_k \lesssim \alpha_k(0) + \eta^2 \alpha_k, \tag{4-56}$$

which in turn yields (4-51) as long as $\eta > 0$ is small enough. □

We now return to the proof of Proposition 4.2. We note that the calculation in the proof of Claim 4.6 also allows us to deduce that

$$2^{k/2}\left\| P_k \int_0^{\delta/N(0)} \frac{e^{-it\sqrt{-\Delta}}}{\sqrt{-\Delta}} u^3(t)\,dt \right\|_{L_x^2(\mathbb{R}^3)} \lesssim \eta^2 \sum_{j\ge k-3} 2^{(k-j)/4} a_j. \tag{4-57}$$

Next, we claim that for any $s_0 \in \big(0, \frac{1}{2}\big]$ we have the estimate

$$\int_{\delta/N(0)}^\infty \left\| \frac{e^{-it\sqrt{-\Delta}}}{\sqrt{-\Delta}} (1-\chi)\left(\frac{x}{c|t|}\right) u^3 \right\|_{\dot H^{1/2+s_0}(\mathbb{R}^3)} dt \lesssim N(0)^{s_0} \delta^{-s_0}, \tag{4-58}$$

where $c > 0$ is a fixed small constant ($c = \frac{1}{4}$ will do) and $\chi \in C_0^\infty(\mathbb{R}^3)$ is radial, $\chi(x) = 1$ for all $|x| \le 1$, and $\chi(x) = 0$ for all $|x| \ge 2$. To prove (4-58), we note that by Sobolev embedding,

$$\left\| \frac{e^{-it\sqrt{-\Delta}}}{\sqrt{-\Delta}}(1-\chi)\left(\frac{x}{c|t|}\right)u^3 \right\|_{\dot H^{1/2+s_0}} = \left\| (1-\chi)\left(\frac{x}{c|t|}\right)u^3 \right\|_{\dot H^{-1/2+s_0}} \lesssim \left\| (1-\chi)\left(\frac{x}{c|t|}\right)u^3 \right\|_{L^p},$$

where $\dfrac{1}{p} = \dfrac{2}{3} - \dfrac{s_0}{3}$. Then using the radial Sobolev embedding, i.e., Lemma 2.2, we have

$$(c|t|)^{1+s_0} \left\| (1-\chi)\left(\frac{x}{c|t|}\right)u^3 \right\|_{L^p} \lesssim \left\| (1-\chi)^{1/3}\left(\frac{x}{c|t|}\right)|x|^{(1+s_0)/3} u \right\|_{L^{3p}}^3 \lesssim \|u\|_{\dot H^{1/2}}^3.$$

Hence,

$$\left\| \frac{e^{-it\sqrt{-\Delta}}}{\sqrt{-\Delta}}(1-\chi)\left(\frac{x}{c|t|}\right)u^3 \right\|_{\dot H^{1/2+s_0}} \lesssim |t|^{-1-s_0}\|u\|_{L_t^\infty \dot H^{1/2}}.$$

Integrating the above in time from $t = \delta/N(0)$ to $t = +\infty$ then yields (4-58).

Once again by the weak convergence result in Lemma 3.2, we have

$$\langle P_k v(0), P_k v(0)\rangle_{\dot H^{1/2}} = \left\langle P_k v(0), P_k\left(\lim_{T_2 \nearrow T_+} \pm \frac{i}{\sqrt{-\Delta}} \int_0^{T_2} e^{i\tau\sqrt{-\Delta}} u^3(\tau)\, d\tau\right)\right\rangle_{\dot H^{1/2}},$$

which for all $T_- < T_1 < 0$ is equal to

$$\begin{aligned}\lim_{T_2 \nearrow T_+} &\left\langle P_k\big(e^{iT_1\sqrt{-\Delta}} v(T_1)\big), \frac{\pm i}{\sqrt{-\Delta}} P_k\left(\int_0^{T_2} e^{i\tau\sqrt{-\Delta}} u^3\, d\tau\right)\right\rangle_{\dot H^{1/2}} \\ &+ \lim_{T_2 \nearrow T_+} \left\langle \frac{1}{\sqrt{-\Delta}} P_k\left(\int_{T_1}^0 e^{it\sqrt{-\Delta}} u^3\, dt\right), \frac{1}{\sqrt{-\Delta}} P_k\left(\int_0^{T_2} e^{i\tau\sqrt{-\Delta}} u^3\, d\tau\right)\right\rangle_{\dot H^{1/2}}. \quad (4\text{-}59)\end{aligned}$$

As $T_1 \searrow T_-$, we note that (4-59) $\to 0$. Indeed, by (4-9),

$$\begin{aligned}\lim_{T_1 \searrow T_-} \lim_{T_2 \nearrow T_+} &\left\langle P_k\big(e^{iT_1\sqrt{-\Delta}} v(T_1)\big), \frac{\pm i}{\sqrt{-\Delta}} P_k\left(\int_0^{T_2} e^{i\tau\sqrt{-\Delta}} u^3\, d\tau\right)\right\rangle_{\dot H^{1/2}} \\ &= \lim_{T_1 \searrow T_-} \lim_{T_2 \nearrow T_+} \left\langle P_k\big(e^{iT_1\sqrt{-\Delta}} v(T_1)\big), P_k\big(v(0) - e^{iT_2\sqrt{-\Delta}} v(T_2)\big)\right\rangle_{\dot H^{1/2}} \\ &= \lim_{T_1 \searrow T_-} \left\langle P_k\big(e^{iT_1\sqrt{-\Delta}} v(T_1)\big), P_k(v(0))\right\rangle_{\dot H^{1/2}} = 0. \quad (4\text{-}60)\end{aligned}$$

Therefore,

$$\langle P_k v(0), P_k v(0)\rangle_{\dot H^{1/2}} = \lim_{T_1 \searrow T_-} \lim_{T_2 \nearrow T_+} \left\langle \frac{1}{\sqrt{-\Delta}} P_k\left(\int_{T_1}^0 e^{it\sqrt{-\Delta}} u^3\, dt\right), \frac{1}{\sqrt{-\Delta}} P_k\left(\int_0^{T_2} e^{i\tau\sqrt{-\Delta}} u^3\, d\tau\right)\right\rangle_{\dot H^{1/2}}.$$

To estimate the right side above, we split each term into two pieces and use the identity

$$\langle A+B, A'+B'\rangle = \langle A+B, A'\rangle + \langle A, A'+B'\rangle - \langle A, A'\rangle + \langle B, B'\rangle, \quad (4\text{-}61)$$

where

$$A := P_k\biggl(\int_{-\delta/N(0)}^{0} \frac{e^{it\sqrt{-\Delta}}}{\sqrt{-\Delta}} u^3\,dt\biggr) + P_k\biggl(\int_{T_1}^{-\delta/N(0)} \frac{e^{it\sqrt{-\Delta}}}{\sqrt{-\Delta}} (1-\chi)\Bigl(\frac{x}{c|t|}\Bigr) u^3\,dt\biggr),$$

$$B := P_k\biggl(\int_{T_1}^{-\delta/N(0)} \frac{e^{it\sqrt{-\Delta}}}{\sqrt{-\Delta}} \chi\Bigl(\frac{x}{c|t|}\Bigr) u^3\,dt\biggr),$$

and $A'$, $B'$ are the analogous quantities in the positive time direction.

We begin by estimating the first two terms on the right side of (4-61). In fact, an identical argument applies to both of these terms, so we only provide details for the term $\langle A+B, A'\rangle$. To begin, we note that by (4-58), we have

$$\biggl\| P_k\biggl(\int_{\delta/N(0)}^{T_+} \frac{e^{i\tau\sqrt{-\Delta}}}{\sqrt{-\Delta}} (1-\chi)\Bigl(\frac{x}{c|\tau|}\Bigr) u^3\,d\tau\biggr)\biggr\|_{\dot H^{1/2}} \lesssim 2^{-k/6} b_k, \tag{4-62}$$

where $b_k = 2^{-k(s_0(k)-1/6)} N(0)^{s_0(k)}$ and we are free to choose any $s_0 = s_0(k) \in \bigl(0, \frac{1}{2}\bigr]$. Setting $s_0(k) = \frac{5}{24}$ if $2^k \geq N(0)$ and $s_0(k) = \frac{1}{8}$ if $2^k < N(0)$, we have

$$b_k := \begin{cases} 2^{-k/24} N(0)^{5/24} \delta^{-5/24} & \text{if } 2^k \geq N(0), \\ 2^{k/24} N(0)^{1/8} \delta^{-1/8} & \text{if } 2^k < N(0). \end{cases} \tag{4-63}$$

Then

$$\|\{b_k\}\|_{\ell^2} \lesssim N(0)^{1/6}. \tag{4-64}$$

Therefore with $b_k$ as in (4-63), we can combine (4-57) and (4-62) to deduce that

$$\biggl\| P_k\biggl(\int_0^{\delta/N(0)} \frac{e^{i\tau\sqrt{-\Delta}}}{\sqrt{-\Delta}} u^3\,dt\biggr)\biggr\|_{\dot H^{1/2}} + \biggl\| P_k\biggl(\int_{\delta/N(0)}^{T_+} \frac{e^{i\tau\sqrt{-\Delta}}}{\sqrt{-\Delta}} (1-\chi)\Bigl(\frac{x}{c|\tau|}\Bigr) u^3\,d\tau\biggr)\biggr\|_{\dot H^{1/2}} \lesssim \eta^2 \sum_{j\geq k-3} 2^{(k-j)/4} a_j + 2^{-k/6} b_k.$$

Then since

$$\frac{\pm i}{\sqrt{-\Delta}} \int_{T_1}^{0} e^{it\sqrt{-\Delta}} u^3\,dt \rightharpoonup v(0) \quad \text{in } \dot H^{1/2} \quad \text{as } T_1 \searrow T_-, \tag{4-65}$$

we can deduce the estimate

$$\begin{aligned} \lim_{T_1\searrow T_-} \lim_{T_2\nearrow T_+} \langle A+B, A'\rangle_{\dot H^{1/2}} = \lim_{T_1\searrow T_-} \lim_{T_2\nearrow T_+} \biggl\langle P_k\biggl(\int_{T_1}^{0} \frac{e^{it\sqrt{-\Delta}}}{\sqrt{-\Delta}} u^3\,dt\biggr), \\ P_k\biggl(\int_0^{\delta/N(0)} \frac{e^{-i\tau\sqrt{-\Delta}}}{\sqrt{-\Delta}} u^3\,d\tau\biggr) + P_k\biggl(\int_{\delta/N(0)}^{T_2} \frac{e^{i\tau\sqrt{-\Delta}}}{\sqrt{-\Delta}} (1-\chi)\Bigl(\frac{x}{c|\tau|}\Bigr) u^3\,d\tau\biggr)\biggr\rangle_{\dot H^{1/2}} \\ \lesssim a_k(0)\biggl(\eta^2 \sum_{j\geq k-3} 2^{(k-j)/4} a_j + 2^{-k/6} b_k\biggr). \end{aligned} \tag{4-66}$$

An identical calculation in the other time direction gives the same estimate for $\langle A, A' + B' \rangle$. Next, we estimate $\langle A, A' \rangle$, again using (4-57) and (4-62). We have

$$\begin{aligned}\Big\langle P_k\Big(\int_0^{\delta/N(0)} \frac{e^{i\tau\sqrt{-\Delta}}}{\sqrt{-\Delta}} u^3\, d\tau\Big) + P_k\Big(\int_{\delta/N(0)}^{T_2} \frac{e^{i\tau\sqrt{-\Delta}}}{\sqrt{-\Delta}} (1-\chi)\Big(\frac{x}{c|\tau|}\Big) u^3\, d\tau\Big), \\ P_k\Big(\int_{-\delta/N(0)}^{0} \frac{e^{it\sqrt{-\Delta}}}{\sqrt{-\Delta}} u^3\, dt\Big) + P_k\Big(\int_{T_1}^{-\delta/N(0)} \frac{e^{it\sqrt{-\Delta}}}{\sqrt{-\Delta}} (1-\chi)\Big(\frac{x}{c|t|}\Big) u^3\, dt\Big)\Big\rangle_{\dot H^{1/2}} \\ \lesssim \Big(\eta^2 \sum_{j\ge k-3} 2^{(k-j)/4} a_j\Big)^2 + 2^{-k/3} b_k^2. \quad (4\text{-}67)\end{aligned}$$

Finally, it remains to estimate $\langle B, B' \rangle$, which is given by

$$\begin{aligned}\int_{\delta/N(0)}^{T_2}\int_{T_1}^{-\delta/N(0)} \Big\langle P_k\Big(\frac{e^{it\sqrt{-\Delta}}}{\sqrt{-\Delta}}\chi\Big(\frac{x}{c|t|}\Big)u^3(t)\Big), P_k\Big(\frac{e^{i\tau\sqrt{-\Delta}}}{\sqrt{-\Delta}}\chi\Big(\frac{x}{c|\tau|}\Big)u^3(\tau)\Big)\Big\rangle_{\dot H^{1/2}} dt\, d\tau \\ = \int_{\delta/N(0)}^{T_2}\int_{T_1}^{-\delta/N(0)} \Big\langle \chi\Big(\frac{x}{c|t|}\Big)u^3(t), P_k^2\Big(\frac{e^{i(\tau-t)\sqrt{-\Delta}}}{\sqrt{-\Delta}}\chi\Big(\frac{\cdot}{c|\tau|}\Big)u^3(\tau)\Big)(x)\Big\rangle_{L^2} dt\, d\tau. \quad (4\text{-}68)\end{aligned}$$

Here we perform an argument similar to our use of the sharp Huygens principle in the proof of (4-22). The kernel of $P_k^2 e^{i(t-\tau)\sqrt{-\Delta}}(\sqrt{-\Delta})^{-1}$ is given by

$$K_k(x) = K_k(|x|) = c\int_0^{2\pi}\int e^{i|x|\rho\cos\theta} e^{i(\tau-t)\rho}\rho^{-1}\phi\Big(\frac{\rho}{2^k}\Big)\rho^2\, d\rho\, \sin\theta\, d\theta, \quad (4\text{-}69)$$

where the integrand is written in polar coordinates on $\mathbb{R}^3$, where $\rho = |\xi|$. The function $\phi(\cdot/2^k)$ above is the Fourier multiplier for the Littlewood–Paley projection, $P_k$, and its support is contained in $\rho \in [2^{k-1}, 2^{k+1}]$. Integration by parts $L \in \mathbb{N}$ times in $\rho$ gives the estimates

$$|K_k(x-y)| \lesssim_L \frac{2^{2k}}{\big\langle 2^k\big|(\tau-t) - |x-y|\big|\big\rangle^L}. \quad (4\text{-}70)$$

In (4-68) we have $|x| \le \frac{1}{4}|t|$, $|y| \le \frac{1}{4}|\tau|$, and therefore $|x-y| \le \frac{1}{4}|\tau - t|$. Thus we have

$$(\tau - t) - |x-y| \ge \tfrac{1}{2}|\tau - t|$$

and hence

$$|K_k(x-y)| \lesssim_L \frac{2^{2k}}{\langle 2^k|\tau-t|\rangle^L}. \quad (4\text{-}71)$$

If $2^k \gg N(0)$, we use (4-71) with $L = 5$ to obtain

$$\begin{aligned}\int_{\delta/N(0)}^{T_2}\int_{T_1}^{-\delta/N(0)} \Big\langle \chi\Big(\frac{x}{c|t|}\Big)u^3(t), P_k^2\Big(\frac{e^{i(\tau-t)\sqrt{-\Delta}}}{\sqrt{-\Delta}}\chi\Big(\frac{\cdot}{c|\tau|}\Big)u^3(\tau)\Big)(x)\Big\rangle_{L^2} dt\, d\tau \\ \lesssim \|u\|_{L_t^\infty L_x^3}^6 2^{-3k}N(0)^3 \lesssim 2^{-3k}N(0)^3 \lesssim 2^{-1/2k}N(0)^{1/2}. \quad (4\text{-}72)\end{aligned}$$

If $2^k \lesssim N(0)$, we use the crude estimate $|K_k(x-y)| \lesssim 2^{2k}$ in the $(t,\tau)$ region $|t-\tau| \lesssim 2^{-k}$, and we use (4-71) with $L=3$ in the region where $|\tau - t| \geq 2^{-k}$. We can then conclude that if $2^k \lesssim N(0)$, we have

$$\int_{\delta/N(0)}^{T_2} \int_{T_1}^{-\delta/N(0)} \left\langle \chi\left(\frac{x}{c|t|}\right) u^3(t),\, P_k^2\left(\frac{e^{i(\tau-t)\sqrt{-\Delta}}}{\sqrt{-\Delta}} \chi\left(\frac{\cdot}{c|\tau|}\right) u^3(\tau)\right)(x)\right\rangle_{L^2} dt\, d\tau \lesssim 1. \tag{4-73}$$

Therefore, (4-66), (4-67), (4-72), and (4-73) imply that

$$a_k^2(0) \lesssim a_k(0)\left(\eta^2 \sum_{j \geq k-3} 2^{(k-j)/4} a_j + 2^{-k/6} b_k\right) + \left(\eta^2 \sum_{j \geq k-3} 2^{(k-j)/4} a_j\right)^2 + 2^{-k/3} b_k^2 + \min(2^{-k/2} N(0)^{1/2}, 1). \tag{4-74}$$

Hence we have

$$a_k(0) \lesssim \eta^2 \sum_{j \geq k-3} 2^{(k-j)/4} a_j + 2^{-k/6} b_k + \min(2^{-k/4} N(0)^{1/4}, 1). \tag{4-75}$$

Using the definitions of $\alpha_k(0)$, $\alpha_k$ and (4-55), we get

$$\alpha_k(0) \lesssim \eta^2 \alpha_k + \sum_j 2^{-|j-k|/8} 2^{-j/6} b_j + \sum_j 2^{-|j-k|/8} 2^{-j/6} \min(2^{-j/12} N(0)^{1/4}, 2^{j/6}).$$

Using (4-51) and choosing $\eta$ small enough, we then have

$$\alpha_k(0) \lesssim \sum_j 2^{-|j-k|/8} 2^{-j/6} b_j + \sum_j 2^{-|j-k|/8} 2^{-j/6} c_j, \tag{4-76}$$

where the $c_j := \min(2^{-j/12} N(0)^{1/4}, 2^{j/6})$ satisfy

$$\|\{c_j\}\|_{\ell^2} \lesssim N(0)^{1/6}. \tag{4-77}$$

By Schur's test, using (4-64) and (4-77), we can finally conclude that

$$\|2^{k/6} \alpha_k(0)\|_{\ell^2} \lesssim N(0)^{1/6}, \tag{4-78}$$

as desired. This finishes the proof, since $\alpha_k(0)$ satisfies (4-47). □

## 5. No energy cascade and even more regularity when $N(t) \equiv 1$

In this section we begin by showing that an energy cascade, i.e., the case $\limsup_{t\to\infty} N(t) = 0$, is impossible. This leaves us with the soliton-like critical element $N(t) = 1$ for all $t \geq 0$. We can then reduce this situation to the case of a soliton-like critical element that is global in both time directions with $N(t) \equiv 1$ for all $t \in \mathbb{R}$. Finally, we show that such a solution is in fact uniformly bounded in $\dot{H}^2 \times \dot{H}^1$, which in turn means that $\vec{u}(t)$ satisfies the compactness property in $\dot{H}^1 \times L^2$.

**5A.** ***No energy cascade.*** We can quickly rule out the case of a critical element $\vec{u}(t)$ with scale $N(t)$ satisfying $\limsup_{t\to\infty} N(t) = 0$. We prove the following consequence of Theorem 4.1 and Proposition 2.6.

**Lemma 5.1.** *Let $\vec{u}(t)$ be a solution to* (1-4) *defined on a time interval $I = (T_-, +\infty)$ with $T_- < 0$ and suppose that $\vec{u}(t)$ has the compactness property on $I$ with $N(t) \le 1$ for all $t \in [0,\infty)$. Then $\limsup_{t\to\infty} N(t) = 0$ is impossible unless $\vec{u}(t) \equiv 0$.*

*Proof.* Since $\vec{u}(t)$ satisfies the conditions of (4-1), we see that

$$\limsup_{t\to\infty} \|\vec{u}(t)\|_{\dot{H}^1\times L^2} \lesssim \limsup_{t\to\infty} N(t)^{1/2} = 0. \tag{5-1}$$

By Sobolev embedding and interpolation, we also have

$$\limsup_{t\to\infty} \|u(t)\|_{L^4} \lesssim \limsup_{t\to\infty} \|u(t)\|_{\dot{H}^{3/4}} \lesssim \limsup_{t\to\infty} \|u(t)\|_{\dot{H}^{1/2}}^{1/2} \|u(t)\|_{\dot{H}^1}^{1/2} \lesssim \limsup_{t\to\infty} N(t)^{1/4} = 0. \tag{5-2}$$

Therefore the conserved energy $E(\vec{u}(t))$ is well-defined and (5-1) and (5-2) imply that we must have $E(\vec{u}(t)) = 0$. If $\vec{u}(t)$ solves the defocusing equation, then $E(\vec{u}(t))$ is given by (1-2) and we can directly conclude that we must have $\vec{u}(t) \equiv 0$. If $\vec{u}(t)$ is a solution to the focusing equation, then we use Proposition 2.6 to deduce that $\vec{u}(t) \equiv 0$. □

**5B.** ***Additional regularity for a soliton-like critical element.*** For the case of a soliton-like critical element, i.e., $N(t) \equiv 1$, the rigidity argument in Section 6 will require that the trajectory $\vec{u}(t)$ be precompact in $\dot{H}^1 \times L^2(\mathbb{R}^3)$ rather that just uniformly bounded in this norm, in time. This is not hard to do given our work in the previous sections.

Let $\vec{u}(t)$ be as in Proposition 3.1 and assume that $N(t) = 1$ for all $t \in [0,\infty)$. Then, without loss of generality, we can assume that $I_{\max}(\vec{u}) = \mathbb{R}$ and we have $N(t) \equiv 1$ for all $t \in \mathbb{R}$. Indeed, let $t_n \to \infty$ be any sequence. Since $\vec{u}(t)$ has the compactness property on $(T_-(\vec{u}), \infty)$ we can find a subsequence, still denoted by $t_n$ so that $\vec{u}(t_n) \to \vec{u}_\infty$ in $\dot{H}^{1/2} \times \dot{H}^{-1/2}$. Then, using the perturbation theory, one can readily check that the solution $\vec{u}_\infty(t)$ with initial data $\vec{u}_\infty(0) = \vec{u}_\infty$ is global in time and has the compactness property on $\mathbb{R}$ with $N(t) = 1$ for all $t \in \mathbb{R}$.

We can now establish the following proposition.

**Proposition 5.2.** *Let $\vec{u}(t)$ be the critical element and assume further that $\vec{u}(t)$ is soliton-like, i.e., $\vec{u}(t)$ is defined globally in time and $N(t) \equiv 1$. Then the trajectory*

$$K := \{\vec{u}(t) \mid t \in \mathbb{R}\} \tag{5-3}$$

*is precompact in $(\dot{H}^{1/2} \times \dot{H}^{-1/2}) \cap (\dot{H}^1 \times L^2)(\mathbb{R}^3)$.*

*Proof.* We prove that in fact we have a uniform-in-time bound on the $\dot{H}^2 \times \dot{H}^1$-norm of $\vec{u}(t)$. We only provide a sketch of this fact, as the proof is nearly identical to the proof of Theorem 4.1. The precompactness of $\{\vec{u}(t) \mid t \in \mathbb{R}\}$ in $\dot{H}^1 \times L^2$ then follows from its precompactness in $\dot{H}^{1/2} \times \dot{H}^{-1/2}$ and interpolation, as we have

$$\|\vec{u}(t)\|_{\dot{H}^1\times L^2} \lesssim \|\vec{u}(t)\|_{\dot{H}^{1/2}\times\dot{H}^{-1/2}}^{2/3} \|\vec{u}(t)\|_{\dot{H}^2\times\dot{H}^1}^{1/3}.$$

First note that by Theorem 4.1, we have

$$\|\vec{u}(t)\|_{\dot{H}^1\times L^2} \lesssim 1. \tag{5-4}$$

**Claim 5.3.** *There exists a $\delta > 0$ such that for all $t_0 \in R$ and for $J := (t_0 - \delta, t_0 + \delta)$, we have*

$$\|u\|_{L_t^2 L_x^\infty(J\times\mathbb{R}^3)} \lesssim 1. \tag{5-5}$$

**Remark 11.** In (5-5) we make use of the endpoint $L_t^2 L_x^\infty$ Strichartz estimate, which is valid in the radial setting; see [Klainerman and Machedon 1993]. However, this use of the endpoint is for convenience only, as it will allow for an upgrade of the uniform bound in $\dot{H}^1 \times L^2$ directly to a uniform bound in $\dot{H}^2 \times \dot{H}^1$. This implies that the trajectory is precompact in $\dot{H}^1 \times L^2$ using the precompactness in $\dot{H}^{1/2} \times \dot{H}^{-1/2}$ and interpolating with the $\dot{H}^2 \times \dot{H}^1$ bound. As we are only interested in proving the compactness property in $\dot{H}^1 \times L^2$, it would also suffice to prove a uniform bound in $\dot{H}^{1+\varepsilon} \times \dot{H}^\varepsilon$, and for this estimate we would not need the endpoint Strichartz estimate.

*Proof of Claim 5.3.* First we note that it suffices to prove the claim for $t_0 = 0$. We apply the endpoint Strichartz estimates, which are valid in the radial setting. Indeed, denote by $Z(J)$ the space $Z(J) := L_t^\infty(J; \dot{H}^1 \times L^2) \cap L_t^2(J; L_x^\infty)$. Then we have

$$\begin{aligned}\|u\|_{Z(J)} &\lesssim \|\vec{u}(0)\|_{\dot{H}^1\times L^2} + \|u^3\|_{L_t^1(J;L_x^2)} \lesssim \|\vec{u}(0)\|_{\dot{H}^1\times L^2} + \|u\|^3_{L_t^3(J;L_x^6)} \\ &\lesssim \|\vec{u}(0)\|_{\dot{H}^1\times L^2} + \delta\|u\|^3_{L_t^\infty(J;\dot{H}_x^1)} \lesssim \|\vec{u}(0)\|_{\dot{H}^1\times L^2} + \delta\|u\|_{Z(J)},\end{aligned} \tag{5-6}$$

where we remark that we have used the Sobolev inequality and the length of $J$ in the third inequality above, and nothing else. In the last inequality we have used (5-4). Choosing $\delta = \delta(\|\vec{u}(0)\|_{\dot{H}^1\times L^2}) > 0$ small enough completes the proof. Note that here it is important that the constant in (5-4) is uniform in $t_0 \in I$. □

The proof of Proposition 5.2 now proceeds exactly as in the proof of Theorem 4.1, except here we seek an $\dot{H}^2$ bound. We give a brief sketch. Let $\vec{v}(t)$ be defined as in (4-6), and $Q_M$ as in (4-10). We prove that

$$\langle Q_M v(t_0), Q_M v(t_0)\rangle_{\dot{H}^2} \lesssim 1 \tag{5-7}$$

for all $M \geq M_0$ with a constant that is uniform in $M$ and in $t_0 \in \mathbb{R}$. Extracting weak limits using Lemma 3.2 as in the proof of Theorem 4.1, we note that it will suffice to prove the following estimate for the "double Duhamel" term:

$$\left|\left\langle Q_M\left(\int_{T_1}^0 e^{it\sqrt{-\Delta}}\nabla(u^3)(t)\,dt\right), Q_M\left(\int_0^{T_2} e^{i\tau\sqrt{-\Delta}}\nabla(u^3)(\tau)\,d\tau\right)\right\rangle_{L^2}\right| \lesssim 1, \tag{5-8}$$

where $T_1 < 0$ and $T_2 > 0$ and the constant above is uniform in such $T_1, T_2$. Note also that above we have set $t_0 = 0$, as again this case will be sufficient.

By (5-5), we see that for $\delta > 0$ as in Claim 5.3, we have

$$\left\|Q_M\left(\int_0^\delta e^{i\tau\sqrt{-\Delta}}\nabla(u^3)(\tau)\,d\tau\right)\right\|_{L^2} \lesssim \int_0^\delta \|\nabla(u^3)\|_{L^2} \lesssim \|\nabla u\|_{L_t^\infty L^2}\|u\|^2_{L_t^2([0,\delta);L_x^\infty)} \lesssim 1. \tag{5-9}$$

Next, by the radial Sobolev embedding, $\| |x|^{3/4} u \|_{L^\infty_x(\mathbb{R}^3)} \lesssim \|u\|_{\dot H^{3/4}(\mathbb{R}^3)}$, we have

$$\left\| (1-\chi)\left(\frac{x}{c|t|}\right) \nabla u^3(t) \right\|_{L^2} \lesssim \frac{1}{c^{3/2}|t|^{3/2}} \|\nabla u\|_{L^\infty_t L^2(\mathbb{R}^3)} \|u(t)\|^2_{\dot H^{3/4}} \lesssim |t|^{-3/2}, \tag{5-10}$$

where $\chi \in C_0^\infty(\mathbb{R}^3)$, radial, satisfies $\chi(x) = 1$ for $|x| \le 1$ and $\chi(x) = 0$ for $|x| \ge 2$, and $c = \frac{1}{4}$. Therefore we have

$$\left\| Q_M \left( \int_\delta^{T_2} e^{i\tau\sqrt{-\Delta}} \left( 1 - \chi\left(\frac{\cdot}{c|\tau|}\right)\right) \nabla(u^3)(\tau)\, d\tau \right) \right\|_{L^2} \lesssim \int_\delta^\infty \left\| \left( 1 - \chi\left(\frac{\cdot}{c|\tau|}\right)\right) \nabla(u^3)(\tau) \right\|_{L^2} \lesssim \delta^{-1/2}. \tag{5-11}$$

Next, using the sharp Huygens principle exactly as in the proof of (4-22), the term

$$\begin{aligned}\Big\langle Q_M \Big( e^{it\sqrt{-\Delta}} \chi\Big(\frac{\cdot}{c|t|}\Big) \nabla(u^3)(t)\, dt \Big), Q_M \Big( e^{i\tau\sqrt{-\Delta}} \chi\Big(\frac{\cdot}{c|\tau|}\Big) \nabla(u^3)(\tau)\, d\tau \Big) \Big\rangle \\ = \Big\langle Q_M \Big( \chi\Big(\frac{\cdot}{c|t|}\Big) \nabla(u^3)(t)\, dt \Big), Q_M \Big( e^{i(\tau-t)\sqrt{-\Delta}} \chi\Big(\frac{\cdot}{c|\tau|}\Big) \nabla(u^3)(\tau)\, d\tau \Big) \Big\rangle\end{aligned} \tag{5-12}$$

is identically 0 for $t < -\delta$ and $\tau > \delta$. With (5-9), (5-11), and (5-12) playing the roles of (4-14), (4-18), and (4-22), the proof now proceeds exactly as the proof of Theorem 4.1. We omit the details. □

## 6. Rigidity via a virial identity

In this section we complete the rigidity argument by proving that a soliton-like critical element (i.e., $N(t) \equiv 1$) cannot exist. Indeed, we prove the following proposition:

**Proposition 6.1.** *Let $\vec u(t) \in (\dot H^{1/2} \times \dot H^{-1/2}) \cap (\dot H^1 \times L^2)(\mathbb{R}^3)$ be a global-in-time solution to* (1-4) *such that the trajectory*

$$K := \{\vec u(t) \mid t \in \mathbb{R}\} \tag{6-1}$$

*is precompact in $(\dot H^{1/2} \times \dot H^{-1/2}) \cap (\dot H^1 \times L^2)(\mathbb{R}^3)$. Then $u(t) \equiv 0$.*

The proposition will follow from a simple argument based on the following virial identity. We will fix a smooth radial cutoff function $\chi \in C_0^\infty(\mathbb{R}^3)$ such that $\chi(r) \equiv 1$ for $0 \le r \le 1$, $\operatorname{supp} \chi \subset [0, 2]$, and $|\chi'(r)| \le C$ for all $r > 0$. For each fixed $R > 0$, we will denote by $\chi_R$ the rescaling

$$\chi_R(r) := \chi(r/R). \tag{6-2}$$

**Lemma 6.2** (virial identity). *Let $\vec u(t) \in \dot H^1 \times L^2(\mathbb{R}^3)$ be a solution to* (1-4). *Then for every $R > 0$,*

$$\begin{aligned}\frac{d}{dt}\langle u_t \mid \chi_R(u + r u_r)\rangle = -E(\vec u)(t) + \int_0^\infty (1-\chi_R)\Big(\frac12 u_t^2 + \frac12 u_r^2 \pm \frac14 u^4\Big) r^2\, dr \\ - \int_0^\infty \Big(\frac12 u_t^2 + \frac12 u_r^2 \pm \frac14 u^4\Big) r \chi_R' r^2\, dr - \int_0^\infty u u_r r \chi_R' r\, dr,\end{aligned} \tag{6-3}$$

*where the bracket $\langle f \mid g\rangle$ is the radial $L^2(\mathbb{R}^3)$ inner product*

$$\langle f \mid g\rangle := \int_0^\infty f(r)g(r)r^2\,dr.$$

*Proof.* The proof follows from (1-4) and integration by parts. □

**Remark 12.** In general, for a semilinear equation of the form

$$u_{tt} - \Delta u = \pm|u|^{p-1}u,$$

one has the formal virial identity

$$\frac{d}{dt}\langle u_t \mid u + x\cdot\nabla u\rangle = -E(u) \pm \left(\frac{p-3}{p+1}\right)\|u\|_{L^{p+1}}^{p+1}.$$

Note that the right side can be bounded from above by a negative constant times the *conserved* energy in the case $1+\sqrt{2} < p \le 3$, yielding a monotone quantity. But in the case $3 < p < 5$, the right side cannot be controlled by the conserved energy in the case of the focusing equation. However, for the range $3 < p < 5$, a different rigidity argument is available, based on the "channels of energy" method developed by Duyckaerts, Kenig, and Merle [2013; 2014]. For an implementation of this strategy for the range $p \in (3, 5)$, see [Shen 2012].

We also note that the virial identities and the argument in this section also readily extend to the nonradial setting.

The proof of Proposition 6.1 will now follow by applying the above lemma to our precompact trajectory $\vec{u}(t)$ in order to show that the energy must be nonpositive. One concludes the proof by noting that a solution to the defocusing equation with nonpositive energy must be identically zero. In the case of the focusing equation, we recall Proposition 2.6, which says that a solution with nonpositive energy must either be identically zero or blow up in both time directions, and the latter is impossible under the hypothesis of Proposition 6.1.

*Proof of Proposition 6.1.* Fix $\eta > 0$. We will show that for $\vec{u}(t)$ as in Proposition 6.1, we have

$$\mathscr{E}(\vec{u}) \le C\eta \tag{6-4}$$

for a fixed constant $C$ which is independent of $\eta$. First, note that since $\{\vec{u}(t) \mid t \in \mathbb{R}\}$ is precompact in $(\dot{H}^{1/2}\times\dot{H}^{-1/2})\cap(\dot{H}^1\times L^2)(\mathbb{R}^3)$, we can find $R_0 = R_0(\eta)$ such that for all $R \ge R_0$ and for all $t \in \mathbb{R}$, we have

$$\int_R^\infty (u_t^2(t) + u_r^2(t))r^2\,dr \le \eta. \tag{6-5}$$

Moreover, due to the embeddings $\dot{H}^{1/2}\cap\dot{H}^1 \hookrightarrow \dot{H}^{3/4} \hookrightarrow L^4$, we can choose $R_0(\eta)$ large enough that we also have

$$\int_R^\infty u^4(t)r^2\,dr \le \eta \tag{6-6}$$

for all $R \geq R_0$ and for all $t \in \mathbb{R}$. Finally, we note that for any $R > 0$ and for any smooth radial function in $\dot{H}^1(\mathbb{R}^3)$, we have

$$\int_R^\infty f^2(r)\,dr + Rf^2(R) = -\int_R^\infty f_r(r) f(r) r\,dr,$$

which can be obtained by integrating by parts. This implies that

$$\int_R^\infty f^2(r)\,dr \leq \int_R^\infty f_r^2(r) r^2\,dr.$$

Therefore, for our precompact trajectory $\vec{u}(t)$, we can use (6-5) to obtain

$$\int_R^\infty u^2(t,r)\,dr \leq \eta \tag{6-7}$$

for all $R \geq R_0(\eta)$ and for all $t \in \mathbb{R}$. Letting $R \geq R_0(\eta)$, we can apply these estimates to the last three terms on the right side of (6-3):

$$\left\langle \int_0^\infty (1-\chi_R)\left(\frac{1}{2}u_t^2 + \frac{1}{2}u_r^2 \pm \frac{1}{4}u^4\right) r^2\,dr \right\rangle \leq C\eta,$$

$$\left\langle \int_0^\infty \left(\frac{1}{2}u_t^2 + \frac{1}{2}u_r^2 \pm \frac{1}{4}u^4\right) r\chi_R' r^2\,dr \right\rangle \leq C\int_R^{2R} \left(\frac{1}{2}u_t^2 + \frac{1}{2}u_r^2 + \frac{1}{4}u^4\right) r^2\,dr \leq C\eta,$$

$$\left\langle \int_0^\infty u u_r r \chi_R' r\,dr \right\rangle \leq \left(\int_R^{2R} u_r^2 r^2\,dr\right)^{1/2} \left(\int_R^{2R} u^2\,dr\right)^{1/2} \leq C\eta.$$

Inserting the above estimates into (6-3) and averaging in time from 0 to $T$, we obtain the estimate

$$E(\vec{u}) \leq C\eta + \frac{1}{T}\left|\langle u_t(T) \mid \chi_R(u(T) + r u_r(T))\rangle\right| + \frac{1}{T}\left|\langle u_t(0) \mid \chi_R(u(0) + r u_r(0))\rangle\right|. \tag{6-8}$$

Now, set $R = T$ above with $T$ large enough that $T \gg R_0(\eta)$. We have

$$\begin{aligned} E(\vec{u}) \leq C\eta + C\frac{1}{T}\int_0^T |u_t(T)|\,|u(T)| r^2\,dr + C\frac{1}{T}\int_0^T |u_t(0)|\,|u(0)| r^2\,dr \\ + C\frac{1}{T}\int_0^T |u_t(T)|\,|u_r(T)| r^3\,dr + C\frac{1}{T}\int_0^T |u_t(0)|\,|u_r(0)| r^3\,dr. \end{aligned} \tag{6-9}$$

We estimate the second and third terms on the right side of (6-9) by

$$\begin{aligned} \frac{1}{T}\int_0^T |u_t|\,|u| r^2\,dr &\leq \frac{1}{T}\left(\int_0^T u_t^2 r^2\,dr\right)^{1/2} \left(\int_0^T |u|^3 r^2\,dr\right)^{1/3} \left(\int_0^T r^2\,dr\right)^{1/6} \\ &\leq C\frac{1}{T^{1/2}} \|u_t\|_{L^2} \|u\|_{\dot{H}^{1/2}} \to 0 \quad \text{as } T \to \infty, \end{aligned}$$

where in the last line we have used the embedding $\dot{H}^{1/2} \hookrightarrow L^3$, the fact that the critical element $\vec{u}(t)$ satisfies $\sup_{t\in\mathbb{R}} \|u(t)\|_{\dot{H}^{\frac{1}{2}}} \lesssim 1$, and $\sup_{t\in\mathbb{R}} \|u_t\|_{L^2} \lesssim 1$. To estimate the fourth and fifth terms in (6-9), we note that for $T \gg R(\eta)$ we have

$$\begin{aligned}\frac{1}{T}\int_0^T |u_t||u_r|r^3\,dr &\le \frac{1}{T}\int_0^{R(\eta)} |u_t||u_r|r^3\,dr + \frac{1}{T}\int_{R(\eta)}^T |u_t||u_r|r^3\,dr \\ &\le \frac{R(\eta)}{T}\|u_t\|_{L^2}\|u\|_{\dot H^1} + \left(\int_{R(\eta)}^T u_t^2 r^2\,dr\right)^{1/2}\left(\int_{R(\eta)}^T u_r^2 r^2\,dr\right)^{1/2} \\ &= \eta + O(T^{-1}) \quad \text{as } T \to \infty.\end{aligned}$$

Thus, letting $T \to \infty$ in (6-9), we obtain

$$E(\vec u) \le C\eta,$$

as desired. Since this holds for all $\eta > 0$, we can conclude that

$$E(\vec u) \le 0. \tag{6-10}$$

In the case that $\vec u(t)$ is a solution to the defocusing equation, we are done, as we can conclude from (6-10) that $\vec u(t) \equiv 0$. In the case that $\vec u(t)$ is a solution to the focusing equation, we note that (6-10) together with Proposition 2.6 implies that either $\vec u(t) \equiv 0$ or $\vec u(t)$ blows up in finite time in both time directions. However, the latter case is impossible, as we have assumed that $\vec u(t)$ is global in time. This completes the proof of Proposition 6.1. □

## 7. Proof of Theorem 1.1

We provide a brief summary of the proof of Theorem 1.1, which is now complete. We argue by contradiction. If Theorem 1.1 were false, we could, by Proposition 3.1, find a critical element, i.e., a *nonzero* solution $\vec u(t)$ to (1-4) with the compactness property in $\dot H^{1/2} \times \dot H^{-1/2}$ on an open interval $I \ni 0$ with scale $N(t)$. By the remarks following the statement of Proposition 3.1, we can reduce to the case $I = (T_-, \infty)$ and $N(t) \le 1$ for $t \in [0, \infty)$. By (4-1), we then have

$$\|\vec u(t)\|_{\dot H^1 \times L^2} \lesssim N(t)^{1/2} \quad \text{for } t \in [0, \infty).$$

Then, since we are assuming $\vec u(t)$ is nonzero, by Section 5A, we can conclude that $N(t) \equiv 1$ for all $t \in [0, \infty)$. We can then ensure that $\vec u(t)$ is global in time for all $t \in \mathbb{R}$, and by Proposition 5.2, we know that $\vec u(t)$ has a precompact trajectory in $\dot H^{1/2} \times \dot H^{-1/2} \cap \dot H^1 \times L^2$. But then Proposition 6.1 shows that $\vec u(t) \equiv 0$, which is a contradiction.

## References

[Bahouri and Chemin 2006] H. Bahouri and J.-Y. Chemin, "On global well-posedness for defocusing cubic wave equation", *Int. Math. Res. Not.* **2006** (2006), Art. ID #54873. MR 2008b:35178 Zbl 1123.35029

[Bahouri and Gérard 1999] H. Bahouri and P. Gérard, "High frequency approximation of solutions to critical nonlinear wave equations", *Amer. J. Math.* **121**:1 (1999), 131–175. MR 2000i:35123 Zbl 0919.35089

[Bizoń and Zenginoğlu 2009] P. Bizoń and A. Zenginoğlu, "Universality of global dynamics for the cubic wave equation", *Nonlinearity* **22**:10 (2009), 2473–2485. MR 2010h:35030 Zbl 1180.35129

[Bizoń et al. 2010] P. Bizoń, P. Breitenlohner, D. Maison, and A. Wasserman, "Self-similar solutions of the cubic wave equation", *Nonlinearity* **23**:2 (2010), 225–236. MR 2011c:35352 Zbl 1184.35209

[Bulut 2012a] A. Bulut, "The defocusing cubic nonlinear wave equation in the energy-supercritical regime", pp. 1–11 in *Recent advances in harmonic analysis and partial differential equations* (Statesboro, GA and Baltimore, MD, 2011), edited by A. R. Nahmod et al., Contemporary Mathematics **581**, American Mathematical Society, Providence, RI, 2012. MR 3013050

[Bulut 2012b] A. Bulut, "Global well-posedness and scattering for the defocusing energy-supercritical cubic nonlinear wave equation", *J. Funct. Anal.* **263**:6 (2012), 1609–1660. MR 2948225 Zbl 1258.35147

[Donninger and Krieger 2013] R. Donninger and J. Krieger, "Nonscattering solutions and blowup at infinity for the critical wave equation", *Math. Ann.* **357**:1 (2013), 89–163. MR 3084344 Zbl 1280.35135 arXiv 1201.3258

[Donninger and Schörkhuber 2012] R. Donninger and B. Schörkhuber, "Stable self-similar blow up for energy subcritical wave equations", *Dyn. Partial Differ. Equ.* **9**:1 (2012), 63–87. MR 2909934 Zbl 1259.35044

[Donninger et al. 2014] R. Donninger, M. Huang, J. Krieger, and W. Schlag, "Exotic blowup solutions for the $u^5$ focusing wave equation in $\mathbb{R}^3$", *Michigan Math. J.* **63**:3 (2014), 451–501. MR 3255688 Zbl 06362112 arXiv 1212.4718

[Duyckaerts and Merle 2008] T. Duyckaerts and F. Merle, "Dynamics of threshold solutions for energy-critical wave equation", *Int. Math. Res. Pap.* **2008** (2008), Art. ID #rpn002. MR 2009m:35329 Zbl 1159.35043

[Duyckaerts et al. 2011] T. Duyckaerts, C. E. Kenig, and F. Merle, "Universality of blow-up profile for small radial type II blow-up solutions of the energy-critical wave equation", *J. Eur. Math. Soc.* **13**:3 (2011), 533–599. MR 2012e:35160 Zbl 1230.35067

[Duyckaerts et al. 2012a] T. Duyckaerts, C. E. Kenig, and F. Merle, "Profiles of bounded radial solutions of the focusing, energy-critical wave equation", *Geom. Funct. Anal.* **22**:3 (2012), 639–698. MR 2972605 Zbl 1258.35148

[Duyckaerts et al. 2012b] T. Duyckaerts, C. E. Kenig, and F. Merle, "Universality of the blow-up profile for small type II blow-up solutions of the energy-critical wave equation: the nonradial case", *J. Eur. Math. Soc.* **14**:5 (2012), 1389–1454. MR 2966655 Zbl 1282.35088

[Duyckaerts et al. 2013] T. Duyckaerts, C. E. Kenig, and F. Merle, "Classification of radial solutions of the focusing, energy-critical wave equation", *Camb. J. Math.* **1**:1 (2013), 75–144. MR 3272053 Zbl 06266536

[Duyckaerts et al. 2014] T. Duyckaerts, C. E. Kenig, and F. Merle, "Scattering for radial, bounded solutions of focusing supercritical wave equations", *Int. Math. Res. Not.* **2014**:1 (2014), 224–258. MR 3158532 Zbl 06340330 arXiv 1208.2158

[Gallagher and Planchon 2003] I. Gallagher and F. Planchon, "On global solutions to a defocusing semi-linear wave equation", *Rev. Mat. Iberoamericana* **19**:1 (2003), 161–177. MR 2004k:35265 Zbl 1036.35142

[Grillakis 1990] M. G. Grillakis, "Regularity and asymptotic behaviour of the wave equation with a critical nonlinearity", *Ann. of Math.* (2) **132**:3 (1990), 485–509. MR 92c:35080 Zbl 0736.35067

[John 1979] F. John, "Blow-up of solutions of nonlinear wave equations in three space dimensions", *Proc. Nat. Acad. Sci. USA* **76**:4 (1979), 1559–1560. MR 80d:35093 Zbl 0397.35046

[Jörgens 1961] K. Jörgens, "Das Anfangswertproblem im Grossen für eine Klasse nichtlinearer Wellengleichungen", *Math. Z.* **77** (1961), 295–308. MR 24 #A323 Zbl 0111.09105

[Keel and Tao 1998] M. Keel and T. Tao, "Endpoint Strichartz estimates", *Amer. J. Math.* **120**:5 (1998), 955–980. MR 2000d:35018 Zbl 0922.35028

[Kenig and Merle 2006] C. E. Kenig and F. Merle, "Global well-posedness, scattering and blow-up for the energy-critical, focusing, non-linear Schrödinger equation in the radial case", *Invent. Math.* **166**:3 (2006), 645–675. MR 2007g:35232 Zbl 1115.35125

[Kenig and Merle 2008] C. E. Kenig and F. Merle, "Global well-posedness, scattering and blow-up for the energy-critical focusing non-linear wave equation", *Acta Math.* **201**:2 (2008), 147–212. MR 2011a:35344 Zbl 1183.35202

[Kenig and Merle 2010] C. E. Kenig and F. Merle, "Scattering for $\dot{H}^{1/2}$ bounded solutions to the cubic, defocusing NLS in 3 dimensions", *Trans. Amer. Math. Soc.* **362**:4 (2010), 1937–1962. MR 2011b:35486 Zbl 1188.35180

[Kenig and Merle 2011] C. E. Kenig and F. Merle, "Nondispersive radial solutions to energy supercritical non-linear wave equations, with applications", *Amer. J. Math.* **133**:4 (2011), 1029–1065. MR 2012i:35244 Zbl 1241.35136

[Kenig et al. 2000] C. E. Kenig, G. Ponce, and L. Vega, "Global well-posedness for semi-linear wave equations", *Comm. Partial Differential Equations* **25**:9-10 (2000), 1741–1752. MR 2001h:35128 Zbl 0961.35092

[Killip and Vişan 2010a] R. Killip and M. Vişan, "Energy-supercritical NLS: critical $\dot{H}^s$-bounds imply scattering", *Comm. Partial Differential Equations* **35**:6 (2010), 945–987. MR 2011m:35357 Zbl 1200.35289

[Killip and Vişan 2010b] R. Killip and M. Vişan, "The focusing energy-critical nonlinear Schrödinger equation in dimensions five and higher", *Amer. J. Math.* **132**:2 (2010), 361–424. MR 2011e:35357 Zbl 1208.35138

[Killip and Vişan 2011] R. Killip and M. Vişan, "The defocusing energy-supercritical nonlinear wave equation in three space dimensions", *Trans. Amer. Math. Soc.* **363**:7 (2011), 3893–3934. MR 2012e:35161 Zbl 1230.35068

[Killip and Vişan 2013] R. Killip and M. Vişan, "Nonlinear Schrödinger equations at critical regularity", pp. 325–442 in *Evolution equations* (Zürich, 2008), edited by D. Ellwood et al., Clay Mathematics Proceedings **17**, American Mathematical Society, Providence, RI, 2013. MR 3098643 Zbl 1298.35195

[Killip et al. 2009] R. Killip, T. Tao, and M. Vişan, "The cubic nonlinear Schrödinger equation in two dimensions with radial data", *J. Eur. Math. Soc.* **11**:6 (2009), 1203–1258. MR 2010m:35487 Zbl 1187.35237

[Killip et al. 2012] R. Killip, B. Stovall, and M. Vişan, "Scattering for the cubic Klein–Gordon equation in two space dimensions", *Trans. Amer. Math. Soc.* **364**:3 (2012), 1571–1631. MR 2869186 Zbl 1252.35203 arXiv 1008.2712

[Klainerman and Machedon 1993] S. Klainerman and M. Machedon, "Space-time estimates for null forms and the local existence theorem", *Comm. Pure Appl. Math.* **46**:9 (1993), 1221–1268. MR 94h:35137 Zbl 0803.35095

[Krieger and Schlag 2014a] J. Krieger and W. Schlag, "Full range of blow up exponents for the quintic wave equation in three dimensions", *J. Math. Pures Appl.* (9) **101**:6 (2014), 873–900. MR 3205646 Zbl 06298840 arXiv 1212.3795

[Krieger and Schlag 2014b] J. Krieger and W. Schlag, "Large global solutions for energy supercritical nonlinear wave equations on $\mathbb{R}^{3+1}$", preprint, 2014. arXiv 1403.2913

[Krieger et al. 2009] J. Krieger, W. Schlag, and D. Tataru, "Slow blow-up solutions for the $H^1(\mathbb{R}^3)$ critical focusing semilinear wave equation", *Duke Math. J.* **147**:1 (2009), 1–53. MR 2010h:58045 Zbl 1170.35066

[Krieger et al. 2013a] J. Krieger, K. Nakanishi, and W. Schlag, "Global dynamics away from the ground state for the energy-critical nonlinear wave equation", *Amer. J. Math.* **135**:4 (2013), 935–965. MR 3086065 Zbl 06203653

[Krieger et al. 2013b] J. Krieger, K. Nakanishi, and W. Schlag, "Global dynamics of the nonradial energy-critical wave equation above the ground state energy", *Discrete Contin. Dyn. Syst.* **33**:6 (2013), 2423–2450. MR 3007693 Zbl 1272.35153

[Krieger et al. 2014] J. Krieger, K. Nakanishi, and W. Schlag, "Threshold phenomenon for the quintic wave equation in three dimensions", *Comm. Math. Phys.* **327**:1 (2014), 309–332. MR 3177940 Zbl 1297.35139

[Lindblad and Sogge 1995] H. Lindblad and C. D. Sogge, "On existence and scattering with minimal regularity for semilinear wave equations", *J. Funct. Anal.* **130**:2 (1995), 357–426. MR 96i:35087 Zbl 0846.35085

[Merle and Zaag 2003] F. Merle and H. Zaag, "Determination of the blow-up rate for the semilinear wave equation", *Amer. J. Math.* **125**:5 (2003), 1147–1164. MR 2004g:35163 Zbl 1052.35043

[Merle and Zaag 2005] F. Merle and H. Zaag, "Determination of the blow-up rate for a critical semilinear wave equation", *Math. Ann.* **331**:2 (2005), 395–416. MR 2005k:35286 Zbl 1136.35055

[Roy 2009] T. Roy, "Adapted linear-nonlinear decomposition and global well-posedness for solutions to the defocusing cubic wave equation on $\mathbb{R}^3$", *Discrete Contin. Dyn. Syst.* **24**:4 (2009), 1307–1323. MR 2010j:35327 Zbl 1178.35267

[Schaeffer 1985] J. Schaeffer, "The equation $u_{tt} - \Delta u = |u|^p$ for the critical value of $p$", *Proc. Roy. Soc. Edinburgh Sect. A* **101**:1-2 (1985), 31–44. MR 87g:35159 Zbl 0592.35080

[Shen 2012] R. Shen, "On the energy subcritical, non-linear wave equation with radial data for $p \in (3, 5)$", preprint, 2012. arXiv 1208.2108

[Sogge 2008] C. D. Sogge, *Lectures on non-linear wave equations*, 2nd ed., International Press, Boston, 2008. MR 2009i:35213 Zbl 1165.35001

[Struwe 1988] M. Struwe, "Globally regular solutions to the $u^5$ Klein–Gordon equation", *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* (4) **15**:3 (1988), 495–513. MR 90j:35142 Zbl 0728.35072

[Tao 2001] T. Tao, "Global regularity of wave maps, I: Small critical Sobolev norm in high dimension", *Int. Math. Res. Not.* **2001**:6 (2001), 299–328. MR 2001m:35200 Zbl 0983.35080

[Tao 2006] T. Tao, *Nonlinear dispersive equations: local and global analysis*, CBMS Regional Conference Series in Mathematics **106**, American Mathematical Society, Providence, RI, 2006. MR 2008i:35211 Zbl 1106.35001

[Tao 2007] T. Tao, "A (concentration-)compact attractor for high-dimensional non-linear Schrödinger equations", *Dyn. Partial Differ. Equ.* **4**:1 (2007), 1–53. MR 2007k:35479 Zbl 1142.35088

[Tao et al. 2007] T. Tao, M. Vişan, and X. Zhang, "Global well-posedness and scattering for the defocusing mass-critical nonlinear Schrödinger equation for radial data in high dimensions", *Duke Math. J.* **140**:1 (2007), 165–202. MR 2010a:35249 Zbl 1187.35246

[Tao et al. 2008] T. Tao, M. Vişan, and X. Zhang, "Minimal-mass blowup solutions of the mass-critical NLS", *Forum Math.* **20**:5 (2008), 881–919. MR 2009m:35495 Zbl 1154.35085

Received 20 May 2014. Revised 12 Oct 2014. Accepted 26 Nov 2014.

BENJAMIN DODSON: bdodson4@jhu.edu
*Department of Mathematics, Johns Hopkins University, 404 Krieger Hall, 3400 N. Charles Street, Baltimore, CA 21218, United States*

ANDREW LAWRIE: alawrie@math.berkeley.edu
*Department of Mathematics, University of California, Berkeley, 970 Evans Hall #3840, Berkeley, CA 94720, United States*

**mathematical sciences publishers** msp

ANALYSIS AND PDE
Vol. 8, No. 2, 2015
dx.doi.org/10.2140/apde.2015.8.499

# COUNTEREXAMPLES TO THE WELL POSEDNESS OF THE CAUCHY PROBLEM FOR HYPERBOLIC SYSTEMS

FERRUCCIO COLOMBINI AND GUY MÉTIVIER

This paper is concerned with the well-posedness of the Cauchy problem for first order symmetric hyperbolic systems in the sense of Friedrichs. The classical theory says that if the coefficients of the system and if the coefficients of the symmetrizer are Lipschitz continuous, then the Cauchy problem is well posed in $L^2$. When the symmetrizer is log-Lipschitz or when the coefficients are analytic or quasianalytic, the Cauchy problem is well posed in $C^\infty$. We give counterexamples which show that these results are sharp. We discuss both the smoothness of the symmetrizer and of the coefficients.



## 1. Introduction

We consider the well-posedness of the Cauchy problem for first order symmetric hyperbolic systems in the sense of Friedrichs [1954], who proved that if the coefficients of the system *and if* the coefficients of the symmetrizer are Lipschitz continuous, then the Cauchy problem is well posed in $L^2$. This has been extended to hyperbolic systems which admit Lipschitzian microlocal symmetrizers (see [Métivier 2014]).

The main objective of this paper is to discuss the necessity of these smoothness assumptions and to provide new counterexamples to the well-posedness. In the spirit of [Colombini and Spagnolo 1989; Colombini and Nishitani 1999], we make a detailed analysis of systems in space dimension one with coefficients which depend only on time. Even more, we concentrate our analysis on the $2 \times 2$ system

$$Lu := \partial_t u + \begin{pmatrix} a(t) & b(t) \\ c(t) & d(t) \end{pmatrix} \partial_x u = \partial_t u + A(t)u. \tag{1-1}$$

The symbol is assumed to be strongly hyperbolic or uniformly diagonalizable, which means that there is a bounded symmetrizer $S(t)$, with $S^{-1}$ bounded, which is positive definite and such that $S(t)A(t)$ is symmetric. This is equivalent to the condition that there is $\delta > 0$ such that

$$\delta((a-d)^2 + b^2 + c^2) \le \tfrac{1}{4}(a-d)^2 + bc. \tag{1-2}$$

Métivier thanks the Centro E. De Giorgi for their hospitality.
*MSC2010:* 35L50.
*Keywords:* hyperbolic systems, the Cauchy problem, nonsmooth symmetrizers, ill-posedness.

If the symmetrizer $S$ and the coefficients are Lipschitz continuous then the Cauchy problem is well posed in $L^2$. Indeed, in this case, solutions on $[0, T] \times \mathbb{R}$ of $Lu = f$ satisfy

$$\|u(t)\|_{L^2} \le C(\|u(0)\|_{L^2} + \|Lu\|_{L^2}) \quad \text{with} \quad C = C_0 \exp\biggl(\int_0^T |\partial_t S(s)|\, ds\biggr). \tag{1-3}$$

Lipschitz smoothness of the symmetrizer is almost necessary for well-posedness in $L^2$, even for very smooth coefficients:

**Theorem 1.1.** *For each modulus of continuity $\omega$ such that $t^{-1}\omega(t) \to +\infty$ as $t \to 0$, there is a system* (1-1) *with coefficients in $\bigcap_{s>1} G^s([0, T])$, with a symmetrizer satisfying*

$$|S(t) - S(t')| \le C\omega(|t - t'|), \tag{1-4}$$

*such that the Cauchy problem is ill posed in $L^2$ in the sense that there is no constant $C$ such that the estimate* (1-3) *is satisfied.*

Here and below, we denote by $G^s([0, T])$ the Gevrey class of functions of order $s$. They are $C^\infty$ functions $f$ such that, for some constant $C$ which depends on $f$,

$$\|\partial_t^j f\|_{L^\infty} \le C^{j+1}(j!)^s \quad \text{for all } j \in \mathbb{N}.$$

This theorem extends to systems a similar result obtained in [Cicognani and Colombini 2006] for the strictly hyperbolic wave equation

$$\partial_t^2 u - a(t)\partial_x^2 u = f. \tag{1-5}$$

Indeed, there is a close parallel between the energy $|\partial_t u|^2 + a(t)|\partial_x u|^2$ for the wave equation and $(S(t)u, u)$ for the system, and, in this case, the smoothness of $S(t)$ plays a role analogous to the smoothness of $a$. For the wave equation, when $a$ is log-Lipschitz, i.e., admits the modulus of continuity $\omega(t) = t|\ln t|$, the Cauchy problem is well posed in $C^\infty$ with a loss of derivatives proportional to time [Colombini et al. 1979]. In intermediate cases between Lipschitz and log-Lipschitz, that is when $(t|\ln t|)^{-1}\omega(t) \to 0$ and $t^{-1}\omega(t) \to +\infty$, the loss of derivative is effective but is arbitrarily small on any interval [Cicognani and Colombini 2006]. The proof of these results extends immediately to systems (1-1) where the smoothness of the symmetrizer plays the role of the smoothness of the coefficient $a$.

The next result extends to systems the result in [Colombini et al. 1979; Colombini and Spagnolo 1989] showing that the log-Lipschitz smoothness of the symmetrizer is a sharp condition for the well-posedness in $C^\infty$, even for $C^\infty$ coefficients.

**Theorem 1.2.** *For each modulus of continuity $\omega$ satisfying $(t|\ln t|)^{-1}\omega(t) \to +\infty$ as $t \to 0$, there are systems* (1-1) *with $C^\infty$ coefficients, with symmetrizers which satisfy the estimate* (1-4), *such that the Cauchy problem is ill posed in $C^\infty$, meaning that, for all $n$ and all $T > 0$, there is no constant $C$ such that the estimate*

$$\|u\|_{L^2} \le C\|Lu\|_{H^n} \tag{1-6}$$

*is satisfied for all $u \in C_0^\infty([0, T] \times \mathbb{R})$.*

In [Colombini and Nishitani 1999] the question of the well-posedness of the Cauchy problem is considered under the angle of the smoothness of the coefficients alone. In this aspect, the analysis is related to the analysis of the weakly hyperbolic wave equation (1-5) (see [Colombini et al. 1983]). If the coefficients are $C^\infty$, the problem is well posed in all Gevrey classes $G^s$, but the well-posedness in $C^\infty$ is obtained only when the coefficients are analytic or belong to a quasianalytic class. Indeed, the next theorem shows that this is sharp.

**Theorem 1.3.** *There are example of systems* (1-1) *on* $[0, T] \times \mathbb{R}$ *with uniformly hyperbolic symbols and coefficients in the intersection of the Gevrey classes* $\bigcap G^s$ *for* $s > 1$, *admitting continuous symmetrizers, such that the Cauchy problem is ill posed in* $C^\infty$.

This theorem improves the similar result obtained in [Colombini and Nishitani 1999], where the counterexample had coefficients in $\bigcap G^s$ for $s > 2$. The same construction can be used to provide a similar improvement to the known result in [Colombini and Spagnolo 1982] for the wave equation:

**Theorem 1.4.** *There are nonnegative functions* $a \in \bigcap_{s>1} G^s([0, T])$ *such that the Cauchy problem for the weakly hyperbolic wave equation* (1-5) *is ill posed in* $C^\infty$.

The theorems above show that the smoothness of *both* the coefficients *and* the symmetrizer play a role in the well-posedness in $C^\infty$. The next theorem is a kind of interpolation between the two extreme cases of Theorem 1.2 and Theorem 1.3.

**Theorem 1.5.** *For all* $s > 1$ *and* $\mu < 1 - 1/s$, *there are examples of systems* (1-1) *on* $[0, T] \times \mathbb{R}$, *with uniformly hyperbolic symbols, coefficients in the Gevrey classes* $G^s$ *and symmetrizers in the Hölder space* $C^\mu$, *such that the Cauchy problem is ill posed in* $C^\infty$.

This leaves open the question of the well-posedness in $C^\infty$ when the coefficients belong to $G^s$ and the symmetrizer to $C^\mu$ with $\mu + 1/s \geq 1$.

We end this introduction with several remarks about symmetrizers for the $2 \times 2$ system (1-1). For simplicity, we assume that the coefficients are real. Write

$$A(t) = \tfrac{1}{2}\mathrm{tr} A(t)\,\mathrm{Id} + A_1(t).$$

Then $A_1^2 = h\,\mathrm{Id}$ with $h = \frac{1}{4}(a-d)^2 + bc$ satisfying (1-2). In particular,

$$\Sigma(t) = A_1^*(t) A_1(t) + h(t)\,\mathrm{Id}$$

is a symmetrizer for $A$ in the sense that $\Sigma$ and $\Sigma A = \frac{1}{2}(\mathrm{tr} A)\Sigma + hA_1^* + hA_1$ are symmetric. In general, $\Sigma$ is *not* a symmetrizer in the sense of Friedrichs, since it is not uniformly positive definite, unless $h > 0$, which means that the system is strictly hyperbolic. More precisely, $\Sigma \approx h\,\mathrm{Id}$. But $\Sigma$ has the same smoothness as the coefficients of $A$.

On the other hand, since the system is uniformly diagonalizable, there are bounded symmetrizers $\Sigma_1(t)$ which are uniformly positive definite. For instance, $h^{-1}\Sigma$ is a bounded symmetrizer. More generally, writing

$$\tfrac{1}{2}(a-d) = h^{1/2} a_1, \quad b = b_1 h^{1/2}, \quad c = c_1 h^{1/2}, \tag{1-7}$$

one has $a_1^2 + b_1c_1 \geq \delta(a_1^2 + b_1^2 + c_1^2) \geq \delta > 0$ and the symmetrizer is of the form

$$\Sigma_1 = \begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix} \quad \text{with} \quad 2a_1\beta = b_1\alpha - c_1\gamma. \tag{1-8}$$

Therefore, there is a cone of positive symmetrizers of dimension 2. Their smoothness depends on the smoothness of $a_1$, $b_1$, $c_1$, that is, of $h^{-1/2}A_1$. There might be better choices than others. For instance, if the system is symmetric, $\Sigma_1 = \mathrm{Id}$ is a very smooth symmetrizer. Our discussion below concerns the smoothness of both $\Sigma$ and $\Sigma_1$ and their possible interplay.

## 2. The counterexamples

We consider systems of the form

$$LU := \partial_t U + \begin{pmatrix} 0 & a(t) \\ b(t) & 0 \end{pmatrix} \partial_x U \tag{2-1}$$

with $a$ and $b$ real. We always assume that it is uniformly strongly hyperbolic, that is, that $\sigma = a/b > 0$ and $1/\sigma$ are bounded. Our goal is to contradict the estimates (1-3) and (1-6). We contradict the analogous estimates which are obtained by Fourier transform in $x$, and, more precisely, we construct sequences of functions $u_k$, $v_k$ and $f_k$ in $C^\infty([0, T])$, vanishing near $t = 0$, satisfying

$$\partial_t u_k + i h_k a(t) v_k = f_k, \quad \partial_t v_k + i h_k b(t) u_k = 0 \tag{2-2}$$

with $h_k \to +\infty$ and such that

$$\frac{\| f_k \|_{L^2}}{\|(u_k, v_k)\|_{L^2}} \to 0 \quad \text{as } k \to \infty \tag{2-3}$$

in the first case, or, for all $j$ and $l$,

$$\frac{\| h_k^j \partial_t^l f_k \|_{L^2}}{\|(u_k, v_k)\|_{L^2}} \to 0 \quad \text{as } k \to \infty \tag{2-4}$$

in the second case. Moreover, the support of these function is contained in an interval $I_k = [t_k, t'_k]$ with $0 < t_k < t'_k$ and $t'_k \to 0$, showing that the problem is ill posed on any interval $[0, T]$ with $T > 0$.

***Exponentially amplified solutions of the wave equation.*** In this section we review and adapt the construction of [Colombini and Spagnolo 1989]. The key remark is that the function $\underline{w}_\varepsilon(t) = e^{-\varepsilon\phi(t)} \cos t$ satisfies

$$\partial_t^2 \underline{w}_\varepsilon + \underline{\alpha}_\varepsilon \underline{w}_\varepsilon = 0 \tag{2-5}$$

if

$$\phi(t) = \int_0^t (\cos s)^2 \, ds, \quad \underline{\alpha}_\varepsilon(t) = 1 + 2\varepsilon \sin 2t - \varepsilon^2 (\cos t)^4. \tag{2-6}$$

The important property of the $\underline{w}_\varepsilon$ is their exponential decay at $+\infty$. More precisely,

$$e^{\varepsilon t/2} \underline{w}_\varepsilon(t) = e^{\varepsilon \sin(2t)/4} \cos t \quad \text{is } 2\pi\text{-periodic}$$

and

$$\underline{w}_\varepsilon(t+2\pi) = e^{-\varepsilon\pi}\,\underline{w}_\varepsilon(t). \tag{2-7}$$

Next, one symmetrizes and localizes this solution. More precisely, consider $\chi \in C^\infty(\mathbb{R})$, supported in $]-7\pi, 7\pi[$, odd, equal to 1 on $[-6\pi, -2\pi]$ and thus equal to $-1$ on $[2\pi, 6\pi]$, and such that, for all $t$, $0 \le \chi(t) \le 1$ and $|\partial_t \chi(t)| \le 1$. For $\nu \in \mathbb{N}$, let

$$\Phi_\nu(t) = \int_0^t \chi_\nu(s)(\cos s)^2\,ds, \qquad \chi_\nu(t) = \chi\Big(\frac{t}{\nu}\Big). \tag{2-8}$$

For $\varepsilon > 0$, $w_{\varepsilon,\nu}(t) = e^{\varepsilon\Phi_\nu(t)}\cos t$ satisfies

$$\partial_t^2 w_{\varepsilon,\nu} + \alpha_{\varepsilon,\nu} w_{\varepsilon,\nu} = 0, \tag{2-9}$$

where

$$\begin{aligned}\alpha_{\varepsilon,\nu}(t) &= 1 + \varepsilon\chi_\nu \sin 2t - \varepsilon\Phi_\nu'' - (\varepsilon\Phi_\nu')^2\\ &= 1 + 2\varepsilon\chi_\nu \sin 2t - \varepsilon\chi_\nu'(\cos t)^2 - \varepsilon^2\chi_\nu^2(\cos t)^4.\end{aligned} \tag{2-10}$$

For $\varepsilon \le \varepsilon_0 = \frac{1}{10}$ and for all $\nu$,

$$|\alpha_{\varepsilon,\nu} - 1| \le \tfrac{1}{2}, \tag{2-11}$$

and we always assume below that the condition $\varepsilon \le \varepsilon_0$ is satisfied. We note also that $\alpha_{\varepsilon,\nu} = 1$ for $|t| \ge 7\nu\pi$, since $\chi_\nu$ vanishes there.

The final step is to localize the solution in $[-6\nu\pi, 6\nu\pi]$. Introduce an odd cut-off function $\zeta(t)$ supported in $]-6\pi, 6\pi[$ and equal to 1 for $|t| \le 4\pi$, and let

$$\tilde{w}_{\varepsilon,\nu}(t) = \zeta\Big(\frac{t}{\nu}\Big) w_{\varepsilon,\nu}(t). \tag{2-12}$$

This function is supported in $[-6\nu\pi, 6\nu\pi]$ and equal to $w_{\varepsilon,\nu}$ on $[-4\nu\pi, 4\nu\pi]$. Then

$$f_{\varepsilon,\nu} = \partial_t^2 \tilde{w}_{\varepsilon,\nu} + \alpha_{\varepsilon,\nu}\tilde{w}_{\varepsilon,\nu} = 2\nu^{-1}\zeta'\Big(\frac{t}{\nu}\Big)\partial_t w_{\varepsilon,\nu} + \nu^{-2}\zeta''\Big(\frac{t}{\nu}\Big) w_{\varepsilon,\nu} \tag{2-13}$$

is supported in $[-6\nu\pi, -4\nu\pi] \cup [4\nu\pi, 6\nu\pi]$.

**Lemma 2.1.** *For all $j$, there is a constant $C_j$ such that, for all $\varepsilon \le \varepsilon_0$ and all $\nu \ge 1$,*

$$\|\partial_t^j f_{\varepsilon,\nu}\|_{L^2} \le C_j \nu^{-1} e^{-\varepsilon\nu\pi} \|\tilde{w}_{\varepsilon,\nu}\|_{L^2}. \tag{2-14}$$

*Proof.* By symmetry, it is sufficient to estimate $f_{\varepsilon,\nu}$ for $t \ge 0$, that is, on $[4\nu\pi, 6\nu\pi]$. On $[2\nu\pi, 6\nu\pi]$, $\chi_\nu = -1$, hence $\Phi_\nu - \phi$ is constant and

$$w_{\varepsilon,\nu}(t) = c_{\nu,\varepsilon}\underline{w}_\varepsilon(t), \qquad c_{\nu,\varepsilon} = e^{\varepsilon\Phi_\nu(2\nu\pi)}.$$

Moreover, on this interval $\alpha_{\varepsilon,\nu} = \underline{\alpha}_\varepsilon$ is bounded with derivatives bounded independently of $\varepsilon$, and hence

$$\|\partial_t^j f_{\varepsilon,\nu}\|_{L^2} \le C_j \nu^{-1} c_{\nu,\varepsilon} \|(\underline{w}_\varepsilon, \partial_t \underline{w}_\varepsilon)\|_{L^2([4\nu\pi, 6\nu\pi])}.$$

By (2-7), this implies

$$\|\partial_t^j f_{\varepsilon,\nu}\|_{L^2} \le C_j \nu^{-1} c_{\nu,\varepsilon} e^{-\varepsilon\nu\pi} \|(\underline{w}_\varepsilon, \partial_t \underline{w}_\varepsilon)\|_{L^2([2\nu\pi, 4\nu\pi])}.$$

On the other hand,

$$\|w_{\varepsilon,\nu}\|_{L^2} \geq c_{\nu,\varepsilon} \|\underline{w}_\varepsilon\|_{L^2([2\nu\pi,4\nu\pi])}.$$

Therefore, it is sufficient to prove that there is a constant $C$ such that, for all $\nu$ and $\varepsilon$,

$$\|(\underline{w}_\varepsilon, \partial_t \underline{w}_\varepsilon)\|_{L^2([2\nu\pi,4\nu\pi])} \leq C \|\underline{w}_\varepsilon\|_{L^2([2\nu\pi,4\nu\pi])}.$$

Using (2-7) again, one has

$$\|(\underline{w}_\varepsilon, \partial_t \underline{w}_\varepsilon)\|^2_{L^2([2\nu\pi,4\nu\pi])} = \sum_{k=0}^{\nu-1} e^{-2(\varepsilon k+\nu)\pi} \|(\underline{w}_\varepsilon, \partial_t \underline{w}_\varepsilon)\|^2_{L^2([0,2\pi])}$$

and

$$\|\underline{w}_\varepsilon\|^2_{L^2([2\nu\pi,4\nu\pi])} = \sum_{k=0}^{\nu-1} e^{-2(\varepsilon k+\nu)\pi} \|\underline{w}_\varepsilon\|^2_{L^2([0,2\pi])}.$$

On $[0, 2\pi]$, the $H^1$ norms of the $\underline{w}_\varepsilon$ are uniformly bounded, while their $L^2$ norms remain larger than a positive constant. □

***Construction of the coefficients and of the solutions.*** For $k \geq 1$, let $\rho_k = k^{-2}$. We consider intervals $I_k = [t_k, t'_k]$ and $J_k = [t'_k, t_{k-1}]$ of the same length $\rho_k = t'_k - t_t = t_{k-1} - t'_k$, starting at $t_0 = 2\sum_{k=1}^{\infty} \rho_k$, and thus such that $t_k \to 0$.

The functions $a$ and $b$ are defined on $]0, t_0]$ as follows: we fix a function $\beta \in C^\infty(\mathbb{R})$ supported in $]0, 1[$ and with sequences $\varepsilon_k$, $\nu_k$ and $\delta_k$ to be chosen later on;

$$\begin{aligned} &\text{on } I_k, \quad \begin{cases} a(t) = \delta_k \alpha_{\varepsilon_k,\nu_k}(-8\pi\nu_k + 16\pi(t-t_k)\nu_k/\rho_k), \\ b(t) = \delta_k, \end{cases} \\ &\text{on } J_k, \quad a(t) = b(t) = \delta_k + (\delta_{k-1} - \delta_k)\beta((t-t'_k)/\rho_k). \end{aligned} \tag{2-15}$$

Because $\alpha_{\varepsilon,\nu} = 1$ for $|t| \geq 7\nu\pi$, the coefficient $a$ equals $\delta_k$ near the endpoints of $I_k$. The use of the function $\beta$ on $J_k$ makes a smooth transition between $\delta_k$ and $\delta_{k-1}$. Therefore, $a$ and $b$ are $C^\infty$ on $]0, t_0]$. The coefficients will be chosen so that they extend smoothly up to $t = 0$.

This is quite similar to the choice in [Colombini and Nishitani 1999], except that we introduce a new sequence $\varepsilon_k$, which is crucial to control the Hölder continuity of $\sigma = a/b$.

We use the family (2-12) to construct solutions of the system supported in $I_k$ for $k$ large. On $I_k$, $b$ is constant and (2-2) reads

$$\partial_t u_k + i h_k \delta_k \alpha_k v_k = f_k, \qquad \partial_t v_k + i h_k \delta_k u_k = 0 \tag{2-16}$$

with

$$\alpha_k(t) = \alpha_{\varepsilon_k,\nu_k}\left(-8\pi\nu_k + \frac{16\pi(t-t_k)\nu_k}{\rho_k}\right).$$

Therefore, a $C^\infty$ solution of (2-2) supported in $I_k$ is

$$u_k(t) = i\partial_t \tilde{w}_{\varepsilon_k,\nu_k}\left(-8\pi\nu_k + \frac{16\pi(t-t_k)\nu_k}{\rho_k}\right), \quad v_k(t) = \tilde{w}_{\varepsilon_k,\nu_k}\left(-8\pi\nu_k + \frac{16\pi(t-t_k)\nu_k}{\rho_k}\right) \tag{2-17}$$

with

$$f_k(t) = 16i\pi\left(\frac{\nu_k}{\rho_k}\right) f_{\varepsilon_k,\nu_k}\left(-8\pi\nu_k + \frac{16\pi(t-t_k)\nu_k}{\rho_k}\right) \tag{2-18}$$

provided that

$$h_k = \frac{16\pi\nu_k}{\rho_k\delta_k}. \tag{2-19}$$

## 3. Properties of the coefficients

We always assume that

$$\varepsilon_k \le \varepsilon_0, \quad \varepsilon_k\nu_k \to +\infty, \quad \delta_k \to 0. \tag{3-1}$$

***Conditions for blow-up.***

**Lemma 3.1.** *If*

$$(\rho_k)^{-1} e^{-\varepsilon_k\nu_k\pi} \to 0, \tag{3-2}$$

*then the blow-up property in* $L^2$, (2-3), *is satisfied.*

*Proof.* By Lemma 2.1,

$$\|f_k\|_{L^2} \le C\rho_k^{-1} e^{-\varepsilon_k\nu_k\pi}\|v_k\|_{L^2}. \qquad \square$$

**Lemma 3.2.** *If*

$$\frac{1}{\varepsilon_k\nu_k}\ln\left(\frac{h_k\nu_k}{\rho_k}\right) \to 0, \tag{3-3}$$

*then the blow-up property in* $C^\infty$, (2-4), *is satisfied.*

*Proof.* By Lemma 2.1, one has

$$\frac{\|\partial_t^l h_k^j f_k\|_{L^2}}{\|(u_k, v_k)\|_{L^2}} \le C_l \nu_k^{-1} h_k^j \left(\frac{16\pi\nu_k}{\rho_k}\right)^{l+1} e^{-\varepsilon_k\nu_k\pi}.$$

This tends to 0 if

$$\varepsilon_k\nu_k\pi - j\ln h_k - (l+1)\ln\left(\frac{\nu_k}{\rho_k}\right) \to +\infty.$$

If (3-3) is satisfied, this is true for all $j$ and $l$. $\square$

***Smoothness of the coefficients.***

**Lemma 3.3.** *If*

$$\frac{\ln(\nu_k/\rho_k)}{|\ln(\delta_k\varepsilon_k)|} \to 0, \tag{3-4}$$

*then the functions $a$ and $b$ are $C^\infty$ up to $t = 0$.*

*Proof.* Both $a$ and $b$ are $O(\delta_k)$ and thus converge to 0 when $t \to 0$. Moreover, for $j \geq 1$,

$$|\partial_t^j a| \leq C_j \begin{cases} \delta_k \varepsilon_k (\nu_k/\rho_k)^j & \text{on } I_k, \\ \delta_k \rho_k^{-j} & \text{on } J_k. \end{cases}$$

The worst situation occurs on $I_k$ and the right-hand side is bounded if

$$j \ln\left(\frac{\nu_k}{\rho_k}\right) - |\ln(\delta_k \varepsilon_k)|$$

is bounded from above. This is true for all $j$ under the assumption (3-4), implying that $a$ is $C^\infty$ on $[0, t_0]$. The proof for $b$ is similar and easier. □

Next, we investigate the possible Gevrey regularity of the coefficients. For that we need to make a special choice of the cut-off functions $\chi$ and $\beta$ which occur in the construction of $a$ and $b$. We can choose them in a class contained in $\bigcap_{s>1} G^s$ and containing compactly supported functions (see, e.g., [Mandelbrojt 1952]). We choose them so that there is a constant $C$ such that, for all $j$,

$$\sup_t (|\partial_t^j \chi(t)| + |\partial_t^j \beta(t)|) \leq C^{j+1} j! (\ln j)^{2j}. \tag{3-5}$$

**Lemma 3.4.** *If* (3-5) *is satisfied then, for* $j \geq 1$,

$$\sup_{t \in I_k \cup J_k} \big(|\partial_t^j a(t)| + |\partial_t^j b(t)|\big) \leq K^{j+1} \delta_k \varepsilon_k \left(\left(\frac{\nu_k}{\rho_k}\right)^j + \left(\frac{1}{\rho_k}\right)^j j! (\ln j)^{2j}\right). \tag{3-6}$$

*Proof.* On $I_k$ we take advantage of the explicit form (2-10) of $\alpha_{\varepsilon,\nu}$: it is a finite sum of sine and cosines with coefficients of the form $\chi(t/\nu)$. Scaled on $I_k$, each derivative of the trigonometric functions yields a factor $\nu_k/\rho_k$, while the derivatives of $\chi_{\nu_k}$ have only a factor $1/\rho_k$. Since $\chi'$ and $\chi^2$ satisfy estimates similar to (3-5), we conclude that $a$ satisfies

$$|\partial_t^j a(t)| \leq \varepsilon_k \delta_k K^j \sum_{l \leq j} \left(\frac{\nu_k}{\rho_k}\right)^{j-l} C^{l+1} l! (\ln l)^{2l},$$

implying the estimate (3-6) on $I_k$. On $I_k$, $b$ is constant. On $J_k$ things are clear by scaling, since the coefficients are functions of $\beta((t - t_k')/\rho_k)$. □

To estimate quantities such as $\delta_k(\nu_k/\rho_k)^j$, we use the following inequalities for $a > 0$ and $x \geq 1$:

$$e^{-x} x^a \leq a^a \tag{3-7}$$

and

$$e^{-e^x} x^a \leq \begin{cases} |\ln a|^a & \text{when } a \geq e, \\ 1 & \text{when } a \leq e. \end{cases} \tag{3-8}$$

**Corollary 3.5.** *Suppose that* $\delta_k = e^{-\eta_k}$ *and that, for* $s > s' > 1$,

$$\left(\frac{\nu_k}{\rho_k}\right) \leq C \eta_k^s \quad \textit{and} \quad \left(\frac{1}{\rho_k}\right)^j \leq C \eta_k^{s'-1}. \tag{3-9}$$

*Then the coefficients belong to the Gevrey class* $G^s$.

*If, for some* $p > 0$ *and* $q > 0$,

$$\eta_k \geq e^{k^q} \quad \textit{and} \quad \left(\frac{\nu_k}{\rho_k}\right) \leq C k^p \eta_k, \tag{3-10}$$

*then the coefficients belong to* $\bigcap_{s>1} G^s$.

*Proof.* We neglect $\varepsilon_k$ and only use the bound $\varepsilon_k \leq \varepsilon_0$. In the first case, we obtain from (3-7) that

$$\delta_k \left(\frac{\nu_k}{\rho_k}\right)^j \leq e^{-\eta_k} (C\eta_k)^{sj} \leq (C' j)^{js}, \quad \delta_k \left(\frac{1}{\rho_k}\right)^j \leq (C'' j)^{j(s'-1)},$$

implying that

$$|\partial_t^j (a, b)| \leq K^{j+1} j^{sj}.$$

In the second case, combining (3-7) and (3-8)

$$e^{-\eta_k} \left(\frac{\nu_k}{\rho_k}\right)^j \leq C'^j j^j k^{pj} e^{-\eta_k/2} \leq C''^j j^j (1 + \ln j)^{pj/q}.$$

Using (3-8) again for the second term, we obtain that

$$|\partial_t^j (a, b)| \leq K^{j+1} j^j (\ln j)^{rj}$$

with $r = \max\{p, 4\}/q$. In particular, the right-hand side is estimated by $K_s^{k+1} j^{js}$ for all $s > 1$, proving that the functions $a$ and $b$ belong to $\bigcap_{s>1} G^s$. □

***Smoothness of the symmetrizer.***

**Lemma 3.6.** *Suppose that* $\omega$ *is a continuous and increasing function on* $[0, 1]$ *such that* $t^{-1}\omega(t)$ *is decreasing. If*

$$\varepsilon_k \leq \omega\left(\frac{\rho_k}{\nu_k}\right) \tag{3-11}$$

*then* $\sigma = a/b$ *satisfies*

$$|\sigma(t) - \sigma(t')| \leq C\omega(|t - t'|). \tag{3-12}$$

*In particular, if* $\mu \leq 1$ *and*

$$\limsup_k \varepsilon_k \left(\frac{\nu_k}{\rho_k}\right)^\mu < +\infty, \tag{3-13}$$

*then* $\sigma$ *is Hölder continuous of exponent* $\mu$. *If*

$$\varepsilon_k \left(\frac{\nu_k}{\rho_k}\right) \leq C \ln\left(\frac{\nu_k}{\rho_k}\right)^\theta, \tag{3-14}$$

*then* $\omega(t) = t|\ln t|^\theta$ *is a modulus of continuity for* $\sigma$.

*Proof.* On $J_k$, $\tilde\sigma = \sigma - 1$ vanishes and, on $I_k$,

$$\tilde\sigma = \varepsilon_k \alpha_{\varepsilon_k,\nu_k}\left(-8\pi\nu_k + \frac{16\pi(t-t_k)\nu_k}{\rho_k}\right),$$

and thus

$$|\tilde\sigma| \le C\varepsilon_k, \quad |\partial_t\tilde\sigma| \le \frac{C\varepsilon_k\nu_k}{\rho_k}. \tag{3-15}$$

Hence, for $t$ and $t'$ in $I_k$,

$$|\tilde\sigma(t) - \tilde\sigma(t')| \le C\varepsilon_k \min\left\{1, \frac{|t-t'|\nu_k}{\rho_k}\right\}.$$

If $\rho_k/\nu_k \le |t-t'|$, we use the first estimate and

$$|\tilde\sigma(t) - \tilde\sigma(t')| \le C\varepsilon_k \le C\omega\left(\frac{\rho_k}{\nu_k}\right) \le C\omega(|t-t'|).$$

If $|t-t'| \le \rho_k/\nu_k$, we use the second estimate and the monotonicity of $t^{-1}\omega(t)$:

$$|\tilde\sigma(t) - \tilde\sigma(t')| \le C\varepsilon_k\left(\frac{\nu_k}{\rho_k}\right)|t-t'| \le C\left(\frac{\nu_k}{\rho_k}\right)\omega\left(\frac{\rho_k}{\nu_k}\right)|t-t'| \le C\omega(|t-t'|).$$

This shows that (3-12) is satisfied when $t$ and $t'$ belong to the same interval $I_k$.

If $t$ belongs to $I_k$ and $t' \in J_k$, then $\tilde\sigma(t') = \tilde\sigma(t'_k) = 0$ and

$$|\tilde\sigma(t) - \tilde\sigma(t')| \le C\omega(|t-t'_k|) \le C\omega(|t-t'|).$$

Similarly, if $t < t'$ and $t$ and $t'$ do not belong to the same $I_k \cup J_k$, there are endpoints $t_j$ and $t_l$ such that $t_j \le t \le t_{j-1} \le t_l \le t' \le t_{l-1}$. Since $\tilde\sigma$ vanishes at the endpoints of $I_k$ and on $J_k$,

$$\begin{aligned}|\tilde\sigma(t) - \tilde\sigma(t')| &\le C|\tilde\sigma(t) - \tilde\sigma(t_j)| + |\tilde\sigma(t) - \tilde\sigma(t_j)|\\ &\le C\omega(|t-t_{j-1}) + C\omega(|t_l - t'|) \le C\omega(|t-t'|),\end{aligned}$$

and the lemma is proved. □

## 4. Proof of the theorems

We now adapt the choice of the parameters $\varepsilon_k$, $\nu_k$ and $\delta_k$ so that the coefficients and the symmetrizer satisfy the properties stated in the different theorems. We will choose two increasing functions, $f$ and $g$, on $\{x \ge 1\}$ and define $\varepsilon_k$ and $\delta_k$ in terms of $\nu_k$ through the relations

$$\frac{\varepsilon_k\nu_k}{\rho_k} = f\left(\frac{\nu_k}{\rho_k}\right), \quad \delta_k = e^{-\eta_k}, \quad \eta_k = g\left(\frac{\nu_k}{\rho_k}\right). \tag{4-1}$$

Recall that $\rho_k = k^{-2}$. The sequence of integers $\nu_k$ will be chosen to converge to $+\infty$, and thus $\nu_k/\rho_k \to +\infty$. The conditions (3-1) are satisfied if, at $+\infty$,

$$f(x) \ll x, \quad g(x) \to +\infty. \tag{4-2}$$

Here $\phi(x) \ll \psi(x)$ means that $\psi(x)/\phi(x) \to \infty$. In particular, the first condition implies that $\varepsilon_k \to 0$, so that the condition $\varepsilon \le \varepsilon_0$ is certainly satisfied if $k$ is large enough.

One has

$$|\ln(\delta_k \varepsilon_k)| = \eta_k + \ln\left(\frac{\nu_k}{\rho_k}\right) + \ln f\left(\frac{\nu_k}{\rho_k}\right).$$

Hence, by Lemma 3.3, the coefficients $a$ and $b$ are $C^\infty$ when

$$\ln x \ll g(x) \ll x, \tag{4-3}$$

since with (4-2) this implies that $|\ln(\delta_k \varepsilon_k)| \sim \eta_k \gg \ln(\nu_k/\rho_k)$.

*Proof of Theorem 1.1.* Given the modulus of continuity $\omega$, we choose $f(x) = x\omega(x^{-1})$. The assumption on $\omega$ is that $f$ is increasing and $f(x) \to +\infty$ at infinity. The essence of the theorem is that $f$ can grow to infinity as slowly as one wants. Lemma 3.6 implies that $\omega$ is a modulus of continuity for $\sigma = a/b$. By Lemma 3.1, the blow-up property (2-3) occurs when

$$k^2 e^{-k^{-2} f(k^2 \nu_k)\pi} \to 0.$$

This condition is satisfied if $\nu_k$ satisfies

$$f(k^2 \nu_k) \ge k^3. \tag{4-4}$$

Let $f_1(x) = \min\{f(x), \ln x\}$. We choose $g(x) = x/f_1(x)$ and $\nu_k$ such that

$$2k^3 \le f_1(k^2 \nu_k) \le 4k^3.$$

Note that this implies (4-4). We show that the conditions (3-10) are satisfied with $p = q = 3$ and $C = 4$ and a suitable choice of $\nu_k$, so that, by Corollary 3.5, the coefficients belong to $\bigcap_{s>1} G^s$ and the theorem is proved.

Indeed, since $f_1(k^2\nu_k) \le 4k^3$, the condition $\nu_k/\rho_k \le 4k^3\eta_k$ is satisfied. Moreover, since $\ln(k^2\nu_k) \ge 2k^3$,

$$\nu_k \ge k^{-2} e^{2k^3} \ge e^{k^3}$$

for $k$ large. □

*Proof of Theorem 1.2.* The proof is similar. Given the modulus of continuity $\omega$, we choose $f(x) = x\omega(x^{-1})$. The assumption on $\omega$ is now that

$$\ln x \ll f(x). \tag{4-5}$$

The essence of the theorem is now that $f(x)/\ln x$ can grow to infinity as slowly as one wants. By Lemma 3.6, $\omega$ is a modulus of continuity for $\sigma = a/b$.

By Lemma 3.2, the blow-up property (2-4) is satisfied if

$$\ln h_k = \eta_k + \ln\left(\frac{\nu_k}{\rho_k}\right) + \ln(16\pi) \ll \varepsilon_k \nu_k;$$

that is, if

$$\rho_k f\left(\frac{\nu_k}{\rho_k}\right) \gg g\left(\frac{\nu_k}{\rho_k}\right) + \ln\left(\frac{\nu_k}{\rho_k}\right). \tag{4-6}$$

Let $\psi(x) = f(x)/\ln x$ and $g(x) = \sqrt{\psi(x)}\ln x$. Then

$$\psi(x) \gg 1, \quad \ln x \ll g(x) \ll f(x).$$

Therefore, the condition (4-6) is satisfied when $\rho_k\sqrt{\psi}(\nu_k/\rho_k) \to +\infty$, and for that it is sufficient to choose $\nu_k$ such that

$$\psi(k^2\nu_k) \geq k^5. \tag{4-7}$$

The condition $g(x) \gg \ln x$ implies that the coefficients are $C^\infty$, and the theorem is proved. □

*Proof of Theorem 1.5.* With $s > 1$ and $0 < \mu < 1 - 1/s$, we choose

$$g(x) = x^{1/s} \ll f(x) = x^{1-\mu}. \tag{4-8}$$

The choice of $f$ implies that $\sigma = a/b \in C^\mu$. The choice of $g$ implies that

$$\frac{\nu_k}{\rho_k} \leq \left(g\left(\frac{\nu_k}{\rho_k}\right)\right)^s = \eta_k^s.$$

With $s' \in ]1, s[$, the condition

$$\rho_k^{-1} \leq \eta_k^{s'-1}$$

is satisfied when $k^2 \leq (k^2\nu_k)^{(s'-1)/s}$, that is, when

$$\nu_k \geq k^{2p}, \quad \text{where} \quad p = \frac{1+s-s'}{s'-1}. \tag{4-9}$$

In this case, Corollary 3.5 implies that the coefficients $a$ and $b$ belong to the Gevrey class $G^s$.

The blow-up property (2-4) is satisfied when (4-6) holds, that is, when

$$k^{-2}(k^2\nu_k)^{1-\mu} \gg (k^2\nu_k)^{1/s},$$

which is true if

$$\nu_k \geq k^{2q}, \quad \text{where} \quad q = \frac{\mu + 1/s}{1 - \mu - 1/s}.$$

Therefore, if $\nu_k \geq k^{2\max\{p,q\}}$, the system satisfies the conclusions of Theorem 1.5. □

*Proof of Theorem 1.3.* The analysis above shows that if one looks for coefficients in $\bigcap_{s>1} G^s$, one must choose $g$, and thus $f$, close to $x$. We choose here

$$g(x) = \frac{x}{(\ln x)^2} \ll f(x) = \frac{x}{\ln x} \ll x$$

Since $f(x)/x \to 0$ at infinity, the symmetrizer is continuous up to $t = 0$, but not in $C^\mu$ for any $\mu > 0$.

The ill-posedness in $C^\infty$ is again guaranteed by the condition (4-6), that is, $\ln(k^2\nu_k) \gg k^2$. In particular, it is satisfied when

$$\nu_k \geq e^{k^3}. \tag{4-10}$$

By Corollary 3.5, to finish the proof of Theorem 1.3 it is sufficient to show that one can choose $\nu_k$ satisfying (4-10) such that $\nu_k/\rho_k \le 4k^6\eta_k$. This condition reads $\ln(k^2\nu_k) \le 2k^3$, or

$$\nu_k \le k^{-2}e^{2k^3}$$

which is compatible with (4-10) if $k$ is large enough. □

*Proof of Theorem 1.4.* Let $a \in \bigcap_{s>1} G^s$ denote the coefficient constructed for the proof of Theorem 1.3. The definition (2-15) shows that $a \ge 0$, and indeed $a > 0$, for $t > 0$. The functions $v_k$ defined at (2-17) are supported in $I_k$ and are solutions of the wave equation (1-5) with source term $f_k$, and we have shown that

$$\frac{\|h_k^j \partial_t^l f_k\|_{L^2}}{\|v_k\|_{L^2}} \to 0 \quad \text{as } k \to \infty. \qquad \square$$

## References


[Cicognani and Colombini 2006] M. Cicognani and F. Colombini, "Modulus of continuity of the coefficients and loss of derivatives in the strictly hyperbolic Cauchy problem", *J. Differential Equations* **221**:1 (2006), 143–157. MR 2006i:35213 Zbl 1097.35092

[Colombini and Nishitani 1999] F. Colombini and T. Nishitani, "Two by two strongly hyperbolic systems and Gevrey classes", *Ann. Univ. Ferrara Sez. VII (N.S.)* **45** (1999), 79–108. MR 2003e:35179 Zbl 0993.35061

[Colombini and Spagnolo 1982] F. Colombini and S. Spagnolo, "An example of a weakly hyperbolic Cauchy problem not well posed in $C^\infty$", *Acta Math.* **148** (1982), 243–253. MR 83m:35085 Zbl 0517.35053

[Colombini and Spagnolo 1989] F. Colombini and S. Spagnolo, "Some examples of hyperbolic equations without local solvability", *Ann. Sci. École Norm. Sup.* (4) **22**:1 (1989), 109–125. MR 90d:35164 Zbl 0702.35146

[Colombini et al. 1979] F. Colombini, E. De Giorgi, and S. Spagnolo, "Sur les équations hyperboliques avec des coefficients qui ne dépendent que du temps", *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* (4) **6**:3 (1979), 511–559. MR 81c:35077 Zbl 0417.35049

[Colombini et al. 1983] F. Colombini, E. Jannelli, and S. Spagnolo, "Well-posedness in the Gevrey classes of the Cauchy problem for a nonstrictly hyperbolic equation with coefficients depending on time", *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* (4) **10**:2 (1983), 291–312. MR 85f:35131 Zbl 0543.35056

[Friedrichs 1954] K. O. Friedrichs, "Symmetric hyperbolic linear differential equations", *Comm. Pure Appl. Math.* **7** (1954), 345–392. MR 16,44c Zbl 0059.08902

[Mandelbrojt 1952] S. Mandelbrojt, *Séries adhérentes, régularisation des suites, applications*, Gauthier-Villars, Paris, 1952. MR 14,542f Zbl 0048.05203

[Métivier 2014] G. Métivier, "$L^2$ well-posed Cauchy problems and symmetrizability of first order systems", *J. Éc. Polytech., Math.* **1** (2014), 39–70. Zbl 06389884


Received 20 Sep 2014. Accepted 9 Jan 2015.

FERRUCCIO COLOMBINI: colombini@dm.unipi.it
*Dipartimento di Matematica, Università di Pisa, Largo B. Pontecorvo 5, I-56127 Pisa, Italy*

GUY MÉTIVIER: guy.metivier@math.u-bordeaux1.fr
*Institut de Mathématiques de Bordeaux, CNRS, UMR 5251, Université de Bordeaux, 351 Cours de la Libération, F-33405 Talence, France*

## Guidelines for Authors

Authors may submit manuscripts in PDF format on-line at the Submission page at msp.org/apde.

**Originality**. Submission of a manuscript acknowledges that the manuscript is original and and is not, in whole or in part, published or under consideration for publication elsewhere. It is understood also that the manuscript will not be submitted elsewhere while under consideration for publication in this journal.

**Language**. Articles in APDE are usually in English, but articles written in other languages are welcome.

**Required items**. A brief abstract of about 150 words or less must be included. It should be self-contained and not make any reference to the bibliography. If the article is not in English, two versions of the abstract must be included, one in the language of the article and one in English. Also required are keywords and subject classifications for the article, and, for each author, postal address, affiliation (if appropriate), and email address.

**Format**. Authors are encouraged to use LaTeX but submissions in other varieties of TeX, and exceptionally in other formats, are acceptable. Initial uploads should be in PDF format; after the refereeing process we will ask you to submit all source material.

**References**. Bibliographical references should be complete, including article titles and page ranges. All references in the bibliography should be cited in the text. The use of BibTeX is preferred but not required. Tags will be converted to the house format, however, for submission you may use the format of your choice. Links will be provided to all literature with known web locations and authors are encouraged to provide their own links in addition to those supplied in the editorial process.

**Figures**. Figures must be of publication quality. After acceptance, you will need to submit the original source files in vector graphics format for all diagrams in your manuscript: vector EPS or vector PDF files are the most useful.

Most drawing and graphing packages (Mathematica, Adobe Illustrator, Corel Draw, MATLAB, etc.) allow the user to save files in one of these formats. Make sure that what you are saving is vector graphics and not a bitmap. If you need help, please write to graphics@msp.org with details about how your graphics were generated.

**White space**. Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

**Proofs**. Page proofs will be made available to authors (or to the designated corresponding author) at a Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

# ANALYSIS & PDE

Volume 8 No. 2 2015





2157-5045(2015)8:2;1-C