



Three-Scale Singular Limits of Evolutionary PDEs

BIN CHENG, QIANGCHANG JU, STEVE SCHOCHET

Communicated by P.-L. LIONS

Abstract

Singular limits of a class of evolutionary systems of partial differential equations having two small parameters and hence three time scales are considered. Under appropriate conditions solutions are shown to exist and remain uniformly bounded for a fixed time as the two parameters tend to zero at different rates. A simple example shows the necessity of those conditions in order for uniform bounds to hold. Under further conditions the solutions of the original system tend to solutions of a limit equation as the parameters tend to zero.

1. Introduction

Many physical systems contain several small parameters, such as the Mach number, Alfvén number, Froude number, Rossby number, etc.. When these parameters are considered to have fixed ratios to one another then the system has two time scales: one induced by the terms containing the small parameters and the other coming from the order-one terms in the equation. The classical theory of singular limits for evolutionary partial differential equations (PDEs) ([1, 4, 5, 7, 9, 12, 13] and numerous papers on particular systems, for example [11]) was developed to treat this case. In order to determine the behavior of solutions when two physical parameters tend to zero in a different manner it is necessary to develop an analogous theory for systems with three time scales. The systems to be considered here have the form

$$A_0(\varepsilon u)u_t + \sum_{j=1}^d A_j(u)u_{x_j} + \frac{1}{\delta}\mathcal{L}u + \frac{1}{\varepsilon}\mathcal{M}u = F(t, x, u), \quad (1.1)$$

where ε and δ are small parameters. As in the theory of two-scale singular limits, the system without the large terms is assumed to be symmetric hyperbolic, and

\mathcal{L} and \mathcal{M} are assumed to be antisymmetric constant-coefficient differential or pseudodifferential operators of order at most one. As for two-scale singular limits [4, 7], parabolic terms of size $O(1)$ could be added to the right side of (1.1), although the complications such terms induce would be greater in the three-scale case.

The fundamental discovery of Klainerman and Majda [7, 8] for two-scale singular limits was that the presence of the small parameter in the matrix A_0 , which occurs naturally in the normalized equations for low Mach number fluid flow, induces a delicate balance. As they showed, this ensures that solutions of (1.1) with $\delta = 1$, having fixed initial data belonging to a Sobolev space of sufficiently high index, exist for a time independent of the small parameter ε and satisfy bounds independent of that parameter, without the need for additional conditions on the large terms or the initial data, such as those assumed in [1, 12] to treat the case when A_0 depends on u rather than εu . Whenever the small parameter δ in (1.1) is not asymptotically smaller than ε , that is, when $\delta \geq c\varepsilon$ for some arbitrarily small positive constant c , then the Klainerman–Majda balance is essentially preserved and their uniform existence result remains valid and requires only cosmetic changes to the proof. Similarly, when A_0 is a constant matrix, as in the rotating shallow water equations ([10, Equation (2.2)]) then the Klainerman–Majda uniform existence result remains valid for arbitrary δ and ε .

Hence we will be concerned here with the case when A_0 does depend nontrivially on εu , and

$$0 < \delta \ll \varepsilon \ll 1. \quad (1.2)$$

Our first main result is a uniform existence theorem under two additional assumptions. The first condition is that

$$\delta \geq c\varepsilon^{1+\frac{1}{s_0}} \quad (1.3)$$

for some positive constant c , where

$$s_0 := \left\lfloor \frac{d}{2} \right\rfloor + 1 \quad (1.4)$$

is the Sobolev embedding exponent in dimension d . The second condition is that the initial data $u_0(x, \varepsilon, \delta)$ are uniformly bounded in the Sobolev space $H^{s_0+1}(\mathbb{D})$ and are “well-prepared” in the usual sense that the initial time derivative

$$u_t(0, x) := A_0(\varepsilon u_0)^{-1} \left[F(0, x, u_0) - \sum_{j=1}^d A_j(u_0)(u_0)_{x_j} - \frac{1}{\delta} \mathcal{L}u_0 - \frac{1}{\varepsilon} \mathcal{M}u_0 \right] \quad (1.5)$$

is uniformly bounded in $H^{s_0}(\mathbb{D})$, with the domain \mathbb{D} being either the whole space \mathbb{R}^d or the torus \mathbb{T}^d . Examples of initial data satisfying this condition are given in (3.10) below. For convenience, we shall henceforth omit the spatial domain in integrals and function spaces throughout the paper. Although (1.3) limits how small δ can be compared to ε , it is consistent with the scaling (1.2) that violates the Klainerman–Majda balance. Moreover, both conditions are necessary, at least for

obtaining uniform bounds on solutions of general systems without Klainerman–Majda balance, as will be shown via a simple explicit example.

Our other main result is a convergence theorem showing, under the additional assumptions described below, that as ε and δ both tend to zero solutions of (1.1) whose initial data converge in H^{s_0+1} tend to the solution of a certain limiting equation. The framework of the convergence theorem is the same as for two-scale singular limits; the bounds of the existence result yield compactness, which implies that every sequence of ε and δ tending to zero while obeying (1.3) has a subsequence for which the solution converges, and convergence without restricting to such subsequences is obtained by showing that the limit satisfies a limit equation for which solutions of initial-value problems are unique. However, both the form and the derivation of the limit equation are more complicated for three-scale singular limits. For the two-scale singular limit obtained when $\delta \equiv 1$, the limit equation is obtained by decomposing (1.1) into the projections onto the null space of \mathcal{M} and onto its orthogonal complement, multiplying the latter by ε and taking the limits of the results. However, in order to obtain the limit equation for the three-scale singular limit in which (1.2), (1.3) holds it is necessary to use perturbation theory to compute some number of terms of the power series in the small parameter $\mu = \frac{\delta}{\varepsilon}$ for the eigenvalues and eigenspaces of $\mathcal{L} + \mu\mathcal{M}$ in Fourier space. The number of terms required and the resulting limit equation depend on the relationship between δ and ε as they both tend to zero. In order to obtain convergence without restricting to subsequences it is necessary to restrict the relationship between δ and ε so as to obtain a specific limit equation. This requires the additional assumption that for some integer $s \geq s_0$ either

$$\frac{\delta}{\varepsilon^{1+\frac{1}{s}}} \rightarrow C > 0 \quad \text{as } \varepsilon \text{ and } \delta \text{ tend to zero} \tag{1.6}$$

or

$$\frac{\delta}{\varepsilon^{1+\frac{1}{s}}} \rightarrow \infty \quad \text{and} \quad \frac{\delta}{\varepsilon^{1+\frac{1}{s+1}}} \rightarrow 0 \quad \text{as } \varepsilon \text{ and } \delta \text{ tend to zero,} \tag{1.7}$$

either of which implies (1.2), (1.3) hold. Note that if $\frac{\delta}{\varepsilon^{1+\frac{1}{r}}} \rightarrow C > 0$ for some $r > s_0$ that is not an integer then (1.7) holds with $s = \lfloor r \rfloor$. The limit equation is different for different values of s and even for different values of C in (1.6), but is the same for all r in $(s, s + 1)$. The reason that the limit equation depends on C is that when (1.6) holds then the limit equation contains a term T_{lim} arising from the power series expansion in δ of $\frac{1}{\delta}(\mathcal{L} + \mu\mathcal{M})$. Moreover, although both \mathcal{L} and \mathcal{M} are both bounded operators from H^1 to L^2 it turns out that T_{lim} may not be, as will be explained in Definition 4.4 and Remark 4.5 below. Such terms do not occur in two-scale singular limits. As a result, the second time derivative of the limit solution may not belong to L^2 , although the limit process ensures that its first time derivative does belong to L^2 .

After presenting the example showing the necessity of our conditions for obtaining uniform bounds in Section 2, the uniform existence theorem will be formulated precisely and proven in Section 3, and the convergence theorem will be formulated precisely and proven in Section 4. Some simple examples of the perturbation procedure and the limit equations will also be presented in that section. In forthcoming

work the results here will be applied to the problem that motivated this research, namely the simultaneous zero Alfvén number and zero Mach number limit of the scaled compressible inviscid MHD equations

$$a(1 + \varepsilon_M r) (\partial_t r + \mathbf{u} \cdot \nabla r) + R(r, \varepsilon_M) \nabla \cdot \mathbf{u} + \varepsilon_M^{-1} \nabla \cdot \mathbf{u} = 0, \quad (1.8a)$$

$$(1 + \varepsilon_M r) (\partial_t \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u}) + R(r, \varepsilon_M) \nabla r + \varepsilon_M^{-1} \nabla r \\ + \nabla \frac{|\mathbf{b}|^2}{2} - \mathbf{b} \cdot \nabla \mathbf{b} = \delta_A^{-1} (\partial_z \mathbf{b} - \nabla b_3), \quad (1.8b)$$

$$\partial_t \mathbf{b} + \mathbf{u} \cdot \nabla \mathbf{b} + (\nabla \cdot \mathbf{u}) \mathbf{b} - \mathbf{b} \cdot \nabla \mathbf{u} = \delta_A^{-1} (\partial_z \mathbf{u} - \mathbf{e}_z \nabla \cdot \mathbf{u}), \quad \nabla \cdot \mathbf{b} = 0, \quad (1.8c)$$

where the small parameters ε_M and δ_A are respectively the Mach number and Alfvén number, the fluid density is $1 + \varepsilon_M r$, its velocity is \mathbf{u} , the magnetic field is $\mathbf{e}_z + \delta_A \mathbf{b}$ with \mathbf{e}_z being the unit vector in the z -direction, and the coefficient functions a and R depend on the constitutive relation giving the fluid pressure as a function of its density.

2. Example

Consider the system

$$a(\varepsilon w) u_t - \frac{1}{\delta} v = 0, \quad a(\varepsilon w) v_t + \frac{1}{\delta} u = 0, \quad w_t = 0, \quad (2.1)$$

which has the form (1.1), together with the initial data

$$u(0, x) = u_0 := \delta, \quad v(0, x) = v_0 := 0, \quad w(0, x) = w_0(x) \quad (2.2)$$

that satisfy the condition that the initial time derivative be uniformly bounded. Arguments that we will make regarding this simpler system could be adapted to more complex versions. For example, the system could be turned into one in which the large terms involve derivatives with respect to an additional spatial variable y by replacing the terms $-\frac{1}{\delta} v$ and $\frac{1}{\delta} u$ by $\frac{1}{\delta} v_y$ and $\frac{1}{\delta} u_y$, respectively, and changing u_0 to $\delta \cos y$. A term containing $\frac{1}{\varepsilon}$ could also be added.

It will be convenient to write the solution to (2.1), (2.2) in terms of

$$z := u + iv, \quad (2.3)$$

which satisfies

$$z_t = -\frac{iz}{\delta a(\varepsilon w_0(x))}, \quad z(0, x) = \delta, \quad (2.4)$$

since $w(t, x) = w_0(x)$. The solution to (2.4) is

$$z(t, x) = \delta e^{-\frac{it}{\delta a(\varepsilon w_0(x))}}. \quad (2.5)$$

Differentiating (2.5) or its derivatives with respect to t produces a term containing a factor $\frac{1}{\delta}$, while differentiating with respect to x produces a term containing a factor $\frac{\varepsilon}{\delta}$ since the x -dependence in the exponent of (2.5) lies inside $a(\varepsilon \cdot)$. Taking

into account the factor of δ in (2.5) that comes from the initial condition, this shows that

$$\partial_x^\ell \partial_t^k z = \frac{\varepsilon^\ell}{\delta^{k+\ell-1}} (z_{k,\ell}(t, x) + o(1)) \tag{2.6}$$

for some function $z_{k,\ell}$ that is not identically zero provided that both a and w_0 genuinely depend on their argument.

The standard existence theory for symmetric hyperbolic systems in spatial dimension d requires obtaining a bound on the H^{s_0+1} norm of solutions. The system (2.1) can be considered to be a system in any dimension, and estimate (2.6) implies that the solution of (2.1), (2.2) will be uniformly bounded in H^{s_0+1} only when $\frac{\varepsilon^{s_0+1}}{\delta^{s_0}}$ is bounded, which requires that (1.3) must hold.

Moreover, if the condition that the initial data must be well prepared is dropped then the initial value of u in (2.2) can be 1 rather than δ , which makes (2.6) more singular by one power of δ . The condition that the H^{s_0+1} norm of the solution be uniformly bounded then requires that $\frac{\varepsilon^{s_0+1}}{\delta^{s_0+1}}$ be bounded, that is, that $\delta \geq c\varepsilon$, so no general result beyond the Klainerman–Majda balance is then possible.

3. Uniform Existence Result

3.1. Scaling

Estimate (2.6) implies that the derivatives of solutions (u, v, w) of (2.1) satisfy corresponding estimates, except that (u, v, w) itself and its pure spatial derivatives are no smaller than $O(1)$ because that is the size of the component w . These estimates suggest that the appropriate norm of solutions of (1.1) to estimate would be

$$\|u\|_{H^{s_0+1}} + \sum_{k=1}^{s_0+1} \sum_{0 \leq |\alpha| \leq s_0+1-k} \frac{\delta^{k+|\alpha|-1}}{\varepsilon^{|\alpha|}} \|D^\alpha \partial_t^k u\|_{L^2}, \tag{3.1}$$

where as usual D^α denotes the spatial derivative $\prod_{j=1}^d \partial_{x_j}^{\alpha_j}$ of order $|\alpha| := \sum_{j=1}^d \alpha_j$. Although our method indeed allows us to estimate the weighted norm (3.1) of solutions, doing so requires keeping an exact count of the spatial derivatives appearing in instances of the Gagliardo–Nirenberg inequalities (3.16) below. In order to avoid the need to count spatial derivatives we will instead perform a simplified estimate by using weights that depend only on the number of time derivatives, with the weight of the term $\partial_t^k u$ and its spatial derivatives equal to ε^k . These weights equal their counterparts in (3.1) for the highest spatial derivative of $\partial_t^k u$, under the assumption that equality holds in (1.3). For lower-order spatial derivatives when that equality holds, or for all cases when strict inequality holds in (1.3), the weights we use are smaller than their counterparts in (3.1). Hence the simplified estimate will be somewhat weaker than the estimate that would be obtained using (3.1). With one exception this difference is of little importance, because estimates of norms of time derivatives of solutions weighted by small constants are simply a means

of obtaining an unweighted estimate for the spatial norms of solutions. The one exception is that the L^2 norm of u_t in (3.1) has weight one and so yields a uniform bound, while the L^2 norm of u_t in the simplified scheme has weight ε and so does not yield a uniform bound. Obtaining a uniform bound for some norm of u_t is important for the convergence theory, and it will turn out that the time evolution of the unweighted L^2 norm of u_t can be estimated in terms of the norms appearing in the simplified estimates, so we will simply adjoin the unweighted L^2 norm of u_t to the simplified scheme of estimates.

However, as is common in the theory of hyperbolic systems, we must modify the standard L^2 and H^s norms to include the coefficient matrix $A_0(\varepsilon u)$ of the time-derivative term in the PDE (1.1), with the argument εu of A_0 taken from some solution to (1.1). We therefore define

$$\begin{aligned}
 \langle v, w \rangle_{A_0} &:= \int v^T A_0(\varepsilon u) w \, dx, \\
 \|v\|_{0, A_0} &:= \sqrt{\langle v, v \rangle_{A_0}}, \\
 \|v\|_{\ell, A_0} &:= \sqrt{\sum_{0 \leq |\alpha| \leq \ell} \|D^\alpha v\|_{0, A_0}^2}, \\
 |||u|||_{s, \varepsilon, A_0} &:= \sqrt{\sum_{k=0}^s \varepsilon^{2k} \|\partial_t^k u(t, \cdot)\|_{s-k, A_0}^2}, \\
 |||u|||_{s, \varepsilon, A_0} &:= \sqrt{|||u|||_{s, \varepsilon, A_0}^2 + \|u_t\|_{0, A_0}^2}.
 \end{aligned} \tag{3.2}$$

The corresponding quantities with the subscript A_0 omitted will denote the standard inner product and norms in which A_0 is replaced by the identity matrix. Assumption 3.3 together with the estimates to be obtained will ensure that the two are equivalent for the time intervals considered here. The definitions in (3.2) are a slight abuse of notation, both since the argument of A_0 is usually not given explicitly but must be understood from the context, and because the value of $|||u|||_{s, \varepsilon, A_0}$ and $|||u|||_{s, \varepsilon, A_0}$ at a given time does not depend solely on the value of u at that time on account of the inclusion of time derivatives.

Remark 3.1. The standard existence theorem for symmetric hyperbolic systems ([9, Ch. 2, Theorem 2.1]) shows that there exists a unique solution, for some time that may depend on δ and ε , to the initial-value problem consisting of (1.1) together with an initial condition $u(0, x) = u_0(x, \delta, \varepsilon) \in H^{s_0+1}$, and, moreover that solution continues to exist as long as its H^{s_0+1} norm remains finite ([9, Ch. 2, Theorem 2.2]). Hence in order to prove that the time of existence can be taken to be independent of δ and ε it suffices to obtain a uniform bound on the H^{s_0+1} norm of the solution. The proof of the existence theorem uses estimates in which the function u appearing inside A_0 in the norm $\|\cdot\|_{\ell, A_0}$ from (3.2) differs from the solution being estimated. However, since in this paper the solutions being estimated are already known to exist the function u appearing inside A_0 in the norms (3.2) will simply be the solution that is being estimated.

Remark 3.2. The standard energy estimates for both symmetric hyperbolic systems without large terms and for singular limits obeying Klainerman–Majda balance involve only spatial derivatives of the solution. The reasons that time derivatives are also needed here and why an unweighted estimate for the time derivative is only obtained in the L^2 norm will be explained after the proof of Theorem 3.6.

3.2. Assumptions and Initial Data

The following standard conditions on the terms appearing in system (1.1) will be assumed, where s_0 is defined in (1.4):

- Assumption 3.3.** 1. The matrices A_0 and the A_j are symmetric and are C^{s_0+1} functions of their arguments.
 2. The matrix A_0 is positive definite; more specifically there are positive constants c_0 and b_0 such that

$$A_0(v) \geq c_0 I \quad \text{for } |v| \leq b_0. \tag{3.3}$$

3. The function $G(t, x) := F(t, x, 0)$ is bounded in H^{s_0+1} uniformly in t , and for $1 \leq k \leq s_0 + 1$ the H^{s_0+1-k} norm of $\partial_t^k G$ is bounded uniformly in t . In addition, the function $H(t, x, u) := \int_0^1 \frac{\partial F}{\partial u}(t, x, \alpha u) d\alpha$ belongs to C^{s_0+1} .
 4. The operators \mathcal{L} and \mathcal{M} are anti-symmetric constant-coefficient differential or pseudodifferential operators of order at most one.

Remark 3.4. The identity

$$F(t, x, u) - F(t, x, 0) = \int_0^1 \frac{d}{d\alpha} F(t, x, \alpha u) d\alpha = \left[\int_0^1 \frac{\partial F}{\partial u}(t, x, \alpha u) d\alpha \right] u$$

together with the definitions in Assumption 3.3 show that

$$F(t, x, u) \equiv G(t, x) + H(t, x, u)u. \tag{3.4}$$

As noted in the introduction, the initial data will be required to be chosen so that $u_t(0, x)$ from (1.5) is uniformly bounded in H^{s_0} . From the PDE (1.1) we see that this well-preparedness condition is equivalent to the condition that

$$\left(\frac{1}{\delta}\mathcal{L} + \frac{1}{\varepsilon}\mathcal{M}\right)u(0, x, \delta, \varepsilon) \text{ be uniformly bounded in } H^{s_0}. \tag{3.5}$$

Under the above conditions Lemma 3.5 below shows that the $||| |||_{s_0+1, \varepsilon, A_0}$ norm of u will be uniformly bounded at time zero. In the statements of both this result and the main theorem we will use the Sobolev embedding constant, that is, the constant K such that

$$\sup_x |v(x)| \leq K \|v\|_{s_0}. \tag{3.6}$$

Lemma 3.5. *Assume that initial data satisfy*

$$\|u_0(x, \delta, \varepsilon)\|_{s_0+1} \leq m_1 \quad \text{and} \quad \|(\frac{1}{\delta}\mathcal{L} + \frac{1}{\varepsilon}\mathcal{M})u_0(x, \delta, \varepsilon)\|_{s_0} \leq m_2 \quad (3.7)$$

for all

$$0 < \varepsilon \leq \varepsilon_0 \quad \text{and} \quad 0 < c_1\varepsilon^{1+\frac{1}{s_0}} \leq \delta \leq 1, \quad (3.8)$$

that Assumption 3.3 holds, and that $\varepsilon_0 K m_1 \leq \frac{b_0}{2}$, where b_0 is defined in Assumption 3.3. Let u be any function such that $u(0, x, \delta, \varepsilon) = u_0(x, \delta, \varepsilon)$, $u_t(0, x, \delta, \varepsilon)$ equals the right side of (1.5) obtained by solving (1.1) for u_t and setting t equal to zero, and the higher time derivatives of u at time zero through order $s_0 + 1$ are determined recursively in similar fashion by solving ∂_t^j of the PDE (1.1) for $\partial_t^{j+1}u$, setting t equal to zero, and substituting in the values of lower time derivatives of u at time zero already so determined.

Then there is a constant M depending only on the spatial dimension d , the constants c_0 from (3.3), m_1 and m_2 from (3.7), and ε_0 and c_1 from (3.8), the norms of \mathcal{L} and \mathcal{M} as operators from H^1 to L^2 , the C^{s_0+1} norms of A_0, A_j , and H over the domain $\{u \mid \leq 2K m_1\}$, and the H^{s_0+1} norm of G , such that at time zero

$$(\| \| \| \| \|_{s_0+1, \varepsilon, A_0}) \Big|_{t=0} \leq M \quad (3.9)$$

for all δ and ε satisfying the above conditions.

Proof. Roughly speaking, the result of the lemma follows from the fact that when $u_t(0, x)$ is $O(1)$ then using the PDE (1.1) plus induction shows that at time zero $\varepsilon^k \partial_t^k u$ is $O(\frac{\varepsilon^k}{\delta^{k-1}})$ for $2 \leq k \leq s_0 + 1$, which is $O(1)$ on account of the assumption that $\delta \geq c_1\varepsilon^{1+\frac{1}{s_0}}$, and hence yields the uniform boundedness of the $\| \| \| \|_{s_0+1, \varepsilon, A_0}$ norm of u at time zero.

More specifically, by repeated applications of the PDE (1.1) to express higher time derivatives in terms of u, u_t and their spatial derivatives, and applications of \mathcal{L} and \mathcal{M} to them, we obtain that, for $2 \leq k \leq s_0 + 1$, the leading-order term of $\partial_t^k u$ is $(\frac{1}{\delta}\mathcal{L} + \frac{1}{\varepsilon}\mathcal{M})^{k-1} \partial_t u \Big|_{t=0}$, which yields the estimate $\varepsilon^k \|\partial_t^k u \Big|_{t=0}\|_{s_0+1-k} \leq c \frac{\varepsilon^k}{\delta^{k-1}}$. To see this note first that the assumptions on the initial data ensure that $A_0 \geq c_0 I$. Applying ∂_t^{k-1} to (1.1), using the invertibility of A_0 to solve the result for $\partial_t^k u$, taking up to $s_0 + 1 - k$ spatial derivatives of the result, and summing the L^2 norms of the results yields a formula for the H^{s_0+1-k} norm of $\partial_t^k u$ in terms of L^2 norms of products of spatial derivatives of lower-order time derivatives of u . Note that coefficients such as $A_j(u)$ can be estimated in the maximum norm in terms of $\|u\|_{s_0}$ and so may be pulled out of those L^2 norms.

For the case $k = 2$ this yields an estimate of $\|u_{tt}\|_{s_0-1}$ in terms of L^2 norms of products of the factors u, u_x, u_t, G and G_t and their spatial derivatives of order at most s_0 , with coefficients of size at most $O(\frac{1}{\delta})$ coming from the presence of $\frac{1}{\delta}$ in the time derivative of (1.1). Since all those factors are bounded in H^{s_0} at time zero, and H^{s_0} is an algebra, this yields the estimate $\|u_{tt} \Big|_{t=0}\|_{s_0-1} \leq \frac{c}{\delta}$.

The analogous expressions for $\|\partial_t^k u\|_{s_0+1-k}$ with $k > 2$ include factors of u_{tt} and possibly higher time derivatives, plus their spatial derivatives. Although u_{tt} and

higher-order time derivatives of u do not belong to H^{s_0} at time zero, the resulting expressions could be estimated by the method used in the proof of Theorem 3.6 to estimate similar expressions. However, it is simpler to use finite induction to express higher-order time derivatives in terms of u and u_t . Since the time derivative in (1.1) is expressed in terms of expressions involving at most one spatial derivative, this again yields an estimate in terms of L^2 norms of products of the factors u, u_x, u_t and their spatial derivatives of order at most s_0 , plus time and spatial derivatives of G of order at most s_0 , this time with coefficients of size at most $O(\frac{1}{\delta^{k-1}})$ since equation (1.1) is used at most $k - 1$ times to express $k - 1$ time derivatives in terms of spatial derivatives. Note that wherever u_t and its spatial derivatives occur the time derivative is left unaltered rather than using (1.1) to express u_t in terms of u , because u_t is $O(1)$ at time zero but the individual terms on the right side of (1.1) may not be. This yields the estimates $\|\partial_t^k u|_{t=0}\|_{s_0+1-k} \leq \frac{c}{\delta^{k-1}}$. As indicated at the beginning of the proof, these estimates together with assumption (3.8) show that the $||| |||_{s_0+1, \varepsilon, A_0}$ norm of u is uniformly bounded at time zero. \square

The well-preparedness condition (3.5) can be achieved, for example, by using initial data of the form

$$u(0, x, \delta, \varepsilon) = u_0(x, \delta, \varepsilon) := \sum_{j=0}^m \left(\frac{\delta}{\varepsilon}\right)^j \tilde{u}_j(x) + \delta U_0(x, \delta, \varepsilon) \tag{3.10}$$

for some nonnegative integer m , with the \tilde{u}_j belonging to H^{s_0+1} and U_0 bounded in that space uniformly in δ and ε . In fact, since

$$\begin{aligned} \left(\frac{1}{\delta} \mathcal{L} + \frac{1}{\varepsilon} \mathcal{M}\right) u_0(x, \delta, \varepsilon) &= \frac{1}{\delta} \mathcal{L} \tilde{u}_0 + \sum_{j=1}^m \frac{\delta^{j-1}}{\varepsilon^j} (\mathcal{L} \tilde{u}_j \\ &\quad + \mathcal{M} \tilde{u}_{j-1}) + \frac{\delta^m}{\varepsilon^{m+1}} \mathcal{M} \tilde{u}_m + O(1), \end{aligned}$$

in view of the scaling assumption (1.2), condition (3.5) will hold provided that

$$\begin{aligned} \mathcal{L} \tilde{u}_0 &= 0, \quad \mathcal{L} \tilde{u}_j = -\mathcal{M} \tilde{u}_{j-1} \text{ for } j = 1, \dots, m, \\ \text{and either } \mathcal{M} \tilde{u}_m &= 0 \text{ or } \delta^m \leq c\varepsilon^{m+1}. \end{aligned}$$

For example, the well-preparedness condition holds when $m = 0$ and $\mathcal{L} \tilde{u}_0 = 0 = \mathcal{M} \tilde{u}_0$, or when $m = s_0$, equality holds in (1.3), $\mathcal{L} \tilde{u}_0 = 0$, and

$$\mathcal{L} \tilde{u}_j = -\mathcal{M} \tilde{u}_{j-1} \text{ for } j = 1, \dots, s_0. \tag{3.11}$$

When the ranges of \mathcal{L} and \mathcal{M} overlap, the condition (3.11) allows more general initial data than would be obtained by requiring that each side of those equations vanish separately.

3.3. Theorem and Proof

Theorem 3.6. *Under the assumptions of Lemma 3.5, there exists a constant T depending only on the quantities that M in that lemma depends on, such that for all ε and δ satisfying (3.8) the solution of the initial-value problem (1.1), $u(0, x, \delta, \varepsilon) = u_0(x, \delta, \varepsilon)$ exists on $[0, T]$ and satisfies $\max_{0 \leq t \leq T} |||u|||_{s_0+1, \varepsilon, A_0} \leq 2M$.*

Proof. The local existence and continuation theorems ([9, Ch 2., Theorems 2.1–2.2]) mentioned in Remark 3.1 ensure that the solution of the initial-value problem exists on some time interval that might depend on δ and ε , and will continue to exist for a time independent of those small parameters provided that it satisfies an H^{s_0+1} estimate independent of them. Hence it suffices to prove such an estimate. Moreover, although the norm $||| \cdot |||_{s_0+1, \varepsilon, A_0}$ used in the estimates below depends on the solution u being estimated, condition (3.3) ensures that the resulting estimate will indeed be uniform. The estimates that will be derived are similar to standard energy estimates for solutions of symmetric hyperbolic systems but require keeping track of the powers of δ and ε that appear in those estimates for the system (1.1).

Applying $D^\alpha \partial_t^k$ with $0 \leq k \leq s_0 + 1$ and $0 \leq |\alpha| \leq s_0 + 1 - k$ to (1.1), taking the inner product with $2D^\alpha \partial_t^k u$, integrating over the spatial variables, integrating by parts in the terms that involve A_j undifferentiated, noting that the terms involving \mathcal{L} or \mathcal{M} drop out on account of the anti-symmetry of those operators, summing over all α satisfying the above-mentioned condition, and multiplying the result by the weight ε^{2k} yields

$$\begin{aligned} \frac{d}{dt} [\varepsilon^{2k} \|\partial_t^k u\|_{s_0+1-k, A_0}^2] &= \frac{d}{dt} \left[\varepsilon^{2k} \sum_{0 \leq |\alpha| \leq s_0+1-k} \int (D^\alpha \partial_t^k u) \cdot A_0(\varepsilon u)(D^\alpha \partial_t^k u) \, dx \right] \\ &= \varepsilon^{2k} \sum_{0 \leq |\alpha| \leq s_0+1-k} \int (D^\alpha \partial_t^k u) \cdot \left[\varepsilon u_t \cdot \nabla_u A_0 + \sum_j u_{x_j} \cdot \nabla_u A_j \right] (D^\alpha \partial_t^k u) \, dx \\ &\quad + 2\varepsilon^{2k} \cdot \sum_{0 \leq |\alpha| \leq s_0+1-k} \int (D^\alpha \partial_t^k u) \cdot \left\{ D^\alpha \partial_t^k (G + Hu) - [D^\alpha \partial_t^k, A_0]u_t - \sum_j [D^\alpha \partial_t^k, A_j]u_{x_j} \right\} dx \\ &\leq \|\varepsilon u_t \cdot \nabla_u A_0 + \sum_j u_{x_j} \cdot \nabla_u A_j\|_{L^\infty} \varepsilon^{2k} \|\partial_t^k u\|_{s_0+1-k}^2 \\ &\quad + c\varepsilon^k \|\partial_t^k u\|_{s_0+1-k} \cdot \\ &\quad \left(\varepsilon^k \|\partial_t^k G\|_{s_0+1-k} + \varepsilon^k \|H\|_{L^\infty} \|\partial_t^k u\|_{s_0+1-k} + \varepsilon^k \left[\sum_{0 \leq |\alpha| \leq s_0+1-k} \|[D^\alpha \partial_t^k, A_0]u_t\|_{L^2}^2 \right]^{1/2} \right. \\ &\quad \left. + \varepsilon^k \left[\sum_{0 \leq |\alpha| \leq s_0+1-k} (\|D^\alpha \partial_t^k, H\|_{L^2}^2 + \sum_j \|[D^\alpha \partial_t^k, A_j]u_{x_j}\|_{L^2}^2) \right]^{1/2} \right), \end{aligned} \tag{3.12}$$

where the inequality is obtained by pulling out $\varepsilon u_t \cdot \nabla_u A_0 + \sum_j u_{x_j} \cdot \nabla_u A_j$ from the first integral in maximum norm, and breaking the second integral into several parts and using the Cauchy–Schwartz inequality in each of them.

Since $A_0 = A_0(\varepsilon u)$ will be differentiated at least once when it appears in any commutator term on the right side of the inequality in (3.12), which yields at least one power of ε , the power of ε in every term appearing on the right side of the

inequality in (3.12) is at least as large as the total number of time derivatives in that term. By the definition of the $||| \cdot |||_{s_0+1, \varepsilon, A_0}$ norm plus the smoothness assumption on A_0 , this implies that in order to bound the right side of the inequality in (3.12) by a continuous function of $|||u|||_{s_0+1, \varepsilon, A_0}$ it suffices to bound all the terms there by a continuous function of $|||u|||_{s_0+1, 1}$ after replacing ε by 1 and replacing A_0 by the identity matrix.

The condition on s_0 ensures that $\|u_t\|_{L^\infty}$ and $\|\nabla u\|_{L^\infty}$ are bounded by a constant times $\|u_t\|_{s_0}$ and $\|u\|_{s_0+1}$, respectively, and those norms are each bounded by $|||u|||_{s_0+1, 1}$. By the smoothness of the A_j , $\sum_{j=0}^d \|\nabla_u A_j\|_{L^\infty} \leq c(\|u\|_{s_0}) \leq \tilde{c}(||u|||_{s_0+1, 1})$ for some continuous function \tilde{c} . This yields the desired estimate for the entire first term on the right side of the inequality in (3.12). The terms on the right side of the inequality in (3.12) in which G and the L^∞ norm of H appear are also so bounded in view of the assumptions of those functions.

There remains to estimate only the terms on the right side of the inequality in (3.12) that involve commutators. Since the factor $\|\partial_t^{k_\ell} u\|_{s_0+1-k_\ell}$ multiplying the norms of the commutators is one of the terms in $|||u|||_{s_0+1, 1}$, only the norms of the commutator terms themselves must be estimated. We can pull out in the L^∞ norm any factor such as $\nabla_u H$ that depends only on t, x and u without derivatives, and the assumptions on the various coefficients ensure that each factor so pulled out is bounded by a continuous function of $\|u\|_{s_0}$ and hence by a continuous function of $|||u|||_{s_0+1, 1}$. Since the presence of the commutator ensures that at least one derivative will be applied to the function appearing in the commutator, the terms arising from the commutators that remain inside the L^2 norms all take the form

$$\left[\int \prod_{\ell=1}^L |D^{\alpha_\ell} \partial_t^{k_\ell} u|^2 dx \right]^{1/2}, \tag{3.13}$$

where $L \geq 2$, $1 \leq |\alpha_\ell| + k_\ell \leq s_0 + 1$, and $\sum_\ell (|\alpha_\ell| + k_\ell) \leq s_0 + 2$. If $|\alpha_\ell| + k_\ell = s_0 + 1$ for some ℓ then only one derivative is applied to the other factor, so that factor can be pulled out in L^∞ norm and estimated by $\|u\|_{s_0+1}$ or $\|u_t\|_{s_0}$, both of which appear in $|||u|||_{s_0+1, 1}$. After pulling out that factor the integral becomes $\int |D^{\alpha_\ell} \partial_t^{k_\ell} u|^2 dx$, which is bounded by $\|\partial_t^{k_\ell} u\|_{s_0+1-k_\ell}^2$, which also appears in $|||u|||_{s_0+1, 1}$. Otherwise $|\alpha_\ell| + k_\ell \leq s_0$ for all ℓ , and by using the multiple-factor version of Hölder’s inequality we will bound the integral in (3.13) by

$$\prod_{\ell=1}^L \left(\int |D^{\alpha_\ell} \partial_t^{k_\ell} u|^{2p_\ell} dx \right)^{\frac{1}{p_\ell}}, \tag{3.14}$$

where the exponents p_ℓ , which will be chosen later, must satisfy

$$1 \leq p_\ell \leq \infty \quad \text{and} \quad \sum_\ell \frac{1}{p_\ell} = 1. \tag{3.15}$$

The integrals in (3.14) will then be bounded via the Gagliardo–Nirenberg inequality (for example, [3, p. 24])

$$\|v\|_{L^p} \leq c \|v\|_r^a \|v\|_{L^2}^{1-a}, \tag{3.16}$$

in which the parameters must satisfy $\frac{1}{p} = \frac{1}{2} - \frac{ar}{d}$, $r \geq 1$, and $0 \leq a < 1$, where as usual d is the spatial dimension. Although a is actually allowed to equal the endpoint value 1 for many values of the other parameters, we will avoid that value in order to obtain a unified proof. The inequality constraint on a will hold provided that $\frac{1}{2} \geq \frac{1}{p} > \frac{1}{2} - \frac{r}{d}$. In order to estimate the integrals in (3.14) we apply (3.16) with $v := D^{\alpha_\ell} \partial_t^{k_\ell} u$, so we will let $r = s_0 + 1 - (|\alpha_\ell| + k_\ell)$, since that is the highest Sobolev norm of $D^{\alpha_\ell} \partial_t^{k_\ell} u$ that is bounded by $\|u\|_{s_0+1,1}$. Since we only use (3.14) when $|\alpha_\ell| + k_\ell \leq s_0$ for all ℓ , the condition $r \geq 1$ will indeed hold. Since the norm of $D^{\alpha_\ell} \partial_t^{k_\ell} u$ appearing in (3.14) is the L^{2p_ℓ} norm, p in (3.16) equals $2p_\ell$. Substituting in these values and multiplying everywhere by two turns the inequality constraint on p into the inequality constraint

$$1 \geq \frac{1}{p_\ell} > 1 - \frac{2(s_0 + 1 - (|\alpha_\ell| + k_\ell))}{d} \tag{3.17}$$

on p_ℓ . We now show that it is possible to choose the p_ℓ such that both (3.15) and (3.17) hold.

Since $|\alpha_\ell| + k_\ell \leq s_0$, the interval to which $\frac{1}{2p_\ell}$ is restricted by (3.17) is nonempty. Since $|\alpha_\ell| + k_\ell \geq 1$ and $\frac{d}{2} \leq s_0 \leq \frac{d}{2} + 1$, the lower limit in (3.17) is negative iff $|\alpha_\ell| + k_\ell = 1$. The inequality in (3.15) is equivalent to

$$1 \geq \frac{1}{p_\ell} \geq 0. \tag{3.18}$$

Combining (3.18) with (3.17) yields

$$1 \geq \frac{1}{p_\ell} > \max \left(0, 1 - \frac{2(s_0 + 1 - (|\alpha_\ell| + k_\ell))}{d} \right), \tag{3.19}$$

where for simplicity we ignore the possibility $p_\ell = \infty$, which will not be needed. Every value of p_ℓ satisfying (3.19) is allowed by both (3.17) and (3.18), so it suffices to show that we can choose values in the intervals in (3.19) that sum to one. That is possible iff the sum of the lower values there is less than one and the sum of the upper values is at least one. The latter condition holds trivially, and as noted above the second expression inside the max in (3.19) is negative iff $|\alpha_\ell| + k_\ell = 1$, so it suffices to show that

$$1 > \sum_{\substack{1 \leq \ell \leq L \\ |\alpha_\ell| + k_\ell \geq 2}} \left[1 - \frac{2(s_0 + 1 - (|\alpha_\ell| + k_\ell))}{d} \right]. \tag{3.20}$$

Let L_2 denote the number of values of ℓ for which $|\alpha_\ell| + k_\ell \geq 2$. If $L_2 = 0$ then the sum on the right side of (3.20) vanishes, so that condition indeed holds. When $L_2 \geq 1$ then (3.20) can be written more explicitly as

$$1 > L_2 \left(1 - \frac{2}{d}(s_0 + 1) \right) + \frac{2}{d} \left[\left(\sum_{\ell=1}^L (|\alpha_\ell| + k_\ell) \right) - (L - L_2) \right]. \tag{3.21}$$

Condition (3.21) can be rewritten as

$$(L_2 - 1)(s_0 - \frac{d}{2}) > (2 - L) + \left(\sum_{\ell=1}^L (|\alpha_\ell| + k_\ell) - (s_0 + 2) \right). \quad (3.22)$$

Since $L_2 \geq 1$ by assumption, $s_0 > \frac{d}{2}$, $L \geq 2$ and $\sum_{\ell=1}^L (|\alpha_\ell| + k_\ell) \leq s_0 + 2$, the left side of (3.22) is non-negative, and the right side there is non-positive. Moreover, since $L \geq 2$, if $L_2 = 1$ then exists an ℓ for which $|\alpha_\ell| + k_\ell = 1$, and in that case the fact that $|\alpha_\ell| + k_\ell \leq s_0$ implies that either $L > 2$ or $\sum_{\ell} (|\alpha_\ell| + k_\ell) < s_0 + 2$. This shows that either the left side of (3.22) is strictly positive or the right side there is strictly negative, and hence that inequality indeed holds.

Summing over $0 \leq k \leq s_0 + 1$ the estimates that we have obtained shows that

$$\frac{d}{dt} |||u|||_{s_0+1, \varepsilon, A_0}^2 \leq c(|||u|||_{s_0+1, \varepsilon, A_0}) \quad (3.23)$$

for some continuous function c .

Finally, differentiating (1.1) with respect to t , taking the inner product of the result with $2u_t$, integrating over the spatial variables, integrating by parts in the terms that involve A_j undifferentiated, and noting that the terms involving \mathcal{L} or \mathcal{M} drop out on account of the anti-symmetry of those operators yields

$$\begin{aligned} \frac{d}{dt} [||u_t||_{0, A_0}^2] &= \frac{d}{dt} \left[\int u_t \cdot A_0(\varepsilon u) u_t \, dx \right] \\ &= \int \partial_t u \cdot \left[\varepsilon u_t \cdot \nabla_u A_0 + \sum_j u_{x_j} \cdot \nabla_u A_j \right] u_t \, dx \\ &\quad + 2 \int u_t \cdot \left\{ \partial_t(G + Hu) - (\varepsilon u_t \cdot \nabla_u A_0) u_t - \sum_j (u_t \cdot \nabla_u A_j) u_{x_j} \right\} dx \\ &\leq \|\varepsilon u_t \cdot \nabla_u A_0 + \sum_j u_{x_j} \cdot \nabla_u A_j\|_{L^\infty} \|\partial_t u\|_0^2 \\ &\quad + c \|\partial_t u\|_0 \left(\|\partial_t G\|_0 + \|H\|_{L^\infty} \|\partial_t u\|_0 + \|\frac{\partial H}{\partial t}\|_{L^\infty} \|u\|_0 + \|u_t\|_0 \|\nabla_u H\|_{L^\infty} \|u\|_{L^\infty} \right. \\ &\quad \left. + \varepsilon \|u_t\|_{L^\infty} \|\nabla_u A_0\|_{L^\infty} \|u_t\|_0 + \|u_t\|_0 \sum_j \|\nabla_u A_j\|_{L^\infty} \|u_{x_j}\|_{L^\infty} \right), \\ &\leq c(|||u|||_{s_0+1, \varepsilon, A_0}) \left[\|u_t\|_0^2 + \|u_t\|_0 \right] \end{aligned} \quad (3.24)$$

where the first inequality follows in similar fashion to (3.12) and the second from Assumption 3.3 plus the definition of the $||| \cdot |||$ norm. Adding (3.24) to (3.23) yields the uniform estimate

$$\frac{d}{dt} |||u|||_{s_0+1, \varepsilon, A_0}^2 \leq c(|||u|||_{s_0+1, \varepsilon, A_0}) \quad (3.25)$$

for some continuous function c . By Lemma 3.5, $|||u|||_{s_0+1, \varepsilon, A_0}$ is bounded uniformly in δ and ε by M at time zero, so the differential inequality (3.24) shows that there is a fixed positive constant T such that $\max_{0 \leq t \leq T} |||u|||_{s_0+1, \varepsilon, A_0} \leq 2M$. \square

Remark 3.7. 1. In the standard energy estimates for spatial derivatives of solutions of systems without large terms and of systems satisfying Klainerman–Majda balance, integrals of the form (3.14) not containing time derivatives are estimated using the Gagliardo–Nirenberg inequality

$$\|D^\alpha v\|_{L^p} \leq c \|v\|_s^a \|v\|_{L^\infty}^{1-a} \tag{3.26}$$

with $p = \frac{2s}{|\alpha|}$, $a = \frac{|\alpha|}{s}$, $s \geq s_0$, and $s > |\alpha|$ instead of (3.16). However, it is not possible to use (3.26) to estimate integrals involving second and higher time derivatives, because the boundedness of $\|u\|_{s_0+1,\varepsilon,A_0}$ does not imply even an ε -dependent bound for $\|\partial_t^k u\|_{L^\infty}$ when $k \geq 2$.

2. The special case of (3.14) and (3.16) in which $p_\ell = 2$, $|\alpha_\ell| + k_\ell = 2$, $p = 4$, $r = 1$, $a = \frac{d}{4}$, and d is either two or three so that $s_0 = 2$ was used previously in [2, §4.1 and Appendix].
3. The expression εu_t appears in the estimates for a purely spatial derivative D^α of u , arising from the commutator term $[D^\alpha, A_0]u_t$. When the spatial derivative terms in the PDE are at most $O(\frac{1}{\varepsilon})$ then substituting for εu_t from the PDE yields a spatial derivative term of order one. Making this substitution allows spatial derivatives to be estimated without requiring estimates of time derivatives, both for systems without large terms and in the Klainerman–Majda theory. However, for the PDE (1.1) with the scaling (1.2) this procedure cannot be used because it yields terms of order $\frac{\varepsilon}{\delta}$, which is large. It is therefore necessary to leave the term εu_t on the right side of the energy estimates for spatial derivatives of u , and this necessitates estimating time derivatives as well. Similarly, two-scale systems for which A_0 depends on u rather than εu also require estimates of time derivatives [1, 12].
4. In a similar fashion, a term containing εu_{tt} appears in estimates for a spatial derivative of u_t on account of the commutator term $[D^\alpha \partial_t, A_0]u_t$. Assuming that u_t is bounded initially but εu_{tt} is large at time zero, this prevents us from obtaining an unweighted estimate for spatial derivatives of u_t . The reason we do obtain an unweighted estimate for u_t itself is that the commutator term $[\partial_t, A_0]u_t$ does not yield any second time derivative.
5. The bound (3.8) on how fast δ can tend to zero compared to ε is only needed to ensure that the $\|u\|_{s_0+1,\varepsilon,A_0}$ norm of the solution is uniformly bounded at time zero. The proof of Theorem 3.6 therefore also yields uniform bounds for a uniform time in the case when the time derivatives of the solution through order $s_0 + 1$ are uniformly bounded at time zero, without the need for assumption (3.8) and without using weights of powers of ε in the norms. In particular, taking $\varepsilon \equiv 1$ and letting $\delta \rightarrow 0$ yields a proof for arbitrary dimensions of the uniform existence theorem stated in [1] but only proven there in the case $d = 1$, for which no Gagliardo–Nirenberg estimates are needed.

4. Convergence

4.1. A Finite-Dimensional Perturbation Result

We begin with a result on perturbations of self-adjoint matrices $\mathcal{T}(\mu) := \frac{1}{\mu^p}(T^{(0,0)} + \mu T^{(0,1)})$, where μ is a small parameter, and p is a positive integer. The result will be used in the proof of the convergence theorem in Section 4.2, where $T^{(0,0)}$ and $T^{(0,1)}$ will stand for the Fourier symbols of operators \mathcal{L} and \mathcal{M} respectively. The result says that there is an orthogonal projection $\mathcal{P}(\mu)$ that commutes with $\mathcal{T}(\mu)$, on whose range $\mathcal{T}(\mu)$ is bounded uniformly and has a limit as $\mu \rightarrow 0$, and on whose null space $\mathcal{T}(\mu)$ is bounded from below by a constant times $\frac{1}{\mu}$ and has a finite expansion in inverse powers of μ .

Lemma 4.1. *Define $\mathcal{T}(\mu) := \frac{1}{\mu^p}T^{(0)}(\mu) := \frac{1}{\mu^p}(T^{(0,0)} + \mu T^{(0,1)})$, where $T^{(0,0)}$ and $T^{(0,1)}$ are operators on a finite dimensional inner-product space X that are either both self-adjoint or both skew-adjoint, μ is a small parameter, and p is a positive integer. Then*

1. *There exists an orthogonal projection operator $\mathcal{P}(\mu)$ that commutes with $\mathcal{T}(\mu)$ for $\mu \neq 0$, is analytic in μ for real μ , and satisfies*

$$\|\mathcal{P}(\mu)\mathcal{T}(\mu)\mathcal{P}(\mu)f\|_X \leq c_1\|f\|_X \tag{4.1}$$

$$\|(I - \mathcal{P}(\mu))\mathcal{T}(\mu)(I - \mathcal{P}(\mu))f\|_X \geq \frac{c_2}{\mu}\|(I - \mathcal{P}(\mu))f\|_X \tag{4.2}$$

for $0 < \mu < \mu_0$, where $\|\cdot\|_X$ is the norm on the space X and μ_0 and the c_j are positive constants.

2. *For $0 \leq j \leq p-1$ there exist commuting orthogonal projection operators $P^{(j)}$ such that the ranges of the complementary projections $I - P^{(j)}$ are mutually orthogonal subspaces, and*

$$\mathcal{P}(0) = \prod_{j=0}^{p-1} P^{(j)} = I - \sum_{j=0}^{p-1} (I - P^{(j)}). \tag{4.3}$$

3. *The $P^{(j)}$ are the orthogonal projection operators onto the null spaces of operators $T^{(j,j)}$, which are determined from $T^{(0)}(\mu) := T^{(0,0)} + \mu T^{(0,1)}$ via the reduction process of [6, §II.2.3]: Specifically, after modifying the notation to facilitate repeated reductions, the $T^{(j,j)}$ are the first terms in the expansions*

$$T^{(j+1)}(\mu) := \frac{1}{\mu} \tilde{P}^{(j)}(\mu)T^{(j)}(\mu)\tilde{P}^{(j)}(\mu) = \sum_{k=0}^{\infty} \mu^k T^{(j+1,j+1+k)}. \tag{4.4}$$

Here $\tilde{P}^{(j)}(\mu) = \tilde{P}^{(j-1)}(\mu)P^{(j)}(\mu)\tilde{P}^{(j-1)}(\mu)$ for $j \geq 0$, $P^{(j)}(\mu)$ is the orthogonal projection onto the direct sum of the eigenspaces of $T^{(j)}(\mu)$ of all eigenvalues of order $o(1)$, and $\tilde{P}^{(-1)}(\mu) := I$. The $T^{(j,k)}$ are all self-adjoint when $T^{(0,0)}$ and $T^{(0,1)}$ are self-adjoint, and are all skew-adjoint when $T^{(0,0)}$ and $T^{(0,1)}$ are skew-adjoint.

4.

$$\lim_{\mu \rightarrow 0} \mathcal{P}(\mu) \mathcal{T}(\mu) = T^{(p,p)}. \tag{4.5}$$

5. The operator $T^{(0,0)}$ is given, and for $1 \leq j \leq 2$,

$$T^{(1,1)} = P^{(0)} T^{(0,1)} P^{(0)} \tag{4.6}$$

$$T^{(2,2)} = -\tilde{P}^{(1)} T^{(0,1)} \left(T^{(0,0)} \right)^{-1}_{\psi} T^{(0,1)} \tilde{P}^{(1)} \tag{4.7}$$

where $\tilde{P}^{(j)} := \prod_{\ell=0}^j P^{(j)}$, and $(M)^{-1}_{\psi}$ denotes the pseudo-inverse of the matrix M , defined by $(C^{-1} \begin{pmatrix} M_{11} & 0 \\ 0 & 0 \end{pmatrix} C)^{-1}_{\psi} := C^{-1} \begin{pmatrix} M_{11}^{-1} & 0 \\ 0 & 0 \end{pmatrix} C$.

Proof. If both $T^{(0,0)}$ and $T^{(0,1)}$ are skew-adjoint then multiplying both of them by i makes them self-adjoint without affecting the projections, so we may assume that they are self-adjoint. Moreover, as noted in [6, §II.6.1], the reduction process preserves self-adjointness and so may be continued without limitation, since the nilpotent factors of the general case are absent. In particular, by [6, Theorem 6.1 in §II.6.1] the eigenvalues of $T^{(j)}(\mu)$ and the projection operators $P^{(j)}(\mu)$ are all analytic for real μ . As in [6, §II.1.3] let $R(z, \mu)$ denote the resolvent $(T^{(0)}(\mu) - z)^{-1}$ wherever z is not an eigenvalue of $T^{(0)}(\mu)$. In view of the analyticity of the eigenvalues, [6, (1.16) in §II.1.4] shows that for μ sufficiently small the operator

$$\mathcal{P}(\mu) = -\frac{1}{2\pi i} \int_{|z|=\mu^{p-\frac{1}{2}}} R(z, \mu) dz$$

is the orthogonal projection onto the direct sum of the eigenspaces of the eigenvalues of $\mathcal{T}(\mu)$ of size at most $O(1)$, and $I - \mathcal{P}(\mu)$ is the orthogonal projection onto the direct sum of the eigenspaces of the eigenvalues of $\mathcal{T}(\mu)$ of size at least $O(\mu^{-1}) \gg 1$. These estimates show that (4.1), (4.2) hold.

We carry out the reduction process of [6, §II.2.3] while choosing the unperturbed eigenvalue zero at every stage. However, we do not want to include the range of $I - P^{(j-1)}$ when considering the zero eigenspace of $T^{(j,j)}$ since that subspace has already been accounted for at previous stages of the reduction process. For this reason we replace the factor $P^{(j)}(\mu)$, which would appear in (4.4) if the corresponding formula [6, (2.37) in §II.2.3] were simply rewritten in our notation, with $\tilde{P}^{(j)}(\mu)$. This corresponds to the suggestion in [6, §II.2.3] to add a constant multiple of $I - P^{(j-1)}$ to $T^{(j,j)}$ but without the need to modify that operator. This procedure yields (4.4).

Since $\mathcal{T}(\mu)$, after multiplication by i if necessary, is self-adjoint for all μ , formula (4.4) implies that all the $T^{(j,k)}$ are then also self-adjoint after that multiplication has been done if necessary. If at any stage of the reduction process the new unperturbed operator $T^{(j+1,j+1)}$ does not have any zero eigenspace except for the range of $(I - P^{(j)})$, then $P^{(j+1)}$ and hence also $\mathcal{P}(0)$ is identically zero, and if $j + 1 < p - 1$ then $T^{(k)}(\mu)$ is simply the zero operator for $j + 1 < k \leq p - 1$.

By the construction of the reduction process, $I - P^{(j)}(\mu)$ is the orthogonal projection onto the direct sum of the eigenspaces of $\mathcal{T}(\mu)$ whose eigenvalues are of

size $O(\mu^{j-p})$. Since the eigenvalues for different values of j are distinct for small enough μ , and $\mathcal{T}(\mu)$ is self-adjoint, the ranges of $I - P^{(j)}(\mu)$ for different values of j are orthogonal to each other. This implies that the $I - P^{(j)}(\mu)$ for different j commute with each other, and hence so do the $P^{(j)}(\mu)$. Since $I - \mathcal{P}(\mu)$ is the orthogonal projection onto the union over $0 \leq j \leq p - 1$ of the eigenspaces of $\mathcal{T}(\mu)$ whose eigenvalues are of size $O(\mu^{j-p})$, and those eigenspaces are orthogonal for distinct j ,

$$I - \mathcal{P}(\mu) = \sum_{j=0}^{p-1} (I - P^{(j)}(\mu)). \tag{4.8}$$

In addition, since the $(I - P^{(j)}(\mu))$ project onto mutually orthogonal subspaces, $(I - P^{(j_1)}(\mu))(I - P^{(j_2)}(\mu)) = 0$ for $j_1 \neq j_2$, which implies that

$$\begin{aligned} \prod_{j=0}^{p-1} P^{(j)}(\mu) &= \prod_{j=0}^{p-1} (I - (I - P^{(j)}(\mu))) \\ &= I - \sum_{j=0}^{p-1} (I - P^{(j)}(\mu)) \\ &\quad + \text{terms with at least two distinct factors } (I - P^{(j_k)}(\mu)) \\ &= I - \sum_{j=0}^{p-1} (I - P^{(j)}(\mu)). \end{aligned} \tag{4.9}$$

Since the eigenvalues of size $O(1)$ of $T^{(j)}(\mu)$ are the perturbations of the nonzero eigenvalues of $T^{(j,j)}$, the continuity of the projections $P^{(j)}(\mu)$ shows that as μ tends to zero the orthogonal projection $I - P^{(j)}(\mu)$ onto the direct sum of the eigenspaces of eigenvalues of $T^{(j)}(\mu)$ that are $O(1)$ tends to the orthogonal projection $I - P^{(j)}$ onto the direct sum of the eigenspaces of eigenvalues of $T^{(j,j)}$ that are nonzero. This shows that

$$P^{(j)}(\mu) \rightarrow P^{(j)} \quad \text{as } \mu \rightarrow 0 \tag{4.10}$$

in the strong (finite-dimensional) operator topology, which is isometric to a suitably normed matrix space. Therefore, taking the limit of (4.8), (4.9) and rearranging yields (4.3). Taking the limit of the identities $P^{(j)}(\mu)P^{(k)}(\mu) = P^{(k)}(\mu)P^{(j)}(\mu)$ yields $P^{(j)}P^{(k)} = P^{(k)}P^{(j)}$, and the orthogonality of the ranges of $I - P^{(j)}(\mu)$ imply the orthogonality of the ranges of $I - P^{(j)}$. This also shows that $P^{(j)}$ is the orthogonal projection onto the null space of $T^{(j,j)}$, where the $T^{(j,j)}$ are the first terms in the expansions (4.4).

Since $\mathcal{P}(\mu)$ is the orthogonal projection onto the direct sum of the eigenspaces of $\mathcal{T}(\mu)$ that are $O(1)$ or $o(1)$, continuing the reduction process one more step yields (4.5).

The formulas for the $T^{(j,j)}$ are obtained by using recursively formula [6, (2.18) in §II.2.2], which in our notation becomes, for the case here in which there are no nilpotents,

$$T^{(j+1,j+n)} = - \sum_{r=1}^n (-1)^r \sum_{\substack{\sum_{\ell=1}^r v_\ell = n \\ \sum_{\ell=1}^{r+1} k_\ell = r-1 \\ v_\ell \geq 1, k_\ell \geq 0}} S^{(j,k_1)} T^{(j,j+v_1)} S^{(j,k_2)} \dots S^{(j,k_r)} T^{(j,j+v_r)} S^{(j,k_{r+1})}, \quad (4.11)$$

$$S^{(j,0)} := -P^{(j)}, \quad S^{(j,\ell)} := \left(\left(T^{(j,j)} \right)_\psi^{-1} \right)^\ell \text{ for } \ell \geq 1. \quad (4.12)$$

In particular, for $j = 0$ and $n = 1$ only the term with $r = 1$ is present in the outer sum in (4.11), and the inner sum then contains only the case where $v_1 = 1$ and $k_1 = 0 = k_2$. Using (4.12), this yields (4.6). An analogous but longer calculation yields (4.7). \square

- Remark 4.2.* 1. Although $T^{(0,k)} \equiv 0$ for $k \geq 2$, $T^{(j,k)}$ may be nonzero for arbitrarily large values of k when $j \geq 1$.
2. Formula (4.11) shows that in order to calculate $T^{(2,2)}$ it is necessary to first calculate $T^{(1,1)}$ and $T^{(1,2)}$, while in order to calculate $T^{(3,3)}$ it would be necessary to first calculate $T^{(1,j)}$ for $1 \leq j \leq 3$ and then $T^{(2,j)}$ for $2 \leq j \leq 3$.

Example 4.3. 1. In the application of Lemma 4.1 to the convergence theorem the operators $T^{(0,0)}$ and $T^{(0,1)}$ will be individual Fourier modes of the operators \mathcal{L} and \mathcal{M} from (1.1). For example, if $\mathcal{L} = \begin{pmatrix} \partial_x & 0 \\ 0 & 0 \end{pmatrix}$ and $\mathcal{M} = \begin{pmatrix} 0 & \partial_y \\ \partial_y & 0 \end{pmatrix}$ then $T^{(0,0)} = \begin{pmatrix} ik & 0 \\ 0 & 0 \end{pmatrix}$ and $T^{(0,1)} = \begin{pmatrix} 0 & i\ell \\ i\ell & 0 \end{pmatrix}$ for some fixed values of k and ℓ . When $k \neq 0$ the projection onto the null space of $T^{(0,0)}$ is $P^{(0)} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$, and formulas (4.6) and (4.7) yield $T^{(1,1)} = P^{(0)} T^{(0,1)} P^{(0)} = 0$ and $T^{(2,2)} = \begin{pmatrix} 0 & 0 \\ 0 & -\frac{i\ell^2}{k} \end{pmatrix}$ since $P^{(1)} = I$ so $\tilde{P}^{(1)} = P^{(0)} P^{(1)} = P^{(0)}$. If $\ell \neq 0$ then $T^{(0,0)}$ and $T^{(2,2)}$ each have one nonzero eigenvalue so the fact that the matrices are of size 2×2 implies that $T^{(j,j)} = 0$ for $j > 2$, while if $\ell = 0$ then $T^{(j,j)} = 0$ for $j \geq 2$. When $k = 0$ but ℓ is nonzero then $T^{(0,0)} = 0$, $P^{(0)} = I$, $T^{(1,1)} = T^{(0,1)}$, $P^{(1)} = 0$, and $T^{(j,j)} = 0$ for $j > 1$, while when both k and ℓ vanish then, for all j , $T^{(j,j)} = 0$ and $P^{(j)} = I$.

2. The operators \mathcal{L} and \mathcal{M} in (1.1) are allowed to have order zero, that is, to be simply multiplication by fixed matrices, and then the operators in the lemma are simply the same operators. For example,

$$T^{(0,0)} = \mathcal{L} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$T^{(0,1)} = \mathcal{M} = \begin{pmatrix} 0 & 0 & 0 & a & b \\ 0 & 0 & 0 & c & d \\ 0 & 0 & im & 0 & 0 \\ -a & -c & 0 & 0 & 0 \\ -b & -d & 0 & 0 & 0 \end{pmatrix}. \quad (4.13)$$

For these operators,

$$T^{(1,1)} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & im & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{and}$$

$$T^{(2,2)} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & bc - ad \\ 0 & 0 & 0 & ad - bc & 0 \end{pmatrix}.$$

When $ad - bc \neq 0$ then all eigenvalues of $T^{(0,0)} + \mu T^{(0,1)}$ have been accounted for, so $T^{(j,j)} = 0$ for $j > 2$. On the other hand, when $ad - bc = 0$ then zero is an eigenvalue of $T^{(0,0)} + \mu T^{(0,1)}$ with multiplicity two for all μ (with the eigenvectors being $(0\ 0\ 0\ -b\ a)^T$ and $(-c\ \mu\ a\ 0\ -(1+b)a)^T$ when $a \neq 0$), so $T^{(j,j)} = 0$ for $j > 1$.

4.2. Theorem and Proof

The following projections and operator will appear in the statement and proof of the convergence theorem. We assume that either (1.6) or (1.7) holds for some integer $s \geq s_0$.

Let \mathcal{L} and \mathcal{M} be operators satisfying the conditions of Assumption 3.3. Let $\widehat{f}(k)$ denote the Fourier transform of f on \mathbb{R}^d or \mathbb{T}^d and let $(g(k))^\vee$ denote the corresponding inverse Fourier transform of $g(k)$. Since \mathcal{L} and \mathcal{M} are constant-coefficient operators there exist functions $\widehat{\mathcal{L}}(k)$ and $\widehat{\mathcal{M}}(k)$ such that $\widehat{\mathcal{L}f} = \widehat{\mathcal{L}}\widehat{f}$ and $\widehat{\mathcal{M}f} = \widehat{\mathcal{M}}\widehat{f}$.

Definition 4.4. For any k , let $\widehat{\mathbb{P}}(k)$ and $\widehat{\mathbb{P}}(\mu)(k)$ denote the projections $\mathcal{P}(0)$ and $\mathcal{P}(\mu)$, respectively, from Lemma 4.1, where $p = s + 1$ when (1.6) holds or $p = s + 2$ when (1.7) holds, $T^{(0,0)} := \widehat{\mathcal{L}}(k)$, and $T^{(0,1)} := \widehat{\mathcal{M}}(k)$. Define the projection \mathbb{P} by $\mathbb{P}f = (\widehat{\mathbb{P}}(k)\widehat{f}(k))^\vee$, and the projection $\mathbb{P}(\mu)$ by $\mathbb{P}(\mu)f = (\widehat{\mathbb{P}}(\mu)(k)\widehat{f}(k))^\vee$. In addition, when (1.6) holds then let $\widehat{T}_{\text{lim}}(k) = C^s T^{(p,p)}$, where C is the constant from (1.6) and $T^{(p,p)}$ is from Lemma 4.1 with $T^{(0,0)}$, $T^{(0,1)}$, and p as mentioned above, and define the operator T_{lim} by $T_{\text{lim}}f = (\widehat{T}_{\text{lim}}(k)\widehat{f}(k))^\vee$. However, when (1.7) holds then define $T_{\text{lim}} = 0$.

Remark 4.5. Since $\widehat{\mathbb{P}}(k)$ is an orthogonal projection for each k and hence bounded by one, \mathbb{P} is an orthogonal projection on L^2 and a bounded operator on H^s for all s . In contrast, although $\widehat{T}_{\text{lim}}(k)$ is a bounded operator for each k the operator T_{lim} may be unbounded. By Lemma 4.1, $\widehat{T}_{\text{lim}}(k)$ is skew-adjoint for each k so T_{lim} is anti-symmetric.

Theorem 4.6. *Assume that the conditions of Theorem 3.6 hold, that δ and ε tend to zero while obeying either (1.6) or (1.7) for some integer $s \geq s_0$, and that $u_0(x, \delta, \varepsilon)$ converges in H^{s_0+1} to $u_{0,0}(x)$ in that limit.*

Then the solution $u(t, x, \delta, \varepsilon)$ of the initial-value problem (1.1), $u(0, x, \delta, \varepsilon) = u_0(x, \delta, \varepsilon)$ converges to the unique solution $U(t, x)$ belonging to $L^\infty([0, T]; H^{s_0+1}) \cap \text{Lip}([0, T]; L^2)$ of

$$\mathbb{P}\left[A_0(0)U_t + \sum_{j=1}^d A_j(U)U_{x_j} + T_{\text{lim}}U - F(t, x, U)\right] = 0, \tag{4.14}$$

$$(I - \mathbb{P})U = 0 \tag{4.15}$$

$$U(0, x) = u_{0,0}(x), \tag{4.16}$$

where \mathbb{P} is the orthogonal projection operator from Definition 4.4 and T_{lim} is the operator defined there.

Proof. The uniform bound for the $||| \cdot |||_{s_0+1, \varepsilon, A_0}$ norm of the solution of (1.1), proven in Theorem 3.6 shows that $\max_{0 \leq t \leq T} [\|u(t, \cdot)\|_{s_0+1}^2 + \|u_t\|_0^2]^{1/2} \leq 2M$, where T and M are as in Theorem 3.6. By Ascoli’s theorem plus the weak-* compactness of $L^\infty([0, T]; H^{s_0+1})$, for every sequence of values of δ and ε tending to zero while satisfying (3.8) there is a subsequence converging weak-* in $L^\infty([0, T]; H^{s_0+1})$ and strongly in $C^0([0, T]; L^2)$ to a limit $U(t, x)$ in $L^\infty([0, T]; H^{s_0+1}) \cap \text{Lip}([0, T]; L^2)$. In particular, this convergence together with the assumption on the convergence of the initial data show that (4.16) holds.

By interpolation between Sobolev spaces, the convergence and bounds obtained so far imply that the subsequence also converges to U in $C^0([0, T]; H^{s_0+1-\mu})$ for any $\mu > 0$, and hence also in $C^0([0, T]; C^1)$. This yields the convergence in at least L^2 of $A_0(\varepsilon u)u_t + \sum_{j=1}^d A_j(u)u_{x_j} - F(t, x, u)$ to $A_0(0)U_t + \sum_j A_j(U)U_{x_j} - F(t, x, U)$.

Now apply the projection $\mathbb{P}(\mu)$ from Definition 4.4 with $\mu = \frac{\delta}{\varepsilon}$ to the PDE (1.1), which yields

$$\begin{aligned} & \frac{1}{\delta} \mathbb{P}\left(\frac{\delta}{\varepsilon}\right) (\mathcal{L} + \frac{\delta}{\varepsilon} \mathcal{M})u \\ &= -\mathbb{P}\left(\frac{\delta}{\varepsilon}\right) \left[A_0(\varepsilon u)u_t + \sum_j A_j(u)u_{x_j} - F(t, x, u) \right]. \end{aligned} \tag{4.17}$$

As noted above, the expression in brackets on the right side of (4.17) converges in $C^0([0, T]; L^2)$ as δ and ε tends to zero in the manner stated in the theorem. Since $\mu := \frac{\delta}{\varepsilon}$ tends to zero in that limit, the projection $\mathbb{P}(\frac{\delta}{\varepsilon})$ converges in the strong operator topology to \mathbb{P} in that limit since the Fourier transform of the former is uniformly bounded and converges pointwise to the Fourier transform of the latter, so for any $f \in L^2$, $\|[\mathbb{P}(\mu) - \mathbb{P}]f\|_{L^2}^2 = \int \|[\widehat{\mathbb{P}(\mu)}(k) - \widehat{\mathbb{P}}(k)]\widehat{f}(k)\|^2 dk$ (or that expression with the integral replaced by a sum if the spatial domain is periodic) tends to zero by (4.10), the Dominated Convergence Theorem and the fact that orthogonal projection operators do not increase vector length. Hence the entire right side of

(4.17) converges in the above limit to $\mathbb{P} \left[A_0(0)U_t - \sum_j A_j(U)U_{x_j} - F(t, x, U) \right]$. This implies that the left side of (4.17) also converges.

When (1.6) holds then that relation plus the definition $p = s + 1$ from Definition 4.4 imply that $\frac{\mu^p}{\delta} = C^{p-1}(1 + o(1))$. Hence Lemma 4.1 shows that

$$[\mathbb{P}(\mu)(\frac{1}{\delta}(\mathcal{L} + \mu\mathcal{M})f)^\wedge = C^{p-1}T^{(p,p)}\widehat{f} + o(1) = \widehat{T_{\text{lim}}f} + o(1). \tag{4.18}$$

Although the Fourier transform of T_{lim} may be unbounded as a function of the Fourier transform variable, (4.18) together with the convergence of u to U shows that the Fourier transform of the left side of (4.17) converges pointwise to the Fourier transform of $T_{\text{lim}}U$. The fact that that left side is known to converge in L^2 implies that its Fourier transform also converges in L^2 . Since the pointwise and L^2 limits of a sequence of functions must coincide when both exist, the Fourier transform of the left side of (4.17) tends in L^2 to the Fourier transform of $T_{\text{lim}}U$, and hence that left side tends to $T_{\text{lim}}U$. The reduction process also shows that the Fourier transform of T_{lim} is in the image of $\widehat{\mathbb{P}}(k)$ for each k , so rearranging the limit of (4.17) yields (4.14). When (1.7) holds instead of (1.6) then that relation plus the definition $p = s + 2$ from Definition 4.4 imply that $\frac{\mu^p}{\delta} = o(1)$, so (4.18) holds with C replaced by zero, and again leads to (4.14) but with $T_{\text{lim}} = 0$.

Now define $\widehat{\mathcal{T}}(\delta, \varepsilon)(k) := \frac{1}{\delta}(\widehat{\mathcal{L}}(k) + \frac{\delta}{\varepsilon}\widehat{\mathcal{M}}(k))$. From the Fourier transform of (1.1), (1.6) or (1.7), and Lemma 4.1,

$$\left| \left(I - \widehat{\mathbb{P}}\left(\frac{\delta}{\varepsilon}\right) \right) \widehat{u}(k) \right| \leq cz(\delta, \varepsilon) \left| \left(I - \widehat{\mathbb{P}}\left(\frac{\delta}{\varepsilon}\right) \right) \widehat{\mathcal{T}}(\delta, \varepsilon)(k) \left(I - \widehat{\mathbb{P}}\left(\frac{\delta}{\varepsilon}\right) \right) \widehat{u}(k) \right|, \tag{4.19}$$

where $z(\delta, \varepsilon) = \mu = \varepsilon^{\frac{1}{s}}$ by (4.2) when (1.6) holds, and $z(\delta, \varepsilon) = \frac{\delta}{\mu^p} \cdot \mu = \left(\frac{\varepsilon^{1+\frac{1}{s}}}{\delta}\right)^s$ by the definition of $\widehat{\mathcal{T}}(\delta, \varepsilon)$ plus (4.2) and the definition of p in terms of s when (1.7) holds. In the former case $\varepsilon^{\frac{1}{s}}$ clearly tends to zero with ε , and in the latter case $\left(\frac{\varepsilon^{1+\frac{1}{s}}}{\delta}\right)^s$ tends to zero by (1.7), that is, in either case $z(\delta, \varepsilon) \rightarrow 0$ as δ and ε tend to zero. Since

$$\begin{aligned} \left| \left(I - \widehat{\mathbb{P}}\left(\frac{\delta}{\varepsilon}\right) \right) \widehat{\mathcal{T}}(\mu)(k) \left(I - \widehat{\mathbb{P}}\left(\frac{\delta}{\varepsilon}\right) \right) \widehat{u}(k) \right| &= \left| \left(I - \widehat{\mathbb{P}}\left(\frac{\delta}{\varepsilon}\right) \right) \widehat{\mathcal{T}}(\mu)(k) \widehat{u}(k) \right| \\ &\leq \left| \widehat{\mathcal{T}}(\mu)(k) \widehat{u}(k) \right| \\ &= \left| \left(A_0(\varepsilon u)u_t + \sum_{j=1}^d A_j(u)u_{x_j} - F(t, x, u) \right)^\wedge(k) \right| \\ &= O(1), \end{aligned}$$

taking the limit of (4.19) as δ and ε tend to zero while satisfying (1.6) or (1.7) shows that $(I - \widehat{\mathbb{P}}(k))\widehat{U}(k) = 0$, which implies (4.15).

To show that a solution of the given smoothness of (4.14), (4.16) is unique, let \mathbb{Q}_R be the projection onto the Fourier modes for which $\widehat{T_{\text{lim}}}(k)$ is bounded by R . Since $\widehat{T_{\text{lim}}}(k)$ is finite for each k , the limit as $R \rightarrow \infty$ of \mathbb{Q}_R is the identity operator. Since \mathbb{Q}_R is a projection onto Fourier modes it commutes with T_{lim} , and \mathbb{P} . Hence, for any R ,

$$\langle \mathbb{Q}_R U, \mathbb{P} T_{\text{lim}} U \rangle = \langle \mathbb{Q}_R \mathbb{P} U, T_{\text{lim}} U \rangle = \langle \mathbb{Q}_R U, T_{\text{lim}} U \rangle = \langle \mathbb{Q}_R U, T_{\text{lim}} \mathbb{Q}_R U \rangle = 0$$

by (4.15) plus the antisymmetry of T_{lim} . Taking the difference of (4.14) for two solutions $U^{(1)}$ and $U^{(2)}$, defining $U := U^{(1)} - U^{(2)}$, taking the L^2 inner product of the result with $\mathbb{Q}_R U$ and letting R tend to infinity therefore yields

$$\begin{aligned} 0 &= \lim_{R \rightarrow \infty} \left\langle \mathbb{Q}_R U, \mathbb{P} \left(A_0(0) U_t \right. \right. \\ &\quad \left. \left. + \sum_j A_j(U^{(1)}) U_{x_j} + \sum_j \left\{ A_j(U^{(2)} + U) - A_j(U^{(2)}) \right\} U_{x_j}^{(2)} \right. \right. \\ &\quad \left. \left. + T_{\text{lim}} U - \left[F(t, x, U^{(2)} + U) - F(t, x, U^{(2)}) \right] \right) \right\rangle \\ &= \left\langle U, \mathbb{P} \left(A_0(0) U_t + \sum_j A_j(U^{(1)}) U_{x_j} + M(t, x, U^{(2)}, \nabla_x U^{(2)}) U \right) \right\rangle \quad (4.20) \\ &= \left\langle \mathbb{P} U, A_0(0) U_t + \sum_j A_j(U^{(1)}) U_{x_j} + M(t, x, U^{(2)}, \nabla_x U^{(2)}) U \right\rangle \\ &= \left\langle U, A_0(0) U_t + \sum_j A_j(U^{(1)}) U_{x_j} + M(t, x, U^{(2)}, \nabla_x U^{(2)}) U \right\rangle, \end{aligned}$$

where

$$\begin{aligned} &\sum_j \left\{ A_j(U^{(2)} + U) - A_j(U^{(2)}) \right\} U_{x_j}^{(2)} \\ &\quad - \left[F(t, x, U^{(2)} + U) - F(t, x, U^{(2)}) \right] \\ &= \int_0^1 \frac{d}{ds} \sum_j A_j(U^{(2)} + sU) U_{x_j}^{(2)} - F(t, x, U^{(2)} + sU) ds \\ &= \left\{ \int_0^1 \sum_j \nabla_v \left[A_j(v) U_{x_j}^{(2)} - F(t, x, v) \right] \Big|_{v=U^{(2)}+sU} ds \right\} U \\ &:= M(t, x, U^{(2)}, \nabla_x U^{(2)}) U. \end{aligned}$$

Since the final expression in (4.20) looks like the L^2 estimate for a symmetric hyperbolic system, we obtain

$$\begin{aligned}
 0 &= \frac{d}{dt} \left(\frac{1}{2} \langle U, A_0(0)U \rangle \right) \\
 &\quad - \left\langle U, \left\{ \sum_j \partial_{x_j} A_j(U^{(1)}) + M(t, x, U^{(2)}, \nabla_x U^{(2)}) \right\} U \right\rangle \\
 &\geq \frac{d}{dt} \left(\frac{1}{2} \langle U, A_0(0)U \rangle \right) - K_1 \langle U, U \rangle \\
 &\geq \frac{d}{dt} \left(\frac{1}{2} \langle U, A_0(0)U \rangle \right) - K_2 \left(\frac{1}{2} \langle U, A_0(0)U \rangle \right)
 \end{aligned} \tag{4.21}$$

for some K_1 and K_2 depending on the $\| \cdot \|_{H^{s_0+1}}$ norms of $U^{(1)}$ and $U^{(2)}$ and the constant c_0 from (3.3). Estimate (4.21) plus the initial condition (4.16) imply that

$$\frac{1}{2} \langle U, A_0(0)U \rangle \leq \left(\frac{1}{2} \langle U(0), A_0(0)U(0) \rangle \right) e^{kt} = 0,$$

which implies that $U \equiv 0$, that is, $U^{(1)} \equiv U^{(2)}$, yielding uniqueness.

As usual, the uniqueness of the limit implies that convergence holds as δ and ε tend to zero while satisfying (1.6) or (1.7) without restricting to a subsequence. \square

Example 4.7. 1. Consider the PDE

$$\begin{pmatrix} u \\ v \end{pmatrix}_t + \frac{1}{\varepsilon^2} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}_x + \frac{1}{\varepsilon} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = 0. \tag{4.22}$$

The relationship $\delta = \varepsilon^2$ does not satisfy (1.3) in dimension two. Nevertheless, as noted in the introduction, the fact that the coefficient matrix of the time derivatives does not depend on u or v implies that solutions of (4.22) satisfy uniform bounds. Let $f(x, y)$ be a function whose gradient belongs to H^3 , and take the initial data to be $u(0, x, y) = u_0(x, y) := -\varepsilon f_y$ and $v(0, x, y) = v_0(x, y) := f_x$. Then $u_t(0, x, y) = 0$ and $v_t(0, x, y) = f_{yy}$, that is, the initial time derivative is bounded. Since the PDE is linear with constant coefficients, it is convenient to express the limit equation in Fourier space. By part 1 of Example 4.3, when $k \neq 0$ then the limit is $\widehat{U}(t, k, \ell) = 0$, $\widehat{V}_t - \frac{i\ell^2}{k} \widehat{V} = 0$, while for $k = 0$ but $\ell \neq 0$ the limit is $\widehat{U}(t, 0, \ell) = 0 = \widehat{V}(t, 0, \ell)$ and for $k = 0 = \ell$ the limit is $\widehat{U}_t(t, 0, 0) = 0 = \widehat{V}_t(t, 0, 0)$. The initial data for the limit are $\widehat{U}(0, k, \ell) = 0$ and $\widehat{V}(0, k, \ell) = ik\widehat{f}(k, \ell)$. When k and ℓ are both nonzero the solution of the limit equation is $\widehat{U}(t, k, \ell) = 0$ and

$$\widehat{V}(t, k, \ell) = ik e^{i\frac{\ell^2}{k}t} \widehat{f}(k, \ell), \tag{4.23}$$

while when $k = 0$ then the limit is $\widehat{U} = 0 = \widehat{V}$. When the spatial domain is \mathbb{R}^2 then $\widehat{T}_{\text{lim}}(k) = \frac{-i\ell^2}{k}$ is unbounded but when the domain is \mathbb{T}^2 then it is bounded since $|k| \geq c$ on the set where it is nonzero. Even when the spatial domain is \mathbb{R}^2 , the fact that $\widehat{V}(t, k, \ell)$ contains a factor of k ensures that \widehat{V}_t is bounded, but \widehat{V}_{tt} will be unbounded if $\widehat{f}(0, 0) \neq 0$. The limit solution (4.23),

which implies the limit equation satisfied by V , can be verified by solving the equation for \widehat{U} and \widehat{V} exactly for $k \neq 0$. This yields

$$\widehat{V} = ik\widehat{f}(k, \ell) \frac{e^{i \frac{(-k+\sqrt{k^2+4\varepsilon^2\ell^2})_t}{2\varepsilon^2}} (k+\sqrt{k^2+4\varepsilon^2\ell^2}) - e^{i \frac{(-k-\sqrt{k^2+4\varepsilon^2\ell^2})_t}{2\varepsilon^2}} (k-\sqrt{k^2+4\varepsilon^2\ell^2})}{2\sqrt{k^2+4\varepsilon^2\ell^2}} + O(\varepsilon),$$

whose limit as $\varepsilon \rightarrow 0$ indeed yields (4.23).

2. Adding the term $-\alpha \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}_y$ to (4.22) changes the limit equation for V to $\widehat{V}_t - \frac{i\ell^2}{k-\alpha\ell} \widehat{V} = 0$. If α is irrational but well-approximated by rationals then the term $\widehat{T}_{\text{lim}}(k) = \frac{-i\ell^2}{k-\alpha\ell}$ may not be bounded by $(|k|+|\ell|)^3$ as that expression tends to infinity, even in the periodic case, so T_{lim} may not be a bounded operator from H^3 to L^2 .
3. Consider the PDE $u_t + u_x + \frac{1}{\delta} \mathcal{L}u + \frac{1}{\varepsilon} \mathcal{M}u = 0$, where $\delta = \varepsilon^{3/2}$ and \mathcal{L} and \mathcal{M} are the matrices discussed in Part 2 of Example 4.3. Since the choice of the relationship between δ and ε makes s in (1.6) equal two and hence p in Definition 4.4 equal three, the projection \mathbb{P} is orthogonal to the non-zero eigenspace of $T^{(2,2)}$ as well as those of $T^{(0,0)}$ and $T^{(1,1)}$. The formula for $T^{(2,2)}$ in Part 2 of Example 4.3 therefore shows that when $m \neq 0$ and $ad - bc \neq 0$ then the limit equation is simply $U = 0$ while when $m \neq 0$ but $ad - bc = 0$ then the limit equation is that the first three components of U vanish and $\partial_t + \partial_x$ of its last two components equal zero. This shows that even the number of nonzero components of the limit cannot be determined simply by looking at the number of components that do not contain large terms nor even by first eliminating all components having terms of order $\frac{1}{\delta}$ and then eliminating those remaining components having terms of order $\frac{1}{\varepsilon}$ not coming from components already eliminated, which works for the system (4.22).

Acknowledgements. Ju is supported by the NSFC (Grants Nos. 11571046, 11471028, 11671-225). Cheng and Ju are supported by the UK Royal Society ‘‘International Exchanges’’ scheme (Award No. IE150886). Ju and Schochet are supported by the ISF-NSFC joint research program (NSFC Grant No. 11761141008 and ISF Grant No. 2519/17). The authors thank the anonymous referee for comments that lead to improvements in the exposition of the paper.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. BROWNING, G., KREISS, H.-O.: Problems with different time scales for nonlinear partial differential equations. *SIAM J. Appl. Math.*, **42**(4), 704–718, 1982
2. CHENG, B.: Improved accuracy of incompressible approximation of compressible Euler equations. *SIAM J. Math. Anal.*, **46**(6):3838–3864, 2014

3. FRIEDMAN, A.: *Partial Differential Equations*. Krieger, Huntington, NY, 1976
4. GALLAGHER, I.: Applications of Schochet's methods to parabolic equations. *J. Math. Pures Appl.*, **77**(10), 989–1054, 1998
5. GRENIER, E.: Pseudo-differential energy estimates of singular perturbations. *Commun. Pure Appl. Math.*, **50**(9), 821–865, 1997.
6. KATO, T.: *A Short Introduction to Perturbation Theory for Linear Operators*. Springer, New York, 1982
7. KLAINERMAN, S., MAJDA, A.: Singular limits of quasilinear hyperbolic systems with large parameters and the incompressible limit of compressible fluids. *Commun. Pure Appl. Math.*, **34**, 481–524, 1981.
8. KLAINERMAN, S., MAJDA, A.: Compressible and incompressible fluids. *Commun. Pure Appl. Math.*, **35**, 629–653, 1982
9. MAJDA, A.: *Compressible fluid flow and systems of conservation laws in several space variables*, Vol. 53 *Applied Mathematical Sciences*. Springer, New York, 1984
10. MAJDA, A., KLEIN, R.: Systematic multiscale models for the tropics. *J. Atmos. Sci.*, **60**(393–408), 2003
11. MÉTIVIER, G., SCHOCHET, S.: The incompressible limit of the non-isentropic euler equations. *Arch. Ration. Mech. Anal.*, **158**, 61–90, 2001
12. SCHOCHET, S.: Asymptotics for symmetric hyperbolic systems with a large parameter. *J. Differ. Equ.*, **75**, 1–27, 1988
13. SCHOCHET, S.: Fast singular limits of hyperbolic PDEs. *J. Differ. Equ.*, **114**(2), 476–512, 1994

B. CHENG

Department of Mathematics,
University of Surrey,
Guildford,
GU2 7XH, UK.
e-mail: b.cheng@surrey.ac.uk.

and

Q. JU

Institute of Applied Physics and Computational Mathematics,
P.O. Box 8009,
Beijing, 100088
China.
e-mail: ju_qiangchang@iapcm.ac.cn

and

S. SCHOCHET

School of Mathematical Sciences,
Tel-Aviv University, 69978 Tel Aviv, Israel.
e-mail: schochet@post.tau.ac.il

(Received April 25, 2017 / Accepted February 5, 2018)

Published online March 7, 2018

© The Author(s) (2018)

This article is an open access publication