CHAPTER 1

Introduction

As we described in the Preface, the purpose of this chapter is to provide some historical background and introductory material. The remaining chapters in the monograph are independent of the present one. Our main goal here is to set the stage for our study of the difficult problem of shock formation in more than one spatial dimension. Readers may also consult the companion survey article [**28**] for additional introductory material.

1.1. Shock formation in one spatial dimension

In Sect. 1.1, we provide a brief overview of shock formation in one spatial dimension. We give some historical background and provide two proofs that shocks often form in solutions to Burgers' equation. The second proof is sharper and has important philosophical parallels with our later study of shock formation in three spatial dimensions. For illustration, we also provide a proof of blow-up for a model quasilinear wave equation under the assumption of plane symmetry.[1] To this end, we embed the equation into the theory of 2×2 strictly hyperbolic genuinely nonlinear systems and provide a standard proof of blow-up for such systems.

1.1.1. A very brief history of shock formation in one spatial dimension. The development of the modern theory of shock waves in solutions to nonlinear hyperbolic PDEs has a rich history filled with false starts and tantalizing turns. For fascinating descriptions of the events leading to the modern theory, we refer the reader to the introduction of [**22**] as well as the survey article [**72**]. Here we only describe the historical results that are most directly connected to the results of the present monograph.

The earliest known observation of shock formation[2] was made by the British physicist/astronomer James Challis [**11**] in his study of the evolution of an ideal gas in one spatial dimension under the isothermal equation of state $p = c_s^2 \rho$, where p is the pressure, ρ is the density, and the constant $c_s > 0$ is the speed of sound. More precisely, Challis considered the compressible Euler equations for ρ and the velocity v:

$$\partial_t \rho + \partial_x (\rho v) = 0, \tag{1.1a}$$

$$\partial_t (\rho v) + \partial_x (\rho v^2) + c_s^2 \partial_x \rho = 0 \tag{1.1b}$$

[1]By plane symmetry, we mean that the solution depends only on a time variable t and a single real spatial variable x.

[2]In PDE literature, shock formation is also known as *wave breaking*.

in conjunction with Bernoulli's equation[3] for the fluid potential Φ, which verifies $v = \partial_x \Phi$:

$$\partial_t \Phi + \frac{1}{2}(\partial_x \Phi)^2 + c_s^2 \ln \rho = 0. \tag{1.2}$$

Forty years earlier, Poisson [**69**] had shown that as a consequence of (1.1a)-(1.2), Φ verifies the following[4] closed equation:

$$\partial_t^2 \Phi + 2(\partial_x \Phi)\partial_x \partial_t \Phi + (\partial_x \Phi)^2 \partial_x^2 \Phi - c_s^2 \partial_x^2 \Phi = 0. \tag{1.3}$$

Poisson also showed that any function Φ verifying an identity of the form

$$\partial_x \Phi = f\left(x + (c_s - \partial_x \Phi)t\right), \tag{1.4}$$

where f is a smooth function, is a solution[5] to (1.3). Note that (1.4) implies that f is the initial condition of $\partial_x \Phi$.

Challis's key observation was that in the case $f(x) = -\sin\left(\frac{\pi}{2}x\right)$, the solution to (1.4) verifies $\partial_x \Phi = 0$ along the line $x = -c_s t$ and $\partial_x \Phi = 1$ along the line $x = -1 - (c_s - 1)t$. Since these lines intersect at the point $(t, x) = (1, -c_s)$, one concludes that *the classical solution must break down there.*

1.1.2. Shock formation in solutions to Burgers' equation and the method of characteristics. Results in the spirit of Challis' blow-up result of Sect. 1.1.1 have been derived for many nonlinear hyperbolic PDEs and initial conditions in one spatial dimension.[6] Here is a far-from-exhaustive collection of examples: Riemann's foundational work [**71**] on the compressible Euler equations in which he invented the method of Riemann invariants, Lax's work [**55**] on scalar conservation laws, his use of Riemann invariants in his study [**54**] of 2×2 strictly hyperbolic genuinely nonlinear systems (see Sect. 1.1.3), Jeffrey's work [**31**] on magnetoacoustics, Jeffrey-Korobeinikov's work [**33**] on nonlinear electromagnetism, Jeffrey-Teymur's work [**32**] on hyperelastic solids, John's extension [**35**] of Lax's work to a larger class of systems in which the method of Riemann invariants cannot be used, Liu's further refinement [**60**] of John's work, John's work [**37**] on spherically symmetric

---

[3] From straightforward computations based on the assumption $v = \partial_x \Phi$ and equations (1.1a)-(1.1b), we find that $\partial_x \left\{\partial_t \Phi + \frac{1}{2}(\partial_x \Phi)^2 + c_s^2 \ln \rho\right\} = 0$. Thus, $\partial_t \Phi + \frac{1}{2}(\partial_x \Phi)^2 + c_s^2 \ln \rho$ is a function of t, which we have set equal to 0 in (1.2).

[4] Equation (1.3) is the Euler-Lagrange equation for Φ. More precisely, in one spatial dimension, and more generally for irrotational solutions in higher dimensions, the compressible Euler equations are equivalent to a quasilinear wave equation for Φ that can written in Euler-Lagrange form. The Lagrangian is $\mathcal{L} = p$, where the pressure p is viewed as a function of the spacetime gradient of Φ; see, for example, [**17**] and [**21**] for further details. For the equation of state $p = c_s^2 \rho$ in one spatial dimension, equation (1.2) implies that $p = c_s^2 \exp\left(-c_s^{-2}\left\{\partial_t \Phi + \frac{1}{2}(\partial_x \Phi)^2\right\}\right)$. The Euler-Lagrange equation is $\partial_t\left(\frac{\partial \mathcal{L}}{\partial(\partial_t \Phi)}\right) + \partial_x\left(\frac{\partial \mathcal{L}}{\partial(\partial_x \Phi)}\right) = 0$, and when $p = c_s^2 \rho$, it can be written in the form (1.3).

[5] To derive the family of solutions (1.4), one can first compute that the function $v + c_s \ln \rho$ is a Riemann invariant (see Sect. 1.1.3) for the system (1.1a)-(1.1b). A special class of solutions is such that this Riemann invariant is constant, which implies that $\partial_x(v + c_s \ln \rho) = 0$. Using this identity and equations (1.1a)-(1.1b), we deduce that these special solutions verify the equation $\partial_t v + (v - c_s)\partial_x v = 0$. This equation, when supplemented with the initial condition $v|_{t=0} = f$, has the solution $v = f\left(x + (c_s - v)t\right)$. Recalling that $v = \partial_x \Phi$, we arrive at Poisson's solutions (1.4).

[6] Under the umbrella of "one spatial dimension," we include problems that are effectively one dimensional, such as the study of spherically symmetric solutions in higher dimensions.

solutions to the equations of elasticity, Klainerman-Majda's work [**49**] on nonlinear vibrating string equations, Bloom's work [**10**] on nonlinear electrodynamics, and Cheng-Young-Zhang's work [**13**] on magnetohydrodynamics and related systems.

In the present section, we use the model case of Burgers' equation to illustrate the essential features of modern proofs of blow-up. We provide two proofs. The first is standard and is effectively a proof by contradiction showing that nontrivial compactly supported (in the spatial variable) smooth solutions cannot be continued indefinitely. The second is much more refined, shows that the singularity is a shock, and is more closely aligned with our later proof of shock formation for solutions to wave equations in three spatial dimensions.

In the following discussion, (t, x) are standard rectangular coordinates on $\mathbb{R}^2$ and ∂_t, ∂_x are the corresponding coordinate partial derivatives. The Cauchy problem for Burgers' equation in one spatial dimension with unknown $\Psi(t, x)$ and initial data $\mathring{\Psi}$ is:

$$\partial_t \Psi + \Psi \partial_x \Psi = 0, \tag{1.5a}$$

$$\Psi(0, x) = \mathring{\Psi}(x). \tag{1.5b}$$

We first present the standard crude analysis, which shows that the solution blows up in finite time but provides limited information about the true nature of the singularity. The argument is based on the method of characteristics, which by definition are the solutions $\gamma(t; x) := (\gamma^0(t; x), \gamma^1(t; x))$ of the following ODE initial value problem:

$$\frac{d}{dt}\gamma^0(t; x) = 1, \qquad \frac{d}{dt}\gamma^1(t; x) = \Psi \circ \gamma(t; x), \tag{1.6a}$$

$$\gamma^0(0; x) = 0, \qquad \gamma^1(0; x) = x. \tag{1.6b}$$

Note that $\{\gamma(\cdot; x)\}_{x \in \mathbb{R}}$ is a set of curves in the plane parametrized by x. To exhibit the blow-up of smooth solutions, we first differentiate (1.5a) with ∂_x and use equation (1.6a) to deduce that along the characteristics, we have

$$\frac{d}{dt}\left\{(\partial_x \Psi) \circ \gamma(t; x)\right\} = -\left\{(\partial_x \Psi) \circ \gamma(t; x)\right\}^2. \tag{1.7}$$

Note that for each fixed x, (1.7) is a Riccati-type[7] ODE for $(\partial_x \Psi) \circ \gamma(\cdot; x)$. In fact, (1.7) is equivalent to $\frac{d}{dt}\left\{\frac{1}{(\partial_x \Psi) \circ \gamma(t; x)}\right\} = 1$. In view of the initial conditions (1.6b), it follows that if $(\partial_x \mathring{\Psi})(x_0) < 0$, then $(\partial_x \Psi) \circ \gamma(t; x_0)$ must blow up by the time $t = -1/(\partial_x \mathring{\Psi})(x_0)$. Note that all nontrivial compactly supported data must lead to blow-up because such data always have at least one point x_0 with $\partial_x \mathring{\Psi}(x_0) < 0$.

One criticism of the above argument is that in its present form, it is actually a proof by contradiction showing that smooth solutions do not persist for all time. That is, additional arguments must be given in order to guarantee that $\partial_x \Psi$ is the true quantity that blows up first and that no other kind of singularity occurs before that. Of course, in the case of Burgers' equation, the additional arguments are simple; we leave the details to the reader. A major difficulty that we must overcome in the present monograph is that the crude proof of blow-up by contradiction presented above does not seem to generalize to quasilinear wave equations in more than

[7] The standard Riccati ODE is $\dot{y} = y^2$.

one spatial dimension. The main reason is that in two or more spatial dimensions, one needs to supplement the method of characteristics with energy estimates (see Sect. 1.2.1). In order to close the energy estimates in the shock formation problem, we must obtain a detailed description of the dynamics near the shock going far beyond a crude description; see Sect. 2.10.4 for an outline of the argument, which is quite involved, even at the heuristic level.

In view of the limitations of the above crude proof of blow-up for solutions to Burgers' equation, we now give a sharper proof that serves as a slightly more realistic caricature of the way in which we prove the main shock formation results of the monograph. Our argument here is based on a version of Christodoulou's geometric framework [**17**], which applies in three spatial dimensions and is the framework that we use throughout most of the monograph. More precisely, because our analysis here involves only one spatial dimension, we use only a small portion of Christodoulou's framework in the present section, a portion that is essentially equivalent to Majda's geometric approach [**61**] to proving shock formation in one spatial dimension. We remark that Majda's work has its roots in the geometric approach of Keller-Ting [**43**]. Our argument here is also similar to the one given by Alinhac in [**3**].

We begin by constructing the key ingredient, which is a new dynamic coordinate $u = u(t,x)$ obtained by solving the following Cauchy problem for a transport equation:

$$\frac{d}{dt}\{u \circ \gamma(t;x)\} = \partial_t u + \Psi \partial_x u = 0, \tag{1.8a}$$

$$u(0,x) = u \circ \gamma(0;x) = x. \tag{1.8b}$$

In (1.8a)-(1.8b), the characteristics γ are the solutions to (1.6a)-(1.6b). The variable u, which we call an *eikonal function*, is an analog of the well-known Lagrangian coordinates that are often used in fluid mechanics. The main idea of the analysis is to study Burgers' equation relative to the coordinate system (t,u), which will reveal important structural features that are not visible relative to the coordinates (t,x). Note that[8]

$$\frac{\partial}{\partial t}|_u = \partial_t + \Psi \partial_x \tag{1.9}$$

and that $\frac{\partial}{\partial t}|_u$ corresponds to differentiation along the characteristics. The most important property of the coordinates (t,u) is that relative to them, Burgers' equation (1.5a) becomes a *linear* PDE:

$$\frac{\partial}{\partial t}|_u \Psi = 0. \tag{1.10}$$

It follows from (1.10) that when the data are smooth, Ψ *and its derivatives of all orders* with respect to $\frac{\partial}{\partial t}|_u$ and $\frac{\partial}{\partial u}|_t$ remain finite for all time. In fact, all of these quantities are constant in t at fixed u. Thus, a singularity can develop in Ψ relative to the (t,x) coordinates *only if the change of variables map from* (t,u) *coordinates to* (t,x) *coordinates becomes degenerate.*

[8]Throughout this section, $\frac{\partial}{\partial t}|_u$ denotes the partial derivative with respect to t at fixed u, and $\frac{\partial}{\partial u}|_t$ has the analogous meaning.

To expand upon the ideas from the previous paragraph, we introduce the *inverse foliation density* $\upmu$, which we define to be

$$\upmu := \frac{1}{\partial_x u}. \tag{1.11}$$

The function $1/\upmu$ measures the density of the level sets of u relative to the lines $\{x = \text{const}\}$. By (1.8b), $\upmu$ is initially 1, and shock formation (that is, the intersection of the characteristics) exactly corresponds to infinite density of the level sets, or equivalently, $\upmu \to 0$. It is not immediately clear that there should be a connection between $\upmu \to 0$ and the blow-up of $\partial_x \Psi$. There is in fact a strong connection, and to illustrate it, we need only to use the chain rule, (1.10), and (1.11) to deduce that

$$\partial_x \Psi = \frac{1}{\upmu} \frac{\partial}{\partial u}|_t \Psi. \tag{1.12}$$

Thus, as long as $\frac{\partial}{\partial u}|_t \Psi \neq 0$, we see that $\partial_x \Psi$ blows up $\iff$ $\upmu \to 0$. Moreover, as we show in the next paragraph, the vanishing of $\upmu$ is exactly tied to having a solution with $\frac{\partial}{\partial u}|_t \Psi < 0$.

To show that shocks indeed form, we first use (1.8a), (1.9), (1.11), and (1.12) to compute that $\upmu$ verifies the following evolution equation, where $[\partial_x, \frac{\partial}{\partial t}|_u]$ denotes the commutator of ∂_x and $\frac{\partial}{\partial t}|_u$:

$$\frac{\partial}{\partial t}|_u \upmu = -\upmu^2 \frac{\partial}{\partial t}|_u \partial_x u = \upmu^2 [\partial_x, \frac{\partial}{\partial t}|_u] u = \upmu \partial_x \Psi = \frac{\partial}{\partial u}|_t \Psi. \tag{1.13}$$

To obtain information about the source term $\frac{\partial}{\partial u}|_t \Psi$ on the right-hand side of (1.13), we simply recall that $\frac{\partial}{\partial u}|_t \Psi$ is constant in t at fixed u. Thus, if $\frac{\partial}{\partial u}|_t \Psi$ is initially negative, we conclude from (1.13) that $\upmu$ will necessarily vanish in finite time along the corresponding characteristic, and by (1.12), $\partial_x \Psi$ will blow-up like $1/\upmu$. In summary, our second proof of blow-up is sharper than the first because *we have identified that the vanishing of the geometric quantity $\upmu$ is the precise condition causing blow-up and because the solution remains regular in (t, u) coordinates.*

REMARK 1.1 (**$\upmu$ plays a key role in the monograph**). To show that shocks form in the wave equations of interest in the present monograph, the main object of study is a quantity that is analogous to the one defined in (1.11) (see (2.27)). In fact, this monograph is primarily about the extension of the above argument to two classes of quasilinear wave equations in three spatial dimensions and the many additional complications that arise in the presence of angular derivatives and dispersion.

We end this section by highlighting another way to think about the dynamics of solutions Ψ to Burgers' equation: Ψ remains regular along the directions tangent to the characteristics but become singular in the transversal directions. As we will see, these basic features are also partially present in the shock forming wave equation solutions that we study later in the monograph. In fact, these features play a critically important role in the analysis.

1.1.3. Two by two strictly hyperbolic genuinely nonlinear systems and their connection to plane symmetric quasilinear wave equations. As we described in the Preface, the main result of this monograph is that shock formation often occurs in small-data solutions to two classes of quasilinear wave equations in three spatial dimensions. A simple example of an equation on $\mathbb{R}^{1+3}$ to which our main result applies is

$$-\partial_t^2\Phi + c^2\Delta\Phi = 0, \tag{1.14}$$

where the factor $c = c(\partial_t\Phi) > 0$ is a smooth function of $\partial_t\Phi$ verifying the following structural assumptions:

$$c(0) = 1, \qquad c'(0) \neq 0. \tag{1.15}$$

In (1.14), Δ denotes the standard Laplacian on $\mathbb{R}^3$, and in (1.15), $c'(p) = \frac{d}{dp}c(p)$. The proof of shock formation for solutions to (1.14) is incredibly more complicated than the proofs of blow-up for Burgers' equation solutions given in Sect. 1.1.2. Not surprisingly, under the assumption of plane symmetry (that is, that the solution depends only on $(t,x) \in \mathbb{R}^{1+1}$), the proof drastically simplifies because we do not need to derive energy estimates and because plane symmetric solutions do not exhibit dispersion. However, the proof still has some new features not found in either of the proofs of blow-up that we gave for Burgers' equation. For this reason, we provide a proof of plane-symmetric blow-up in this section. Our proof is a standard adaption of the first (crude) proof (by contradiction) of blow-up for Burgers' equation given in Sect. 1.1.2. We note that is possible to give a sharper proof of plane-symmetric shock formation in the spirit of our second proof of blow-up for Burgers' equation; one could proceed by making straightforward modifications to the sharp proof of shock formation for spherically symmetric solutions to quasilinear wave equations in three spatial dimensions given in the companion survey article [**28**].

A simple way to prove blow-up for plane symmetric solutions to (1.14) is to show that under the assumptions (1.15), if $\partial_t\Phi$ is sufficiently small, then (1.14) is equivalent to a member of an important class of PDEs known as 2×2 *strictly hyperbolic genuinely nonlinear systems*; we will verify the equivalence below. In [**54**], Lax proved the first general blow-up results for such systems by extending, with the help of Riemann invariants, the method of characteristics that we used in Sect. 1.1.2 in our crude proof of blow-up for solutions to Burgers' equation. John later developed an approach that allowed him to extend [**35**] Lax's blow-up results to a larger class of systems in one spatial dimension in which the method of Riemann invariants cannot be used. Moreover, in the case of the nonlinear vibrating string equations in one spatial dimension, Klainerman and Majda showed [**49**] that the genuine nonlinearity condition is not necessary.

We now provide a proof of Lax's blow-up results [**54**] and show that it implies blow-up for plane symmetric solutions to (1.14). In the remainder of this section, lowercase Latin indices take on the values 1 and 2, lowercase Greek indices take on the values 0 and 1, and we use Einstein's summation convention. Moreover, we use the notation $(x^0, x^1) = (t, x)$, $\partial_0 = \partial_t$, and $\partial_1 = \partial_x$.

We will study the following Cauchy problem for a 2×2 nonlinear hyperbolic system:

$$\partial_t U_j + M_j^a(U)\partial_x U_a = 0, \qquad (j = 1, 2), \tag{1.16a}$$

$$U_j|_{t=0} = \mathring{U}_j. \tag{1.16b}$$

We say that (1.16a) is a *strictly hyperbolic system* on the domain $\mathcal{U} \subset \mathbb{R}^2$ if for every $U \in \mathcal{U}$, the 2×2 matrix $M_j^i(U)$ has two distinct eigenvalues $\lambda_1(U) < \lambda_2(U)$.

Under the assumption of strict hyperbolicity, for $i = 1, 2$, we let $r^{(i)}$ and $l_{(i)}$ respectively denote the Euclidean-unit length right eigencovector and the Euclidean-unit length left eigenvector (which are unique up to an overall sign) corresponding to λ_i. In particular, these quantities verify the following systems of equations (with no summation over i):

$$M_b^a r_a^{(i)} = \lambda_i r_b^{(i)}, \qquad (b, i = 1, 2), \tag{1.17a}$$

$$M_a^b l_{(i)}^a = \lambda_i l_{(i)}^b, \qquad (b, i = 1, 2). \tag{1.17b}$$

We use the following critically important identity in our subsequent analysis:

$$r_a^{(i)} l_{(j)}^a = 0, \qquad (i \neq j). \tag{1.18}$$

To derive (1.18), we contract equation (1.17a) against $l_{(j)}^b$ and use equation (1.17b) to deduce the identity $\lambda_j r_a^{(i)} l_{(j)}^a = \lambda_i r_a^{(i)} l_{(j)}^a$ (with no summation over i or j). The desired result (1.18) follows from this identity and the strict hyperbolicity assumption.

We say that the strictly hyperbolic system (1.16a) is a *genuinely nonlinear system* on the domain $\mathcal{U} \subset \mathbb{R}^2$ if for every $U \in \mathcal{U}$ and $i = 1, 2$, we have (with no summation over i):

$$r^{(i)} \cdot D_U \lambda_i \neq 0. \tag{1.19}$$

In (1.19) and in the remainder of this section, $D_U f = \left(\frac{\partial f}{\partial U_1}, \frac{\partial f}{\partial U_2} \right)$ denotes the gradient of f viewed as a function of U and $\cdot$ denotes the Euclidean dot product. Thus, in component form, equation (1.19) reads $r_a^{(i)} \frac{\partial \lambda_i}{U_a} \neq 0$ (with summation over a but *not* over i).

We now show that under the assumption of plane symmetry and under the assumptions (1.15), when $\partial_t \Phi$ is sufficiently small, equation (1.14) is equivalent to a 2×2 genuinely nonlinear strictly hyperbolic system. In plane symmetry, equation (1.14) reduces to

$$-\partial_t^2 \Phi + c^2(\partial_t \Phi)\partial_x^2 \Phi = 0. \tag{1.20}$$

To derive the equivalent 2×2 genuinely nonlinear strictly hyperbolic system, we introduce the variables

$$U_1 := \partial_t \Phi, \qquad U_2 := \partial_x \Phi. \tag{1.21}$$

It is straightforward to see that for C^2 solutions, equation (1.20) is equivalent to equation (1.16a), where the matrix $M = M(U)$ has the components

$$\begin{pmatrix} M_1^1 & M_1^2 \\ M_2^1 & M_2^2 \end{pmatrix} = \begin{pmatrix} 0 & -c^2 \\ -1 & 0 \end{pmatrix}. \tag{1.22}$$

To verify the genuine nonlinearity and strict hyperbolicity conditions for the matrix (1.22) when $\partial_t \Phi$ is sufficiently small, we simply recall that $c = c(\partial_t \Phi) = c(U_1)$ and compute that

$$\lambda_1 = -c, \qquad \lambda_2 = c, \tag{1.23a}$$

$$r^{(1)} = \frac{1}{\sqrt{1+c^{-2}}}(1, c^{-1}), \qquad r^{(2)} = \frac{1}{\sqrt{1+c^{-2}}}(1, -c^{-1}), \tag{1.23b}$$

$$l_{(1)} = \frac{1}{\sqrt{1+c^{2}}}(1, c), \qquad l_{(2)} = \frac{1}{\sqrt{1+c^{2}}}(1, -c), \tag{1.23c}$$

$$r_a^{(1)} \frac{\partial \lambda_1}{U_a} = -\frac{1}{\sqrt{1+c^{-2}}} c', \qquad r_a^{(2)} \frac{\partial \lambda_2}{U_a} = \frac{1}{\sqrt{1+c^{-2}}} c'. \tag{1.23d}$$

As is well known, to analyze genuinely nonlinear strictly hyperbolic systems, it is convenient to use *Riemann invariants.* Given such a system, for $i = 1, 2$, we define a Riemann invariant $w_i = w_i(U_1, U_2)$ to be any function that is constant along the integral curves[9] of $r^{(i)}$. Equivalently, w_i verifies the following equation (with no summation over i):

$$r^{(i)} \cdot D_U w_i = 0. \tag{1.24}$$

Under the assumption that $D_U w_i \neq 0$ for $i = 1, 2$, w_1 and w_2 form a system of *state space coordinates* that can be used in place of U_1 and U_2. The main advantage is that relative to a coordinate system of Riemann invariants, the Cauchy problem (1.16a)-(1.16b) takes the following simple form:[10]

$$L_{(2)} w_1 = 0, \qquad L_{(1)} w_2 = 0, \tag{1.25a}$$

$$w_1|_{t=0} = \mathring{w}_1, \qquad w_2|_{t=0} = \mathring{w}_2, \tag{1.25b}$$

where the two *characteristic vectorfields* $L_{(i)}$ are defined as follows:

$$L_{(i)} := \partial_t + \lambda_i \partial_x. \tag{1.26}$$

We now verify the equivalence of the systems (1.16a) and (1.25a) when w_1 and w_2 form a coordinate system of Riemann invariants. To this end, we first claim that the genuine nonlinearity condition (1.19) is equivalent to

$$\frac{\partial \lambda_i}{\partial w_j} = \frac{\partial \lambda_i}{\partial U_a} \frac{\partial U_a}{\partial w_j} \neq 0, \qquad (i \neq j). \tag{1.27}$$

The equality in (1.27) follows from the chain rule. To verify the $\neq 0$ aspect of the claim, we note that (1.19) and (1.24) imply that $D_U \lambda_i$ and $D_U w_i$ are not parallel or anti-parallel. Moreover, by the chain rule, when $i \neq j$, we have $\frac{\partial w_i}{\partial U_a} \frac{\partial U_a}{\partial w_j} = 0$. Thus, when $i \neq j$, the one-form with components $\left(\frac{\partial U_1}{\partial w_j}, \frac{\partial U_2}{\partial w_j} \right)$ is perpendicular to $D_U w_i$ and hence, by the previous observation, not perpendicular to $D_U \lambda_i$. Clearly this is equivalent to the desired $\neq 0$ statement in (1.27).

[9] More precisely, w_i is constant along the integral curves of the dual of the covector $r^{(i)}$ with respect to the standard Euclidean metric on the state space $\mathbb{R}^2$.

[10] Throughout the monograph, if X is a vectorfield and f is a scalar-valued function, then $Xf := X^\alpha \partial_\alpha f$ denotes the derivative of f in the direction X. The X^α are the components of X relative to the spacetime coordinate partial derivative vectorfield frame $\{\partial_\alpha\}$. In particular, the first equation in (1.25a) is $L_{(2)}^0 \partial_0 w_1 + L_{(2)}^1 \partial_1 w_1 = \partial_t w_1 + \lambda_i \partial_x w_1 = 0$.

We now give the proof that the first equation in (1.25a) is a consequence of the system (1.16a). The proofs that the second equation in (1.25a) also follows from (1.16a) and that the reverse implication holds are similar, and we omit those details. To proceed, we first use the chain rule and equation (1.16a) to deduce that

$$\begin{aligned} \partial_t w_1 + \lambda_2 \partial_x w_1 &= \frac{\partial w_1}{\partial U_a} \partial_t U_a + \lambda_2 \frac{\partial w_1}{\partial U_a} \partial_x U_a \\ &= -\frac{\partial w_1}{\partial U_a} M_a^b \partial_x U_b + \lambda_2 \frac{\partial w_1}{\partial U_a} \partial_x U_a \\ &= \frac{\partial w_1}{\partial U_a} \left\{ -M_a^b + \lambda_2 \delta_a^b \right\} \partial_x U_b, \end{aligned} \tag{1.28}$$

where δ_a^b is the standard Kronecker delta. We now claim that $\frac{\partial w_1}{\partial U_a} \left\{ -M_a^b + \lambda_2 \delta_a^b \right\} = 0$, which implies the desired first equation in (1.25a). This claim follows from the orthogonality condition (1.18) and the definition (1.24) of a Riemann invariant, which together guarantee that $D_U w_1$ is proportional to $l_{(2)}$, and from equation (1.17b).

In the remainder of this section, we assume that the solution remains inside of a compact subset of the region of state space in which the system is strictly hyperbolic and genuinely nonlinear. Under this assumption, we will show that when the data $(\mathring{w}_1, \mathring{w}_2)$ (given in (1.25b)) are compactly supported and nontrivial, the solution to (1.25a) blows up in finite time. This shows in particular that the plane symmetric wave equation (1.20) exhibits finite-time blow-up. For definiteness, in view of the $\neq 0$ statement in (1.27), we assume that there is a constant $C > 0$ and a point x_0 such that

$$\frac{\partial \lambda_2}{\partial w_1} > C, \tag{1.29}$$

$$\partial_x \mathring{w}_1(x_0) < 0. \tag{1.30}$$

We will show that (1.29)-(1.30) imply finite-time blow-up for w_1. Moreover, by making straightforward modifications to our proof, we could deduce that related blow-up results hold for w_1 if the signs in (1.29)-(1.30) are altered in a compatible fashion, and that w_2 blows up if (1.29) is replaced with a signed condition on $\frac{\partial \lambda_1}{\partial w_2}$ and (1.30) is replaced with a compatible signed condition on $\partial_x \mathring{w}_2$. It is straightforward to check that for nontrivial compactly supported data, at least one of these conditions leading to blow-up must occur.

Our proof of the blow-up of w_1 is an adaption of our first (crude) proof of blow-up for Burgers' equation given in Sect. 1.1.2. In the remaining discussion in this section, we view the characteristic speeds λ_i as functions of the Riemann invariants (w_1, w_2) but we often suppress this functional dependence. Moreover, we view w_1 and w_2 as functions of the standard coordinates (t, x). As in our first proof of blow-up for Burgers' equation, the main idea is to exploit the fact that $\partial_x w_1$ verifies a Riccati-type equation. The new complication is that the equation is coupled to w_2. Specifically, we differentiate the first equation in (1.25a) with ∂_x and use the chain rule, definition (1.26), and the identity $\partial_x w_2 = \frac{1}{\lambda_2 - \lambda_1} L_{(2)} w_2$

(valid when w_2 verifies the second equation in (1.25a)) to deduce that

$$L_{(2)}\partial_x w_1 + \frac{\partial \lambda_2}{\partial w_1}(\partial_x w_1)^2 + \left\{ \frac{1}{\lambda_2 - \lambda_1}\frac{\partial \lambda_2}{\partial w_2} L_{(2)} w_2 \right\} \partial_x w_1 = 0. \tag{1.31}$$

To derive the desired blow-up result for $\partial_x w_1$, we control the term in braces on the left-hand side of (1.31) and show that it remains bounded. To this end, we define, much as in (1.8a)-(1.8b), the family of characteristic curves $\gamma_{(2)}(t;x) = (\gamma_{(2)}^0(t;x), \gamma_{(2)}^1(t;x))$ corresponding to the vectorfield $L_{(2)}$ to be the solutions to the following ODE initial value problems:

$$\frac{d}{dt}\gamma_{(2)}^0(t;x) = 1, \qquad \frac{d}{dt}\gamma_{(2)}^1(t;x) = \lambda_2 \circ \gamma_{(2)}(t;x), \tag{1.32a}$$

$$\gamma_{(2)}^0(0;x) = 0, \qquad \gamma_{(2)}^1(0;x) = x. \tag{1.32b}$$

As a first quantitative step in our proof of blow-up, we now derive simple uniform bounds from above and below for w_1 and w_2. To this end, we note that by the chain rule and (1.32a), the first equation in (1.25a) can be expressed as

$$\frac{d}{dt}\left\{ w_1 \circ \gamma_{(2)}^0(t;x) \right\} = 0. \tag{1.33}$$

Integrating (1.33) starting from $t = 0$, we find that at any time t of classical existence, we have

$$\min_{x\in\mathbb{R}} w_1 \circ \gamma_{(2)}^0(t;x) = \min_{x\in\mathbb{R}} \mathring{w}_1(x), \qquad \max_{x\in\mathbb{R}} w_1 \circ \gamma_{(2)}^0(t;x) = \max_{x\in\mathbb{R}} \mathring{w}_1. \tag{1.34}$$

Similarly, by integrating along characteristic curves corresponding to the vectorfield $L_{(1)}$, we find that (1.34) also holds with w_1 everywhere replaced with w_2. Thus, we deduce that there exists a constant $C > 0$ depending on the data such that in any region of classical existence, we have

$$-C \leq w_i \leq C. \tag{1.35}$$

Our proof of blow-up relies on using an integrating factor $\iota(t;x)$ to simplify the structure of equation (1.31). Specifically, we define

$$\iota(t;x) := \exp\left(\int_{w=\mathring{w}_2(x)}^{w=w_2\circ\gamma_{(2)}(t;x)} \left\{ \frac{1}{\lambda_2 - \lambda_1}\frac{\partial \lambda_2}{\partial w_2} \right\} (\mathring{w}_1(x), w)\, dw \right). \tag{1.36}$$

Note that by (1.32b), we have $\iota(0;x) = 1$. Note also that (1.35) and the strict hyperbolicity assumption $\lambda_1 < \lambda_2$ imply that there exists a constant $c > 1$ (depending on the data) such that in any region of classical existence, we have

$$c^{-1} < \iota(t;x) < c. \tag{1.37}$$

The main step in the proof of blow-up for solutions $\partial_x w_1$ to equation (1.31) is to derive the following equation:

(1.38)

$$\begin{aligned}(\partial_x w_1) \circ \gamma_{(2)}(t;x) &= (\partial_x \mathring{w}_1)(x)\iota^{-1}(t;x) \\ &\quad \times \left\{ 1 + (\partial_x \mathring{w}_1)(x) \int_{s=0}^{t} \iota^{-1}(s;x) \frac{\partial \lambda_2}{\partial w_1} \circ \gamma_{(2)}(s;x)\, ds \right\}^{-1}.\end{aligned}$$

Once we have obtained (1.38), the desired finite-time blow-up of $\partial_x w_1$ easily follows from (1.29)-(1.30) and (1.37).

It remains for us to derive equation (1.38). To this end, we use equations (1.33) and (1.32b) to deduce that

$$w_1 \circ \gamma_{(2)}(t,x) = \mathring{w}_1(x). \tag{1.39}$$

Hence, from definition (1.26), equations (1.32a) and (1.39), and the chain rule, we deduce the following identity involving the term in braces on the left-hand side of (1.31):

$$\begin{aligned} \frac{d}{dt} \int_{w=\mathring{w}_2(x)}^{w=w_2\circ\gamma_{(2)}(t;x)} & \left\{ \frac{1}{\lambda_2-\lambda_1} \frac{\partial \lambda_2}{\partial w_2} \right\} (\mathring{w}_1(x), w)\, dw \\ &= \left\{ \frac{1}{\lambda_2-\lambda_1} \frac{\partial \lambda_2}{\partial w_2} L_{(2)} w_2 \right\} \circ \gamma_{(2)}(t;x). \end{aligned} \tag{1.40}$$

Using (1.32a), (1.36), (1.40), and the chain rule, we can rewrite the Riccati-type equation (1.31) as

$$\frac{d}{dt}\left(\frac{1}{\iota(t;x)(\partial_x w_1)\circ\gamma_{(2)}(t;x)} \right) = \frac{1}{\iota(t;x)} \frac{\partial \lambda_2}{\partial w_1} \circ \gamma_{(2)}(t;x). \tag{1.41}$$

Finally, integrating equation (1.41) with respect to time from 0 to t, using the initial conditions (1.32b), and carrying out straightforward computations, we arrive at the desired equation (1.38). This concludes our proof of blow-up of solutions to (1.25a)-(1.25b).

1.2. New aspects in more than one spatial dimension

In Sect. 1.2, we discuss some fundamental issues that distinguish the study of quasilinear hyperbolic PDEs in two or more spatial dimensions from the case of one spatial dimension. In particular, we discuss the need for energy estimates and the possible presence of dispersion. Both of these features play a key role in our derivation of the main shock formation results of this monograph. Next, for nonlinear wave equations, we discuss the role that dispersion plays in modern proofs of small-data global existence. Moreover, in the case of three spatial dimensions, we discuss Klainerman's (classic) null condition and outline why nonlinearities verifying it allow for small-data global existence. The wave equations that we study in Chapters 2-23 have nonlinearities that fail the null condition, which ultimately leads to small-data shock formation.

1.2.1. The energy method and local well-posedness. For quasilinear hyperbolic equations, a fundamental difference between one and more than one spatial dimension is that in the latter case, one cannot rely exclusively on the method of characteristics. Even basic local well-posedness[11] results are based on the availability of a priori L^2-based energy estimates; there are no known approaches that avoid them. The use of the energy method to prove local well-posedness has a long history with a huge number of contributors. In Sect. 1.2.1, we review basic local well-posedness theory for the well-known class of symmetric hyperbolic systems, with an emphasis on the role of the energy method. As we describe in Sect. 1.2.2, wave equations can be viewed as a special case of symmetric hyperbolic systems for which a *much larger* class of energy estimates is available. Although in the present section we discuss only symmetric hyperbolic systems in detail, we briefly mention

[11] By "well-posedness", we mean existence, uniqueness, and continuous dependence on the initial conditions.

several notable classes of equations (with some overlap) to which the energy method has been applied.

- In the case of quasilinear wave equations, relevant for the present monograph, Schauder [**73**] applied energy methods to linearized versions of the equations and employed a fixed point argument to prove local well-posedness. This is the earliest known example of this kind of argument, which is now standard.
- For *strictly hyperbolic* first-order systems, which by definition have characteristic speeds that are strictly separated, Petrovskii [**68**] discovered a coercive energy identity based on the Fourier transform and used it to prove local well-posedness. Gårding [**24**, **25**], using ideas from Leray's work [**56**] (see the next item), extended the well-posedness result to higher-order strictly hyperbolic scalar equations.
- Leray developed the theory of *Leray hyperbolicity* [**56**], which allows for operators having principal parts of different orders within the same system.
- As we have mentioned, there is a well-posedness theory for the well-known *symmetric hyperbolic systems*, which are not generally strictly hyperbolic. The general theory for such systems was developed by Friedrichs [**23**]. We review this theory later in this section.
- Christodoulou introduced [**16**] *regularly hyperbolic* PDEs, which are a class of well-posed Euler-Lagrange equations for maps between a domain manifold $\mathcal{M}$ and a target manifold $\mathcal{N}$. He developed a geometric energy method framework for such PDEs that extends the multiplier method described in Sect. 1.2.2 and leads to a large family of energy estimates. The full theory takes into account the structure of both $\mathcal{M}$ and $\mathcal{N}$.

We now review the basic theory of local well-posedness for symmetric hyperbolic systems. We start by providing their definition.

DEFINITION 1.2 (**Symmetric hyperbolic systems**). A PDE system[12]

$$A_I^{J\alpha}(U)\partial_\alpha U_J = F_I(U), \qquad (I = 1, 2, \cdots, m) \tag{1.42}$$

on the domain[13] $\mathbb{R}^{1+n}$ in the unknowns $U = (U_1, \cdots, U_m)$ is said to be *symmetric hyperbolic* in an open subset $\mathcal{U} \subset \mathbb{R}^m$ if for each fixed $U \in \mathcal{U}$, we have the symmetry property

$$A_I^{J\alpha}(U) = A_J^{I\alpha}(U), \qquad (I, J = 1, 2, \cdots, m), \ (\alpha = 0, 1, \cdots, n), \tag{1.43}$$

and furthermore, there exists a one-form $\xi = (\xi_0, \xi_1, \cdots, \xi_n)$ on $\mathbb{R}^{1+n}$ (generally varying from point to point) such that

$$\text{the } m \times m \text{ matrices } \xi_\alpha A_I^{J\alpha}(U), (I, J = 1, \cdots, m), \text{ are positive definite.} \tag{1.44}$$

REMARK 1.3 (**Symmetrizable systems**). The theory of symmetric hyperbolic systems can also be extended to apply to symmetrizable systems, which are

[12]Throughout the remainder of the monograph, we use Einstein's summation convention. Lowercase Greek "spacetime" indices vary over $0, 1, \cdots, n$ and lowercase Latin "spatial" indices vary over $1, 2, \cdots, n$. Moreover, starting in Chapter 2, we have $n = 3$. In Sect. 1.2.1, capital Latin indices correspond to the target and vary over $1, 2, \cdots, m$.

[13]Throughout Sect. 1.2.1, $(x^0 = t, x^1, \cdots, x^n)$ are standard rectangular coordinates on $\mathbb{R}^{1+n}$ and $\partial_\alpha = \dfrac{\partial}{\partial x^\alpha}$ are the corresponding coordinate partial derivatives.

systems that can be transformed into symmetric hyperbolic form; see, for example, [**22**]. In many cases, the transformation procedure involves a nonlinear state-space change of variables of the form $U \to f(U)$, where $f : \mathbb{R}^m \to \mathbb{R}^m$.

The main step in proving well-posedness for the system (1.42) is to obtain Sobolev estimates for solutions to the following linearized version of (1.42):

$$A_I^{J\alpha}(\widetilde{U})\partial_\alpha U_J = F_I(\widetilde{U}), \qquad (I = 1, 2, \cdots, m). \tag{1.45}$$

In (1.45), $\widetilde{U} : \mathbb{R}^{1+n} \to \mathcal{U} \subset \mathbb{R}^m$ is a background function[14] and U is the unknown. The Sobolev estimates follow from an energy identity for solutions to (1.45), which we now explain. In the ensuing discussion, we abbreviate

$$|U|_E := \sqrt{(E^{-1})^{IK} U_I U_K},$$

$$\langle A^\alpha(\widetilde{U})U, U\rangle_E := (E^{-1})^{IK} A_I^{J\alpha}(\widetilde{U}) U_J U_K,$$

and

$$\left\langle \left\{ \partial_\alpha \left(A^\alpha(\widetilde{U}) \right) \right\} U, U \right\rangle_E := (E^{-1})^{IK} \left\{ \partial_\alpha \left(A_I^{J\alpha}(\widetilde{U}) \right) \right\} U_J U_K,$$

where $E_{IJ} = \text{diag}(1, 1, \cdots, 1)$ is the standard Euclidean metric on $\mathbb{R}^m$. To derive the energy identities, we use a framework that is related to a robust, more powerful geometric framework for deriving generalized energy estimates for wave equations. The geometric framework plays a critical role in this monograph, and we provide an overview of it in Sect. 1.2.2. The main idea behind deriving the energy identities is to apply the divergence theorem in a suitable spacetime region with the help of the *compatible energy current* vectorfield $J^\alpha_{(Energy)}[U]$ on $\mathbb{R}^{1+n}$ defined by

$$J^\alpha_{(Energy)}[U] := \langle A^\alpha(\widetilde{U})U, U\rangle_E, \qquad (\alpha = 0, 1, \cdots, n). \tag{1.46}$$

To proceed, we use (1.43) and (1.45) to compute that

$$\partial_\alpha J^\alpha_{(Energy)}[U] = \left\langle \left\{ \partial_\alpha A^\alpha(\widetilde{U}) \right\} U, U \right\rangle_E + 2\langle U, F(\widetilde{U})\rangle_E. \tag{1.47}$$

We can obtain coercive energy identities in spacetime regions whose boundary is the union of "spacelike" hypersurfaces, which are hypersurfaces with co-normals ξ that verify the analog of (1.44) for the linear system (1.45). We now explain a particular case, sufficient for deducing local well-posedness on a small spacetime patch. We consider spacetime subsets foliated by the level sets of a scalar-valued *time function* $\uptau$. That is, we assume that there is a C^1 scalar-valued function $\uptau$ defined on a subset of $\mathbb{R}^{1+n}$ such that the level sets

$$H_\uplambda := \{p \in \mathbb{R}^{1+n} \mid \uptau(p) = \uplambda\} \tag{1.48}$$

have the property that $\xi_\alpha A_I^{J\alpha}(\widetilde{U})$ is positive definite at every point in $H_\uplambda$, where ξ is the one-form that is (Euclidean) metric-dual[15] to the future-directed[16] Euclidean-unit normal vectorfield along $H_\uplambda$. Here, by Euclidean metric, we mean the standard one on $\mathbb{R}^{1+n}$ defined by $e_{\alpha\beta} = \text{diag}(1, 1, \cdots, 1)$. For convenience, we assume here

[14] In a proof of well-posedness via an iteration scheme, the role of the background function is played by a known iterate, and the role of U is played by the next iterate, which is to be solved for; see just below Prop. 1.5.

[15] That is, $\xi_\alpha = e_{\alpha\beta} N^\beta$, where N is the future-directed Euclidean unit normal vectorfield to $H_\uplambda$.

[16] Throughout the monograph, future-directed vectors V are such that $V^0 > 0$, where, relative to the rectangular coordinates, $V = V^\alpha \partial_\alpha$.

that the H_λ are compact and have a common C^1 boundary, denoted by ∂H (see Figure 1). We now derive energy estimates on regions $\mathcal{M}_\lambda \subset \mathbb{R}^{1+n}$ defined by (see Figure 1)

$$\mathcal{M}_\lambda := \cup_{\lambda' \in [0,\lambda]} H_{\lambda'}. \tag{1.49}$$

The most geometric way of formulating the divergence theorem in the present context involves the well-known *co-area formula* (see, for example, [**53**]), which states that for functions f, we have the following integral identity on the region (1.49):

$$\int_{\mathcal{M}_\lambda} f \, d\varpi_e = \int_{\lambda'=0}^{\lambda} \int_{H_{\lambda'}} f |d\tau|_e^{-1} \, d\sigma_{\lambda'} \, d\lambda'. \tag{1.50}$$

In (1.50), $d\varpi_e$ is the spacetime volume form[17] corresponding to the Euclidean metric $e_{\alpha\beta}$ and $d\sigma_\lambda$ is the volume form corresponding to the Riemannian metric ${}^{(\lambda)}h_{\alpha\beta}$ induced on[18] H_λ by $e_{\alpha\beta}$. Moreover, $d\tau = (\partial_t \tau, \partial_1 \tau, \cdots, \partial_n \tau)$ is the spacetime gradient one-form of τ and $|d\tau|_e = \sqrt{(e^{-1})^{\alpha\beta} \partial_\alpha \tau \partial_\beta \tau}$ is its Euclidean norm.

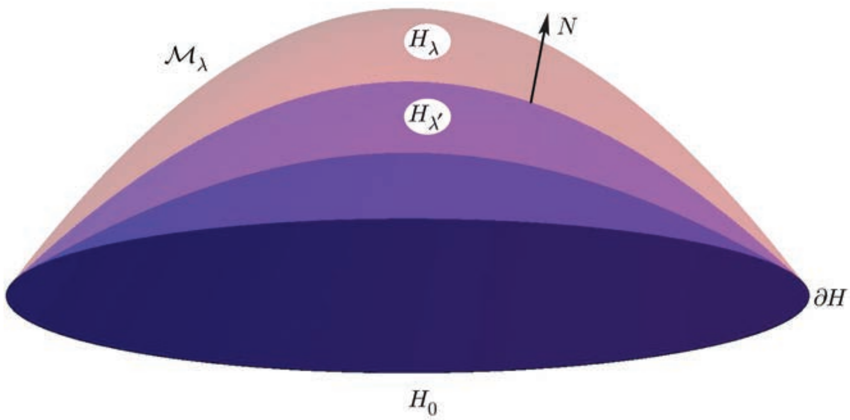

FIGURE 1. Foliation of the spacetime region $\mathcal{M}_\lambda$ by $H_{\lambda'}$, $\lambda' \in [0, \lambda]$. N denotes the Euclidean-unit normal to $H_{\lambda'}$, which is dual to ξ via the Euclidean metric.

The fundamental energy corresponding to the linear system (1.45) and the hypersurface H_λ is

$$\mathbb{E}(\lambda) := \int_{H_\lambda} \langle \xi_\alpha A^\alpha(\widetilde{U}) U, U \rangle_E \, d\sigma_\lambda = \int_{H_\lambda} \xi_\alpha J^\alpha_{(Energy)}[U] \, d\sigma_\lambda, \tag{1.51}$$

where $J^\alpha_{(Energy)}[U]$ is defined in (1.46). In view of our above assumptions on H_λ, we see that there exists a constant $C > 1$ depending on $\widetilde{U}$ such that

$$C^{-1} \int_{H_\lambda} |U|_E^2 |d\tau|_e^{-1} \, d\sigma_\lambda \leq \mathbb{E}(\lambda) \leq C \int_{H_\lambda} |U|_E^2 |d\tau|_e^{-1} \, d\sigma_\lambda. \tag{1.52}$$

[17]Relative to the standard rectangular coordinates on $\mathbb{R}^{1+n}$, we have $d\varpi_e = dx^0 \cdots dx^n$.

[18]We have ${}^{(\lambda)}h_{\alpha\beta} = e_{\alpha\beta} - \xi_\alpha \xi_\beta$. Furthermore, relative to rectangular coordinates on $\mathbb{R}^{1+n}$, we have $\xi_\alpha = \dfrac{\partial_\alpha \tau}{|d\tau|_e}$.

Applying the divergence theorem with the current (1.46) over the region $\mathcal{M}_\lambda$ and using the identity (1.47), we obtain the following *energy identity*:

(1.53)

$$\mathbb{E}(\lambda) = \mathbb{E}(0) + \int_{\lambda'=0}^{\lambda} \int_{H_{\lambda'}} \left\{ \left\langle \left(\partial_\alpha A^\alpha(\widetilde{U}) \right) U, U \right\rangle_E + 2\langle U, F(\widetilde{U}) \rangle_E \right\} |d\tau|_e^{-1} \, d\sigma_{\lambda'} \, d\lambda'.$$

We remark that in our proof of shock formation for solutions to wave equations, we rely on intricate geometric energy identities in the spirit of (1.53); see Prop. 10.13.

Using (1.52), (1.53), and the elementary inequality $ab \leq \frac{1}{2}(a^2 + b^2)$, we see that there exists a constant $C > 1$ depending on $\widetilde{U}$ and its C^1 norm such that

$$\mathbb{E}(\lambda) \leq \mathbb{E}(0) + C \int_{\lambda'=0}^{\lambda} \mathbb{E}(\lambda') \, d\lambda' + C \int_{\lambda'=0}^{\lambda} \int_{H_{\lambda'}} F^2(\widetilde{U}) |d\tau|_e^{-1} \, d\sigma_{\lambda'} \, d\lambda'. \tag{1.54}$$

From (1.54) and Gronwall's inequality, we deduce the fundamentally important a priori estimate

$$\mathbb{E}(\lambda) \leq \left\{ \mathbb{E}(0) + C \int_{\lambda'=0}^{\lambda} \int_{H_{\lambda'}} F^2(\widetilde{U}) |d\tau|_e^{-1} \, d\sigma_{\lambda'} \, d\lambda' \right\} \exp(C\lambda). \tag{1.55}$$

As we describe below, the estimate (1.55) is the main ingredient in the proof of local well-posedness for symmetric hyperbolic systems. In the next proposition, we provide a basic local well-posedness result. For convenience, we restrict our attention to the case in which the one-form ξ from Def. 1.2 has the simple form $\xi = (1, \mathbf{0}_{1\times n})$ relative to rectangular coordinates and hence $H_\lambda = \Sigma_\lambda$, where in the remaining discussion,

$$\Sigma_t := \left\{ (s, x^1, x^2, \cdots, x^n) \in \mathbb{R}^{1+n} \mid s = t \right\} \simeq \mathbb{R}^n \tag{1.56}$$

denotes the flat hypersurface in $\mathbb{R}^{1+n}$ of constant time t.

Remark 1.4 (**The spaces $H_e^s(\Sigma_t)$ and $L_e^2(\Sigma_t)$**). Throughout the monograph, we use the notation $H_e^s(\Sigma_t)$ to denote the standard Sobolev space of order s corresponding to rectangular spatial coordinate partial derivatives along Σ_t, where the volume form inherent in the corresponding norm $\| \cdot \|_{H_e^s(\Sigma_t)}$ is the one induced by the standard Euclidean metric e on Σ_t; see Sect. B.17. A similar remark applies to the Lebesgue space $L_e^2(\Sigma_t)$ and the corresponding norm $\| \cdot \|_{L_e^2(\Sigma_t)}$. This distinction is important because later in the monograph (see Sect. 3.19), we define another norm, denoted by $\| \cdot \|_{L^2(\Sigma_t)}$, where the volume form that we use in defining it is, near a shock singularity, drastically different than the one induced by e.

Proposition 1.5 (**Local well-posedness and continuation criteria for symmetric hyperbolic systems**). *Assume that the system (1.42) is symmetric hyperbolic on the open set $\mathcal{U} \subset \mathbb{R}^m$ and has vanishing inhomogeneous term*[19] *$F_I(U) \equiv 0$. Let $s > n/2+1$ and let $\mathring{U} \in H_e^s(\Sigma_0)$ be initial data. Assume that $\mathring{U}(\Sigma_0)$ is contained in a compact subset $\mathfrak{K}$ of interior($\mathcal{U}$) such that for $U \in \mathfrak{K}$, the $m \times m$ matrix $A_J^{I0}(U)$, $(I, J = 1, \cdots, m)$, is positive definite. Then there exists a $T > 0$,*

[19]The proposition can easily be adapted so as to apply when $F_I(U) \not\equiv 0$, under suitable assumptions on $F_I(U)$.

depending only on $\|\mathring{U}\|_{H^s_e}(\Sigma_0)$ *and* $\mathfrak{K}$*, such that these data launch a unique classical solution* U *existing on the slab* $[0,T)\times\mathbb{R}^n$ *which has the regularity property*

$$U \in C([0,T), H^s_e(\Sigma_t)). \tag{1.57}$$

The solution depends continuously on the data. Moreover, there is a compact set $\mathfrak{K}' \subset$ *interior*$(\mathcal{U})$ *with* $\mathfrak{K} \subset$ *interior*$(\mathfrak{K}')$ *such that* $U([0,T)\times\mathbb{R}^n) \subset \mathfrak{K}'$*. Furthermore,* U *verifies the following estimate for* $t \in [0,T)$*:*

$$\begin{aligned} \|U\|_{H^s_e(\Sigma_t)} \leq f\left(\left\|\mathring{U}\right\|_{H^s_e(\Sigma_0)}; \mathfrak{K}'; s\right) \\ \times \exp\left(C_{\mathfrak{K}';s}\int_{t'=0}^{t}\sum_{\alpha=0}^{n}\|\partial_\alpha U\|_{L^\infty(\Sigma_{t'})}\,dt'\right), \end{aligned} \tag{1.58}$$

where f *is a function that is continuous and increasing in its first argument and that vanishes when* $\mathring{U} = 0$*. Moreover, there exists a maximal slab* $[0, T_{(Lifespan)})\times\mathbb{R}^n$ *on which the above properties, including the regularity property* (1.57)*, hold. Finally, if* $T_{(Lifespan)} < \infty$*, then either* $\sum_{\alpha=0}^{n}\int_{t'=0}^{T_{(Lifespan)}}\|\partial_\alpha U\|_{L^\infty(\Sigma_{t'})}\,dt' = \infty$ *or* $U([0,T)\times\mathbb{R}^n)$ *escapes every compact subset of interior*$(\mathcal{U})$ *as* $T \uparrow T_{(Lifespan)}$*.*

The proof of the proposition is standard and most of the details can be found, for example, in [**30**, Chapter VI]. The main idea is to define iterates[20] $\{U^{(k)}\}_{k=0}^\infty$ where $U^{(0)} := \mathring{U}$ and $U^{(k+1)}$ is determined in terms of $U^{(k)}$ by solving the linearized system $A_I^{J\alpha}(U^{(k)})\partial_\alpha(U_J^{(k+1)} - \mathring{U}) = -A_I^{Ja}(U_J^{(k)})\partial_a\mathring{U}$ with initial data $U_J^{(k+1)}|_{\Sigma_0} - \mathring{U} = 0$. One then derives energy estimates along the lines of (1.55) to deduce, with the help of the standard Sobolev calculus, uniform estimates for[21] $\{\sup_{t\in[0,T)}\{\|U^{(k)} - \mathring{U}\|_{H^s_e(\Sigma_t)}\}_{k=0}^\infty$ whenever T is sufficiently small. In particular, standard Sobolev embedding implies that $\sum_{\alpha=0}^n\|\partial_\alpha U\|_{L^\infty(\Sigma_{t'})} \leq C\|U\|_{H^s_e(\Sigma_{t'})}$, which provides control over the argument of exp on the right-hand side of (1.58). Next, one can derive similar uniform estimates based on the previous estimates and use them to show that the sum $\sum_{k=0}^\infty\sup_{t\in[0,T)}\left\|U^{(k+1)} - U^{(k)}\right\|_{L^2_e(\Sigma_t)}$ is finite, which implies the convergence of the iterates in L^2. One can then show that the limit U is a classical solution and obtain additional information, including estimates on its derivatives up to order s. The proofs are based on similar arguments and standard results from functional analysis.

1.2.2. Square integral-type dispersion for wave equations based on generalized energy estimates. A major aspect permeating the study of quasilinear hyperbolic PDEs in higher dimensions is that dispersion is sometimes present, and it can cause the solution to decay (at least initially) and thus delay or prevent the formation of singularities. In the opposite direction, depending on the structure of the equations, it is possible that other kinds of singularities besides shocks might develop. For example, a major open problem is whether or not vorticity blow-up occurs in the compressible Euler equations. Although Sideris gave, in the case of three spatial dimensions, a proof by contradiction that an open set of initial data generate solutions that blow-up [**74**], we have reason to suspect that his result is

[20]Alternatively, one could formulate a similar proof based on the contraction mapping principle.

[21]In a rigorous proof, one has to smooth the data during this step to compensate for the loss of one derivative of the data in the term on the right-hand side of the linearized system.

detecting shocks rather than vorticity blow-up; below we elaborate on this suspicion. Specifically, Sideris showed that under suitable convexity assumptions on the fluid equation of state, small-data blow-up occurs; see also [**26**] for related large-data blow-up results for the relativistic Euler equations. Sideris' proof was based on using virial identity arguments to indirectly rule out the possibility of global smooth solutions. That is, he constructed spatially averaged quantities that, on the one hand are smooth when the solution is smooth, but on the other hand must blow up in finite time as a consequence of some differential inequalities verified by solutions. In particular, Sideris' proof did not reveal the blow-up mechanism. It is important to note that Sideris' condition for blow-up involves an integral condition on the data that resembles the one given by Christodoulou in [**17**] for the full vorticity-containing relativistic Euler equations in the small-data regime. The key point is that Christodoulou showed [**17**, Chapter 14] that his condition leads to the development of a large irrotational region in which a shock singularity forms; see also [**21**] for a similar result in the case of the irrotational nonrelativistic compressible Euler equations. For these reasons, we speculate that Siderian blow-up is caused by the same mechanism.

Dispersive estimates play an especially important role in the study of solutions to nonlinear wave equations in[22] two or more spatial dimensions. In particular, the issue of global existence versus finite-time blow-up for solutions is often decided by a competition between the dispersive nature of the corresponding linear problem and the strength of the nonlinearities.[23] In the shock forming wave equation solutions (in three spatial dimensions) that we study later in this monograph, the dispersion just barely fails to suppress the singularity formation. However, an important and perhaps surprising fact is that dispersion, both L^2-type and pointwise, plays a key role in our proof of the shock formation. In fact, relative to a dynamic system of coordinates related to the eikonal function u (see (1.8a)) used in our study of Burgers' equation, the shock forming solutions exhibit dispersive behavior at the lower derivative levels, all the way up to and including the shock; see Sect. 2.6 for an overview. Because the dispersive behavior is so important, we dedicate Sects. 1.2.2 and 1.2.3 to providing some general background information on wave dispersion. In this section, we focus on L^2-type dispersion.

We begin by motivating some of the quasilinear wave equations that we study in detail starting in Chapter 2. To avoid the difficult problem, mentioned above, of analyzing vorticity in solutions to the compressible Euler equations, one can study irrotational solutions. As we mentioned in Footnote 4 on pg. 2, for such solutions, the compressible Euler equations reduce to single quasilinear wave equation of Euler-Lagrange type for the scalar-valued fluid potential. In Sect. 2.11.2, we provide additional details on the structure of the wave equation in the case of the relativistic Euler equations. In the present section, rather than studying the equations of irrotational fluid mechanics, we instead study a closely related (as described in Appendix A) but seemingly more geometric family of covariant quasilinear wave equations of the form

$$\square_{g(\Psi)}\Psi = \mathcal{N}(\Psi, \partial\Psi) \tag{1.59}$$

[22] In one spatial dimension, the linear wave equation $-\partial_t^2\Phi + \partial_x^2\Phi = 0$ is effectively a transport equation whose solutions do not disperse.

[23] That is, often the nonlinear terms do not enhance the decay rates corresponding to the linear problem.

on the domain $\mathbb{R}^{1+n}$. In equation (1.59), $\square_g := (g^{-1})^{\alpha\beta}\mathscr{D}^2_{\alpha\beta}$ is the covariant wave operator[24] corresponding to g, $\mathscr{D}$ is the Levi-Civita connection corresponding to g, and $\mathcal{N}(\Psi, \partial\Psi)$ is a nonlinearity consisting of quadratic and higher-order terms. The notation $g = g(\Psi)$ means that relative to standard rectangular coordinates $(x^0 = t, x^1, x^2, \cdots, x^n)$ on $\mathbb{R}^{1+n}$, the components $g_{\mu\nu}(\Psi)$ are given smooth functions of Ψ. We assume that there is an open set $\mathcal{H}$ (of "hyperbolicity") such that for $\Psi \in \mathcal{H}$, the metric $g(\Psi)$ is Lorentzian[25] and, for convenience, that relative to the rectangular coordinates, that[26] $g_{00}(\Psi) < 0$.

Note that $\square_{g(\Psi)}\Psi$, when expanded relative to the rectangular coordinates, involves both quasilinear and semilinear terms. We remark that under certain assumptions that we describe starting in Chapter 2, when $n = 3$, the main shock formation results of this monograph apply to solutions of (1.59).

Before discussing L^2-type dispersion, we first discuss local well-posedness for equation (1.59). One can prove a basic well-posedness result for equation (1.59) by introducing the "new" independent variables $\partial_\nu\Psi := U_\nu$, $(\nu = 0, 1, \cdots, n)$, and then reformulating the equation as a (first-order) symmetric hyperbolic system in Ψ and U; one can then apply Prop. 1.5, which leads to the following proposition; see also Prop. 21.1, which provides a detailed version of local well-posedness that is relevant for the main results of the monograph.

PROPOSITION 1.6 (**Local well-posedness and continuation criteria for quasilinear wave equations of type** (1.59)). *Let $s > n/2 + 1$ and let $\mathcal{H} \subset \mathbb{R}$ be the set mentioned just below equation* (1.59). *Let $(\mathring{\Psi}, \mathring{\Psi}_0) = (\Psi|_{\Sigma_0}, \partial_t\Psi|_{\Sigma_0}) \in H_e^{s+1}(\Sigma_0) \times H_e^s(\Sigma_0)$ be initial data for the wave equation* (1.59) *and assume that there is a compact subset $\mathfrak{K} \subset$ interior$(\mathcal{H})$ such that $\Psi(\Sigma_0) \subset \mathfrak{K}$. Then there exists a $T > 0$, depending only on $\sum_{a=1}^n \|\partial_a\mathring{\Psi}_0\|_{H_e^s(\Sigma_0)} + \|\mathring{\Psi}_0\|_{H_e^s(\Sigma_0)}$ and $\mathfrak{K}$, such that these data launch a unique classical solution existing on the slab $[0,T) \times \mathbb{R}^n$ which has the regularity property*

$$\Psi \in C([0,T), H_e^{s+1}(\Sigma_t)), \qquad \partial_t\Psi \in C([0,T), H_e^s(\Sigma_t)). \tag{1.60}$$

The solution depends continuously on the data. Moreover, there is a compact set $\mathfrak{K}' \subset$ interior$(\mathcal{H})$ with $\mathfrak{K} \subset$ interior$(\mathfrak{K}')$ such that $\Psi([0,T)\times\mathbb{R}^n) \subset \mathfrak{K}'$. Furthermore, Ψ verifies the following estimate for[27] *$t \in [0,T)$:*

$$\begin{aligned}\sum_{\alpha=0}^n \|\partial_\alpha\Psi\|_{H_e^s(\Sigma_t)} \leq f\left(\sum_{a=1}^n \left\|\partial_a\mathring{\Psi}\right\|_{H_e^s(\Sigma_0)} + \left\|\mathring{\Psi}_0\right\|_{H_e^s(\Sigma_0)}; \mathfrak{K}'; s\right) \\ \times \exp\left(C_{\mathfrak{K}';s}\int_{t'=0}^t \sum_{\alpha=0}^n \|\partial_\alpha\Psi\|_{L^\infty(\Sigma_{t'})}\, dt'\right),\end{aligned} \tag{1.61}$$

where f is a function that is continuous and increasing in its first argument and that vanishes when the data are trivial. Moreover, there exists a maximal slab

[24] Relative to an arbitrary coordinate system, $\square_g\Psi = \frac{1}{\sqrt{|\det g|}}\partial_\alpha(\sqrt{|\det g|}(g^{-1})^{\alpha\beta}\partial_\beta\Psi)$.

[25] By Lorentzian, we mean that relative to arbitrary coordinates, $g_{\alpha\beta}$ is an $(1+n)\times(1+n)$ matrix of signature $(-,+,+,\cdots,+)$.

[26] The inequality $g_{00}(\Psi) < 0$ is equivalent to the timelike character of the vectorfield $\partial_t = \frac{\partial}{\partial x^0}$.

[27] An estimate for $\|\Psi\|_{L_e^2(\Sigma_t)}$ can be obtained by integrating the estimate (1.61) for $\|\partial_t\Psi\|_{L_e^2(\Sigma_t)}$ with respect to time.

$[0, T_{(Lifespan)}) \times \mathbb{R}^n$ *on which the above properties, including the regularity property* (1.60)*, hold. Finally, if* $T_{(Lifespan)} < \infty$*, then either*

$$\sum_{\alpha=0}^{n} \int_{t'=0}^{T_{(Lifespan)}} \|\partial_\alpha \Psi\|_{L^\infty(\Sigma_{t'})} \, dt' = \infty$$

or $\Psi([0,T) \times \mathbb{R}^n)$ *escapes every compact subset of interior*$(\mathcal{H})$ *as* $T \uparrow T_{(Lifespan)}$.

As we described below Prop. 1.5, the standard proof of Prop. 1.6 relies on using Sobolev embedding to estimate the norm in the argument of exp in equation (1.61) as follows:

$$\sum_{\alpha=0}^{n} \|\partial_\alpha \Psi\|_{L^\infty(\Sigma_{t'})} \leq C \sum_{\alpha=0}^{n} \|\partial_\alpha \Psi\|_{H_e^s(\Sigma_{t'})}.$$

It is important to note that in some cases, the Sobolev exponent s from Prop. 1.6 has been markedly improved by using refined tools such as Strichartz and bilinear estimates to replace the above crude Sobolev estimate. The most advanced result along these lines is the recent proof of the bounded L^2 curvature conjecture [**52**], which for the Einstein-vacuum equations of general relativity with $n = 3$ essentially leads to local well-posedness for $(\mathring{\Psi}, \mathring{\Psi}_0) \in H_e^2(\Sigma_0) \times H_e^1(\Sigma_0)$. This remarkable work is a major extension of earlier work [**51**] of Klainerman-Rodnianski, which proved local well-posedness for the Einstein-vacuum equations with $n = 3$ relative to wave coordinates[28] for data verifying, for any $\epsilon > 0$, $(\mathring{\Psi}, \mathring{\Psi}_0) \in H_e^{2+\epsilon}(\Sigma_0) \times H_e^{1+\epsilon}(\Sigma_0)$. Their work was extended by Smith-Tataru [**75**], who showed that when $n \in \{3, 4, 5\}$, quasilinear wave equations of the form

$$(g^{-1})^{\alpha\beta}(\Psi)\partial_\alpha\partial_\beta\Psi = \mathcal{N}^{\alpha\beta}(\Psi)\partial_\alpha\Psi\partial_\beta\Psi$$

are locally well-posed for data verifying, for any $\epsilon > 0$, $(\mathring{\Psi}, \mathring{\Psi}_0) \in H_e^{(n+1)/2+\epsilon}(\Sigma_0) \times H_e^{(n-1)/2+\epsilon}(\Sigma_0)$; see also [**79**] for an alternate proof in the case $n = 3$. It is also important to note that Lindblad showed [**57**] that when $n = 3$, it is not possible to further lower the Sobolev exponent without additional structure on the nonlinearities (such as the structure present in the case of the Einstein-vacuum equations).

The symmetric hyperbolic framework, which we outlined in Sect. 1.2.1, is generally not well-suited for proving results going beyond local well-posedness. For wave equations, one can derive significantly more sophisticated L^2-type estimates based on the *vectorfield multiplier* and *vectorfield commutator methods*, which are collectively known as the *vectorfield method*. The family of *generalized energy estimates* afforded by these methods is much larger than the family of energy estimates afforded by the symmetric hyperbolic framework in the sense that one has great freedom in the choice of multipliers, differential operator commutators, and surfaces of integration. Both methods are important for deriving the dispersive properties of waves and play a key role in our proofs of the main results of this monograph. In the present section, we focus on the multiplier method. We first review the basic framework, which is applicable to general nonlinear wave equations. For illustration, we then show how the multiplier method can be used to derive L^2-type dispersive estimates for solutions to the linear wave equation. In Sect. 1.2.3, we focus on the commutator method which, when combined with the multiplier method,

[28]In wave coordinates, the Einstein-vacuum equations are equivalent to a system of wave equations of the form $(g^{-1})^{\alpha\beta}(\Psi)\partial_\alpha\partial_\beta\Psi^I = \mathcal{N}^{\alpha\beta}(\Psi)\partial_\alpha\Psi\partial_\beta\Psi^I$, where $\Psi = \{\Psi^I\}$ is an array consisting of the components of the metric.

can be used to derive refined pointwise decay estimates exhibiting the directionally dependent dispersive properties of linear waves.

We now outline the main ideas behind the multiplier method. We consider the nonlinear wave equation (1.59) because of its relevance for the main results of this monograph; see Chapter 10 for a detailed version of the method for this equation in the context of the problem of shock formation. At the heart of the approach lies the *energy-momentum tensorfield* $Q[\Psi]$, a type $\binom{0}{2}$ tensorfield on $\mathbb{R}^{1+n}$ defined by (see Footnote 12 on pg. 12 regarding our index conventions)

$$Q_{\mu\nu}[\Psi] := \mathscr{D}_\mu \Psi \mathscr{D}_\nu \Psi - \frac{1}{2} g_{\mu\nu} (g^{-1})^{\alpha\beta} \mathscr{D}_\alpha \Psi \mathscr{D}_\beta \Psi. \tag{1.62}$$

The quantity $Q[\Psi]$ has the following well-known key properties (the first is easy to prove while for the second, readers may consult [**18**] for the main ideas behind the proof):

(**1**) For solutions to (1.59), we have[29]

$$\mathscr{D}_\alpha Q^\alpha_{\ \nu}[\Psi] = \mathcal{N}(\Psi, \partial\Psi) \mathscr{D}_\nu \Psi. \tag{1.63}$$

(**2**) For future-directed (see Footnote 16 on pg. 13), causal[30] vectors V and W, we have the *dominant energy condition*

$$Q_{\alpha\beta}[\Psi] V^\alpha W^\beta \geq 0. \tag{1.64}$$

As we will now explain, the properties (1.63) and (1.64), when combined with the divergence theorem, allow one to derive a large family of coercive L^2-type energy estimates. This is the multiplier method in action.

For bookkeeping purposes in the divergence theorem, it is convenient to introduce the following *compatible current* vectorfield ${}^{(V)}J^\nu[\Psi]$, which depends on an auxiliary *multiplier vectorfield* V:

$${}^{(V)}J^\nu[\Psi] := Q^\nu_{\ \alpha}[\Psi] V^\alpha. \tag{1.65}$$

Using (1.63) and the symmetry of $Q[\Psi]$, we compute that for *solutions* to (1.59), we have the following identity:

$$\mathscr{D}_\alpha {}^{(V)}J^\alpha[\Psi] = \frac{1}{2} Q^{\alpha\beta}[\Psi] {}^{(V)}\pi_{\alpha\beta} + \mathcal{N}(\Psi, \partial\Psi) V\Psi. \tag{1.66}$$

In (1.66), $V\Psi = V^\alpha \partial_\alpha \Psi$ denotes the V directional derivative of Ψ (see Footnote 10 on pg. 8),

$${}^{(V)}\pi_{\mu\nu} := \mathcal{L}_V g_{\mu\nu} = \mathscr{D}_\mu V_\nu + \mathscr{D}_\nu V_\mu \tag{1.67}$$

denotes the deformation tensor[31] of V, and $\mathcal{L}_V$ denotes Lie differentiation with respect to V (see Def. 3.55).

---

[29] Here and throughout the remainder of the monograph, we lower and raise indices with g and g^{-1}.

[30] By definition, causal vectors V are such that $g(V,V) := g_{\alpha\beta} V^\alpha V^\beta \leq 0$.

[31] The second equality in (1.67) is a consequence of the torsion-free property of the connection $\mathscr{D}$.

By integrating the identity (1.66) over a suitable spacetime region $\mathcal{M} \subset \mathbb{R}^{1+n}$ bounded by spacelike[32] and/or null[33] hypersurfaces and applying the divergence theorem, we obtain an energy identity, in analogy with (1.53). By (1.65), the corresponding integrals along the bounding hypersurfaces appearing in the divergence theorem feature the integrands $Q_{\alpha\beta}[\Psi]V^\alpha N^\beta$, where N is the unit normal[34] vectorfield N along the bounding hypersurface. Hence, by (1.64) and the fact that $Q[\Psi]$ is quadratic in the derivatives of Ψ, we infer that *the integrals along the bounding hypersurfaces are coercive in the derivatives of* Ψ whenever V and N are both future-directed and causal. It is because the integrals are coercive that we refer such identities as *generalized energy identities* and the resulting estimates as *generalized energy estimates*. In contrast, *the theory of symmetric hyperbolic systems provides (up to scalar function multiples) only one compatible current that leads to a coercive energy identity*, namely the compatible energy current (1.46).

REMARK 1.7 (**Connection between $^{(V)}\pi$ and Noether's theorem**). Note that by the identity (1.66) and the divergence theorem argument described above, the tensorfield $^{(V)}\pi_{\mu\nu}$ is connected to the availability (or not) of *conservation laws* for solutions to the wave equation. In particular, if the right-hand side of (1.59) vanishes and if V is a Killing field[35] of g, then (1.66) implies that $\mathscr{D}_\alpha {}^{(V)}J^\alpha[\Psi] = 0$. In particular, there is a conserved quantity associated to the current (1.65); see below for some simple but important examples. This phenomenon may be viewed as a geometric version of *Noether's theorem*, tailored to wave equations. We note that most metrics g do not admit any Killing fields. Thus, generally speaking, in order to prove a global result for a nonlinear problem, one must construct vectorfields V that not only yield coercive energies (upon integrating the corresponding compatible current over a suitable domain), but that also are such that one can control the error integral generated by the first term[36] on the right-hand side of (1.66).

For the purpose of illustration, we now derive some coercive energy identities for solutions to the standard linear wave equation[37]

$$\square_m \Psi = 0, \tag{1.68}$$

where m is the Minkowski metric on $\mathbb{R}^{1+n}$. The simplest nontrivial example occurs when $V = \partial_t$ (relative to standard rectangular coordinates) and $\mathcal{M} = \cup_{t'=0}^t \Sigma_{t'}$ (see (1.56) for the definition of Σ_t). It is straightforward to compute that $^{(\partial_t)}\pi = 0$, $N = \partial_t$, and $^{(\partial_t)}J^\alpha[\Psi]N_\alpha = \frac{1}{2}\sum_{\alpha=0}^n (\partial_\alpha\Psi)^2$. The resulting energy identity yields the

[32]A spacelike hypersurface Σ is such that at each point, the future-directed unit normal N is timelike (that is, it has a negative length as measured by g). In particular, we have $g(N,N) = -1$.

[33]A *null hypersurface* has normal vectors that are tangent to the hypersurface and that have length 0 as measured by g. A particular kind of null hypersurface, namely null cones, forms the lateral boundaries of the integration region that we use in our proof of shock formation for wave equations in three spatial dimensions; see Prop. 10.13.

[34]The phrase "unit normal" is not accurate if the bounding hypersurface is null, since in this case any normal is null (that is, it has length 0 as measured by g).

[35]By definition, Killing fields of g are vectorfields such that $^{(V)}\pi = 0$.

[36]In some cases, this first term contains a piece with a favorable sign that can be used to control other error terms. This is the case in the shock formation problem that we study later in the monograph; see the discussion surrounding equation (2.53).

[37]Relative to standard rectangular coordinates $\{x^\alpha\}_{\alpha=0,1,\cdots,n}$ on $\mathbb{R}^{1+n}$, $\square_m = (m^{-1})^{\alpha\beta}\partial_\alpha\partial_\beta$, where $(m^{-1})^{\alpha\beta} = \mathrm{diag}(-1,1,1\cdots,1)$ is the inverse Minkowski metric.

well-known conservation of the L^2 norm of the gradient:

$$\sum_{\alpha=0}^{n}\int_{\Sigma_t}(\partial_\alpha\Psi)^2\,d^nx=\sum_{\alpha=0}^{n}\int_{\Sigma_0}(\partial_\alpha\Psi)^2\,d^nx, \tag{1.69}$$

where $d^nx := dx^1\cdots dx^n$.

Although it is of fundamental importance, the energy identity (1.69) does not contain any information about the dispersive behavior of linear waves. A much more refined L^2-type estimate, exhibiting some aspects of wave dispersion, can be obtained by replacing the multiplier ∂_t with the Minkowskian *Morawetz multiplier* K, defined by

$$K := (t^2+r^2)\partial_t + 2tr\partial_r = \frac{1}{2}(t+r)^2 L_{(Flat)} + \frac{1}{2}(t-r)^2\underline{L}_{(Flat)}. \tag{1.70}$$

In (1.70),

$$L_{(Flat)} := \partial_t + \partial_r, \qquad \underline{L}_{(Flat)} := \partial_t - \partial_r \tag{1.71}$$

are the standard (future) outgoing/ingoing radial Minkowski-null[38] vectorfields. Above and throughout,

$$r = \sqrt{\sum_{a=1}^{n}(x^a)^2} \tag{1.72}$$

is the standard Euclidean radial coordinate on $\mathbb{R}^n$ and

$$\partial_r = \frac{x^a}{r}\partial_a \tag{1.73}$$

is the standard Euclidean radial vectorfield. The vectorfield K is a conformal Killing field of m, which means that its deformation tensor is a scalar function multiple of m. More precisely, simple calculations yield (see (1.67)) that

$$^{(K)}\pi_{\mu\nu} = 4tm_{\mu\nu}. \tag{1.74}$$

As a consequence of (1.66) and (1.74), we have that[39] $\mathscr{D}_\alpha {}^{(K)}J^\alpha[\Psi] \neq 0$, which, in light of Remark 1.7, means that $^{(K)}J$ does not directly lead to a conservation law. However, as is described in [**48**], we can modify $^{(K)}J[\Psi]$ by adding correction terms to create a divergence-free vectorfield. Specifically, we define (see Footnote 29 on pg. 20 regarding the notation)

$$^{(K+Correction)}J^\nu[\Psi] := {}^{(K)}J^\nu[\Psi] + (n-1)t\Psi\partial^\nu\Psi - \frac{n-1}{2}\Psi^2\partial^\nu t, \tag{1.75}$$

and straightforward calculations based on (1.66) and (1.74) yield that for solutions to (1.68), we have

$$\mathscr{D}_\alpha {}^{(K+Correction)}J^\alpha[\Psi] = 0. \tag{1.76}$$

[38]By Minkowski-null, we mean that $m(L_{(Flat)}, L_{(Flat)}) = m(\underline{L}_{(Flat)}, \underline{L}_{(Flat)}) = 0$, where $m(V,W) := m_{\alpha\beta}V^\alpha W^\beta$.

[39]Note that when the metric is equal to the Minkowski metric m, the operator $\mathscr{D}_\alpha$ agrees, relative to rectangular coordinates, with the standard partial derivative operator ∂_α.

As in (1.69), we can integrate the identity (1.76) over the spacetime domain $[0,t] \times \mathbb{R}^n$ and invoke the divergence theorem to obtain the following conservation law (again with $N = \partial_t$):

$$\int_{\Sigma_t} {}^{(K+Correction)}J^\alpha[\Psi] N_\alpha \, d^n x = \int_{\Sigma_0} {}^{(K+Correction)}J^\alpha[\Psi] N_\alpha \, d^n x. \tag{1.77}$$

The identity (1.77) is useful only if the integrals $\int_{\Sigma_t} \cdots$ are coercive. As we now describe, they are in fact coercive in a rather strong sense when $n \geq 3$ due to the weights inherent in the definition (1.70) of K. The main estimate of interest to us is that when $n \geq 3$, there is a constant $C > 0$ depending on n such that

$$\begin{aligned} &\int_{\Sigma_t} {}^{(K+Correction)}J^\alpha[\Psi] N_\alpha \, d^n x \\ &\geq \frac{1}{C} \int_{\Sigma_t} (t+r)^2 (L_{(Flat)}\Psi)^2 \, d^n x + \frac{1}{C} \int_{\Sigma_t} (t-r)^2 (\underline{L}_{(Flat)}\Psi)^2 \, d^n x \\ &\quad + \frac{1}{2} \int_{\Sigma_t} (t^2+r^2) |\nabla\!\!\!/_{\not{h}} \Psi|^2 \, d^n x + \frac{1}{C} \int_{\Sigma_t} \frac{t^2+r^2}{r^2} \Psi^2 \, d^n x. \end{aligned} \tag{1.78}$$

In (1.78), $\nabla\!\!\!/_{\not{h}}$ is the Levi-Civita connection of $\not{h}$, the Riemannian metric on the Euclidean spheres of constant t and r induced by the Minkowski metric m. Thus, $|\nabla\!\!\!/_{\not{h}}\Psi|$ is the size[40] of the angular derivatives of Ψ. The proof of (1.78) is more involved than the proof of (1.69) and requires some new ingredients. Below we will provide a proof using arguments given in [**48**]. Equation (1.77) and inequality (1.78) together exhibit some important L^2-type dispersive properties of linear waves showing that, for example, the gradient of Ψ cannot remain concentrated (in the $L^2(\mathbb{R}^n)$ sense) in any compact set of $\mathbb{R}^n$ as $t \to \infty$. It also shows that $L_{(Flat)}\Psi$ and $\nabla\!\!\!/_{\not{h}}\Psi$ enjoy stronger decay properties than $\underline{L}_{(Flat)}\Psi$. Estimates in this vein play an important role in many nonlinear problems, including our analysis of shock forming waves.[41]

To prove (1.78), we first use (1.62), (1.65), (1.75), and the fact that $N = \partial_t$ along Σ_t to compute that (see the proof of Lemma 10.7 for inspiration on how to carry out the computations)

$$\begin{aligned} &{}^{(K+Correction)}J^\alpha[\Psi] N_\alpha \\ &= \frac{1}{4}(t+r)^2 (L_{(Flat)}\Psi)^2 + \frac{1}{4}(t-r)^2 (\underline{L}_{(Flat)}\Psi)^2 + \frac{1}{2}(t^2+r^2)|\nabla\!\!\!/_{\not{h}}\Psi|^2 \\ &\quad + (n-1)t\Psi\partial_t\Psi - \frac{n-1}{2}\Psi^2. \end{aligned} \tag{1.79}$$

---

[40] The implicit metric in the norm $|\nabla\!\!\!/_{\not{h}}\Psi|$ is $\not{h}$.

[41] Our analysis in the shock formation problem for wave equations takes place in a region trapped in between two outgoing null cones. Because the width of the region is not too large, it turns out that in order to close our estimates, we can rely on a simplified Morawetz multiplier that does not involve an analog of the term $\frac{1}{2}(t-r)^2\underline{L}_{(Flat)}$ from the definition (1.70); see (10.11b). However, our analog of the identity (1.77) also involves square integrals along the outgoing cones that provide new coercive information going beyond that provided by the Σ_t integrals; see Prop. 10.13 and Lemma 14.1. These additional "cone fluxes" play a critical role in closing the energy estimates in the shock formation problem.

It is not immediately apparent from (1.79) that $\int_{\Sigma_t} {}^{(K+Correction)}J^\alpha[\Psi]N_\alpha \, d^n x$ is coercive in the sense of (1.78). Our proof of (1.78) relies on the following two identities for the product $t\Psi\partial_t\Psi$, which are easy to verify (see Footnote 12 on pg. 12 regarding our summation convention):

$$t\Psi\partial_t\Psi = \Psi(t\partial_t\Psi + r\partial_r\Psi) + \frac{n}{2}\Psi^2 - \frac{1}{2}\partial_a\left\{x^a\Psi^2\right\}, \tag{1.80a}$$

$$t\Psi\partial_t\Psi = \frac{t}{r}\Psi(r\partial_t\Psi + t\partial_r\Psi) + \frac{n-2}{2}\frac{t^2}{r^2}\Psi^2 - \frac{1}{2}\partial_a\left\{t^2\frac{x^a}{r^2}\Psi^2\right\}, \tag{1.80b}$$

as well as the identity

$$\begin{aligned}\frac{1}{4}(t+r)^2(L_{(Flat)}\Psi)^2 &+ \frac{1}{4}(t-r)^2(\underline{L}_{(Flat)}\Psi)^2 \\ &= \frac{1}{2}(t\partial_t\Psi + r\partial_r\Psi)^2 + \frac{1}{2}(r\partial_t\Psi + t\partial_r\Psi)^2.\end{aligned} \tag{1.81}$$

To proceed, we decompose the factor $n-1$ from (1.79) as

$$n - 1 = \alpha + \beta,$$

where we will choose the positive constants α and β below. We then rewrite the product $(n-1)t\Psi\partial_t\Psi$ from (1.79) as $\alpha t\Psi\partial_t\Psi + \beta t\Psi\partial_t\Psi$, and we express $\alpha t\Psi\partial_t\Psi$ as α times the right-hand side of (1.80a) and $\beta t\Psi\partial_t\Psi$ as β times the right-hand side of (1.80a). Next, integrating by parts over Σ_t on the terms $\alpha t\Psi\partial_t\Psi$ and $\beta t\Psi\partial_t\Psi$ and using (1.80a), (1.80b), and (1.81), we derive the following identity:

$$\begin{aligned}&\int_{\Sigma_t} {}^{(K+Correction)}J^\alpha[\Psi]N_\alpha \, d^n x \\ &= \frac{1}{2}\int_{\Sigma_t} (t\partial_t\Psi + r\partial_r\Psi)^2 + 2\alpha\Psi(t\partial_t\Psi + r\partial_r\Psi) + \{\alpha n - (n-1)\}\Psi^2 \, d^n x \\ &+ \frac{1}{2}\int_{\Sigma_t} (r\partial_t\Psi + t\partial_r\Psi)^2 + 2\beta\frac{t}{r}\Psi(r\partial_t\Psi + t\partial_r\Psi) + \beta(n-2)\frac{t^2}{r^2}\Psi^2 \, d^n x \\ &+ \frac{1}{2}\int_{\Sigma_t} (t^2 + r^2)|\nabla\!\!\!/_{m}\Psi|^2 \, d^n x.\end{aligned} \tag{1.82}$$

We now choose

$$\alpha := \frac{n}{2}, \qquad \beta := \frac{n-2}{2} \tag{1.83}$$

in (1.82). It is straightforward to check that under this choice, when $n \geq 3$, the cross terms $2\alpha\cdots$ and $2\beta\cdots$ in the first two integrands on the right-hand side of (1.82) are dominated by the positive definite terms $(t\partial_t\Psi + r\partial_r\Psi)^2$, $(r\partial_t\Psi + t\partial_r\Psi)^2$, and Ψ^2. In view of the identity (1.81), it follows that when $n \geq 3$, there exists a $C > 0$ such that the desired estimate (1.78) holds.

1.2.3. Pointwise dispersion in solutions to wave equations. In this section, we complement the L^2-type dispersive estimates derived in Sect. 1.2.2, which were based on the multiplier method, with pointwise dispersive (decay) estimates that rely on the multiplier and commutator methods. As is well known [**48**], solutions to the linear wave equation $\square_m\Psi = 0$ in $\mathbb{R}^{1+n}$ with initial data having

Fourier support[42] contained in the dyadic shell $\{\xi \in \mathbb{R}^n \mid \frac{1}{2} \leq |\xi| \leq 2\}$ verify the *basic dispersive estimate*[43]

$$|\Psi(t,x)| \leq C(1+t)^{\frac{1-n}{2}} \sum_{\alpha=0}^{n} \|\partial_\alpha \Psi\|_{W_e^{\frac{n-1}{2},1}(\Sigma_t)}, \qquad (t,x) \in [0,\infty) \times \mathbb{R}^n. \tag{1.84}$$

The standard proof of (1.84) is based on the method of stationary phase applied to the Fourier representation of the solution. Although the estimate (1.84) is useful in linear theory, there are two ways in which it is inadequate for studying the kinds of quasilinear equations of interest to us in this monograph. The first is that L^1-type norms are not suitable, for in quasilinear problems, the only kinds of regularity propagated at top-order are L^2-based. The second is that (1.84) fails to capture more refined, directionally-dependent dispersive behavior, which often plays a critical role in the study of solutions to nonlinear equations. In particular, as we describe in Sect. 2.6, the refined behavior plays an important role in our proof of shock formation.

Fortunately, Klainerman developed [**46**] the *commutator method*, a robust L^2-type vectorfield approach to deriving refined pointwise dispersive estimates. To explain his approach, which is connected to the one that we use throughout this monograph, we first introduce a *Minkowskian null (vectorfield) frame*

$$\{L_{(Flat)} := \partial_t + \partial_r,\ \underline{L}_{(Flat)} := \partial_t - \partial_r,\ X_{1;(Flat)}, \cdots, X_{n-1;(Flat)}\}, \tag{1.85}$$

where $\cup_{a=1}^{n-1}\{X_{a;(Flat)}\}$ is a locally defined vectorfield frame spanning the tangent space of the $n-1$-dimensional Euclidean spheres $\{t = \text{const}\} \cap \{r = \text{const}\}$ whenever $r > 0$. In its most basic form, the commutator method is based on commuting the wave equation with a subset of Killing and conformal Killing fields[44] of the Minkowski metric. A convenient set of commutators is the $\mathscr{Z}_{(Flat)}$, whose elements can be expressed relative to rectangular coordinates as follows:

$$\mathscr{Z}_{(Flat)} := \cup_{\alpha=0}^{n}\{\partial_\alpha\} \bigcup \cup_{1 \leq i < j \leq n}\{x^i \partial_j - x^j \partial_i\} \bigcup \cup_{i=1}^{n}\{x^i \partial_t + t\partial_i\} \cup \{x^\alpha \partial_\alpha\}. \tag{1.86}$$

In (1.86), the ∂_α are the translations, the $x^i\partial_j - x^j\partial_i$ are the Euclidean rotations, the $x^i\partial_t + t\partial_i$ are the Lorentz boosts, and $x^\alpha\partial_\alpha$ is the scaling vectorfield. All vectorfields in $\mathscr{Z}_{(Flat)}$ are Killing except for the scaling vectorfield, which is conformal Killing and verifies $\mathcal{L}_{x^\alpha\partial_\alpha} m = {}^{(x^\alpha\partial_\alpha)}\pi = 2m$.

---

[42] The restriction on the Fourier support of the data can be removed by replacing the Euclidean Sobolev norm $\|\cdot\|_{W_e^{\frac{n-1}{2},1}(\Sigma_t)}$ on the right-hand side of (1.84) with an appropriate $L^1(\mathbb{R}^n)$-type Besov norm of the data.

[43] We note that $\|f\|_{W_e^{\frac{n-1}{2},1}(\Sigma_t)} := \sum_{|\vec{I}| \leq \frac{n-1}{2}} \left\|\partial^{\vec{I}} f\right\|_{L_e^1(\Sigma_t)}$, where $\vec{I}$ denotes a multi-index corresponding to repeated differentiation with respect to the rectangular spatial coordinate vectorfields ∂_i.

[44] We recall that Killing fields are vectorfields whose deformation tensors (1.67) vanish while conformal Killing fields have deformation tensors that are a scalar-valued function times the Minkowski metric.

A simple calculation reveals that a general vectorfield V has the following commutation property with the Minkowskian wave operator $\square_m$:

$$[\square_m, V]\Psi = (\mathscr{D}_\alpha {}^{(V)}\pi^{\alpha\beta})\mathscr{D}_\beta\Psi - \frac{1}{2}(\mathscr{D}^\alpha \mathrm{tr}_m {}^{(V)}\pi)\mathscr{D}_\alpha\Psi + {}^{(V)}\pi^{\alpha\beta}\mathscr{D}^2_{\alpha\beta}\Psi, \tag{1.87}$$

where $[P, Q]$ denotes the commutator of P and Q and $\mathrm{tr}_m {}^{(V)}\pi = (m^{-1})^{\alpha\beta\,(V)}\pi_{\alpha\beta}$ is the trace of ${}^{(V)}\pi$ (see (1.67) for the definition of ${}^{(V)}\pi$). An important consequence of (1.87) is that the vectorfields in $\mathscr{Z}_{(Flat)}$ enjoy the following good commutation properties with $\square_m$:

$$[\square_m, Z_{(Flat)}] = 0, \qquad \text{if } Z_{(Flat)} \in \mathscr{Z}_{(Flat)}\backslash\{x^\alpha\partial_\alpha\}, \tag{1.88a}$$

$$[\square_m, x^\alpha\partial_\alpha] = 2\square_m. \tag{1.88b}$$

In particular, if

$$\square_m\Psi = 0, \tag{1.89}$$

then $\square_m \mathscr{Z}^{\vec{I}}_{(Flat)}\Psi = 0$ for any multi-index $\mathscr{Z}^{\vec{I}}_{(Flat)}$ of vectorfield operators constructed out of the elements of $\mathscr{Z}_{(Flat)}$. Thus, for any integer $N \geq 0$, we deduce from (1.69) that solutions to (1.89) verify (see Remark 1.4 regarding the notation for the norms)

$$\sum_{|\vec{I}|\leq N}\sum_{\alpha=0}^{n}\left\|\partial_\alpha \mathscr{Z}^{\vec{I}}_{(Flat)}\Psi\right\|_{L^2_e(\Sigma_t)} = \sum_{|\vec{I}|\leq N}\sum_{\alpha=0}^{n}\left\|\partial_\alpha \mathscr{Z}^{\vec{I}}_{(Flat)}\Psi\right\|_{L^2_e(\Sigma_0)}, \tag{1.90}$$

where the right-hand side of (1.90) is determined by the data $(\Psi|_{\Sigma_0}, \partial_t\Psi|_{\Sigma_0})$.

Some important pointwise dispersive properties of linear waves now follow immediately from (1.90) and the well-known Klainerman-Sobolev inequality,[45] the first version of which was proved in [**46**], while the version (1.91) is found in [**76**]:

(1.91)

$$(1+|t+r|)^{\frac{n-1}{2}}(1+|t-r|)^{1/2}\sum_{\alpha=0}^{n}|\partial_\alpha\Psi(t,x)| \leq C\sum_{|\vec{I}|\leq\frac{n+2}{2}}\sum_{\alpha=0}^{n}\left\|\partial_\alpha \mathscr{Z}^{\vec{I}}_{(Flat)}\Psi\right\|_{L^2_e(\Sigma_t)}.$$

Moreover, the estimate (1.91) also holds with $\partial_\alpha\Psi$ replaced by Ψ on the left and $\partial_\alpha \mathscr{Z}^{\vec{I}}_{(Flat)}\Psi$ replaced by $\mathscr{Z}^{\vec{I}}_{(Flat)}\Psi$ on the right. The proof of (1.91) is based on modifying the standard proof of Sobolev embedding to take advantage of the weights inherent in the definitions of the Euclidean rotations, Lorentz boosts, and the scaling vectorfield. Moreover, as a corollary, in the region $\{t \geq 0\}$, we have the following decay estimates for the derivatives of $\partial_\alpha\Psi$ with respect to the null frame vectorfields

[45] Inequality (1.91) is also known as the global Sobolev inequality.

(see [**76**] for a proof):

$$(1.92a)\quad (1+|t+r|)^{\frac{n+1}{2}}(1+|t-r|)^{1/2}\sum_{\alpha=0}^{n}\left|L_{(Flat)}\partial_\alpha\Psi(t,x)\right| \leq C\sum_{|\vec{I}|\leq\frac{n+4}{2}}\sum_{\alpha=0}^{n}\left\|\partial_\alpha\mathscr{Z}_{(Flat)}^{\vec{I}}\Psi\right\|_{L^2_e(\Sigma_t)},$$

$$(1.92b)\quad (1+|t+r|)^{\frac{n+1}{2}}(1+|t-r|)^{1/2}\sum_{\alpha=0}^{n}\sum_{a=1}^{n-1}\left|X_{a;(Flat)}\partial_\alpha\Psi(t,x)\right| \leq C\sum_{|\vec{I}|\leq\frac{n+4}{2}}\sum_{\alpha=0}^{n}\left\|\partial_\alpha\mathscr{Z}_{(Flat)}^{\vec{I}}\Psi\right\|_{L^2_e(\Sigma_t)},$$

$$(1.92c)\quad (1+|t+r|)^{\frac{n-1}{2}}(1+|t-r|)^{3/2}\sum_{\alpha=0}^{n}\left|\underline{L}_{(Flat)}\partial_\alpha\Psi(t,x)\right| \leq C\sum_{|\vec{I}|\leq\frac{n+4}{2}}\sum_{\alpha=0}^{n}\left\|\partial_\alpha\mathscr{Z}_{(Flat)}^{\vec{I}}\Psi\right\|_{L^2_e(\Sigma_t)}.$$

Note that (1.92a) and (1.92b) imply that the derivatives of $\partial_\alpha\Psi$ with respect to the vectorfields $\{L_{(Flat)}, X_{1;(Flat)}, \cdots, X_{n-1;(Flat)},\}$, which span the tangent space of the outgoing Minkowski null cones $\{t-r=\text{const}\}$ when $r>0$, enjoy the extra decay rate factor of $(1+t+r)^{-1}$ relative to the Klainerman-Sobolev inequality (1.91). In contrast, $\underline{L}_{(Flat)}$ is transversal to the outgoing Minkowski null cones, and $\underline{L}_{(Flat)}\partial_\alpha\Psi$ gains only the decay rate factor $(1+|t-r|)^{-1}$ relative to the Klainerman-Sobolev inequality. The proofs of (1.92a)-(1.92c) are based on the Klainerman-Sobolev inequality (1.91) together with algebraic identities involving weighted combinations of the vectorfields in $\mathscr{Z}_{(Flat)}$. For example, the derivations of (1.92a) and (1.92c) rely in part on the identities

$$(1.93a)\quad (t+r)L_{(Flat)} = x^\alpha\partial_\alpha + \frac{x^a}{r}\left\{x^a\partial_t + t\partial_a\right\},$$

$$(1.93b)\quad (t-r)\underline{L}_{(Flat)} = x^\alpha\partial_\alpha - \frac{x^a}{r}\left\{x^a\partial_t + t\partial_a\right\}.$$

1.2.4. Almost global existence and global existence via the classic null condition. In Sects. 1.2.2 and 1.2.3, we studied some of the dispersive properties of linear waves. In the present section, we discuss the effect that dispersion has on small-data solutions to nonlinear wave equations, with an emphasis on the case of three spatial dimensions. For convenience, we restrict our attention to Cauchy problems of the form

$$(1.94a)\quad (g^{-1})^{\alpha\beta}(\Phi,\partial\Phi)\partial_\alpha\partial_\beta\Phi = \mathcal{N}(\Phi,\partial\Phi),$$

$$(1.94b)\quad (\Phi|_{t=0},\partial_t\Phi|_{t=0}) = (\mathring{\Phi},\mathring{\Phi}_0).$$

We assume that relative to rectangular coordinates $(x^0=t,x^1,\cdots,x^n)$ on $\mathbb{R}^{1+n}$,

$$(1.95)\quad g_{\alpha\beta}(0,0) = m_{\alpha\beta},$$

$$(1.96)\quad \mathcal{N}(0,0) = \frac{\partial\mathcal{N}}{\partial\Phi}(0,0) = \frac{\partial\mathcal{N}}{\partial(\partial\Phi)}(0,0) = 0,$$

where $m_{\alpha\beta} = \mathrm{diag}(-1, 1, 1, \cdots, 1)$ is the Minkowski metric. Note that (1.95)-(1.96) imply that equation (1.94a) is a quadratic perturbation of the linear wave equation. We also note that the local well-posedness proposition (Prop. 1.6) can be extended to apply to equation (1.94a) with one notable change: the Sobolev exponent must be increased to $s > n/2 + 2$ because of the dependence of g on the first derivatives of Φ in (1.94a).

A primary feature of the solution to the nonlinear problem (1.94a)-(1.94b) is that when the data are small, the linear dispersive effects described in Sects. 1.2.2 and 1.2.3 initially dominate the dynamics. Consequently, the solution decays for a long time, even if a singularity eventually forms. The first results capturing the effect of dispersion on the solution's lifespan in three spatial dimensions were proved by John and Klainerman [**34**, **44**]. Under some mild assumptions on the structure of the nonlinearities, they used dispersive estimates to show that if the data and a certain number their derivatives are of small size $\mathring{\epsilon}$ relative to a suitable weighted Sobolev norm,[46] then the classical lifespan of the solution is at least $\mathcal{O}(\exp(c\mathring{\epsilon}^{-1}))$. This kind of long-time existence result is often referred to as *almost global existence*. The shock forming solutions that we study in this monograph in fact form their first singularity around the time $\mathcal{O}(\exp(c\mathring{\epsilon}^{-1}))$. Klainerman later simplified and extended these results [**46**] using the aforementioned vectorfield commutator and multiplier methods. His proof of almost global existence is similar to the proof of Prop. 1.6, but with some important new features. It is based on commuting the wave equation (1.94a) with the vectorfields in $\mathscr{Z}_{(Flat)}$ defined in (1.86), exploiting the commutation properties (1.88a)-(1.88b), deriving energy estimates using the multiplier ∂_t (much as in (1.69)) in order to control the quantity on the left-hand side of (1.90), and using the Klainerman-Sobolev inequality (1.91) to bound the factor

$$\exp\left(\int_{t'=0}^{t} \sum_{\alpha=0}^{n} \|\partial_\alpha \Psi\|_{L^\infty(\Sigma_{t'})}\, dt'\right) \tag{1.97}$$

from (1.61), which also appears in the energy estimates of [**46**].

In four spatial dimensions, the pointwise dispersive decay provided by the Klainerman-Sobolev inequality (1.91) suggests that the factor (1.97) is uniformly bounded, which in turn suggest small-data global existence. In fact, (1.91) is strong enough to yield small-data global existence for all quadratic nonlinearities except those of the form[47] Φ^2. In five or more spatial dimensions, the inequality (1.91) is strong enough to yield small-data global existence for all quadratic nonlinearities.

In three spatial dimensions, the question of which quadratic nonlinearities allow for small-data global existence is much more delicate. Klainerman identified an important class of nonlinearities with a now-famous structural property allowing for small-data global existence. He called this property the (classic) *null condition* [**45**] (see Footnote 22 on pg. xix regarding the terminology). Distinct proofs of

[46] Roughly, the norms are equivalent to the right-hand side of (1.91) at $t = 0$.

[47] The main obstacle in handling nonlinearities of the form Φ^2 is that energy estimates in the spirit of (1.69) for the nonlinear equation do not directly provide bounds for the quantities $\|\mathscr{Z}_{(Flat)}^{\vec{I}}\Phi\|_{L_e^2(\Sigma_t)}$, which are needed to control solutions to the $\mathscr{Z}_{(Flat)}^{\vec{I}}$-commuted equation. One can derive estimates for these terms by integrating estimates for $\|\partial_t \mathscr{Z}_{(Flat)}^{\vec{I}}\Phi\|_{L_e^2(\Sigma_t)}$ in time, but in four spatial dimensions, the time integration results in a t-dependent factor that is large enough to spoil the proof of global existence.

small-data global existence for nonlinearities verifying the null condition were given by Klainerman [**47**] and Christodoulou [**15**]. There are several equivalent ways to describe it. Our description here is different than Klainerman's original one and is instead closely tied to the decay rates for linear waves outlined in Sect. 1.2.2. We remark that although the null condition can be formulated for systems of equations (see [**47**]), we restrict our attention here to the scalar wave equation (1.94a).

Specifically, the null condition is verified by the nonlinearities in equation (1.94a) if one Taylor expands them around $\mathbf{0}$ to quadratic order, decomposes all derivatives relative to the null frame (1.85), and finds that the following terms are *absent*: **i)** $(\underline{L}_{(Flat)}\Phi)^2$, $(L_{(Flat)}\Phi)^2$; **ii)** $\Phi\underline{L}_{(Flat)}\underline{L}_{(Flat)}\Phi$, $\Phi L_{(Flat)}L_{(Flat)}\Phi$. The canonical example of a quadratic term verifying Klainerman's null condition is $(m^{-1})^{\alpha\beta}\partial_\alpha\Phi\partial_\beta\Phi$, where m is the Minkowski metric; see Sect. 2.4 for a proof of this fact. The main point is that when these terms are absent, the only kinds of quadratic terms that are present have at least one factor that, thanks to the estimates (1.92a)-(1.92c) and their higher-order analogs, are expected to decay faster than $(1+t)^{-1}$. For clarity, we note that the products $(\underline{L}_{(Flat)}\Phi)^2$ and $\Phi\underline{L}_{(Flat)}\underline{L}_{(Flat)}\Phi$ are possible obstructions to global existence in the region $\{t \geq 0\}$, while the products $(L_{(Flat)}\Phi)^2$ and $\Phi L_{(Flat)}L_{(Flat)}\Phi$ are possible obstructions to global existence in the region $\{t \leq 0\}$.

In three spatial dimensions, a glaring question is: what is the global behavior of small-data solutions to equation (1.94a) when the null condition fails? Note that the Klainerman-Sobolev inequality (1.91) suggests that when $n = 3$ and the data are small, the factor

$$\sum_{\alpha=0}^{n}\int_{t'=0}^{t}\|\partial_\alpha\Psi\|_{L^\infty(\Sigma_{t'})}\,dt'$$

from (1.61) is logarithmically divergent as $t \to \infty$. One suspects that in the absence of special structure such as the null condition, this divergence can lead to the formation of a singularity. This suspicion is in fact borne out in many cases: for many scalar wave equations failing the null condition, including the ones that we study in this monograph, singularities eventually form. At the beginning of Chapter 2, we survey the blow-up results that are known when the classic null condition fails and place the sharp small-data shock formation results of the present monograph in context.