Preface

A central issue surrounding the study of quasilinear hyperbolic PDEs is that classical solutions, generated by smooth initial conditions, can develop singularities in finite time. In principle, many different kind of singularities are possible. The subject of this monograph is perhaps the most well-known type: shocks. Our work here primarily concerns scalar quasilinear wave equations, the main reasons being **1)** they arise in many important mathematical, physical, and geometric contexts; **2)** the last few decades have led to the development of advanced machinery tailored to such equations; and **3)** their characteristic hypersurfaces[1] are relatively simple[2] and can be analyzed using tools and concepts from the well-developed theory of Lorentzian geometry.

Before the groundbreaking work of S. Alinhac [**2**, **4**–**6**], very little was known about shock formation except in problems that are effectively one (spatial) dimensional. One-dimensional shock formation results are of course classical. We describe some of the most important such results in Chapter 1. As we describe in detail in Chapter 2, Alinhac proved finite-time shock formation for a class of wave equations in two and three spatial dimensions. Roughly, his proof applied to wave equations of the form[3]

$$(h^{-1})^{\alpha\beta}(\partial\Phi)\partial_\alpha\partial_\beta\Phi = 0$$

whenever the nonlinear terms *fail* to satisfy the null condition, which was first formulated by S. Klainerman [**45**] in three spatial dimensions and later in two spatial dimensions by Alinhac [**1**, **2**]. In the above wave equation, $h = h(\partial\Phi)$ is a Lorentzian metric such that $h(\partial\Phi = 0) = m$, where m is the standard Minkowski metric on $\mathbb{R}^{1+n}$, $n \in \{2, 3\}$. Alinhac's main results concern solutions generated by small initial data that belong to a suitable Sobolev space and that verify a nondegeneracy condition. The foundation of his proof was a new system of *geometric coordinates* tied to an *eikonal function* u, which by definition is a solution to the *eikonal equation*, that is, the hyperbolic PDE

$$(h^{-1})^{\alpha\beta}(\partial\Phi)\partial_\alpha u\partial_\beta u = 0,$$

supplemented by appropriate initial conditions. The level sets of u are true characteristics (as opposed to approximate ones) corresponding to the nonlinear wave equation.

---

[1] In the study of wave equations, characteristic hypersurfaces are often referred to as "null hypersurfaces" in view of their connection to the Lorentzian notion of a null vectorfield. More generally, they are often referred to as simply "the characteristics."

[2] Roughly, in our study of wave equations, the characteristics are a family of "true curved cones" corresponding to the dynamic Lorentzian metric of the wave equation.

[3] The equation is to be interpreted as an equation given relative to standard rectangular coordinates, which we describe at the beginning of Chapter 2. Moreover, here and throughout, we use Einstein's summation convention.

Eikonal functions are perhaps best known for the central role they played in Christodoulou-Klainerman's celebrated proof [**19**] of the stability of Minkowski spacetime.[4] That work was the first instance in which eikonal functions were used to prove a global nonlinear result for a hyperbolic PDE. In the study of shock formation, the behavior of u, its properties, and their connection to the behavior of the solution variable[5] lie at the heart of the analysis. This was the case in Alinhac's work and in Christodoulou's work (described two paragraphs below), and it remains true in the present monograph as well. As we will see, in the problem of shock formation, the eikonal function plays an even more important role than it does in the proof of stability of Minkowski spacetime; there is an alternate proof [**59**] of the stability of Minkowski spacetime, due to Lindblad-Rodnianski, that relies on an approximate eikonal function corresponding to the Minkowski metric rather than a true eikonal function. This alternate approach leads to remarkable simplifications in the analysis because the characteristics associated to the Minkowski metric are much simpler.[6] In contrast, in the problem of shock formation, the formation of the shock and the corresponding blow-up of the solution are *exactly tied to the blow-up of the first rectangular coordinate partial derivatives of a true eikonal function.* Thus, there is little hope of finding an alternate proof that avoids the use of a true eikonal function (and, as we will see, the weighty baggage that accompanies it).

It is well-known that energy estimates are an unavoidable aspect of the study of quasilinear wave equations in more than one spatial dimension. To close the energy estimates in his proof of shock formation, Alinhac relied on a Nash-Moser iteration scheme featuring a free boundary. The presence of the free boundary is connected to the blow-up time of the iterates, which can vary, albeit slightly. A fundamental aspect of his proof, which is also present in Christodoulou's work and the present monograph, is that *the solution remains regular*[7] *relative to the geometric coordinates* mentioned two paragraphs above. In the equations studied by Alinhac, the singularity occurs in the second rectangular coordinate partial derivatives of Φ and is tied to the degeneration of the geometric coordinate system relative to the rectangular one. At the same time, the degeneracy is tied to the intersection of the characteristics. In discussing Alinhac's results and related ones, we often refer to the intersection of the characteristics as "the formation of a shock." Our intention in using this terminology is to highlight the following fundamentally important aspect of Alinhac's work: he proved that Φ and its first rectangular coordinate partial derivatives $\partial_\alpha \Phi$ remain bounded all the way up to the singularity. In particular, the singularity occurs at the level of the first rectangular coordinate partial derivatives of the Lorentzian metric components $h_{\alpha\beta}(\partial\Phi)$ and not in the $h_{\alpha\beta}(\partial\Phi)$ themselves; this feature is of fundamental importance for closing the proof.

Although Alinhac's approach is compellingly short, it has some limitations, which we describe in detail in Sect. 2.11.1. In particular, his framework allows one

---

[4] Roughly, [**19**] is a small-data global existence result for the Einstein-vacuum equations.

[5] In the present work, the solution variable is the solution to the wave equation.

[6] In this context, the characteristics associated to the Minkowski metric are the usual flat Minkowski light cones.

[7] Actually, the solution does not necessarily remain regular at the high derivative levels: a fundamental aspect of the proof is that the high-order energies are allowed to blow-up as the shock forms; see three paragraphs below.

to follow the solution only to the first singularity[8] *and not further*. In his 2007 monograph [**17**], D. Christodoulou proved, for a sub-class of Alinhac's wave equations,[9] a breakthrough result that significantly sharpened Alinhac's results and eliminated the drawbacks of his approach. Specifically, Christodoulou's work applies to the wave equations that arise in irrotational relativistic[10] (compressible) fluid mechanics in three spatial dimensions. In this context, the equations are known as the irrotational relativistic Euler equations. Christodoulou assumed that the data have small H^N norm, where N is a sufficiently large (nonexplicit) integer. To deduce the shock formation, he also assumed that the data verify a signed integral inequality, distinct from the nondegeneracy condition of Alinhac mentioned above. Christodoulou's framework allows one to do much more than follow the solution to the first singularity: it provides a complete picture of a portion of the maximal development[11] of the data including a description of the behavior of the solution along the boundary;[12] this is the main advantage of Christodoulou's framework. One of the key reasons that Christodoulou was able to sharpen Alinhac's results is that he was able to close the energy estimates without invoking a Nash-Moser iteration scheme. Moreover, his estimates do not involve a free boundary. Instead, Christodoulou developed a forwards approach relative to a set of geometric coordinates that, like Alinhac's, are tied to a true eikonal function. By forwards approach, we mean that Christodoulou derives traditional global-existence-type estimates for a Cauchy problem in the geometric coordinates, relative to which the solution remains rather smooth (see, however, footnote 7). As in Alinhac's results and those of the present monograph, the blow-up occurs in certain rectangular coordinate partial derivatives of the solution and is tied to the degeneracy of the change of variables map from geometric to rectangular coordinates.

In Christodoulou's framework, the degeneracy mentioned at the end of the previous paragraph and the corresponding blow-up of the solution are mediated by the vanishing of a quantity known as the *inverse foliation density*, which we denote by $\upmu$. Roughly, $\upmu$ is the reciprocal of a derivative of the eikonal function u. Geometrically, $1/\upmu$ is a measure of the density of the level sets of u. As we will see starting in Chapter 2, the vanishing of $\upmu$ is equivalent to the intersection of the characteristics and the blow-up of the eikonal function's first rectangular coordinate partial derivatives, as we mentioned above. Moreover, *these degeneracies are exactly tied to the formation of a singularity in the solution to the wave equation*, much like in the classic example of Burgers' equation (see Sect. 1.1.2). The study of $\upmu$ and the prospect of its vanishing are the main themes of [**17**] and the present work. For reasons that we describe two paragraphs below, the most compelling advantage of Christodoulou's framework is that it allows one to construct the portion of the set

[8]Roughly speaking, Alinhac's proof works when there is such a unique first singularity; his nondegeneracy conditions on the data ensure that this is the case.

[9]One has to take into account some simple differences in normalization, described in Sect. 2.11.2, in order to see that Christodoulou's equations fall under the scope of Alinhac's work.

[10]In [**21**], Christodoulou-Miao extended the result to the nonrelativistic case.

[11]Roughly speaking, the maximal development is the largest classical solution that is uniquely determined by the data.

[12]The boundary can be very complicated and, in particular, it is not contained in a constant-time hypersurface.

$\{\mu = 0\}$ that corresponds to the part of the boundary[13] along which the solution blows up.[14] Generally, the set $\{\mu = 0\}$ "evolves" into a spacetime region lying to the future of the constant-time hypersurface of first blow-up and thus it lies to the future of the region that Alinhac was able to probe.

It turns out that Christodoulou's sharp description is accompanied by severe technical difficulties: the high-order energies are allowed to blow-up as $\mu \to 0$. This difficulty is fundamentally tied to the regularity theory of the eikonal function. As was first shown in [**19**], to control the top derivatives of u, one must invoke a nontrivial procedure based on elliptic estimates and modified quantities, the latter being special combinations of terms that satisfy evolution equations with a good structure. In the problem of shock formation, this procedure introduces, at the top order, a difficult factor of $1/\mu$ into the energy identities; *this factor is the reason that the high-order energies are allowed to blow up.* A related but distinct difficulty is that *one needs to show that the low-order energies remain bounded all the way up to* $\{\mu = 0\}$. This latter step is essential for establishing, via Sobolev embedding, the basic uniform L^∞ estimates that allow one to control error terms and to treat the problem as a traditional one in which one derives global-existence-type estimates (relative to the geometric coordinates). These are the main technical difficulties that one encounters in the problem of shock formation à la Christodoulou and they are a primary reason that the work is technical and lengthy. In Chapter 2, we provide an extended overview of these issues, especially in view of the fact that they do not arise in any other context in the study of nonlinear wave equations.

Christodoulou's framework is compelling for the geometric insight it provides into the formation of shocks and for the sharp description that it yields. In addition, his approach is fundamentally important for a related problem: it turns out that his sharp description of the solution near the boundary of the maximal development is an essential ingredient in setting up the *shock development problem*, which is the problem of **1)** continuing the solution to Euler's equations (in a weak sense, subject to appropriate jump and entropy conditions) beyond the first singularity and **2)** at the same time, constructing the shock hypersurface[15] across which the solution is discontinuous.[16] The shock development problem in relativistic fluid mechanics was recently solved in spherical symmetry [**20**]. Away from symmetry, the problem remains open and is expected to be exceptionally difficult.

We mention here an important problem related to the shock development problem, which A. Majda solved [**62–64**] in the early 1980s. Specifically, for hyperbolic systems of conservation laws with suitable structure in more than one spatial dimension, Majda solved the *shock front problem*. That is, in a suitable Sobolev

[13]Roughly, a subset of $\{\mu = 0\}$ corresponds to the singular portion of the maximal development of the data.

[14]The boundary of the maximal development also contains another portion, along which the solution does not blow up; see Sect. 2.11.2 and Theorem 2.19 in particular.

[15]Part of the set $\{\mu = 0\}$ turns out to be a hypersurface portion along which, in the classical formulation of the irrotational Euler equations, certain solution derivatives blow up. However, $\{\mu = 0\}$ does not correspond to the physically correct hypersurface of discontinuity. The physically correct hypersurface of discontinuity propagates at supersonic speed starting from the first spacetime point where μ vanishes and develops "before" the set $\{\mu = 0\}$ has a chance to form. The physically correct hypersurface can be derived only by imposing the weak formulation of the full compressible Euler equations (without the assumption of irrotationality) starting from the time of first blow-up and by assuming suitable jump and entropy conditions.

[16]One expects the solution to be smooth on either side of the shock hypersurface.

framework, he proved a local existence result starting from an initial discontinuity given across a smooth hypersurface[17] subset of the Cauchy hypersurface. We stress that the initial hypersurface of discontinuity is prescribed. In contrast, in the shock development problem mentioned in the previous paragraph, the spacetime hypersurface of discontinuity is fully dynamic, emerging from singularity-free initial data. In the shock front problem, the data must verify suitable jump conditions, entropy conditions, and higher-order compatibility conditions. As in the shock development problem, the shock front problem features a free boundary: the shock hypersurface,[18] which is one of the unknowns. Majda's work also required an additional assumption[19] on the data that seems to be necessary for the stability of the corresponding linearized problem.

We now describe the origin and motivation behind the present monograph. In his work [**17**], Christodoulou exploited various special structures enjoyed by the wave equations of irrotational relativistic fluid mechanics, structures which Alinhac did not use in his proof of shock formation. In particular, the wave equations in [**17**] derive from a Lagrangian[20] and are invariant under the Poincaré group; just below equation (2.11), we describe some ways in which Christodoulou used these structures in his proof. In studying Christodoulou's work [**17**], the author discovered that it is possible to use his framework to close the proof of shock formation for a larger class of equations and without relying on these special structures. In particular, his framework can be extended to treat all of the wave equations studied by Alinhac. A somewhat surprising fact, which plays a fundamental role in our analysis, is that they all have a special null structure, *even though they fail to satisfy Klainerman's null condition*. This special structure is not visible relative to the standard formulation of the wave equation, *but becomes visible upon reformulating it as a system of geometric wave equations*; see Lemmas A.9 and A.16 for the main results in this direction. It is this realization that led to the present work. Our work here also generalizes and unifies earlier work on singularity formation initiated by F. John in the 1970s and continued by L. Hörmander and many others.

More precisely, in the present monograph, we extend Christodoulou's framework and use it to prove that shock singularities often develop in initially small, regular solutions to two important classes[21] of quasilinear wave equations in three spatial dimensions. Specifically, we study **i)** covariant scalar wave equations of the form $\square_{g(\Psi)}\Psi = 0$ and **ii)** Alinhac's noncovariant scalar wave equations, that is, wave equations of the form $(h^{-1})^{\alpha\beta}(\partial\Phi)\partial_\alpha\partial_\beta\Phi = 0$. Our main result shows that whenever the nonlinear terms fail Klainerman's (classic) null condition,[22] shocks develop in solutions arising from an open set of small data. Hence, within the classes **i)** and **ii)**, our work can be viewed as a sharp converse to a fundamental result, due separately to Christodoulou [**15**] and Klainerman [**47**], which showed that

[17] This hypersurface is co-dimension two when viewed as a subset of spacetime.

[18] The shock hypersurface is a co-dimension one subset of spacetime.

[19] The assumption is automatically verified for the nonrelativistic Euler equations under the adiabatic equations of state $p = A\rho^\gamma$, where $A > 0$ and $\gamma > 1$ are constants.

[20] That is, the equations in [**17**] are Euler-Lagrange equations.

[21] It turns out that the two classes of equations are more closely related than one might expect; see the discussion below equation (2.9) and in Appendix A.

[22] Readers should take care not to confuse Klainerman's null condition with the future strong null condition and past strong null condition introduced in Appendix A and mentioned in Remark 2.3.

when the null condition is verified in three spatial dimensions, small-data global existence holds. Roughly, we give the same sharp description of the solution that Christodoulou gave in [**17**]. However, to avoid lengthening the monograph, we did not give a full description of the boundary of the maximal development nor the behavior of the solution along it. For readers interested in those details, we remark that the estimates proved in our main Theorem 22.1 are sufficient for invoking the arguments of [**17**, Chapter 15] in which Christodoulou reveals properties of the maximal development. That is, with modest additional effort, our results could be extended to give the same sharp description of the maximal development that Christodoulou gave in [**17**, Chapter 15].

In proving our main results, we have taken substantial steps that go beyond replicating the proofs given by Christodoulou in [**17**]. This is partly out of necessity, as the general class of equations that we treat leads to new kinds of error terms that are not present in [**17**]. However, we have also developed alternate strategies that greatly simplify certain aspects of the proof. One big simplification is that we are more selective in our use of geometry. That is, we use sharp, fully geometric decompositions only for treating the most delicate terms. Another simplification is that our bootstrap argument is very straightforward in view of the fact that we have organized the monograph in a linear fashion (see two paragraphs below). We have also developed alternate approaches to deriving some of the difficult top-order estimates by reducing them to other top-order estimates. This spares one a great deal of effort; see, for example, the discussion at the beginning of Sect. 15.1 in which we describe a simplified approach for obtaining estimates for the top-order derivatives of μ.

We now give an overview of the content and organization of the monograph. Chapter 1 sets the stage for the rest of monograph but is independent of the remaining chapters. It contains historical background, a discussion of shock formation in solutions to Burgers' equation, a discussion of singularity formation in 2×2 strictly hyperbolic genuinely nonlinear systems, an overview of wave dispersion in higher dimensions and its connection to global and almost global existence results, an overview of the vectorfield method (including the multiplier and commutator methods) for deriving generalized energy estimates, and a discussion of the null condition. In Chapter 2, we describe the main results of the monograph, place them in context, and provide an extended overview of the most important aspects of the proofs. In the remaining chapters, we develop the machinery and estimates needed to prove the two theorems of the monograph, which are located in Chapters 22 and 23. Roughly, in the first theorem (the main one, which is difficult to prove), we show that the solution must persist unless μ vanishes, and we derive sharp a priori estimates that hold as long as μ remains positive. In the second theorem, which is a relatively easy consequence of the first one, we exhibit an open set of data such that μ does in fact vanish in finite time, thus yielding a shock singularity. Our analysis in Chapters 3–23 applies to covariant wave equations of the form $\square_{g(\Psi)}\Psi = 0$, while in Appendix A, we outline how to extend the results to the class of noncovariant wave equations studied by Alinhac. In Appendix B, we summarize the notation and conventions used in Chapters 2–23.

Chapters 3–23 are interdependent and are designed to be read consecutively. That is, this part of the monograph constitutes one long bootstrap-type proof, presented in chronological order. In Sect. 2.13, we give an overview of the contents of each chapter and provide suggestions on how to read the monograph, both for expert and novice readers.

This monograph is mostly self-contained but relies extensively on basic concepts from Lorentzian and Riemannian geometry such as Levi-Civita connections, fundamental forms, curvature, pullbacks, etc. We anticipate that there are readers with knowledge in fluid mechanics, conservation laws, and/or PDEs but who are unfamiliar with those geometric concepts. Such readers can find introductory geometric material, suitable for reading almost all[23] of the present monograph, in select portions of the books [**66**, **67**, **78**]. Novice geometers should bear in mind that at the end of the day, one aims to derive estimates, and that the geometry merely provides a framework for organizing calculations and revealing analytic structural features that would otherwise be difficult to detect.

We close by highlighting the following wide-open question:

> In more than one spatial dimension, to what extent can the results of this monograph be generalized to quasilinear systems featuring multiple speeds of propagation, such as the equations of elasticity, the equations of magnetohydrodynamics, the equations of crystal optics, the Euler–Einstein equations of cosmology, or even coupled systems of wave equations featuring two or more metrics with strictly separated[24] speeds of propagation?

Jared Speck

[23]Our short proof of geometric Sobolev embedding, presented in Chapter 18, relies on a handful of more advanced results from geometry.

[24]This roughly corresponds to the presence of two or more distinct families of characteristic hypersurfaces.