COURANT 14

PETER D. LAX

LECTURE NOTES

Hyperbolic Partial Differential Equations

American Mathematical Society
Courant Institute of Mathematical Sciences

Courant Lecture Notes in Mathematics

Hyperbolic Partial Differential Equations

Peter D. Lax
Courant Institute of Mathematical Sciences
With an Appendix by Cathleen S. Morawetz

14 Hyperbolic Partial Differential Equations

Courant Institute of Mathematical Sciences
New York University
New York, New York

American Mathematical Society
Providence, Rhode Island

2000 *Mathematics Subject Classification.* Primary 35L05, 35L10, 35L15, 35L20, 35L25, 35L30, 35L35, 35L40, 35L45, 35L50, 35L55, 35L60, 35L65, 35L67, 35P25.

For additional information and updates on this book, visit
www.ams.org/bookpages/cln-14

Library of Congress Cataloging-in-Publication Data

Lax, Peter D.
Hyperbolic partial differential equations / Peter D. Lax.
p. cm. — (Courant lecture notes, ISSN 1529-9031 ; 14)
Includes bibliographical references.
ISBN-13: 978-0-8218-3576-0 (alk. paper)
1. Differential equations, Hyperbolic. I. Title.

QA377.L387 2006
515′.3535--dc22 2006050151

∞ The paper used in this book is acid-free and falls within the guidelines established to ensure permanence and durability.
Visit the AMS home page at `http://www.ams.org/`

10 9 8 7 6 5 4 3 2 1 11 10 09 08 07 06

Contents

Foreword

The theory of hyperbolic equations is a large subject, and its applications are many: fluid dynamics and aerodynamics, the theory of elasticity, optics, electromagnetic waves, direct and inverse scattering, and the general theory of relativity.

The first seven chapters of this book, based on notes of lectures delivered at Stanford in the spring and summer of 1963, deal with basic theory: the relation of hyperbolicity to the finite propagation of signals, the concept and role of characteristic surfaces and rays, energy, and energy inequalities.

The structure of solutions of equations with constant coefficients is explored with the help of the Fourier and Radon transforms. The existence of solutions of equations with variable coefficients with prescribed initial values is proved using energy inequalities. The propagation of singularities is studied with the help of progressing waves.

Chapter 8 of the second part describes finite difference approximations of hyperbolic equations. This subject is obviously of great importance for applications, but also intriguing for the theorist. The proof of stability of difference schemes is analogous to the derivation of energy estimates, but much more sophisticated.

Chapter 9 presents a streamlined version of the Lax-Phillips scattering theory. The last section describes the Pavlov-Faddeev analysis of automorphic waves, and their mysterious connection to the Riemann hypothesis.

Chapter 10, the only one dealing with nonlinear waves, is about hyperbolic systems of conservation laws, an active research area today. We present the basic concepts and results.

Five brief appendices sketch topics that are important or amusing, such as Huygens' principle, a theory of mixed initial and boundary value problems, and the use of nonstandard energy identities.

I hope that this book will serve well as an introduction to the multifaceted subject of hyperbolic equations.

Peter Lax
New York
February 2006

CHAPTER 1

Basic Notions

The wave equation is the prototype of a hyperbolic equation

$$u_{tt} - ku_{xx} = 0\,, \quad k \text{ positive.} \tag{1.1}$$

To put the positivity of k into evidence, we set $k = c^2$; then (1.1) becomes

$$u_{tt} - c^2 u_{xx} = 0\,. \tag{1.1$'$}$$

This equation governs the transverse motion of a flexible elastic string, the constant k being the ratio of the tension T and the linear density ρ. Observe that c has the dimension of velocity.

We expect, in analogy with the motion of finite systems of particles, that the motion of the string is determined once we specify its initial position and velocity:

$$u(x, 0) = a(x), \quad u_t(x, 0) = b(x)\,. \tag{1.2}$$

This is indeed so; to find the solution, we write the wave equation in operator form,

$$Lu = 0\,,$$

and then factor the operator L. We get

$$L = D_t^2 - c^2 D_x^2 = (D_t + cD_x)(D_t - cD_x)\,.$$

Each linear factor on the right is directional differentiation, along the lines $x = x_0 \pm ct$, respectively. Integrating along these lines successively and making use of the initial conditions (1.2), we get, after a brief calculation, the following explicit expression for u:

$$u(x, t) = \frac{a(x + ct)}{2} + \frac{a(x - ct)}{2} + \frac{1}{2c} \int_{x-ct}^{x+ct} b(s)ds\,. \tag{1.3}$$

This derivation shows that the lines $x = x_0 \pm ct$ (called *rays*) play a special role for the wave equation.

Suppose we prescribe u and u_t not at $t = 0$ but along some curve $t = p(x)$. Then the above procedure can still be used to determine the solution u *as long as every ray of both families cuts the curve in exactly one point*. We call such curves *spacelike*; the analytic condition[1] sufficient for a curve $t = p(x)$ to be spacelike is that $|p'|$ be less than $1/c$ at every point.

[1] The condition $|p'| > 1/c$ also guarantees this; but since it turns out to be a special feature of the one-dimensional situation, we leave it aside.

We shall now use formula (1.3) to study the manner in which the solution depends on the initial data. The following features—some qualitative, some quantitative—are of importance:

(1) The motion is uniquely determined by the initial data.
(2) The initial data can be prescribed as arbitrary, infinitely differentiable functions with compact support, and the corresponding u is infinitely differentiable in x and t.
(3) The principle of superposition holds.
(4) Influence propagates with speed $\leq c$.

We shall show that properties (1)–(4) imply the following further properties:

(5) Motion is governed by a partial differential equation.
(6) Sharp signals propagate along rays.
(7) Energy is conserved.
(8) Spacelike manifolds have the same properties as the manifolds $t = \text{const}$.

We shall further show that motions depend continuously on their initial data; this implies that the governing equation is of a special type called *hyperbolic*.

The first two properties follow immediately from formula (1.3). The third property follows from the linearity of the wave equation. Property (4) means that, as evidenced by formula (1.3), the value of u at x, t is not influenced by the initial values outside the interval $(x - ct, x + ct)$. The formula also shows that, as asserted under (6), the influence of the endpoints is stronger than that of the interior; this will be made more precise later. To give meaning to (7) we have to define energy. In analogy with mechanics, we define

$$E_{\text{kinetic}} = \frac{1}{2}\rho \int u_t^2\, dx\,, \qquad E_{\text{potential}} = \frac{1}{2}T \int u_x^2\, dx\,.$$

The total energy is then (using $T/\rho = c^2$)

$$E_{\text{total}} = \frac{1}{2}\rho \int u_t^2 + c^2 u_x^2\, dx.$$

From the explicit formula for u we can verify that the *energy density* $u_t^2 + c^2 u_x^2$ *is the sum of a function of* $x + ct$ *and of* $x - ct$; the integral of such a function is indeed independent of time.

We shall give now a derivation of the law of conservation of energy for arbitrary spacelike surfaces; this second method is applicable in rather general situations.

Let P_1 and P_2 be two spacelike curves; multiply equation (1.1′) by u_t and integrate over the domain contained between P_1 and P_2. We get

$$0 = \iint_{P_1}^{P_2} u_t u_{tt} - c^2 u_t u_{xx}\, dx\, dt\,.$$

Integrating by parts with respect to x in the second term, we get

$$0 = \iint_{P_1}^{P_2} u_t u_{tt} + c^2 u_{tx} u_x\, dx\, dt - c^2 \int u_t u_x \xi\, ds \bigg|_{P_1}^{P_2},$$

where ξ, τ denote the x, t components of the unit normal drawn in the positive t direction. The remaining double integrals are both perfect t derivatives and so they can be integrated with respect to t. The result is

$$E(P_1) = E(P_2),$$

where we define the energy $E(P)$ contained in u on the curve P as

$$E(S) = \frac{1}{2}\int_P \left(\tau u_t^2 - 2c^2\xi u_t u_x + c^2\tau u_x^2\right)ds.$$

We recall now that P is spacelike if

$$\tau > c|\xi|.$$

From the above form of E we deduce that *the energy density* (*and thus the total energy*) *along P is positive definite if and only if P is spacelike.*

The law of conservation of energy gives another proof of the result that initial data along a spacelike curve uniquely determine the motion. We shall see later that energy conservation is the basic tool for constructing solutions of general hyperbolic equations.

Property (8) follows easily from an explicit representation for u in terms of the values of u and u_t along a spacelike curve.

In the next chapter we shall investigate a class of media whose motions have properties (1)–(4). We shall show that such motions are governed by partial differential equations satisfying a certain algebraic condition. We shall show that, conversely, solutions of differential equations satisfying that algebraic property have properties (1)–(8). It is perhaps surprising that the qualitative assumptions (1)–(4) have such quantitative consequences as (5), (6), and (7).

CHAPTER 2

Finite Speed of Propagation of Signals

We shall be dealing with motions of continuous media. The state of a medium at any point $x = (x_1, \dots, x_k)$ of Euclidean space and any time t is specified by the values of n variables (which in concrete cases are quantities like density, pressure, velocity, strength of electric and magnetic fields, etc.). We shall denote the state variables by $u_1, \dots, u_n$, and their totality as a single vector u.

The state of the medium at any given time is a vector function $u(x)$. We stipulate that

(1) all C_0^∞ vector functions $f(x)$ describe a state of the medium;
(2) the state of the medium at any given time s determines its state $u(x, t)$ at all future and past times.

These functions $u(x, t)$ describe all possible motions of the medium. Knowledge of these motions makes it possible to describe those points q in space-time that are influenced by the state of the medium at a point $p = (x, s)$.

DEFINITION 2.1 The point $p = (x, s)$ is said *to influence the point* q if, given any spatial neighborhood O of x and any space-time neighborhood D of q, there exist two motions u_1 and u_2 which at the time s are equal outside of O but which are unequal at some point in D.

If the motions are linear, i.e., if the superposition of two motions is also a motion—which we assume hereafter—then the last part of the definition can be rephrased as follows: *There exists a motion u that at time s is zero outside O but which is nonzero somewhere in D.*

Having defined influence we can further define: The *domain of influence* of p is the set of all points q influenced by it. The *domain of dependence* of a point q is the set of all points p influencing it.

EXERCISE Show that domains of influence and dependence are closed sets.

A surface $t = f(x)$ is called *spacelike* if no point on it influences another. It is called *strictly spacelike* if there exists a positive quantity δ such that all segments connecting a point p of the surface to any point q sufficiently close to p and influenced by it makes an angle greater than δ with the tangent plane at p.

We assume now that influence propagates with speed $\leq c$ in the following sense: Whenever

$$|x_p - x_q| > c|t_p - t_q|,$$

p does not influence q; here $|x|$ denotes the Euclidean length of x.

This assumptions has the following consequences:

(1) If the medium at time s has compact support, it has compact support at all other times.
(2) Every surface $t = p(x)$ where $|\text{grad}\, p| < 1/c$ is strictly spacelike.

THEOREM 2.2 *Denote by A the intersection of the domain of dependence of q with $t = r$. Suppose that the data of a motion u at time r are zero in an open set G containing A; then u is zero in some neighborhood of q.*

PROOF: Since no point p at time r and outside of G belongs to the domain of dependence of q, to each such point p there exists a spatial neighborhood G_p of p and a neighborhood D_p of q such that any motion which at time s is zero outside G_p is zero inside D_p. Since influence propagates with finite speed, there is a spatial neighborhood G_∞ of ∞ with the same property. By compactness there exists a finite, smooth partition of unity subordinate to the above covering, i.e., a finite number of functions $\varphi_j(x)$ such that

(1) $\sum \varphi_j(x) \equiv 1$,
(2) each φ_j is smooth, and
(3) the support of each φ_j is contained in one of the sets G_p or G_∞.

Denote the data of u at time s by f:

$$u(x, s) = f(x).$$

Multiplying (1) by f we get

$$\sum \varphi_j f = \sum f_j = f.$$

Notice that if the support of p_j lies in G then $f_j = p_j f$ is zero since f was assumed to be zero in G. Denote by u_j the motion with initial state f_j; by the principle of superposition

$$\sum u_j = u.$$

Each motion u_j vanishes in some neighborhood of q, and therefore their sum vanishes in the intersection of these neighborhoods. □

The theorem justifies calling the set A a *domain of determinancy* of q at time s. Later we shall show that the intersection of the domain of dependence of q with any spacelike surface is likewise a domain of determinacy of q on the surface.

The following result is called the *Huygens wave construction*:

THEOREM 2.3 *Denote by (x, t) any point in space-time. Denote by s any time $> t$, and by $K(x, t; s)$ the set of points at time s that are influenced by (x, t). Let r be any time between t and s; we claim that $K(x, t; s)$ is contained in the set of all points at time s that are influenced by points in $K(x, t; r)$.*

EXERCISE Prove Theorem 2.3. We note that in some interesting cases the containment is proper; see Appendix A.

The Differential Equations of Motion. We have assumed that $u(x, t)$ is uniquely determined everywhere in terms of its value at time s, s any value. In particular, the value of its t-derivative u_t at time s is determined. We denote the operator relating $u_t(s)$ to $u(s)$ by $G = G(s)$:

$$u_t = Gu . \tag{2.1}$$

G is an operator mapping the space of all C^∞ differentiable functions of x with compact support into itself. We claim that G is a *local operator*, in the following sense: *The support of Gf is contained in the support of f.* Another way of saying this is: If f vanishes for all x in some open set U, so does Gf.

To show that G is local, we note that if $u(x, s)$ vanishes in the open set U, then since influence propagates with speed less than c, $u(x, s+h)$ vanishes at all points x whose distance from the complement of U is greater than $c|h|$. This shows that $u_t = Gu$ vanishes in U, i.e., that G is local.

According to a theorem of Peetre every linear operator mapping C^∞ into C^∞ that is local is a partial differential operator with coefficients that are differentiable functions of x. Since G depends in general on s, so will its coefficients; it is easy to show that the dependence of the coefficients on s is also differentiable.

We have thus shown that motions which have properties (1)–(4) satisfy a partial differential equation. We turn now to the problem of characterizing algebraically the partial differential equations satisfied by such motions. We shall treat first the special case when the motions are translation invariant in the following sense: *If $u(x, t)$ is a motion, then $u(x-y, t-s)$ is also a motion for any fixed y, s.*

The differential equations governing homogeneous motion have, it is easy to show, constant coefficients. We proceed now to solve these equations by taking the Fourier transform (FT) in x. Denote as usual the FT by the symbol $\tilde{}$:

$$\tilde{f}(\xi) = \int f(x) e^{-ix\xi}\, dx ,$$

ξ being the dual vector variable of x. Denote by D the vector operator of differentiation with respect to $x_1, \dots, x_n$. The operator G can be written as a polynomial in D,

$$G(D) = \sum A_j D^j ,$$

where A_j is a matrix, j a multi-index, and D^j the symbolic power $D_1^{j_1}, \dots, D_k^{j_k}$. According to the well-known rule, under FT differentiation goes into multiplication by $i\xi$, so

$$\widetilde{Gu} = G(i\xi)\tilde{u} . \tag{2.2}$$

Since motions are C^∞ and have, for fixed t, compact support in x, they have spatial Fourier transforms. Taking the Fourier transform of both sides in (2.1) we get, using (2.2),

$$\tilde{u}_t = G(i\xi)\tilde{u} .$$

This ordinary differential equation has the unique solution with initial value $\tilde{f}(\xi)$

$$\tilde{u}(\xi, t) = e^{tG(i\xi)}\tilde{f}(\xi), \tag{2.3}$$

from which u itself can be determined by inversion.

In the process of deriving this explicit formula for the solution we have proved this uniqueness theorem:

THEOREM 2.4 *If a differentiable solution of* (2.1) *has compact support for each t, and if it vanishes at $t = 0$, then it vanishes for all time.*

It follows that not only does every motion satisfy differential equation (2.1), but every solution of (2.1) that has compact support for each t is a motion.

THEOREM 2.5 *Motions depend continuously on their initial data.*

PROOF: Consider the set of all C^∞ initial data f with support in some compact set K. They form a Frechet space F, which is a complete metric linear space. The corresponding motions u, restricted to the strip $-1 \le t \le 1$, belong to the Frechet space U of C^∞ functions in this strip whose support lies in the domain of influence of K. The mapping relating initial values with support in K to motions restricted to the strip has the following properties:

(1) it is linear,
(2) it maps the whole space F into the space U,
(3) its graph is closed.

Property (3) means that if f_n is a sequence in F converging to f, and if the corresponding motions u_n tend to an element u in U, then this limit u is the motion corresponding to f. To verify that this is indeed so, we observe that the limit u is a solution of the differential equation governing the motion, that it has compact support, and that its initial value is f. According to the uniqueness theorem, u is the motion with initial data f.

We recall that continuity means that if f_n and all its derivatives tend to f, then u_n and all its derivatives tend to u.

We appeal now to the *closed graph* theorem: A transformation which has properties (1)–(3) is continuous. This completes the proof of Theorem 2.5. □

COROLLARY 2.6 *There exists a constant M and an integer m such that for all motions u whose initial data f are supported in K, and for all x,*

$$|u(x,t)| \le M|f|_m, \quad |t| \le 1. \tag{2.4}$$

Here $|f|_m$ denotes the maximum of the initial value f and its partial derivatives up to order m.

PROOF: Suppose not; then there exists a sequence of initial data f_m supported in K such that $|f_m|_m$ tends to zero but the maximum value of u_m in the strip $|t| \le 1$ doesn't. This clearly contradicts the continuity of the dependence of u on f. □

We take now K to be the hypercube $|x_j| \le 1$, $j = 1, \dots, k$, and define the initial value f with the aid of the following auxiliary function p:

$$p(x) = \begin{cases} (1-x^2)^m & \text{for } |x| \le 1, \\ 0 & \text{for } |x| > 1. \end{cases}$$

We set

$$f(x) = \prod p(x_j)h,$$

where h is a vector to be chosen later. We take the Fourier transform of f, and integrate by parts m^k times. We get

$$\tilde{f}(\xi) = 2^m m! \prod \frac{\sin \xi_j}{\xi_j^m} + O\left(\frac{1}{\prod |\xi_j|^m}\right) \tag{2.5}$$

as $|\xi| \to \infty$.

We turn now to $u(x, \pm 1)$. Since its initial value is supported in $|x_j| \leq 1$, and since signals propagate with speed $\leq c$, $u(x, \pm 1)$ is supported in $|x_j| \leq 1 + c$. Therefore we can estimate $\tilde{u}(\xi, \pm 1)$ in terms of $u(x, \pm 1)$ as follows:

$$\tilde{u}(\xi, \pm 1) \leq (2 + 2c)^k |u(\cdot, \pm 1)|_{\max}.$$

$|u(\cdot, \pm 1)|_{\max}$ is bounded by inequality (2.4); therefore

$$|\tilde{u}(\xi, \pm 1)| \leq (2 + 2c)^k |f|_m.$$

The norm $|f|_m$ depends only on m and k; therefore we can rewrite the above estimate as

$$|\tilde{u}(\xi, \pm 1)| \leq \text{const}, \tag{2.6}$$

where the constant depends only on M, m, k and c.

We use now formula (2.3) to express $\tilde{u}(\xi, \pm 1)$ in terms of $\tilde{f}$. We denote the eigenvalues of the matrix $G(w)$ by $\sigma(w)$, the corresponding eigenvector by h. Then the eigenvalue of $G(i\xi)$ is $\sigma(i\xi)$, and (2.3) becomes

$$\tilde{u}(\xi, \pm 1) = e^{\pm \sigma(i\xi)} h \tilde{f}(\xi). \tag{2.7}$$

Using (2.5) to express $\tilde{f}$ and the estimate (2.6) for $\tilde{u}$ we deduce from (2.7) that for all real ξ

$$\prod |\sin \xi_j| |e^{\sigma(i\xi)}| \leq \text{const} \prod |\xi_j|^m. \tag{2.8}$$

Similarly we define $g(x)$ as

$$g(x) = \prod x_j p(x_j) h$$

and deduce as before that as $|\xi| \to \infty$,

$$\tilde{g}(\xi) = 2^m m! \prod \frac{\cos \xi_j}{\xi_j^m} + O\left(\frac{1}{\prod |\xi_j|^m}\right).$$

We deduce as before that

$$\prod |\cos \xi_j| |e^{\sigma(i\xi)}| \leq \text{const} \prod |\xi_j|^m. \tag{2.8'}$$

Combining (2.8) and (2.8′) we deduce that

$$|e^{\sigma(i\xi)}| \leq \text{const} \prod |\xi_j|^m,$$

which implies that

$$|\text{Re}\, \sigma(i\xi)| \leq \text{const} \log(2 + |\xi|) \tag{2.9}$$

for all real ξ.

The eigenvalues σ are roots of the *characteristic polynomial* P:

$$P(\xi, \tau) = \det[\tau I - G(\xi)] .$$

THEOREM 2.7 *The characteristic polynomial of the differential equation satisfied by translation invariant motions propagating with finite speed has the following properties*:

(i) *its degree in* τ *equals its degree in* τ, ξ, *and*
(ii) *its roots* $\tau = \sigma(\xi)$ *satisfy the inequality*

$$|\mathrm{Re}\, \sigma(i\xi)| \leq \mathrm{const} \log(2 + |\xi|) \quad \textit{for } \xi \textit{ real.}$$

PROOF: Property (ii) follows by taking ξ real in (2.9).

The proof of (i) is based on the fact that the Fourier transform of a function u of compact support,

$$\tilde{u}(\xi) = \int u(x) e^{-ix\xi}\, dx,$$

is defined for all complex ξ, and is of exponential growth:

$$|\tilde{u}(\xi)| \leq \mathrm{const}\, e^{a|\mathrm{Im}\,\xi|}.$$

Since signals propagate with finite speed, the solution $u(x, t)$ has compact support for all t. According to formula (2.3), its Fourier transform $\tilde{u}$ is

$$\tilde{u}(\xi, t) = e^{tG(i\xi)} \tilde{f}(\xi).$$

Therefore the eigenvalues of $G(i\xi)$ grow at most linearly in ξ:

$$|\sigma(\xi)| \leq \mathrm{const}(1 + |\xi|) . \tag{2.10}$$

Write now

$$P(\xi, \tau) = \sum_0^n a_{n-\nu} \tau^\nu, \quad a_0 = 1 ,$$

a_ν polynomial in ξ. Since $a_{n-\nu}$ is the sum of products of the roots ν at a time, it follows from (2.10) that

$$|a_{n-\nu}(\xi)| \leq \mathrm{const}(1 + |\xi|^\nu) .$$

This shows that $a_{n-\nu}(\xi)$ is a polynomial of degree at most ν in ξ. Since this is true for arbitrary ξ_0, it follows that $a_{n-\nu}$ is a polynomial of degree at most ν in ξ. □

Condition (ii) imposes a restriction on the roots $\tau(\xi)$ for $|\xi|$ large; for such values the highest-order terms $P_0(\xi, \tau)$ dominate. P_0 is a form of degree n, called the *characteristic form* associated with the differential equation. The null set of $P_0(\xi, \tau) = 0$ is called the *characteristic variety*.

It is easy to show, that the following relation holds between properties of the roots of P and P_0:

THEOREM 2.8

(i) *If the roots of the characteristic equation satisfy inequality* (ii) *of Theorem* 2.7, *then* $P_0(\xi, \tau) = 0$ *has, for* ξ *real, only real roots* τ.

(ii) *If the characteristic form P_0 has for real nonzero ξ only real roots τ which furthermore are distinct, then the roots $\sigma(\xi)$ of the polynomial satisfy*

$$|\operatorname{Re}\sigma(i\xi)| \leq \text{const} \quad \textit{for all real } \xi .$$

Condition (ii) in Theorem 2.8 is slightly stronger than condition (ii) in Theorem 2.7. Gårding has shown by purely algebraic methods that the two conditions are equivalent.

DEFINITION 2.9 A partial differential equation $u_t - Gu$ is called (strictly) hyperbolic if its characteristic polynomial has properties (i) and (ii) in Theorem 2.7.

We give now an invariant formulation of the property of forms discussed above; this is important for it frees us from a fixed space-time frame:

DEFINITION 2.10 Let P_0 be a form of degree n in the variables ζ. A real vector ν is called *(strictly) hyperbolic* for P_0 if for any real vector ζ not parallel to ν the polynomial in s, $P_0(s\nu + \zeta)$, has n real (distinct) roots.

A form is called (strictly) *hyperbolic* if there exists a (strictly) hyperbolic vector for it. In this terminology Theorem 2.7 asserts that the characteristic form of the differential equations of motions investigated in these notes is hyperbolic, and that the normal to the hyperplane $t = 0$ is hyperbolic. In the next chapter we shall see that the hyperbolic directions are precisely those which are normal to a spacelike hypersurface.

Examples.

EXAMPLE 1

$$G = \sum_1^k A^j D_j + B$$

with the A^j symmetric, real $n \times n$ matrices, and B an arbitrary real matrix. The equation $u_t = Gu$, G of the above form, is called a *symmetric hyperbolic system* of first order.

We shall verify now that condition (ii) holds for the eigenvalues τ of the matrix

$$G(\xi) = \sum A^j \xi_j + B .$$

When B is zero we obtain, using the result that real symmetric matrices have real eigenvalues, that for ξ real the eigenvalues τ are real. For nonzero B we use the following result about symmetric matrices: *If A is a symmetric matrix and B an arbitrary matrix, then the eigenvalues of $A + B$ have imaginary part not greater than $\|B\|$.*

PROOF: Let w be an eigenvector of $A + B$ of length 1 with eigenvalue τ:

$$(A + B)w = \tau w .$$

Take scalar product with w;

$$(Aw, w) + (Bw, w) = \tau .$$

Take the imaginary part of the equation above. Since A is symmetric,

$$\operatorname{Im}(Bw, w) = \operatorname{Im}\tau .$$

The assertion follows now by the Schwarz inequality. □

When this estimate is applied to $G(\xi)$, we deduce that the eigenvalues of $G(\xi)$ satisfy condition (ii). Observe that symmetric hyperbolic systems need not be strictly hyperbolic; more about this in Appendix C. We note that Maxwell's equations of electromagnetism form a symmetric hyperbolic system.

EXAMPLE 2 $n = 2$.

$$u_t = v , \quad v_t = c^2 \sum u_{x^k x^k} = c^2 \Delta u .$$

Eliminating v, we get

$$u_{tt} - c^2 \Delta u = 0 ,$$

the familiar wave equation, the prototype of hyperbolic equations. The characteristic polynomial

$$\tau^2 - c^2\xi^2 , \quad \xi^2 = \xi_1^2 + \cdots + \xi_k^2 ,$$

is homogeneous, and its roots

$$\tau = \lambda(\xi) = \pm c(\xi^2)^{1/2}$$

are real and distinct for ξ real and $\neq 0$.

EXAMPLE 2′ $n = 2$.

$$u_t = v , \quad v_t = \sum a_{ij} u_{x^i x^j} - b_i u_{x^i} - cu .$$

Eliminating v, we get the second-order equation

$$u_{tt} - \sum a_{ij} u_{x^i x^j} + b_i u_{x^i} + cu = 0 .$$

The characteristic form of this equation is

$$\tau^2 - \sum a_{ij} \xi_i \xi_j ,$$

which has real and distinct roots for ξ real and $\neq 0$ if and only if the quadratic form $\sum a_{ij}$ is positive definite.

EXAMPLE 3 $n = 2$.

$$u_t = v , \quad v_t = u_x .$$

Eliminating v, we get $u_{tt} - u_x = 0$. The characteristic polynomial is $\tau^2 - \xi$, whose roots are $\tau = \xi^{1/2}$. Clearly condition (ii) is violated, so that this equation is *not* hyperbolic. The characteristic form is $P_0 = \tau^2$ whose roots, $\tau = 0$, are real but not distinct. So P_0 is hyperbolic, but not strictly hyperbolic.

EXAMPLE 4 $k = 2$.

$$P(\xi, \tau) = \sum_0^n P_j(\xi, \tau) ,$$

ξ, τ scalars, and P_j homogeneous of degree $n - j$. Suppose that $P_0(1, \tau) = 0$ has real roots τ_k of multiplicity n_k, $\sum n_k = n$. E. E. Levi has shown that if τ_k

is a root of degree $n_k - j$ of P_j, $j = 1, 2, \ldots$, then the initial values $\partial_t^m u(x, 0)$, $m = 0, 1, \ldots, n-1$ of a solution u of $Pu = 0$ can be prescribed as arbitrary C^∞ functions. Anneli Lax has shown that Levi's condition is necessary for the existence of solutions. Svensson has extended this result to any number of space variables.

EXAMPLE 5 n any integer, $k = n(n+1)/2$. Consider the following form P_0 of degree n in the k variables

$$\zeta_{ij}, \quad i, j = 1, \ldots, n, \quad \zeta_{ij} = \zeta_{ji} : P_0(\zeta) = \det |\zeta_{ij}|.$$

We claim that this form is hyperbolic; the hyperbolic vectors ν are precisely those for which the symmetric matrix $\nu = (\nu_{ij})$ is positive definite. For this equation $P_0(s\nu + \zeta) = 0$ means precisely that the matrix $s\nu + \zeta$ is not invertible, which implies that there exists a nonzero vector w such that $s\nu w = -\zeta w$; i.e., s is an eigenvalue of ζ with respect to ν. But it is well-known from matrix theory that the eigenvalues of any symmetric matrix ζ with respect to a positive definite matrix ν are all real. On the other hand, for ν nonpositive there exists a matrix ζ with complex eigenvalues with respect to ν.

EXAMPLE 6 n any integer, $k = n^2$. Consider the following form P_0 of degree n in the k variables

$$\zeta'_{ij}, \zeta''_{ij}, \quad i, j = 1, \ldots, n, \quad \zeta'_{ij} = \zeta'_{ji}, \quad \zeta''_{ij} = -\zeta''_{ji},$$
$$P_0(\zeta', \zeta'') = \det |\zeta'_{ij} + i\zeta''_{ij}|.$$

An analysis similar to the above shows that P_0 is hyperbolic; the hyperbolic vectors ν are precisely those for which the Hermitian matrix ν is positive definite.

Examples 5 and 6 are due to Lars Gårding.

EXAMPLE 7 Given a hyperbolic form P_0 of degree n in the variables ξ, τ, take G as the companion matrix of P_0 regarded as polynomial in τ; i.e.,

$$G(\xi) = \begin{vmatrix} 0 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 0 & 0 & 1 \\ -a_n & -a_{n-1} & \cdots & \cdots & \cdots & -a_1 \end{vmatrix}$$

where $a_j = a_j(\xi)$ is the coefficient of τ^{n-j} in P_0. Clearly the characteristic form of the operator $\partial/\partial t - G$ is P_0.

In all these examples it was possible to associate an order d_j with each of the variables u_j so that in the i^{th} equation

$$\frac{\partial}{\partial t} u_i = \sum g_{ij}(D) u_j,$$

the total order of each term on the right does not exceed $d_i + 1$, the order of the left side; i.e., the *order* of g_{ij} *does not exceed* $d_i - d_j + 1$. From now on we shall confine our studies to operators of this kind.

The *principal part* G_0 of an operator G as above is defined as the matrix formed by the terms of order $d_i - d_j + 1$. The characteristic form of $\frac{\partial}{\partial t} - G$ can be expressed in terms of the principal part of G as follows:

$$P_0(\xi, \tau) = \det(\tau I - G_0(\xi)).$$

The operator $\frac{\partial}{\partial t} - G$ is called hyperbolic if the order of g_{ij} satisfies the above condition, and if the roots of its characteristic polynomial satisfy condition (ii) of Theorem 2.7. It is called strictly hyperbolic if $G_0(\xi)$ has real and distinct eigenvalues for ξ real.

EXAMPLE 8 The equations governing the motion of elastic media are a hyperbolic system of second-order partial differential equations.

In 1957 the author showed that part of Theorem 2.8 holds also for differential equations governing motions with properties (1)–(4) that are not translation invariant. In this case the characteristic form P_0 depends on x and t as well; the roots τ of $P_0 = 0$ are required to be real for all real ξ and all x and t. The definitive form of this result is due to Mizohata.

References

Gårding, L. The solution of Cauchy's problem for two totally hyperbolic linear differential equations by means of Riesz integrals. *Ann. of Math. (2)* 48: 785–826, 1947.

______. Linear hyperbolic partial differential equations with constant coefficients. *Acta Math.* 85: 1–62, 1951.

Lax, A. On Cauchy's problem for partial differential equations with multiple characteristics. *Comm. Pure Appl. Math.* 9: 135–169, 1956.

Lax, P. D. Asymptotic solutions of oscillatory initial value problems. *Duke Math. J.* 24: 627–646, 1957.

Levi, E. E. Caratteristiche multiple e problema di Cauchy. *Ann. Mat. Pura Appl. (3)* 16: 161–201, 1909.

Peetre, J. Réctification à l'article "Une caractérisation abstraite des opérateurs différentiels". *Math. Scand.* 8: 116–120, 1960.

Svensson, S. L. Necessary and sufficient conditions for the hyperbolicity of polynomials with hyperbolic principal part. *Ark. Mat.* 8: 145–162, 1969.

CHAPTER 3

Hyperbolic Equations with Constant Coefficients

In this chapter we shall show that if $u_t - Gu = 0$ is a strictly hyperbolic equation with constant coefficients, then its solutions represent motions that have properties (1)–(8) listed in Chapter 1. The verification of property (6) will be carried out in Chapter 7.

3.1. The Domain of Influence

Properties (1), (2), and (4) assert that, given initial data which are infinitely differentiable and of compact support, there exists exactly one solution of the differential equation that has the prescribed data and that has compact support in x for all t. As shown in Chapter 2, this solution is easily constructed by Fourier transformation. Denote, as customary,

$$\tilde{f}(\xi) = \int e^{-ix\xi} f(x)\, dx \,. \tag{3.1}$$

Taking the Fourier transform of both sides of the differential equation (2.1) and using (2.2) gives

$$\frac{\partial}{\partial t}\tilde{u}(\xi, t) = G(i\xi)\tilde{u}(\xi, t)\,.$$

Integrate this ordinary differential equation:

$$\tilde{u}(\xi, t) = e^{tG(i\xi)}\tilde{u}(\xi, 0)\,. \tag{3.2}$$

Since the differential equation is assumed to be hyperbolic, $G(i\xi)$ satisfies condition (ii) of Theorem 2.7, and this assures us that the exponential in (3.2) grows at most like a power of ξ. Then the function on the right in (3.2) decays fast as $|\xi| \to \infty$, and so has a Fourier inverse. This inverse clearly satisfies the differential equation, has the prescribed initial values, and is unique in the class of solutions admitted. What remains to be shown is that it has compact support in x for all values of t. We shall deduce this on the basis of Plancherel and Polya generalization of the Paley-Wiener theorem, which characterizes the Fourier transforms of functions with compact support.

Let $f(x)$ be m times differentiable, and suppose that its support is the closed bounded set K. Denote by $s_K(\xi)$ the *support function* of K, defined by all real vectors ξ as

$$s_K(\xi) = \max_{x\in K} x\xi\,.$$

Note that $s_K(\xi)$ is positive homogeneous: $s_K(\alpha\xi) = \alpha s_K(\xi)$ for α positive, and is a convex function of ξ. The Fourier transform

$$\tilde{f}(\zeta) = \int_K e^{-ix\zeta} f(x)dx$$

of such a function f is also well-defined for complex values of ζ and is an entire analytic function of ζ. Integrating by parts m times and estimating the resulting integral by replacing the integrand by its maximum value leads to the following estimate:

$$|\tilde{f}(\zeta)| \le \frac{\text{const}}{1+|\zeta|^m} e^{s(\eta)}, \quad s = s_K, \eta = \operatorname{Im}\zeta. \tag{3.3}$$

The converse of this result is the following:

THEOREM 3.1 (Theorem of Plancherel and Polya) *Suppose $\tilde{f}(\zeta)$ is an entire analytic function that satisfies inequality* (3.3) *with some positive homogeneous function $s(\eta)$. Then $f(x)$, its Fourier inverse, is zero at all points outside the set of points $\{y\}$ that satisfy $y\eta \le s(\eta)$ for all real η.*

SKETCH OF PROOF: If x does not belong to this set, then for some real ω

$$x\omega > s(\omega). \tag{3.4}$$

In the Fourier inversion formula

$$f(x) = \int \tilde{f}(\xi) e^{ix\xi}\, d\xi,$$

change the path of integration to $\xi + i\rho\omega$; by Cauchy's theorem the value of the integral is unchanged. Relations (3.3) and (3.4) show that the value of the integral is less than any ϵ for ρ large positive. □

In order to apply this to the situation at hand, we have to estimate the rate of growth of $\exp G(i\zeta)$ for complex ζ. We make use of the following variant of the Phragmén-Lindelöf principle.

PRINCIPLE (Phragmén-Lindelöf Principle) Let $h(z)$ be an analytic function in the upper half-plane satisfying the following:

(i) $|h(z)| \le m$ for z real.
(ii) $|h(iy)| \le me^{sy}$ for y large positive.
(iii) $|h(z)| \le Me^{S|z|}$ for all z in the upper half-plane, then

$$|h(z)| \le me^{s\operatorname{Im} z} \tag{3.5}$$

in the whole upper half-plane.

In most applications the first two estimates are delicate, the third is crude. The salient point of the lemma is that the constants in the crude estimate do not appear in the conclusion (3.5). For proof, see the well-known text of Ahlfors on function theory.

We shall also make use of the following estimate from linear algebra: Let T be a square matrix with a complete set of eigenvectors. Then

$$|T| \le \text{const}\, r\,, \tag{3.6}$$

where r is the absolute value of the largest eigenvalue of M and the constant depends only on the determinant of the normalized eigenvectors of T and the order of the matrix. This follows from the representation of T as RDR^{-1} where R is the matrix whose columns are the eigenvectors of T.

For ω real, denote the largest eigenvalue of $G_0(\omega)$ by $\sigma_{\max}(\omega)$.

LEMMA 3.2 *There exist constants m and N such that*

(i) $|e^{G(i\xi)}| \le m(1+|\xi|^N)$ *for all real* ξ.

(ii) *For any real ξ and η and any given positive ϵ,*

$|e^{G(i\xi - y\eta)}| \le m(1+y^N)e^{[\sigma_{\max}(-\eta)+\epsilon|\eta|]y}$ *for y sufficiently large positive.*

(iii) *There exist constants M and S such that*

$$|e^{G(\zeta)}| \le Me^{S|\zeta|} \quad \textit{for all complex } \zeta.$$

SKETCH OF PROOF: Part (i) follows from inequality (2.9). To verify (ii), use the fact that for y large, positive the eigenvalues of $G(i\xi - y\eta)$ differ by $o(y)$ from the eigenvalues of $G_0(-y\eta)$. These eigenvalues are homogeneous functions of $y\eta$ of degree 1; so the largest eigenvalue of $G_0(-y\eta)$ is $y\sigma_{\max}(-\eta)$. The estimate in (ii) is a consequence of this observation.

To prove (iii), we recall from the end of Chapter 2 that the entry $g_{ij}(\zeta)$ of $G(\zeta)$ has degree $\le d_i - d_j + 1$. It is convenient to represent G by a homogeneous matrix H defined as

$$H(\zeta) = \zeta^{-1}D^{-1}(\zeta)G(\zeta)D(\zeta)\,,$$

where $D(\zeta)$ is a diagonal matrix with diagonal elements ζ^{d_i}. It follows that the elements of H have nonpositive degree, and that for $|\zeta|$ large, H has the form

$$H(\zeta) = D^{-1}G_0(\omega)D + 0(1/|\zeta|)\,, \tag{3.7}$$

where $\omega = \zeta/|\zeta|$ and G_0 is the principal part of G. Since by assumption G is strictly hyperbolic, G_0 has distinct eigenvalues; it follows that it has n linearly independent eigenvectors. Thus the determinant $d(\omega)$ of the normalized eigenvectors of $G_0(\omega)$ is nonzero for each ω; since the unit sphere is compact, it follows that $d(\omega)$ is bounded away from zero uniformly for all ω on the unit sphere. Then by continuity it follows from (3.7) that for $|\zeta|$ large enough $H(\zeta)$ has linearly independent eigenvectors whose determinant is bounded away from zero, so by (3.6) for $|\zeta|$ large enough

$$|e^{\zeta H(\zeta)}| \le \text{const}\, e^{\rho(\zeta)|\zeta|}$$

where $\rho(\zeta)$ is the largest real part of the eigenvalues of $H(\zeta)$. Using this estimate in

$$e^{G(\zeta)} = De^{\zeta H(\zeta)}D^{-1}$$

gives inequality (iii) of Lemma 3.2. The rest follow by using the strict hyperbolicity of G. □

Let ξ and η be arbitrary real vectors; define

$$h(z) = \frac{1}{(z+i)^N} e^{G(i\xi + iz\eta)} .$$

It follows from Lemma 3.2 that $h(z)$ satisfies the hypotheses of the Phragmén-Lindelöf principle with $s = \sigma_{\max}(-\eta) + \epsilon|\eta|$. So

$$|h(z)| \leq m e^{\sigma_{\max}(-\eta)+\epsilon|\eta|} .$$

Put $z = i$; we get,

$$|e^{G(i\zeta)}| \leq m(|\zeta|^N + 1) e^{\sigma_{\max}(-\eta)+\epsilon|\eta|} \tag{3.8}$$

for all complex ζ, where $\eta = \operatorname{Im}\zeta$.

Suppose that the initial values $u(x, 0)$ have their support inside the sphere $|x| \leq r$; then $\tilde{u}(\xi, 0)$ is bounded by

$$|\tilde{u}(\zeta, 0)| \leq \frac{\text{const}}{1 + |\zeta|^K} e^{r|\eta|} . \tag{3.9}$$

Combining (3.8) and (3.9) to estimate $\tilde{u}(\xi, 1)$ as given by formula (3.2), we conclude that

$$|\tilde{u}(\zeta, 1)| \leq \frac{\text{const}}{1 + |\zeta|^m} e^{\sigma_{\max}(-\eta)+(\varepsilon+r)} ,$$

where $\eta = \operatorname{Im}\zeta$. Using this estimate in the theorem of Plancherel and Polya, we conclude: $u(x, 1)$ *is zero if x lies outside the set of points satisfying for all* η

$$x\eta \leq \sigma_{\max}(-\eta) + r ,$$

provided that its initial values $u(x, 0)$ are zero outside a sphere of radius r around the origin.

Denote by K the intersection of the domain of influence of the origin and the hyperplane $t = 1$. It follows from the assertion above that *the support function of K, defined as $s_K(\eta) = \max_{x\in K} x\eta$, is not greater than $\sigma_{\max}(-\eta)$*. We claim:

THEOREM 3.3 *The support function of K is equal to $\sigma_{\max}(-\eta)$.*

PROOF: Denote by v the eigenvector of $G_0(-\omega)$ which corresponds to the eigenvalue $\sigma_{\max}(-\omega)$. Then, as our analysis before shows, given any ϵ, for ρ large enough positive

$$|e^{G(-\rho\omega)} v| \geq e^{\rho\sigma_{\max}(-\omega)-\epsilon\rho} . \tag{3.10}$$

Choose the scalar function $a(x)$ with support in a small sphere around $x = 0$ and so[1] that its Fourier transform does not decrease too fast in the direction $i\omega$:

$$|\tilde{a}(i\rho\omega)| > e^{-\epsilon\rho} , \quad \rho > 0 . \tag{3.11}$$

Put $u(x, 0) = a(x)v$; then expressing $\tilde{u}$ by (3.2) and using estimates (3.10) and (3.11), we get

$$|\tilde{u}(i\rho\omega, 1)| \geq e^{\rho\sigma_{\max}(-\omega)-2\epsilon\rho} . \tag{3.12}$$

[1]This is easily done; in fact, whenever $a(x)$ vanishes for $|x| \geq \epsilon$, $\limsup_{\rho\to\infty} \log|\hat{a}(i\rho\omega)|/\rho \geq -\epsilon$, and this suffices.

On the other hand, by inequality (3.3) every solution whose initial values have sufficiently small support satisfies for ρ large positive

$$|\tilde{u}(i\rho\omega, 1)| < \text{const}\, e^{\rho s_K(\omega)+\epsilon\rho}. \tag{3.13}$$

Comparing (3.12) and (3.13) we get

$$\sigma_{\max}(-\omega) < s_K(\omega) + 3\epsilon.$$

Since ϵ is arbitrary, this shows that

$$\sigma_{\max}(-\omega) \leq s_K(\omega),$$

which, together with the previous inequality, proves Theorem 3.3. □

Being the maximum of linear functions, the support function of a closed, bounded set is a convex function. Therefore we have

COROLLARY 3.4 *$\sigma_{\max}(\eta)$ is a convex function of η.*

Knowing the support function of K, we can determine the convex hull $\widehat{K}$ of K: it is the set of points y satisfying for all η

$$y\eta \leq \sigma_{\max}(-\eta). \tag{3.14}$$

When the equation in question is strictly hyperbolic, $\sigma_{\max}(-\eta)$ is a regular algebraic function. In this case the boundary of $\widehat{K}$ contains no straight line segments. According to the theory of convex sets, it follows that every boundary point of $\widehat{K}$ is an extreme point, and so belongs to K.

For not strictly hyperbolic equations, on the other hand, it can happen that not every boundary point of the set defined by (3.14) belongs to the domain of influence K of the origin; see, e.g., examples 5 and 6 in Chapter 2.

What *interior* points of the set defined by (3.14) belong to K is a delicate question; in Section 3.5, as well as in Appendix A, we shall give examples where not all interior points of $\widehat{K}$ belong to K.

3.2. Spacelike Hypersurfaces

Let P be a hyperplane in x, t space through the origin whose equation is

$$z\nu = 0, \tag{3.15}$$

where z and ν stand for (x, t) and (ξ, τ), respectively.

THEOREM 3.5 *P is strictly spacelike if for all real ξ*

$$0 < \frac{\tau - \sigma_{\max}(\xi)}{\tau + \sigma_{\max}(-\xi)} < \infty. \tag{3.16}$$

PROOF: Let z be any point in the domain of influence of the origin; we take for simplicity the t-coordinate of z to be one: $z = (x, 1)$. Then x belongs to K and so by the result derived at the end of the last section

$$-\sigma_{\max}(\xi) \leq x\xi \leq \sigma_{\max}(-\xi). \tag{3.17}$$

Since $z\nu = x\xi + \tau$, it follows from (3.17) that

$$\tau - \sigma_{\max}(\xi) \le z\nu \le \tau + \sigma_{\max}(-\xi)\,. \tag{3.18}$$

If (3.16) is satisfied, (3.18) shows that $z\nu$ lies between two numbers of the same sign and so is not zero. By (3.15) it follows that z does not lie on P, so P is spacelike. The restriction of the t-coordinate of z to be 1 is irrelevant. We conclude that P is strictly spacelike; this completes the proof of Theorem 3.5. □

When the underlying equation is strictly hyperbolic, and there is more than one space variable, the converse of Theorem 3.5 holds. For, since $\sigma_{\max}(-\xi)$ is the support function of K, there are points x and x' in K for which

$$\sigma_{\max}(-\xi) = x\xi\,, \quad \sigma_{\max}(\xi) = -x'\xi\,.$$

If condition (3.16) fails there are three possibilities:

(i) $\tau + x\xi = 0$,
(ii) $\tau + x'\xi = 0$, or
(iii) $\tau + x\xi$ and $\tau + x'\xi$ have opposite signs.

The points x and x' are boundary points of K. For a strictly hyperbolic equation the boundary of K is the boundary of the convex hull of K, and so for more than one space variable it is a connected set. Therefore, in case (iii) there is a point $x'' \in K$ for which $\tau + x''\xi = 0$. So one of the points $(x, 1)$, $(x', 1)$, or $(x'', 1)$ lies on the hyperplane defined by (3.15), and so P is not spacelike.

EXAMPLE Suppose that G is a symmetric first-order operator

$$G = \sum A_j D_j\,,$$

A_j symmetric matrices. Here $\sigma_{\max}(\xi)$ denotes the largest eigenvalue of

$$G(\xi) = \sum \xi_j A_j\,.$$

Condition (3.16) means that $\tau - \sigma_{\max}(\xi)$ and $\tau + \sigma_{\max}(-\xi)$ are of the same sign, which means that the smallest and largest eigenvalue of the characteristic matrix

$$I\tau - G(\xi) \tag{3.19}$$

are of the same sign. This is the same as saying that the matrix (3.19) is definite; thus according to Theorem 3.5, (ξ, τ) is normal of a spacelike hyperplane if and only if (3.19) is definite.

Next we observe that *the definiteness of* (3.19) *is sufficient for the direction* $\nu = (\xi, \tau)$ *to be hyperbolic.* For hyperbolicity means that for any $\zeta = (\eta, \kappa)$ not parallel to ν

$$P_0(s\nu + \zeta)$$

vanishes for n real values of s. The vanishing of P_0 means that the characteristic matrix $(s\tau + \kappa)I - G(s\xi + \eta)$ is not invertible, i.e., that it annihilates some nonzero vector v. Since $G(\xi)$ depends linearly on ξ, this can be written in the form

$$[\kappa I - G(\eta)]v = -s[\tau I - G(\xi)]v\,.$$

According to the spectral theory of symmetric matrices, if the matrix $\kappa I - G(\xi)$ is definite, this eigenvalue equation has only real roots s.

The above result is not surprising since hyperbolic directions were introduced as a natural generalization of the prototype (0,1). Indeed the foregoing results hold in general:

THEOREM 3.6 *The direction $\nu = (\xi, \tau)$ is strictly hyperbolic if the hyperplane whose normal is ν is strictly spacelike.*

PROOF: To prove this theorem, we factor the characteristic form

$$P_0(\xi', \tau') = \prod_{j=1}^{n} (\tau' - \sigma_j(\xi')),$$

where $\sigma_j(\xi')$ are arranged in decreasing order,

$$\sigma_1(\xi') > \sigma_2(\xi') > \cdots > \sigma_n(\xi')$$

for ξ' real and $\neq 0$. Since P_0 is homogeneous

$$\sigma_j(a\xi') = \begin{cases} a\sigma_j(\xi') & \text{for } a \text{ positive} \\ a\sigma_{n-j+1}(\xi') & \text{for } a \text{ negative} \end{cases} \tag{3.20}$$

and

$$\sigma_{\max}(\xi') = \sigma_1(\xi'), \qquad \sigma_{\max}(-\xi') = -\sigma_n(\xi'). \tag{3.21}$$

Suppose that ν is spacelike; we wish to show that ν is then strictly hyperbolic, i.e. that the equation

$$P_0(s\nu + \zeta) = 0 \tag{3.22}$$

is satisfied for n different real values of s for any vector $\zeta = (\eta, \kappa)$ not parallel to ν. Using the factored form of P_0, we see that (3.22) vanishes if and only if one of its factors does:

$$s\tau + \kappa - \sigma_j(s\xi + \eta) = 0. \tag{3.22j}$$

Different indices j furnish different roots s_j since the functions σ_j are distinct except at the origin; since ζ is not parallel to ν, $s\xi + \eta \neq 0$.

We shall show now that for each j (3.22j) has a real root by showing that the function on the left in (3.22j) has opposite sign for large positive and negative values of s. By (3.20) we can write the function on the left as

$$s\left[\tau - \sigma_j\left(\xi + \frac{\eta}{s}\right) + \frac{\kappa}{s}\right] \quad \text{for } s \text{ positive,}$$
$$s\left[\tau - \sigma_{n-j+1}\left(\xi + \frac{\eta}{s}\right) + \frac{\kappa}{s}\right] \quad \text{for } s \text{ negative.}$$

For $|s|$ large this can be written as

$$\begin{aligned} &s[\tau - \sigma_j(\xi) + \epsilon] \quad \text{for } s \text{ positive,} \\ &s[\tau - \sigma_{n-j+1}(\xi) + \epsilon] \quad \text{for } s \text{ negative.} \end{aligned} \tag{3.23}$$

Since ν was assumed spacelike, according to Theorem 3.5 condition (3.16) holds. Using (3.21) we can state (3.16) in the form: *$\tau - \sigma_1(\xi)$ and $\tau - \sigma_n(\xi)$ have the same sign.*

Since σ_1 and σ_n are the extreme roots, it follows that for all j, $\tau - \sigma_j(\xi)$ has the same sign. But then the function (3.23) has opposite sign for large positive and large negative values of s, which proves that each must vanish for some real value of s. □

The converse of Theorem 3.6 is left as an exercise.

The geometric picture behind this analytic discussion will be discussed in Section 3.4.

Let $u(x,t)$ be a solution with compact support in x. The boundary M of the support is a point set in x, t space that separates disturbed from undisturbed regions and thus can be thought of as a wave front. What is the shape of these possible wave fronts?

Denote by $s(\xi, t)$ the support function of the support of $u(x,t)$ at time t. It follows from Theorem 2.2 and Theorem 3.3 that for t positive

$$s(\xi, t) \leq s(\xi, 0) + t\sigma_{\max}(-\xi) . \tag{3.24}$$

Suppose that the sign of equality holds in (3.24) and that the boundary M of the support of u is a smooth surface in x, t space. Then an easy geometric argument shows that the normal to M is of the form

$$(\xi, -\sigma_{\max}(-\xi)) = (\xi, \sigma_{\min}(\xi)) .$$

This shows that the normal to M lies in the characteristic variety; such a surface is called a *characteristic surface*, see Section 3.4. We shall see in Chapter 7 that, as the above discussion indicates, characteristic surfaces play an important role in the more detailed description of motions.

The following useful result is an immediate consequence of Theorem 3.6:

THEOREM 3.7 *Let P be a spacelike hyperplane; through every $(k-1)$–dimensional linear subspace L in P there pass n distinct characteristic hyperplanes.*

PROOF: Let ν be the normal to P, and let L be the intersection of P with the hyperplane whose normal is ζ, ζ not parallel to ν. According to Theorem 3.6, the equation

$$P_o(s\nu + \zeta) = 0$$

has n distinct real solutions s_j. Clearly the hyperplanes with normal $s_j\nu + \zeta$ are characteristic, and they pass through L. □

In Section 3.4 we shall give a generalization of Theorem 3.7 where P is replaced by an arbitrary spacelike surface, L by any smooth $(k-1)$–dimensional submanifold, the characteristic hyperplane by characteristic hypersurfaces. In Section 4.2 we further extend this result to equations whose coefficients may vary with x and t.

3.3. The Initial Value Problem on Spacelike Hypersurfaces

THEOREM 3.8 *Let S be any smooth, strictly spacelike surface, given as $t = p(x)$. Let $u_0(x)$ be a smooth vector-valued function with compact support. Then there exists a motion $u(x, t)$ whose value on S equals u_0:*

$$u(x, p(x)) = u_0(x)\,.$$

Furthermore, u is uniquely determined.

PROOF: We shall construct a solution of the differential equation of the motion $u_t = Gu$ that has compact support in x and that equals u_0 on S; by the uniqueness theorem such a solution is a motion.

For simplicity we shall treat the special case when G is a first-order operator:

$$G = \sum A_j D_j + B\,.$$

We introduce $s = t - p(x)$ as a new variable; denote by $v(x, s)$ the function $u(x, s + p(x))$. Using the chain rule, we get

$$u_t = v_s\,, \quad u_x = v_x - v_s p_x\,.$$

In terms of v the differential equation is

$$v_s = \sum A_j(v_{x_j} - v_s p_{x_j})\,,$$

which can be rewritten as

$$\Big(I + \sum p_j A_j\Big) v_s = Gv\,.$$

Since the surface S is assumed to be strictly spacelike, according to Theorem 3.6 its normal ν is hyperbolic, which shows that ν does not lie on the characteristic variety. Since ν is equal to $(-p_x, 1)$, this shows that *the coefficient of v_s above is nonsingular* and so we solve for v_s:

$$v_s = Hv\,, \tag{3.25}$$

where H is a first-order differential operator in the x-variables whose coefficients are independent of s.

Repeated differentiation with respect to s gives

$$\frac{\partial^n}{\partial s^n} v = H^n v.$$

Define the functions $u_n(x)$, $n = 1, 2, \ldots,$ by

$$u_n = H^n u_0\,. \tag{3.26}$$

Let N be any positive integer and define v_N as the sum

$$v_N(x, s) = \sum_0^N \frac{s^n}{n!} u_n\,.$$

Clearly v_N satisfies the differential equation

$$\frac{\partial}{\partial s} v_N - H v_N = -\frac{s^N}{N!} u_{N+1}\,.$$

Returning to the original variables and denoting $v_N(x, t - p(x))$ by $r_N(x, t)$, we find that r_N satisfies a differential equation of the form

$$\frac{\partial}{\partial t} r_N - G r_N = (t - p)^N k_N . \tag{3.27}$$

Furthermore, r_N has compact support in x and equals u_0 on S; so r_N is a "near solution" to our problem. We shall make it into an exact solution by subtracting from it a suitable function w_N. For this we need the following:

LEMMA 3.9 *Given any smooth function $g(x, t)$ with compact support, the inhomogeneous differential equation*

$$w_t - G w = g \tag{3.28}$$

has a solution satisfying the initial condition

$$w(x, a) = 0 . \tag{3.28'}$$

PROOF: Taking the Fourier transform in the space variables we reduce (3.28) and (3.28′) to an initial value problem for an ordinary differential equation. □

We return to the proof of Theorem 3.8. Let's assume for sake of simplicity that $p(x)$ is bounded, that is, that $a^- < p(x) < a^+$ for all x. We take for the inhomogeneous term g in (3.28) the function g^- defined as follows:

$$g^-(x, t) = \begin{cases} (t - p)^N k_N & \text{for } t > p(x) \\ 0 & \text{for } t \le p(x). \end{cases}$$

Denote by w^- the solution of (3.28) with $g = g^-$ and $a = a^-$. For this choice of g^-, w^- satisfies the homogeneous equation

$$w_t^- - G w^- = 0$$

for $a^- \le t \le p(x)$. Since S is spacelike, every point (x, t) in this region belongs to the domain of determinacy of the initial plane $t = a^-$, and so, since $w^-(x, a^-)$ was chosen to be zero, $w^-(x, t)$ is zero in this region; in particular, $w^-(x, t) = 0$ on S.

Similarly, we define g^+ as

$$g^+(x, t) = \begin{cases} 0 & \text{for } t \ge p(x) \\ (t - p)^N k_N & \text{for } t < p(x), \end{cases}$$

and denote by w^+ the solution of (3.28), (3.28′) with $g = g^+$ and $a = a^+$. According to the previous argument, $w^+(x, t) = 0$ for $t \ge p(x)$. Since g^- and g^+ are N times differentiable, w^- and w^+ can be made as smooth as we wish by taking N large enough.

Define u as follows:

$$u(x, t) = \begin{cases} r_N - w^- & \text{for } t > p(x) \\ r_N - w^+ & \text{for } t < p(x). \end{cases} \tag{3.29}$$

Since r_N satisfies equation (3.27) and $w^{\pm}$ satisfy $(3.28)^{\pm}$, it follows that u satisfies $u_t - Gu = 0$. Since both w^+ and w^+ are zero on S, and since $r_N = u_0(x)$ on S, it follows that $u = u_0(x)$ on S. This proves the first part of Theorem 3.8.

To prove uniqueness, let u be a solution of $u_t = Gu$ that is zero on S. Define u^- by

$$u^-(x,t) = \begin{cases} u & \text{for } t \geq p(x) \\ 0 & \text{for } t \leq p(x). \end{cases} \tag{3.29'}$$

Equation (3.26) shows that not only u but all its derivatives are zero on S, so the function u^- defined above has continuous derivatives of all orders; furthermore, u^- satisfies the equation $u_t^- = Gu^-$. Since $a^- < p(x)$, $u_-(x, a^-)$ is, according to (3.29'), zero for all x. Therefore, according to the basic property of hyperbolic motions, $u^-(x,t) = 0$ for all x, t. This shows, in view of (3.29'), that $u(x,t) = 0$ for $t \geq p(x)$. That $u(x,t) = 0$ for $t \leq p(x)$ can be deduced analogously. □

If S and u_0 are C^∞, then, since N is arbitrary, u too is C^∞. The simplifying assumption that S lies between $t = a^-$ and $t = a^+$ is not hard to remove; we leave it as an exercise to the reader. By the uniqueness theorem, u defined by (3.29) is independent of N.

3.4. Characteristic Surfaces

For simplicity we shall deal with first-order hyperbolic systems of the form

$$Lu = D - tu - \sum A_j D_j u = 0, \tag{3.30}$$

A_j constant $n \times n$ matrices, u a vector function, $D_t = \partial/\partial t$, and $D_j = \partial/\partial x_j$. Given the value of u at $t = 0$, we can determine all derivatives of u with respect to x from equation (3.30) and its derivatives with respect to t.

A hyperplane in x, t space is called *free* if given the values of a solution u of (3.30) on the hyperplane we can determine all partial derivatives of u with respect to x and t. A hyperplane that is not free is called *characteristic* with respect to the operator l defined in (3.30).

To derive an algebraic criterion for a hyperplane to be characteristic we introduce new variables y, s in terms of which the hyperplane is given by $s = 0$. We set

$$s = t\tau + x\xi, \quad y = x;$$

then

$$D_t = \tau D_s, \quad D_j = \xi_j D_s + D_{y_j}.$$

Setting this into (3.30) we obtain

$$(\tau I - \sum \xi_j A_j) D_s u - \sum A_j D_{y_j} u = 0.$$

Clearly $s = 0$ is characteristic if and only if the matrix $\tau I - \sum \xi_j A_j$ is not invertible; here (ξ, τ) is normal to the hyperplane.

Denote by $\sigma(\xi)$ any one of the eigenvalues of $\sum \xi_j A_j$. Then the condition for the hyperplane with normal (ξ, τ) to be characteristic can be written as

$$\tau - \sigma(\xi) = 0. \tag{3.31}$$

Note that a spacelike hyperplane is never characteristic.

The notion of characteristic can be extended to hypersurfaces:

DEFINITION 3.10 A hypersurface in x, t space is characteristic for the operator L if all its tangent hyperplanes are characteristic for L

The significance of characteristic hypersurfaces for the propagation of signals is that they are the carriers of discontinuities. A discontinuous solution is defined as solution in the *weak sense* as follows:

A piecewise differentiable function $u(x, t)$ that has a discontinuity across a hypersurface S is called a *weak solution* of equation (3.30) if for all C^∞ functions w of compact support

$$\int_{\mathbb{R}^k \times \mathbb{R}} u L^* w \, dx \, dt = 0, \tag{3.32}$$

where L^* denotes the adjoint of L,

$$L^* = -D_t + \sum D_j A_j .$$

When u is everywhere differentiable, we can integrate (3.32) by parts to obtain

$$\int (Lu) w \, dx \, dt = 0.$$

Since w is an arbitrary smooth function with compact support, it follows that a differentiable function u is a genuine solution of (3.30) if and only if it is a weak solution.

The argument above shows that at all points where a weak solution u is differentiable, it satisfies pointwise the equation $Lu = 0$. In particular, a piecewise differentiable weak solution that has a discontinuity across a surface S satisfies $Lu = 0$ on either side of S.

Take now any open set that is intersected by S; denote by G_1 and G_2 the parts of G that lie on opposite sides of S (see Figure 3.1). Let w be any smooth function whose support lies in the closure of G. We write (3.32) as a sum

$$\int_{G_1} u L^* w \, dx \, dt + \int_{G_2} u L^* w \, dx \, dt = 0. \tag{3.32'}$$

We integrate by parts each term. Since $Lu = 0$ on either side of S, and since $w = 0$ at those boundary points of G_1 and G_2 that do not lie on S, we get

$$\int_S \left(\tau I - \sum \xi_j A\right)[u] w \, dS = 0, \tag{3.33}$$

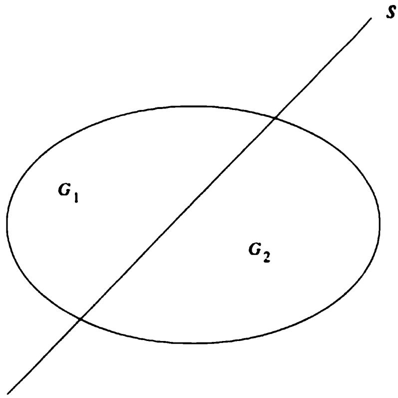


FIGURE 3.1

where (ξ, τ) is the normal to S, $[u]$ the difference in the values of u on the two sides of S, and dS the surface area element. Since the value of w on S is an arbitrary smooth function, it follows from (3.33) that

$$\left(\tau I - \sum \xi_j A_j\right)[u] = 0 \tag{3.33'}$$

on S. If S were noncharacteristic, the matrix $\tau I \sum \xi_j A_j$ would be invertible, and (3.33′) would imply that $[u] = 0$. This shows that *jump discontinuities can occur only across characteristic surfaces.*

We turn now to the construction of characteristic surfaces. We shall describe these surfaces implicitly by $\varphi(x, t) = \text{const}$; we shall assume that these surfaces are characteristic for all values of the constant. The normal to the surface $\varphi = \text{const}$ is $(D_x\varphi, D_t\varphi)$; setting this into (3.31) gives

$$D_t\varphi - \sigma(D_x\varphi) = 0, \tag{3.34}$$

a nonlinear partial differential equation for φ, called the *eikonal equation.*

Since σ is a homogeneous function of order 1 of its arguments, it satisfies the relation

$$\sigma(\xi) = \sum \sigma_j \xi_j, \tag{3.35}$$

where $\sigma_j = \partial\sigma/\partial\xi_j$. Setting this into (3.34) gives

$$D_t\varphi - \sum \sigma_j D_j\varphi = 0. \tag{3.36}$$

This implies that φ is constant along the curves defined by

$$\frac{dx_j}{dt} = -\sigma_j(D_x\varphi). \tag{3.37}$$

We shall now show that $D_x\varphi$—and therefore σ_j—are constant along such a curve, and therefore these curves are straight lines. To see this differentiate equation

(3.34) with respect to x_i; we get

$$0 = D_i D_t \varphi - \sigma_j D_i D_j \varphi = D_t \varphi_i - \sigma_j D_j \varphi_i ,$$

where φ_i abbreviates $D_i \varphi$. The constancy of φ_i along the curves (3.37) follows, therefore these curves are straight lines.

Next we show how to put together the values of φ along these straight line solutions of (3.37) to construct a solution φ of (3.34). Choose $\varphi_0(x) = \varphi(x, 0)$ as any C^∞ function of x. From each point x of $\mathbb{R}^k$ there issues a straight line in (x, t) space defined by equation (3.37), where the values of $D_x \varphi$ are those of $D_x \varphi_0$. Suppose that the first derivatives φ_0 are uniformly bounded in $\mathbb{R}^k$; then it is not hard to show that there is a time T such that these straight lines fill the slab $\mathbb{R}^k \times (-T, T)$ in a one-to-one fashion. We then define $\varphi(x, t)$ along each line to be equal to φ_0 at the point where the line starts. Clearly, φ satisfies (3.34), and so the hypersurfaces $\varphi(x, t) = \text{const}$ are characteristics.

THEOREM 3.11 *Through any smooth $(k - 1)$–dimensional manifold in $\mathbb{R}^k$ there pass n characteristic surfaces.*

PROOF: Choose φ_0 to be 0 on the prescribed $(k - 1)$–dimensional manifold, and φ to be the solution of one of the n eikonal equations (3.34). □

A hyperbolic operator L of form (3.30) has n characteristic fields $\sigma^{(1)}, \dots, \sigma^{(n)}$ corresponding to the n real eigenvalues of $\sum \xi_j A_j$. Therefore, through any given smooth $(k - 1)$–dimensional manifold in $\mathbb{R}^k$ there pass n characteristic surfaces, one of each field $\sigma^{(i)}$. i We now describe an especially important characteristic surface, a *characteristic cone*. These are formed by the set of straight lines defined by (3.37), all issuing from the same point, say the origin (0, 0), in the direction $-D\sigma(\omega)$, $D\sigma$ the gradient of $\sigma(\omega)$, where $\sigma(\omega)$ is one of the eigenvalues of $\sum \omega_j A_j$, and ω runs through all unit vectors. Define H to be the intersection of this cone with the hyperplane $t = 1$. H consists of the points $-D\sigma(\omega)$, $|\omega| = 1$. Define $\rho(\omega) = \sigma_{\max}(-\omega)$; then

$$H : D\rho(\omega), \quad |\omega| = 1 . \tag{3.38}$$

We are particularly interested in the characteristic cone corresponding to the largest eigenvalue $\sigma_{\max}$. We recall from Corollary 3.4 that $\sigma_{\max}(\xi)$ is a convex function; therefore so is $\rho(\xi)$. Since σ and ρ are positive homogeneous as well, it follows that ρ is subadditive,

$$\rho(\omega + \eta) \le \rho(\omega) + \rho(\eta) .$$

Replace η by $\varepsilon\xi$, ε any positive number,

$$\rho(\omega + \varepsilon\xi) \le \rho(\omega) + \varepsilon\rho(\xi) .$$

At $\varepsilon = 0$ equality holds. Therefore at $\varepsilon = 0$ the derivative of the left side with respect to ε is less than or equal to the derivative of the right side,

$$\sum \rho_j(\omega)\xi_j \le \rho(\xi) . \tag{3.39}$$

We now recall from the beginning of this chapter the notion of the support function of a compact set H in $\mathbb{R}^k$:

$$s_H(\xi) = \max_{x \in H} x\xi .$$

For the set H defined by (3.38) it follows from (3.39) that $s_H(\xi) \leq \rho(\xi)$. On the other hand, it follows from (3.35) that for $\omega = \xi$ the sign of equality holds in (3.39); this shows that

$$s_H(\xi) = \rho(\xi) = \sigma_{\max}(-\xi) .$$

The argument above shows that through every point of H lies a supporting hyperplane; i.e., all points of H lie on one side of it. It follows that H is a convex hypersurface; i.e., H and its interior, consisting of the set of points

(3.38′) $$(r\rho_1(\omega), \dots, r\rho_k(\omega)), \quad |\omega| = 1, \ 0 \leq r \leq 1 ,$$

form a convex set in $\mathbb{R}^k$.

We now turn to the set of points K that belong to the domain of influence of the origin, and lie on the hyperplane $t = 1$. According to Theorem 3.3, the support function of this set K is $\sigma_{\max}(-\xi)$, the same as the support function of the set H. It then follows from the hyperplane separation theorem that

THEOREM 3.12 *The domain of influence of the origin is contained inside or on the characteristic cone corresponding to $\sigma_{\max}$ issuing from the origin.*

EXERCISE Let $P(D_x, D_t)$ be an n^{th}-order scalar operator. Show that $t\sigma + x\xi = 0$ is characteristic for $Pu = 0$ if and only if $P(\xi, \sigma) = 0$.

3.5. Solution of the Initial Value Problem by the Radon Transform

In this section we shall express in a fairly direct fashion solutions of hyperbolic equations in terms of their initial data. The prototype for the type of expression we are looking for is furnished by

$$u(x, t) = h(x - ct) + k(x + ct)$$

for solutions of the one-dimensional wave equation $u_{tt} - c^2 u_{xx} = 0$. Direct verification shows that every function of the above form is a solution of the wave equation, and by choosing h and k appropriately we can satisfy initial conditions imposed on u.

Let L be any scalar partial differential operator with constant coefficients and homogeneous of order n; i.e., L does not contain terms of order lower than n. Such equations have special solutions, called *plane waves*, of the form

(3.40) $$u(x, t) = h(x\omega + t\tau) .$$

One can verify immediately that (3.40) is a solution of $Lu = 0$ if the vector (ω, τ) is real and satisfies the characteristic equation $L(\omega, \tau) = 0$, and $h(s)$ is any function of the real variable s.

If L is hyperbolic, we may take for (ω, τ) any point on one of the $[\frac{n+1}{2}]$ real branches of the characteristic variety. The question is: Can every solution, or at least those which have compact support in x, be expressed as a superposition of plane waves? The answer is yes, and the demonstration consists in showing that we can satisfy arbitrary initial conditions by suitably chosen linear combinations of plane waves. We shall carry out the details for the classical case of the wave equation

$$u_{tt} - \Delta u = 0 . \tag{3.41}$$

In this case every plane wave can be written as $h(x\omega - t)$, $|\omega| = 1$; so we are looking for solutions in the form

$$u(x,t) = \int_{|\omega|=1} h(x\omega - t, \omega) d\omega . \tag{3.42}$$

The function $h(s, \omega)$ has to be chosen so that the initial conditions

$$\begin{aligned} u(x,0) &= f_1(x) = \int h(x\omega, \omega) d\omega , \\ u_t(x,0) &= f_2(x) = -\int h'(x\omega, \omega) d\omega , \end{aligned} \tag{3.43}$$

are satisfied.

The contribution of the odd part of h to the first integral in (3.43), and of its even part to the second integral, is zero. part. Therefore, in order to solve (3.43) it is sufficient to solve the following problem:

Given a function $p(x)$, find an even function $\ell(s, \omega)$ such that

$$p(x) = \int \ell(x\omega, \omega) d\omega \quad \text{and} \quad \ell(-s, -\omega) = \ell(s, \omega) . \tag{3.44}$$

The solution of this problem is furnished by the *Radon transform*, whose theory has been expounded by Fritz John, Helgason, and Gelfand-Graev-Vilenkin. We shall outline the theory in $\mathbb{R}^k$, k odd and > 1.

We start with the Fourier representation of f:

$$f(x) = c \int \tilde{f}(\xi) e^{ix\xi} \, d\xi , \tag{3.45}$$

where $\tilde{f}$ is the Fourier transform of f,

$$\tilde{f}(\xi) = c \int f(x) e^{-ix\xi} \, dx . \tag{3.46}$$

So as not to weary the reader (and the author), the letter c in these and subsequent formulas denotes the right constant. In formula (3.46) we express ξ in polar coordinates as $\xi = \rho\omega$ and write

$$\tilde{f}(\rho\omega) = c \int f(x) e^{-i\rho\omega x} \, dx .$$

Carrying out the x-integration first on the hyperplane $\omega \cdot x = s$, we get

$$\tilde{f}(\rho\omega) = c \int \hat{f}(s, \omega) e^{-i\rho s} \, ds , \tag{3.47}$$

where

$$\hat{f}(s,\omega) = \int_{\omega x = s} f(x)\,dS\,. \tag{3.48}$$

$\hat{f}$ is the *Radon transform* of f. Formula (3.47) shows that $\tilde{f}(\rho\omega)$ is the Fourier transform of $\hat{f}(s,\omega)$ with respect to s.

In terms of polar coordinates, we can write the Fourier representation (3.45) of f as follows:

$$f(x) = c\iint \tilde{f}(\rho\omega)e^{ix\omega\rho}\rho^{k-1}\,d\rho\,d\omega\,. \tag{3.49}$$

We take the ρ-integration over all of $\mathbb{R}$, at the cost of cutting the constant c in half. It is at this point that we exploit the evenness of $k-1$.

Inverting the Fourier transform (3.47), we get

$$\hat{f}(s,\omega) = c\int \tilde{f}(\rho\omega)e^{i\rho s}\,d\rho\,. \tag{3.47'}$$

Applying $D_s = \frac{\partial}{\partial s}$, $k-1$ times, we get

$$D_s^{k-1}\hat{f}(s,\omega) = c\int \tilde{f}(\rho\omega)\rho^{k-1}e^{i\rho s}\,d\rho\,.$$

Setting this into (3.49), we get

$$f(x) = c\int D_s^{k-1}\hat{f}(x\omega,\omega)\,d\omega\,, \tag{3.50}$$

a representation of form (3.44), with $\ell(s,\omega) = D_s^{k-1}\hat{f}(s,\omega)$.

We list now the properties of the Radon transform:

THEOREM 3.13

(i) *$\hat{f}(s,\omega)$ is an even function,*

$$\hat{f}(-s,-\omega) = \hat{f}(s,\omega)\,.$$

(ii) *The Parseval relation holds:*

$$\int |f(x)|^2\,dx = c\int \left|D_s^{\frac{k-1}{2}}\hat{f}\right|^2 ds\,d\omega\,. \tag{3.51}$$

(iii) *Every even function $m(s,\omega)$ for which $D_s^{(k-1)/2}m$ lies in L^2 is the Radon transform of some f in L^2.*

(iv) $\widehat{Df} = -\omega D_s\hat{f}$.

(v) $\widehat{\Delta f} = D_s^2\hat{f}$.

PROOF: Properties (i), (iv), and (v) follow from formula (3.48) for $\hat{f}$. To deduce (ii) and (iii), apply $D_s^{(k-1)/2}$ to (3.47′) and use the one-dimensional Parseval relation to obtain, after integration with respect to ω,

$$\int \left|D_s^{\frac{k-1}{2}}\hat{f}\right|^2 ds\,d\omega = c\int |\tilde{f}(\rho\omega)|^2\rho^{k-1}\,d\rho\,d\omega\,.$$

The Parseval relation between the L^2 norm of f and $\tilde{f}$ completes the derivation of (3.51). □

From the explicit expression (3.48) for the Radon transform, we can read off the following important consequence:

THEOREM 3.14 *If $f(x)$ vanishes for $|x| > r$, its Radon transform $\hat{f}(s,\omega)$ vanishes for $|s| > r$.*

We return now to expression (3.42) for the solution of the wave equation. Choose the function h as

$$h(s,\omega) = D_s^{k-1}\hat{f}_1 - D_s^{k-2}\hat{f}_2 . \tag{3.52}$$

Setting this choice of h into formula (3.43) we see that u defined by (3.42) has the assigned initial values. Suppose now that both f_1 and f_2 are zero for $|x| > r$; then by Theorem 3.14 both $\hat{f}_1$ and $\hat{f}_2$ are zero for $|s| > r$; it follows that $h(s,\omega)$ is zero for $|s| > r$. Looking at the explicit expression (3.42) for u in terms of h, we see that if $|x| < |t| - r$, then the integrand on the right is zero for all ω. So we have proven the following:

LEMMA 3.15 *If the initial data of u are zero for $|x| > r$, then $u(x,t)$ is zero for $|x| < |t| - r$.*

THEOREM 3.16 *In an odd number of space dimensions k, $k > 1$, the domain of influence of the origin for the wave equation* (3.41) *consists of the double cone $|x| = |t|$.*

PROOF: According to the results of Section 3.1, if $|x| > |t|$, then (x,t) does not belong to the domain of influence of the origin. Suppose that $|x| < |t|$; to show that x, t lies outside the domain of influence of the origin, we have to show that if the initial data of u are zero outside a ball of radius r, r small enough, then u is zero at x, t. But this follows from Lemma 3.15. □

Define the energy of the initial data f_1, f_2 of a solution of the wave equation as

$$\mathrm{E} = \int (|D_x f_1|^2 + f_2^2)\,dx . \tag{3.53}$$

We claim that

$$\mathrm{E} = c\int \left|D_s^{\frac{k-1}{2}}\ell\right|^2 ds\,d\omega , \tag{3.54}$$

where ℓ is defined as

$$\ell = D_s\hat{f}_1 - \hat{f}_2 ; \tag{3.55}$$

note that ℓ is related to h defined in (3.52) by $h = D_s^{k-2}\ell$. To deduce this we combine parts (ii) and (iv) of (3.51) to get

$$\int |D_x f_1|^2\,dx = c\int \left|D_s^{\frac{k+1}{2}}\hat{f}_1\right|^2 ds\,d\omega , \tag{3.54$_1$}$$

while

$$\int f_2^2\,dx = c\int \left|D_s^{\frac{k-1}{2}}\hat{f}_2\right|^2 ds\,d\omega. \tag{3.54$_2$}$$

$D_s^{\frac{k+1}{2}}\hat{f}_1$ and $D_s^{\frac{k-1}{2}}$ are the even and odd parts of $D_s^{\frac{k-1}{2}}\ell$, so they are orthogonal. Therefore adding (3.54$_1$) and (3.54$_2$) yields (3.54).

We can use formula (3.54) to express the energy contained in the data of u at any time t. If follows from formula (3.42) and $h = D_s^{k-2}\ell$ that

$$D_s\hat{u}(t) - \hat{u}_t(t) = \ell(s-t,\omega).$$

Expression (3.54) does not change if ℓ is subjected to a translation in s; thus we have proven

THEOREM 3.17 *The energy of a solution u of the wave equation* (3.41) *defined as*

$$\mathrm{E} = \int \left(|D_x u(x,t)|^2 + u_t^2(x,t)\right)dx$$

is independent of t.

Theorem 3.16 is the celebrated *Huygens principle*. It is *false* in an even number of space variables.

Given a general hyperbolic equation, we can, following the method outlined above for the wave equation, express solutions of it as superposition of plane waves:

$$u(x,t) = \sum_{j=1}^{<\frac{n+1}{2}} \int_{|\omega|=1} h_j(x\omega + \tau_j(\omega)t,\omega)\,d\omega. \tag{3.42$'$}$$

The functions $h_j(s,\omega)$ can be expressed as linear combinations of the integrals with respect to s of the Radon transforms of the initial data of u. The details are left to the reader.

There are further and more delicate generalizations of the Huygens principle; see Appendix A.

At the beginning of this section, we set out to express solutions of hyperbolic equations in terms of their initial data. The expressions we have found are in terms of the Radon transform of the initial data. Using expression (3.48) for these Radon transforms yields a formula for solutions directly in terms of their initial data; however, this formula is best interpreted in the language of distributions. The details will be carried out, with a slightly different twist, in Chapter 7.

3.6. Conservation of Energy

In this last section of Chapter 3 we shall discuss conservation of energy. For simplicity, we shall deal with symmetric hyperbolic systems, i.e., equations of the form

$$u_t = \sum A_j u_{x_j} = Gu, \tag{3.56}$$

where the A_j are real symmetric matrices.

THEOREM 3.18 *Let u be a solution of* (3.56) *with compact support in x. Then the quantity*

$$\|u(t)\|^2 = \int_{R^k} |u^2(x,t)|^2\,dx \tag{3.57}$$

is independent of t.

Quantity (3.57) is called *energy*.

In view of the importance of this result, we give two different proofs.

PROOF 1: Denote as usual by (u, v) the L_2 scalar product of the vector functions u and v with respect to the space variables. Then quantity (3.57) can be written as (u, u). Differentiate with respect to t to get, using the differential equation (3.56),

$$\frac{d}{dt}(u,u) = (u_t, u) + (u, u_t) = (Gu, u) + (u, Gu) = ([G + G^*]u, u)\,, \tag{3.58}$$

where G^* denotes the adjoint of G. Since $A_j^* = A_j$,

$$G^* = -\frac{\partial}{\partial x^j} A_j^* = -G\,;$$

i.e., the operator G is *antisymmetric*. This shows that the expression (3.58) is zero. □

PROOF 2: The Fourier transform of u at time t can be expressed explicitly in terms of the Fourier transform of the initial values:

$$\tilde{u}(\xi, t) = e^{i(\sum \xi_j A_j)t}\tilde{u}(\xi, 0)\,.$$

Since $i(\sum \xi_j A_j)t$ is an antisymmetric matrix, its exponential is unitary; so

$$|\tilde{u}(\xi, t)|^2 = |\tilde{u}(\xi, 0)|^2\,.$$

Integrating with respect to ξ and using the fact that Fourier transformation preserves the L_2 norm, we get (3.57). □

In Section 3.5 we proved the conservation of energy for solutions of the wave equation using a representation of solutions in terms of the Radon transform of their initial data. This proof can be extended to solutions of hyperbolic equations of any order.

In the next chapter we shall take up the problem of formulating and proving a law of conservation of energy for hyperbolic equations with variable coefficients.

References

Ahlfors, L. V. *Complex analysis*. An introduction to the theory of analytic functions of one complex variable. McGraw-Hill, New York–Toronto–London, 1953.

Gelfand, I. M., Gindikin, S. G., and Graev, M. I. *Selected topics in integral geometry*. Translations of Mathematical Monographs, 220. American Mathematical Society, Providence, R.I., 2003.

Helgason, S. *The Radon transform*. Progress in Mathematics, 5. Birkhäuser, Boston, 1980.

John, F. *Plane waves and spherical means applied to partial differential equations*. Wiley Interscience, New York–London, 1955.

Paley, R. E. A. C., and Wiener, N. *Fourier transforms in the complex domain*. Reprint of the 1934 original. American Mathematical Society Colloquium Publications, 19. American Mathematical Society, Providence, R.I., 1987.

Plancherel, M., and Polya, G. Fonctions entières et intégrales de Fourier multiples. *Comment. Math. Helv.* 9: 224–248, 1937; 10: 110–163, 1938.

CHAPTER 4

Hyperbolic Equations with Variable Coefficients

This chapter is about linear hyperbolic equations with smoothly variable coefficients. Such an equation is called *strictly hyperbolic* if for every choice of y, s the equation with constant coefficients frozen at y, s is strictly hyperbolic. We shall not give a general definition of a nonstrictly hyperbolic equation although we shall present at least one example.

The characteristic form $P_0(z, \zeta)$ of a hyperbolic equation with variable coefficients is a form of degree n in $\zeta = (\xi, \tau)$ whose coefficients depend on $z = (x, t)$. Throughout this chapter we shall assume that the coefficients are infinitely differentiable functions of z.

Section 4.1 contains the theory of hyperbolic equations in one space variable. Section 4.2 describes characteristic surfaces. Sections 4.3, 4.4, and 4.5 present energy inequalities for solutions of symmetric hyperbolic systems, second-order hyperbolic equations, and higher-order hyperbolic equations, respectively, and the uniqueness theorems that follow from them.

4.1. Equations with a Single Space Variable

In this section we give a thumbnail sketch of the theory for one space variables, which is very much simpler than for many space variables. We shall treat first-order systems; general hyperbolic equations can be turned into first-order systems by introducing the higher derivatives as new unknowns; this is no longer possible if there are more space variables. Courant-Hilbert, vol. II, chap. V.

A first-order system is of the form

$$u_t = Au_x + Bu \,. \tag{4.1}$$

We assume that the coefficient matrices A and B are infinitely differentiable functions of x and t, and that (4.1) is strictly hyperbolic. The latter means that for every x, t, the matrix A has real and distinct eigenvalues.

We want to construct solutions of (4.1) with prescribed initial values,

$$u(x, 0) = u_0(x)\,; \tag{4.2}$$

unless specified otherwise, we assume that $u_0(x)$ is an infinitely differentiable function of x.

Denote by $V = V(z)$ the matrix whose columns are the right eigenvectors of A, normalized in some convenient fashion so that V depends infinitely differentiably on z. V satisfies the eigenvalue equation

$$AV = VT\,,$$

where T is the diagonal matrix with elements $\sigma_1, \dots, \sigma_n$. Introduce a new unknown v related to u by

$$u = Vv\,.$$

The equation satisfied by v is

$$v_t = T v_x + C v\,, \tag{4.1'}$$

where $C = V^{-1}BV + V^{-1}AV_x - V^{-1}V_t$. Componentwise this equation reads

$$v_{j,t} = \sigma_j v_{j,x} + \sum c_{jk} v_k\,. \tag{4.1j}$$

The differentiated terms can be combined into a single directional derivative

$$\frac{dv_j}{dj} = \sum c_{jk} v_k \tag{4.3j}$$

where

$$\frac{d}{dj} = D_t - \sigma_j D_x\,. \tag{4.4}$$

Equations (4.3j) and (4.4) constitute a system of ordinary differentiable equations but along different curves. Such equations can be solved by methods used to solve ordinary differential equations, e.g., Picard iteration. A slightly different twist is needed at the end.

Let C_j denote the trajectory of the j^{th} direction field through some point (y, s); i.e., $C_j = (x(t), t)$, where $x(t)$ is a solution of the differential equation

$$\frac{dx(t)}{dt} = -\sigma_j(x, t)\,. \tag{4.5}$$

These curves are *characteristic curves.*

Integrate (4.3j) along the j^{th} characteristic curve C_j between the point y, s and the intercept of C_j and the initial line $t = 0$; we obtain an integral relation

$$v_j(y, s) = v_j(y_j, 0) + \int_0^s \sum c_{jk} v_k \, dt\,. \tag{4.6}$$

We abbreviate this as

$$v = v_0 + Kv\,, \tag{4.6'}$$

where v_0 is determined by the initial values of v and K is an integral operator.

Take some fixed point (x_0, t_0); the two extreme characteristic curves C_1 and C_n issuing from it and the initial line $t = 0$ together bound a curved triangular region Δ; it can easily be shown that any point in Δ can be connected to the initial line by a characteristic curve of the j^{th} kind, $j = 1, 2, \dots, n$, which lies entirely in Δ. That means that the integral operator K defined by (4.6′) maps functions defined in Δ into functions defined in Δ. It is easy to show that K maps continuous functions defined in Δ into continuous functions in Δ, and that it is of Volterra type:

$$|K^m| \le \frac{\text{const}\, M^m}{m!}\,, \tag{4.7}$$

where K^m is the m^{th} power of the operator K, and $|K|$ is the operator norm with respect to the maximum over Δ.

EXERCISE Prove (4.7).

It follows from (4.7) that $I - K$ is invertible,

$$(I - K)^{-1} = \sum_0^\infty K^m,$$

so (4.6′) has a unique solution.

Since a solution of the initial value problem for the partial differential equation is a solution of the integral equation (4.6), it follows that the former has at most one solution. We still need to show that every solution of the integral equation (4.6) is a solution of the partial differential equation.

An immediate consequence of the integral relation (4.6) is that the j^{th} component of v has a directional derivative that satisfies (4.3j). It is not quite obvious, however, that the solution v has continuous partial derivatives. To show this we proceed as follows:

Let C^1 be the space of functions defined in Δ that have continuous first partial derivatives in the closure of Δ. C^1 is a complete normed linear space. It is easy to show that the operator K maps C^1 into itself and that it is of Volterra type, i.e., that (4.7) is satisfied in the sense of the C^1 norm. Then we conclude as before that $I - K$ is invertible so that (4.6′) has a unique solution in C^1. This solution is of course the same as the one constructed before.

This argument can be repeated for the class of m times differentiable functions, m arbitrary, and leads to the following existence theorem:

THEOREM 4.1 *Suppose that the initial function $u_0(x)$ is infinitely differentiable; then the initial value problem* (4.1)–(4.2) *has a uniquely defined, infinitely differentiable solution. If u_0 has continuous derivatives up to order $m, m \geq 1$, there exists a solution with continuous partial derivatives up to order m.*

Our method of construction yields the following:

COROLLARY 4.2 *The domain of influence of any point is the region contained between the two extreme characteristics issuing from it.*

The construction of solutions of hyperbolic equations in more than one space variable is harder than in the one-dimensional case. We shall give two existence proofs: the first one, in Chapter 6, is entirely indirect and is based on inequalities derived in Sections 4.3–4.5. The second one, in Chapter 7, is more constructive and gives some further information about the manner of dependence of solutions on initial data.

4.2. Characteristic Surfaces

We shall again deal with first-order hyperbolic systems of the form

$$Lu = D_t u = \sum A_j D_j u = 0, \tag{4.8}$$

A_j matrices of order $n \times n$ that are C^∞ functions of x and t.

The operator L in (4.8) is strongly hyperbolic if all eigenvalues σ of $\sum \xi_j A_j$ are real and distinct for all real choices of ξ. Since the matrices A_j are functions of x and t, so are the eigenvalues:

$$\sigma = \sigma(x, t, \xi). \tag{4.9}$$

As in Section 3.4, we define a surface S to be *characteristic* for L if at every point x, t on S, the normal (ξ, τ) to S satisfies

$$\tau - \sigma(x, t, \xi) = 0. \tag{4.10}$$

A weak solution of equation (4.8) is defined in the same way as in Section 3.4 for equations with constant coefficients:

$$\int u L^* w \, dx \, dt = 0, \tag{4.11}$$

where L^* is the adjoint of L , holds for all C^∞ functions w with compact support, and just as for equations with constant coefficients, discontinuities of piecewise continuous weak solutions can occur only along characteristic surfaces.

We now turn to the construction of characteristic surfaces. As before we shall describe them implicitly by $\varphi(x, t) = \text{const.}$ The normal to such a surface is $(D_x\varphi, D_t\varphi)$; setting this into (4.10) gives

$$D_t\varphi - \sigma(x, t, D_x\varphi) = 0, \tag{4.12}$$

called the *eikonal equation.*

Since σ is a homogeneous function of ξ of order 1,

$$\sigma(\xi) = \sum \sigma_j \xi_j, \tag{4.13}$$

where $\sigma_j = \partial\sigma/\partial\xi$. Setting this into (4.12) gives

$$D_t\varphi - \sum \sigma_j D_j \varphi = 0. \tag{4.14}$$

This implies that φ is constant along the curves defined by

$$\frac{dx_j}{dt} = -\sigma_j(x, t, D_x\varphi). \tag{4.15}$$

In order to determine these curves we need a differential equation for $D_x\varphi$ along such a curve. This can be obtained by differentiating (4.12) with respect to x_i; we get

$$D_i D_t \varphi - \sum \sigma_j D_i D_j \varphi - D_i \sigma = 0,$$

which can be rewritten as

$$D_t \xi_i - \sum \sigma_j \xi_j = D_i \sigma. \tag{4.16}$$

Here ξ_i denotes $D_i\varphi$, and $D_i\sigma$ is the partial derivative of $\sigma(x, t, \xi)$ with respect to x_i. We combine (4.15) and (4.16) into a system of ordinary differential equations:

$$\begin{aligned} \frac{dx_j}{dt} &= -\frac{\partial\sigma}{\partial\xi_j}(x, t, \xi), \\ \frac{d\xi_i}{dt} &= \frac{\partial\sigma}{\partial x_i}(x, t, \xi). \end{aligned} \tag{4.17}$$

This is a Hamiltonian system of differential equations that can be solved uniquely once the initial values of x and ξ are specified, at, say, $t = 0$.

Solutions of (4.17) are called *bicharacteristic*; the projection of a bicharacteristic into x space is called a *ray*.

Just as in the case of constant coefficients, we can build characteristic surfaces out of rays. We choose $\varphi_0(x) = \varphi(x, 0)$ as any C^∞ function of x. From each point x_0 there issues a ray, obtained by solving the Hamiltonian system (4.17) with the initial value of ξ_i given by $D_i\varphi_0(x)$. It is not hard to show that if the first derivatives of φ_0 are uniformly bounded in $\mathbb{R}^k$, the rays cover some slab $\mathbb{R}^k \times (-T, T)$ in a one-to-one fashion. We define $\varphi(x, t)$ to be equal to the value of $\varphi_0(y)$ at the point y where the ray through x, t originates.

EXERCISE Verify that for φ defined this way the surfaces $\varphi =$ const are characteristic.

The analogues of characteristic cones are *conoids*, defined similarly as in the case of equations with constant coefficients. They are formed by all the rays issuing from a single point x_0, t_0, as the initial values of ξ range over all unit vectors ω. Of particular interest are the characteristic cones corresponding to the largest eigenvalue $\sigma_{\max}(x, t, \xi)$. In Section 4.3 we shall show, for symmetric first-order systems, that any point influenced by x_0, t_0 is contained inside or on the characteristic conoid corresponding to $\sigma_{\max}$ issuing from x_0, t_0.

It should be noted that the characteristic surfaces constructed in this section, including the characteristic conoids, exist only for a finite time interval. Eventually they develop wrinkles and other singularities. Therefore the description of the domain of influence in terms of the characteristic conoids works only for a limited time interval. The domain of influence for all time can be obtained by combining the local time description with the Huygens wave construction; see Theorem 2.3.

We shall encounter the eikonal equation and bicharacteristics again in Section 7.2 on progressing waves.

4.3. Energy Inequalities for Symmetric Hyperbolic Systems

In this section we shall derive so-called energy inequalities for solutions of symmetric hyperbolic systems of first-order equations, i.e., equations of the form

$$Lu = u_t - Gu = 0 \tag{4.18}$$

where

$$G = \sum A_j D_j + B, \quad D_j = \frac{\partial}{\partial x^j}, \tag{4.19}$$

A, B matrices depending smoothly on x and t, A_j symmetric. These are analogues to energy identities derived in Section 3.6.

As usual, we shall denote the L_2 inner product of u and v with respect to x by (u, v), and the L_2 norm of u by $\|u\|$.

THEOREM 4.3 *Let u be a solution of* (4.18) *with compact support in x. Then*

$$\|u(t)\| \le e^{M|t|}\|u(0)\|, \tag{4.20}$$

where the constant M depends on the magnitude of the symmetric part of B and of the first derivative of A_j with respect to x_j.

The proof is based on the following simple lemma:

LEMMA 4.4 *Let G be an operator whose domain and range lie in a Hilbert space. Suppose that G is almost antisymmetric in the sense that $G + G^*$ is a bounded operator, say*

$$\|G + G^*\| \leq 2M, \tag{4.21}$$

where G^* denotes the Hilbert space adjoint of G. Then every solution of

$$u_t = Gu$$

satisfies the energy inequality (4.20).

PROOF OF LEMMA 4.4: To prove the lemma we form $E(t) = (u, u)$, where $(\cdot, \cdot)$ is the scalar product in Hilbert space. Differentiating with respect to t gives

$$\frac{dE}{dt} = (u_t, u) + (u, u_t) = (Gu, u) + (u, Gu) = (u, G + G^* u)\,.$$

Estimating the expression on the right by the Schwarz inequality and using the fact that $G + G^*$ is bounded by $2M$ gives

$$\left|\frac{dE}{dt}\right| \leq 2ME\,.$$

Multiplying this differential inequality by e^{-Mt} and integrating with respect to t gives

$$e^{-2MT}E(T) \leq E(0) \leq e^{2MT}E(T)\,,$$

as asserted in (4.20). □

PROOF OF THEOREM 4.3: To prove the theorem we have to verify that G as given by (4.19), its domain consisting of smooth functions with compact support, is almost antisymmetric. The adjoint is easily computed:

$$G^* = -\sum D_j A_j + B^*\,;$$

adding this to (4.19) gives

$$G + G^* = \sum (A_j D_j - D_j A_j) + B + B^* = -\sum A_{j,j} + B + B^* \tag{4.22}$$

where $A_{j,j} = D_j A_j$. □

Observe that if expression (4.22) is zero, then we may take M to be zero. This means that the solutions are isometric; i.e., $\|u(t)\|$ is a constant.

Theorem 4.3 implies that if $\|u\|$ is zero at time $t = 0$, then it is zero for all times. This shows that solutions with compact support in x are uniquely determined by their initial values. We shall show now how to modify the above proof to obtain a truly local uniqueness theorem.

Let P_1 and P_2 denote pieces of a pair of hypersurfaces in x, t space that have the same edge; that is, P_1 and P_2 are two smooth imbeddings in x, t space of the

k-dimensional unit ball that agree at the boundary. The domain O bounded by P_1 and P_2 is called a *lens-shaped domain*, P_1 and P_2 its faces.

We shall call a surface in x, t space *weakly spacelike* with respect to the operator (4.18)–(4.19) if at every point its normal $\nu = (\xi, \tau)$ satisfies the following condition: *the matrix*

$$\tau I - \sum \xi_j A_j \tag{4.23}$$

is nonnegative.

THEOREM 4.5 *Let O denote a lens-shaped domain whose face P_2 is weakly spacelike with respect to an operator L. Let u be a solution of $Lu = 0$ that vanishes on P_1; then u vanishes in O.*

PROOF: Let ν denote the *outward* normal to the boundary of O. Then the matrix (4.23) is positive on one of the faces, say P_2 of O, and negative on the other. We introduce a new variable v in place of u,

$$u = e^{\lambda t} v .$$

The equation satisfied by v is

$$Lv + \lambda v = 0 . \tag{4.18'}$$

Take the scalar product of the equation by v; the differentiated terms can be written as

$$v \cdot v_t = \frac{1}{2}(v \cdot v)_t , \quad v \cdot Av_x = \frac{1}{2}(v \cdot Av)_x - \frac{1}{2} v \cdot A_x v .$$

Integrate over O and perform integration by parts; we get the following identity:

$$\begin{aligned} 0 &= \iint_O v \cdot (Lv + \lambda v) dx\, dt \\ &= \frac{1}{2} \int_{P_1 \cup P_2} [v \cdot v\tau - v \cdot \sum A_j v \xi_j] dS \\ &\quad + \iint_O \left[\frac{1}{2} \sum v \cdot A_{j,j} v - v \cdot Bv + \lambda v \cdot v \right] dx\, dt . \end{aligned} \tag{4.24}$$

Since u is assumed to be zero on P_1, so is v, and therefore the surface integral over P_1 on the right side of (4.24) is zero. Since P_2 is weakly spacelike the integrand in the remaining surface integral over P_2 is nonnegative, and if λ is chosen large enough, so is the integrand in the integral over O. This shows that the right side of (4.24) is positive unless v is zero in O. This proves the theorem. □

We show now how to use Theorem 4.5 to find, or at least put bounds on, the domain of influence of points.

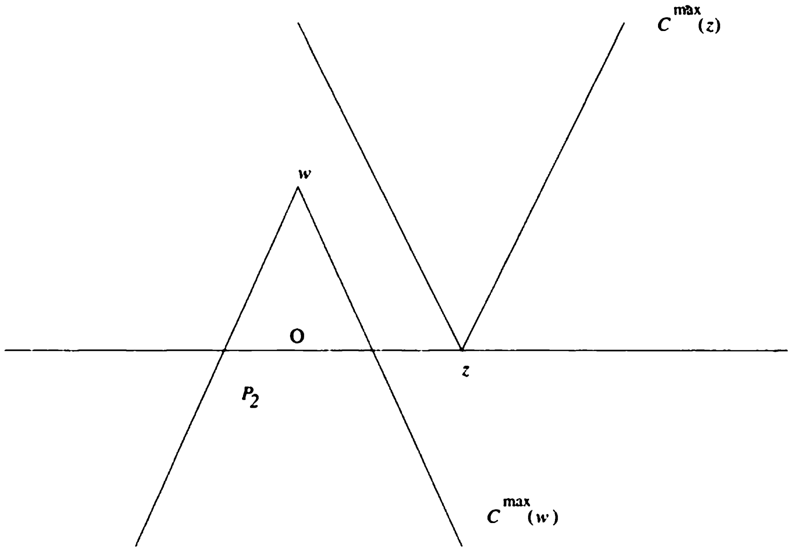


FIGURE 4.1

THEOREM 4.6 *Let $z = (y, s)$ be any point, $C_+^{\max}(z)$ the characteristic conoid corresponding to $\sigma_{\max}$ issuing from z in the direction $t > s$. We claim that all points $t > s$ influenced by z lie inside or on $C_+^{\max}(z)$.*

PROOF: We have to show that no point $w = (x, r), r > s$, outside $C_+^{\max}(z)$ is influenced by z. For this we need the following:

LEMMA 4.7 *If w lies outside the forward characteristic conoid $C_+^{\max}(z)$ through z, then z lies outside the backward characteristic conoid $C_-^{\max}(w)$ through w.*

PROOF: If x' is sufficiently far away from y, z lies outside $C_-^{\max}(x', r)$. Since the exterior of $C_-^{\max}(z)$ is connected, if for some (x, s) in the exterior of $C_+^{\max}(z)$ the conoid $C_-^{\max}(x, r)$ contained in z, there would be a point $w''(x'', r)$ outside $C_+^{\max}(z)$ for which z would lie on $C_-^{\max}(w'')$. $C_-^{\max}(w'')$ consists of rays issuing from w''; this ray would then coincide with one of the rays issuing from z. But then w'' would lie on $C_+^{\max}(z)$, a contradiction. □

To prove Theorem 4.6 we note that the characteristic conoids $C_-^{\max}(w)$ are weakly spacelike. To see this we use equation (4.10): $\tau = \sigma_{\max}(x, t, \xi)$. From this it follows that $\tau I - \sum \xi_j A_j$ is nonnegative. Consider now the lens-shaped domain bounded by $C_-^{\max}(w)$ and that portion P_2 of the hyperplane $t = s$ that is contained inside $C_-^{\max}(w)$. It follows from Theorem 4.5 that if a solution u of equation (4.18) is zero on P_2, then it is zero in O , and in particular $u(w) = 0$. Since z lies outside of P_2, it follows that w is not influenced by z. □

A similar theory can be developed for equations of the form

$$Hu_t - Gu = 0,$$

where G is given by (4.19), and H is a positive symmetric matrix function of x and t.

4.4. Energy Inequalities for Solutions of Second-Order Hyperbolic Equations

In this section we shall derive energy inequalities for solutions of second-order hyperbolic equations

$$u_{tt} - \sum a_{ij} u_{x_i x_j} + b_i u_{x_i} + cu = 0, \tag{4.25}$$

a_{ij} positive definite. Let u be a solution of this equation that has compact support in x. Multiply by u_t and integrate over the slab $x \in \mathbb{R}^k$, $0 \le t \le T$. Integrating the second term by parts with respect to x_j gives

$$\iint \left[u_t u_{tt} + \sum a_{ij} u_{tx_j} u_{x_i} + a_{ij,j} u_t u_{x_i} + b_i u_t u_{x_i} + c u_t u \right] dx\, dt = 0. \tag{4.26}$$

Here $a_{ij,j}$ abbreviates $\frac{\partial}{\partial x_j} a_{ij}$. The first term can be written as a perfect t-derivative:

$$\frac{1}{2}(u_t)_t^2.$$

Since $a_{ij} = a_{ji}$, the second term can be written as

$$\sum \frac{1}{2}(a_{ij} u_{x_i} u_{x_j})_t - \frac{1}{2} a_{ij,t} u_{x_i} u_{x_j}.$$

Substituting these expressions into (4.26) and carrying out the integrations with respect to t gives the following identity:

$$\int q(u) dx \Big|_0^T + \iint Q(u) dx\, dt = 0 \tag{4.27}$$

where

$$q(u) = \frac{1}{2} u_t^2 + \frac{1}{2} \sum a_{ij} u_{x_i} u_{x_j} \tag{4.28}$$

and Q is a quadratic form in the first derivatives of u and u itself.

As shown at the end of Chapter 2, the hyperbolicity of (4.25) is equivalent to the positive definiteness of the quadratic form (4.28). So using the Schwarz inequality we can estimate Q in terms of q:

$$Q \le \text{const}(q + u^2). \tag{4.29}$$

Integrating (4.29) with respect to x gives

$$\int Q\, dx \le \text{const}\left(\int q\, dx + \int u^2\, dx \right). \tag{4.30}$$

Since u is of compact support in x, by a well-known inequality

$$\int u^2\, dx \le \text{const} \int u_x^2\, dx,$$

the constant depending on the size of the support of u. Substituting this into (4.30) gives

$$\int Q\,dx \le M \int q\,dx, \tag{4.31}$$

M some constant.

Let us denote by $E(t)$ the quantity

$$E(t) = \int q(u)dx. \tag{4.32}$$

Using (4.31) we can turn (4.27) into the inequality

$$E(T) \le E(0) + M \int_0^T E(t)dt. \tag{4.33}$$

It is well-known and easy to show that a function which satisfies the integral inequality (4.33) for all T is bounded by

$$E(T) \le e^{MT} E(0). \tag{4.34}$$

Denote by the symbol $\|u\|_1$ the Sobolev norm

$$\|u\|_1^2 = \int u_t^2 + \sum u_{x_j}^2\, dx. \tag{4.35}$$

Since the quadratic form q is positive definite,

$$\text{const}\,\|u(t)\|_1^2 \le E(t) \le \text{const}\,\|u(t)\|_1^2. \tag{4.36}$$

Combining (4.34) and (4.36) yields the following:

THEOREM 4.8 *Every solution of hyperbolic equation* (4.25) *that has compact support in x satisfies the inequality*

$$\|u(t)\|_1 \le e(t)\|u(0)\|_1, \tag{4.37}$$

where $e(t)$ is an exponential function of t.

Theorem 4.8 implies a global uniqueness theorem. Just as in Section 3.2, we may in the proof above replace the slab by any lens-shaped domain whose faces are spacelike and obtain a local uniqueness theorem. Thus we can derive the analogues of Theorem 4.6.

4.5. Energy Inequalities for Higher-Order Hyperbolic Equations

In this section we shall present Leray's beautiful derivation of energy inequalities for solutions of strictly hyperbolic equations of order n, n arbitrary. Let the equation be

$$Lu = 0. \tag{4.38}$$

The derivation will parallel that for the second-order case: we shall multiply (4.38) by Mu, M an operator of order $n-1$, integrate over the slab $0 \le t \le T$, integrate by parts, and pull out boundary integrals that are positive definite.

First we need the following generalization of Green's theorem:

LEMMA 4.9 *Let L and M be a pair of partial differential operators with smooth variable coefficients of order n and n − 1, respectively, and let O be any smoothly bounded domain, ∂O its boundary. Then an identity of the form*

$$\iint\limits_O MuLu\,dV = \int\limits_{\partial O} q(u)dS + \iint\limits_O Q(u)dV \tag{4.39}$$

holds for every smooth function u, where q(u) is a quadratic form in the partial derivatives of u of order n − 1 depending linearly on the components of the normal to ∂O, and Q a quadratic form in the partial derivatives of u of order n − 1 and less.

PROOF: A typical term in $\iint MuLu$ is of the form

$$\iint aD_1\cdots D_{n-1}uD_n\cdots D_{2n-1}u\,.$$

Integrate by parts, transferring alternately the n differentiations on the right factor to the left factor and the $n-1$ differentiations from the left factor to the right factor. We get

$$\iint MuLu = -\iint LuMu + \int q + \iint Q\,,$$

where q and Q are as before. The minus sign occurs because the number of integration by parts, $2n-1$ in all, is odd. This gives the desired formula (4.39). □

REMARK. For n greater than two and for more than two independent variables, the quadratic forms q and Q are not uniquely determined by M and L; their form depends on the order in which the integration by parts is carried out. Nevertheless the values of the integrals

$$\int\limits_{\partial O} q(u)dS \quad\text{and}\quad \iint\limits_O Q(u)dx\,dt \tag{4.40}$$

are uniquely determined by M and L.

It will be convenient in what follows to deal with complex-valued solutions; for these the analogue of Green's identity (4.39) is easily derivable from (4.38) applied to the real and imaginary parts of u separately. We get

$$\operatorname{Re}\iint Mu\overline{Lu}\,dV = \int\limits_{\partial O} q_h(u)dS + \iint Q_h(u)dV\,, \tag{4.39h}$$

where q_h and Q_h are the Hermitian forms induced by the quadratic forms q and Q, i.e., for $u = v + iw$,

$$q_h(u) = q(v) + q(w), \quad \text{etc.}$$

From now on we shall omit the subscripts h.

Take the domain O to be the slab $0 \le t \le T$, and suppose that u is a solution with compact support of $Lu = 0$. Then (4.39h) gives

$$\int q(u)dx\Big|_0^T + \iint Q(u)dx\,dt = 0\,. \tag{4.41}$$

The coefficients of L, M, and q vary smoothly with x and t. Fix any point in the slab and denote by L_0, M_0, and q_0 the operators formed with constant coefficients localized at that point. We take L_0 and M_0 to contain only terms of order n and $n-1$, respectively. The coefficients of q are bilinear functions of the coefficients of the highest-order terms in L and M. Therefore for L_0 and M_0 we can derive Green's identity

$$\text{Re} \iint M_0 u \overline{L_0 u}\, dV = \int q_0(u) dx \Big|_0^T , \tag{4.42}$$

provided that we perform the integration by parts in the same order as we did for L and M.

We are going to evaluate the left side of (4.42) by Fourier transformation in the x-variable. By Parseval's relation

$$\int M_0 u \overline{L_0 u}\, dx = \int \widetilde{M_0 u}\, \overline{\widetilde{L_0 u}}\, d\xi ,$$

where the symbol $\widetilde{\ }$ denotes Fourier transformation. Assume that u is smooth and has compact support in x; then

$$\widetilde{M_0 u} = M_0(i\xi, D_t)\tilde{u} , \quad \widetilde{L_0 u} = L_0(i\xi, D_t)\tilde{u} .$$

Abbreviating $\tilde{u}$ by w we get

$$\iint M_0 u \overline{L_0 u}\, dx\, dt = \iint M_0(i\xi, D_t) w \overline{L_0(i\xi, D_t) w}\, d\xi\, dt . \tag{4.43}$$

By assumption, L is strictly hyperbolic, which means that L_0 is; so L_0 can be factored as

$$L_0(\xi, \tau) = \Pi(\tau - \sigma_j(\xi))$$

and distinct σ_j real for ξ real. So

$$L_0(i\xi, D_t) = \Pi(D_t - i\sigma_j(\xi)) .$$

We shall keep ξ fixed at some real value and not write it out. Express M_0 by Lagrange interpolation at the roots σ_j of L_0. The interpolating polynomials are

$$L_k = \prod_{j \neq k} (D_t - i\sigma_j) ;$$

write

$$M_0 = \sum a_k L_k . \tag{4.44}$$

The coefficients a_k are easily evaluated:

$$a_k = \frac{M_0(\sigma_k)}{L_k(\sigma_k)} .$$

The quantities $L_k(\sigma_k) = \prod_{j \neq k}(\sigma_j - \sigma_k)$ *alternate in sign* as k goes from 1 to n. This proves the following:

LEMMA 4.10 *Suppose that the coefficient of τ^{n-1} in $M_0(\tau)$ is positive; then all the quantities a_k in (4.44) are positive if and only if the roots of $M_0(\tau)$ separate those of $L_0(\tau)$.*

From the definition of L_k it follows that

$$L_0 = (D_t - i\sigma_k)L_k\,. \tag{4.45}$$

Using (4.44) and (4.45) we get the identity

$$M_0 w\overline{L_0 w} = \left(\sum a_k L_k w\right)\overline{L_0 w} = \sum a_k L_k w \overline{(D_t - i\sigma_k)L_k w}\,.$$

Take the real part of both sides; since a_k and σ_k are real, we get

$$\operatorname{Re} M_0 w\overline{L_0 w} = \operatorname{Re}\sum a_k L_k w D_t\overline{L_k w} = \frac{1}{2}\sum a_k D_t|L_k w|^2\,.$$

Integrate over the slab; using (4.43) we get the identity

$$\operatorname{Re}\iint M_0 u\overline{L_0 u}\,dx\,dt = \frac{1}{2}\int \sum a_k|L_k w|^2 d\xi\Big|_0^T. \tag{4.46}$$

Compare (4.42) and (4.46); the left sides are identical and the right sides of both consist of the difference of two quantities which only depend on the values of u and its derivatives at $t = T$ and 0, respectively. From this it follows that

$$\int q_0(u)dx = \frac{1}{2}\int \sum a_k|L_k w|^2\,d\xi\,. \tag{4.47}$$

Suppose now that we can find M so that for every real ξ the roots $M_0(\xi, \tau)$ separate those of $L_0(\xi, \tau)$; then according to Lemma 4.10 the quantities a_k are positive. Since the interpolating polynomials L_k form a base for all polynomials of degree $n-1$, for every real ξ the integrand on the right in (4.47) is positive definite:

$$\frac{1}{2}\sum a_k|L_k w|^2 \geq \sum_0^{n-1} c_\ell|D_t^\ell w|^2,$$

where c_ℓ are positive. It is easy to show that c_ℓ is homogeneous of degree $2(n - \ell - 1)$ in ξ; so there exists a positive constant c independent of ξ so that

$$\frac{1}{2}\sum a_k|L_k w|^2 \geq c\sum |\xi|^{2(n-\ell-1)}|D^\ell w|^2\,.$$

Integrate this with respect to ξ; for the left side we have the identity (4.47); the right side can be evaluated by Parseval's formula in terms of the square integrals of the derivatives of u, the Fourier inverse of w. The resulting inequality is

$$\int q_0(u)dx \geq c\int \sum_{|\alpha|=n-1} |D^\alpha u|^2\,dx = c\|u\|_{n-1}^2, \tag{4.48}$$

where D^α denotes any partial differentiation with respect to x and t of order $|\alpha|$, and the quantity $\|u\|_{n-1}$, the $(n-1)$ Sobolev norm of u, is defined by (4.48).

Inequality (4.48) shows that for operators with constant coefficients the boundary integrals in Green's formula are indeed positive provided that the roots of M_0 separate those of L_0. Next we show that for operators with variable coefficients they are nearly positive. We need two lemmas of which the first is merely an application of the Schwarz inequality:

LEMMA 4.11

$$\int q(u)dx \le \text{const}\, \|u\|_{n-1}^2, \tag{4.49}$$

where the value of the constant depends only on the magnitude of the coefficients of q.

The next lemma is due to Gårding:

LEMMA 4.12 *Let L and M be a pair of hyperbolic operators with variable coefficients such that for every x, t in the slab and every real ξ, the roots in τ of $M_0(\xi, \tau)$ separate those of $L_0(\xi, \tau)$. Then there exist two positive constants c and C such that*

$$\int q(u)dx \ge c\|u\|_{n-1}^2 - C\|u\|_0^2, \tag{4.48$'$}$$

where q is the quadratic form associated by Lemma 4.9 with L and M and where

$$\|u\|_0^2 = \int |u(x)|^2 dx .$$

PROOF: We take first the case that the support of u is small; denote by q_0 the localization of q at some point z_0 in the support of u. Write

$$q = q_0 + q_1 .$$

By Lemma 4.11

$$\int q_1(u)dx \le \epsilon\|u\|_{n-1}^2, \tag{4.50}$$

where ϵ is an upper bound for the coefficients of q_1 in the support of u. The coefficients of q_1 are small in the neighborhood of the point z_0; since we have assumed that u vanishes outside this neighborhood, it follows that ϵ in (4.50) is small. Combining (4.48) and (4.50) we get

$$\int q = \int q_0 + \int q_1 \ge c\|u\|_{n-1}^2 - \epsilon\|u\|_{n-1}^2,$$

which is inequality (4.48′) with $c - \epsilon$ in place of c and $C = 0$.

The general case can be reduced to this by a partition of unity. We need the following well-known lemma from calculus:

LEMMA 4.13 *For every smooth function u with compact support there exists a constant a, depending only on the size of the support of u and on n, such that*

$$\|D^j u\| \le a\|u\|_{n-1} \quad \textit{for } |j| < n - 1 . \tag{4.51}$$

Also, given any positive ϵ, no matter how small, there exists a constant b such that

$$\|D^j u\| \le \epsilon\|u\|_{n-1} + b\|u\|_0 \quad \textit{for } |j| < n - 1 . \tag{4.52}$$

Let $\{p_j\}$ be a set of smooth functions with small support such that

$$\sum p_j^2(x) \equiv 1 . \tag{4.53}$$

Define

$$u_j = p_j u ; \tag{4.54}$$

since q is quadratic,

$$q(u_j) = q(p_j u) = p_j^2 q(u) + r_j(q) , \tag{4.55}$$

where r_j is linear in derivatives of u of order $n-1$, and quadratic in derivatives of order less than $n-1$. Summing (4.55) and using (4.53), we get

$$\sum q(u_j) = q(u) + r(u) ,$$

where $r(u) = \sum r_j(u)$. Integrating over x gives

$$\int q(u) = \sum \int q(u_j) + \int r(u) . \tag{4.56}$$

Equation (4.54) shows that each u_j has small support and therefore by our previous derivation (4.48′) holds with $C = 0$:

$$\int q(u_j) \geq c\|u_j\|_{n-1}^2 . \tag{4.57}$$

Estimating $\int r(u)$ by the Schwarz inequality and using (4.52) of Lemma 4.13 gives

$$\left|\int r(u)\right| \leq \epsilon\|u\|_{n-1}^2 + K\|u\|_0^2 , \tag{4.58}$$

K some constant. Putting (4.57) and (4.58) into (4.56) gives

$$\int q(u)dx \geq c\sum \|u_j\|_{n-1}^2 - \epsilon\|u\|_{n-1}^2 - K\|u\|_0^2 . \tag{4.59}$$

Analogously to (4.56) we have

$$\|u\|_{n-1}^2 = \sum \|u_j\|_{n-1}^2 + \int r'(u) .$$

Estimating the second term on the right as in (4.58) and combining the resulting inequality with (4.59) gives

$$\int q(u)dx \geq c\|u\|_{n-1}^2 - 2\epsilon\|u\|_{n-1}^2 - (K + K')\|u\|_0^2 ,$$

which, for ϵ small enough, is the desired inequality (4.48′). This completes the proof of Lemma 4.12. □

Next we show how Lemma 4.12 can be used to derive energy inequalities: Using inequality (4.51) we get for functions u with compact support

$$\left|\iint Q(u)\right| \leq \text{const} \int_0^T \|u(t)\|_{n-1}^2 \, dt . \tag{4.60}$$

Denote

$$\int q(u)dx = E(t).$$

Let u be a solution of $Lu = 0$ with compact support in x. Green's formula (4.41) can be written as

$$E(T) = E(0) - \iint Q(u).$$

Using inequality (4.48′) for $E(T)$, (4.49) for $E(0)$, and (4.60) for $Q(u)$, we get

$$\|u(T)\|_{n-1}^2 \le \text{const}\,\|u(0)\|_{n-1}^2 + \text{const}\int_0^T \|u(t)\|_{n-1}^2\,dt + C\|u(T)\|_0^2. \tag{4.61}$$

On the other hand, by differentiation and application of the Schwarz inequality and (4.51), we get

$$\frac{d}{dt}\|u(t)\|_0^2 = (u, u_t) + (u_t, u) \le 2\|u\|_0\,\|u_t\| \le \text{const}\,\|u\|_{n-1}^2.$$

Integrate from zero to T:

$$\|u(T)\|_0^2 \le \|u(0)\|_0^2 + \text{const}\int_0^T \|u\|_{n-1}^2\,dt. \tag{4.62}$$

Multiply (4.62) by C and add it to (4.61); using the abbreviation

$$F(t) = \|u(t)\|_0^2 + \|u(t)\|_{n-1}^2, \tag{4.63}$$

the resulting inequality can be written as

$$F(T) \le \text{const}\,F(0) + \text{const}\int_0^T F(t)dt.$$

As is well-known, this implies that F satisfies the inequality

$$F(T) \le c(T)F(0), \tag{4.64}$$

$c(T)$ an exponential function of T.

The last piece of information is contained in the following:

LEMMA 4.14 *Given a hyperbolic operator L with variable coefficients, there exist hyperbolic operators M of one order lower such that M_0 separates the roots of L_0 for every real ξ and for every x, t.*

PROOF: Set

$$M(z, \xi, \tau) = \frac{\partial}{\partial \tau} L(z, \xi, \tau).$$

□

REMARK. Differentiating L in any hyperbolic direction produces an M with the desired property. For two space variables, see Appendix B.

Combining Lemma 4.14 with Lemma 4.12 gives the following:

THEOREM 4.15 *Let L be a strictly hyperbolic operator with variable coefficients, and u a solution of $Lu = 0$ with compact support in x. Then energy inequality* (4.64) *is satisfied, where F is defined by* (4.63).

An immediate corollary is that solutions are uniquely determined by their initial data. This is only a global uniqueness theorem, since we have to assume a priori that u has compact support in x. It is no longer clear how to modify the argument in order to obtain a truly local energy estimate and uniqueness theorem since the use of the Fourier transformation needs all of space. In Chapter 6 we shall show how global energy estimates lead to a global existence theorem from which *local* uniqueness follows by a classical argument of Holmgren.

References

Friedrichs, K. O. Symmetric hyperbolic linear differential equations. *Comm. Pure Appl. Math.* 7: 345–392, 1954.

Friedrichs, K. O., and Lewy, H. Über die Eindeutigkeit und das Abhängigkeitsgebiet der Lösungen beim Anfangswertproblem linearer hyperbolischer Differentialgleichungen. *Math. Ann.* 98: 192–204, 1927.

Gårding, L. Solution directe du problème de Cauchy pour les équations hyperboliques. *La thèorie des équations aux dérivées partielles. Nancy, 9-15 avril 1956,* 71–90.

Leray, J. *Hyperbolic differential equations.* The Institute for Advanced Study, Princeton, N.J., 1953, 1955.

CHAPTER 5

Pseudodifferential Operators and Energy Inequalities

This chapter contains a simple and direct derivation due to Calderón of energy inequalities for solutions of any strictly hyperbolic equation. The main tool used is a ring of operators $\mathcal{R}$ that constitutes a natural extension of partial differential operators with variable coefficients.

Let G be a matrix partial differential operator of order m in the space variables with smoothly variable coefficients; its characteristic matrix $g(x,\xi) = c(G)$ is obtained by discarding all terms of order lower than m and replacing D by ξ. Clearly g is a homogeneous function of degree m.

Those properties of partial differential operators that we want to retain for the ring $\mathcal{R}$ can be expressed in terms of the relation of G to g:

(i) If G_1, G_2 are both of degree m, then

$$c(G_1 + G_2) = c(G_1) + c(G_2), \quad c(G_1G_2) = c(G_1)c(G_2).$$

(ii) $c(G^*) = (-1)^m c^*(G)$, where G^* denotes the adjoint of G with respect to the L_2 scalar product, c^* the matrix adjoint.

(iii) The order of $DG - GD$ is $\leq$ order of G.

(iv) The order of $GH \leq$ order of G + order of H.

We turn now to constructing the operators of class $\mathcal{R}$.

Let $\mathcal{K}$ denote the class of $n \times n$ matrix-valued functions $k(x,\xi)$ defined for all real x and ξ, except $\xi = 0$, that have the following properties:

(a) k is homogeneous of degree zero in ξ,

(b) k is independent of x for $|x| > R$, and

(c) k is infinitely differentiable in x, ξ for $\xi \neq 0$.

$\mathcal{K}$ forms a star algebra under pointwise addition and multiplication, and conjugation defined as taking the adjoint of k at each point. We shall associate to each k of $\mathcal{K}$ an operator K denoted as $s(k)$ mapping the space of square-integrable vector-valued functions u with n components into itself, as follows:

(1) If k is independent of ξ, K is multiplication of $u(x)$ by $k(x)$.

(2) If k is independent of x, K is multiplication of the Fourier transform of u by $k(\xi)$:

$$K = F^{-1}k(\xi)F.$$

(3) If $k(x,\xi) = \sum a_j(x)k_j(\xi)$, then we put $K = \sum A_jK_j$, A_j and K_j being defined as in (1) and (2).

It is not hard to show that every function of class $\mathcal{K}$ can be expanded in a series of the above form, and that the corresponding operator series defining K converges.

EXERCISE Show that the definition of K is independent of the expansion used.

Each operator K is associated with a function k in $\mathcal{K}$, called the *symbol* of K. We define the operator Λ as follows:

$$\Lambda = F^{-1}|\xi|F.$$

Clearly

$$\Lambda^2 = F^{-1}|\xi|^2 F = -\Delta.$$

We define the ring $\mathcal{R}$ to consist of all operators of the form $K\Lambda^m$, $K = s(k)$, k in $\mathcal{K}$. Partial differential operators can be expressed in terms of Λ and operators in $\mathcal{R}$:

LEMMA 5.1 *Let G be a matrix partial differential operator homogeneous of order m, g its characteristic matrix, and define*

$$g_h(x,\xi) = \frac{g(x,\xi)}{|\xi|^m}.$$

Then

$$G = i^m G_h \Lambda^m$$

where G_h is that operator whose symbol is g_h.

DEFINITION 5.2 $\mathcal{C}$ is the class of all operators C mapping L_2 into L_2 that are compact, and for which ΛC and $C\Lambda$ are bounded.

The basic theorem is the following:

THEOREM 5.3 *The mapping $s : k \to K$ of $\mathcal{K}$ into a ring of bounded operators defined above is a star isomorphism* mod $\mathcal{C}$, *i.e.,*

(i) $s(k_1 + k_2) = s(k_1) + s(k_2)$, $s(k_1k_2) = s(k_1)s(k_2) + C$, $C \in \mathcal{C}$.
(ii) $s(k^*) = s(k)^* + C$, $C \in \mathcal{C}$.
(iii) *For every K in $\mathcal{K}$, $K\Lambda - \Lambda K$ is bounded.*
(iv) *If K belongs to $\mathcal{K}$ and C to $\mathcal{C}$, then KC also belongs to $\mathcal{C}$.*

COROLLARY 5.4 $s(\Pi k_i) = \Pi s(k_i) + C$, *$C$ in $\mathcal{C}$.*

For proof, see Hörmander or M. Taylor. Calderón has shown how to use operators of class $\mathcal{R}$ to derive energy estimates for solutions of strictly hyperbolic equations. Although the method is applicable in general, for simplicity we shall restrict the discussion to first-order systems, i.e., to systems of the form

$$u_t = Gu = \sum A_j D_j + Bu, \tag{5.1}$$

A_j, B matrices depending on x and t; denote $(x,t) = z$.

DEFINITION 5.5 Equation (5.1) is strictly hyperbolic if and only if $a(\xi, z) = \sum \xi_j A_j(z)$ has real and distinct eigenvalues for all real values of z and ξ, $\xi \neq 0$.

In Section 3.2 we proved the following simple theorem:

THEOREM 5.6 *Let G be an operator in Hilbert space such that $G+G^*$ is bounded; then solutions of*

$$u_t = Gu \tag{5.2}$$

satisfy the energy inequality

$$\|u(T)\| \le e(T)\|u(0)\|, \tag{5.3}$$

e some function of T.

An operator G satisfying the above condition is called *almost antisymmetric* and so is equation (5.2).

In Section 3.2 we took the scalar product of the Hilbert space to be the L_2 scalar product

$$(u, v) = \int u\bar{v}\,dx .$$

For this scalar product, equation (5.1) is almost antisymmetric if and only if the coefficients A_j are symmetric. Thus, in order to be able to use Theorem 5.6 to derive energy inequalities for nonsymmetric first-order equations, we may have to alter the scalar product. Instead of changing the scalar product we shall change the function u:

$$Ku = v . \tag{5.4}$$

This is equivalent to changing the scalar product. The equation satisfied by v is

$$v_t = KGK^{-1}v - K_tK^{-1}v . \tag{5.5}$$

The task is to choose the operator K so that the equation satisfied by v is almost antisymmetric with respect to the L_2 scalar product. To construct a suitable K we need two lemmas. Denote by a the homogenized characteristic matrix of the operator G:

$$a(z, \xi) = \sum \frac{\xi_s}{|\xi|} A_j(z) . \tag{5.6}$$

LEMMA 5.7 *If (5.1) is strictly hyperbolic, there exists a real, symmetric invertible matrix function $k(x, t, \xi)$ of class $\mathcal{K}$, depending smoothly on t, such that*

$$kak^{-1} = s \tag{5.7}$$

is real and symmetric, where a is the symbol defined in (5.6).

Proof of this lemma will be given at the end of this section.

The operator K whose symbol is k constructed above will serve to symmetrize equation (5.1). Note that since k depends differentiably on t, K_t is a bounded operator.

LEMMA 5.8 *Denote by K and K_1 the operators whose symbols are k and k^{-1}. Then KGK_1 is almost antisymmetric with respect to the L_2 scalar product.*

PROOF: According to Lemma 5.1

$$G = iA\Lambda\,,$$

where A is the operator associated with $a(z,\xi)$ defined in (5.6). Then since, according to Theorem 5.3(iii), the commutator of K_1 and Λ is bounded,

$$KGK_1 = iKA\Lambda K_1 = iKAK_1\Lambda + \text{bounded operator.} \tag{5.8}$$

Denote by S the operator whose symbol is s, defined in (5.7). According to the Corollary 5.4 to Theorem 5.3, it follows from (5.7) that

$$KAK_1 = S + C\,, \quad C \text{ in } \mathcal{C}.$$

It follows from the above and (5.8) that

$$KGK_1 = S\Lambda + \text{bounded operator.} \tag{5.9}$$

Since by Lemma 5.7, $s^* = s$, it follows from Theorem 5.3(ii) that

$$S^* = S + C\,, \quad C \text{ in } \mathcal{C}. \tag{5.10}$$

Taking the adjoint of both sides of (5.9) we get, using the fact that $\Lambda^* = -\Lambda$, (5.10), and Theorem 5.3(iii), that

$$\begin{aligned}(KGK_1)^* = -\Lambda S^* + \text{bd. op.} &= -\Lambda S + \text{bd. op.}\\ &= -S\Lambda + \text{bd. op.} = -KGK_1 + \text{bd. op.}\end{aligned}$$

This completes the proof of Lemma 5.8. □

THEOREM 5.9 *Let* (5.1) *be strictly hyperbolic, and u a solution of it in the slab $0 \le x \le T$ with compact support in x. Then u satisfies an energy inequality*

$$\|u(T)\| \le e(T)\|u(0)\|\,,$$

the norm being the L_2 norm.

PROOF: It follows from Theorem 5.3(i) that

$$K_1K = I + C\,, \quad C \text{ in } \mathcal{C}. \tag{5.11}$$

Suppose now that K is invertible; multiplying (5.11) by K^{-1} from the right we get

$$K_1 = K^{-1} + CK\,,$$

so by Theorem 5.3(iv), K_1 differs from K^{-1} by an operator in $\mathcal{C}$. From this it follows that

$$KGK^{-1} = KGK_1 + \text{bd. op.},$$

since according to Lemma 5.8 the operator on the right is almost antisymmetric, so is the operator on the left. Since K_t is bounded, this shows that equation (5.5) satisfied by $v = Ku$ is almost antisymmetric. According to Theorem 5.6 we conclude that v satisfies an energy inequality; does u. if K is invertible, so does u.

If K is not invertible, we claim that we can add an operator in $\mathcal{C}$ to K and make it invertible. To see this we note that since according to Lemma 5.7 k is real and symmetric, it follows from Theorem 5.3(ii) that K differs from the symmetric operators $(K + K^*)/2$ by an operator in $\mathcal{C}$. Clearly, if we replace K by its symmetric part, Lemma 5.8 and (5.11) remain valid. We claim that for the symmetric K the origin is an isolated point of the spectrum of finite multiplicity. For otherwise there would be an orthonormal sequence of elements u_j such that

$$\|Ku_j\| \to 0.$$

By (5.11)

$$(I + C)u_j = K_1 K u_j,$$

so

$$\|u_j + Cu_j\| \to 0. \tag{5.12}$$

The sequence u_j, being orthonormal, tends to zero weakly. Since C is a compact operator, it maps this into a sequence converging to zero strongly. By (5.12) it follows that $\|u_j\| \to 0$, a contradiction.

Once we know that zero is an isolated point of the spectrum of K, we can add to K a degenerate operator M (finite-dimensional range) which belongs to C such that $K + M$ is invertible. This completes the proof of Theorem 5.9. □

There remains to prove Lemma 5.7. Multiply (5.7) by k from both sides; we get

$$k^2 a = ksk. \tag{5.13}$$

Since k and s are supposed to be symmetric, so is the right side of (5.13); since k is invertible, $k^2 = p$ is positive definite. Thus (5.13) can be written as

$$pa = r, \tag{5.14}$$

p positive definite, r symmetric. Conversely, if we can solve (5.14), then $k = p^{1/2}$ solves (5.7).

To solve (5.14) we make use of the basic assumption that (5.1) is strictly hyperbolic, i.e., that at each point the eigenvalues of a are real and distinct. From this it follows that a can be made diagonal, i.e., that there exists a real nonsingular matrix m such that

$$mam^{-1} = d, \quad d \text{ diagonal}.$$

Multiplying this by m^* on the left and m on the right, we get

$$m^*ma = m^*dm,$$

which is (5.14) with $p = m^*m$, $r = m^*dm$.

In this way we can solve (5.14) by smooth functions p and r in the neighborhood of every point of $\mathbb{R}^k \times S^{k-1}$. Let

$$\sum q_j \equiv 1$$

be a partition of unity by smooth *nonnegative* scalar functions q_j with small support. We can find smooth solutions of

$$p_j a = r_j \tag{5.14j}$$

in an open set containing the support of each q_j. Multiplying (5.14j) by q_j and summing gives a solution of (5.14) in the large with

$$p = \sum q_j p_j, \quad r = \sum q_j r_j.$$

As a last step we observe that if p is smooth, so is $k = \sqrt{p}$ provided we take the positive square root. In order for k to belong to the class $\mathcal{K}$ it has to be independent of x for $|x|$ large enough. This will be the case if $a(x, \xi)$ is independent of x for $|x|$ large; this can always be accomplished by altering the differential operator outside the support of the function u for which we are deriving the energy inequality (5.3).

NOTES.

- An alternative way to proceed is to take conclusion (5.7) of Lemma 5.7 as the definition of hyperbolicity of (5.1); in this way we can admit operators with multiple characteristics.
- Petrowsky has used Fourier series to symmetrize hyperbolic equations of degree n. His technique was extremely unwieldy.

References

Calderón, A.-P. Uniqueness in the Cauchy problem for partial differential equations. *Amer. J. Math.* 80: 16–36, 1958.

Hörmander, L. *The analysis of linear partial differential operators*. III. Pseudo-differential operators, chapter 18. Grundlehren der Mathematischen Wissenschaften, 274. Springer, Berlin, 1985.

Kohn, J. J., and Nirenberg, L. An algebra of pseudo-differential operators. *Comm. Pure Appl. Math.* 18: 269–305, 1965.

Petrowsky, I. G. Über das Cauchysche Problem für Systems von partiellen Differentialgleichungen. *Mat. Sb.* 2(44): 815–868, 1937.

Taylor, M. E. *Pseudodifferential operators*. Princeton Mathematical Series, 34. Princeton University, Princeton, N.J., 1981.

CHAPTER 6

Existence of Solutions

In this chapter we shall use energy inequalities to show that the initial value problem for hyperbolic equations has a solution. For the sake of simplicity we shall deal with the simplest case, symmetric first-order systems introduced in Section 4.3:

$$L = D_t - \sum A_j D_x + B\,, \tag{6.1}$$

where A_j and B are C^∞ matrix functions of $x = (x_1, \ldots, x_k)$ and t, and the A_j are symmetric.

THEOREM 6.1 (Main Existence Theorem) *Let s be a C^∞ vector function the slab $B : x$ in $\mathbb{R}^k$, $-T \le t \le T$, f a C^∞ function of x; then the initial value problem*

$$Lu = s\,, \quad u(0, x) = f(x)\,, \tag{6.2}$$

has a C^∞ solution in the slab.

This result will be obtained as a corollary of the following, more precise existence theorem. Let n be a whole number, and define by H_n the Sobolev space of vector functions $v(x, t)$ defined in the above slab B with finite Sobolev norm of order n, where

$$\|v\|_n^2 = \int_B \sum_{|\alpha| \le n} |D^\alpha v|^2 \, dx\, dt\,. \tag{6.3}$$

We denote by $H_n^{(0)}$ the corresponding Sobolev space of vector functions of the space variables alone.

THEOREM 6.2 *Let s be a vector function of class H_n in the slab, and f a vector function of class $H_n^{(0)}$. Then the initial value problem* (6.2) *has a solution u of class H_n in the slab.*

Clearly, by letting n tend to ∞ in Theorem 6.2, we obtain Theorem 6.1.

6.1. Equivalence of the Initial Value Problem and the Periodic Problem

As a first step we alter the coefficients A_j, B, and the data s and f for $|x|$ and $|t|$ large so that they become periodic functions of x and t. Since solutions of a hyperbolic equation depend locally on the data and coefficients, altering them far away will not alter the solution locally.

The periodic problem for the hyperbolic operator L is to find for s in H_n a solution w in H_n of $Lw = s$ that is periodic in x and t. We show now that the periodic problem and the initial value problem are equivalent. Suppose we can solve the initial value problem; denote $S(t)$ the operator that maps the initial values of solutions of $Lu = 0$ into their value at t:

$$S(t) : u(0) \to u(t)\,.$$

According to Theorem 4.3 the operators $S(t)$ are bounded in the $\|u\|_0$ norm. We claim that they are also bounded in the $\|u\|_n$ norms. This can be seen by differentiating the equation $Lu = 0$ with respect to x; the derivatives of u, $D^\alpha u$, $|\alpha| \le n$, satisfy a symmetric hyperbolic system of differential equations, whose solutions can be estimated as in Section 4.3.

The change of variables $u = e^{pt}v$, p a positive number, results in a system of equations for v whose zero-order term has the coefficient $B + pI$, and whose solution operator $S(T)$ has, for p large enough, $\|\cdot\|_n$ norm less than 1. We assume that L has this form.

To solve the periodic problem $Lu = s$, $u(0) = u(T)$, we first solve the initial value problem $Lv = s$, $v(0) = 0$. Let w denote the solution of the initial value problem $Lw = 0$, $w(0) = f$, f yet to be chosen. We set $u = v + w$; clearly $Lu = L(v+w) = Lv + Lw = s$, $u(0) = v(0) + w(0) = f$, and $u(T) = v(T) + S(T)f$. So the periodicity condition is

$$f = v(T) + S(T)f.$$

Since the norm of S is < 1, this has a unique solution f.

Next we show how the solutions of the initial value problem can be obtained from solutions of the periodic problem.

LEMMA 6.3 *Given any f in $H_n^{(0)}$, there exists a g in the domain of $S(T)$ such that $g - S(T)g = f$.*

PROOF: We construct an auxiliary function v in H_n with these properties,

$$v(0) = 0\,, \quad v(T) = f\,, \quad Lv \text{ in } H_n\,,$$

as follows: We determine the first $n + 1$ derivatives of v with respect to t at $t = T$ so that Lv and its first n t-derivatives are zero at $t = T$. From $Lv = v_t - Gv = 0$, we get $v_t(T) = Gf$. From $D_t^j Lv = D_t^{j+1} - D_t^j Gv = 0$, we determine recursively $(D_t^{j+1}v)(T)$, $j \le n$. Then we set

$$v(x,t) = h(t)\left[f + \sum_1^n (D_t^j v)(T)\frac{(t-T)^{j'}}{j!}\right],$$

where $h(t)$ is a C^∞ function such that

$$h(t) = \begin{cases} 1 & \text{for } \frac{2}{3}T < t < T \\ 0 & \text{for } 0 \le t < \frac{1}{3}T. \end{cases}$$

Clearly, v and Lv belong to H_n.

Next we define w as the periodic solution of $Lw = Lv$. Their difference satisfies $L(w - v) = 0$. Denote by g the value of w at $t = 0$ and $t = T$; then $w - v$

at $t = 0$ is g, and at $t = T$ is $g - f$. Therefore

$$S(T)g = g - f\,. \tag{6.4}$$

□

Next we show that the domain of $S(T)$ is dense in $H_n^{(0)}$; for if not, there would be a nonzero f in $H_n^{(0)}$ orthogonal to the domain of $S(T)$. Taking the norm of both sides of (6.4) and using the orthogonality of f and g we get, denoting $\|\cdot\|_n^0$ as $\|\cdot\|$,

$$\|f\|^2 + \|g\|^2 = \|Sg\|^2\,.$$

Since the norm of S is < 1, the right side is $< \|g\|^2$; this implies that $\|g\| = 0$, and so $\|f\| = 0$, a contradiction.

Since the operator $S(T)$ is bounded and densely defined, it follows that its closure is defined on the whole space H_n^0 and is the solution operator for $Lu = 0$.

EXERCISE Use the solution of the periodic problem to solve the inhomogeneous initial value problem (6.2).

6.2. Negative Norms

All functions in this section and the next are periodic in x and t. For any vector function a we define the $-n$ norm as follows:

DEFINITION 6.4 For any vector function a in H_0 we define

$$\|a\|_{-n} = \sup_w \frac{(w, a)_0}{\|w\|_n}\,. \tag{6.5}$$

In words, $\|a\|_{-n}$ is the norm of $\ell(w) = (w, a)_0$ regarded as a linear functional on H_n, where $(\cdot, \cdot)_0$ is the H_0 scalar product.

DEFINITION 6.5 H_{-n} is the completion of H_0 in the $\|\cdot\|_{-n}$ norm.

REMARK. H_{-n} can be identified as a subspace of the space of periodic distributions.

The scalar product $(w, b)_0$ can be defined by closure for every w in H_n and every b in H_{-n}. It follows from (6.5) that

$$(w, b)_0 \le \|w\|_n \|b\|_{-n}\,. \tag{6.6}$$

It follows that for fixed w in H_n, $(w, b)_0$ is a bounded linear functional on H_{-n}.

Define the n-Laplacian operator Δ_n as

$$\Delta_n = \sum_{|\alpha| \le n} (-1)^{|\alpha|} D^{2\alpha}\,.$$

For any C^∞ periodic function u integration by parts gives

$$(\Delta_n u, u)_0 = \sum_{|\alpha| \le n} (D^\alpha u, D^\alpha u)_0 = \|u\|_n^2\,. \tag{6.7}$$

LEMMA 6.6 *Given any periodic distribution b in H_{-n}, there is a unique periodic function c in H_n such that*

$$\Delta_n c = b\,, \tag{6.8}$$

$$\|c\|_n = \|b\|_{-n}\,. \tag{6.9}$$

PROOF: First take b in H_0 and solve equation (6.8) by, say, Fourier series. Taking the H_0 scalar product of equation (6.8) with c gives, using (6.7) and (6.6),

$$\|c\|_n^2 = (\Delta_n c, c)_0 = (c, b)_0 \le \|c\|_n \, \|b\|_{-(n)}$$

from which $\|c\|_n \le \|b\|_{-n}$ follows. Next take the H_0 scalar product of equation (6.8) with any w in H_n and integrate by parts:

$$\begin{aligned}(w, b)_0 = (w, \Delta_n c)_0 &= \sum_{|\alpha|\le n} (D^\alpha w, D^\alpha c)_0 \\ &\le \sum \|D^\alpha w\|_0 \, \|D^\alpha c\|_0 \le \|w\|_n \, \|c\|_n\,,\end{aligned}$$

where in the last two steps we have used the Schwarz inequalities. This shows that $\|c\|_n$ is an upper bound of the linear functional $\ell(w) = (w, b)_0$ in the H_n norm. According to (6.5), the exact upper bound of $\ell(w)$ is $\|b\|_{-n}$. Therefore $\|b\|_{-n} \le \|c\|_n$; since we have already derived the opposite inequality, (6.9) follows.

So far we have assumed that b lies in H_0; any b in H_{-n} can be approximated by a sequence of functions in H_0, and c obtained as a limit in the H_n norm. □

THEOREM 6.7 *Every bounded linear functional ℓ on H_{-n} can be expressed as $\ell(b) = (w, b)_0$, w in H_n.*

PROOF: Inequality (6.6) can be stated in the following words: for any w in H_n the linear functional $\ell(b) = (w, b)_0$ is bounded, and its norm is $\le \|w\|_n$. We shall show now that its norm equals $\|w\|_n$. We take first the case when w lies in H_{2n}; we define b as $\Delta_n w$. By (6.7)

$$\ell(b) = (w, b) = (w, \Delta_n w) = \|w\|_n^2$$

By (6.8) and (6.9), $\|b\|_{-n} = \|w\|_n$. So the identity above implies

$$\ell(b) = \|w\|_n \|b\|_{-n}\,;$$

this shows that the norm of ℓ is $\ge \|w\|_n$. Combined with our previous upper bound we conclude that the norm of ℓ is $\|w\|_n$. For w merely in H_n, we reach the same conclusion by approximating w.

It follows that the space of linear functionals of form $\ell(b) = (w, b)_0$ is a closed linear subspace of the dual of H_{-n}. If it were not all of the dual, there would be a nonzero b in H_{-n} such that $(w, b)_0 = 0$ for all w in H_n. Since H_{-n} is the completion of H_0, for any $\varepsilon > 0$ there is an a in H_0 such that $\|b - a\|_{-n} < \varepsilon$. By (6.5) there is a w in H_n, $\|w\|_n = 1$, such that

$$(w, a)_0 \ge \|a\|_{-n} - \varepsilon\,. \tag{6.10}$$

Using (6.6), (6.10), and $\|b - a\|_{-n} < \varepsilon$, we get

$$(w, b)_0 = (w, a)_0 + (w, b - a)_0 \geq \|a\|_{-n} - \varepsilon - \|b - a\|_{-n}$$
$$\geq \|b\|_{-n} - 3\varepsilon\,.$$

If we choose $\varepsilon < \|b\|_{-n}/3$, it follows that $(w, b)_0 > 0$, contrary to our assumption. It follows that the assumption is wrong, as asserted in Theorem 6.7 □

6.3. Solution of the Periodic Problem

THEOREM 6.8 *Let s be a periodic vector function in H_n. The equation $Lw = s$ has a periodic solution w in H_n.*

PROOF: Denote by L^* the transpose of the operator L. L^* is of the same general form as L; in particular, we can assume that its zero-order term has been augmented by pI, p a positive number as large as we desire.

LEMMA 6.9 *L^* is bounded from below in the sense that for all b in H_{-n+1}*

$$\|L^* b\|_{-n} \geq \text{const}\,\|b\|_{-n} \tag{6.11}$$

with some positive constant.

PROOF: Take first b to be C^∞. According to Lemma 6.6, b can be represented as $b = \Delta_n c$, so that (6.9) holds. Denote $L^* b = a$; then

$$(a, c)_0 = (L^* b, c)_0 = (L^* \Delta_n c, c)_0\,. \tag{6.12}$$

Integrating by parts, using the fact that L^* is of the form $-ID_t + \sum A_j D_j + B + pI$, where the A_j are symmetric matrices, we obtain, see (4.22),

$$(L^* \Delta_n c, c)_0 = Q(c, c) + p(\Delta_n c, c)_0\,, \tag{6.13}$$

where Q is a quadratic form of the derivatives of c up to order n integrated over the period parallelogram. Using (6.7) we deduce that the right side of (6.13) is bounded from below by const $\|c\|_n^2$. The left side of (6.12) is, by (6.6), bounded from above by $\|c\|_n \|a\|_{-n}$. Combining the upper and lower bounds of (6.12) we deduce that

$$\|a\|_{-n} \geq \text{const}\,\|c\|_n.$$

Using (6.9) gives inequality (6.11).

Since every b in H_{-n+1} can be approximated by C^∞ functions, we obtain Lemma 6.9. □

Inequality (6.11) shows that the mapping $b \to L^* b = a$ is a one-to-one mapping of H_{-n+1}, boundedly, into H_{-n}. Given s in H_n, we define on the range of L^* the linear functional $\ell(a)$ as

$$\ell(a) = (s, b)_0\,, \tag{6.14}$$

and extend it boundedly to all a in H_{-n}. According to Theorem 6.7, ℓ can be represented as $\ell(a) = (w, a)_0$, $w \in H_n$; setting this in (6.14) shows that for all b in H_{-n+1}

$$(s, b)_0 = (w, a)_0 = (w, L^* b)_0\,.$$

The right side can be rewritten to give

$$(s, b)_0 = (Lw, b)_0 ,$$

valid for all b in H_{-n+1}; it follows that $s = Lw$, as claimed in Theorem 6.8. □

NOTES.

- In this proof we did not use the fact that the coefficient of D_t is the identity, only that it is a symmetric matrix. So Theorem 6.8 is not really about hyperbolic equations, but about *symmetric positive* operators; their theory is due to Friedrichs.
- The theory of distribution uses the duality of C^∞ and the space of distributions; here we have used the duality of H_n and H_{-n}.

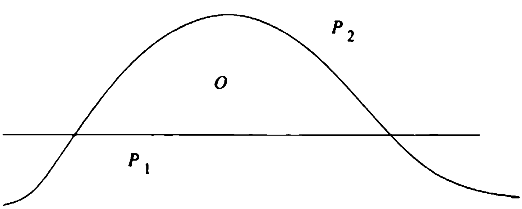
This streamlined existence proof is from the author's 1955 paper. Other proofs based on a priori L^2 inequalities can be found in Friedrich's paper. A much earlier proof based on a priori estimates was given by Schauder in 1937.

Earlier existence theorems relied on some form of approximation to the Riemann function; see Chapter 7. In particular, Hadamard dealt with the fact that the Riemann function is a distribution by introducing the ingenious concept of the finite part of a seemingly divergent integral.

6.4. A Local Uniqueness Theorem

In this section we use a method of Holmgren to deduce a uniqueness theorem from the existence theorem derived in the previous sections.

We use the notion of a lens-shaped domain O introduced in Section 4.3, bounded by two spacelike hypersurfaces P_1 and P_2. We start with a spacelike hypersurface $t = f(x)$, $x \in \mathbb{R}^k$, where the function $f(x)$ is ≥ 0 on a compact subset P_1 of $\mathbb{R}^k$, and < 0 on the rest of $\mathbb{R}^k$; P_2 is that portion of the hypersurface that lies above P_1; O is the domain contained between P_1 and P_2 as in the figure below:



THEOREM 6.10 *Denote by u a solution in O of $Lu = 0$ that equals 0 in P_1; then $u = 0$ on P_2.*

PROOF: Denote by L^* the adjoint of $L = D_t - \sum A_j D_x + B$. By Green's theorem, for any differentiable function v,

$$(6.15)\qquad \iint\limits_{O} (Lu)v\,dx\,dt = \iint\limits_{O} uL^*v\,dx\,dt + \int\limits_{P_2} (n_t I - \sum n_j A_j)u \cdot v\,dS - \int\limits_{P_1} u \cdot v\,dx\,,$$

where n_t and n_j are the components of the normal to P_2.

By assumption $Lu = 0$ in O and $u = 0$ on P_1; if we choose v to satisfy $L^*v = 0$, (6.15) becomes

$$(6.16)\qquad \int\limits_{P_2} (n_t I - \sum n_j A_j)u \cdot v\,dS = 0\,.$$

Choose the initial values of v on the spacelike surface to be equal to $(n_t I - \sum n_j A_j)u$ on P_2, and extend it to the rest of the surface to satisfy the periodicity condition imposed in Section 6.1. For this choice we deduce from (6.16) that $(n_t I - \sum n_j A_j)u = 0$ on P_2. Since P_2 is spacelike, the matrix $n_t I - \sum n_j A_j$ is invertible; therefore $u = 0$ on P_2, as claimed. □

NOTE. Theorem 6.10 can be used to give estimates on the domain of dependence and influence of points. In particular, we can derive the analogue of Theorem 4.6 for first-order hyperbolic systems that are not necessarily symmetric.

References

Friedrichs, K. O. Symmetric hyperbolic linear differential equations. *Comm. Pure Appl. Math.* 7: 345–392, 1954.

Lax, P. D. On Cauchy's problem for hyperbolic equations and the differentiability of solutions of elliptic equations. *Comm. Pure Appl. Math.* 8: 615–633, 1955.

Schauder, J. Cauchy'sches Problem für partielle Differentialgleichungen erster Ordnung. Anwendung einiger sich auf die Absolutbeträge der Lösungen beziehenden Abschätzungen. *Comment. Math. Helv.* 9: 263–283, 1936/37.

CHAPTER 7

Waves and Rays

Introduction

In Section 7.1 we shall show that the initial value problem for distribution initial data has a unique distribution solution. In Section 7.2 we show how to approximate smooth solutions by so-called progressive waves. In Section 7.3 we show that progressive waves make sense for solutions with distribution initial data, and in Section 7.4 we apply these notions to study the propagation of singularities.

The rest of this introduction is a brief review of the relevant portions of the theory of distributions.

Square brackets, parentheses, and curly brackets denote the L^2 scalar product in the slab $\mathbb{R}^k \times (-T, T)$, $\mathbb{R}^k$, and $(-T, T)$, respectively,

$$[u, v] = \iint u(x, t)\bar{v}(x, t)dx\,dt\,,$$

$$(u, v) = \int u(x)\bar{v}(x)dx\,,$$

$$\{u, v\} = \int u(t)\bar{v}(t)dt\,.$$

We denote by $\mathcal{D}$, $\mathcal{D}_x$, and $\mathcal{D}_t$, respectively, the space of C^∞ functions with compact support of x, t, of x, and of t, respectively, in the slab $\mathbb{R}^k \times (-T, T)$. We denote by $\mathcal{E}$ the space of C^∞ functions of x, t in the closed slab $-T \le t \le T$ that have bounded support.

The duals of these spaces are denoted by primes, and the associated bilinear form by square, round, and curly brackets, respectively. We shall refer to elements of $\mathcal{D}'$, $\mathcal{D}'_x$, and $\mathcal{D}'_t$ as distributions in the slab, space, and time, respectively.

Just as any uniformly continuous function defined in a slab can be regarded as a continuous function of t whose values are continuous functions of x, so every distribution defined in a slab can be regarded as a distribution in t whose values are distributions in x. More precisely, we make the following definition:

DEFINITION 7.1 For any d in $\mathcal{D}'$ and u in $\mathcal{E}$, $m = (u, d)$ can be defined as a distribution in $\mathcal{D}'_t$ by setting $\{p, m\} = [pu, d]$ for every p in $\mathcal{D}_t$.

EXERCISE Verify that as defined above, m is a distribution in $\mathcal{D}'_t$.

LEMMA 7.2

(i) *Let d be any distribution in $\mathcal{D}'$, and u any function in $\mathcal{E}$. Then*

$$D_t(u, d) = (D_t u, d) + (u, D_t d)\,. \tag{7.1}$$

(ii) *Let G be any partial differential operator with respect to the x-variables whose coefficients are C^∞ functions, G^* its adjoint. Then*

$$(Gu, d) = (u, G^*d). \tag{7.2}$$

EXERCISE Prove Lemma 7.2.

LEMMA 7.3 *Let d be a distribution in $\mathcal{D}'$ that satisfies a partial differential equation of the form $D_t d - Gd = 0$ in the slab $\mathbb{R}^k \times (-T, T)$, where G is a linear partial differential operator with respect to x, with C^∞ coefficients. Then d is a C^∞ function of t in the sense that for every u in $\mathcal{E}$, $m = (u, d)$ is a C^∞ function of t.*

PROOF: Since d is in $\mathcal{D}'$, for every R and $S < T$ there is an integer N and a constant such that

$$|[v, d]| \leq \text{const}\, |v|_N \tag{7.3}$$

for every v in $\mathcal{D}$ whose support is contained in $|x| \leq R$, $|t| \leq S$. Here $|v|_N = \max_{|\alpha| \leq N} |D^\alpha v|$.

Take any p in $\mathcal{D}_t$ whose support is contained in $[-S, S]$. We shall show that for $m = (u, d)$,

$$|\{p, m\}| \leq \text{const}\, |p|_0, \tag{7.4}$$

where the constant depends on S and on the function u. First we take the special case when p satisfies the following N linear conditions:

$$\int t^j p\, dt = 0, \quad j = 0, 1, \ldots, N-1. \tag{7.5}$$

Then we may represent p as $D_t^N q$, q a C^∞ function whose support lies in $[-S, S]$. We write

$$\begin{aligned} \{p, m\} &= [pu, d] \equiv [p, ud] = [D_t^N q, ud] = (-1)^N [q, D_t^N (ud)] \\ &= \sum (-1)^N \binom{N}{\ell} \left[q, (D_t^{N-\ell} u)(D_t^\ell d)\right]. \end{aligned} \tag{7.6}$$

Next we make use of the fact that d satisfies $D_t d = Gd$; therefore $D_t^\ell d = G^\ell d$, and we rewrite the right side of (7.6) as

$$(-1)^N \sum \binom{N}{\ell} \left[q D_t^{N-\ell} u, G^\ell d\right] = (-1)^N \sum \binom{N}{\ell} \left[q G^{*\ell} D_t^{N-\ell} u, d\right].$$

Now we apply inequality (7.3) with $v = qG^{*\ell} D_t^{N-\ell} u$ to obtain

$$|\{p, m\}| \leq \text{const}\, |q|_N$$

where the constant depends on $|u|_{2N}$. Since $D_t^N q = p$, $|q|_N \leq \text{const}\, |p|_0$, this proves inequality (7.4) for all p that satisfy (7.5). For p in general we construct a function r in $\mathcal{D}_t$ whose support lies in $[-S, S]$ so that

$$\int t^j (p - r) dt = 0, \quad |r|_N \leq \text{const}\, |p|_0.$$

We then set $p = D_t^N q + r$ and proceed as before.

The same technique can be used to derive the estimate

$$|\{p, D_t^i m\}| \leq \text{const}\,|p|_0 \tag{7.4$_i$}$$

for any positive integer i. From this inequality we conclude, using the Riesz representation theorem, that $D_t^i m$ is a signed measure. Since this holds for all i, m is a C^∞ function. □

REMARK. The partial differential equation that d is required to satisfy in Lemma 7.3 need not be hyperbolic.

7.1. The Initial Value Problem for Distributions

We shall study hyperbolic first-order operators of the form

$$L = D_t - G = D_t - \sum A_j D_j - B\,, \tag{7.7}$$

whose coefficient matrices A_j and B are C^∞ functions of x and t.

THEOREM 7.4 *The initial value problem*

$$L^* d = -D_t d - G^* d = 0\,, \quad d(0) = \ell\,, \tag{7.8}$$

has a unique distribution solution d for every prescribed initial value ℓ in $\mathcal{D}'_x$.

PROOF: The statement makes sense, since according to Lemma 7.3 every distribution d that satisfies a partial differential equation of form (7.8) is a C^∞ function of t with values in $\mathcal{D}'_x$.

Next we show the following:

LEMMA 7.5 *Let d be a distribution solution of $L^* d = 0$, and u a C^∞ solution of $Lu = 0$. Then (u, d) is independent of t.*

PROOF: Differentiate (u, d) with respect to t; using relations (7.1) and (7.2) as well as (7.7) and (7.8) we get

$$D_t(u, d) = (D_t u, d) + (u, D_t d) = (Gu, d) - (u, G^* d) = 0\,.$$

□

We can use Lemma 7.5 to define the solution $d(s)$ of equation (7.8) as follows: For every f in D_x we set

$$(f, d(s)) = (u(0), \ell)\,, \tag{7.9}$$

where u is the solution of $Lu = 0$ whose value at time s is f: $u(s) = f$. We shall show that $d(s)$ as defined by (7.9) is a distribution. Clearly, (7.9) is a linear functional of f, since $u(0)$ depends linearly on f. In addition, it is a continuous functional of f, since $u(0)$ depends continuously on f in all the H_n norms, as shown in Chapter 6. But then it follows by Sobolev's inequality that $u(0)$ depends continuously on f in the C^n norm for functions f whose support is contained in some fixed compact set S.

LEMMA 7.6 *$(f, d(s))$ as defined by (7.9) is a Lipschitz continuous function of s.*

PROOF: Take any value $r \neq s$, and denote by v the solution of $Lv = 0$ whose value at r is f: $v(r) = f$. Then by (7.9)

$$(f, d(r)) = (v(0), \ell),$$

so

$$(f, d(s)) - (f, d(r)) = (u(0) - v(0), \ell). \tag{7.10}$$

We have seen that the H_{n+1} norm of a solution v can be estimated in terms of the H_{n+1} of its initial value at $t = r$:

$$\|v\|_{n+1} \leq \text{const} \|f\|_{n+1}.$$

We write

$$v(s) - v(r) = \int_r^s D_t v \, dt,$$

so we can estimate

$$\|v(s) - v(r)\|_n \leq |r - s| \|v\|_{n+1} \leq \text{const} |r - s| \|f\|_{n+1}.$$

Since $v(r) = f$, we deduce that

$$\|v(s) - f\|_n \leq \text{const} |r - s| \|f\|_{n+1}. \tag{7.11}$$

The function $v - u$ satisfies the equation $L(v - u) = 0$, so we can estimate its value at $t = 0$ in terms of its value of $t = s$:

$$\begin{aligned} \|v(0) - u(0)\|_n &\leq \text{const} \|v(s) - u(s)\|_n \\ &= \text{const} \|v(s) - f\|_n \leq \text{const} |r - s| \|f\|_{n+1}, \end{aligned} \tag{7.12}$$

where in the last step we have used the inequality (7.11).

The support of f lies in some compact set in $\mathbb{R}^k$; therefore, since signals propagate with finite speed, $u(0)$ and $v(0)$ are supported in compact sets. Since ℓ is a distribution in $\mathcal{D}_x$, it follows that for g in $\mathcal{D}_x$ supported in some compact set,

$$|(g, \ell)| \leq \text{const} |g|_j$$

for some positive integer j. By Sobolev's inequality $|g|_j \leq \text{const} \|g\|_n$ for $n > j + k/2$; therefore

$$|(g, \ell)| \leq \text{const} \|g\|_n. \tag{7.13}$$

We choose now $g = u(0) - v(0)$; using (7.10), (7.12), and (7.13) we get

$$|(f, d(s)) - (f, d(r))| \leq \text{const} \|u(0) - v(0)\|_n \leq \text{const} |r - s| \|f\|_{n+1}.$$

This proves the Lipschitz continuity of $(f, d(s))$. □

It follows from Lemma 7.6 that $d(s)$ as defined by (7.9) is a distribution in $\mathcal{D}'$; by (7.9), its value at $s = 0$ is ℓ. We shall show now that d satisfies $L^*d = 0$. We rewrite equation (7.9) as

$$(u(s), d(s)) = (u(0), d(0)),$$

valid for all solutions of $Lu = 0$. Now we apply the *converse* of Lemma 7.5: if d is a distribution in $\mathcal{D}$ such that $(u(s), d(s))$ is independent of s for all C^∞ solutions of

$Lu = 0$, then d satisfies $L^*d = 0$. The proof is the same as the proof of Lemma 7.5 run backwards.

This completes the proof of the existence theorem, Theorem 7.4. □

EXERCISES

(1) Show that if $\{\ell_j\}$ is a sequence of distributions in $\mathcal{D}'_x$ that converge to the distribution ℓ, then the solutions d_j of $L^*d_j = 0$ with initial value ℓ_j tend to d, the solution of $L^*d = 0$ with initial value ℓ.

(2) Show that signals carried by distribution solutions of a hyperbolic equation propagate with finite speed.

NOTE. For a *matrix*-valued distribution M we define (u, M) as a distribution in t whose values are row-vector-valued, the j^{th} component being (u, c_j), where c_j is the j^{th} column of M. We define L^*M by letting L^* act on each column vector of M.

DEFINITION 7.7 The *Riemann function*—more precisely, Riemann distribution—for the hyperbolic operator L is the matrix distribution $R(x, t; y, s)$ that satisfies in the x, t-variables $L^*R = 0$, and has prescribed data $\delta(x - y)I$ at time s, where I is the identity matrix,

$$R(x, s; y, s) = \delta(x - y)I\,. \tag{7.14}$$

THEOREM 7.8

(i) *For any C^∞ solution of $Lu = 0$, and for any t,*

$$u'(y, s) = \int u(x, t) \cdot R(x, t; y, s)dx\,, \tag{7.15}$$

where the integration is taken in the sense of distributions, and u' denotes the transpose of u.

(ii) *Each row r_i of R satisfies in the variables y, s the equation*

$$L'r_i' = 0\,.$$

PROOF:

(i) It follows from Lemma 7.5 that for each column c_j of R, $(u(t), c_j(t))$ is independent of t. By (7.14), all components except the j^{th} of $c_j(s)$ are zero, and the j^{th} is $\delta(x - y)$. So $(u(t), c_j(t)) = (u(s), c_j(s)) = u_j(y, s)$, the j^{th} component of $u(y, s)$. This proves (7.15)

(ii) Formally, apply the operator $L'_{y,s}$ to (7.15); the left side is zero. Since for any fixed time t, $u(x, t)$ can be prescribed arbitrarily, it follows that $L'_{y,s}$ annihilates R, which is the same as $L'_{y,s}$ annihilating the rows of R.

To make this formal argument rigorous, take any C^∞ test function $w'(y, s)$ of compact support, multiply (7.15) by L'^*w, and integrate with respect to y and s. The left side is zero, and we argue as before. □

Of course, the usefulness of this explicit expression depends on what we know about the Riemann function; its mere existence tells us nothing. The next two sections will be devoted to this task.

7.2. Progressing Waves

In Section 3.5 we showed that expressing the initial value in terms of its Radon transform leads to an expression of the corresponding solution of a hyperbolic differential equation with constant coefficients as an integral of plane waves, that is, functions of the form $h(x \cdot \omega + \sigma t, \omega)r$, where $\sigma(\omega)$ is an eigenvalue of $C(\omega) = \sum \omega_j A_j$, r the corresponding eigenvector, and ω a unit vector in $\mathbb{R}^k$. When the coefficients A_j are functions of x and t, we replace plane waves by *progressing waves*, which are sums of terms of the form

$$h(\varphi(x, t))v(x, t), \tag{7.16}$$

where h is an arbitrary function, φ a scalar function, called the *phase*, satisfying the *eikonal equation* described below in (7.18), and v a right eigenvector.

Set hv given by (7.16) into the equation $L(hv) = 0$, L defined by (7.7):

$$L(hv) = h'[\varphi_t I - C(D\varphi)]v + hLv, \tag{7.17}$$

where $C(\omega) = C(\omega, x, t) = \sum \omega_j A_j(x, t)$, $D\varphi$ the x-gradient of φ, and $h' = D_s h(s)$. We choose φ and v so that the first term on the right in (7.17) is zero. To avoid the trivial case $v = 0$, φ_t is required to be an eigenvalue $\sigma = \sigma(D\varphi, x, t)$ of $C(D\varphi, x, t)$:

$$D_t\varphi = \sigma(D\varphi, x, t), \tag{7.18}$$

and $v = ar$, where r is the corresponding normalized eigenvector of $C(D\varphi, x, t)$, and a some scalar-valued function called the *amplitude*, which will be determined in the next step.

It follows from (7.17) and (7.18) that the function hv satisfies the inhomogeneous equation

$$L(hv) = hLv.$$

We shall add to it a function of the form $h_1(\varphi)v_1$ chosen so that it approximates a solution of $L(h_1v_1) = -hLv$. Analogously to (7.17),

$$L(h_1v_1) = h_1'(\varphi_t I - C(D\varphi)v_1 + h_1Lv_1.$$

We choose h_1 as an antiderivative of h, $h_1' = h$, and v_1 so that

$$(\varphi_t I - C(D\varphi))v_1 = -Lv. \tag{7.19}$$

Since the matrix $\varphi_t I - C(D\varphi)$ is not invertible, equation (7.19) has a solution only if the right side is orthogonal to the left null vector ℓ of $\varphi_t I - C(D\varphi)$, that is, $\ell Lv = 0$. Since v is of form ar,

$$\ell Lv = \ell r a_t - \sum \ell A_j r a_{x_j} + a\ell Lr.$$

Normalizing the left eigenvector so that $\ell r = 1$ we can write the above equation as

$$\frac{d}{dt}a + ca = 0, \tag{7.20}$$

where

$$\frac{d}{dt} = D_t - \ell A_j r D_j \quad \text{and} \quad c = \ell Lr. \tag{7.21}$$

The initial value of a is available to be chosen appropriately.

The solution v_1 of (7.19) is determined only modulo $a_1 r$; the function a_1 will be determined at the next step of approximation.

We construct recursively a sequence of correction terms, all of the form $h_j(\varphi)v_j$, where the functions h_j satisfy $h_j' = h_{j-1}$, and the v_j satisfy

$$[\varphi_t I - C(D\varphi)]v_j = -Lv_{j-1}\,. \tag{7.19$_j$}$$

This determines v_j modulo a multiple of the right eigenvector $a_j r$; the compatibility condition for solving (7.19$_j$) gives an inhomogeneous differential equation for a_{j-1} analogous to (7.20).

The partial sum $u_N = \sum_0^N h_j v_j$ satisfies the equation

$$Lu_N = h_N L v_N\,. \tag{7.22}$$

We recall from Section 4.2 how to solve the eikonal equation (7.18). Differentiate (7.18) with respect to x_j, denoting $D_j\varphi$ by φ_j we get

$$D_t\varphi_j = \sigma_{x_j} + \sum \sigma_{\varphi_k} D_k\varphi_j\,.$$

This can be rewritten as

$$\dot{\varphi}_j = \sigma_{x_j} \tag{7.23}$$

where $\dot{}$ is differentiation with respect to t along the curve

$$\dot{x}_j = -\sigma_{\varphi_j}\,. \tag{7.23$'$}$$

Note that (7.23)–(7.23′) constitute a Hamiltonian system. A solution of it is called a *bicharacteristic*. The projection of a bicharacteristic into $\mathbb{R}^k$ is called a *characteristic ray*, or just a *ray*.

From the definition of $\sigma(\omega)$ as an eigenvalue of $C(\omega) = \sum \omega_j A_j$, it follows that σ is a homogeneous function of ω of degree 1. Therefore Euler's relation holds:

$$\sigma(\omega) = \sum \omega_k \sigma_{\omega_k}\,.$$

Setting this into the eikonal equation (7.18) and setting $\omega = D\varphi$, we get

$$D_t\varphi = \sum \varphi_k \sigma_{\varphi_k}\,.$$

Using the dot notation (7.23′), we can rewrite this as

$$\dot{\varphi} = 0\,.$$

In words: *the phase φ is constant along rays.*

We have now the solution of the initial value problem for the eikonal equation in hand.

THEOREM 7.9 *Let $\varphi_0(x)$ be a C^∞ function in $\mathbb{R}^k$ whose first derivatives are bounded by M. The eikonal equation* (7.18) *has a unique C^∞ solution φ in the slab $-T < t < T$ whose initial value is φ_0. The value of T depends on M.*

PROOF: Prescribing φ_0 provides the initial data for φ_j required to solve the Hamiltonian system (7.23)–(7.23′). For T not too large the rays cover the slab $\mathbb{R}^k \times (-T, T)$ in a one-to-one fashion. Take φ to be constant along rays; this defines φ in the slab as a solution of the eikonal equation with the prescribed initial value. □

NOTE. The solution is *local* in time; in general, singularities, called *caustics*, develop after a finite time.

Equation (7.20) is a differential equation for the amplitude along a curve defined by

$$\dot{x}_j = -\ell A_j r\,. \tag{7.24}$$

We show now that this curve is a ray, that is, the same as defined by (7.23′). To see this, take the eigenvalue equation

$$C(\omega)r = \Big(\sum \omega_k A_k\Big)r = \sigma r$$

and differentiate it with respect to ω_j:

$$A_j r + C(\omega) r_{\omega_j} = \sigma_{\omega_j} r + \sigma r_{\omega_j}\,.$$

Multiply on the left by the left eigenvector ℓ:

$$\ell A_j r + \ell C(\omega) r_{\omega_j} = \sigma_{\omega_j} \ell r + \sigma \ell r_{\omega_j}\,.$$

Since $\ell C(\omega) = \sigma \ell$, and $\ell r = 1$ by normalization, we deduce that

$$\ell A_j r = \sigma_{\omega_j}\,.$$

Setting $\omega = D\varphi$, we see that the right sides of (7.24) and (7.23′) are equal.

Finally, we show how to use progressing waves to solve approximately initial value problems. We saw in Section 3.5 that all initial functions can be represented as a superposition of functions of the form $g(x \cdot \omega)$, ω a unit vector. Accordingly, we choose the initial value of the phase φ to be $x \cdot \omega$. We then decompose $g(x \cdot \omega)$ according to the right eigenvectors $r_j(x, \omega)$ of $C(\omega, x, 0)$:

$$g(x \cdot \omega) = \sum g_j r_j\,,$$

where $g_j = \ell_j g = \sum \ell_j^{(m)} g^{(m)}$, ℓ_j the normalized left eigenvectors, and where the superscript m denotes the m^{th} component. The first term of the progressing wave, given by (7.16), is at $t = 0$ of the form $h(x \cdot \omega)ar$. In our case we take $h(s) = g^{(m)}(s)$ and choose $a(0) = \ell_j^{(m)}(x, \omega)$ to be the initial value of a; then we sum the waves over all j and m.

The second and subsequent terms of the progressing waves are formed as follows: We subtract from $h_1 v_1$ traveling waves of the form $h_1 \sum b_j r_j$ and choose the initial values of b_j so that $\sum b_j r_j$ equals the initial value of v_1.

The last step, after taking the sum of all these progressing waves, is to integrate this sum with respect to ω over the unit sphere. The result is an approximate solution u of $Lu = 0$ with the prescribed initial values. We make this statement more precise.

THEOREM 7.10 *Given initial data f in class C^M and with compact support, we can construct using progressing waves, an approximate solution u of class C^K, $u(0) = f$, for which Lu belongs to C^K, K arbitrary, M sufficiently large.*

We shall not present a direct proof of this proposition; rather we shall show how to use progressive waves to construct an approximation to the Riemann function. To prepare the way, we present in the next section a discussion of compound distributions and their integrals.

7.3. Integrals of Compound Distributions

We shall confine the discussion to the situation at hand, where the variables x, t lie in a slab B, and the parameter ω is a point of the unit sphere $\mathbb{S}$.

DEFINITION 7.11 Let m be a distribution in a single variable, and φ a C^∞ scalar-valued function, defined in the slab B, whose gradient is nowhere zero. The *compound distribution* $m(\varphi)$ is defined as follows:

Let c be a C^∞ function with compact support in B. If the support of c is small enough, it is possible to introduce coordinates $w_1, \dots, w_{k+1}$ in an open set containing the support of c, where $w_1 = \varphi$. Then we set

$$[c, m(\varphi)] = \int c(z(w))m(w_1)J(w)dw_1 \cdots dw_{k+1}, \tag{7.25}$$

where $J(w)$ is the Jacobian determinant of the mapping $w \leftarrow z = (x, t)$. The dw_1 integration is taken in the sense of distributions, all the others in the ordinary sense.

When the support of c is not small, c can be decomposed by using a sufficiently fine partition of unity, as $c = \sum c_j$, where each c_j has small support. We then define $[c, m(\varphi)]$ as $\sum [c_j, m(\varphi)]$.

EXERCISES

(1) Show that the definition (7.25) is independent of the particular mapping $w \to z$ employed.
(2) Show that the definition of $[c, m(\varphi)]$ is independent of the partition of unity employed to decompose c.

Compound distributions have the customary properties:

LEMMA 7.12

(i) *The chain rule holds: $Dm(\varphi) = m'(\varphi)D\varphi$.*
(ii) *If m is C^∞ in a neighborhood of $\varphi(z_0)$, then $m(\varphi)$ is C^∞ in a neighborhood of z_0.*
(iii) *If the x-gradient of $\varphi(x, t)$ is nonzero, then for any u in $\mathcal{D}_x$, $(u, m(\varphi))$ is a C^∞ function of t.*

EXERCISE Prove Lemma 7.12.

We turn now to distribution-valued functions $\ell(\omega)$ defined on the unit sphere $\mathbb{S}$; that is, for each ω in $\mathbb{S}$, $\ell(\omega)$ lies in $\mathcal{D}'(B)$. We call $\ell(\omega)$ *integrable* over $\mathbb{S}$ if for each u in $\mathcal{D}(B)$ the integral

$$\int_{\mathbb{S}} [u, \ell(\omega)] d\omega \tag{7.26}$$

exists and is a continuous linear functional of u. For integrable functions $\ell(\omega)$, the integrated distribution

$$\int_{\mathbb{S}} \ell(\omega) d\omega = \ell$$

is defined by setting $[u, \ell]$ equal to (7.26).

LEMMA 7.13

(i) *Let $\ell(\omega)$ be an integrable distribution-valued function, and L a partial differential operator with C^∞ coefficients. Then $L\ell(\omega)$, too, is an integrable function, and*

$$L \int \ell \, d\omega = \int L\ell \, d\omega .$$

(ii) *If $\omega \to \omega'$ is a C^∞ invertible map, then*

$$\int \ell(\omega) d\omega = \int \ell(\omega) J(\omega') d\omega'.$$

EXERCISE Prove Lemma 7.13.

We shall apply these notions to the terms that arise in the approximate solutions of initial value problems by progressing waves. These are of the form

$$\ell = \int a(z, \omega) m(\varphi(z, \omega)) d\omega , \tag{7.27}$$

where $a(z, w)$ and $\varphi(z, \omega)$ are vector and scalar C^∞ functions of their arguments, the z gradient of φ is nonzero, and m is a distribution of a single argument. It is not hard to show that (7.27) is an integrable compound distribution in $\mathcal{D}'(B)$. Recall that the singular support of a distribution is the smallest closed set outside of which the distribution is C^∞.

THEOREM 7.14 *Let z_0 be a point in B with the following property: the gradient of φ with respect to ω is nonzero at all points ω for which $\varphi(z_0, \omega)$ lies in the singular support K of m. Then the distribution* (7.27) *is C^∞ in a neighborhood of z_0.*

PROOF: Apply a sufficiently fine partition of unity to decompose a as the sum $\sum a_j$ on the unit sphere where the support of each a_j lies in a small ω set S_j. This decomposes the distribution (7.27) as the sum of distribution ℓ_j, each of form (7.27). For each of these distributions there are two cases:

(i) $\varphi(z_0, \omega)$ does not belong to K for any ω in S_j.
(ii) For some ω_j in S_j, $\varphi(z_0, \omega_j)$ lies in K.

In case (i) ℓ_j is obviously a C^∞ function. In case (ii) it follows from the assumption about z_0 that we may for all points in a small neighborhood of z_0 introduce new parameters $\zeta_1, \ldots, \zeta_{k-1}$ in S_j so that $\varphi = \zeta_1$. We then express ℓ_j as an integral with respect to the ζ-variables as

$$\ell_j = \int b(z, \zeta) J(z, \zeta) m(\zeta_1) d\zeta_1 \cdots d\zeta_{k-1}$$

where $b(z, \zeta) = a(z, \omega(\zeta, z))$, and J is the Jacobian of the mapping $\zeta \to \omega$. The integral with respect to ζ_1 is taken in the sense of distributions; the result is a C^∞ function of z and the rest of the ζ. This proves that ℓ_j is a C^∞ function of z. □

We conclude this section with the Radon transform representation of δ in $\mathbb{R}^k$, k odd.

EXERCISES Take in the definition (3.48) of the Radon transform $f = \delta$.

(1) Show that $\tilde{\delta} = \delta$; here δ on the left is the δ-function in $\mathbb{R}^k$, on the right in $\mathbb{R}$. Therefore by formula (3.50) for the inversion of the Radon transform

$$\delta(x) = c \int \delta^{(k-1)}(x \cdot \omega) d\omega, \tag{7.28}$$

where $\delta^{(k-1)}$ denotes the $(k-1)^{\text{st}}$ derivative of δ.

(2) Give a proper proof of formula (7.28).

7.4. An Approximate Riemann Function and the Generalized Huygens Principle

We recall from Section 7.1 the Riemann function R for the hyperbolic operator L as the matrix solution of the hyperbolic equation $L^* R = 0$, whose value at time s is $\delta(x - y)I$, where δ is the delta function in $\mathbb{R}^k$ and I the $n \times n$ identity matrix. Formula (7.28), and the theory of progressing waves developed in Section 7.2, gives an approximation R_N to R that is a sum of terms of the following form:

$$\int a(x, t, \omega) h(\varphi(x, t, \omega)) d\omega, \tag{7.29}$$

where h is a derivative or integral of some order of δ in $\mathbb{R}$, and a is a C^∞ matrix-valued function of its arguments. The calculations in Section 7.2 show that $L^* R_N$ belongs to C^M, M arbitrary, provided that N is taken large enough, and the initial value of R_N is δI.

The exact Riemann function R differs from R_N by a smooth function; this can be seen as follows: denote by E_N the solution of the initial value problem

$$L^* E_N = L^* R_N, \quad E_N(0) = 0.$$

Clearly, $R = R_n - E_N$ satisfies $L^* R = 0$, $R(0) = \delta I$, which characterizes it as the Riemann function. The function E_N can be made as smooth as we wish by taking N large enough. Note, furthermore, that R is zero for x, t outside the domain of influence of (y, s).

THEOREM 7.15 *The Riemann function $R(x, t; y, s)$ is a C^∞ function of (x, t) and (y, s) except when (x, t) and (y, s) are connected by a ray.*

PROOF: Since the correction term E_N can be made C^M by taking N large enough, it suffices to show that R_N is a C^∞ function of (x, t) except on the rays connecting (x, t) and (y, s). We shall show that each term (7.29) contained in R_N has this property. We appeal to Theorem 7.14, according to which such a term is C^∞ at all points (x, t) provided that the gradient of φ with respect to ω is nonzero at all points where $\varphi(x, t, \omega)$ belongs to the singular set of the distribution h. Since h is obtained from the δ distribution by differentiation or integration, these are the points where $\varphi(x, t, \omega) = 0$.

To verify this condition, we calculate the partial derivatives of φ with respect to ω. Denote by g such a partial derivative, $g = \partial\varphi/\partial\omega_j$. Differentiate the eikonal equation (7.18) with respect to ω_j; we get

$$D_t g - \sigma_{\varphi_k} D_{x_k} g = 0 .$$

According to equation (7.23′) this equation says that the derivative of g along a ray is zero; therefore g is constant along this ray. At $t = s$ the phase function was taken to be $(x - y) \cdot \omega$; therefore $g = \partial\varphi/\partial\omega_j = x_j - y_j$. Suppose (x, t) does not lie on any of the rays issuing from (y, s); origin; then at least one of the $x_j - y_j$ is nonzero, which shows that the ω-gradient of φ is nonzero at x, t.

To conclude that the gradient of φ on $|\omega| = 1$ is nonzero, we observe that φ is a homogeneous function of ω of first degree. This shows that the derivative of φ with respect to $|\omega|$ equals φ, which equals 0. Therefore the gradient of φ on $|\omega| = 1$ is not equal to 0.

According to Theorem 7.8, the transpose of $R(x, t; y, s)$ is the Riemann function for L^* in the variables y, s. It follows then from what we have just proven that R is a C^∞ function of (y, s) as well, provided that (y, s) is not connected to (x, t) by a ray. □

We come now to the main result of this chapter.

THEOREM 7.16 *Let u be a distribution solution of the strictly hyperbolic equation $Lu = 0$ with C^∞ coefficients. Let (y, s) be a point with the property that the initial value $f(x) = u(x, 0)$ of u is C^∞ in a neighborhood of all points x that lie on a ray emanating from (y, s); then u is C^∞ in a neighborhood of (y, s).*

PROOF: Decompose the initial data as $f = f_1 + f_2$, where f_1 is zero in an open set containing all points x that lie on a ray emanating from (x_0, t_0), and f_2 is C^∞ everywhere. Then $u = u_1 + u_2$, where u_i is the solution of $Lu_i = 0$, $u_i(x, 0) = f_i$, $i = 1, 2$. Clearly, u_2 is C^∞ everywhere.

To show that u_1 is C^∞ in a neighborhood of y, s, we approximate the distribution f_1 by a sequence of C^∞ functions h_m by setting

$$h_m(x) = \int f_1(y) p(m(x - y)) m^k \, dy ,$$

where $p(y)$ is a C^∞ function of compact support, $\int p \, dy = 1$. Since $f_1 = 0$ in an open set containing all points x that lie on a ray issuing from (x_0, t_0), the same is true for all functions h_m for m large enough. Denote by v_m the solution of

$Lv_m = 0$, $v_m(x,0) = h_m(x)$; clearly v_m tends in the sense of distributions to u_1. We represent v_m in terms of the Riemann function as in (7.15):

$$v_m(y,s) = \int R(x,0;y,s)h_m(x)dx\,. \tag{7.30}$$

According to Theorem 7.15, $R(x,0;y,s)$ is a C^∞ function of x away from the intersection of rays emanating from (y,s) with the initial hyperplane. Since f_1, and all functions h_m are zero in an open set containing this exceptional set, in formula (7.30) we can pass to the limit $m \to \infty$:

$$u(y,s) = \int R(x,0)f_1(x)dx\,. \tag{7.30$'$}$$

We appeal once more to Theorem 7.15 on the differentiability of R with respect to y,s to conclude that u is C^∞ for the set of (y,s) specified in Theorem 7.16. □

According to Theorem 3.16, in an odd number of space dimensions the value at some point (x,t) of a solution of the wave equation doesn't depend on the values of the initial data except at those points where the characteristic rays issuing from (x,t) intersect the initial hyperplane. This property is called, after its discoverer, the Huygens principle. Theorem 7.16 says that the smoothness at (y,s) of solutions of any strictly hyperbolic equation with C^∞ coefficients depends on the smoothness of the initial data only at those points where the characteristic rays issuing from (y,s) intersect the initial hyperplane. Theorem 7.16 is called the *Huygens principle in a generalized sense*.

Representation (7.28) of the δ function holds only for an odd number k of space dimensions; an extension to even k is not hard. More troublesome is that our proof is local in time, for solutions of the eikonal equation are local in time. The extension of Theorem 7.16 to large times is due to Donald Ludwig.

The last result of this section is about the propagation of discontinuities.

THEOREM 7.17 *Let f be initial data that are C^∞ on either side of a C^∞ hypersurface K lying in the initial plane, and f and its partial derivatives have jump discontinuities across K. Then the distribution solution of $Lu = 0$, $u(x,0) = f$, is C^∞ except on the n characteristic hypersurfaces issuing from the discontinuity K; u and its partial derivatives have jump discontinuities across these characteristic surfaces.*

SKETCH OF PROOF: Describe the discontinuity K by an equation $\varphi(x) = 0$, φ a scalar function. Given any integer M, we can decompose f as $f = f_1 + f_2$, where f_1 is a finite sum of the form $a_j(x)m(\varphi(x))$, where a_j is a C^∞ vector function, m is some integral of the δ-function, and f_2 belongs to C^M. We can, using the technique of Section 7.2, construct progressive waves w_j whose initial value is f_1 and which consists of terms of the form

$$a_j(x,t)m(\varphi_j(x,t))\,, \quad j = 1,\ldots,n.$$

Here $\varphi_j(x,t)$ is the solution of the j^{th} eikonal equation with initial value φ. We have shown that φ_j is constant along rays; therefore $\varphi_j(x,t) = 0$ only on the rays issuing from those points x for which $\varphi(x) = 0$, and this set is the hypersurface K along which the discontinuities of the initial data occur. The rays issuing from K form the characteristic hypersurfaces issuing from K. It follows that w is C^∞ except on the characteristic surfaces issuing from K. Since w differs by a C^M function from the solution of $Lu_1 = 0$, $u_1(x, 0) = f_1(x)$, it follows that u_1 is smooth away from the characteristic surfaces. The solution u_2 of $Lu_2 = 0$, $u_2(0, x) = f_2(x)$, is smooth everywhere, so Theorem 7.17 follows. □

REMARK 7.18. In Courant-Lax we have shown how to deduce the generalized Huygens principle from Theorem 7.17, for which we gave a direct proof.

REMARK 7.19. Hörmander has introduced the important notion of wave front set, and has shown that it propagates along rays. This is a further generalization of Theorems 7.16 and 7.17.

REMARK 7.20. There is an important extension due to Melrose and Taylor of the generalized Huygens principle to solutions of mixed initial and boundary value problems; see the expository article by Taylor.

References

Courant, R., and Lax, P. D. The propagation of discontinuities in wave motion. *Proc. Nat. Acad. Sci. U.S.A.* 42: 872–876, 1956.

Hörmander, L. Uniqueness theorems and wave front sets for solutions of linear differential equations with analytic coefficients. *Comm. Pure Appl. Math.* 24: 671–704, 1971.

Lax, P. D. Asymptotic solutions of oscillatory initial value problems. *Duke Math. J.* 24: 627–646, 1957.

Ludwig, D. Exact and asymptotic solutions of the Cauchy problem. *Comm. Pure Appl. Math.* 13: 473–508, 1960.

Taylor, M. E. Propagation, reflection, and diffraction of singularities of solutions to wave equations. *Bull. Amer. Math. Soc.* 84(4): 589–611, 1978.

CHAPTER 8

Finite Difference Approximation to Hyperbolic Equations

8.1. Consistency

We shall discuss first-order symmetric hyperbolic systems

$$D_t u = \sum A_j D_j u \,, \tag{8.1}$$

where $u = u(x, t)$ is a vector variable, and the A_j are real symmetric matrices that depend smoothly on x but not on t. D_j denotes partial differentiation with respect to x_j, $j = 1, \dots, k$.

Problems in physics and engineering often call for numerical approximations of solutions whose initial values are given. The most powerful and most general methods for constructing the needed approximation is to discretize the variables x and t, replace derivatives by difference quotients, and solve the resulting equations in a finite number of variables. We seek to approximate the values of the function u only at points of a lattice in x, t space. Denote by h a multi-index $(h_1, \dots, h_k)$, and denote by n an integer. Denote the spatial and time scales by δ and ϵ, respectively, and denote by u_h^n an approximation to the value of u at the point $x = h\delta$ and time $t = n\epsilon$.

A *difference equation* connects the values of u_h^n at various points of the lattice. Here we shall discuss the simplest schemes, explicit two-level ones; these express u_h^{n+1} in terms of the values of u_h^n at the previous time level. For simplicity we first treat the case when there is only one space variable. Then equation (8.1) takes the form

$$D_t u = A D_x u \,; \tag{8.2}$$

h is a scalar index, and the difference scheme is of the form

$$u_h^{n+1} = \sum C_j u_{h+j}^n \,. \tag{8.3}$$

Here C_j are matrices that are functions of $x = h\delta$; the integer j ranges over a set of a finite number of indices.

To relate the difference equation (8.3) to the differential equation (8.2), and the matrices C_j to the matrix A, we express u_h^{n+1} and u_{h+j}^n in terms of u and its derivatives at $x = h\delta$, $t = n\epsilon$, using Taylor's formula:

$$u_h^{n+1} = u_h^n + D_t u \epsilon + O(\epsilon^2) \,, \tag{8.4}$$

$$u_{h+j}^n = u_h^n + j D_x u \delta + O(\delta^2) \,. \tag{8.5}$$

Setting these into the difference equation (8.3) gives

$$u_h^n + D_t u\epsilon + O(\epsilon^2) = \sum C_j u_h^n + \sum jC_j D_x u\delta + O(\delta^2)\,. \tag{8.6}$$

According to equation (8.2), $D_t u = AD_x u$. Setting this into (8.6) and writing $\epsilon = \lambda\delta$, we deduce the following two *consistency conditions* for the coefficients C_j of the difference scheme:

$$\sum C_j = I\,, \quad \sum jC_j = \lambda A\,. \tag{8.7}$$

If these are satisfied, it follows from (8.6) that

$$\left|u_h^{n+1} - u(h\delta, (n+1)\epsilon)\right| \le O(\epsilon^2 + \delta^2)\,.$$

We shall keep the ratio λ of ϵ to δ constant so that we can write the above estimate as

$$\left|u_h^{n+1} - u(h\delta, (n+1)\epsilon)\right| \le O(\delta^2)\,. \tag{8.8}$$

We now consider the simplest case: only C_{-1} and C_1 are different from zero. Then the two consistency conditions (8.7) yield a unique value for the two coefficients:

$$C_{-1} = \frac{1}{2}(I - \lambda A), \quad C_1 = \frac{1}{2}(I + \lambda A)\,. \tag{8.9}$$

The resulting difference scheme is called the *Lax-Friedrichs scheme*:

$$u_h^{n+1} = C_{-1}u_{h-1}^n + C_1 u_{h+1}^n\,. \tag{8.10}$$

Note the following important distinction between the differential equation (8.3) and the approximation (8.10): whereas the initial value problem for (8.3) can be solved equally well for t positive or t negative, (8.10) is set up to solve the initial value problem only in the positive time direction.

8.2. Domain of Dependence

The value of u_i^n is determined by the values of u_{h-1}^{n-1} and u_{h+1}^{n-1}; these in turn are determined by the values of u_{h-2}^{n-2}, u_h^{n-2}, and u_{h+2}^{n-2}. Continuing in this fashion backward in time, we find that u_i^n is determined by the initial values u_{h-n}^0, u_{h-n+2}^0, $\dots$, u_{h+n}^0. We can express this result as follows:

The *domain of dependence* of u_h^n on the initial data consists of all lattice points in the interval $[(h-n)\delta, (h+n)\delta]$.

In Section 4.1 we showed that the domain of dependence of $u(x,t)$, $t > 0$, on the initial data lies in the interval cut out of the initial line by the leftmost and rightmost characteristic curves. These characteristic curves propagate with speed $-\sigma_{\min}$ and $-\sigma_{\max}$, the smallest and largest eigenvalues of the matrix A. When A does not depend on x and t, these characteristic curves are straight lines, and the interval they cut out of the initial line is $[x + \sigma_{\min}t, x + \sigma_{\max}t]$.

Setting $x = h\delta$, $t = n\epsilon$, the domain of dependence of x, t for the difference scheme consists of integer multiples of δ on the interval $[(h-n)\delta, (h+n)\delta] = [x - t/\lambda, x + t/\lambda]$; here we have used $n\delta = n\epsilon\delta/\epsilon = t/\lambda$, since λ was defined as ϵ/δ.

In their seminal paper Courant, Friedrichs, and Lewy observed that if the scheme is to converge to the solution of the differential equation, then the domain of dependence of the difference scheme must include all points of the domain of dependence of the differential equation. That means that

$$\lambda^{-1} \geq \sigma_{\max}, \quad -\lambda^{-1} \leq \sigma_{\min}, \tag{8.11}$$

the celebrated CFL condition.

8.3. Stability and Convergence

It turns out that the CFL condition is not only necessary but also sufficient for convergence of the LF scheme. The most important step in proving convergence is to prove the *stability* of the scheme. Stability means that the L^2 norm of the solution of the difference scheme at the n^{th} step, $\|u^n\|^2 = \delta \sum_h |u_h^n|^2$, is bounded by a constant multiple of the L^2 norm of the initial data $\|u^0\|^2 = \delta \sum_h |u_i^0|^2$; the constant is typically an exponential function of $t = n\epsilon$.

To derive such an estimate we note that the CFL condition (8.11) implies that the matrices C_{-1} and C_1 in (8.9) are *nonnegative*. The stability of the LF scheme (8.10) follows from the following general theorem due to Friedrichs:

THEOREM 8.1 (Stability Theorem) *Consider a difference scheme of the form*

$$u_h^{n+1} = \sum_j C_j u_{h+j}^n, \quad j \text{ ranging over a finite set,} \tag{8.12}$$

where the coefficients C_j are symmetric matrix functions that satisfy

(8.13a) $C_j \geq 0,$

(8.13b) $\sum_j C_j = I,$

(8.13c) *C_j is a Lipschitz continuous function of $x = h\delta$.*

Then the scheme is stable in the sense that

$$\|u^n\|_\delta \leq e^{O(M)n\delta} \|u^0\|, \tag{8.14}$$

where $\|u^n\|_\delta^2 = \delta \sum |u_h^n|^2$, and M is the Lipschitz constant.

PROOF: Take a vector w, to be specified later, and take the scalar product of (8.12) with w; we get

$$u_h^{n+1} \cdot w = \sum_j C_j u_{h+j}^n \cdot w. \tag{8.15}$$

Since C_j is a nonnegative symmetric matrix, by the Schwarz inequality

$$C_j u \cdot w \leq \sqrt{C_j u \cdot u}\,\sqrt{C_j w \cdot w};$$

applying the arithmetic-geometric mean inequality on the right, we get

$$C_j u \cdot w \leq \frac{1}{2} C_j u \cdot u + \frac{1}{2} C_j w \cdot w.$$

Take $u = u_{h+j}^n$ and use the above inequality to estimate the right side of (8.15); we get

$$u_h^{n+1} \cdot w \leq \frac{1}{2} \sum_j C_j u_{h+j}^n \cdot u_{h+j}^n + \frac{1}{2} \sum C_j w \cdot w. \tag{8.16}$$

Since $\sum_j C_j = I$, the second sum on the right adds up to $\frac{1}{2} w \cdot w$. Now choose $w = u_h^{n+1}$; multiplying (8.16) by 2 gives

$$|u_h^{n+1}|^2 \leq \sum_j C_j u_{h+j}^n \cdot u_{h+j}^n .$$

Sum this over all h and multiply by δ; we get

$$\|u^{n+1}\|_\delta^2 \leq \delta \sum_{h,j} C_j u_{h+j}^n \cdot u_{h+j}^n .$$

Introduce $h + j = k$ as the index of summation in place of h; we get

$$\|u^{n+1}\|_\delta^2 \leq \delta \sum_{j,k} C_j u_k^n \cdot u_k^n , \tag{8.17}$$

where $C_j = C_j(h\delta) = C_j((k - j)\delta)$. Sum first with respect to j; $\sum C_j(k\delta) = I$, and since C_j are Lipschitz continuous and j ranges over a finite set,

$$\sum_j C_j(k\delta - j\delta) = I + O(M)\delta .$$

Setting this into (8.17) gives

$$\|u^{n+1}\|_\delta^2 \leq (1 + O(M)\delta)\|u^n\|_\delta^2 . \tag{8.18}$$

Using (8.18) repeatedly we obtain

$$\|u^n\|_\delta \leq (1 + O(M)\delta)^n \|u^0\|_\delta \leq e^{O(M)\delta n} \|u^0\|_\delta .$$

This proves the stability of the scheme. □

Obviously, Theorem 8.1 holds in any number of space variables.

We now show how to combine the error estimate (8.8) and the stability estimate (8.14) to prove convergence of the scheme as δ tends to zero.

Denote by $S(\epsilon) = S$ the operator that relates the initial data of solutions of (8.2) to their values at time ϵ. Since we have assumed that the matrix A does not depend on t, the operator that relates the initial values of solutions to their value at $t = n\epsilon$ is S^n. Denote by C_δ the operator that relates the u^{n+1} to u^n by equation (8.12). Estimate (8.14) can be expressed so: *the L^2 norm of the operator C_δ^n is $\leq e^{O(M)\delta n}$.*

Denote by P the operation of discretization that maps a continuous function $u(x)$ into its values at the lattice points $h\delta$:

$$P : u(x) \to \{u(h\delta)\} .$$

We convert the error estimate (8.8) into a norm estimate:

LEMMA 8.2 *Consider a difference scheme of form* (8.12) *whose coefficients* C_j *satisfy the consistency condition* (8.7) *with the differential equation* (8.2). *Denote by* g *any twice-differentiable function of* x *of compact support; then*

$$\|PSg - C_\delta Pg\|_\delta \leq O(\delta^2)\,. \tag{8.19}$$

PROOF: In inequality (8.8) set $u(x, n\epsilon) = g$. Square (8.8), sum over h, and multiply by δ. Since g is of compact support, the number of terms in the sum is $O(\delta^{-1})$; taking the square root yields (8.19). □

Here is the basic approximation theorem:

THEOREM 8.3 *Consider a hyperbolic differential equation of form* (8.2) *and a difference scheme of form* (8.12) *that is consistent with the equation in the sense of* (8.7), *and that is stable in the sense of* (8.14). *Then for any sufficiently differentiable initial value* f *of compact support, the difference between the solution of the differential equation and the approximation furnished by the scheme* (8.12) *is*

$$\left\|PS^N f - C_\delta^N Pf\right\|_\delta \leq O(\delta)e^{O(N\delta)}\,. \tag{8.20}$$

PROOF: We start with the identity

$$PS^N f - C_\delta^N Pf = \sum_{j=0}^{N-1} C_\delta^j (PS - C_\delta P) S^{N-j+1} f\,. \tag{8.21}$$

When f is sufficiently smooth, $S^{N-j+1} f$ is twice differentiable; so inequality (8.19) gives

$$\left\|(PS - C_\delta P)S^{N-j+1} f\right\| \leq O(\delta^2)\,. \tag{8.22}$$

Since the scheme is assumed to be stable,

$$\left\|C_\delta^j\right\| \leq \|C_\delta\|^j \leq (1 + O(M)\delta)^j \leq e^{O(M)\delta j}\,. \tag{8.23}$$

Since (8.2) is hyperbolic, $\|S^k\| \leq e^{O(k\epsilon)}$. We can use (8.22) and (8.23) to estimate the right side of (8.21):

$$\left\|PS^N f - C_\delta^N Pf\right\| \leq \sum_0^{N-1} e^{O(M)\delta j} O(\delta)^2\,;$$

since $N = t/\epsilon = t/\lambda\delta$, this is $\leq e^{O(M)t} O(\delta)$. This proves (8.20). □

NOTES.

(1) Lemma 8.2 and Theorem 8.3 hold equally for functions of k space variables, except that then we need $k+1$ consistency conditions of form (8.7). The proof is the same.

(2) The requirement that the coefficients A and C_j be independent of t is inessential. The statement of Theorem 8.3 becomes

$$\left\|P\prod S_j f - \prod C_{\delta,j} Pf\right\|_\delta \leq O(\delta)\,;$$

the proof is the same.

(3) We have shown in Theorem 8.1 that schemes with nonnegative coefficients are stable. It follows that such schemes converge to solutions of differential equations with which they are consistent. In particular, the LF scheme converges.

8.4. Higher-Order Schemes and Their Stability

Error estimate (8.20) shows that if we need an accurate estimate of the solution, we have to choose δ very small, which means that the evaluation of $C_\delta^N f$ takes many steps. An intelligent way to reduce the number of steps needed is to choose the coefficients C_j of the difference scheme (8.3) so that u_i^{n+1} is a better approximation of the exact solution than (8.8). This can be achieved by using Taylor polynomials of higher order in the place of (8.4) and (8.5) :

$$u_h^{n+1} = u_h^n + D_t u\epsilon + \frac{1}{2} D_t^2 u\epsilon^2 + O(\epsilon^3),$$

$$u_{h-j}^n = u_h^n + jD_x u\delta + \frac{j^2}{2} D_x^2 u\delta^2 + O(\delta^3).$$

Setting into

$$u_h^{n+1} = \sum C_j u_{h+j}^n$$

gives

$$u_h^n + D_t u\epsilon + \frac{1}{2} D_t^2 u\epsilon^2 + O(\delta^3) = \sum C_j u_h^n \sum jC_j D_x u\delta + \frac{1}{2} \sum j^2 C_j D_x^2 u\delta^2 + O(\delta^3). \tag{8.24}$$

For simplicity, we first take the case when the coefficient matrix A in equation (8.2) is independent of x as well as t. In this case differentiation of

$$D_t u = AD_x u$$

gives

$$D_x D_t u = AD_x^2 u, \quad D_t^2 u = AD_t D_x u = A^2 D_x^2 u.$$

Setting these relations into (8.24) and equating the coefficients of u, $D_x u$, $D_x^2 u$ on the two sides gives

$$\sum C_j = I, \quad \sum jC_j = \lambda A, \quad \sum j^2 C_j = \lambda^2 A^2, \tag{8.25}$$

where $\epsilon = \lambda\delta$ as before. These three equations for the C_j should be solvable if we take all C_j but C_{-1}, C_0, C_1, equal to zero. Indeed, the reader can easily verify that

$$C_{-1} = \frac{1}{2}(\lambda^2 A^2 - \lambda A), \quad C_0 = I - \lambda^2 A^2, \quad C_1 = \frac{1}{2}(\lambda^2 A^2 + \lambda A), \tag{8.26}$$

is the unique solution of the equations. The resulting difference scheme is called the *Lax-Wendroff scheme*:

$$u_h^{n+1} = C_{-1} u_{h-1}^n + C_0 u_h^n + C_1 u_{h+1}^n. \tag{8.27}$$

The domain of dependence of this scheme is essentially the same as for the LF scheme (8.9), (8.10). Therefore the CFL condition is again a *necessary* condition for the convergence of the scheme

$$\lambda^{-1} \geq \sigma_{\max}, \quad -\lambda^{-1} \leq \sigma_{\min}, \tag{8.11$_1$}$$

where $\sigma_{\max}$ and $\sigma_{\min}$ denote the largest and smallest eigenvalues of the symmetric matrix A. But this time the CFL condition does *not* imply that the matrices C_j are nonnegative. In fact, it is easy to see that at least one of the matrices C_{-1} or C_1 has a negative eigenvalue, except in the trivial cases when the eigenvalues of A are 1 or -1 and the corresponding matrices C_1 has eigenvalues of 1 or 0 and C_{-1} has eigenvalues of 0 or 1. I call these cases trivial because then equation (8.2) can be solved explicitly.

One might surmise that if one doesn't impose at the outset that only C_{-1}, C_0, C_1 are different from zero, one could satisfy the consistency relations (8.25) by nonnegative matrices; alas, this hope is in vain.

THEOREM 8.4 *The consistency relations* (8.25) *cannot be satisfied by nonnegative matrices except in trivial cases.*

PROOF: First we take the scalar case; for simplicity we take $\lambda = 1$ and a positive:

$$\sum c_j = 1, \quad \sum jc_j = a, \quad \sum j^2 c_j = a^2, \tag{8.28}$$

c_j and a scalars. Suppose $c_j \geq 0$; by the Schwarz inequality and (8.28),

$$a = \sum jc_j = \sum \left(jc_j^{1/2}\right)c_j^{1/2} \leq \left(\sum j^2 c_j\right)^{1/2}\left(\sum c_j\right)^{1/2} = a.$$

Equality holds only if the two vectors $\{jc_j^{1/2}\}$ and $\{c_j^{1/2}\}$ are proportional; that means that exactly one c_j is $\neq 0$, and therefore by (8.28) equal to 1. This is one of the trivial cases.

We now turn to the matrix case. Denote by w one of the eigenvectors of norm 1 of the symmetric matrix A. Let (8.25) act on w and form the scalar product with w. Denoting $C_j w \cdot w$ by c_j and the corresponding eigenvalue of A by a, we obtain equation (8.28). If all the C_j were nonnegative matrices, the c_j would be nonnegative numbers, and so by the result above we are in one of the trivial cases. □

The positivity of the coefficients C_j was an essential ingredient of the proof of the stability theorem, Theorem 8.1. To prove the stability of the LW scheme (8.26), (8.27) we need a new method. When A and thereby C_{-1}, C_0, C_1 are independent of x, we can use the Fourier transform. Define the Fourier transform of $\tilde{u}^n$ of the array $\{u_h^n\}$ by

$$\tilde{u}^n(\theta) = \sum u_h^n e^{ih\theta}. \tag{8.29}$$

Define the *symbol* of the difference scheme (8.12) to be

$$C(\theta) = \sum C_j e^{ij\theta}. \tag{8.30}$$

Taking the Fourier transform of (8.12) we obtain

$$\tilde{u}^{n+1}(\theta) = \bar{C}(\theta)\tilde{u}^n(\theta).$$

Using this relation repeatedly gives

$$\tilde{u}^n(\theta) = \bar{C}(\theta)^n \tilde{u}^0(\theta). \tag{8.31}$$

We now recall Parseval's relation, which says that

$$\|u^n\|^2 = \delta \int |\tilde{u}^n(\theta)|^2 \bar{d}\theta, \tag{8.32}$$

where $\bar{d}\theta$ is $d\theta/2\pi$. Combining (8.31) and (8.32) gives

$$\|u^n\|^2 = \delta \int |\bar{C}(\theta)^n \tilde{u}^0(\theta)|^2 \bar{d}\theta, \tag{8.33}$$

which leads immediately to the following stability criterion, valid in any number of space dimensions:

THEOREM 8.5 *The difference scheme*

$$u_h^{n+1} = \sum C_j u_{h+j}^n \tag{8.12}$$

with constant coefficients is stable if the norm of the powers of its symbol $C(\theta)$ *defined by* (8.30) *are uniformly bounded* :

$$|C^n(\theta)| \le K \tag{8.34}$$

for some K, *for all* θ, *and all* n.

PROOF: The norm of u^n is easily estimated from (8.33) when (8.34) is available:

$$\|u^n\|_\delta^2 \le \delta K^2 \int |\tilde{u}^0(\theta)|^2 \bar{d}\theta = K^2 \|u^0\|_\delta^2.$$

In the last step we used Parseval's relation (8.32). This inequality can be expressed in terms of the operator C_δ linking u^{n+1} to u^n: the norm of the n^{th} power of C_δ is $\le K$ for all n. □

Here is a useful necessary and a sufficient condition for (8.34) to hold:

COROLLARY 8.6

(i) *If* $C(\theta)$ *satisfies* (8.34), *then the eigenvalues of* $C(\theta)$ *are* ≤ 1 *in absolute value for all* θ.

(ii) *If* $|C(\theta)| \le 1$ *for all* θ, *then condition* (8.34) *holds with* $K = 1$.

PROOF:

(i) The spectral radius of $C(\theta)$ equals $\lim_{n\to\infty} |C^n(\theta)|^{1/n}$.

(ii) This condition is obvious.

□

We now show how to use criterion (ii) to prove the stability of the LW scheme.

THEOREM 8.7 *The LW scheme* (8.26) *and* (8.27), *with constant coefficients, is stable provided that the CFL condition is satisfied.*

PROOF: Abbreviate λA as A; then the CFL condition says that the eigenvalues of A lie between -1 and 1.

Next we compute the symbol of the LW scheme:

$$C(\theta) = \frac{1}{2}(A^2 - A^4)e^{-i\theta} + I - A^2 + \frac{1}{2}(A^2 + A)e^{i\theta} \\ = A^2(\cos\theta - 1) + I + iA\sin\theta \tag{8.35}$$

and

$$C^*(\theta)C(\theta) = (A^2(\cos\theta - 1) + I)^2 + A^2\sin^2\theta \\ = A^4(\cos\theta - 1)^2 + A^2(2(\cos\theta - 1) + \sin^2) + I \\ = (A^4 - A^2)(\cos\theta - 1)^2 + I\,. \tag{8.36}$$

Since the eigenvalues of A are ≤ 1 in absolute value, $A^4 - A^2$ is negative; this proves that $C^*(\theta)C(\theta) \leq I$.

Take any vector w; then

$$|C(\theta)w|^2 = (C(\theta)w, C(\theta)w) = (w, C^*(\theta)C(\theta)w) \leq (w, w) = |w|^2\,;$$

in the last step we have used $C^*(\theta)C(\theta) \leq I$. The above inequality asserts that $|C(\theta)| \leq 1$ for all θ. By part (ii) of Corollary 8.6 it follows that the scheme is stable. □

NOTES.

(1) Condition (i) of Corollary 8.6 on the spectrum of the symbol of a difference scheme is the discrete analogue of the condition of hyperbolicity on the symbol of the differential operator (8.2). Both were derived by Fourier analysis of the growth of solutions.

(2) We have proven the stability of the LW scheme in the L^2 norm. In Section 4.1 we have shown that solutions of hyperbolic equations in one space variable are stable in all the L^p norms, $1 \leq p \leq \infty$. However, it was shown by Brenner et al. that the LW scheme is *not* stable in any L^p norm except for $p = 2$. In contrast, the LF scheme is stable in all L^p norms.

(3) Condition (i) of Corollary 8.6 is necessary for stability of schemes in any number of dimensions. It is called the *von Neumann condition*; von Neumann formulated it in his study of the discretization of the equations of fluid dynamics. He conjectured that the criterion guarantees stability even for schemes for nonlinear equations. Similar stability conditions were also formulated by Olga Ladyzhenskaya.

We now turn to the intriguing problem of showing that von Neumann's criterion, mildly strengthened, implies the stability of difference schemes with variable coefficient.

THEOREM 8.8 *Suppose that the matrix $A(x)$ depends Lipschitz continuously on x. Then the scheme* (8.26), (8.27) *is stable if the CFL condition is satisfied.*

PROOF: We have already computed the symbol $C(\theta)$ of this scheme; formula (8.36) gives

$$C^*(\theta)C(\theta) = I + (A^4 - A^2)(\cos\theta - 1)^2 . \tag{8.37}$$

The CFL condition requires that the eigenvalues of $A(x)$ lie between -1 and 1 for all x.

We define $D(\theta) = (\cos\theta - 1)I$; then (8.37) can be rewritten as

$$C^*(\theta)C(\theta) + (A^2 - A^4)D^*(\theta)D(\theta) = I . \tag{8.38}$$

We define the matrix function K as

$$K(\varphi,\theta) = C^*(\varphi)C(\theta) + (A^2 - A^4)D^*(\varphi)D(\theta) . \tag{8.39}$$

LEMMA 8.9 *$K(\varphi,\theta)$ has the following properties:*

(i) *The integral operator with kernel $K(\varphi,\theta)$ acting on vector-valued functions is symmetric and positive,*

(ii) *$K(\theta,\theta) = I$,*

(iii) *$K(0,\theta) = C(\theta)$ and,*

(iv) *$K(\varphi,\theta)$ depends Lipschitz continuously on x.*

PROOF:

(i) Denote by K the integral operator with kernel $K(\varphi,\theta)$. Using formula (8.39) we get

$$\begin{aligned}(KU,U) &= \iint K(\varphi,\theta)U(\theta)d\theta \cdot U(\varphi)d\varphi \\ &= \iint C^*(\varphi)C(\theta)U(\theta)\cdot U(\varphi)d\theta\, d\varphi \\ &\quad + \iint (A^2 - A^4)D^*(\varphi)D(\theta)U(\theta)\cdot U(\varphi)d\theta\, d\varphi \\ &= \int C(\theta)U(\theta)d\theta \cdot \int C(\varphi)U(\varphi)d\varphi \\ &\quad + \int BD(\theta)U(\theta)d\theta \cdot \int BD(\varphi)U(\varphi)d\varphi .\end{aligned} \tag{8.40}$$

Here B denotes a square root of $A^2 - A^4$, and we have used the fact that B commutes with D. Clearly (8.40) is positive or zero.

(ii) This follows from (8.38).

(iii) This follows from the fact that $C(0) = I$ and $D(0) = 0$.

(iv) This follows from the fact that the ingredients of $K(\varphi,\theta)$, powers of A, are Lipschitz continuous.

□

Set in (8.40)

$$U(\theta) = \sum a_m e^{im\theta} , \quad U(\varphi) = \sum a_\ell e^{i\ell\varphi} ;$$

we get

$$\int K(\varphi,\theta)U(\theta)\cdot U(\varphi)\bar{d}\theta\,\bar{d}\varphi = \sum K_{m\ell}a_m\cdot a_\ell\,,$$

where

$$K_{m\ell} = \int K(\varphi,\theta)e^{i(m\theta-\ell\varphi)}\,\bar{d}\theta\,\bar{d}\varphi\,. \tag{8.41}$$

Since $K(\varphi,\theta)$ is a positive kernel, it follows that the block matrix $\mathcal{K} = (K_{\ell m})$ is positive. The Fourier coefficients of $K(\varphi,\theta)$ are given by formula (8.41); therefore K itself can be expressed as

$$K(\varphi,\theta) = \sum K_{m\ell}e^{i(\ell\varphi-m\theta)}\,. \tag{8.41'}$$

Property (ii) of Lemma 8.9, $K(\theta,\theta) = I$, can be expressed using (8.41′) in terms of the $K_{m\ell}$ as follows:

$$\sum_\ell K_{\ell\ell} = I\,,\qquad \sum_{m-\ell=r} K_{m\ell} = 0 \quad \text{for } r\neq 0\,. \tag{8.42}$$

Property (iii) of Lemma 8.9, $K(0,\theta) = C(\theta)$, can be expressed in terms of the $K_{m\ell}$, using formulas (8.41′) and (8.30), as follows:

$$\sum_\ell K_{m\ell} = C_{-m}\,. \tag{8.43}$$

We are now ready to tackle the stability of the scheme (8.26), (8.27). For simplicity denote u_h^n as u_h and u_h^{n+1} as v_h. Then we can write the scheme as

$$v_h = \sum_j C_j u_{h+j}\,. \tag{8.44}$$

Let w denote a vector to be specified later, and take the scalar product of (8.44) with w; since the C_j are symmetric, we have

$$v_h\cdot w = \sum_j C_j u_{h+j}\cdot w = \sum_j u_{h+j}\cdot C_j w\,.$$

Using relation (8.43) and switching j to $-j$, we get

$$v_h\cdot w = \sum_{j,\ell} u_{h-j}\cdot K_{j\ell}w\,.$$

Since the block matrix $K_{j\ell}$ is positive, we can apply the Schwarz inequality on the right:

$$(v_h\cdot w)^2 \le \Big(\sum_{j,\ell} u_{h-j}\cdot K_{j\ell}u_{h-\ell}\Big)\Big(\sum_{j,\ell} w\cdot K_{j\ell}w\Big)\,.$$

It follows from (8.42) that

$$\sum_{j,\ell} w\cdot K_{j\ell}w = w\cdot w\,;$$

so the above inequality says that

$$(v_h \cdot w)^2 \le \Big(\sum_{j,\ell} u_{h-j} \cdot K_{j\ell} u_{h-\ell}\Big) w \cdot w .$$

Now we choose $w = v_h$ and obtain

$$|v_h|^2 \le \sum_{j,\ell} u_{h-j} \cdot K_{j\ell} u_{h-\ell} .$$

Sum this over h and multiply by δ; we get

$$\|v\|_\delta^2 \le \delta \sum_{h,j,\ell} u_{h-j} \cdot K_{j\ell} u_{h-\ell} .$$

We now recall that the matrix $K_{j\ell}$ depends Lipschitz continuously on x. Writing $x = h\delta$ and introducing $h - j = k$ as a new index of summation, we get

$$\|v\|_\delta^2 \le \delta \sum_{k,j,\ell} u_k \cdot K_{j\ell}(h\delta) u_{k+j-\ell} . \tag{8.45}$$

On the right side replace $K_{j\ell}(h\delta)$ by $K_{j\ell}(k\delta)$; then using relations (8.42) we realize that the right side of (8.45) thus modified is $\delta \sum u_k \cdot u_k = \|u\|^2$. Since $K_{j\ell}(x)$ is Lipschitz continuous and since j ranges over a finite set of integers (in fact, $-1, 0, 1$), the error committed by the replacement is $O(\delta)\|u\|^2$. So we deduce from (8.45) that

$$\|v\|_\delta^2 \le (1 + O(\delta))\|u\|_\delta^2 . \tag{8.46}$$

This proves the stability of (8.26), (8.27). □

The reader will no doubt observe that we have in hand the elements of a proof of a much more general stability result.

THEOREM 8.10 *The difference scheme*

$$v_h = \sum_j C_j u_{h+j} , \quad |j| \le N , \tag{8.3}$$

is stable provided that its symbol satisfies these conditions:

(i) $\sum C_j = I$,

(ii) $C(x, \theta)$ *depends Lipschitz continuously on* x,

(iii) $C^*(\theta)C(\theta) \le I$, *with the inequality holding at all* θ *except* $\theta = 0$,

(iv) *for* θ *near* 0,

$$C^*(\theta)C(\theta) \le I - Q(x)\theta^{2q} + O(\theta^{2q+1}) ,$$

where $Q(x)$ *is positive definite, and* q *a natural number.*

NOTE. It follows from formula (8.36) that for the LW scheme $q = 2$ provided that in the CFL condition (8.11) the strict inequality holds.

For the LF scheme (8.9), (8.10)

$$C(\theta) = I\cos\theta + iA\sin\theta ,$$

so

$$C^*(\theta)C(\theta) = I\cos^2\theta + A^2\sin^2\theta \le I - (I - A^2)\sin^2\theta\,.$$

This shows that for the LF scheme $q = 1$ provided that the CFL condition holds strictly.

PROOF: First we treat the scalar case. It follows from assumption (iii) of Theorem 8.10 that $I - C^*(\theta)C(\theta)$ is a nonnegative trigonometric polynomial of degree $2N$. According to a classical theorem of Fejér and F. Riesz, a nonnegative trigonometric polynomial of degree $2N$ can be represented as the absolute value squared of a trigonometric polynomial $D(\theta)$ of degree N. It follows therefore that

$$C^*(\theta)C(\theta) + D^*(\theta)D(\theta) = I\,.$$

It is not hard to show that if $C(x,\theta)$ satisfies condition (iv) of Theorem 8.10 and depends Lipschitz continuously on x, then so does $D(x,\theta)$. This allows us to define the kernel $K(\varphi,\theta)$ as

$$K(\varphi,\theta) = C^*(\varphi)C(\theta) + D^*(\varphi)D(\theta)\,;$$

then the proof proceeds as the proof of Theorem 8.8 via Lemma 8.9.

This last argument can be extended to the matrix-valued case thanks to Murray Rosenblatt's extension of the Fejér-Riesz theorem to matrix-valued trigonometric polynomials. □

NOTE. One can avoid appealing to the Fejér-Riesz theorem by choosing $D(\theta)$ as the positive square root of $I - C^*(\theta)C(\theta)$. Defined this way, $D(\theta)$ is not a trigonometric polynomial, but its Fourier coefficients die down fast enough so that crucial estimate (8.46) can be deduced from inequality (8.45).

We now return to the LW scheme.

When the coefficient A in equation (8.2), $D_t u = A D_x u$, depends on x, the LW scheme (8.26) has to be modified slightly. Differentiating the equation gives

$$\begin{aligned} D_x D_t u &= A D_x^2 u + A_x D_x u\,,\\ D_t^2 u &= A D_t D_x u = A^2 D_x^2 u + A A_x D_x u\,. \end{aligned}$$

Setting these relations into (8.24) and equating the coefficients of u, $D_x u$, $D_x^2 u$ on the two sides gives

$$\sum C_j = I\,,\quad \sum jC_j = A + \delta A A_x\,,\quad \sum j^2 C_j = A^2\,.$$

Here we have chosen $\epsilon = \delta$, i.e., $\lambda = 1$, for the sake of simplicity. These equations are easily solved for C_{-1}, C_0, C_1:

$$\text{(8.47)}\quad C_{-1} = \frac{1}{2}(A^2 - A - \delta A A_x)\,,\quad C_0 = I - A^2\,,\quad C_1 = \frac{1}{2}(A^2 + A + \delta A A_x)\,.$$

This scheme is merely a perturbation of the scheme (8.26), (8.27); we shall now show that its stability can be deduced from the former with the help of the following:

THEOREM 8.11 *Let F be an operator in a Hilbert space whose powers are bounded in norm: $\|F^n\| \leq K$. Let G be a bounded operator, $\|G\| \leq M$. Then*

$$\|(F+\delta G)^n\| \leq K(1+MK\delta)^n .$$

PROOF: By the noncommutative binomial theorem,

$$(F+\delta G)^n = F^n + \delta \sum F^j G F^{n-j-1} + \cdots .$$

The norm of the sums on the right is bounded by

$$K + nK^2 M\delta + \binom{n}{2} K^3 M^2 \delta^2 + \cdots = K(1+MK\delta)^n .$$

□

Denote by F the operator linking $\{u_h^n\}$ to $\{u_h^{n+1}\}$ by formulae (8.26) and (8.27), and denote by G the operator with coefficients $C_- = -AA_x/2$, $C_0 = 0$, and $C_1 = AA_x/2$. The operator C_δ with coefficients given by formula (8.47) can then be expressed as $F + \delta G$. According to Theorem 8.8, F is stable and therefore $|F^n| \leq K$ for $n \leq N$. By Theorem 8.11, $F + \delta G$ is stable too.

Denote as before by S the operator that relates the initial data of solutions to their values at $t = \epsilon$, and by P the discretization operator:

$$P : u(x) \rightarrow \{u(h\delta)\} .$$

Denote the operator $F + \delta G$ as C_δ. In analogy with Lemma 8.2, we deduce from (8.24) and (8.47) that for any three-times-differentiable function f of compact support

$$\|PSf - C_\delta P f\| \leq O(\delta^3) . \tag{8.48}$$

From this and the stability of C_δ, we deduce, as in the proof of Theorem 8.3, that

$$\left\|PS^N f - C_\delta^N P f\right\| \leq O(\delta^2) e^{O(N\delta)} . \tag{8.49}$$

This proves the *convergence* of the LW scheme in the L^2 norm. Note that since the error in (8.48) is $O(\delta^3)$, the approximation error (8.49) is $O(\delta^2)$, a significant improvement over (8.20) in Theorem 8.3.

Gil Strang constructed different schemes of arbitrary high order of accuracy that are stable under a suitable restriction on λ.

8.5. The Gibbs Phenomenon

Let's take another look at the LW scheme; for simplicity take $\lambda = 1$, and the scalar case:

$$v_h = \frac{1}{2}(a^2 - a)u_{h-1} + (1 - a^2)u_h + \frac{1}{2}(a^2 + a)u_{h+1} .$$

The CFL condition requires a to be less than 1 in absolute value. We choose *discontinuous* initial data

$$u_h = \begin{cases} 0 & \text{for } h < 0 \\ 1 & \text{for } h \geq 0. \end{cases}$$

From the definition of v_h above

$$v_h = \begin{cases} 0 & \text{for } h < -1 \\ \frac{1}{2}(a^2 + a) & \text{for } h = -1 \\ 1 + \frac{1}{2}(a - a^2) & \text{for } h = 0 \\ 1 & \text{for } h > 0. \end{cases}$$

It follows from this formula for v_h that when a is negative, v_{-1} undershoots the initial data; for a positive an overshoot occurs at $h = 0$. Let's take a to be positive; the amount of overshoot is $\frac{1}{2}(a - a^2)$. The maximum of this quantity occurs at $a = \frac{1}{2}$, where it amounts to $\frac{1}{8}$, about 12% of the magnitude of the jump in the initial data.

The explanation of the overshoot is that the LW scheme is designed to give a good approximation to the solution of the initial value problem at $t = 0$ for three-times-differentiable initial data. When the scheme is applied to discontinuous initial data, the features designed to give a good approximation for smooth data produce instead gross distortions.

The same happens when a function is approximated by the partial sums of its Fourier series. For smooth functions the partial sums give an excellent approximation to the function. When applied to a discontinuous function, the features that produce an excellent approximation lead instead to an over and undershoot of about 8% of the size of the discontinuity, called the *Gibbs phenomenon*.

A clever strategy for avoiding a Gibbs phenomenon was devised by Ami Harten and Gideon Zwass. They proposed monitoring the smoothness of the solution being computed, and at points where the solution appear to lack smoothness, switching from a high-order scheme to a low-order scheme. We shall illustrate how the method works by using LW as the high-order scheme and LF as the low-order one.

The criterion that tests the smoothness of the computed solution u at h is the ratio

$$\frac{u_h - u_{h-1}}{u_{h+1} - u_h} = r_h\,. \tag{8.50}$$

When r_h is near 1, the solution is smooth at h; when r_h differs appreciably from 1, then either u is not differentiable at h, or u_ℓ has a maximum or minimum at or near h.

Using (8.50) we can express

$$u_h = \frac{1}{1+r}u_{h-1} + \frac{r}{1+r}u_{h+1}\,.$$

Setting this into the LW scheme

$$v_h = \frac{a^2 - a}{2}u_{h-1} + (1 - a^2)u_h + \frac{a^2 + a}{2}u_{h+1}\,,$$

we get

$$v_h = \left(\frac{a^2 - a}{2} + \frac{1 - a^2}{1+r}\right)u_{h-1} + \left(\frac{a^2 + a}{2} + \frac{r(1 - a^2)}{1+r}\right)u_{h+1}\,. \tag{8.51}$$

For r positive the coefficient of u_{h-1} in (8.51) is a decreasing function of r, while the coefficient of u_{h+1} is an increasing function of r. Therefore in the interval $\frac{1}{2} \le r \le 2$, the minimum value of the coefficient of u_{h-1} is $(2 - 3a + a^2)/6$, while the minimum of the coefficient of u_{h+1} is $(2 + 3a + a^2)/6$. Both of these quantities are *positive* in the interval $-1 < a < 1$.

We form a hybrid of the LW and LF schemes by setting

$$C = sC^W + (1 - s)C^F . \tag{8.52}$$

Here $s = s(r)$ is a *switch* that turns off the LW scheme when $r = r_h$ differs appreciably from 1, and turns on the LF scheme. The function $s(r)$ is chosen to be equal to 1 at $r = 1$ and 0 for $r \le \frac{1}{2}$ and $r \ge 2$. In this range we have shown that the LW scheme expresses v_h as a linear combination of u_{h-1} and u_{h+1} with positive coefficients that add up to 1. The LF scheme also has this property, therefore so does the hybrid scheme (8.52). It follows in particular that

$$\min(u_{h-1}, u_h, u_{h+1}) \le v_h \le \max(u_{h-1}, u_h, u_{h+1}) .$$

We can express this property in the following words:

THEOREM 8.12 *When the CFL condition is satisfied, the hybrid scheme* (8.52) *is stable in the L^∞ norm $|u|_\infty = \max_h |u_h|$, and does not exhibit the Gibbs phenomenon.*

Hybrid schemes of the above type have been constructed for symmetric hyperbolic systems in any number of space dimensions by Xu-Dong Liu, and applied to quasi-linear hyperbolic systems of conservation laws as well. Solutions of these can develop spontaneous discontinuities—shocks, whereas discontinuities of solutions of linear hyperbolic equations always originate in initial discontinuities.

Since hybrid schemes are nonlinear even when used to solve linear equations, the simple proof given for Theorem 8.3, which shows that stability implies convergence, cannot be used. Not much has been proven about the convergence of hybrid schemes, but the evidence of numerical calculations shows that they work very well indeed.

In the next section we outline another—linear—approach for the approximation of solutions of linear hyperbolic equations that contain discontinuities.

8.6. The Computation of Discontinuous Solutions of Linear Hyperbolic Equations

The error estimate (8.49) shows that for smooth initial data the approximate solution furnished by the LW scheme differs from the exact solution by $O(\delta^2)$. When the cruder LF scheme is used, the error, as estimated by (8.20), is only $O(\delta)$. A natural question arises: when the initial data are piecewise-smooth but contain discontinuities, does it pay to use the more accurate LW scheme? In this section we show that the answer is yes; Michael Mock and the author devised a way of preventing the gross errors that arise at the discontinuity from polluting the calculation in the smooth regions. The tool is *preprocessing* the initial data and *postprocessing* the numerical answer.

We start with a quadrature formula going back to Newton:

THEOREM 8.13 *Let $s(x)$ be a C^∞ scalar function with bounded support defined on $\mathbb{R}_+$. Given any positive integer q, there is a quadrature formula that is accurate of order q, of the form*

$$\int_0^\infty s(x)dx = \delta \sum_0^\infty w_h s(h\delta) + O(\delta^q), \tag{8.53}$$

where the weights w_h depend on q, but $w_h = 1$ for $h \geq q$.

Here are the values of the weight for low values of q:

q	w_0	w_1	w_2	w_3
2	$\frac{1}{2}$	1	1	1
3	$\frac{3}{8}$	$\frac{9}{8}$	1	1
4	$\frac{3}{8}$	$\frac{7}{6}$	$\frac{23}{24}$	1

For a derivation see any old-fashioned text on numerical analysis.

The following is an immediate consequence:

COROLLARY 8.14 *Let s be a piecewise C^∞ function on $\mathbb{R}$ with compact support and a discontinuity at $x = 0$. Then for any positive integer q,*

$$\int_{-\infty}^\infty s(x)dx = \delta \sum w'_h s(h\delta) + O(\delta^q), \tag{8.53$'$}$$

where $w'_0 = 2w_0$, $w'_h = w_{|h|}$, and

$$s(0) = \frac{s(0_-) + s(0_+)}{2}.$$

We shall study discontinuous solutions of first-order hyperbolic systems of PDEs in one space variable, not necessarily symmetric:

$$D_t u = A D_x u + B, \tag{8.54}$$

where A and B are C^∞ matrix-valued functions of x. Discontinuous solutions of (8.54) satisfy the equation in the sense of distribution. An equivalent formulation is this: Let $v(x, t)$ be a C^∞ solution with compact support in x of the *adjoint equation*

$$D_t v = D_x A^* v - B^* v. \tag{8.54*}$$

Then

$$(u(t), v(t)) \stackrel{\text{def}}{=} \int u(x, t) \cdot v(x, t) dx$$

is independent of t. Another way of expressing this is

$$(u(T), v(T)) = (u(0), v(0)) \quad \text{for all } T. \tag{8.55}$$

Take any two-level forward difference scheme to approximate solutions of (8.54):

$$u_h^{n+1} = \sum C_j u_{h+j}^n, \quad C_j = C_j(h). \tag{8.56}$$

Take any other function v_h^n defined on the lattice, form the scalar product of (8.56) with v_h^{n+1}, and sum over all h; we get

$$\sum_h u_h^{n+1} \cdot v_h^{n+1} = \sum_{h,j} C_j u_{h+j}^n \cdot v_h^{n+1} = \sum_{k,j} u_k^n \cdot C_j^*(k-j) v_{k-j}^{n+1};$$

in the last step we have introduced $k = h + j$ as a new index to be summed over.

If the lattice function v_h^n satisfies the adjoint relation

$$v_k^n = \sum C_j^*(k-j) v_{k-j}^{n+1}, \tag{8.56*}$$

then the above identity can be stated as follows:

$$(u^{n+1}, v^{n+1})_\delta = (u^n, v^n)_\delta, \tag{8.57}$$

where the scalar product $(\ ,\ )_\delta$ is the L^2 scalar product over the lattice

$$(u, v)_\delta = \delta \sum u_h \cdot v_h.$$

It follows from (8.57) that for any N,

$$(u^N, v^N)_\delta = (u^0, v^0)_\delta, \tag{8.55$_\delta$}$$

a discrete analogue of (8.55).

DEFINITION The scheme (8.56) approximates differential equation (8.54) with *accuracy of order* q if for all C^∞ solutions $u(x,t)$ of (8.54) that have compact support in x, the following holds: Define u_h^n as $u(h\delta, n\epsilon)$, u_h^{n+1} by formula (8.56), and w_h^{n+1} as $u(h\delta, (n+1)\epsilon)$; then

$$\|w^{n+1} - u^{n+1}\|_\delta \le O(\delta^{q+1}).$$

REMARK. Formula (8.19) shows that the LF scheme is accurate of order $q = 1$, and formula 8.47 shows that the LW scheme is accurate of order $q = 2$.

THEOREM 8.15

(i) *If the scheme* (8.56) *approximates the differential equation* (8.54) *with accuracy q, then the adjoint scheme* (8.56*) *approximates the adjoint of equation* (8.54) *with the same accuracy q.*

(ii) *If scheme* (8.56) *is stable, so it its adjoint* (8.56*).

EXERCISE Prove Theorem 8.15.

The proof offered for Theorem 8.3 also proves the following more general result:

THEOREM 8.16 *Suppose that the difference scheme* (8.56) *approximates the differential equation* (8.54) *with accuracy of order q and is stable. Take any C^∞ initial function $f(x)$ of compact support; denote by $u(x,t)$ the solution of equation* (8.54) *with initial value $f(x)$, and denote by u_h^n the solution of the difference scheme* (8.56) *with initial value $u_h^0 = f(h\delta)$. Then for any time T with $N\epsilon = T$,*

$$\|Pu(T) - u^N\|_\delta \le O(\delta^q);$$

here $Pu(T) = \{u(h\delta, T)\}$.

We now turn to initial data f that are piecewise C^∞ with a discontinuity at $x = 0$. The first step is to *preprocess* the initial data. Define u_h^0 as follows:

$$u_h^0 = w_h' f(h\delta), \tag{8.58}$$

where the w_h' are defined in Corollary 8.14.

THEOREM 8.17 *Impose the same assumptions on the difference scheme as in Theorem* 8.15, *but take the initial data f to be piecewise C^∞ with a discontinuity at $x = 0$. Denote by $u(x,t)$ the distribution solution of equation* (8.54) *with initial value f, and by u_h^n the solution of the difference scheme with preprocessed initial data* (8.58). *Then for any time $T = N\epsilon$ and any C^∞ vector-function $g(x)$ of compact support,*

$$\int u(x,T)\cdot g(x)dx = \delta\sum u_h^N\cdot g(h\delta) + O(\delta^q). \tag{8.59}$$

PROOF: Denote by $v(x,t)$ the solution of the equation adjoint to (8.54) whose value at T is $g(x)$:

$$v(x,T) = g(x).$$

Now we use relation (8.55); noting that $v(x,T) = g(x)$ and $u(x,0) = f(x)$, we get

$$\int u(x,T)\cdot g(x)dx = \int f(x)\cdot v(x,0)dx. \tag{8.60}$$

Next we note that the adjoint difference scheme (8.56*) is a two-level *backward* difference scheme; denote by v_h^n that solution of (8.56*) whose value at N is $g(k\delta)$:

$$v_k^N = g(k\delta).$$

According to Theorem 8.15, the difference scheme (8.56*) approximates the adjoint of (8.54) with accuracy of order q and is stable. Therefore by Theorem 8.16 applied backwards in time,

$$\|Pv(0) - v^0\|_\delta \le O(\delta^q), \tag{8.61}$$

where $Pv(0) = \{v(h\delta, 0)\}$.

We now use relation (8.55_δ); noting that $v_h^N = g(h\delta)$ and, by (8.58), $u_h^0 = w_h' f(h\delta)$, we get

$$\delta\sum u_h^N\cdot g(h\delta) = \delta\sum w_h' f(h\delta)\cdot v_h^0. \tag{8.60*}$$

Next we show that the right side of (8.60*) differs by $O(\delta^q)$ from

$$\delta\sum w_k' f(h\delta)\cdot v(h\delta,0). \tag{8.62}$$

We estimate the difference of the two by using the Schwarz inequality and estimate (8.61):

$$\delta\sum w_h' f(h\delta)\cdot\left[v_h^0 - v(h\delta,0)\right] \le \|f'\|_\delta\,\|v^0 - Pv(0)\|_\delta = O(\delta^q).$$

Here f' denotes $\{w_h' f(h\delta)\}$.

We now apply (8.53′) of Corollary 8.14 to $s(x) = f(x) \cdot v(x, 0)$ to conclude that (8.62) differs by $O(\delta^q)$ from

$$\int f(x) \cdot v(x, 0) dx .$$

This show that the right side of (8.60*) differs from the right side of (8.60) by $O(\delta^q)$. But then the left sides differ by the same amount; this proves estimate (8.59) of Theorem 8.17. □

Theorem 8.17 can be stated in the following words: the solution u_h^N of the difference equation with preprocessed initial data (8.58) differs in the *weak sense* from the exact solution $u(x, T)$ by $O(\delta^q)$.

As the last step, we show how to use the weak error bound (8.59) to obtain pointwise estimates of the solution $u(x, T)$. We have shown in Chapter 7 that discontinuities propagate along characteristics. Therefore we know the exact locations of the discontinuities of $u(x, t)$. We now show how to use Theorem 8.17 to reconstruct with good accuracy the values of $u(y, T)$ at points not too close to the discontinuities of u at $t = T$.

Denote by $m(x)$ an auxiliary function with the following properties:

(i) $m(x)$ is $q + 1$ times differentiable and is supported on $[-1, 1]$.

(ii)

$$\int m(x) dx = 1 , \quad \int x^j m(x) dx = 0 , \quad j = 1, \ldots, q - 1 . \tag{8.63}$$

The function $m_\ell(x) = m(x/\ell)/\ell$ has the corresponding properties:

(i′) $m_\ell(x)$ is supported on $[-\ell, \ell]$, and its derivative of order $q + 1$ is bounded by $O(\ell^{-q-2})$.

(ii′) same as (ii).

Suppose that the interval $[y - \ell, y + \ell]$ is free of discontinuities of $u(x, T)$. A good approximation to $u(y, T)$ is furnished by the weighted mean

$$\int u(x, T) m_\ell(x - y) dx .$$

Changing to $x - y = z$ as variables of integration, and approximating $u(x, T)$ by its Taylor polynomial at y, we get

$$\begin{aligned} &\int u(x, T) m_\ell(x - y) dx \\ &\quad = \int u(z + y, T) m_\ell(z) dz \\ &\quad = \int \left[u(y) + z D_x u(y) + \cdots + z^{q-1} D_x^{q-1}(y) + O(\ell^q)\right] m_\ell(z) dz \\ &\quad = u(y) + O(\ell^q) ; \end{aligned} \tag{8.64}$$

in the last step we have used the identities (8.63) as they apply to m_ℓ.

We now apply the conclusion of Theorem 8.17 (8.59) to $g(x) = m_\ell(x - y)$. The coefficient of δ^q in $O(\delta^q)$ in (8.59) is bounded by $\max|D_x^{q+1} g|$; according to (i′), for $g = m_\ell$ this quantity if bounded by $O(\ell^{-q-2})$. So we conclude that

$$\int u(x,T)m_\ell(x-y)dx = \delta \sum u_h^N m_\ell(h\delta - y) + O(\ell^{-q-2}\delta^q).$$

Combining this with (8.64) we conclude that

$$u(y,T) = \delta \sum u_h^N m_\ell(h\delta - y) + O(\ell^q) + O(\ell^{-q-2}\delta^q). \tag{8.65}$$

We choose ℓ so that the two error terms in (8.65) are equal:

$$\ell^q = \ell^{-q-2}\delta^q, \quad \text{so} \quad \ell = \delta^{\frac{q}{2q+2}}$$

and (8.65) asserts that

$$u(y,T) = \delta \sum u_h^N m_\ell(h\delta - y) + O\big(\delta^{\frac{q^2}{2q+2}}\big).$$

For q large the error term approximates $\delta^{q/2}$.

Similar approximations can be obtained for derivatives of $u(y,T)$.

When y is located at a point of discontinuity, we can estimate $u(y+0,T)$ by choosing for m a $q+1$ times differentiable function supported on $[-1,1]$ that satisfies instead of (ii) the relations

$$\text{(ii}'\text{)} \quad \begin{aligned} &\int_{-1}^{0} m(x)dx = 0, \quad \int_0^1 m(x)dx = 1, \\ &\int_{-1}^{0} x^j m(x)dx = 0 = \int_0^1 x^j m(x)dx = 0, \quad j = 1, 2, \ldots, p-1. \end{aligned}$$

The case $q = 2$ has been done by Majda and Osher; they proved a pointwise estimate.

8.7. Schemes in More Than One Space Variable

We shall study difference approximations to symmetric hyperbolic systems in two and more space variables:

$$D_t u = AD_x u + BD_y u. \tag{8.66}$$

A and B are real, symmetric matrices that depend smoothly on x and y. We shall study difference approximations of the same form as before

$$u_h^{n+1} = \sum C_j u_{h+j}^n, \tag{8.67}$$

but this time h and j are multi-indices with $j = (j_1, j_2)$. Here u_h^n is an approximation of the solution sought at $(x,y) = h\delta$, $t = \epsilon n$. We derive the consistency conditions between (8.53) and (8.54) by approximating u_h^{n+1} and u_{h+j}^n using Taylor's formula, and using the differential equation (8.66) to relate the t-derivative to the x and y derivatives. The resulting relations are the two-dimensional analogues of equation (8.7):

$$\sum C_j = I, \quad \sum j_1 C_j = \lambda A, \quad \sum j_2 C_j = \lambda B, \tag{8.68}$$

where λ is the ratio of ϵ to δ.

The analogue of the LF scheme is to set equal to zero all coefficients C_j except those pertaining to the nearest neighbor of h; we denote these as C_{E}, C_{W}, C_{N}, and C_{S}, in obvious notation. Since these four matrix coefficients are restricted by only three equations, there is a redundancy. We choose, in analogy with (8.9),

$$\begin{aligned} C_{\mathrm{W}} &= \frac{1}{2}\left(\frac{1}{2}I - \lambda A\right), & C_{\mathrm{E}} &= \frac{1}{2}\left(\frac{1}{2}I + \lambda A\right), \\ C_{\mathrm{S}} &= \frac{1}{2}\left(\frac{1}{2}I - \lambda B\right), & C_{\mathrm{N}} &= \frac{1}{2}\left(\frac{1}{2}I + \lambda B\right). \end{aligned} \tag{8.69}$$

Choose λ so small that

$$\lambda|A| < \frac{1}{2}, \quad \lambda|B| < \frac{1}{2}; \tag{8.70}$$

then the four matrices C_{W}, C_{E}, C_{S}, and C_{N} are *positive*. It follows therefore from Theorem 8.1 that with this choice the scheme (8.69) is stable, provided that A and B are Lipschitz continuous functions of x and y.

Take the case that A and B are independent of x and y. In Chapter 3 we determined the convex hull of the domain of dependency K of the point $x = y = 0$, $t = 1$, on the initial plane $t = 0$ for solutions of the PDE (8.66). We found that the support function $s_K(\xi, \eta)$ of K is

$$s_K(\xi, \eta) = \max \sigma(\xi A + \eta B),$$

where $\sigma(M)$ denotes an eigenvalue of M. Take $\xi = \eta = \pm 1$; we conclude that

$$\begin{aligned} s_K(\pm 1, \pm 1) &\le |A + B|, \\ s_K(\mp 1, \pm 1) &\le |A - B|. \end{aligned} \tag{8.71}$$

The domain of dependence of the point (0, 0, 1) on the initial plane for solutions of the difference equation (8.67), (8.69) consists of the lattice points contained in the rectangle whose vertices are $(\pm n\delta, 0)$, $(0, \pm n\delta)$, where $n\epsilon = 1$; so the vertices are $(\pm\lambda^{-1}, 0)$, $(0, \pm\lambda^{-1})$. Denote by $s_\lambda(\xi, \eta)$ the support function of this rectangle; clearly

$$s_\lambda(\pm 1, \pm 1) = \lambda^{-1}, \quad s_\lambda(\mp 1, \pm 1) = \lambda^{-1}. \tag{8.71$_\delta$}$$

The CFL condition says that for a convergent scheme the domain of dependence of the difference scheme must include the domain of dependence of the differential equation. Comparing (8.71$_\delta$) with (8.71) this is satisfied if

$$|A + B| < \lambda^{-1}, \quad |A - B| < \lambda^{-1}. \tag{8.72}$$

Condition (8.70), sufficient for stability—and thereby for convergence— implies condition (8.72), but is more stringent.

NOTE. The most general four-point scheme satisfying the three consistency conditions (8.68) is of the form

$$C_W = \frac{1}{2}\left(\frac{1}{2}I + M - \lambda A\right), \qquad C_E = \frac{1}{2}\left(\frac{1}{2}I + M + \lambda A\right),$$
$$C_S = \frac{1}{2}\left(\frac{1}{2}I - M - \lambda B\right), \qquad C_N = \frac{1}{2}\left(\frac{1}{2}I - M + \lambda B\right),$$

M any symmetric matrix. The choice $M = \frac{1}{2}(A^2 - B^2)$ gives a stable scheme under condition (8.72).

We now turn to second-order schemes, analogues of the LW scheme. The second-order consistency conditions, analogues of (8.25), are, for $\lambda = 1$, as follows:

(8.73)
$$\sum C_j = I, \quad \sum j_1 C_j = A, \quad \sum j_2 C_j = B,$$
$$\sum j_1^2 C_j = A^2, \quad \sum j_2^2 C_j = B^2, \quad \sum j_1 j_2 C_j = AB + BA.$$

These are six conditions; they can no longer be satisfied by setting $C_j = 0$ except for $j = 0$ and the four nearest neighbors. We have to use all eight neighbors, which allows for many choices. But, just as in the one-dimensional case, see Theorem 8.4, it is impossible to satisfy (8.73) with all positive C_j, except in trivial cases.

A straightforward way of constructing a scheme that is second-order accurate is to expand $u(x, t + \delta)$ into a Taylor series,

$$u(x, t + \delta) = u(x, t) + D_t u(x, t)\delta + \frac{1}{2}D_t^2 u(x, t)\delta^2 + O(\delta^3),$$

and then use differential equation (8.66) to express the t derivatives in terms of x, y derivatives:

$$u(x, t + \delta) = u(x, t) + ((AD_x + BD_y)u)\delta$$
$$+ \frac{1}{2}(A^2 D_x^2 + (AB + BA)D_x D_y + B^2 D_y^2)u + O(\delta^3).$$

We approximate $D_x u$ by $(u_E - u_W)/2\delta$, $D_y u$ by $(u_N - u_S)/2\delta$, $D_x^2 u$ by $(u_E - 2u_0 + u_W)/2\delta^2$, $D_y^2 u$ by $(u_N - 2u_n + u_S)/2\delta^2$, and $D_x D_y u$ by $(u_{NE} - u_{NW} - u_{SE} + u_{SW})/4\delta^2$. This determines the nine coefficients C_j in the two-dimensional LW difference scheme. The *symbol* $C(\theta, \varphi)$ of the scheme, defined as

$$C(\theta, \varphi) = \sum C_j e^{ij\cdot(\theta,\varphi)}$$

is easily calculated:

(8.74)
$$C(\theta, \varphi) = I - A^2(1 - \cos\theta) - B^2(1 - \cos\varphi)$$
$$- \frac{1}{2}(AB + BA)\sin\theta\sin\varphi + i(A\sin\theta + B\sin\varphi).$$

Compare this with the symbol of the one-dimensional LW scheme given by formula (8.35). The main difference is that in the two-dimensional case both the Hermitian and anti-Hermitian parts of C are polynomials in both A and B. Since

A and B do not in general commute, the matrix function $C(\theta, \varphi)$ is *not* a normal matrix. Consequently, the von Neumann necessary condition, as explained in Corollary 8.6, that the eigenvalues of $C(\theta, \varphi)$ be ≤ 1 in absolute value, no longer implies that the norm of $C(\theta, \varphi)$, given in (8.74), is ≤ 1. And, in fact, the norm of $C(\theta, \varphi)$ is *not* less than one. However, Burt Wendroff and the author showed the uniform boundedness of the powers of $C(\theta, \varphi)$ by employing the notion of the *numerical range* of a matrix, defined as follows:

DEFINITION The *numerical range* of a matrix M is the set of all complex numbers of the form $Mw \cdot w$ as w takes on all unit vectors with *complex* entries.

The properties of the numerical range that we need are contained in

THEOREM 8.18

(i) *The numerical range of a Hermitian symmetric matrix is the interval on the real axis between its smallest and largest eigenvalues.*
(ii) *The numerical range of any matrix M includes all its eigenvalues.*
(iii) *If the numerical range of M lies in the unit disk, then the norm of M is ≤ 2.*
(iv) *If the numerical range of M lies in the unit disk, so does the numerical range of all its powers M^n.*

PROOF:

(i) is just the variational characterization of the smallest and largest eigenvalues of a Hermitian matrix.
(ii) follows if we choose w to be a normalized eigenvector of M; $Mw \cdot w = \sigma w \cdot w = \sigma$.
(iii) The real and imaginary parts of the numerical range of M are the numerical range of the Hermitian and anti-Hermitian parts of M. Therefore by part (i), the Hermitian and anti-Hermitian parts of M have norm ≤ 1, from which $|M| \leq 2$ follows.
(iv) is the celebrated Halmos-Berger-Pearcy theorem. □

NOTE. Hausdorff and Toeplitz proved that the numerical range of any matrix is a closed, convex subset of $\mathbb{C}$; we shall not need this interesting result in what follows.

THEOREM 8.19 *The numerical range of the symbol $C(\theta, \varphi)$ of the two-dimensional LW scheme* (8.74) *lies in the unit disk, provided that*

$$|A| \leq \frac{1}{8}, \quad |B| \leq \frac{1}{8}.$$

PROOF: Separate C into its Hermitian and anti-Hermitian parts R and J:

$$C = R + iJ,$$

where

$$R = I - K,$$

$$K = A^2(1 - \cos\theta) + B^2(1 - \cos\varphi) + \frac{1}{2}(AB + BA)\sin\theta \sin\varphi,$$

and

(8.75) $$J = A\sin\theta + B\sin\varphi\,.$$

Using the abbreviation

$$1-\cos\theta = e\,,\ 1-\cos\varphi = f\,,$$

we can write K as

(8.76) $$K = \frac{1}{2}A^2e^2 + \frac{1}{2}B^2f^2 + \frac{1}{2}J^2\,.$$

Our aim is to estimate $Cw\cdot w$ for any unit vector w:

$$Cw\cdot w = Rw\cdot w + iJw\cdot w = r + ij\,;$$

r and j are real, therefore

$$|Cw\cdot w|^2 = r^2 + j^2\,.$$

By the Schwarz inequality

$$j^2 = (Jw\cdot w)^2 \le |Jw|^2\,.$$

Since $R = I - K$, $r = Rw\cdot w = 1 - Kw\cdot w$. Using formula (8.76) we get

$$r = 1 - \frac{1}{2}|Aw|^2e^2 - \frac{1}{2}|Bw|^2f^2 - \frac{1}{2}|Jw|^2\,.$$

Squaring this gives

$$r^2 = 1 - a^2e^2 - b^2f^2 - |Jw|^2 + (Kw\cdot w)^2\,,$$

where we have used the abbreviations

(8.77) $$|Aw| = a\,,\ |Bw| = b\,.$$

Adding to this the above estimate for j, we get

(8.78) $$r^2 + j^2 \le 1 - a^2e^2 - b^2f^2 + |Kw\cdot w|^2\,.$$

Using the original definition of K, we have

$$Kw\cdot w = a^2e + b^2f + \operatorname{Re} Aw\cdot Bw\sin\theta\sin\varphi\,.$$

Estimating the last term by the Schwarz inequality gives

$$|Kw\cdot w| \le a^2e + b^2f + ab\sin\theta\sin\varphi\,.$$

Estimate the last term by the arithmetic-geometric inequality:

$$|Kw\cdot w|^2 \le a^2e + b^2f + \frac{a^2\sin^2\theta + b^2\sin^2\varphi}{2}\,.$$

Recalling that $e = 1-\cos\theta$, $f = 1-\cos\varphi$, $\frac{\sin^2\theta}{2} \le 1-\cos\theta$, and $\frac{\sin^2\varphi}{2} \le 1-\cos\varphi$, we get

$$|Kw\cdot w| \le 2a^2e + 2b^2f\,;$$

squaring this gives

$$|Kw\cdot w|^2 \le 8a^4e^2 + 8b^4f^2\,.$$

Setting this into (8.78) gives

$$r^2 + j^2 \le 1 - a^2(1 - 8a^2)e^2 - b^2(1 - 8b^2)f^2.$$

The right side is ≤ 1 if $8a^2$ and $8b^2$ are ≤ 1. Recalling the definition (8.77) of a and b, these inequalities amount to $|A| \le \frac{1}{8}$, $|B| \le \frac{1}{8}$; this completes the proof of Theorem 8.19. □

We now combine parts (iii) and (iv) of Theorem 8.18 to conclude that if A and B have norms $\le \frac{1}{8}$, all powers of the 2D LW scheme are ≤ 2. We now appeal to Theorem 8.5 to conclude that when A and B are independent of x and y, the 2D LW scheme is stable, and so by Theorem 8.3 convergent, with an approximation error $O(\delta^2)$.

As already remarked, the six consistency conditions (8.73) for second-order accuracy do not determine the nine coefficients C_j uniquely. It is easy to verify that

$$C'_{\mathrm{NE}} = C'_{\mathrm{NW}} = C'_{\mathrm{SE}} = C'_{\mathrm{SW}} = -\frac{M}{4},$$

$$C'_{\mathrm{N}} = C'_{\mathrm{W}} = C'_{\mathrm{S}} = C'_{\mathrm{E}} = \frac{M}{2},$$

$$C'_{O} = -M,$$

satisfy the *homogeneous* equations (8.73). The symbol of this scheme is

$$-M(1 - \cos\theta)(1 - \cos\varphi).$$

COROLLARY 8.20 *The numerical range of* $C(\theta, \varphi) - \frac{A^2+B^2}{2}(1 - \cos\theta)(1 - \cos\varphi)$, *where* $C(\theta, \varphi)$ *is defined by* (8.74), *lies in the unit disk, provided that* $A^2 + B^2 \le \frac{l}{2}$.

EXERCISE Verify Corollary 8.20.

It follows, as before, that the modified LW scheme whose symbol is given in Corollary 8.20 is stable, and therefore convergent, when A and B are independent of x and y, and satisfy $A^2 + B^2 \le \frac{l}{2}$.

The challenging task is to prove the convergence of the schemes discussed above when A and B are functions of x and y. We face this issue in the next section.

8.8. The Stability of Difference Schemes

As mentioned, one difficulty in proving the stability of difference schemes in several space variables is that the symbol of the schemes is no longer a normal operator. Another difficulty is that nonnegative functions of many variables may not have representations as squares, or even as sums of squares of smooth functions. So a new method of proof is needed.

It is analytically very convenient—although artificial from the numerical point of view—to make the difference operators act on functions of continuous spatial variables. Thus the approximate solution u^n at time $t = n\delta$ is a function of the space variables x in $\mathbb{R}^k$, and the operator relating u^n to u^{n+1} is of the form

$$u^{n+1} = \left(\sum C_j(x)T_\delta^j\right)u^n, \tag{8.79}$$

where T_δ^j is the translation operator

$$\left(T_\delta^j u\right)(x) = u(x + j\delta).$$

The summation in (8.79) is over a finite number of multi-indices j. The *symbol* of the scheme is defined as

$$C(\xi) = \sum C_j(x)e^{ij\xi};$$

ξ is a multivariable $\xi_1, \dots, \xi_k$, and $j\xi = j_1\xi_1 + \cdots + j_k\xi_k$.

The key estimate, due to Louis Nirenberg and the author, is

THEOREM 8.21 *Let P_δ be a difference operator of the form*

$$P_\delta = \sum_{|j|\leq N} P_j(x)T_\delta^j,$$

$P_j(x)$ symmetric matrix functions that depend twice differentiably on x. Suppose that the symbol of P_δ,

$$P(x,\xi) = \sum P_j(x)e^{ij\xi},$$

is Hermitian and nonnegative:

$$P(x,\xi) \geq 0.$$

Then

$$\operatorname{Re} P_\delta \geq -\text{const}\,\delta,$$

where Re P_δ *denotes the Hermitian part of P_δ:*

$$2\operatorname{Re} P_\delta = P_\delta + P_\delta^*,$$

*where * denotes the adjoint in the Hilbert spaces $L^2(\mathbb{R}^k)$.*

NOTE. Unlike the one-dimensional case, where the coefficients of the scheme had to be merely Lipschitz continuous, here the coefficients of the scheme—and therefore of the hyperbolic equation whose solution they approximate—have to be twice differentiable.

PROOF: The first step is to localize the problem. Let $\{\varphi_j(x)\}$ be a Gårding type partition of unity:

$$\sum \varphi_h^2(x) \equiv 1, \tag{8.80}$$

where the support of each φ_h has diameter $O(\sqrt{\delta})$. The first derivatives of the φ_h are then $O(1/\sqrt{\delta})$, and so

$$|\varphi_h(x) - \varphi_h(y)| \leq O(1/\sqrt{\delta})|x - y|. \tag{8.81}$$

Our aim is to estimate from below the real part of $(P_\delta u, u)$; by definition

$$(P_\delta u, u) = \int \sum P_j(x)u(x + j\delta) \cdot u(x)dx.$$

LEMMA 8.22 *Define* $u_h = \varphi_h u$; *then* $\sum(P_\delta u_h, u_h)$ *differs by* $O(\delta)\|u\|^2$ *from* $(P_\delta u, u)$ *for any* u *in* L^2.

PROOF:

$$(8.82)\quad (P_\delta u, u) - \sum_h (P_\delta u_h, u_h) = \int \sum_j P_j(x)u(x+j\delta)\cdot u(x)\Big(1 - \sum_h \varphi_h(x+j\delta)\varphi_h(x)dx\Big).$$

Using the identity (8.80) we see that

$$(8.83)\qquad 1 - \sum_h \varphi_h(x+j\delta)\varphi_h(x) = \frac{1}{2}\sum_h (\varphi_h(x+j\delta) - \varphi_h(x))^2 .$$

Using (8.81), and the fact that at any point y only a finite number of $\varphi_h(y)$ are $\neq 0$, we conclude that the sum (8.83) is $O(j^2\delta)$. Set this into (8.82), and Lemma 8.22 follows.

□

LEMMA 8.23 *Let* P_δ *be a difference operator whose symbol* $P(\xi)$ *is Hermitian and independent of* x. *Denote by* $\varphi(x)$ *be a real scalar function with Lipschitz constant* K. *Then for all* u *in* L^2

$$(8.84)\qquad |\mathrm{Re}(\varphi P_\delta u, \varphi u) - \mathrm{Re}(P_\delta \varphi u, \varphi u)| \le O(K^2\delta^2)\|u\|^2 .$$

PROOF: Since $P(\xi)$ is Hermitian, its coefficients satisfy $P_{-j} = P_j^*$. The amount A inside the absolute value sign in (8.84) can be written as

$$(8.85)\qquad A = \mathrm{Re}\int \sum P_j u(x+j\delta)\cdot u(x)(\varphi(x) - \varphi(x+j\delta))\varphi(x)dx .$$

Replacing $x + j\delta$ by x as a variable of integration, and replacing j by $-j$ we get

$$A = \mathrm{Re}\int \sum P_j^* u(x)\cdot u(x+j\delta)(\varphi(x+j\delta) - \varphi(x))\varphi(x+j\delta)dx ,$$

where we have used $P_{-j} = P_j^*$. Since

$$P_j^* u(x)\cdot u(x+j\delta) = u(x)\cdot P_j u(x+j\delta),$$

$$(8.85')\qquad A = \mathrm{Re}\int \sum u(x)\cdot P_j u(x+j\delta)(\varphi(x+j\delta) - \varphi(x))\varphi(x+j\delta)dx$$

The factors involving u in (8.85) and (8.85′) are complex conjugates of each other. Adding them gives

$$A = \frac{1}{2}\mathrm{Re}\int \sum P_j u(x+j\delta)\cdot u(x)(\varphi(x) - \varphi(x+j\delta))^2 dx + O(K\delta^2),$$

from which we deduce that

$$|A| \le O(K^2\delta^2)\|u\|^2 .$$

□

The next result is a version of the Schwarz inequality:

LEMMA 8.24 *Let M and N be Hermitian matrices, and assume that for all vectors v*

$$|Mv \cdot v| \le Nv \cdot v\,. \tag{8.86}$$

Then for any pair of vectors v, w

$$|Mv \cdot w| \le (Nv \cdot v)^{\frac{1}{2}}(Nw \cdot w)^{\frac{1}{2}}\,. \tag{8.86'}$$

PROOF:

$$(N+M)(v+w)\cdot(v+w) \ge 0\,,$$
$$(N-M)(v-w)\cdot(v-w) \ge 0\,.$$

Add and divide by 2:

$$Nv\cdot v + 2\,\mathrm{Re}\,Mv\cdot w + Nw\cdot w \ge 0\,.$$

Replace w by $te^{i\varphi}w$, t real, and argue as usual about the discriminant of a nonnegative quadratic function. □

The next lemma is the local version of Theorem 8.21:

LEMMA 8.25 *P_δ is an operator as in* Theorem 8.21, *and v a function whose support has diameter $O(\sqrt{\delta})$. Then*

$$\mathrm{Re}(P_\delta v, v) \ge O(\delta)\|v\|^2\,.$$

PROOF: To simplify notation, suppose the support of v is centered at the origin. Then

$$(P_\delta v, v) = \int_{|x|<c\sqrt{\delta}} \sum P_j(x)v(x+j\delta)\cdot v(x)dx\,.$$

Since $P_j(x)$ is assumed twice differentiable, we approximate it by two terms of its Taylor series:

$$P_j(x) = P_j(0) + \sum P_j^\ell(0)x_\ell + O(|x|^2)\,;$$

so for $|x| < c\sqrt{\delta}$

$$\left|P_j(x) - P_j(0) - \sum P_j^\ell(0)x_\ell\right| \le O(\delta)\,.$$

Denote by P_δ^0 and P_δ^ℓ the operators

$$P_\delta^0 = \sum P_j(0)T^j\,, \quad P_\delta^\ell = \sum P_j^\ell(0)T^j\,.$$

For all v supported in $|x| \le c\sqrt{\delta}$,

$$(P_\delta v, v) = \left(P_\delta^0 v, v\right) + \sum\left(x_\ell P_\delta^\ell v, v\right) + O(\delta)\|v\|^2.$$

So to prove Lemma 8.25 we have to show that

$$Q = \left(P_\delta^0 v, v\right) + \mathrm{Re}\sum_\ell\left(x_\ell P_\delta^\ell v, v\right) \ge -O(\delta)\|v\|^2\,. \tag{8.87}$$

Since P_δ^0 and P_δ^ℓ are operators with constant coefficients, we can use Parseval's formula for the Fourier transform to express Q :

$$(P_\delta^0 v, v) = \int P^0(\xi)\tilde{v}(\xi)\cdot\tilde{v}(\xi)d\xi\,, \tag{8.88}$$

where $P^0(\xi)$ is the symbol of P_δ^0, and $\tilde{v}(\xi)$ the Fourier transform of v:

$$\tilde{v}(\xi) = \int v(x)e^{-ix\xi}\bar{d}x\,.$$

Similarly, denoting $x_\ell v$ by v^ℓ we can write

$$(x_\ell P^\ell v, v) = \int P^\ell(\xi)\tilde{v}(\xi)\cdot\tilde{v}^\ell d\xi\,. \tag{8.88'}$$

We look now at the Taylor approximation to the symbol of P_δ:

$$P(x,\xi) = P^0(\xi) + \sum x_\ell P^\ell(\xi) + O(x^2)\,.$$

By hypothesis, $P(x,\xi)\ge 0$; setting $x_\ell = \pm\sqrt{\delta}$, all the other $x_m = 0$ into the Taylor approximation we deduce that

$$P^0(\xi) \pm \sqrt{\delta}P^\ell(\xi) \ge O(\delta)\,.$$

Adding $O(\delta)$ to $P^0(\xi)$ alters Q only by $O(\delta)\|v\|^2$ and is acceptable; consider it done. That turns the above inequality into

$$P^0(\xi) \pm \sqrt{\delta}P^\ell(\xi) \ge 0\,.$$

We apply now Lemma 8.24, with $M = \sqrt{\delta}P^\ell(\xi)$, $N = P^0(\xi)$, and conclude that

$$\sqrt{\delta}\,\left|P^\ell\tilde{v}\cdot\tilde{v}^\ell\right| \le (P^0\tilde{v}\cdot\tilde{v})^{\frac{1}{2}}(P^0\tilde{v}^\ell\cdot\tilde{v}^\ell)^{\frac{1}{2}}\,.$$

We estimate the right side by the arithmetic-geometric inequality:

$$\left|P^\ell\tilde{v}\cdot\tilde{v}^\ell\right| \le \frac{1}{2m}\,P^0\tilde{v}\cdot\tilde{v} + \frac{m}{2\delta}\,P^0\tilde{v}^\ell\cdot\tilde{v}^\ell\,,$$

where m is the number of components of the vector v. Set (8.88) and (8.88′) into the definition (8.87) of Q, and use the above inequality for the integrand in the second term. We get the following lower bound for Q:

$$\begin{aligned} Q &\ge (P_\delta^0 v, v) - \sum_\ell\left[\frac{1}{2m}\,(P_\delta^0 v, v) + \frac{m}{2\delta}\,(P_\delta^0 v^\ell, v^\ell)\right] \\ &= \frac{1}{2}\,(P_\delta^0 v, v) - \frac{m}{2\delta}\sum (P_\delta^0 v^\ell, v^\ell)\,; \end{aligned} \tag{8.89}$$

v^ℓ is defined as $x_\ell v$. According to Lemma 8.23, with $\varphi = x_\ell$,

$$(P_\delta^0 v^\ell, v^\ell) = (P_\delta^0 x_\ell v, x_\ell v) = (P_\delta^0 v, x_\ell^2 v) + O(\delta^2)\|v\|^2\,,$$

Summing over all ℓ, and denoting $\sum x_\ell^2 = r^2$ we set this in the second term of (8.89):

$$
\begin{aligned}
Q &\geq \frac{1}{2}\left(P_\delta^0 v, v\right) - \frac{1}{2}\operatorname{Re}\left(P_\delta^0 v, \frac{mr^2}{\delta} v\right) + O(\delta)\|v\|^2 \\
&= \frac{1}{2}\operatorname{Re}\left(P_\delta^0 v, \left(1 - \frac{mr^2}{\delta}\right) v\right) + O(\delta)\|v\|^2 .
\end{aligned} \tag{8.90}
$$

The support of the function v is contained in the ball $|x| \leq c\sqrt{\delta}$. Choose c to be $1/\sqrt{2m}$, and introduce the function φ defined as

$$
\varphi(r) = \begin{cases} \sqrt{1 - mr^2/\delta} & \text{for } r < c\sqrt{\delta} \\ \sqrt{1/2} & \text{for } r \geq c\sqrt{\delta} . \end{cases}
$$

In terms of φ we can rewrite (8.90) as

$$
Q \geq \frac{1}{2}\operatorname{Re}\left(P_\delta^0 v, \varphi^2 v\right) + O(\delta)\|v\|^2. \tag{8.91}
$$

The function φ is Lipschitz continuous with Lipschitz constant $K = O(1/\sqrt{\delta})$. So according to Lemma 8.23

$$
\begin{aligned}
\left(P_\delta^0 v, \varphi^2 v\right) &= \left(P_\delta^0 \varphi v, \varphi v\right) + O(K^2\delta^2)\|v\|^2 \\
&= \left(P_\delta^0 \varphi v, \varphi v\right) + O(\delta)\|v\|^2 .
\end{aligned}
$$

P_δ^0 is a difference operator with constant coefficients whose symbol is positive. Therefore

$$
\left(P_\delta^0 \varphi v, \varphi v\right) = \int P(0,\xi)\widetilde{\varphi v} \cdot \widetilde{\varphi v}\, d\xi
$$

is positive. Setting this into (8.91) we conclude that

$$
Q \geq O(\delta)\|v\|^2.
$$

But we have seen in (8.87) that this lower bound for Q implies Lemma 8.25. □

According to Lemma 8.22

$$
(P_\delta u, u) \geq \sum (P_\delta u_h, u_h) + O(\delta)\|u_h\|^2,
$$

where $u_h = \varphi_h u$. By Lemma 8.25 applied to $v = u_h$

$$
(P_\delta u_h, u_h) \geq O(\delta)\|u_h\|^2 ;
$$

setting this into the estimate above we get

$$
(P_\delta u, u) \geq O(\delta) \sum \|u_h\|^2 = O(\delta)\|u\|^2,
$$

where in the last step we have used (8.80). This completes the proof of Theorem 8.21. □

We deduce now two important stability theorems from Theorem 8.21.

THEOREM 8.26 *The difference scheme*

$$u^{n+1} = \Big(\sum_{|j|\le N} C_j(x)T_\delta^j\Big)u^n \tag{8.79}$$

is stable, provided that its symbol $C(x,\xi)$ satisfies these conditions:

(i) *$C(x,\xi)$ is a twice differentiable function of x.*

(ii) *$|C(x,\xi)| \le 1$ for all x and ξ.*

PROOF: We shall show that the norm of C_δ is $\le 1 + O(\delta)$, C_δ being the operator relating u^n to u^{n+1}. This is sufficient for stability.

$$\|C_\delta u\|^2 = (C_\delta u, C_\delta u) = (u, C_\delta^* C_\delta u)\,.$$

Since the coefficients of C_δ are Lipschitz continuous, $C_\delta^* C_\delta$ differs by $O(\delta)$ from the operator whose symbol is $C^*(x,\xi)C(x,\xi)$. Define

$$P(x,\xi) = I - C^*(x,\xi)C(x,\xi)$$

and denote by P_δ the operator whose symbol is $P(x,\xi)$. By assumption (ii), $C^*C \le I$, therefore $P(x,\xi) \ge 0$, and so by Theorem 8.21, Re $P_\delta \ge O(\delta)$,

$$\|u\|^2 - \|C_\delta u\|^2 = (u, (I - C_\delta^* C_\delta)u) = (u, P_\delta u) + O(\delta)\|u\|^2\,.$$

The left side is real; the real part of the right side is $\ge O(\delta)\|u\|^2$, therefore $\|C_\delta u\|^2 \le (1 + O(\delta))\|u\|^2$. □

NOTE. Theorem 8.26 should be compared to Theorem 8.10. Here the symbol is assumed twice differentiable in x, in the earlier one only once. On the other hand, there is no need for an analogue of condition (iv).

THEOREM 8.27 *The difference scheme*

$$u^{n+1} = \Big(\sum_{|j|\le N} C_j(x)T_\delta^j\Big)u^n \tag{8.79}$$

is stable provided that its symbol $C(x,\xi)$ satisfies these conditions:

(i) *$C(x,\xi)$ is twice differentiable function of x.*

(ii) *The numerical range of $C(x,\xi)$ lies in the unit disk for all x,ξ.*

PROOF: We shall show that the numerical range of the operator C_δ lies inside a disk of radius $1 + O(\delta)$ centered at the origin in the complex plane. What we have to show is that for all u in L^2,

$$|(C_\delta u, u)| \le (1 + O(\delta))\|u\|^2\,.$$

This is equivalent to the following: for all complex numbers z, $|z| = 1$,

$$\operatorname{Re} z(C_\delta u, u) \le (u,u) + O(\delta)\|u\|^2.$$

This can be written as

$$\operatorname{Re}((I - zC_\delta)u, u) \ge O(\delta)\|u\|^2\,. \tag{8.92}$$

The symbol of $I - zC_\delta$ is $I - zC(x, \xi)$; it follows from assumption (ii) that the real part of $I - zC(x, \xi)$ is nonnegative. Therefore, by Theorem 8.21 the Hermitian part of the operator whose symbol is $I - zC(x, \xi)$ is $\geq O(\delta)$; this proves (8.92).

We appeal now to part (iv) of Theorem 8.18, The Halmos-Berger-Pearcy theorem. which says that if the numerical range of an operator C_δ in Hilbert space lies in a disk of radius $1 + O(\delta)$, then the numerical range of its powers C_δ^n lies in a disk of radius $(1 + O(\delta))^n$. It follows from then by part (iii) of Theorem 8.18 that

$$\|C_\delta^n\| \leq 2(1 + O(\delta))^n .$$

This proves the stability of the scheme (8.79). □

NOTE. It follows from Theorem 8.27, with the help of Theorem 8.19 and Theorem 8.11, that the LW scheme for the symmetric hyperbolic equation $D_t u = AD_x u + BD_y u$ is stable, provided that the coefficients A and B are twice differentiable functions of x and y, are independent of t, and $|A| \leq \frac{1}{8}$, $|B| \leq \frac{1}{8}$.

When A and B depend on t as well, the proof presented above breaks down; however, numerical experience indicates that the LW scheme is stable even in the time dependent case.

Theorems 8.26 and 8.27 are not the last word in stability theory. To go beyond them we need to introduce *nonlocal* difference operators of the form

$$W_\delta = \sum W_j T^j, \tag{8.93}$$

where W_j are not zero for $|j| > N$, but *tend to zero rapidly* as $|j| \to \infty$. This is the same as saying that

$$W(\xi) = \sum W_j e^{ij\xi} \tag{8.93'}$$

is a *smooth* periodic function if ξ. We discuss first the simpler case when the W_j—and therefore $W(\xi)$—are independent of x.

We take the case that, for each ξ, $W(\xi)$ is a symmetric positive definite matrix. We define the W_δ-norm of a function u in L^2 by

$$\|u\|_W^2 = (W_\delta u, u) . \tag{8.94}$$

Denote as before by $\tilde{u}(\xi)$ the Fourier transform of u. Then

$$\widetilde{W_\delta u} = W(\delta\xi)\tilde{u}(\xi) ;$$

by Parseval's theorem

$$\|u\|_W^2 = (W_\delta u, u) = \int W(\delta\xi)\,\tilde{u} \cdot \tilde{u}\, d\xi . \tag{8.94'}$$

By hypothesis, for all ξ

$$c_1^2 I \leq W(\xi) \leq c_2^2 I , \tag{8.95}$$

c_1, c_2 positive constants. So it follows that

$$c_1\|u\| \leq \|u\|_W \leq c_2\|u\| . \tag{8.95'}$$

In words: ***the L^2 norm and the W_δ-norm are equivalent.***

It follows from formula (8.94′) that if $W(0) = I$, $\lim_{\delta\to 0} \|u\|_W = \|u\|$.

It follows from the equivalence of the two norms that stability of a difference scheme in the W-norm implies its stability in the L^2 norm.

Denote by C_δ a difference operator $C_\delta u = v$ of the form

$$v_h = \sum C_j u_{h+j} . \tag{8.96}$$

Denote by $C(\xi)$ the symbol of C_δ:

$$C(\xi) = \sum C_j e^{ij\xi} . \tag{8.96'}$$

THEOREM 8.28 *Denote by $C(\xi)$ the symbol of a difference scheme with constant coefficients. Suppose there exists a smooth, periodic function $W(\xi)$ whose values are symmetric positive definite matrices that satisfies*

$$C^*(\xi)W(\xi)c(\xi) \leq W(\xi). \tag{8.97}$$

Then the scheme is stable.

PROOF: We apply formula (8.94′) to $v = C_\delta u$;

$$\|v\|_W^2 = \int W(\delta\xi)\tilde{v}(\xi) \cdot \tilde{v}(\xi) d\xi ,$$

where

$$\tilde{v}(\xi) = \int v(x)e^{-ix\xi}dx = \sum \int C_j u(x+\delta j)e^{-ix\xi}dx$$
$$= \int \sum C_j u(y)e^{-iy\xi+i\delta j\xi}dy = C(\delta\xi)\tilde{u}(\xi) .$$

Setting this into the formula for $\|v\|_W^2$ we get

$$\|v\|_W^2 = \int W(\delta\xi)C(\delta\xi)\tilde{u}(\xi) \cdot C(\delta\xi)\tilde{u}(\xi)d\xi$$
$$= \int C^*(d\xi)W(\delta\xi)C(\delta\xi)\tilde{u}(\xi) \cdot \tilde{u}(\xi)d\xi .$$

Using inequality (8.97) we deduce that

$$\|C_\delta u\|_W \leq \|u\|_W .$$

It follows that the W_δ-norm of any power of C_δ is ≤ 1:

$$\left\|C_\delta^n u\right\|_W \leq \|u\|_W .$$

It follows from (8.95′) that

$$\left\|C_\delta^n u\right\| \leq c_1^{-1}\left\|C_\delta^n u\right\|_W$$
$$\leq c_1^{-1}\|u\|_W \leq c_1^{-1}c_2\|u\| .$$

This proves the stability of the scheme C_δ. □

We sketch now how to extend Theorem 8.28 to schemes with variable coefficients.

THEOREM 8.29 *Denote by $C(x,\xi)$ the symbol of the difference scheme*

$$C_\delta = \sum C_j(x)T^j.$$

Suppose that there exists a function $W(x,\xi)$ with the following properties:

(i) *The values of W are symmetric, positive definite matrices.*

(ii) *$W(x,\xi)$ is a twice differentiable function of x.*

(iii) *$W(x,\xi)$ is a smooth, periodic function of ξ.*

(iv) *$C^*(x,\xi)W(x,\xi)C(x,\xi) \le W(x,\xi)$ for all x and ξ.*

Conclusion: the scheme C_δ is stable.

PROOF: We shall use repeatedly the following:

LEMMA 8.30

(i) *Let A_δ and B_δ denote difference operators*

$$A_\delta = \sum A_j(x)T^j, \quad B_\delta = \sum B_j(x)T,$$

whose symbols

$$A(x,\xi) = \sum A_j(x)e^{ij\xi}, \quad B(x,\xi) = \sum B_j(x)e^{ij\xi},$$

are smooth function of x and ξ. Then the operator whose symbol is $A(x,\xi)B(x,\xi)$ differs by $O(\delta)$ from the operator product $A_\delta B_\delta$.

(ii) *The operator whose symbol is $A^*(x,\xi)$ differs by $O(\delta)$ from the operator A_δ^*.*

We leave it to the reader to find a proof for this lemma.

We define the W_δ norm as follows: denote by $V(x,\xi)$ the positive square root of W, and by V_δ the associated difference operator. Define

$$\|u\|_W = \|V_\delta u\|. \tag{8.98}$$

Then

$$\|u\|_W^2 = (V_\delta u, V_\delta u) = (u, V_\delta^* V_\delta u).$$

Denote by W_δ the operator whose symbol is $W(x,\xi)$. The product $V_\delta^* V_\delta$ differs by $O(\delta)$ from the operator W_δ, according to Lemma 8.30; so

$$\|u\|_W^2 = (u, W_\delta u) + O(\delta)\|u\|^2. \tag{8.98'}$$

Now we are ready to show the equivalence of the W_δ-norm and the L^2 norm for small δ. It follows from formula (8.98), since V_δ is an operator bounded independently of δ, that $\|u\|_W \le \text{const}\|u\|$. To derive an inequality in the opposite direction we recall the lower bound $c_1^2 I \le W(\xi)$ in (8.95). We choose a positive number $c < c_1^2$, and decompose W as $W = cI + U^2$, where U is the positive square root of $W - cI$. Since c is less than the lower bound of W, U is a smooth function of ξ and a twice differentiable function of x; the values of U are symmetric matrices. Using Lemma 8.30 we get the following decomposition of W_δ:

$$W_\delta = cI + U_\delta^2 + O(\delta),$$

and

$$U_\delta^* = U_\delta + O(\delta)\,.$$

Setting these in (8.98′) we get

$$\|u\|_W^2 = c\|u\|^2 + \left(u,\, U_\delta^2 u\right) + O(\delta)\,. \tag{8.99}$$

The second term on the right can be written as

$$\left(U_\delta^* u,\, U_\delta u\right) = (U_\delta u,\, U_\delta u) + O(\delta) = \|U_\delta u\|^2 + O(\delta).$$

Setting this into (8.99) we conclude that

$$\|u\|_W^2 \geq c\|u\|^2 + O(\delta)\,;$$

so for δ small enough $\|u\|_W \geq (c/2)\|u\|^2$ follows. This completes the proof of the equivalence of the W_δ-norm with the L^2 norm.

We turn now to proving the stability of C_δ in the W_δ-norm. We need the full power of Theorem 8.21, as presented in Lax and Nirenberg, where $P(x,\xi)$ need not be a trigonometric polynomial in ξ, but is merely required to be a smooth function of ξ; being twice differentiable is enough for the conclusion of Theorem 8.21, namely that if $P(x,\xi)$ is symmetric and nonnegative, then Re $P_\delta \geq -K\delta$.

We apply this result to

$$P = W - C^* W C,$$

by assumption (iv) of Theorem 8.29 nonnegative. We write, using (8.98′)

$$\begin{aligned}\|u\|_W^2 - \|C_\delta u\|_W^2 &= (u,\, W_\delta u) - (C_\delta u,\, W_\delta C_\delta u) + O(\delta)\\ &= (u,\, (W_\delta - C_\delta^* W_\delta C_\delta)u) + O(\delta)\\ &= (u,\, P_\delta u) + O(\delta)\,.\end{aligned}$$

The left side is real, and the real part of the right side is $\geq O(\delta)\|u\|^2$. Using the equivalence of the W_δ-norm and the L^2 norm we deduce that

$$\|C_\delta u\|_W \leq (1 + O(\delta))\|u\|_W\,.$$

But then

$$\left\|C_\delta^n u\right\|_W \leq (1 + O(\delta))^n \|u\|_W\,.$$

For $n \leq T/\delta$, all the operators C_δ^n are uniformly bounded in the W_δ-norm. But then, since the W_δ-norm is equivalent with the L^2 norm, we conclude as in the proof of Theorem 8.28 that C_δ is stable in the L^2 norm. □

For what schemes C_δ is there a $W(x,\xi)$ that satisfies the hypotheses of Theorem 8.29 is an open question. Note that a necessary condition is that the eigenvalues of $C(x,\xi)$ not exceed 1 in absolute value.

References

Berger, C. A. A strange dilation theorem. *Notices Amer. Math. Soc.* 12(5): 590, 1965. Abstract 625-152.

Brenner, P., Thomée, V., and Wahlbin, L. B. *Besov spaces and applications to difference methods for initial value problems.* Lecture Notes in Mathematics, 434. Springer, Berlin–New York, 1975.

Courant, R., Friedrichs, K. O., and Lewy, H. Über die pertiellen Differenzengleichungen der mathematischen Physik. *Math. Ann.* 100: 32–74, 1928.

Friedrichs, K. O. Symmetric hyperbolic linear differential equations. *Comm. Pure Appl. Math.* 7: 345–392, 1954.

Halmos, P. R. *A Hilbert space problem book.* Van Nostrand, Princeton, N.J.–Toronto–London, 1967.

Kato, T. Some mapping theorems for the numerical range. *Proc. Japan Acad.* 41: 652–655, 1965.

Lax, P. D. On the stability of difference approximations to solutions of hyperbolic equations with variable coefficients. *Comm. Pure Appl. Math.* 14: 497–520, 1961.

Lax, P. D., and Nirenberg, L. On stability for difference schemes: A sharp form of Gårding's inequality. *Comm. Pure Appl. Math.* 19: 473–492, 1966.

Lax, P. D., and Wendroff, B. Difference schemes for hyperbolic equations with high order of accuracy. *Comm. Pure Appl. Math.* 17: 381–398, 1964.

Liu, X. D., and Lax, P. Positive schemes for solving multi-dimensional hyperbolic systems of conservation laws. *Comput. Fluid Dyn. J.* 5(2): 133–156, 1996.

Majda, A., and Osher, S. Propagation of error into regions of smoothness for accurate difference approximations to hyperbolic equations. *Comm. Pure Appl. Math.* 30: 671–705, 1977.

Mock, M. S., and Lax, P. D. The computation of discontinuous solutions of linear hyperbolic equations. *Comm. Pure Appl. Math.* 31: 423–430, 1978.

Pearcy, C. An elementary proof of the power inequality for the numerical radius. *Michigan Math. J.* 13: 289–291, 1966.

Rosenblatt, M. A multi-dimensional prediction problem. *Ark. Mat.* 3: 407–424, 1958.

Strang, G. Accurate partial difference methods. I. Linear Cauchy problems. *Arch. Rational Mech. Anal.* 12: 392–402, 1963.

CHAPTER 9

Scattering Theory

Scattering theory studies obstacles in space—objects, potentials—by comparing the propagation of waves in the presence of the obstacle with the propagation of waves in free space. The information available is the asymptotic behavior of waves as time goes from $-\infty$ to ∞. This is expressed as the scattering operator, whose precise definition is given in the pages that follow.

The aim of scattering theory is twofold. The first, called the direct problem, is to prove the existence of the scattering operator. The second, called the inverse problem, is to reconstruct the scatterer from the scattering operator. Solving the inverse problem is of great importance in situations when direct measurements of the scatterer are not possible.

In this chapter we shall study the scattering of acoustic waves by an obstacle in space; only the direct problem will be discussed.

9.1. Asymptotic Behavior of Solutions of the Wave Equation

In this section we shall discuss one of the simplest scattering problems governed by the wave equation in the exterior of an obstacle denoted by B:

$$u_{tt} - \Delta u = 0 \quad \text{outside } B. \tag{9.1}$$

B is a smoothly bounded domain in $\mathbb{R}^3$, contained in a ball of radius R around the origin. On B the solution u is required to be zero:

$$u(x, t) = 0 \quad \text{for } x \text{ on } \partial B. \tag{9.2}$$

We shall be studying solutions u of finite energy, that is, those for which

$$E = \int (u_t^2 + u_x^2) dx$$

is finite, where the integration is over the exterior of B. The standard technique for computing energy—multiplying (9.1) by u_t and integrating by parts—shows that if u has finite energy at, say, $t = 0$, then it has the same energy for all other times.

The mixed initial-boundary value problem (9.1), (9.2), and (9.3),

$$u(x, 0) = g_1(x), \quad u_t(x, 0) = g_2(x), \tag{9.3}$$

can be put into the form of a symmetric system discussed in Appendix D by introducing the first derivatives of u as new unknowns:

$$u_t = w, \quad u_x = p, \quad u_y = q, \quad u_z = r.$$

These four functions satisfy the following system of equations:

$$w_t - p_x - q_y - r_z = 0\,, \quad p_t - w_x = 0\,, \qquad (9.4)$$
$$q_t - w_y = 0\,, \quad r_t - w_z = 0\,.$$

These equations form a symmetric hyperbolic system, and the boundary condition $p = 0$ is minimally nonnegative. We invite the reader to verify these statements. Thus the existence of a solution is assured; its uniqueness is guaranteed by the law of conservation of energy.

Signals governed by the wave equation propagate with speed 1. Therefore it is reasonable to expect that as t tends to infinity, most of the signal is propagated far away from the obstacle. Here is a precise way of stating this.

Given any bounded subset C of the exterior of B, the energy of u contained in C,

$$\int_C (u_t^2 + u_x^2)dx$$

tends to zero as t tends to ∞.

If so, the obstacle plays less and less of a role for u as t tends to ∞; thus u behaves more and more as a solution of the wave equation in free space. To put this more precisely, there exists a solution v_- of the wave equation in free space of finite energy such that

$$\lim_{t\to\infty} \int \left((u - v)_t^2 + (u - v)_x^2\right) dx = 0\,, \qquad (9.5)_+$$

where the integration is over all space, and $u(x, t)$ is set $= 0$ inside the obstacle.

Since the wave equation is invariant under time reversal, an entirely similar state of affairs is expected to hold as t tends to $-\infty$. That is, there is a solution v_- of the wave equation in free space of finite energy for which

$$\lim_{t\to-\infty} \int \left((u - v_-)_t^2 + (u - v_-)_x^2\right) dx = 0\,. \qquad (9.5)_-$$

It is appropriate at this time to introduce the space H of all initial data g of finite energy in the exterior of B that are zero on the boundary of B. We define the energy norm of $g = \{g_1, g_2\}$ as

$$\|g\|_E^2 = \int \left(g_{1x}^2 + g_2^2\right) dx\,,$$

with integration over the exterior of B.

We denote by $U(t)$ the operator that relates the initial data g of a solution u of the mixed initial-boundary value problem to its data at time t. According to the law of conservation of energy, $U(t)$ maps H into H isometrically. Since $U(t)$ is invertible, its inverse being $U(-t)$, the operators $U(t)$ are unitary. Furthermore, they form a one-parameter group:

$$U(s + t) = U(s)U(t)\,.$$

Similarly, we define the space H_0 of all initial data of finite energy in the whole space $\mathbb{R}^3$; the energy norm is defined as before, except that the integration is over the whole space.

H can be embedded in H_0 by setting $g(x) = 0$ inside the obstacle; thus H is a subspace of H_0. The operator that relates the initial data of solutions of the wave equation in $\mathbb{R}^3$ to their data at time t is denoted by $U_0(t)$. They map H_0 onto H_0, and form a one-parameter group of unitary operators.

Let g denote any initial data in H. The relations $(9.5)_\pm$ can be expressed as follows:

There exist data g_+ and g_- in H_0 such that

$$\lim_{t\to\infty} \|U(t)g - U_0(t)g_+\|_E = 0\,, \quad \lim_{t\to-\infty} \|U(t)g - U_0(t)g_-\|_E = 0\,. \tag{9.6}$$

Here g_+ and g_- are the initial values of v_+ and v_-.

Since the operator $U_0(t)$ is an isometry, we deduce from (9.6) that

$$\lim_{t\to\infty} U_0(-t)U(t)g = g_+\,, \quad \lim_{t\to-\infty} U_0(-t)U(t)g = g_-\,; \tag{9.7}$$

here the limit is in the sense of the E-norm.

Denote by W_+ and W_- the operators relating g to g_+ and g_-; these operators, called the *wave operators*, map H into H_0. Since U_0 and U preserve the E-norm, so do their strong limits W_+ and W_-.

Replace t in (9.7) by $t + s$; using the group property we can write

$$U_0(-t-s)U(t+s)g = U_0(-s)U_0(-t)U(t)U(s)g\,.$$

Letting t tend to ∞ we get, using (9.7), that

$$W_+ = U_0(-s)W_+U(s)\,.$$

Multiplying this relation by $U_0(s)$, we get

$$U_0(s)W_+ = W_+U(s)\,, \tag*{$(9.8)_+$}$$

and similarly

$$U_0(s)W_- = W_-U(s)\,. \tag*{$(9.8)_-$}$$

Suppose that W_- maps H onto H_0. Then, since W_- is an isometry, it maps H 1-to-1 onto H_0; its inverse maps H_0 onto H. Multiply $(9.8)_-$ by W_-^{-1} on the left; we get

$$W_-^{-1}U_0(s)W_- = U(s)\,. \tag{9.9}$$

This shows that the groups U and U_0 are unitarily equivalent. Set (9.9) into the right side of $(9.8)_+$ and multiply by W_-^{-1} on the right; we get

$$U_0(s)W_+W_-^{-1} = W_+W_-^{-1}U_0(s)\,. \tag{9.10}$$

The operator $W_+W_-^{-1}$, mapping H_0 into H_0, is the scattering operator alluded to in the introduction to this chapter; it is denoted by S:

$$S = W_+W_-^{-1}\,.$$

Relation (9.10) can be rewritten as

(9.10)′ $$U_0(s)S = SU_0(s);$$

it says that the scattering operator commutes with the group U_0. If we make the further supposition that W_+ maps H onto H_0, we conclude that S is a unitary operator mapping H_0 onto itself.

To gain further knowledge of the nature of the scattering operator, we recall from Section 3.5 of Chapter 3 the representation of solutions of the wave equation in k-dimensional space, k odd, in terms of the Radon transform of their initial data $g = \{g_1, g_2\}$. For $k = 3$ it goes as follows:

Denote the Radon transform of g_1 and g_2 by $\hat{g}_1$ and $\hat{g}_2$. Define

(9.11) $$k_0 = \partial_s^2 \hat{g}_1 - \partial_s \hat{g}_2;$$

then

(9.12) $$\|g\|_E^2 = \int k_0^2 \, ds \, d\omega,$$

and the solution $u_0(x, t)$ with initial value g is given by

(9.13) $$u_0(x, t) = \int k_0(x \cdot \omega - t, \omega) \, d\omega.$$

The data g are represented by the function $k_0(s)$ defined by (9.11); it follows from (9.13) that $U_0(t)g$ is represented by $k_0(s - t)$. For this reason (9.11) is called the *translation representation* of H_0 for the group U_0.

According to (9.10), S commutes with U_0. It follows that in the translation representation, S acts as *convolution*:

(9.14) $$(Sk_0)(r) = \int S(r - s) k_0(s, \omega) \, ds.$$

LEMMA 9.1 *The operator-valued function $S(t)$ representing the scattering operator in the translation representation is supported on* $(-\infty, 2R)$.

PROOF: Let $k_0(s, \omega)$ be a function supported on $(-\infty, -R]$, and g the initial data it represents. It follows from formula (9.13) that the solution $u_0(x, t)$ with initial value g is zero for $t < 0$ and $|x| < R - t$. Such a solution is zero on the obstacle for $t < 0$, and therefore is a solution of the mixed initial-boundary value problem as well. It follows that for $t < 0$

$$U_0(-t)U(t)g = g.$$

Letting t tend to $-\infty$, we conclude that $W_- g = g$.

Let $l(s, \omega)$ be a function supported on $[R, \infty)$, and f the initial data it represents. Arguing as above, replacing $t < 0$ by $t > 0$, we conclude that $W_+ f = f$. Using these relations, we deduce that

$$\begin{aligned}(Sg, f)_E = (W_+ W_-^{-1} g, f)_E &= (W_+ g, f)_E \\ &= (g, W_+^* f)_E = (g, W_+^{-1} f)_E = (g, f)_E;\end{aligned}$$

in the step before the last we used the fact that W_+ is unitary and therefore $W_+^* = W_+^{-1}$.

Next we use the fact that the translation representation is an isometry, and therefore

$$(Sg, f)_E = (g, f)_E = (k_0, l_0) = 0,$$

since the supports of k_0 and l_0 are disjoint. Next we use formula (9.14) for the translation representer of Sg to write

$$0 = (Sg, f)_E = \iint S(r - s)k_0(s)l_0(r)ds\,dr,$$

for all k_0 supported on $(-\infty, -R]$ and all l_0 supported on $[R, \infty)$. It follows from this that $S(t) = 0$ for $t > 2R$. □

Taking the Fourier transform of the translation representation gives a spectral representation of the unitary group U_0. Take the Fourier transform on (9.14); it follows that in the spectral representation the scattering operator acts as *multiplication* by an operator valued function $M(\sigma)$, the Fourier transform of the function $S(t)$ appearing in (9.14). Since $S(t)$ is supported in $(-\infty, 2R)$, $M(\sigma)$ is analytic in the lower half-plane.

We stop at this point and remind the reader that nothing has been proved rigorously. The whole edifice erected in this section is based on the supposition that its wave operators, defined as strong limits, exist and are unitary.

Rigorous proofs of these suppositions will be presented in Section 9.4, in terms of an abstract setup described in Sections 9.2 and 9.3.

9.2. The Lax-Phillips Scattering Theory

The scene is a Hilbert space H and a strongly continuous group $U(t)$ of unitary operators, $U(t + s) = U(t)U(s)$, mapping H onto H.

We start with a theorem of Sinai.

THEOREM 9.2 *Let $U(t)$ be a strongly continuous group of unitary operators mapping a Hilbert space onto itself. Let F_- be a closed subspace of H, called an* incoming subspace, *related to $U(t)$ as follows:*

$$\text{(i)}\quad U(t)F_- \subset F_- \quad \textit{for } t < 0,$$

$$\text{(9.15}_-\text{)}\qquad \text{(ii)}\quad \bigcap U(t)F_- = \{0\}$$

$$\text{(iii)}\quad \bigcup U(t)F_- \quad \textit{is dense in } H.$$

Then H can be represented as $L^2(\mathbb{R}, N)$, N some auxiliary Hilbert space, so that

(a) *F_- is represented as $L^2(\mathbb{R}_-, N)$.*
(b) *The action of $U(t)$ is right translation; that is, if h in H is represented by the function $k(s)$, $U(t)h$ is represented by $k(s - t)$.*

The representation is essentially unique; any two representations are related by a constant unitary factor $N \to N$.

Similarly, let F_+ be a closed subspace of H, called an outgoing subspace, *related to $U(t)$ as follows*:

$$(9.15_+)\qquad \begin{aligned} &\text{(i)}\ \ U(t)F_+ \subset F_+ \quad \textit{for } t > 0\,, \\ &\text{(ii)}\ \ \bigcap U(t)F_+ = \{0\}\,, \\ &\text{(iii)}\ \ \bigcup U(t)F_+ \quad \textit{is dense in } H\,. \end{aligned}$$

Then H can be represented as $L^2(\mathbb{R}, N)$ so that F_+ is represented as $L^2(\mathbb{R}_+, N)$ and the action of $U(t)$ is right translation.

Curiously, in the applications we make in this chapter, the translation representations are constructed explicitly, without appeal to Sinai's theorem. But our thinking has been guided throughout this development by being aware of Sinai's theorem.

Suppose the unitary group $U(t)$ acting on the Hilbert space H has both an incoming and an outgoing subspace F_- and F_+. By Sinai's theorem, there are two translation representations of $U(t)$, $H \leftrightarrow L^2(\mathbb{R}, N_+)$ and $H \leftrightarrow L^1(\mathbb{R}, N_-)$. Since the Fourier transform of a translation representation is a spectral representation, the dimension of the auxiliary space N is the multiplicity of the spectrum. Therefore N_+ and N_- have the same dimension, and so may be taken as the same space.

Let h be any element of H, k_- and k_+ its incoming and outgoing representers. We denote by S the operator relating the two:

$$(9.16)\qquad Sk_- = k_+\,;$$

S is called the *scattering operator* associated with $U(t)$, F_-, and F_+.

THEOREM 9.3 *Let $U(t)$ be a strongly continuous unitary group acting on the Hilbert space, and let F_- and F_+ be incoming and outgoing subspaces that are orthogonal to each other. Then the scattering operator S has the following properties*:

(i) *S is a unitary map of $L^2(\mathbb{R}, N)$ onto itself.*
(ii) *S commutes with translation.*
(iii) *S maps $L^2(\mathbb{R}_-, N)$ into itself.*

PROOF:

(i) Since k_- and k_+ represent the same element of H isometrically, S is an isometry. Since it maps $L^2(\mathbb{R}, N)$ onto itself, it is unitary.

(ii) Since $k_-(s-t)$ and $k_+(s-t)$ both represent $U(t)h$, S maps translates of k_- onto translates of k_+.

(iii) Any k_- in $L^2(\mathbb{R}_-, N)$ represents an element of F_-. By assumption such an element is orthogonal to F_+. Therefore its outgoing representer k_+ is orthogonal to the functions representing F_+, which is $L^2(\mathbb{R}_+, N)$. Therefore k_+ is supported on $\mathbb{R}_-$.

□

We pass now from the translation representations to the corresponding spectral representations by taking their Fourier transforms:

$$a_-(\sigma) = \int k_-(s)e^{is\sigma}\,ds, \quad a_+(\sigma) = \int k_+(s)e^{is\sigma}\,ds. \tag{9.17}$$

We denote by $\widetilde{S}$ the operator linking a_- to a_+,

$$\widetilde{S}a_- = a_+ . \tag{9.18}$$

In what follows we need the vector version of the Paley-Wiener theorem:

THEOREM 9.4 *The Fourier transform of a vector-valued function $a(s)$ supported on $\mathbb{R}_-$ and square integrable has an analytic extension to $\mathbb{C}_-$ with the following properties*:

(i) *For fixed $\eta > 0$, $a(\sigma - i\eta)$ is an L^2 function of σ on $\mathbb{R}$; as $\eta \to \infty$, the L^2 norm of $a(\sigma - i\eta)$ tends to 0.*
(ii) *As η tends to 0, $-a(\sigma - i\eta)$ tends to a in the L^2 norm.*
(iii) *Conversely, every function of a with properties* (i) *and* (ii) *is the Fourier transform of an $L^2(\mathbb{R}_-, N)$ function.*

The proof in the scalar case is nothing more than an application of the Cauchy integral theorem. The extension to the vector-valued case is straightforward.

We denote by A_- the Fourier transform of $L^2(\mathbb{R}_-, N)$, and by A_+ the Fourier transform of $L^2(\mathbb{R}_+, N)$. Analogously to A_-, functions in A_+ have analytic extensions in the upper half-plane $\mathbb{C}_+$.

Theorem 9.3 has a straightforward version for the spectral representation:

THEOREM 9.5

(i) *$\widetilde{S}$ is a unitary mapping of $L^2(\mathbb{R}, N)$ onto itself.*
(ii) *$\widetilde{S}$ commutes with multiplication by bounded, continuous scalar functions.*
(iii) *$\widetilde{S}$ maps A_- into itself.*

PROOF: Only (ii) needs a ghost of a proof. According to part (ii) of Theorem 9.3, the operator S commutes with translation. It follows that $\widetilde{S}$ commutes with multiplication by $e^{i\sigma t}$ for all real t. By forming linear combinations of these exponentials we deduce that $\widetilde{S}$ commutes with multiplication by any continuous, bounded scalar function. □

THEOREM 9.6 *The operator $\widetilde{S}$ defined in* (9.18) *is multiplication by an operator valued function $M(\sigma)$, $N \to N$, with the following properties*:

(i) *$M(\sigma)$ is unitary for almost all real σ.*
(ii) *$M(\sigma)$ is the boundary value of an operator-valued function defined and holomorphic in the lower half-plane $\mathbb{C}_-$ defined as* $\operatorname{Im}\sigma < 0$.
(iii) *For each ζ in $\mathbb{C}_-$, $M(\zeta)$ is a contraction mapping N into N.*

PROOF: We tackle (ii) first. According to part (iii) of Theorem 9.5, if the incoming spectral representer belongs to A_-, so does the outgoing representer. We claim that for any ζ in $\mathbb{C}_-$ the value of $a_+(\zeta)$ is determined by the value of $a_-(\zeta)$.

To prove this it suffices to show that if $a_-(\zeta) = 0$, then $a_+(\zeta) = 0$. We factor such an a_- as

$$a_-(\sigma) = \frac{\sigma - \zeta}{\sigma + \zeta} h(\sigma).$$

It follows from part (iii) of Theorem 9.4 that $h(\sigma)$ belongs to A_-. Since $\widetilde{S}$ commutes with multiplication by bounded, continuous functions

$$a_+\sigma = \widetilde{S}a_- = \widetilde{S}\frac{\sigma - \zeta}{\sigma + \zeta} h = \frac{\sigma - \zeta}{\sigma + \zeta} \widetilde{S}h.$$

Since h belongs to A_-, so does $\widetilde{S}h$; setting $\sigma = \zeta$ in the above equation shows that $a_+(\zeta) = 0$. This shows that that $a_-(\zeta)$ determines the value of $a_+(\zeta)$.

Since $\widetilde{S}$ is linear, $a_+(\zeta)$ depends linearly on $a_-(\zeta)$:

$$a_+(\zeta) = M(\zeta)a_-(\zeta), \tag{9.19}$$

$M(\zeta)$ a linear operator mapping $N \to N$.

To show that $M\sigma$ is a strongly analytic function of ζ, set $a_-(\sigma) = \frac{1}{\sigma - i}n$, n any vector in N. Clearly a_- belongs to A_-; therefore so does a_+. Set this pair into (9.19):

$$a_+(\sigma) = \frac{1}{\sigma - i} M(\sigma)n,$$

since $a_+(\sigma)$ is an analytic function in $\mathbb{C}_-$, so is $M(\sigma)n$.

We turn now to proving part (iii). Let n be any vector in N, $|n|$ its norm. ζ is any point in $\mathbb{C}_-$, $\operatorname{Im}\zeta = -\eta$. Set

$$a_-(\sigma) = \frac{2i\eta}{\bar{\zeta} - \sigma} n;$$

a_- belongs to A_-, and

$$a_-(\zeta) = n, \quad \|a_-\| = 2\sqrt{\pi\eta}\,|n|. \tag{9.20}$$

Set $a_+ = \widetilde{S}a_-$; by Theorem 9.4, part (iii), a_+ belongs to A_-. We express the value of $a_+(\zeta)$ using the residue theorem:

$$a_+(\zeta) = \frac{1}{2\pi i} \int \frac{a_+(\sigma)}{\sigma - \zeta} d\sigma \tag{9.21}$$

over a contour that goes around ζ. We choose a rectangular contour from the point l on the real axis to $-l$, then to $-l - il$, to $l - il$, then back to l. Since $a_+(\sigma)$ tends to zero as σ tends to ∞, as $l \to \infty$ we are left with the integral along the real axes. We estimate that integral by the Schwarz inequality:

$$|a_+(\zeta)| \leq \frac{1}{2\pi} \|a_+\| \left\| \frac{1}{\sigma - \zeta} \right\| = \frac{1}{2\sqrt{\pi\eta}} \|a_+\|. \tag{9.22}$$

According to part (i) of Theorem 9.5, $\widetilde{S}$ is an isomorphism: $\|a_+\| = \|a_-\|$. Setting this into (9.22) and using (9.20) we get

$$|a_+(\zeta)| \leq \frac{1}{2\sqrt{\pi\eta}} 2\sqrt{\pi\eta}\,|n| = |a_-(\zeta)|.$$

This proves that $M(\zeta)$ is a contraction for every ζ in the lower half-plane.

The rest of the proof is basic theory of analytic functions. Take any two vectors n and p in N; $(M(\sigma)n, p)$, where $(\cdot, \cdot)$ denotes the scalar product in N, is a bounded analytic function in $\mathbb{C}_-$. For such functions

$$\lim_{\eta \to 0}(M(\lambda - i\eta)n, p) \tag{9.23}$$

exists for almost all real λ. Take a dense denumerable set of n and p; the limit (9.23) exists on a set of λ whose complement is of measure zero. Then, since $M(\sigma)$ is a contraction, it follows that the limit (9.23) exists for all n and p in N for almost all λ. We denote this weak limit as $M(\lambda)$; clearly, $M(\lambda)$ is a contraction a.e.

For any vector n in N, $a_-(\sigma) = \frac{n}{\sigma - i}$ belongs to A_-; therefore so does $a_+ = \widetilde{S}a_-$. We have shown that $\widetilde{S}$ acts as multiplication by $M(\sigma)$; therefore

$$M(\sigma)a_-(\sigma) = \frac{1}{\sigma - i}M(\sigma)n = a_+(\sigma). \tag{9.24}$$

Set $\sigma = \lambda - i\eta$, λ real. As $\eta \to 0$, the right side tends, according to part (ii) of Theorem 9.4, strongly to $a_+(\lambda)$. It follows that $M(\lambda - i\eta)n$ tends strongly to $M(\lambda)$ for almost all λ.

Since $\widetilde{S}$ is an isometry, $\|a_-\| = \|a_+\|$. Using the form above of a_- and a_+, this means that

$$|n|^2 \int \frac{d\lambda}{|\lambda - i|^2} d\lambda = \int \frac{|M(\lambda)n|^2}{|\lambda - i|^2} d\lambda .$$

Since $|M(\lambda)n| \le |n|$ for almost all λ, the sign of equality must hold for almost all λ.

It follows from (9.24) as $\eta \to 0$ that $a_+(\lambda) = M(\lambda)a_-(\lambda)$ a.e. Since $\widetilde{S}$ is unitary, it follows that $M(\lambda)$ is invertible, and therefore unitary, for a.a. λ. This completes the proof of Theorem 9.6. □

The function $M(\sigma)$ is called the *scattering metric*.

9.3. The Associated Semigroup

The setting is the same as in Section 9.2.

Let $U(t)$ be a strongly continuous, one-parameter group of unitary operators acting on a Hilbert space H. Let F_- and F_+ be a pair of incoming and outgoing subspaces in the sense of (9.15_-) and (9.15_+); furthermore, we assume that F_- and F_+ are orthogonal to each other. Denote by P_- and P_+ orthogonal projections onto the orthogonal complements F_- and F_+, respectively. Denote by K the orthogonal complement in H of $F_- \oplus F_+$. Define $Z(t)$ by

$$Z(t) = P_+U(t)P_-, \quad t \ge 0. \tag{9.25}$$

THEOREM 9.7 *$Z(t)$ is a one-parameter semigroup of contractions on K, and that $Z(t)$ tends strongly to zero as $t \to \infty$.*

PROOF: Clearly each $Z(t)$ is a contraction. To show that it maps K into K we have to show that for every k in K, $Z(t)k$ is orthogonal to F_- and F_+ when t is positive.

Since P_- is the identity on K, $Z(t)k = P_+U(t)k$. We claim that $U(t)k$ is orthogonal to F_-; to see this take any f_- in F_- and write

$$(9.26)\qquad (U(t)k, f_-)_E = (k, U^*(t)f_-)_E = (k, U^{-1}(t)f_-)_E = (k, U(-t)f_-)_E,$$

where $(\cdot,\cdot)_E$ denotes the scalar product in H. By (9.15$_-$), $U(-t)$ takes F_- into itself, and therefore the scalar product on the right in (9.26) is zero. Since $P_+U(t)k$ differs from $U(t)k$ by a vector in F_+, assumed orthogonal to F_-, it follows that $P_+U(t)k$, too, is orthogonal to F_-.

Since the range of P_+ is orthogonal to F_+, $P_+U(t)k$ is orthogonal to F_+. This completes the proof that $Z(t)$ maps K into K.

Next we show that the operators $Z(t)$ form a semigroup. By definition

$$Z(t)Z(s)k = P_+U(t)P_-P_+U(s)P_-k.$$

Since k is orthogonal to F_-, and since P_+ and P_- commute, we can rewrite the right side as

$$P_+U(t)P_+P_-U(s)k.$$

We have shown above that $U(s)k$ is orthogonal to F_-, so P_- above can be dropped. We are left with

$$P_+U(t)P_+U(s)k = P_+U(t)(U(s)k + f_+) = P_+U(t)U(s)k + P_+U(t)f_+.$$

By (9.15$_+$), $U(t)$ maps F_+ into itself; therefore P_+ kills the second term and leaves us with $P_+U(t+s)k = Z(t+s)k$.

To show that $Z(t)$ tends to zero strongly, we present Z in the outgoing translation representation. Since k in K is orthogonal to F_+, its outgoing translator representer K_+ is supported on $\mathbb{R}_-$. $U(t)$ translates k_+, and P_+ removes that part of $U(t)k_+$ that is supported in $[0, t]$. So the action of $Z(t)$ on k_+ is translation followed by restriction to $\mathbb{R}_-$; clearly $\|Z(t)k_-\|$ tends to zero. □

We establish now an interesting and important relation between the semigroup $Z(t)$ and the scattering matrix $M(\sigma)$ defined in Section 9.2. That some relation exists is not surprising, since both are built out of the same ingredients.

We recall the concept of the infinitesimal generator G of a semigroup $Z(t)$:

$$(9.27)\qquad Gk = \lim_{t\to 0} \frac{Z(t)k - k}{t}.$$

The domain of G is the set of k in K for which the limit on the right in (9.27) exists in the sense of convergence in the norm of K. It follows easily from the semigroup property that if k belongs to the domain of G, so does $Z(t)k$, and $GZ(t)k = Z(t)Gk$. Put in another way,

$$(9.27)'\qquad \frac{d}{dt}Z(t)k = GZ(t)k.$$

THEOREM 9.8 *A complex number γ, $\operatorname{Re}\gamma < 0$, belongs to the point spectrum of G, $Gk = \gamma k$, iff $Z(t)k = e^{\gamma t}k$.*

We leave the proof of this proposition to the reader.

THEOREM 9.9 *A complex number γ, $\operatorname{Re}\gamma < 0$, belongs to the spectrum of the infinitesimal generator G of the semigroup $Z(t)$ iff $M^*(i\bar{\gamma})$ has a nontrivial nullspace. The dimension of the nullspace equals the multiplicity of the eigenvalue,*

PROOF: Let k be an eigenvector of $G : Gk = \gamma k$. Then $Z(t)k = e^{\gamma t}k$. Let k_+ denote the outgoing translation representer of k. Since k belongs to K, it is orthogonal to $F+$; therefore k_+ is supported on $\mathbb{R}_-$. As we have seen earlier, the action of $Z(t)$ is right translation followed by restriction to $\mathbb{R}_-$. Since k satisfies $Z(t)k = e^{\gamma t}k$,

$$k_+(s-t) = e^{\gamma t}k_+(s), \quad s < 0.$$

It follows that

$$k_+(s) = \begin{cases} e^{\gamma s}n, & s < 0, \\ 0, & 0 < s. \end{cases}$$

The outgoing spectral representer of k is the Fourier transform of k_+:

$$f_+(\sigma) = \frac{1}{i\sigma - \gamma} n,$$

The incoming spectral representer is

$$f_-(\sigma) = M^{-1}(\sigma) f_+(\sigma) = \frac{1}{i\sigma - \gamma} M^{-1}(\sigma)n.$$

For σ real, $\sigma = \bar{\sigma}$ and $M^{-1}(\sigma) = M^*(\sigma)$; so for σ real

$$f_-(\sigma) = \frac{1}{i\sigma - \gamma} M^*(\bar{\sigma})n. \tag{9.28}$$

Since k_- is supported on $\mathbb{R}_+$, its Fourier transform belongs to A_+ and thus has an analytic extension to the upper half-plane. Formula (9.28) gives a meromorphic extension of $f_-(\sigma)$ to $\mathbb{C}_+$; $M^*(\bar{\sigma})n$ is analytic for σ in $\mathbb{C}_+$, but $\frac{1}{i\sigma-\gamma}$ has a pole at $\sigma = -i\gamma$. So for $f_-(\sigma)$ to be analytic $M^*(\bar{\sigma})n$ must vanish at $\sigma = -i\gamma$:

$$M^*(i\bar{\gamma})n = 0.$$

It follows furthermore that to each eigenfunction of G there corresponds a nullvector of $M^*(i\bar{\gamma})$.

The proof can be run backwards to deduce the converse proposition. □

REMARK. Suppose that the scattering matrix $M(\sigma)$ is invertible at all but a discrete set of points σ in $\mathbb{C}_-$, and that it is continuous on the real axis. Then $M(\sigma)$ has a meromorphic continuation into the upper half-plane, given by the operator version of the Schwarz reflection principle:

$$M(\sigma) = M^*(\bar{\sigma})^{-1} \quad \text{for } \operatorname{Im}\sigma > 0.$$

9.4. Back to the Wave Equation in the Exterior of an Obstacle

The setting is the same as in Section 9.1.

H is the space of all initial data g of finite energy in the exterior of an obstacle B contained in the ball $|x| \leq R$. The *energy norm* of $g = \{g_1, g_2\}$ is

$$\|g\|_E^2 = \int (g_{1x}^2 + g_2^2) dv , \tag{9.29}$$

the integration over the exterior of B. The operator $U(t)$ relates the initial data g of a solution u of the wave equation in the exterior of B, $u = 0$ on ∂B, to its data at time t. The operators $U(t)$ form a strongly continuous one-parameter group of unitary operators mapping H onto H.

A solution u is called *incoming* if $u(x, t) = 0$ in the backward cone $|x| \leq R - t$, $t < 0$. *Outgoing* solutions are those that are zero in the forward cone $|x| \leq R + t$, $t > 0$. The initial data of incoming solutions are denoted as F_-; those of outgoing solutions as F_+.

THEOREM 9.10 *F_- and F_+ defined above are incoming and outgoing subspaces for the one-parameter group $U(t)$ defined above. That is,*

$$(9.30)_- \qquad \begin{aligned} &\text{(i)} \ U(t)F_- \subset F_- \quad \textit{for } t < 0 , \\ &\text{(ii)} \ \bigcap U(t)F_- = \{0\} , \\ &\text{(iii)} \ \bigcup U(t)F_- \quad \textit{is dense in } H , \end{aligned}$$

and similarly

$$(9.30)_+ \qquad \begin{aligned} &\text{(i)} \ U(t)F_+ \subset F_+ \quad \textit{for } t > 0 , \\ &\text{(ii)} \ \bigcap U(t)F_+ = \{0\} , \\ &\text{(iii)} \ \bigcup U(t)F_+ \quad \textit{is dense in } H . \end{aligned}$$

Furthermore, F_- and F_+ are orthogonal to each other.

PROOF: We recall from Section 9.1 the space H_0 of all data with finite energy defined in the whole space $\mathbb{R}^3$, and the operator $U_0(t)$ that relates the initial data of a solution of the wave equation in $\mathbb{R}^3$ to its data at time t. We further recall the translation representation (9.11) of H_0 furnished by the Radon transform of the data. The solution u_0 of the wave equation in terms of the translation representer k_0 is given by formula (9.13):

$$u_0(x, t) = \int k_0(x \cdot \omega - t, \omega) d\omega , \tag{9.13}$$

where k_0 is defined by (9.11). The energy norm of u equals the L^2 norm of k_0; see (9.12).

It follows from formula (9.13) that when $k_0(s, \omega) = 0$ for $s \leq 0$, $u_0(x, t) = 0$ in the forward cone $|x| < t$, $0 < t$. Likewise, when $k_0(s, \omega) = 0$ for $s \geq 0$, $u_0(x, t) = 0$ in the backward cone $|x| < -t$, $t < 0$. □

We now show the converse of these propositions.

LEMMA 9.11 *Suppose that $u_0(x,t)$ is a solution of the wave equation in all space and time, has finite energy, and is zero in the forward cone $|x| < t$, $t > 0$. Then the translation representer $k_0(s,\omega)$ of its initial data is zero for $s < 0$. Similarly, if $u_0(x,t) = 0$ in the backward cone $|x| < -t$, $t < 0$, the translation representer $k_0(s,\omega)$ of its initial data is zero for $s > 0$.*

PROOF: We shall deal with C^∞ solutions. We can achieve this by mollifying the solution in the t-variable:

$$v_0(x,t) = \int u_0(r,x)\phi(t-r)dr\,,$$

where ϕ is a C^∞ function, $\int \phi\, dr = 1$, and ϕ is supported on the interval $[-\varepsilon, 0]$. The translation representer of v is

$$k_\phi(s) = \int k_0(r,\omega)\phi(s-r)dr\,. \tag{9.31}$$

This is a $C^\infty \cap L^2$ function of s; it follows from formula (9.13) that $v_0(x,t)$ is C^∞. Since ϕ is supported on $\mathbb{R}_-$, if u_0 is zero in the forward cone, so is v_0.

Since $v_0(x,t) = 0$ in the forward cone, all its spacial derivatives are zero on the positive t-axis:

$$D_x^n v_0(x,t)\big|_{x=0} = 0 \quad \text{for } t \geq 0\,, \tag{9.32}$$

where n is a multi-index. We can compute these derivatives from (9.13):

$$D_x^n v_0(x,t)\big|_{x=0} = \int \omega^n D_s^{|n|} k_0(-t,\omega)d\omega\,.$$

Multiply this by a smooth test function $a(t)$ supported on a finite interval of $\mathbb{R}_+$. It follows from (9.32) that

$$\int_0^\infty \omega^n k_0(-t,\omega) D_t^{|n|} a(t)dt = 0\,.$$

Any smooth function $b(t)$ compactly supported on $\mathbb{R}_+$ can be approximated in the $L^2(\mathbb{R}_+)$ norm by functions of the form $D_t^n a(t)$. Therefore

$$\int k_0(-t,\omega)\omega^n b(t)dt = 0$$

for all smooth $b(t)$ supported on a bounded interval on $\mathbb{R}_+$. Since functions of the form $\sum c_n w^n b_n(t)$ are dense in $L^2(t,\omega)$, $t > 0$, it follows that $k_0(t,\omega) = 0$ for $t > 0$.

It follows from (9.31) that as the support $[-\varepsilon, 0]$ of the mollifying function ϕ tends to zero, k_ϕ tends to k_0; therefore $k_0(-t,\omega) = 0$ for $t > 0$, as asserted in Lemma 9.11. The second assertion of Lemma 9.11 follows analogously. □

We have observed in Section 9.1 that H can be regarded as a subspace of H_0. Denote by g_+ any member of the outgoing subspace F_+ of H. Clearly, $U_0(t)g_+ = U(t)g_+$ for $t > 0$, therefore $U_0(-R)g_+$ is zero in the forward cone $|x| < t$, $t > 0$, and so by Lemma 9.11 its translation representer is zero in $\mathbb{R}_-$. It follows that the translation representer k_+ of g_+ is zero on $s < R$. Similarly, the translation representer of any g_- in F_- is zero for $-R < s$. Thus the supports of k_+ and of k_- are disjoint. Therefore the L^2 scalar product $(k_-, k_+) = 0$. Since the translation representation is isometric, $(g_-, g_+)_E = 0$. This proves that F_- and F_+ are orthogonal.

Next we shall deal with $(9.30)_+$; $(9.30)_-$ can be treated similarly. Properties (i) and (ii) in $(9.30)_+$ are obviously true. Part (iii) is trickier and requires seven lemmas. We start with the following observation:

Let C be a bounded set containing the obstacle. Denote by $\|g\|_{E,C}$ the *local energy norm*:

$$\|g\|_{E,C}^2 = \int_C (g_{1x}^2 + g_2^2)dx \,. \tag{9.33}$$

We claim that property (iii) in $(9.30)_+$ implies local energy decay, that is, that for every h in H

$$\lim_{t\to\infty} \|U(t)h\|_{E,C} = 0 \,. \tag{9.34}$$

PROOF: If property (iii) holds, then given any h in H and any $\varepsilon > 0$, there exists an element g of F_+ and a time T such that

$$\|h - U(T)g\|_E < \varepsilon \,. \tag{9.35}$$

Since C is bounded, it is contained in some ball around the origin of radius K, The solution with initial value in F_+ is zero in the ball $|x| \le R + t$. Therefore if $t > |T| + K$, $U(t+T)g = 0$ in C.

Since $U(t)$ is an isometry, it follows from (9.35) that

$$\|U(t)h - U(t+T)g\|_E < \varepsilon \,.$$

Since local energy is less than total energy,

$$\|U(t)h - U(t+T)g\|_{E,C} < \varepsilon \,.$$

But since for $k > |T| + K$, $U(t+T)$ is zero in C,

$$\|U(t)h\|_{E,C} < \varepsilon \,.$$

This proves (9.34). □

Of greater interest is the converse implication:

LEMMA 9.12 *Suppose that for all h in H and all bounded domains C*

$$\lim_{s\to\infty} \inf \|U(s)h\|_{E,C} = 0 \,. \tag{9.36}$$

Then property (iii) *in* $(9.30)_+$ *holds.*

PROOF: We argue indirectly. The union of $U(t)F_+$ is a linear subspace of H; if it were not dense in H, there would be a nonzero p in H orthogonal to it:

$$(p, U(t)g_+)_E = 0 \quad \text{for all } g_+ \text{ in } F_+ \,. \tag{9.37}$$

Since $U(t)$ is unitary, $U^*(t) = U^{-1}(t) = U(-t)$. So the above relation can be rewritten as

$$(U^*(t)p, g_+)_E = (U(-t)p, g_+) = 0 \quad \text{for all } g_+ \text{ in } F_+ .$$

Writing s for $-t$, we can express this relation in words: For all s, $U(s)p$ is orthogonal to F_+. □

LEMMA 9.13 *For all q in H that are orthogonal to F_+*

(i) $U_0(t)q = 0$ *in the cone* $|x| < -t - R$ *for* $t < -R$, *and*

(ii) $U_0(t)q = U(t + 2R)U_0(-2R)q$ *for* $t < -2R$.

PROOF: We have seen earlier in this section that the free space translation representation of F_+ is $L^2[R, \infty)$. Therefore the translation representer of any element in H that is orthogonal to F_+ is supported on $(-\infty, R]$. Part (i) then follows from the explicit formula (9.13) expressing solutions in free space in terms of their translator representers. Since for $t < -2R$ this solution is zero on the obstacle, it is a solution of the mixed problem as well; this proves part (ii). □

LEMMA 9.14 *For any h in H*

(i) $U_0(-2R)h = U(-2R)$ *for* $|x| > 3R$,

(ii) $\|U_0(-2R)h\|_{E,3R} \le \|h\|_{E,5R}$, $\|U(-2R)h\|_{E,3R} \le \|h\|_{E,5R}$.

PROOF: Both statements express the fact that signals are propagated with speed ≤ 1. Part (i) says that solutions of the initial-boundary value problem at time t are unaffected by the boundary condition at points x whose distance to the boundary is greater than $|t|$. Part (ii) says that the energy contained inside a ball of radius $3R$ at time $-2R$ comes from the energy contained in the solution inside the ball of radius $5R$ at time 0. We leave it to the reader to give a formal derivation of this estimate. □

Next we make use of hypothesis (9.36) of local energy decay: given any ε, there exist s arbitrarily large such that

$$\|U(s)p\|_{E,5R} < \varepsilon ; \tag{9.38}$$

here $\| \; \|_{E,5R}$ denotes the energy contained inside the ball of radius $5R$.

Let's apply part (ii) of Lemma 9.14 to $q = U(s)p$; using estimate (9.38) we get

$$\|U_0(-2R)U(s)p\|_{E,3R} < \varepsilon , \quad \|U(s - 2R)p\|_{E,3R} < \varepsilon .$$

By the triangle inequality

$$\|U_0(-2R)U(s)p - U(s - 2R)p\|_{E,3R} < 2\varepsilon . \tag{9.39}$$

According to part (i) of Lemma 9.14 applied to $h = U(s)p$,

$$U_0(-2R)U(s)p = U(s - 2R)p \quad \text{for } |x| \ge 3R .$$

Combining this with (9.39), we get the estimate

$$\|U_0(-2R)U(s)p - U(s - 2R)p\|_E < 2\varepsilon . \tag{9.39$'$}$$

According to part (i) of Lemma 9.13 applied to $q = U(s)p$, $U_0(-2R)U(s)p$ belongs to H. Therefore $U(2R-s)$ acts on $U_0(-2R)U(s)p - U(s-2R)p$; since it is an isometry, it follows from (9.39)′ that

$$\|U(2R-s)U_0(-2R)U(s)p - p\|_E < 2\varepsilon . \tag{9.40}$$

Now apply part (ii) of Lemma 9.13 with $t = -s$ and $q = U(s)p$; we conclude that for $s > 2R$, $U(2R-s)U_0(-2R)U(s)p = U_0(-s)U(s)p$. According to part (i) of Lemma 9.13, $U_0(-s)U(s)p$ is zero for $|x| < s - R$. Combine this with (9.40); we get

$$\|p\|_{E,s-R} < 2\varepsilon .$$

According to hypothesis (9.36), s can be taken arbitrarily large, so it follows that $\|p\|_E < \varepsilon$. Since ε can be taken arbitrarily small, it follows that $\|p\|_E = 0$, a contradiction to $p \neq 0$. This completes the proof of Lemma 9.12.

We turn now to proving hypothesis (9.36) of Lemma 9.12, that for all h in H,

$$\liminf_{s\to\infty} \|U(s)h\|_{E,G} = 0 \tag{9.36}$$

for all compact subsets C. We need the following theorem of Wiener:[1]

THEOREM 9.15 *Let dm be a signed measure on $\mathbb{R}$ that has finite total variation. Denote by $\widetilde{m}(t)$ its Fourier transform:*

$$\widetilde{m}t = \int e^{it\lambda}\, dm(\lambda) . \tag{9.41}$$

Suppose that dm contains no point measure; then the mean value of $|\widetilde{m}|^2$ is zero.

$$\lim_{T\to\infty} \frac{1}{T}\int_{-T}^{T} |\widetilde{m}(t)|^2\, dt = 0 . \tag{9.42}$$

For proof, see, e.g., section 5.2 in *Scattering Theory* by Lax and Phillips.

We return now to the unitary group $U(t)$ that describes solutions of the wave equation in the exterior of an obstacle that have finite energy. Denote by A the infinitesimal generator of this group; A is an anti-self-adjoint operator on H. Let $dP(\lambda)$ denote the spectral resolution of iA; $dP(\lambda)$ is a projection-valued measure. The group $U(t)$ generated by A is the Fourier transform of this measure:

$$U(t) = \int e^{i\lambda t}\, dP(\lambda) . \tag{9.43}$$

We only need the weak form of the representation:

$$(U(t)h, g)_E = \int e^{i\lambda t}\, d(P(\lambda)h, g) \tag{9.43$_w$}$$

for any pair of elements g and h of H.

LEMMA 9.16 *Suppose that the spectrum of A is free of point eigenvalues. Then there is a sequence t_m tending to ∞ such that $U(t_m)$ tends weakly to zero, that is, for any pair of elements g and h in H,*

$$\lim (U(t_m)h, g)_E = 0 . \tag{9.44}$$

[1]The suggestion to use Wiener's theorem is due to Karel de Leeuw.

PROOF: Formula $(9.43)_w$ says that $U(t)h, g)_E$ is the Fourier transform of the measure $dm = d(P(\lambda)h, g)$. The total variation of m is $\|h\|_E\|g\|_E$, and we have supposed—supposition to be proved subsequently—that dP contains no point measure. So by Wiener's theorem, 9.15, the mean value of its Fourier transform squared is zero:

$$\lim_{T\to\infty} \frac{1}{T}\int_{-T}^{T} |(U(t)h, g)_E|^2\, dt = 0. \tag{9.45}$$

We choose a denumerable set $\{f_j\}$ of elements that are dense in H. It follows from (9.45) that for any n there is a T such that $T > n$ and

$$\int_{-T}^{T} |(U(t)f_i, f_j)_E|^2\, dt < \frac{T}{3n^3},$$

$i = 1, \dots, n$, $j = 1, \dots, n$. It follows that for each i, j, $|(U(t)f_i, f_j)_E| \le 1/n$ except on a set of measure $\le T/3n^2$. Therefore there is a t_n contained in $(T/2, T)$ such that

$$|(U(t_n)f_i, f_j)_E| < \frac{1}{n}$$

for $i, j = 1, \dots, n$. Since the $\{f_j\}$ are dense in H, it follows that (9.44) holds for all h and g. □

Next we show how to deduce strong local energy decay (9.36) from weak energy decay (9.44):

LEMMA 9.17 *For any h in H, there is a sequence $s_n \to \infty$ such that for any compact set C*

$$\lim_{n\to\infty} \|U(s_n)h\|_{E,C} = 0.$$

PROOF: It suffices to prove this for a dense set of h; we may take this dense set to consist of smoothed data of the form

$$\int \phi(t-r)U(r)h\, dr = h_\phi,$$

ϕ a C_0^∞ function. For such h the solution $u(x,t)$ of the wave equation with initial data h is C^∞ in t. In particular, u_t satisfies the wave equation, and so the sum of the L^2 norms of u_{xt} and u_{tt} over the exterior of the obstacle is the same for all t. Since u satisfies the wave equation, the L^2 norm of $\Delta u = u_{tt}$ is uniformly bounded for all t. Standard elliptic techniques give estimates of the L^2 norm of all second derivatives of u in any bounded domain C adjacent to the boundary of the obstacle in terms of the L^2 norm of Δu. By Rellich's compactness theorem the functions u_x and u_t belong to a precompact set in the L^2 norm over C. It follows that the sequence $U(t_n)h$ has a subsequence $U(s_n)h$ which converges in the $\|\cdot\|_{E,C}$ norm. But we have seen in Lemma 9.16 that the weak limit of this sequence is zero. Since weak and strong limits are equal, it follows that $\|U(s_n)h\|_{E,C}$ tends to zero, as claimed in Lemma 9.17. □

We have now proved hypothesis (9.36) underlying Lemma 9.12 under the additional hypothesis that the infinitesimal generator A of $U(t)$ has no point spectrum. We supply now the last step in this chain of reasoning leading to the proof of parts $(9.30)_-$ and $(9.30)_+$ of Theorem 9.10.

LEMMA 9.18 *A has no point eigenvalue.*

PROOF: We argue indirectly; suppose that for some nonzero g in H, and σ real, $Ag = \sigma g$. Then $g = (a, i\sigma a)$, and

$$U(t)g = e^{i\sigma t} g\,,$$

so the solution u of the wave equation with initial value g is of the form $u(x,t) = e^{i\sigma t}a(x)$. It follows that $a(x)$ satisfies

$$\Delta a + \sigma^2 a = 0\,. \tag{9.46}$$

We claim that this equation has no solution in the exterior of the obstacle that has finite, nonzero energy. We take first the case $\sigma = 0$. To see this multiply (9.46) by a and integrate by parts. Since $a(x) = 0$ in the boundary of the obstacle, we get that

$$\int a_x^2\,dx = 0\,.$$

It follows that $\|g\|_E = 0$, a contradiction.

For $\sigma \neq 0$ there is a sharper result due to Rellich and Vekua: (9.46) has no nonzero solution that is square integrable at infinity; we don't need to require that the solution a vanish on the boundary of the obstacle. To see this we expand the solution $a(x)$ into a series of spherical harmonics:

$$a(x) = \sum b_n(r) Y_n(\omega)\,.$$

The coefficients b_n satisfy the ordinary differential equation:

$$b_n'' + \frac{2}{r}b_n' - \frac{n^2+n}{r^2}b_n + \sigma^2 b_n = 0\,; \tag{9.47}$$

here $'$ denotes $\frac{d}{dr}$. Square integrability of $a(x)$ implies that the functions rb_n are square integrable. Introduce rb_n as a new variable:

$$rb_n = c\,, \quad b_n = \frac{c}{r}\,.$$

Then

$$b_n' = \frac{c'}{r} - \frac{c}{r^2}\,, \quad b_n'' = \frac{c''}{r} - \frac{2c'}{r^2} + \frac{2c}{r^3}\,;$$

by (9.47) c satisfies

$$c'' - \frac{n^2+n}{r^2}c + \sigma^2 c = 0\,. \tag{9.48}$$

By assumption, c is square integrable up to ∞; we shall show now that so is c'. Multiply (9.48) by c and integrate over $[r, l]$, where l may be taken arbitrarily large. We get, after integration by parts, that

$$cc'\Big|_r^l - \int_r^l c'^2 + \int_r^l \left(\sigma^2 - \frac{n^2+m}{s^2}\right)c^2 = 0\,. \tag{9.49$_l$}$$

Since the integral of c^2 from r to ∞ is finite, there is a sequence l tending to ∞ such that the derivative of c^2 at l tends to zero. Passing to the limit through such a sequence l in (9.49)$_l$ we conclude that c' is square integrable up to ∞.

Next multiply (9.48) by $2c'$ and integrate over $[r, l]$, and let l tend to ∞ over a sequence for which both c and c' tend to zero. We get

$$0 = -c'(r)^2 - (n^2+n)\int_r^\infty \frac{2cc'}{s^2}\,ds - \sigma^2 c^2(r)\,.$$

Integrate the middle term by parts:

$$0 = -c'(r)^2 + \frac{n^2+n}{r^2}c^2(r) - (n^2+n)\int_r^\infty \frac{2c^2}{s^3} - \sigma^2 c^2(r)\,.$$

For r large, the sum of the second and the fourth terms, $(\frac{n^2+n}{r^2} - \sigma^2)c^2(n)$, is negative, as are the first and third terms. This is a contradiction, proving that (9.48) has no nonzero solution square integrable up to ∞. □

And this, dear reader, completes the proof of assertions in parts (iii), (9.30)$_+$ and (9.30)$_-$ of Theorem 9.10.

From Theorem 9.10 we conclude by Sinai's theorem, Theorem 9.2 in Section 9.2, that there exist incoming and outgoing representations, $h \leftrightarrow k_-$ and $h \leftrightarrow k_+$, for $U(t)$ acting on H. These representations can also be obtained from the translation representation $g \leftrightarrow k_0$ of $U_0(t)$ acting in free space. We define the incoming representer k_- of any given g in F_- as $k_-(s) = k_0(s+R)$. We then define the representer k_- of $U(t)g$, g in F_- as $k_0(s+R-t)$. Since by part (iii) of (9.30)$_-$ $U(t)g$ are dense in H, this defines $h \leftrightarrow k_-$ for all h in H. So Sinai's theorem is not needed.

The outgoing representation $j \leftrightarrow k_+$ can be defined similarly.

Appendix E contains an astonishingly simple proof of local energy decay for bodies that are star-shaped.

9.5. The Semigroup Associated with Scattering by an Obstacle

Given a pair of orthogonal incoming and outgoing subspaces F_- and F_+ for a one-parameter group $U(t)$ of unitary operators, we have in Section 9.3 (see Theorem 9.7), constructed a semigroup $Z(t) = P_+U(t)P$. Here P_- is the projection that removes the F_- component; P_+ removes the F_+ component.

For waves scattered by an obstacle, with F_- and F_+ defined in Section 9.4, we can motivate this construction as follows:

Data in F_- do not undergo scattering until they are carried out of F_- by $U(t)$, $t > 0$. We remove them to start the scattering process immediately.

Data in F_+ do not undergo any further scattering; removing them focuses attention on the scattering process.

For scattering by an obstacle, $Z(t)$ has the following important property:

THEOREM 9.19 *$Z(t)$ is the semigroup defined by* (9.25), *and G its generator defined by* (9.27).

(i) *Every positive number κ belongs to the resolvent set of G, and the resolvent $(\kappa I - G)^{-1}$ is given by the formula*

$$(\kappa I - G)^{-1}k = \int_0^\infty Z(s)e^{-\kappa s}k\,ds\,, \tag{9.50}$$

where k is any element of K, the space on which $Z(t)$ acts.

(ii) *$Z(2R)(\kappa I - G)^{-1}$ is a compact operator.*

PROOF: The intuitive derivation of formula (9.50) is to write $Z(t)$ in the exponential form $Z(t) = e^{Gt}$ and to carry out the integration in (9.50) in a formal fashion. The rigorous proof is not hard. Denote by T the operator defined on the right side of (9.50):

$$Tk = \int_0^\infty Z(s)e^{-\kappa s}k\,ds\,.$$

We claim that the range of T belongs to the domain of G. To see this we form the difference

$$\begin{aligned} Z(t)Tk - Tk &= \int_0^\infty Z(T+s)e^{-\kappa s}k\,ds - \int_0^\infty Z(s)e^{-\kappa s}k\,ds \\ &= \int_t^\infty Z(r)e^{-\kappa r+\kappa t}k\,dr - \int_0^\infty Z(s)e^{-\kappa s}k\,ds \\ &= (e^{\kappa t}-1)\int_t^\infty Z(r)e^{-\kappa r}k\,dr - \int_0^t Z(s)e^{-\kappa s}k\,ds\,. \end{aligned}$$

Clearly the difference quotient (9.27)

$$\frac{Z(t)Tk - Tk}{t}$$

tends as $t \to 0$ to $\kappa Tk - k$. So

$$GTk = \kappa Tk - k\,,$$

from which (9.50) follows.

REMARK. In a similar fashion we can prove that for κ positive

$$(\kappa I - A)^{-i}h = \int_0^\infty U(s)e^{-\kappa s}h\,ds\,, \tag{9.50$'$}$$

where A denotes the infinitesimal generator of the group $U(t)$, and h is any element of H.

To prove part (ii) we need the following lemmas:

LEMMA 9.20 *Define the operator N as*

$$N = U(2R) - U_0(2R)\,. \tag{9.51}$$

(i) $\|N\|_E \le 2$.
(ii) *For any h in H, $Nh = 0$ for $|x| \ge 3R$.*
(iii) $\|Nh\|_E \le 2\|h\|_{E,5R}$.

(iv) *N maps the orthogonal complement of F_-, defined by* (9.30)$_-$, *into itself.*
(v) *N maps F_+, defined by* (9.30)$_+$, *into itself.*
(vi) $P_+U(2R)P_- = P_+NP_-$.

PROOF: Since both $U(2R)$ and $U_0(2R)$ are unitary operators, their norm is 1; so part (i) follows by the triangle inequality.

Parts (ii) and (iii) follow from Lemma 9.14.

To see why (iv) is true, we note that both $U_0(-2R)$ and $U(-2R)$ map F_- into itself. Therefore their adjoints $U_0(2R)$ and $U(2R)$ map the orthogonal complement of F_- into itself; but then so does their difference N.

To show (v) we note that $U(2R) = U_0(2R)$ on F_+, and both map F_+ into itself; then so does their difference N.

Finally, for (vi), in the free translation representation F_- and F_+ are represented by $L^2(-\infty, -R]$ and $L^2[R, \infty)$. The orthogonal complement of F_- is represented by $L^2[-R, \infty)$, whose translate by $2R$ represents F_+. This shows that $P_+U_0(2R)P_- = 0$; (vi) is an immediate consequence of this and the definition (9.51) of N. □

LEMMA 9.21

(i) *A set of elements h of H that satisfy an inequality of the form*

(9.52) $$\|Ah\|_E + \|h\|_E \le \text{const}$$

is precompact in any local energy norm $\|\cdot\|_{E,G}$.

(ii) $(\kappa I - A)^{-1}$ *maps the unit ball* $\|g\|_E \le 1$ *into a set that is compact in any local energy norm.*

PROOF: A is the 2×2 matrix operator $\left(\begin{smallmatrix} 0 & I \\ \Delta & 0 \end{smallmatrix}\right)$. Inequality (9.52) can be expressed as

$$\|D_x h_2\| + \|\Delta h_1\| + \|D_x h_1\| + \|h_2\| \le \text{const},$$

where $\|\cdot\|$ denotes the L^2 norm over the exterior of the obstacle. Since h_1 is zero on the boundary, the L^2 norms of all of its second derivatives over any bounded domain adjoining the boundary can be estimated in terms of the L^2 norm of Δh_1. The conclusion then follows from Rellich's compactness criterion.

(ii) We shall show that all elements of the form $h = (\kappa I - A)^{-1}g$, $\|g\|_E \le 1$, satisfy an inequality of form (9.52), in particular that

$$\|Ah\|_E = \|A(\kappa I - A)^{-1}g\|_E \le \text{const}\ .$$

This follows from the identity

$$A(\kappa I - A)^{-1} = -I + \kappa(\kappa I - A)^{-1}\ .$$

The conclusion of (ii) then follows from part (i). □

LEMMA 9.22

(9.53) $$P_-(\kappa I - A)^{-1}P_- = (\kappa I - A)^{-1}P_-\ .$$

PROOF: By definition, for t positive $U(-t)$ maps F_- into itself. Therefore its adjoint $U(t)$ maps the orthogonal complement of F_- into itself. It follows that $P_-U(t)P_- = U(t)P_-$ for all $t \geq 0$. Multiply this relation by $e^{-\kappa t}$ and integrate from 0 to ∞, and (9.53) results. □

We are now ready to complete the proof of Theorem 9.19. The following series of identities are based on the definition of $Z(t)$, that P_+ and P_- commute, that $P_+U(2R)P_+ = P_+U(2R)$, and on (9.51) and (9.53):

$$\begin{aligned} Z(2R)(\kappa I - G)^{-1}k &= Z(2R)\int_0^\infty Z(s)e^{-\kappa s}k\,ds \\ &= P_+U(2R)P_-P_+\int_0^\infty U(s)P_-e^{-\kappa s}k\,ds \qquad (9.54) \\ &= P_+NP_-(\kappa I - A^{-1})P_-k = P_+N(\kappa I - A)^{-1}P_-k\,. \end{aligned}$$

By part (ii) of Lemma 9.21, $(\kappa I - A)^{-1}$ maps the unit ball $\|h\|_E \leq 1$ into a set precompact in the $\|\cdot\|_{E.5R}$ norm. So part (ii) of Theorem 9.19 follows from the representation (9.54) of $Z(2R)(\kappa I - G)^{-1}$ as $P_+N(\kappa I - A)^{-1}P_-$; for according to part (iii) of Lemma 9.20, N is a bounded map from the space normed by $\|\cdot\|_{E.5R}$ into the space normed by $\|\cdot\|_E$. □

It is not hard to deduce from Theorem 9.19 that

(i) the generator G of the semigroup has a pure point spectrum,
(ii) the eigenvalues of G have no finite point of accumulation, and
(iii) the eigenvalues γ of G have negative real part.

Part (iii) follows from the assertion in Theorem 9.7 that $Z(t)$ tends strongly to 0 as $t \to \infty$.

More refined results about the spectrum of G depend on the details of the shape of the obstacle. The first result of this kind was given by Cathleen Morawetz, who proved that if the obstacle is star-shaped, then in any compact set C solutions with finite energy of the exterior problem for the wave equations decay uniformly as t^{-1}:

$$\|U(t)h\|_{E.C} \leq \frac{\text{const}}{t}\|h\|_E\,. \qquad (9.55)$$

From this it is not hard to deduce that the local decay is exponential:

THEOREM 9.23 *If the obstacle is star-shaped, then $Z(t)$ decays exponentially*

$$\|Z(t)\| \leq \text{const}\, e^{-\infty t}, \quad \alpha > 0\,. \qquad (9.56)$$

For proof we refer to Lax, Morawetz, and Phillips (1963), Morawetz (1961), or Appendix E.

The crucial property of the obstacle turns out to be whether it can confine rays indefinitely. A ray is a polygonal line in the exterior of the obstacle B, straight except at boundary points, where it is reflected according to the classical law of reflection. Denote by $L(B)$ the supremum of the lengths of all rays originating at a point of the sphere $|x| = R$ and remaining in $|x| \leq R$. Of course, $L(B)$ could be ∞.

THEOREM 9.24

(i) *If* $L(B) < \infty$, *then* $Z(t)$ *is a compact operator for* $t > L(B) + 12R$.

(ii) *If* $L(B) = \infty$, *then* $\|Z(t)\|_E = 1$ *for all* t.

PROOF: According to an important result of Melrose and Taylor, discussed in Appendix D, singularities of solutions of mixed initial-boundary value problems propagate along rays. This result can be formulated as follows:

THEOREM 9.25 *Let* C_1 *and* C_2 *be closed bounded sets in the exterior of the obstacle. Suppose no ray originating in* C_1 *at time* 0, *traveling with speed* 1, *lies at time* T *in* C_2. *Then* $U(T)$ *maps data of* E*-norm* ≤ 1 *supported in* C_1 *into a precompact set in the norm* $\|\cdot\|_{E,C_2}$.

In the application we wish to make of this result we take C_1 to consist of those points of the exterior that belong to the ball $|x| \leq 3R$. We take C_2 to consist of those points that belong to the ball $|x| \leq 5R$. It is easy to see that any ray originating in C_1 and traveling with speed 1 leaves C_2 by the time $T = L(B)+8R$. We factor now $Z(T+4R)$ as follows:

$$\begin{aligned} Z(T+4R) &= Z(2R)Z(T)Z(2R) \\ &= P_+U(2R)P_-P_+U(T)P_-P_+U(2R)P_- \\ &= P_+U(2R)P_-U(T)P_+U(2R)P_- \,. \end{aligned} \tag{9.57}$$

The legitimacy of omitting P_+ and P_- in the middle has been explained in Section 9.3 where the semigroup property of $Z(t)$ is shown.

We now make use of property (vi) in Lemma 9.20 to rewrite (9.57) as

$$Z(T+4R) = P_+NP_-U(T)P_+NP_- \,. \tag{9.57$'$}$$

According to part (iv) of Lemma 9.20 N maps the orthogonal complement of F_- into itself; so do $U(T)$ and P_+. Therefore we may omit the third factor P_- on the right in (9.57)′. According to part (v), N maps F_+ into itself, as do $U(T)$ and P_- Therefore we may drop the sixth factor P_+ on the right in (9.57)′. Altogether we get

$$Z(T+4R) = P_+NU(T)NP_- \,. \tag{9.58}$$

According to part (ii) of Lemma 9.20, N maps H into data supported in the ball $|x| \leq 3R$. According to Theorem 9.25 $U(T)$ maps such data into a set that is precompact in the norm $\|\cdot\|_{E,5R}$.

We appeal now to Lemma 9.20, part (iii), to conclude that $NU(T)N$ maps the unit ball $\|h\|_E \leq 1$ into a precompact set in the $\|\cdot\|_E$ norm. In view of formula (9.58), this proves part (i) of Theorem 9.24.

Part (ii) was proved by Jim Ralston by constructing solutions whose bulk travels along rays. □

An obstacle for which $L(B) < \infty$ is called *nonconfining*.

According to Theorem 9.24, for a nonconfining obstacle $Z(t)$ is compact for t large enough. Therefore for such t, $Z(t)$ has a pure point spectrum that accumulates only at 0. Since $\|Z(t)\|_E \leq 1$, the eigenvalues are ≤ 1 in absolute value. By the spectral mapping theorem, the same holds for all t.

For $t > 0$, $Z(t)$ has no eigenvalue of absolute value 1, since it was shown in Theorem 9.7 of Section 9.3 that $Z(t)$ tends to 0 strongly as $t \to \infty$. So the spectral radius ρ of $Z(1)$ is less than 1. By the formula for the spectral radius

$$\lim_{t\to\infty} \|Z(n)\|_E^{1/n} = \rho;$$

this shows that $Z(t)$ decays exponentially as $t \to \infty$.

THEOREM 9.26 *When B is a nonconfining obstacle, $U(t)h$ decays exponentially on any compact set, for any h in H.*

PROOF: $U(t)h = Z(t)h$ for $|x| < R$. Since $Z(t)h$ decays exponentially, so does $U(t)h$ for $|x| \le R$. But R is an arbitrary number. □

9.6. Analytic Form of the Scattering Matrix

By now we know quite a bit about the scattering matrix—it is analytic in the lower half of the complex plane, meromorphic in the upper half-plane. It is high time we learned how to calculate it. To this end we shall turn to the spectral representations.

The free space spectral representation is essentially the Fourier transform. Define

$$e_0(x, \sigma, \omega) = e^{-i\sigma x\cdot\omega}\{1, i\sigma\}. \tag{9.59}$$

e_0 satisfies the differential equation

$$A_0 e_0 = i\sigma e_0, \tag{9.60}$$

where A_0 is the generator of the group $U_0(t)$:

$$\frac{d}{dt}U_0(t) = A_0 U_0(t), \quad A_0 = \begin{pmatrix} 0 & I \\ \Delta & 0 \end{pmatrix}. \tag{9.61}$$

Since A_0 is the generator of a group of unitary operators, it is antisymmetric: $A_0^* = -A_0$.

THEOREM 9.27 *We define the spectral representer of any C_0^∞ data g_0 in H_0 as*

$$a_0(\sigma, w) = (g_0, e_0)_E. \tag{9.62}$$

This is a spectral representation of $U_0(t)$.

PROOF: The representer of $U_0(t)g_0$ is

$$(U_0(t)g_0, e_0)_E.$$

We shall show that it satisfies an ordinary differential equation

$$\begin{aligned}\frac{d}{dt}(U_0(t)g_0, e_0)_E &= \left(\frac{d}{dt}U_0(t)g_0, e_0\right)_E \\ &= (A_0U_0(t)g_0, e_0)_E = -(U(t)g_0, A_0e_0)_E = i\sigma(U_0(t)g_0, e_0)_E.\end{aligned}$$

In this derivation we have used (9.61), the antisymmetry of A_0, and that e_0 is an eigenfunction of A_0.

The solution of the above differential equation is

$$(U_0(t)g_0, e_0)_E = e^{i\sigma t}(g_0, e_0)_E ,$$

proving that (9.62) is a spectral representation of $U_0(t)$. The isometry of (9.62) follows from the Parseval relation for the Fourier transform. Using the definition of the energy scalar product we write $g_0 = \{g_1, g_2\}$:

$$(g_0, e_0)_E = (D_x g_1, D_x e^{-i\sigma x\cdot\omega}) - i\sigma(g_2, e^{-i\sigma x\cdot\omega}) .$$

Integrating by parts the first term gives

$$\sigma^2(g_1, e^{-i\sigma x\cdot\omega i}) - i\sigma(g_2, e^{-i\sigma x\cdot\omega}) ,$$

where $(\cdot, \cdot)$ denotes the L^2 scalar product. So

$$(g_0, e_0)_E = \sigma^2\tilde{g}_1(\sigma\omega) - i\sigma\tilde{g}_2(\sigma\omega) , \tag{9.63}$$

where $\tilde{}$ denotes the Fourier transform. The first term is an even function of σ and ω, the second term odd. The sum of their L^2 norms is, by Parseval,

$$\|D_x g_1\|^2 + \|g_2\|^2 .$$

Finally, we show that a_0, defined by (9.62), is the Fourier transform with respect to s of the translation representer k_0 defined in (9.11) as

$$k_0(s, \omega) = -\partial_s^2\hat{g}_1 + \partial_s\hat{g}_2 ,$$

where $\hat{}$ denotes the Radon transform. Denote by F the Fourier transform with respect to s; applying F to k_0 gives

$$Fk_0 = \sigma^2 F\hat{g}_1 - i\sigma F\hat{g}_2 .$$

As we have shown in Section 3.5 of Chapter 3, $F\hat{g} = \tilde{g}$; so we conclude from (9.63) that

$$Fk_0 = a_0 .$$

□

We turn now to the much more interesting task of constructing incoming and outgoing spectral representations for the group $U(t)$. To make matters simple, we shall treat scattering by obstacles that do not confine rays. In this case $U(t)h$ decays exponentially on any compact subset of the exterior of the obstacle, as shown in Theorem 9.26.

The incoming and outgoing spectral representations will be of a similar form as (9.62):

$$a_-(\sigma, w) = (h, e_-)_E , \tag*{(9.64)$_-$}$$

where e_- is an eigenfunction of A, the generator of the group $U(t)$:

$$Ae_- = i\sigma e_- . \tag*{(9.65)$_-$}$$

e_- will be constructed as a perturbation of e_0:

$$e_- = e_0 + f_+ . \tag*{(9.65)$'_-$}$$

To serve its purpose, the function f_+ has to have the following properties:

(i) For $(9.65)_-$ to hold, f_+ has to satisfy the differential equation

$$Af_+ = i\sigma f_+ . \tag{9.66}$$

(ii) e_- has to be zero on ∂B; therefore f_+ must satisfy

$$f_+(x) = -e_0(x) \quad \text{on } \partial B .$$

(iii) In order that the free space and incoming spectral representations be the same for all h in F_-, f_+ must be orthogonal to F_-.

THEOREM 9.28 *There exists a unique f_+ that has all three properties* (i)–(iii) *listed above.*

PROOF: We start by choosing a smooth cutoff function $\zeta(x)$ with the following properties:

$$\zeta(x) = \begin{cases} 0 & \text{for } |x| \geq R , \\ 1 & \text{in an open set containing } B . \end{cases}$$

We set

$$f_+ = -\zeta e_0 + p . \tag{9.67}$$

In order for f_+ to satisfy (9.66) p must satisfy

$$(A - i\sigma)p = (A - i\sigma)\zeta e_0 . \tag{9.68}$$

Denote the right side of (9.68) by g. A simple calculation shows that

$$g = (A - i\sigma)\zeta e_0 = \{0, r\} ,$$

where

$$r = (\Delta\zeta - 2i\sigma D_x\zeta \cdot \omega)e^{i\sigma x\cdot\omega} .$$

The support of r lies outside of B and inside $|x| \leq R$, so that g belongs to H.

We define p by the explicit formula

$$p = -\int_0^\infty e^{-i\sigma t}U(t)g\,dt . \tag{9.69}$$

We claim that f_+ given by (9.67) and (9.69) has properties (i), (ii), and (iii) above. First we note that the integral converges for all x, since, according to Theorem 9.26, $U(t)g$ decays exponentially as a function of t. Since g is a smooth function of x, the x-derivatives of $U(t)g$ also decay exponentially. Therefore x differentiation of p, given by formula (9.69), can be carried out under the integral sign:

$$\begin{aligned}(A - i\sigma)p &= -\int_0^\infty e^{-i\sigma t}(A - i\sigma)U(t)g\,dt \\ &= -\int_0^\infty e^{-i\sigma t}\left(\frac{d}{dt} - i\sigma\right)U(t)g\,dt \\ &= -\int_0^\infty \frac{d}{dt}(e^{-i\sigma t}U(t)g)\,dt = g .\end{aligned}$$

Since g belongs to H, so do $U(t)g$, and therefore they are zero on the boundary. But then so is their integral. We have thereby shown that f_+ defined by (9.67) has properties (i) and (ii).

What remains to be shown is property (iii). We shall show that for all smooth h_- in F_- of compact support,

$$(f_+, h_-)_E = 0 .$$

This scalar product makes sense, since h_- is assumed to have compact support. Using the definition (9.67) of f_+ we can write the equation above as

$$-(\zeta e_0, h_-)_E + (p, h_-)_E = 0 .$$

Since the cutoff function ζ is zero for $|x| > R$, the supports of ζe_0 and h are disjoint, and so the first term above is zero. We turn now to the second term; using the definition of p we write

$$(p, h_-)_E = -\left(\int_0^\infty e^{-i\sigma t} U(t) g \, dt, h_-\right)_E = -\int_0^\infty e^{-i\sigma t} (U(t)g, h_-)_E \, dt$$
$$= -\int_0^\infty e^{-i\sigma t} (g, U(-t)h_-)_E \, dt .$$

In the last step we have used the fact that the adjoint of $U(t)$ is $U(-t)$. $U(-t)$ maps F_- into itself; g is zero outside of the ball $|x| = R$; therefore the value of the integrand above is zero for all $t \geq 0$. But then so is their integral.

To nail down this argument we need the following piece of information:

LEMMA 9.29 *The set of smooth elements h_- in F_- that have compact support is dense in F_-.*

We shall present a proof of this result at the end of this section.

We leave it to the reader to verify that the f_+ we constructed is independent of the choice of the cutoff function ζ.

Next we verify that

$$a_-(\sigma, \omega) = (h, e_-)_E \tag{9.64}$$

gives the incoming spectral representation of $U(t)$. Here h can be any smooth data in H of compact support.

Set in (9.64) $U(t)h$ in place of h:

$$a_-(\sigma, \omega, t) = (U(t)h, e_-)_E .$$

Differentiate with respect to t, integrate by parts, and use the fact that e_- is an eigenfunction of A:

$$\begin{aligned} \frac{d}{dt} a_-(\sigma, \omega, t) &= (AU(t)h, e_-)_E \\ &= -(U(t)h, Ae_-)_E = i\sigma (U(t)h, e_-)_E \\ &= i\sigma a_-(\sigma, \omega, t) . \end{aligned}$$

It follows that $U(t)h$ is represented by $e^{i\sigma t} a_-(\sigma, \omega)$. From part (iii) it follows that the representer of any h in F_- is the same as its free space spectral representer. So the representation is isometric for h in F_-; it follows that it is isometric for all $U(t)h$. According to the basic Theorem 9.10, these are dense in H, so we can, by closure, define a spectral representer for all h in H.

We are now ready to complete the proof of Theorem 9.28 and show that properties (i), (ii), and (iii) completely determine f_+. Denote by q the difference of two choices for f_+. Since both choices give the same spectral representation, it follows that $(h, q)_E = 0$ for all h in H that have compact support. It follows from the fact that $e_- + f_+$ is an eigenfunction that q is of the form $\{v, i\sigma v\}$. Now choose $h = \{0, \zeta v\}$, where ζ is a real, nonnegative cutoff function at ∞, and we get that

$$(\zeta v, v) = 0,$$

from which $v = 0$ follows. □

The most important information about f_+ is its asymptotic behavior for $|x|$ large. To determine that, we shall represent f_+ in terms of its translation representer in free space. Since the energy of f_+ is infinite, we have to generalize the concept of translation representer.

Extend f_+ smoothly into the interior of B; f_+ thus extended satisfies the equation

$$(A_0 - i\sigma)f_+ = g_0,$$

where g_0 is supported on the obstacle B. We can solve this equation as we have solved (9.68):

$$f_+ = -\int_0^\infty e^{-i\sigma t} U_0(t) g_0 \, dt. \tag{9.70}$$

By Huygens' principle, $U_0(t)g_0$ is zero for $|x| < t - R$; therefore for x in a compact set the integral is over a finite range of t. Define f_+^T as

$$f_+^T = -\int_0^T e^{-i\sigma t} U_0(t) g_0 \, dt. \tag{9.70^T}$$

It follows that

$$f_+^T(x) = f_+(x) \quad \text{for } |x| < T - R. \tag{9.71}$$

We denote by $k_0(s, \omega')$ the free space translation representer of g_0. The angular variable is denoted as ω'; g_0 depends on σ and ω as parameters, and therefore so does k_0. Since g_0 is supported in the ball $|x| \leq R$, k_0 is supported in $-R \leq s \leq R$.

By definition (9.70)^T, the free space translation representer of f_+^T, is

$$-\int_0^T e^{-i\sigma t} k_0(s - t, \omega') dt = j_0^T(s, \omega'),$$

which can be rewritten as

$$j_0^T = -e^{-i\sigma s} \int_{s-T}^{s} e^{i\sigma b} k_0(b, \omega') db. \tag{9.72^T}$$

f_+^T can be expressed in terms of its translation representer:

$$f_+^T(x)_1 = \int j_0^T(x \cdot \omega', \omega') d\omega'; \tag{9.73}$$

a similar expression holds for the second component of f_+^T.

The support of j_0^T lies in $-R \le s \le R + T$. Let h_- be any element of F_-, $k_-(s, \omega')$ its free space translation representer; k_- is supported on $s \le -R$, disjoint from the support of j_0^T. Therefore

$$(h_-, f_+^T)_E = (k_-, j_0^T) = 0. \tag{9.74}$$

We are now ready to let $T \to \infty$. According to (9.71) $f_+^T(x)$ tends to $f_+(x)$. We let $T \to \infty$ in $(9.72)^T$ and define

$$j_0(s, \omega') = -e^{-i\sigma s} \int_{-\infty}^{s} e^{i\sigma b} k_0(b, \omega') db. \tag{9.72}$$

Since $k_0(b, \omega') = 0$ for b outside $[-R, R]$,

$$j_0(s, \omega') = e^{-i\sigma s} n(s, \omega'), \tag{9.75}$$

where

$$n(s, \omega') = \begin{cases} 0 & \text{for } s < -R, \\ n(\omega') & \text{for } R < s; \end{cases} \tag{9.75$'$}$$

here

$$n(\omega') = \int_{-R}^{R} e^{i\sigma b} k_0(b, \omega')\, db.$$

If we compare $(9.72)^T$ and (9.72), we conclude that for $T - R > s$,

$$j_0^T(s, \omega') = j_0(s, \omega').$$

Therefore choosing $T > |x| + R$ we conclude from this, (9.71), and (9.73) that for every x,

$$f_+(x)_1 = \int j_0(x \cdot \omega', \omega') d\omega'. \tag{9.76}$$

Suppose h_- belongs to F_- and has compact support. Then by (9.71), $f_+(x) = f_+^T(x)$ on the support of $h_-(x)$ for T large enough, and so it follows from (9.74) that

$$(h_-, f_+)_E = 0;$$

this shows that f_+, as defined by (9.70), has the required property (iii).

The asymptotic behavior of $f_+(x)$ for $|x| = r$ large can be determined from formula (9.76). Take x to be $r(1, 0, 0)$, and parametrize ω' as

$$\omega' = (\cos\phi, \sin\phi\cos\psi, \sin\phi\sin\psi), \tag{9.77}$$

ϕ in $[0, \pi]$, ψ in $[0, 2\pi]$, $d\omega' = \sin\phi\, d\phi\, d\psi$, and $x \cdot \omega' = r\cos\phi$. We set this in (9.76), using the definition (9.72) of j_0:

$$f_+(x)_1 = -\int_0^{\pi} e^{-i\sigma r\cos\phi} \sin\phi \int_{-\infty}^{r\cos\phi} e^{i\sigma b} k_0(b, \omega') db\, d\phi\, d\psi.$$

Introduce $\cos\phi = c$ as a new variable of integration:

$$f_+(x)_1 = -\int_{-1}^{1} e^{-i\sigma rc} \int_{-\infty}^{rc} e^{i\sigma b} k_0(b, \omega') db\, dc\, d\psi.$$

We integrate by parts with respect to c, integrating the first factor $e^{-i\sigma rc}$, differentiating the second factor

$$\int_{-\infty}^{rc} e^{i\sigma b} k_0(b, \omega')\,db\,.$$

Since $k_0(b, \omega')$ is zero for $|b| > R$, the second factor is zero for $c = -1$ when $r > R$. For $c = 1$ it is equal to

$$n(\omega') = \int_{-\infty}^{\infty} e^{i\sigma b} k_0(b, \omega')\,db\,.$$

So we get

$$f_+(x)_1 = \frac{e^{-i\sigma r}}{r} m_+ - \frac{1}{i\sigma} \int_{-1}^{1} k_0(rc, \omega')\,dc\,d\psi\,, \qquad m_+ = m_+(\sigma, \omega)\,.$$

We claim that the second term in the above formula for $f_+(x)_1$ is $O(1/r^2)$. To see this we note that outside the range $|c| \le R/r$ the integrand $k_0(rc, \omega')$ is zero. In this range we may replace ω', given in formula (9.77), by $\omega(\psi) = (0, \cos\psi, \sin\psi)$, with an error that is $O(1/r^2)$. In the remaining integral we introduce $s = rc$ as a new variable of integration; we get

$$\frac{1}{i\sigma r} \int_{-\infty}^{\infty} k_0(s, \omega(\psi))\,ds\,d\psi\,.$$

But it follows from formula (9.11) that the s-integral is, for fixed ω, equal to zero. This proves that as x tends to ∞,

$$f_+(x)_1 = \frac{e^{-i\sigma r}}{r} m_+ + O\left(\frac{1}{r^2}\right), \tag{9.78}$$

where $|x| = r$.

So far we have taken x to be $r(1, 0, 0)$. But clearly the argument can be extended to any x. In the general case, m in (9.78) will be a function of $\theta = x/r$, and of ω and σ, which enter the function g_0, and therefore also k_0, as parameters.

In Theorem 9.28, f_+ was characterized by the following conditions:

(i) $f_+ = \{f, i\sigma f\}$, where f satisfies $\Delta f + \sigma^2 f = 0$ outside B.
(ii) $f(x) = -e^{-i\sigma x\cdot\omega}$ on ∂B.
(iii) f_+ is orthogonal to F_-.

We have shown that such an f exists, and that it is uniquely characterized by these conditions. The asymptotic behavior of $f(x)$ for large x is given by formula (9.78).

We shall show now that condition (iii) may be replaced by (9.78).

THEOREM 9.30 *There is exactly one solution of the reduced wave equation in the exterior of B with prescribed boundary values* (ii) *whose asymptotic behavior is of the form* (9.78).

PROOF: The asymptotic behavior of $\frac{\partial}{\partial r} f$ as $|x| \to \infty$ can be determined by the same argument as was used for f itself. The result is the same as that obtained by differentiating (9.78) with respect to r:

$$\partial_r f(x) \simeq -\frac{i\sigma e^{-i\sigma r}}{r} m + O\left(\frac{1}{r^2}\right). \tag{9.78$_r$}$$

Combine (9.78) and (9.78)$_r$; as $x \to \infty$,

$$|\partial_r f + i\sigma f| = O\left(\frac{1}{r^2}\right). \tag{9.78$'$}$$

(9.78)$'$ is called the *Sommerfeld radiation condition.*

Suppose that there are two functions satisfying (i), (ii), and (9.78)$'$. Their difference, again denoted as f, satisfies $\Delta f + \sigma^2 f = 0$ in the exterior of B, and $f = 0$ on ∂B. Furthermore, f satisfies the asymptotic relation ((9.78)$'$) as $r \to \infty$.

Denote by S_r the sphere $|x| = r$.

$$\int_{S_r} |\partial_r f + i\sigma f|^2 dS = \int_{S_r} |\partial_r f|^2 + \sigma^2 |f|^2 dS + i\sigma \int_{S_r} (f\partial_r \bar{f} - \partial_r f \bar{f}) dS .$$

We claim that the third term on the right is zero. To see this, integrate $\bar{f}(\Delta f + \sigma^2 \bar{f})$ in the region contained between B and S_r. Since $f = 0$ on ∂B, after integration by parts we obtain the desired result.

Integrate the identity above from R to ∞. By estimate (9.78)$'$ the left side is finite; therefore so is the right side; this shows that f is L^2 at ∞. But according to the Rellich-Vekua theorem, Lemma 9.18, this implies that $f \equiv 0$. □

We turn now to a physical interpretation of f_+. What happens when a wave impinges on an obstacle? We take the wave to be a harmonic plane wave $e^{i\sigma(t-\omega\cdot x)}$, traveling in the direction ω, with frequency $\sigma/2\pi$. To make this wave impinge on B we choose a smooth cutoff function ξ,

$$\xi(x) = \begin{cases} 0 & \text{on an open set containing } B\,, \\ 1 & \text{for } R \le |x|\,, \end{cases}$$

and take the initial value ξe_0 for the exterior problem. We shall show that as $T \to \infty$,

$$U(T)\xi e_0 \to e^{i\sigma T} e_- , \tag{9.79}$$

uniformly on any compact set in the exterior of the obstacle B.

PROOF: By (9.65)$'_-$, $e_- = e_0 + f_+$; therefore $e_- - \xi e_0 = (1 - \xi)e_0 + f_+$. Taking $1 - \xi$ to be ζ and using (9.67) we get

$$e_- - \xi e_0 = p .$$

Applying $U(T)$ we get

$$e^{i\sigma T} e_- - U(T)\xi e_0 = U(T)p . \tag{9.79$'$}$$

Using the definition (9.69) of p we can write

$$U(T)p = -\int_0^\infty e^{-i\sigma t}U(t+T)g\,dt$$
$$= e^{i\sigma T}\int_T^\infty e^{-i\sigma s}U(s)g\,ds\,.$$

According to Theorem 9.26, $U(s)g$ tends to zero exponentially as $s \to \infty$, uniformly on compact sets. Therefore so does $U(T)p$ as $T \to \infty$. This shows that (9.79)′ tends to zero as $T \to \infty$; from this (9.79) follows.

The signal $e^{i\sigma T}e_-$ consists of two parts, $e^{i\sigma T}e_0$ plus $e^{i\sigma T}f_+$. Our analysis shows that $e^{i\sigma T}e_0$ is the incident signal, $e^{i\sigma T}f_+$ is the reflected signal.

In an experimental setup the wave $U(t)\xi e_0$ is sent in, and the reflected wave $e^{i\sigma t}f_+$ is measured. Since the measurement is made at some distance from the obstacle, only the leading term of the reflected wave

$$\frac{e^{i\sigma(t-r)}}{r}m_+(\theta,\omega;\sigma)\,, \tag{9.80}$$

is measured.

Note that (9.80) is an *outgoing spherical wave.*

The *inverse problem* of scattering theory is to determine the shape of the obstacle from the scattering data $m_+(\theta, w; \sigma)$.

THEOREM 9.31 *The scattering data $m_+(\theta,\omega;\sigma)$ uniquely determine the obstacle.*

PROOF: Take two obstacles B and B', with e_- and e'_- the incoming eigenfunctions. According to (9.65)$'_-$,

$$e_- = e_0 + f_+, \quad e'_- = e_0 + f'_+\,.$$

Their difference is

$$e_- - e'_- = f_+ - f'_+\,.$$

Denote by m_+ and m'_+ the terms appearing in the asymptotic description (9.78) of f_+ and f'_+. Suppose that for some fixed value of ω and σ, $m_+(\theta,\omega;\sigma) = m'_+(\theta,\omega;\sigma)$ for all θ. Then the difference $e_- - e'_-$ is $O(1/r^2)$ as $x \to \infty$. This implies that $e_- - e'_-$ is square integrable around ∞. Since the components of e_- and e'_- satisfy the reduced wave equation (9.46), it follows from the Rellich-Vekua theorem (see Lemma 9.18) that $e_- - e'_-$ is zero in a neighborhood of ∞. But since solutions of the reduced wave equation are analytic functions of x, it follows that $e_- - e'_- = 0$ at every point of the set in the exterior of both B and B' that is connected to infinity within this set.

The set of points that belong to B' but lie outside of B is the union of connected components G_j. The boundary if each G_j belongs either to the boundary of B or of B'; therefore e_- is zero on the boundary of each G. The first component of e_- is a solution of the reduced wave equation, and therefore an eigenfunction in G of $-\Delta$, equal to zero on the boundary of G, with eigenvalue σ^2. If both B and B' are contained in a ball of radius R, so is G, and we can estimate from above the number N of eigenvalues $\leq \sigma^2$.

Suppose $m_+(\theta, \omega; \sigma) = m'_+(\theta, \omega; \sigma)$ for all θ and $N + 1$ values of ω. That would give $N + 1$ eigenfunctions of $-\Delta$ with eigenvalue σ^2, too many, a contradiction, into which we were led by assuming that $B \neq B'$. □

The inverse problem of scattering is to reconstruct the obstacle B from the observed scattering amplitudes $m_+(\theta, \omega; \sigma)$ for all θ and a finite set of ω. This would require measuring the scattered waves for all θ, that is, all around the obstacle. This is not practical. However, one of the results of scattering theory, the reciprocity law, says that $m_+(\theta, \omega; \sigma)$ is a *symmetric function* of θ and ω:

$$m_+(\theta, \omega; \sigma) = m_+(\omega, \theta; \sigma).$$

REMARK. Merely proving that a certain set of scattering data uniquely determines the scatterer is a far cry from solving the inverse problem. That task calls for algorithms to actually generate the obstacle from the scattering data. We refer the reader to chapter 9 in Michael Taylor's book on scattering theory (1996) for some thoughts on this subject.

We turn now to the scattering matrix $M(\sigma)$. Since both the incoming and outgoing spectral representations are perturbations of the free space spectral representation, it is no surprise that the scattering matrix is of the form

$$M(\sigma) = I + K(\sigma),$$

where K is an integral operator. The kernel of K can be expressed in terms of the scattering amplitude $m_+(\theta, \omega; \sigma)$.

We conclude this section by presenting a proof of Lemma 9.29, that the set of smooth elements h_- in F_- that have compact support is dense in F_-. We argue indirectly: suppose not. Then there would be a nonzero k in F_- that is orthogonal to all h_-.

It is easy to construct data h_- in F_- that have compact support. Take any g in H_0 that is supported in $|x| < T - R$, and set $h_- = U_0(-T)g$. By Huygens' principle, h_- belongs to F_-, and since signals propagate with speed ≤ 1, h_- is supported in $|x| \leq 2T - R$. All such h_- would be orthogonal to k:

$$(U_0(-T)g, k)_E = 0.$$

Since the adjoint of $U_0(-T)$ is $U_0(T)$,

$$(g, U_0(T)k)_E = 0.$$

We may take for the second component of g any smooth function supported in $|x| < T - R$, and take the first component zero. It follows that the second component of $U_0(T)k$ is zero for $|x| < T - R$.

Denote by $u_0(x, t)$ the solution of the wave equation in free space with initial data k. We have shown that $\partial_t u_0(x, t) = 0$ in the cone $|x| < t - R$. It follows that $u_0(x, t) = u_0(x)$ for $|x| < t - R$. Since $\partial_t^2 u_0 = 0$ for $|x| < t - R$, it follows from $\partial_t^2 u_0 - \Delta u_0 = 0$ that $u_0(x)$ is a harmonic function. Since $\int_{|x| \leq R} |\partial_x u_0|^2 \, dx$ is less than the energy of k, $u_0(x)$ has finite energy. But a harmonic function defined in the exterior of a compact set that has finite energy is a constant. We claim that this constant is zero.

The proof is based on the following estimate for functions f defined in $\mathbb{R}^3$, differentiable and of compact support:

For every R,

$$\int_{|x|\leq R} |f(x)|^2\,dx \leq \frac{R^2}{2}\int |f_x|^2\,dx\,. \tag{9.81}$$

To derive this inequality write for $x = r\theta$,

$$f(x) = -\int_r^\infty \frac{d}{dr} f(r\theta)dr\,.$$

By the Schwarz inequality

$$|f(x)|^2 \leq \int_r^\infty r^{-2}\,dr \int_r^\infty f_r^2 r^2\,dr = r^{-1}\int_r^\infty f_r^2 r^2\,dr\,.$$

Integrating with respect to θ on the unit sphere gives

$$\int |f(r\theta)|^2\,d\theta \leq r^{-1}\int_{|x|\geq r} f_r^2\,dx \leq r^{-1}\int |f_x|^2\,dx\,.$$

Multiply this inequality by r^2 and integrate dr from 0 to R to obtain the estimate (9.81).

Inequality (9.81) holds for any function $f(x)$ that is the limit in the norm ($\int f_x^2\,dx$) of C_0^1 functions. The harmonic function $u_0(x)$ is such a limit, and therefore

$$\int_{|x|\leq R} u_0(x)^2\,dx \leq \frac{R^2}{2}\int |\partial_x u_0|^2\,dx \leq R^2|k|_E^2\,. \tag{9.82}$$

We have already shown that $u_0(x) \equiv$ const. If that constant were nonzero, the left side of (9.82) would grow like R^3, contradicting (9.82) for R large enough.

We have thus shown that $u_0(x, t) = 0$ in the cone $|x| < t - R$. Since k belongs to F_-, $u_0(x, t) = 0$ in the cone $|x| < -t + R$. It follows then from Lemma 9.11 that the translation representer of k is zero, and therefore so is k itself, contrary to assumption. This completes the indirect proof of Lemma 9.29. □

9.7. Scattering of Automorphic Waves

Boris Pavlov and Ludwig Faddeev have given a beautiful analysis of the scattering of automorphic waves.

The scene of action is hyperbolic plane $\mathbb{H}$. In the Poincaré model this is the upper x, y plane, $y > 0$, equipped with the Riemannian metric

$$ds^2 = \frac{dx^2 + dy^2}{y^2}\,. \tag{9.83}$$

Motions in the hyperbolic plane can be expressed as fractional linear transformations in the complex variable $z = x + iy$:

$$z \to \frac{az+b}{cz+d}\,, \quad a, b, c, d \text{ real}, \quad ad - bc = 1\,. \tag{9.84}$$

THEOREM 9.32

(i) *The mappings* (9.84) *preserve the metric* (9.83).

(ii) *The hyperbolic* L^2 *norm*

$$\iint u^2 \frac{dx\,dy}{y^2}$$

is invariant under the mappings (9.84).

(iii) *The hyperbolic Dirichlet form*

$$D(u) = \iint (u_x^2 + u_y^2)dx\,dy$$

is invariant under the mappings (9.84).

(iv) *The Laplace-Beltrami operator*

$$\Delta_{\mathbb{H}} = y^2 \Delta \tag{9.85}$$

is invariant under the mappings (9.84).

Proof is left as an exercise for the reader.

The renormalized Laplace-Beltrami operator is

$$L = \Delta_{\mathbb{H}} + \frac{1}{4}, \tag{9.86}$$

and the hyperbolic wave equation is

$$u_{tt} - Lu = 0. \tag{9.87}$$

The conservation is energy is derived just as in the Euclidean case. Multiply (9.87) by $2u_t$ and integrate by parts. We get

$$0 = 2\int_{\mathbb{H}} u_t\left(u_{tt} - y^2\Delta u - \frac{1}{4}u\right)\frac{dx\,dy}{y^2} = \frac{d}{dt}\int_{\mathbb{H}}\left(\frac{u_t^2}{y^2} + u_x^2 + u_y^2 - \frac{1}{4y^2}u^2\right)dx\,dy.$$

This proves that

$$E(u) = \int_{\mathbb{H}}\left(u_x^2 + u_y^2 - \frac{1}{4y^2}u^2 + \frac{u_t^2}{y^2}\right)dx\,dy \tag{9.88}$$

is a constant of the motion.

A useful form for energy is

$$E(u) = -(u, Lu) + (u_t, u_t), \tag{9.88'}$$

where (,) denotes the L^2 scalar product in the metric (9.83).

For data with compact support, energy is positive, in spite of the presence of a negative term in (9.88). Integration by parts gives

$$\int_0^\infty \left(u_y + \frac{1}{2y}u\right)^2 dy = \int_0^\infty \left(u_y^2 + \frac{uu_y}{y} + \frac{u^2}{4y^2}\right)dy$$
$$= \int_0^\infty \left(u_y^2 - \frac{u^2}{2y^2} + \frac{u^2}{4y^2}\right)dy = \int_0^\infty \left(u_y^2 - \frac{u^2}{4y^2}\right)dy.$$

The Pavlov-Faddeev theory deals with automorphic solutions of the hyperbolic wave equation, with respect to to the *modular group*, a subgroup of the group Γ of motions (9.84) characterized by the condition that *the coefficients a, b, c, d are integers*. A function $u(x, y)$ defined in the hyperbolic plane is called *automorphic* with respect to Γ if $u(\gamma(x, y)) = u(x, y)$ for all motions γ in Γ.

Automorphic functions are hyperbolic analogues of periodic functions in Euclidean space. The analogue of a period parallelogram is a *fundamental domain*, defined as follows:

A subset P of $\mathbb{H}$ is called a *fundamental domain* for Γ if

(i) every point of $\mathbb{H}$ can be mapped into a point of P by some γ in Γ,
(ii) only boundary points of P are mapped into each other by a $\gamma \neq I$ in Γ.

It is a classical fact that set T defined by the inequalities $-\frac{1}{2} \leq x \leq \frac{1}{2}$, $x^2 + y^2 \geq 1$ is a fundamental domain for Γ. Its boundaries, $x = -\frac{1}{2}$, $x = \frac{1}{2}$, $x^2 + y^2 = 1$ are *geodesics* of the metric (9.83), as the reader may verify. Thus T is a *geodesic triangle*, one of whose vertices lies at ∞. Nevertheless, the area of T is finite; the reader is invited to verify these facts. The mapping $x, y \to x + 1, y$ carries the left edge of T onto the right edge. The mapping $z \to -\frac{1}{z}$ carries the bottom edge of T onto itself.

NOTE. These two maps generate Γ.

It follows from property (i) of a fundamental domain that the image of T under any γ in the modular group is again a fundamental domain, and that the union of $\gamma(T)$, $\gamma \in \Gamma$, is the whole hyperbolic plane $\mathbb{H}$. These facts lead to the following construction of continuous automorphic functions $u(x, y)$:

Take any continuous function $u(x, y)$ defined on the fundamental domain T, and which satisfies the following automorphic boundary conditions:

$$u\left(\frac{1}{2}, y\right) = u\left(-\frac{1}{2}, y\right), \quad u(x, y) = u(-x, y) \quad \text{for } x^2 + y^2 = 1. \tag{9.89}$$

We then define $u(x, y)$ for any (x, y) in $\mathbb{H}$ as $u(\gamma(x, y))$, where γ in Γ carries (x, y) into T.

We can similarly construct differentiable automorphic functions by imposing automorphic boundary conditions on the derivatives of u as well.

The objects of our study are automorphic solutions of the hyperbolic wave equation (9.86), (9.87). Given automorphic initial data $g = \{u(0), u_t(0)\}$ in $\mathbb{H}$, we denote by $u(x, y, t)$ the solution in $\mathbb{H} \times \mathbb{R}$ of the hyperbolic wave equation (9.87) with initial data g. Since the wave equation is invariant under hyperbolic motions, $u(\gamma(x, y), t)$ also is a solution of the wave equation. Since g is automorphic $u(\gamma(x, y), t)$ has the same initial values as $u(x, y, t)$, and therefore is $\equiv u(x, y, t)$. This shows that if the initial data of $u(x, y, t)$ are automorphic, then $u(x, y, t)$ is automorphic for all t.

DEFINITION H is the space of all automorphic data $h = \{h_1, h_2\}$ with the following properties:

(i) h_1 and its first derivatives are square integrable over T.
(ii) h_2 is square integrable over T.

The energy $E_T(h)$ of data h in H is defined as

$$E_T(h) = \int_T \left[(\partial_x h_1)^2 + (\partial_y h_1)^2 - \frac{1}{4y^2}h_1^2 + \frac{h_2^2}{y^2}\right] dx\, dy . \tag{9.90}$$

NOTE. Energy is not positive for all h in H; take for instance $h = \{1, 0\}$.

Let $u(t)$ be the solution of the hyperbolic wave equation with initial data h in H: $\{u(0), u_t(0)\} = h$. Denote by $U(t)$ the operator $h \to \{u(t), u_t(t)\}$. The operators $U(t)$ form a one-parameter group; each $U(t)$ preserves energy; this follows from the standard argument.

The operator L acting on automorphic functions square integrable in T has a very rich point spectrum with negative eigenvalue:

$$Lg = -\mu^2 g , \quad \mu \text{ real} . \tag{9.91}$$

The corresponding eigendata of $U(t)$ are

$$e_+ = \{g, i\mu g\} , \quad e_- = \{g, -i\mu g\} . \tag{9.91'}$$

The corresponding solutions of the wave equation,

$$g(x, y)e^{i\mu t} \quad \text{and} \quad g(x, y)e^{-i\mu t} ,$$

are standing waves and so do not contribute to the scattering process. A scattering theory for automorphic waves has to be built on data that are orthogonal to the eigenfunctions f. We denote by H_c the space of such data. We show now how to construct data in H_c.

LEMMA 9.33 *For any automorphic function* $h(x, y)$ *define*

$$\bar{h}(y) = \int_{-1/2}^{1/2} h(x, y) dx , \quad y > 0 . \tag{9.92}$$

For eigenfunctions g *satisfying* (9.91), $\bar{h}(y) \equiv 0$.

PROOF: Integrate with respect to x the eigenvalue equation (9.91); the result is the ordinary differential equation

$$y^2 \bar{h}_{yy} + \frac{1}{4}\bar{h} + \mu^2 \bar{h} = 0 . \tag{9.93}$$

Two solutions of this equation are $y^{1/2+i\mu}$ and $y^{1/2-i\mu}$; all others are linear combinations of them. Since $h(x, y)$ is square integrable over T, $\bar{h}(y)$ is a square integrable function of y over $[1, \infty)$ with respect to dy/y^2. But the only square integrable solution of (9.93) is $\bar{h} = 0$. □

Let $u(x, y, t)$ be any automorphic solution of the hyperbolic wave equation. We define, as before, its x-average $\bar{u}(y, t)$ as

$$\bar{u}(y, t) = \int_{-1/2}^{1/2} u(x, y, t) dx , \quad y > 0 .$$

Integrate the hyperbolic wave equation with respect to x. It follows that $\bar{u}$ satisfies the *averaged wave equation*

$$\bar{u}_{tt} - y^2\bar{u}_{yy} - \frac{1}{4}\bar{u} = 0. \tag{9.94}$$

The change of variables

$$\bar{u} = y^{1/2}v, \quad y = e^s, \tag{9.95}$$

turns this equation into

$$v_{tt} - v_{ss} = 0. \tag{9.94'}$$

We can factor this equation as

$$(\partial_t + \partial_s)(v_t - v_s) = 0,$$

from which it follows that $v_s - v_t$ is a function of $s - t$:

$$v_s - v_t = r_+(s - t). \tag{9.96$_+$}$$

Going back to (9.95) we can express the function $r_+(s)$ in terms of $\bar{u}$:

$$r_+(s) = \partial_s e^{-s/2}\bar{h}_1(e^s) - e^{-s/2}\bar{h}_2(e^s), \tag{9.97$_+$}$$

where $\{h_1, h_2\}$ are the initial values of u. We call $r_+(s)$ the *outgoing translation representer* of $h = \{h_1, h_2\}$; we shall denote it as

$$r_+ = R_+h.$$

It follows from (9.96$_+$) that

$$R_+U(t)h = r_+(s - t), \tag{9.96'}$$

that is, R_+ transmutes the action of $U(t)$ into translation.

It follows from Lemma 9.33 that the translation representer of e_+ and e_- defined in (9.91′) is zero.

Automorphic solutions of the hyperbolic wave equation that are independent of x satisfy equation (9.94). The transformation (9.95) turns this into the classical wave equation (9.94′), which has solutions $m(s - t)$ that propagate only in one direction. The corresponding solution of (9.94) is

$$u_+(y, t) = y^{1/2}m(\log y - t). \tag{9.98$_+$}$$

We choose for $m(s)$ a differentiable function with compact support contained in $0 < s$. Then for $t \geq 0$, $u_+(y, t)$ satisfies the automorphic boundary conditions (9.89) on the boundary of T, and so can be extended as an automorphic function to all of $\mathbb{H}$.

The initial data of u_+ in T are

$$f_+ = \{y^{1/2}m(\log y), -y^{1/2}m'(\log y)\}, \tag{9.99$_+$}$$

where m' denotes the derivative of m.

These solutions $u_+(y, t)$ are *outgoing* in the sense that given any compact set in T, $u_+(y, t) = 0$ in this set for t large enough. Thus for $t > \log a$, $u(y, t) = 0$ for $y < a$.

DEFINITION F_+ consists of all data f_+ of the form (9.99_+) where m in C_0^2 is supported on $0 < s$.

THEOREM 9.34

(i) *F_+ satisfies the first two properties of outgoing subspaces that are listed in* (9.15_+):

$$U(t)F_+ \subset F_+ \text{for } t > 0; \quad \cap U(t)F_+ = \{0\}.$$

(ii) *Every f_+ in F_+ is E-orthogonal to every e_+ and e_- defined in* (9.91′).

(iii) *The energy of f_+ defined by* (9.98_+) *is*

$$E(f_+) = 2\int_0^\infty (m'(s))^2\,ds. \tag{9.100}$$

(iv) *The outgoing translation representer of f_+ is $2m'(s)$.*

PROOF:

(i) Follows from (9.98_+).

(ii) Since f_+ is independent of x and zero for $y \leq 1$, the integration in

$$(f_+, e_+)_E \quad \text{and} \quad (f_+, e_-)_E$$

acts only on e_+ and e_-. By Lemma 9.33, $\bar{e}_+$ and $\bar{e}_-$ are zero.

(iii) From (9.99_+) we get for the first component of f_+

$$f_y = y^{-1/2}m'(\log y) + \frac{1}{2}y^{-1/2}m(\log y); \tag{9.101}$$

therefore, according to (9.88),

$$\begin{aligned} E(f_+) &= \int_1^\infty \left(y^{-1}\left(m' + \frac{1}{2}m\right)^2 - \frac{1}{4y}m^2 + y^{-1}m'^2\right)dy \\ &= \int_1^\infty (2y^{-1}m'^2 + ymm')dy = \int_0^\infty 2m'(s)^2 ds + \int_0^\infty mm'\,ds \\ &= 2\int_0^\infty (m'(s))^2\,ds. \end{aligned}$$

(iv) From (9.99_+), for $s > 0$

$$f_1(e^s) = e^{s/2}m(s), \quad f_2(e^s) = -e^{s/2}m'(s).$$

Setting this into formula (9.97_+) for R_+f_+, we conclude that for $s > 0$

$$R_+f_+ = r_+(s) = 2m'(s).$$

For $t > 0$ and for $s > 0$,

$$U(t)f = \{y^{1/2}m(\log y - t), -y^{1/2}m'(\log y - t).$$

Setting this into the formula (9.96′) for $R_+U(t)f_+$ gives

$$R_+U(t)f_+ = r_+(s-t) = 2m'(s-t).$$

Since this holds for all $t > 0$, $s > 0$,

$$r_+(s) = 2m'(s) \tag{9.102}$$

holds for all s. □

The *incoming translation representer* $r = R_-h$ of any h in H is defined analogously to (9.97_+) as

$$r_-(s) = \partial_s e^{-s/2}\bar{h}_1(e^s) + e^{-s/2}\bar{h}_2(e^s)\,. \tag{9.97$_-$}$$

R_- transmutes the action of $U(t)$ into translation to the left:

$$R_-U(t)h = r_+(s+t)\,.$$

F_- is defined analogously to F_+; it consists of the initial data of solutions

$$u_-(y,t) = y^{1/2}n(\log y + t)\,,$$

where n is a differentiable function of compact support contained in $0 < s$. The initial data of u_- are

$$f_- = \{y^{1/2}n(\log y),\, y^{1/2}n'(\log y)\}\,. \tag{9.99$_-$}$$

The analogue of Theorem 9.34 holds for the space F_-.

THEOREM 9.35 *F_- and F_+ are E-orthogonal to each other.*

PROOF: This is a simple calculation. Using formula (9.101), and its analogue for f_-, in the definition of the energy scalar product, we have

$$\begin{aligned}(f_+, f_-)_E &= \int_1^\infty \left[y^{-1}\left(m' + \frac{1}{2}m\right)\left(n' + \frac{1}{2}n\right) - \frac{1}{4y}mn - \frac{1}{y}m'n'\right]dy \\ &= \frac{1}{2}\int_1^\infty (m'n + mn')\frac{dy}{y} = \frac{1}{2}\int_0^1 (m'n + mn')ds = 0\end{aligned}$$

□

The operator L has a single square integrable eigenfunction g with positive eigenvalue:

$$g \equiv 1\,, \quad Lg = \frac{1}{4}g\,.$$

This gives rise to two eigenfunctions of $U(t)$:

$$\begin{aligned} p_+ &= \{1, 1/2\}\,, & p_- &= \{1, -1/2\}\,, \\ U(t)p_+ &= e^{t/2}p_+\,, & U(t)p_- &= e^{-t/2}p_-\,. \end{aligned} \tag{9.103}$$

LEMMA 9.36 *p_- is E-orthogonal to F_+, and p_+ is E-orthogonal to F_-.*

PROOF: This is again a calculation based on (9.99_+):

$$(f_+, p_-)_E = -\frac{1}{4}\left(y^{1/2}m, 1\right) + \left(y^{1/2}m', \frac{1}{2}\right)$$
$$= \int_1^\infty \left(-\frac{y^{-3/2}}{4} m(\log y) + \frac{y^{-3/2}}{2} m'(\log y)\right) dy$$
$$= \int_0^\infty \left(-\frac{e^{-s/2}}{4} m(s) + \frac{e^{-s/2}}{2} m'(s)\right) ds = 0,$$

as may be seen by integrating the second term by parts.

$(f_-, p_+)_E = 0$ can be proved similarly. □

Denote by F_+^0 those elements f of F_+ that are orthogonal to p_+. Then $U(t)f$, too, is orthogonal to p_+, and by Lemma 9.36 to p_- as well. The energy norm is positive on the orthogonal complement of p_+ and p_-.

DEFINITION H_+ is the closure in the energy norm of $U(t)F_+^0$.

THEOREM 9.37 *H_+ consists of all f in H that are E-orthogonal to the eigendata e_+ and e_- defined in* (9.91′), *and p_+, p_- defined in* (9.103).

SKETCH OF PROOF: According to part (ii) of Theorem 9.34, every f_+ in F_+ is E-orthogonal to the eigenfunctions e_+ and e_-. By Lemma 9.36, f_+ is E-orthogonal to p_-, and by definition of F_+^0, f_+ in F_+^0 is orthogonal to p_+ as well. Since $U(t)$ preserves the E-scalar product,

$$0 = (f_+, e_\pm)_E = (U(t)f_+, U(t)e_\pm)_E = e^{\mp i\mu t}(U(t)f_+, e_\pm)_E$$

it follows that $U(t)f_+$ is E-orthogonal to e_+, e_-, as well as to p_- and p_+. But then so is every f in H_+, the closure of $U(t)F_+^0$. The content of Theorem 9.37 is that, conversely, the E-orthogonal complement of F_+^0 is spanned by these eigenfunctions. This can be shown by demonstrating that $(A - kI)^{-1}$ acting on the E-orthogonal complement of F_+^0 is a compact operator. For complete proof the reader is referred to Lax and Phillips (1985); below we merely give the key lemma. □

LEMMA 9.38 *Denote by H^0 the space of all g in H that are E-orthogonal to p_+ and p_-. Such a g is E-orthogonal to F_+^0 iff $R_+g \equiv 0$.*

PROOF: Rewrite formula (9.97_+) for the outgoing translation representer in terms of the variable y instead of s:

$$R_+g = y^{1/2}D_y\bar{g}_1 - \frac{1}{2}y^{-1/2}\bar{g}_1 - y^{-1/2}\bar{g}_2.$$

Denote by f_1 and f_2 the components of f_+. Using formula (9.99_+) we get

$$(f_+, g)_E = \int \left[D_y f_1 D_y \bar{g}_1 - \frac{1}{4y^2} f_1\bar{g}_1 + \frac{1}{y^2} f_2\bar{g}_2\right] dy$$
$$= \int \left[y^{-1/2}\left(m' + \frac{1}{2}m\right)D_y\bar{g}_1 - \frac{y^{-3/2}}{4} m\bar{g}_1 - y^{-3/2}m'\bar{g}_2\right] dy. \tag{9.104}$$

We integrate the second term by parts:

$$\int \frac{1}{2} y^{-1/2} m D_y \bar{g}_1 dy = \frac{1}{2} \int y^{-3/2} \left(\frac{1}{2} m - m' \right) \bar{g}_1 dy ;$$

set this into (9.104):

$$(f_+, g)_E = \int m' \left[y^{-1/2} D_y \bar{g}_1 - \frac{1}{2} y^{-3/2} \bar{g}_1 - y^{-3/2} \bar{g}_2 \right] dy = \int m' (\log y) R_+ g \frac{dy}{y} .$$

Switch to $s = \log y$ as variable:

$$(f_+, g)_E = \int m'(s) R_+ g(s) ds . \tag{9.104'}$$

It follows from (9.104′) that if $R_+ g(s) \equiv 0$, then $(f_+, g)_E = 0$ for all f_+ in F_+^0. Conversely, if (9.104′) is zero for all $m'(s)$ satisfying orthogonality of f_+ to p_+:

$$\int e^{-s/2} m'(s) ds = 0 ,$$

it follows that $R_+ g(s) \equiv 0$. □

We define similarly F_-^0, and H_- as the closure of $U(t) F_-$. The analogue of Theorem 9.37 holds:

H_- consists of all f in H that are E-orthogonal to e_+, e_-, d_+, d_-.

These two results show that H_+ and H_- have the same E-orthogonal complement. But then H_+ and H_- are the same space; we denote $H_+ = H_-$ as H_c, for it is the subspace on which $U(t)$ has a continuous spectrum.

We are now back in the framework of LP scattering theory presented in Section 9.2. The one-parameter group of unitary operator $U(t)$ acts on H_c. There is an incoming and an outgoing subspace F_-^0 and F_+^0, E-orthogonal, having the usual properties. Each gives rise to a translation representation given by R_+, and R_- with a sign reversal.

Note that these representations have been constructed explicitly, without appealing to Sinai's theorem.

We can describe the scattering process as follows:

Let u be an incoming wave, defined for $t \le 0$ by

$$u(x, y, t) = y^{-1/2} n(\log y + t) , \tag{9.105}$$

n(s) supported on $s \ge 0$. For $t > 0$, the wave impinges on the edge $x^2 + y^2 = 1$ of the fundamental triangle T, and is scattered by the geometry of T. For t tending to infinity $u(t)$ becomes outgoing, as may be seen by the following argument: Denote by h the initial values of u defined by (9.105). Given any $\varepsilon > 0$, there is an $f_+(\varepsilon)$ in F_+^0 and a $t(\varepsilon)$ such that

$$\|h - U(-t(\varepsilon)) f_+(\varepsilon)\|_E < \varepsilon .$$

Since $U(t)$ preserves energy,

$$\|U(t(\varepsilon))h - f_+(\varepsilon)\|_E < \varepsilon .$$

This shows that $u(x, y, t(\varepsilon))$ is very nearly outgoing.

The scattering operator relates the incoming and outgoing translation representations, and the scattering matrix relates their Fourier transformations. Since the continuous spectrum has multiplicity one—the translation representers are scalar valued functions—the scattering matrix is a scalar valued function. Pavlov and Faddeev have determined $M(\sigma)$ as

$$M(\sigma) = \frac{\Gamma(\frac{1}{2})\Gamma(i\sigma)\zeta(2i\sigma)}{\Gamma(\frac{1}{2}+i\sigma)\zeta(1+2i\sigma)}, \tag{9.106}$$

where ζ is the Riemann ζ function. Denote by z the zeros of ζ; the zeros of $M(\sigma)$ are located at $\sigma = z/2i$; so the Riemann hypothesis is true iff all zeros of $M(\sigma)$ have imaginary part $-i/4$.

Let $Z(t)$ be the LP semigroup associated with the scattering process; see Section 9.3. According to Theorem 9.9, the spectrum of the infinitesimal generator G of $Z(t)$ consists of those complex numbers γ for which $i\bar{\gamma}$ is a zero of $M(\sigma)$.

According to Phillips' spectral mapping theorem for semigroups, if γ belongs to the spectrum of the generator G of $Z(t)$ $e^{\gamma t}$ belongs to the spectrum of $Z(t)$. Using this fact Pavlov and Faddeev examined the following intriguing possibility.

Suppose we can show that $Z(t)$ decays exponentially:

$$\lim_{t\to\infty} \frac{1}{t} \log \|Z(t)\| = \alpha < 0. \tag{9.107}$$

Since the spectrum of $Z(t)$ lies in the disk of radius $\|Z(t)\|$,

$$|e^{t\gamma}| \le e^{(\alpha+\varepsilon)t} \tag{9.108}$$

for all γ in the spectrum of G; since $\varepsilon \to 0$ as $t \to \infty$, it follows from (9.107) that

$$\operatorname{Re}\gamma \le \alpha. \tag{9.108'}$$

The zeros of $M(\sigma)$ are $i\bar{\gamma}$, and therefore the imaginary part of the zeros of $M(\sigma)$ are $\le \alpha$. It follows that the zeros $z = 2i\sigma$ of the ζ function have real part $\ge -2\alpha$. If one could prove (9.107) for $\alpha = -1/4$, the Riemann hypothesis would follow.

Scattering theory is a big subject. In the bibliography below we list mainly those items that were used in the text.

References

Kato, T. *Perturbation theory for linear operators*. Springer, New York, 1966.

Lax, P. D., Morawetz, C. S., and Phillips, R. S. Exponential decay of solutions in the exterior of a star-shaped obstacle. *Comm. Pure Appl. Math.* 16: 477–486, 1963.

Lax, P. D., and Phillips, R. S. *Scattering theory*. Academic Press, New York–London, 1967.

———. Decaying modes for the wave equation in the exterior of an obstacle. *Comm. Pure Appl. Math.* 22: 737–787, 1969.

———. *Scattering theory for automorphic functions*. Princeton University Press, Princeton, N.J., 1976.

———. Translation representation for automorphic solutions of the wave equation in non-Euclidean spaces; the case of finite volume. *Trans. Amer. Math. Soc.* 289: 715–735, 1985.

———. Translation representation for solutions of the non-euclidean wave equation. *Comm. Partial Differential Equations* 2: 395–438, 1977.

Melrose, R. B. *Geometric scattering theory.* Cambridge University Press, Cambridge, 1995.

Morawetz, C. S. The decay of solutions of the exterior initial-boundary value problem for the wave equation. *Comm. Pure Appl. Math.* 14: 561–568, 1961.

Morawetz, C. S., Ralston, J., and Strauss, W. Decay of solutions of the wave equation outside nontrapping obstacles. *Comm. Pure Appl. Math.* 30: 447–508, 1977.

Pavlov, B. S.; Faddeev, L. D. Scattering theory and automorphic functions. (Russian) Boundary value problems of mathematical physics and related questions in the theory of functions, 6. *Zap. Naučn. Sem. Leningrad. Otdel. Mat. Inst. Steklov (LOMI)* 27: 161–193, 1972.

Ralston, J. Propagation of singularities and the scattering matrix. *Singularities in boundary value problems (Proc. NATO Adv. Study Inst., Maratea, 1980)*, 169–184. NATO Adv. Study Inst. Ser., Ser. C: Math. Phys. Sci., 65. Reidel, Dordrecht-Boston, 1981.

Reed, M., and Simon, B. *Methods of modern mathematical physics.* 4 vols. Academic Press, New York–London, 1972–1978.

Taylor, M. *Partial differential equations.* II. *Qualitative studies of linear equations.* Applied Mathematical Sciences, 116. Springer, New York, 1996.

CHAPTER 10

Hyperbolic Systems of Conservation Laws

No monograph on hyperbolic equations is complete without some discussion of a nonlinear instance. Systems of conservation laws in a single space variable is a well-rounded subject, suitable for presentation in an introductory monograph.

The Euler equations of compressible flow, discussed in Section 10.6, are an important example of a hyperbolic system of conservation laws.

In phenomena governed by nonlinear hyperbolic equations signals propagate with finite speed, just as in the linear case. Singularities propagate along characteristics, but unlike the linear case, can arise spontaneously, leading to the formation of shocks. Time is not reversible as for linear equations but future and past are different, as they are in real life. There is a substantial loss of information as time moves forward, which can be interpreted as an increase of entropy.

The basic existence theory of solutions of hyperbolic conservation laws in a single space variable is due to Jim Glimm (1965). It is a scandal of mathematical physics that, apart from isolated results, no comparable theory exists for more space variables.

10.1. Scalar Equations; Basics

Many phenomena in the solutions of nonlinear hyperbolic equations are already manifested by solutions of the simplest nonlinear equation

$$u_t + uu_x = 0\,. \tag{10.1}$$

The left side can be interpreted as a directional derivative, leading to the following form of (10.1):

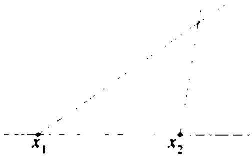
$$\frac{du}{dt} = 0\,, \quad \frac{dx}{dt} = u\,. \tag{10.2}$$

The first equation says that u is constant on the curve, call it characteristic, along which we are differentiating; the second equation says that the speed u with which this curve propagates is constant. Therefore the characteristic curves are straight lines.

Given initial values $u(x,0) = u_0(x)$, we draw a straight line from each point $(x,0)$ traveling with speed $u_0(x)$. If $u_0(x)$ is differentiable, these lines cover a slab $0 \le t \le T$ in a one-to-one fashion and provide these a solution of the initial value problem.

EXERCISE Prove the last statement.

When $u_0(x)$ is an increasing, smooth function of x, the straight lines cover the whole half-plane $x, t, t \geq 0$ in a one-to-one fashion. But consider the case when $u_0(x_1) > u_0(x_2)$ for some pair x_1, x_2, where $x_1 < x_2$. In this case the characteristics issuing from x_1 and x_2 intersect at time $t_c = (x_2 - x_1)/(u_0(x_1) - u_0(x_2))$:



at the point of intersection $u(x, t_c)$ would have to be equal to both $u_0(x_1)$ and $u_0(x_2)$, an impossibility. So there is no continuous solution beyond the time t_c.

A similar breakdown of solutions of the Euler equations of incompressible flow, observed in the middle of the nineteenth century caused a crisis in fluid dynamics. The leading theorists, Stokes and Airy among them, have grappled with the problem. Its resolution came from the greatest mathematician, Riemann, who pointed out that the equations of fluid dynamics are integral conservation laws of the following kind.

Denote by $u = u(x, t)$ the density of some quantity that obeys a conservation law, and by $f = f(x, t)$ the flux that transports that quantity. Then the total amount of that quantity contained in any smoothly bounded domain C changes at the rate at which that quantity is transported across the boundary of C:

$$\frac{d}{dt} \int_C u(x, t)dx = - \int_{\partial C} f \cdot n \, dS . \tag{10.3}$$

The minus sign occurs because n denotes, as usual, the outward normal to ∂C.

If u and f are differentiable, the differentiation on the left can be carried out under the integral sign, and the boundary integral on the right can be transformed by the divergence theorem, leading to

$$u_t + \operatorname{div} f = 0 . \tag{10.4}$$

Riemann then pointed out that whereas the differential equation (10.4) makes sense only for differentiable functions, the integral form (10.3) is meaningful for a much larger class, such as discontinuous ones.

Today of course we can, using the theory of distribution, make sense of the differential equation. Riemann, in 1860, had anticipated that theory by almost a hundred years.

Riemann proposed that solutions can by continued beyond the time when singularities form, as solutions in the integral sense (10.3). Furthermore he derived the law of propagation of discontinuities of solutions in the integral sense. We shall carry out the calculation in one space dimension; the differential form of the conservation law is

$$u_t + f_x = 0 , \tag{10.5}$$

where f is a function of u and of other densities. Let $x = y(t)$ be a curve along which u is discontinuous; since f is a function of u, it too is discontinuous. We shall denote by u^l, u^r and f^l, f^r the values of u and f at the left and right sides of the discontinuity.

We take in (10.3) the domain C to be an interval $[a, b]$ containing $y(t)$. Then (10.3) says that

$$\frac{d}{dt}\int_a^b u(x,t)dx = f^a - f^b . \tag{10.6}$$

We break up the integral on the left as

$$\int_a^b u\,dx = \int_a^{y(t)} u\,dx + \int_{y(t)}^b u\,dx .$$

Its t-derivative is

$$\int_a^y u_t\,dx + y_t u^l - y_t u^r + \int_y^b u\,dx .$$

Expressing u_t to the left and right of y from (10.5) and carrying out the integration we get

$$f^a - f^l + y_t(n^l - u^r) + f^r - f^l .$$

Setting this into (10.6) and solving for $y_t = s$, the speed of propagation of the discontinuity, we get

$$s = \frac{[f]}{[u]} , \tag{10.7}$$

where [ ] denotes the jump in the quantity in brackets upon crossing the discontinuity.

Equation (10.7) is, or should be, called the Riemann-Rankin-Hugoniot condition (RRH).

We return now to equation (10.1) and write it as a conservation law

$$u_t + \left(\frac{1}{2}u^2\right)_x = 0 . \tag{10.8}$$

For this conservation law the RRH condition, with $f(u) = \frac{1}{2}u^2$,

$$s = \frac{u^l + u^r}{2} . \tag{10.8'}$$

Consider now the following initial value problem:

$$u_0(x) = \begin{cases} 1 & \text{for } x \le 0 \\ 1 - x & \text{for } 0 < x < 1 \\ 0 & \text{for } 1 \le x . \end{cases} \tag{10.9}$$

For $t < 1$ this initial value problem has the solution

$$u(x,t) = \begin{cases} 1 & \text{for } x \le t \\ \frac{x-1}{t-1} & \text{for } t \le x < 1 \\ 0 & \text{for } 1 \le x . \end{cases} \tag{10.10}$$

The characteristic lines issuing from the interval $0 \le x \le 1$ all intersect at the point $(1, 1)$. So no solution exists in the classical sense for $t \ge 1$. But there is a discontinuous solution that satisfies the jump condition (10.8′):

$$u(x,t) = \begin{cases} 1 & \text{for } x \le 1 + \frac{t}{2} \\ 0 & \text{for } 1 + \frac{t}{2} < x \,. \end{cases} \tag{10.10′}$$

The function $u(x, t)$ defined by (10.10) for $t \le 1$ and by (10.10′) for $t \ge 1$ is a solution of the initial value problem (10.9) for the conservation law (10.8).

Consider now the initial value problem

$$u_0(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } 0 < x \,. \end{cases} \tag{10.11}$$

The discontinuous solution

$$u(x,t) = \begin{cases} 0 & \text{for } x < \frac{t}{2} \\ 1 & \text{for } \frac{t}{2} < x \end{cases} \tag{10.12}$$

satisfies the jump condition (10.8′), and is therefore a solution in the integral sense of the conservation law (10.8).

On the other hand, the function

$$v(x,t) = \begin{cases} 0 & \text{for } x \le 0 \\ \frac{x}{t} & \text{for } 0 \le x \le t \\ 1 & \text{for } t < x \end{cases} \tag{10.12′}$$

is a continuous solution of the differential form (10.1) of the conservation law (10.8). These examples show that within the class of solutions in the integral sense an initial value problem may have several solutions. Therefore this class has to be narrowed. Recall that we were forced to introduce discontinuous solutions because of the collision of characteristics. Therefore we only accept discontinuities that separate two characteristics that otherwise would impinge on each other; that is, if

$$u^l > s > u^r \,. \tag{10.13}$$

We turn now to general scalar conservation laws

$$u_t + f(u)_x = 0 \,. \tag{10.14}$$

The differential form of this equation is

$$u_t + a(u)u_x = 0 \,, \quad a(u) = \frac{df}{du} \,. \tag{10.14′}$$

This equation is genuinely nonlinear if the derivative of $a(u)$ with respect to u is nonzero. This requires f to be strictly convex or concave; for definiteness we take $f(u)$ to be *convex*. The condition of admissibility, the same for convex and concave f, is

$$a(u^l) > s > a(u^r) \,. \tag{10.13′}$$

In addition to its geometric meaning this condition has a physical meaning; this will be explained in Section 10.4.

For a flux function that is neither convex nor concave, Oleĭnik has given the correct admissibility condition.

A discontinuity that satisfies the RRH condition (10.7) and the *entropy condition* (10.13′) is called a *shock*.

A solution u of (10.14) all of whose discontinuities are shocks is called an *admissible solution*.

10.2. The Initial Value Problem for Admissible Solutions

We start with the following result of Kruzkov; see also Keyfitz.

THEOREM 10.1 *Let u and v be a pair of admissible solutions of the same conservation law* (10.14), *$f(u)$ convex. We claim that*

$$|u(t) - v(t)|_{L^1} = \int |u(x,t) - v(x,t)|dt \tag{10.15}$$

is a decreasing function of t.

PROOF: Denote by $y_1(t), \dots,$ the points where $u(x,t) - v(x,t)$ as a function of x changes sign. We rewrite (10.15) as

$$\sum \delta_n \int_{y_n}^{y_{n+1}} (u(x,t) - v(x,t))dx\,, \tag{10.15′}$$

where $\delta_n = 1$ if the integrand in (10.15′) is positive, and $\delta_n = -1$ if it is negative. Differentiate (10.15′) with respect to t; we get

$$\begin{aligned} &\frac{d}{dt}|u(t) - v(t)|_{L^1} \\ &= \sum \delta_n \Bigg[\int_{y_n}^{y^{n+1}} (u_t - v_t)dx \\ &\qquad + (u - v)(y_{n+1})\frac{dy_{n+1}}{dt} - (u - v)(y_n)\frac{dy_n}{dt} \Bigg]. \end{aligned} \tag{10.16}$$

Expressing u_t, v_t as $-f(u)_x$, $-f(v)_x$ we get for (10.16)

$$\begin{aligned} &\sum \delta_n \Bigg[f(v(y_{n+1})) - f(u(y_{n+1})) - f(v(y_n)) + f(u(y_n)) \\ &\qquad + (u - v)(y_{n+1})\frac{dy_{n+1}}{dt} - (u - v)(y_n)\frac{dy_n}{dt} \Bigg] \\ &= \sum \delta_n \Bigg[f(v) - f(u) + (u - v)\frac{dy}{dt} \Bigg]_{y_n}^{y_{n+1}} . \end{aligned} \tag{10.16′}$$

If the change in sign of $u - v$ at y_{n+1} occurs at a point of continuity of both u and v, the contribution to (10.16′) at y_{n+1} is zero. Suppose u has a shock at y_{n+1} and v is continuous there; for convex f, by (10.13′)

$$u^l > v > u^r\,. \tag{10.17}$$

Then the sign of $u - v$ in $[y_n, y_{n+1}]$ is positive, so $\delta_n = 1$, and the contribution to (10.16′) at y_{n+1} is

$$f(v) - f(u^l) + (u^l - v)\frac{f(u^l) - f(u^r)}{u^l - u^r},$$

where we have used the RRH relation (10.7). We rewrite the formula above as

$$f(v) - \left[\frac{v - u^r}{u^l - u^r} f(u^l) + \frac{u^l - v}{u^l - u^r} f(u^r)\right]. \tag{10.18}$$

Since, by (10.17), v lies between u_r and u_l, and since $f(u)$ is convex, the value of $f(v)$ is less than the value at v of the linear function that interpolates f between u_l and u_r. This shows that (10.18) is a negative quantity.

The continuation to (10.16′) at y_n is likewise negative. The same is true when both u and v are discontinuous at y_n. □

An important consequence of Theorem 10.1 is

COROLLARY 10.1′ *Admissible solutions are uniquely determined by their initial data.*

We turn now to the problem of existence of admissible solutions with given initial values. We start with the observation that a convex function $f(u)$ lies above its tangent lines:

$$f(u) \geq f(v) + a(v)(u - v), \tag{10.19}$$

where $a(v) = f'(v)$.

Let $u(x, t)$ be a smooth solution of (10.14), and suppose that its initial value $u_0(x)$ is zero for x large negative. Then the same is true for $u(x, t)$ for any t for which $u(x, t)$ is defined. We define $U(x, t)$ as

$$U(x, t) = \int_{-\infty}^{x} u(z, t)dz. \tag{10.20}$$

Conversely

$$u = U_x. \tag{10.20′}$$

Integrating the equation

$$u_t + f(u)_x = 0$$

from $-\infty$ to x gives

$$U_t + f(U_x) = 0; \tag{10.21}$$

here we assumed that $f(0) = 0$. We apply inequality (10.19) with $u = U_x$:

$$-U_t = f(u) \geq f(v) + a(v)(u - v),$$

which we can rewrite as

$$U_t + a(v)U_x \leq g(v), \tag{10.22}$$

where

$$g(v) = a(v)v - f(v). \tag{10.22′}$$

Denote by y the point where the line $dx/dt = a(v)$ intersects the initial line $t = 0$:

$$\frac{x-y}{t} = a(v)\,. \tag{10.23}$$

Denote the inverse of the function $a(u)$ by b; we obtain from (10.23) that

$$b\left(\frac{x-y}{t}\right) = v\,. \tag{10.23'}$$

The left side of (10.22) is a directional derivative of U; integrate (10.22) from 0 to t:

$$U(x,t) \le U(y,0) + tg(v)\,. \tag{10.24}$$

Expressing v from (10.23′) gives

$$g(v) = g\left(b\left(\frac{x-y}{t}\right)\right). \tag{10.25}$$

Denote the function $g(b(s))$ as $h(s)$. From (10.22′) we get, using $f' = a$, that

$$g_v = a'(v)v\,;$$

therefore by the chain rule

$$\frac{d}{ds}h(s) = g'(b)b' = a'(v)b'(s)v\,.$$

Since a and b are inverse of each, and $b(s) = v$, it follows that $a'(v)b'(s) = 1$, which leaves

$$h'(s) = b(s)\,. \tag{10.26}$$

This determines h up to a constant. Since $f(0) = 0$, it follows from (10.22′) that $g(0) = 0$. Denote $a(0)$ by c; then $b(c) = 0$, and so

$$h(c) = g(b(c)) = g(0) = 0\,. \tag{10.26'}$$

Note that b is an increasing function, and so h is convex.

Coming back to (10.24), we use (10.25) to rewrite it as

$$U(x,t) \le U(y,0) + th\left(\frac{x-y}{t}\right). \tag{10.27}$$

Since (10.22) holds for arbitrary v, (10.27) holds for arbitrary y. Since for $v = u(x,t)$ equality holds in (10.22) along the whole characteristic issuing from x, t, for each x, t there is a value y for which equality holds in (10.27). We summarize:

THEOREM 10.2 *Let $u(x,t)$ be a smooth solution of* (10.14):

$$u_t + f(u)_x = 0\,.$$

Then

$$u(x,t) = b\left(\frac{x-y}{t}\right), \tag{10.28}$$

where $y = y(x, t)$ minimize,

$$U_0(y) + th\left(\frac{x-y}{t}\right) = G(x, y, t). \tag{10.29}$$

Here b is the inverse function of $a(u) = df/du$, $f(u)$ is convex,

$$\frac{dh}{du} = b(u), \quad h(c) = 0 \quad \text{for } c = a(0), \tag{10.30}$$

and

$$U_0(y) = \int_{-\infty}^{y} u_0(z)dz, \quad u_0(x) = u(x, 0). \tag{10.31}$$

Since the derivation of Theorem 10.2 used the integrated form (10.21) of the differential equation (10.14), it is not surprising that the formula (10.28) also holds for distribution solutions.

THEOREM 10.3 *Let $u_0(x)$ be any bounded, L^1 function on the x-axis.*

(i) *The function $u(x, t)$, defined by formulas* (10.28), (10.29), *is an admissible solution of equation* (10.14) *with initial value $u_0(x)$.*
(ii) *All discontinuities of $u(x, t)$ are shocks.*

PROOF OF THEOREM 10.3(i): Since $u_0(x)$ is integrable $U_0(y)$ as defined by (10.31) is a bounded, continuous function; $h(u)$ is a convex function that achieves its minimum at $u = c$. Therefore the function $G(x, y, t)$, defined by (10.29), achieves its minimum in y at some point or points.

We can combine (10.28) and (10.29) into a single formula:

$$u(x, t) = L^1 \lim_{N\to\infty} u_N(x, t), \tag{10.32}$$

where

$$u_N(x, t) = \frac{\int b(\frac{x-y}{t})e^{-NG(x,y)}\,dy}{\int e^{-NG(x,t)}\,dy}. \tag{10.32'}$$

Similarly

$$f(u) = L^1 \lim_{N\to\infty} f_N, \tag{10.33}$$

where

$$f_N(x, t) = \frac{\int f(b(\frac{x-y}{t}))e^{-NG}\,dy}{\int e^{-NG}\,dy}. \tag{10.33'}$$

We can express u_N and f_N in terms of the function

$$V_N(x, t) = \log \int e^{-NG}\,dy \tag{10.34}$$

as

$$u_N = -\frac{1}{N}\frac{\partial}{\partial x}V_N, \quad f_N = \frac{1}{N}\frac{\partial}{\partial t}V_N, \tag{10.34'}$$

where we use the relations

$$h' = b, \quad sb(s) - h(s) = f(b(s)).$$

EXERCISE Verify these relations.

If follows from (10.34′) that

$$\partial_t u_N + \partial_x f_N = 0 .$$

Let $N \to \infty$; using (10.32) and (10.33) we conclude that the equation

$$u_t + f(u)_x = 0$$

is satisfied in the sense of distributions.

It follows from (10.32′) that $u_N(x, 0)$ tends to u_0. This completes the proof of the first part of Theorem 10.3. □

To prove the second part, that all discontinuities of u are shocks, we need

LEMMA 10.4 *For t fixed, denote by $y(x)$ any value of y where $G(x, y)$ achieves its minimum; $y(x)$ is a nondecreasing function of x.*

PROOF: We shall show that for $x_1 < x_2$, $G(x_2, y)$ does not take on its minimum for $y < y_1$, where $G(x_1, y)$ takes on its minimum at y_1; that is,

$$U_0(y_1) + th\left(\frac{x_1 - y_1}{t}\right) < U_0(y) + th\left(\frac{x_1 - y}{t}\right). \tag{10.35}$$

Next we apply Jensen's inequality to the convex function $h(s)$; since $x_1 < x_2$, $y < y_1$,

$$h\left(\frac{x_2 - y_1}{t}\right) + h\left(\frac{x_1 - y}{t}\right) < h\left(\frac{x_1 - y_1}{t}\right) + h\left(\frac{x_2 - y}{t}\right).$$

Multiply this by t and add to (10.35); after cancellation we get

$$G(x_2, y_1) < G(x_2, y) .$$

This proves that $G(x_2, y)$ does not take on its minimum for $y < y_1$. □

PROOF OF THEOREM 10.3(ii): Using formula (10.28) and the fact that $b(s)$ is an increasing function, and that $x_1 < x_2$, $y_1 \le y_2$ we get

$$u(x_1, t) = b\left(\frac{x_1 - y_1}{t}\right) \ge b\left(\frac{x_1 - y_2}{t}\right),$$

$$u(x_2, t) = b\left(\frac{x_2 - y_2}{t}\right).$$

Subtract the first from the second:

$$u(x_2, t) - u(x_1, t) \le b\left(\frac{x_2 - y_2}{t}\right) - b\left(\frac{x_1 - y_2}{t}\right).$$

Denote by k an upper bound for b'; then the right side is $\le k(x_2 - x_1)/t$, therefore since $x_2 - x_1$ is positive,

$$\frac{u(x_2, t) - u(x_1, t)}{x_2 - x_1} \le \frac{k}{t}. \tag{10.36}$$

So $u(x, t)$ satisfies a one-sided Lipschitz condition; it follows that at a discontinuity, $u_l > u_r$. □

From the explicit representation of the solution we can extract information about the dependence of solutions on their initial data.

THEOREM 10.5 *Denote by $S(t)$ the operator that relates admissible solutions of* (10.14) *to their initial data. The operators $S(t)$ form a semi-group, that is, for all $s, t > 0$*

$$S(s+t) = S(s)S(t)\,.$$

PROOF: For u_0 in L^1 and bounded, define

$$U(x,t) = \int_{-\infty}^{x} u(z,t)dz\,.$$

We claim that

$$U(x,t) = G(x, y(x,t), t)\,,$$

where G is defined by (10.29) and $y(x,t)$ minimizes $G(x,y,t)$. For smooth solutions this follows from the derivation of Theorem 10.2; for solutions in Theorem 10.3 we argue as follows: define

$$U_N(x,t) = \int_{-\infty}^{x} u_N(z,t)dz\,.$$

Since u_N tends to u in the L^1 norm, it follows that U_N tends to U uniformly. It follows from (10.34′) that $U_N = -V_N/N$. But it follows from (10.34) that V_N/N tends to $-G(x, y(x,t), t)$. Denote $S(t)u_0$ as $u(t)$; by Theorem 10.3,

$$u(x,t) = b\left(\frac{x-y}{t}\right),$$

where $y(x,t)$ minimizes

$$U_0(y) + th\left(\frac{x-y}{t}\right).$$

Similarly, denote $S(s+t)u_0$ as $u(s+t)$; then

$$u(x, s+t) = b\left(\frac{x-y}{s+t}\right), \tag{10.37}$$

where y minimizes

$$U_0(y) + (s+t)h\left(\frac{x-y}{s+t}\right). \tag{10.37′}$$

Denote $S(s)u(t)$ as $v(s)$; then

$$v(x,s) = b\left(\frac{x-z}{s}\right), \tag{10.38}$$

where z minimizes

$$U(z,t) + sh\left(\frac{x-z}{s}\right). \tag{10.38′}$$

We have shown above that

$$U(z,t) = \min_y U_0(y) + th\left(\frac{z-y}{t}\right).$$

Setting this into (10.38′) characterizes $z(x, s)$ as the minimizer of

$$\min_z \min_y \left[U_0(y) + th\left(\frac{z-y}{t}\right) + sh\left(\frac{x-z}{s}\right)\right]. \tag{10.39}$$

We perform the minimization first with respect to z; since $h' = b$, we get

$$b\left(\frac{z-y}{t}\right) - b\left(\frac{x-z}{s}\right) = 0.$$

Since b is a strictly increasing function,

$$\frac{z-y}{t} = \frac{x-z}{s}; \tag{10.39′}$$

from this

$$z = \frac{tx + sy}{s+t}.$$

Setting this into (10.38) gives

$$v(x, s) = b\left(\frac{x-s}{s+t}\right). \tag{10.40}$$

Setting (10.39′) into (10.39) gives

$$\min_y \left[U_0(y) + (s+t)h\left(\frac{x-y}{s+t}\right)\right],$$

the same minimization problem as (10.37′). Comparing (10.37) and (10.40) we conclude that

$$u(x, s+t) = v(x, s).$$

This proves the semigroup property. □

The next result shows that the admissible solution of the nonlinear equation (10.14) shares some of the properties of solutions of linear equations.

THEOREM 10.6 *Suppose that the initial value $u_0(x)$ of a solution u is in L^1, and is bounded: $|u_0(x)| \le m$ for all x; then*

(i) $|u(x, t)| \le m$ *for all x and all t,*

(ii) *signals propagate with speed* $\le \max\{|a(-m)|, a(m)\}$.

We leave it to the reader to deduce these properties from the explicit formula for the solution.

More interesting are results about solutions of nonlinear conservation laws that have no counterpart for linear equations.

THEOREM 10.7

(i) *Suppose that the initial value u_0 of u is in L^1; then for $t > 0$, $u(x, t)$ is bounded for all x, and*

$$|u(x, t)| \le \frac{\text{const}}{\sqrt{t}} \tag{10.41}$$

as $t \to \infty$.

(ii) *Take the set of all initial values of u_0 that are supported in a given finite interval, and whose L^1 norm is ≤ 1. For $t > 0$, the corresponding admissible solutions $u(x,t)$ form a precompact set in L^1.*

PROOF:

(i) Since u_0 is in L^1, $-C \leq U_0(y) \leq C$ for all y, $C = |u_0|_{L^1}$. We recall from (10.29) that

$$G(x,y,t) = U_0(y) + th\left(\frac{x-y}{t}\right);$$

the function $h(s)$ reaches its minimum at $s = c$, $h(c) = 0$. Since h is convex,

$$h(s) \geq q(s-c)^2. \tag{10.42}$$

For any value of x and t, the choice of $y_0 = x - ct$ yields

$$G(x, y_0, t) = U_0(y_0) \leq C.$$

Therefore the minimizer y of $G(x,y,t)$ must make $G \leq C$. Using (10.42) we have for any y

$$G(x,y,t) \geq -C + tq\left(\frac{x-y}{t} - c\right)^2.$$

In order for the right side to be $\leq C$, we must have

$$tq\left(\frac{x-y}{t} - c\right)^2 \leq 2C,$$

so for the minimizing y

$$\left|\frac{x-y}{t} - c\right| \leq \sqrt{\frac{2C}{qt}}. \tag{10.43}$$

Since $b(c) = 0$,

$$|b(s)| \leq k|s-c|, \tag{10.44}$$

where k is an upper bound for the derivative of b. Setting (10.43) into (10.44) we get

$$\left|b\left(\frac{x-y}{t}\right)\right| \leq k\sqrt{\frac{2C}{qt}}$$

since by (10.28)

$$u(x,t) = b\left(\frac{x-y}{t}\right);$$

(10.41) follows.

(ii) We claim that for $t > 0$, all $u(x,t)$ are supported in an interval somewhat larger than the support of u_0, and that all $|u(x,t)|$ are bounded. This can be deduced by the type of argument presented in the proof of (i). It follows from inequality (10.36) that

$$u(x,t) - \frac{kx}{t}$$

is a decreasing function of x. It is not hard to show, and is left to the reader, that a set of decreasing functions uniformly bounded on a finite interval is precompact in the L^1 norm. □

Luc Tarter has established the much deeper result of compactness for solutions of hyperbolic systems of a pair of conservation laws.

Theorem 10.7 shows that most of the information contained in the initial data is lost with the passage of time. A more precise measure of this loss has been given by DeLellis and Golse.

From the fact that

$$u_t + f_x = 0$$

is satisfied in the sense of distributions it follows that $\int u(x,t)dx$ is a *conserved quantity*, that is, its value is independent of t. Somewhat surprisingly, there is a second conserved quantity:

THEOREM 10.8 *Let u denote a solution whose initial value u_0 is in L^1. Denote as before by $U(x,t)$ the integral of $u(x,t)$:*

$$U(x,t) - \int_{-\infty}^{x} u(z,t)dz\,.$$

We claim that

$$\inf_x U(x,t)$$

is a conserved quantity.

PROOF: We have shown in the proof of Theorem 10.5 that

$$U(x,t) = \min_y U_0(y) + th\left(\frac{x-y}{t}\right).$$

Since h is a nonnegative function, it follows that

$$U(x,t) \geq \inf_y U_0(y)\,.$$

Suppose that $U_0(y)$ achieves its minimum at y_m; then $U(x,t)$ achieves it minimum at $x_m = y_m + ct$, and it is equal to the minimum of $U_0(y)$. The equality of the infima can be proved analogously. □

The next result shows that there are only these two conserved quantities that depend continuously on the initial data in the L^1 topology.

THEOREM 10.9

(i) *Let p and q denote two nonnegative numbers, and d a positive number. Define*

$$N(x,t;p,q) = \begin{cases} (x/t - c)/d & \text{for } -\sqrt{pt} < x - ct < \sqrt{qt} \\ 0 & \text{otherwise}. \end{cases} \tag{10.45}$$

Define $f(u) = cu + du^2/2$. Then N is an admissible solution of

$$N_t + f(N)_x = 0$$

with two shocks.

(ii) *Conversely, let $f(u)$ be a convex function, $f(0) = 0$, $f'(0) = c$, $f''(0) = d$. Let u be an admissible solution of*

$$u_t + f(u)_x = 0$$

whose initial value u_0 belongs to L^1. Define

$$p = -2d \inf \int_{-\infty}^{y} u_0(x)dx\,, \quad q = 2d \sup \int_{y}^{\infty} u_o(x)dx\,.$$

We claim that

$$\lim_{t\to\infty} |u(t) - N(t;p,q|_{L^1} = 0\,. \tag{10.45'}$$

PROOF: Part (i) is a simple calculation; part (ii) follows from the explicit description given in Theorem 10.3 of admissible solutions in terms of their initial data. We leave the working out to the reader. □

It follows from the asymptotic description (10.45′) that any conserved quantity for admissible solutions is a function of p and q.

10.3. Hyperbolic Systems of Conversation Laws

These are systems of equations of the form

$$\partial_t u_i + \partial_x f_i = 0\,, \quad i = 1,\dots,n\,; \tag{10.46}$$

each f_i is a function of all the $u_1, \dots, u_n$. Denote by u the column vector of the u_i; then (10.46) can be written as a quasi-linear system

$$u_t + A(u)u_x = 0\,, \tag{10.46'}$$

where the rows of A are the gradients of the f_i:

$$A_{ij} = \frac{\partial f_i}{\partial u_j}\,. \tag{10.47}$$

We assume that the system (10.46′) is hyperbolic, that is, that for all values of u, $A(u)$ has real and distinct eigenvalues $a_k(u)$,

$$a_1 < a_2 < \cdots < a_n\,.$$

We denote by $r_k(u)$ the right eigenvectors:

$$Ar_k = a_k r_k\,. \tag{10.48}$$

We need (10.46′) to be *genuinely nonlinear*; we require not only that the eigenvalues a_k should depend on u, that is $\operatorname{grad} a_k \neq 0$, but also that it be not orthogonal to r_k: $\operatorname{grad} a_k \cdot r_k \neq 0$. We normalize r_k so that

$$r_k \cdot \operatorname{grad} a_k = 1 . \tag{10.49}$$

If for a characteristic field $r_k \cdot \operatorname{grad} a_k \equiv 0$, we call the k^{th} characteristic field *linearly degenerate*.

Let u be a solution of the systems of conservation laws (10.46) that is piecewise continuous. Then across a discontinuity the RRH jump condition for each conservation law must be satisfied:

$$s[u_k] = [f_k], \quad k = 1, \dots, n , \tag{10.50}$$

where s is the speed of propagation of the discontinuity.

Just as in the case of scalar conservation laws, not all discontinuous solutions are admissible. We recall from (10.13′) that for scalar conservation the admissibility condition was

$$a(u^l) > s > a(a^r) ,$$

where s is the speed with which the discontinuity propagates, and $a(u^l)$, $a(u^l)$ the characteristic speeds on the left and right sides of the discontinuity.

Analogously, for systems we require that there be an index k, $1 \leq k \leq n$, such that

$$a_k(u^l) > s > a_k(u^r) , \tag{10.51}$$

and

$$a_{k-1}(u^l) < s < a_{k+1}(u^r) . \tag{10.51′}$$

For weak shocks, (10.51′) follows from (10.51). It follows that characteristics with speed $a_k(u^l), a_{k+1}(u^r), \dots, a_n(u^l)$, starting to the left of the discontinuity, and characteristics with speed $a_1(u^r), \dots, a_k(u^r)$ starting to the right, impinge on the curve across which u is discontinuous, carrying $n-k+1$, and k, a total of $n+1$ pieces of information to each point on the curve of discontinuity. This information, together with the n jump conditions (10.50), suffices to determine uniquely the $2n+1$ quantities u^l_j, u^r_j, $j = 1, \dots, n$, and s.

A discontinuity across which (10.50), (10.51), and (10.51′) are satisfied is called a *k-shock*. A piecewise continuous distribution solution, all of whose discontinuities are shocks, is an *admissible* solution.

We describe now all states u^r near u^l that can be connected to $u(l)$ through a k-shock.

THEOREM 10.10 *The set of states u^r near u^l that can be connected to u^l through a k-shock form a smooth one-parameter family of states $u(p)$, $0 \geq p \geq -\varepsilon$, $u(0) = u^l$; the shock speed s is also a smooth function of p.*

A proof can be based on bifurcation theory.

We shall now calculate the first two derivatives of $u(p)$ at $p = 0$. Differentiating the jump relations (10.50) with respect to p gives

$$\dot{s}[u] + s\dot{u} = \dot{f} = A\dot{u} ; \tag{10.52}$$

here the symbol $\dot{}$ denotes differentiation with respect to p.

At $p = 0$, $[u] = 0$, so

$$s(0)\dot{u}(0) = A(u^l)\dot{u}(0).$$

For $\dot{u}(0) \neq 0$ this can be satisfied only if $s(0)$ is an eigenvalue of $A(u^l)$:

$$s(0) = a_k(u^l), \quad \dot{u}(0) = \text{const}\, r_k(u^l). \tag{10.53}$$

By appropriate choice of parameter the constant can be chosen to be one.

Differentiating (10.52) once more and setting $p = 0$ we get

$$s\ddot{u} + 2\dot{s}\dot{u} = A\ddot{u} + \dot{A}\dot{u}.$$

Set (10.53) in this relation; dropping the subscript k we get

$$a\ddot{u} + 2\dot{s}r = A\ddot{u} + \dot{A}r. \tag{10.54}$$

To determine $\dot{s}(0)$ and $\ddot{u}(0)$ we differentiate the relation

$$ar = Ar, \quad u = u(p),$$

with respect to p; we get

$$a\dot{r} + \dot{a}r = A\dot{r} + \dot{A}r.$$

Subtract this from (10.54):

$$a(\ddot{u} - \dot{r}) + (2\dot{s} - \dot{a})r = A(\ddot{u} - \dot{r}). \tag{10.55}$$

Take the scalar product of this vector equation with the left eigenvector l of A corresponding to the eigenvalue a; we get

$$2\dot{s} - \dot{a} = 0. \tag{10.56}$$

Since $\dot{a} = \operatorname{grad} a \cdot \dot{u} = \operatorname{grad} a \cdot r = 1$ according to the normalization (10.49),

$$\dot{a}(0) = 1, \quad \dot{s}(0) = \frac{1}{2}. \tag{10.56$'$}$$

Setting (10.56) into (10.55) shows that $\ddot{u} - \dot{r}$ is an eigenvector of A

$$\ddot{u} - \dot{r} = \text{const}\, r; \tag{10.57}$$

by a suitable reparametrization we can make that constant 0. So

$$\ddot{u}(0) = \dot{r}(0). \tag{10.57$'$}$$

According to (10.56′), $\dot{s}(0)$ is positive; it follows that the left side of inequality (10.51),

$$a_k(u^l) > s(p),$$

holds for p small and negative.

Since $\dot{a}(0) = 1 = 2\dot{s}(0)$, the right side of inequality (10.51),

$$s(p) > a_k(u(p))$$

holds for p negative and small.

For shocks of medium strength an additional condition due to Liu must be imposed.

We turn now to an important class of continuous solutions of the equation (10.46′), *centered rarefaction waves*. These are solutions that are functions of

$$\frac{x - x_0}{t - t_0};$$

x_0, t_0 is the center of the wave. Waves centered at (0, 0) are of the form

$$u(x,t) = w\left(\frac{x}{t}\right), \tag{10.58}$$

w a vector valued function so chosen that (10.46′) is satisfied. Denote x/t by q, and denote differentiation with respect to q by $'$. Then

$$u_t = -w'\frac{x}{t^2}, \quad u_x = \frac{w'}{t}.$$

Setting this into

$$u_t + A(u)u_x = 0$$

gives

$$[A(w) - q]w' = 0.$$

This equation is satisfied if

$$q = a(w(q)), \quad w'(q) = r(w(q)). \tag{10.59}$$

Differentiate the first relation; using the second relation, and the normalization (10.49) we get

$$1 = \operatorname{grad} a \cdot w' = \operatorname{grad} a \cdot r = 1,$$

which shows that the relations (10.59) are consistent. The second equation in (10.59),

$$w' = r(w) \tag{10.59'}$$

is an ordinary differential equation; we specify its initial value at $q_0 = a(u^l)$ as

$$w(q_0) = u^l.$$

A solution will exist for q close enough to q_0.

We define now the following continuous, piecewise smooth solution:

$$u(x,t) = \begin{cases} u^l & \text{for } x \le q_0 T \\ w(x/t) & \text{for } q_0 t < x < (q_0 + \varepsilon)t \\ u^r & \text{for } (q_0 + \varepsilon)t \le x, \end{cases} \tag{10.60}$$

where

$$u^r = w(q_0 + \varepsilon).$$

Note that the solution (10.12′) is a special instance of a centered rarefaction wave.

Take $a(u)$ to be $a_k(u)$; we say that the states u^l and u^r are connected by the *centered k-rarefaction wave* (10.60). We summarize:

THEOREM 10.11 *The set of states u^r near u^l that can be connected to u^l by a centered k-rarefaction wave form a smooth one-parameter family.*

Introduce as new parameter $p = q - q_0$; we can then combine the states u^r that can be connected to u^l through either a k-shock or a k-rarefaction wave into a one-parameter family $u(p)$, $-\varepsilon < p < \varepsilon$.

THEOREM 10.12 *This one-parameter family is twice differentiable.*

PROOF: We only have to verify that the two one-parameter families out of which we have built $u(p)$ have the same first and second derivatives at $p = 0$.

According to formulas (10.53) and (10.57′), for the $p \leq 0$ branch

$$\dot{u}(0) = r\,, \quad \ddot{u}(0) = \dot{r}\,.$$

For the branch $p \geq 0$ we deduce from (10.59′), and by differentiating (10.59′), that

$$w'(0) = r\,, \quad w''(0) = r'\,,$$

the same as for the other branch. □

THEOREM 10.13 *Suppose that the k^{th} characteristic field is degenerate in the sense that*

$$\operatorname{grad} a_k \cdot r_k \equiv 0\,. \tag{10.61}$$

Given u^l, there is a one-parameter family of states u^r such that

$$u(x,t) = \begin{cases} u^l & \text{for } x < st \\ u^r & \text{for } st < x\,, \end{cases} \tag{10.62}$$

where

$$s = a_k(u^l)\,,$$

is a solution in the distribution sense of the conservation laws (10.46).

PROOF: We shall construct a one-parameter family of states $u(p)$ so that for $u^r = u(p)$ the jump conditions (10.50) are satisfied. To this end we differentiate the jump conditions with respect to p; since s is independent of p, we get

$$s\dot{u} = A\dot{u}\,. \tag{10.63}$$

We set $s = a_k(u^l)$, and require that

$$\dot{u} = r_k(u)\,, \quad u(0) = u^l\,.$$

We claim that $a_k(u)$ is constant along this one-parameter family; this follows from

$$\dot{a}_k = \operatorname{grad} a_k \cdot \dot{u} = \operatorname{grad} a_k \cdot r_k = 0\,,$$

where we have used (10.61).

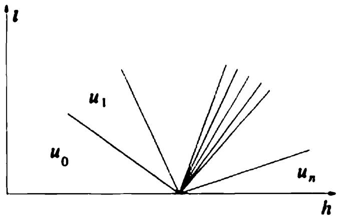
It follows that (10.63) is satisfied for all p, from which the jump conditions follow. □

Solutions of form (10.62) are called *contact discontinuities*.

We now have all ingredients in hand to construct an admissible solution of the *Riemann initial value problem*:

$$u_0(x) = \begin{cases} u_0 & \text{for } x < 0 \\ u_n & \text{for } 0 < x\,. \end{cases} \tag{10.64}$$

THEOREM 10.14 *Suppose that states u_0 and u_n are sufficiently close. Then the initial value problem with u_0 given by* (10.64) *has an admissible solution, consisting of $n + 1$ constant states $u_0, u_1, \ldots, u_m$, separated by a shock, a centered rarefaction wave, or a contact discontinuity, one of each family.*



PROOF: The state u_0 can be connected through a 1-wave to a one-parameter family $u_1(p_1)$ of states to the right of u_0; u_1 in turn can be connected through a 2-wave to a one-parameter family $u_2(p_1, p_2)$ of states to the right of u_1. Continuing in this fashion we can connect u_0 through a succession of n waves to an n-parameter family $u_n(p_1, \ldots, p_n)$ of states.

According to (10.59′),

$$\frac{\partial u}{\partial p_k} = r_k\,;$$

since the r_k are linearly independent, it follows from the implicit function theorem that $u_n(p_1, \ldots, p_n)$ maps u_0 one-to-one onto a neighborhood of u_0. □

How large this neighborhood is is an interesting and important problem for the equations of fluid dynamics.

Sergei Godunov had used the solution of the Riemann initial value problem for solving approximately arbitrary initial value problems, as follows: the given initial value $u_0(x)$ is approximated by a piecewise constant initial function $u^\delta(x)$,

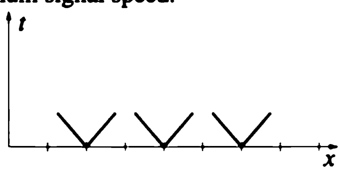
$$u^\delta(x) = m_h\,, \quad h\delta < x < (h+1)\delta \quad h = 0, \pm 1, \ldots, \tag{10.65}$$

where δ is the spatial discretization, a small number, and m_h is the average of $u_0(x)$ over the h^{th} interval $[h\delta, (h+1)\delta]$.

The initial value problem with initial value u^δ given by (10.65) can be solved exactly. At each point $h\delta$ we have to solve a Riemann initial value problem. The waves issuing from two neighboring points of discontinuity $h\delta$ and $(h+1)\delta$ will not interact as long as

$$t a_{\max} \le \frac{\delta}{2}\,,$$

where $a_{\max}$ is the maximum signal speed.



So the solutions of the Riemann problems can be combined into an exact solution.

At time $t = \delta/2a_{\max}$ we replace this exact solution by averaging with one that is piecewise constant, and repeat the process.

Numerical experiments strongly suggest that Godunov's method supplies good approximations to exact solutions of the equations of compressible flow. Leveque and Temple have proved convergence of the method for certain 2×2 systems, but proof of convergence of fluid dynamics is still lacking.

In 1965 James Glimm suggested the following modification of Godunov's method: instead of defining the quantities m_j as the average of the initial values on the interval $[h\delta, (h+1)\delta]$, set

$$m_j = u(h\delta + r\delta, t),$$

where r is a number chosen randomly in the unit interval $[0, 1]$. At the next time step another number r is chosen at random, and so on. Glimm then proved that with probability 1 the approximate solutions constructed in this fashion converge to an exact solution. The main ingredient of the proof if a new, powerful estimate for the approximate solutions.

For his proof to work, Glimm assumed that the given initial value $u_0(x)$ has small oscillation and small total variation. Robin Young showed how to prove convergence when the total variation of u_0 is not small; see also Schochet.

What about uniqueness of solutions of initial value problems? Recently Bressan and Bianchini have proved the following vast generalization of Theorem 10.1:

THEOREM 10.15 *Let u and v be a pair of admissible solutions of a hyperbolic system of conservation laws, whose initial values belong to L^1. Then*

$$\int |u(x,t) - v(x,t)|dx \le \text{const} \int |u(x,0) - v(x,0)|dx. \tag{10.66}$$

Uniqueness of solutions with prescribed initial values is a corollary. See also Bressan and LeFloch.

Liu and DiPerna have studied the behavior of solutions of genuinely nonlinear systems of conservation laws for t large. Under some mild assumptions on the initial values, the analogues of Theorems 10.7 and 10.9, derived for scalar conservation laws, holds:

(i) Solutions decay as $1/\sqrt{t}$ as t tends to ∞.

(ii) As $t \to \infty$

$$|u(t) - N(t; p, q)|_{L^1} \to 0,$$

where $N(x, t; p, q)$ is a superposition on n functions of form (10.45), depending on $2n$ parameters p, q.

10.4. The Viscosity Method and Entropy

The equations of compressible flow consist of the laws of conservation of mass, momentum, and energy; they form a hyperbolic system of conservation laws. When viscosity is included, the equations become partly parabolic. The solutions of the inviscid equations, including solutions with shocks, are the limits of the solutions of the viscous equation as the coefficient of viscosity tends to zero. This has inspired the *vanishing viscosity method* for solving the initial values problem for hyperbolic systems of conservation laws

$$(10.46') \qquad u_t + A(u)u_x = 0 \quad A_{ij} = \frac{\partial f_i}{\partial u_j}, \quad u(x,0) = u_0(x),$$

as the limit $\varepsilon \to 0$ of solutions of the parabolic system

$$(10.67) \qquad u_t + A(u)u_x = \varepsilon u_{xx}$$

with the same prescribed initial values u_0.

Let $u(x,t)$ be a smooth solution of (10.46′) for $t < T$, its initial value $u_0(x)$ in L^1. Denote by $u^\varepsilon(x,t)$ the solution of the parabolic system (10.67) with initial value $u_0(x)$. It can be shown that for $t < T$,

$$(10.68) \qquad \lim_{\varepsilon\to 0} u^\varepsilon(x,t) = u(x,t).$$

But more is true; suppose $u(x,t)$ is an admissible solution of the system of conservation laws (10.46); then (10.68) holds in the L^1 sense for each t. This is a deep, recent result; see Bianchini and Bressan. We shall not present the proof, but only sketch an argument why limits of solutions of (10.67) that are discontinuous would satisfy the admissibility condition (10.51). We rewrite equation (10.67) in conservation form

$$(10.67') \qquad u_t + f(u)_x = \varepsilon u_{xx}.$$

Denote by u^ε its solution whose initial value is given by $u(x,t)$ the limit of u^ε as $\varepsilon \to 0$. Suppose u has a shock traveling with speed s; denote by u^l and u^r the values of u on the two sides of the shock. The jump conditions

$$s(u^l - u^r) = f(u^l) - f(u^r)$$

are satisfied, and so are the admissibility conditions (10.51) and (10.51′):

$$(10.69) \qquad \begin{aligned} a_k(u^l) > s > a_k(u^r), \\ a_{k-1}(u^l) < s < a_{k+1}(u^r). \end{aligned}$$

It is plausible to suppose that near the shock, $u^\varepsilon(x,t)$ has approximately the shape of a traveling wave of the form

$$(10.70) \qquad u^\varepsilon(x,t) \simeq v\left(\frac{x - st}{\varepsilon}\right),$$

where

$$(10.71) \qquad v(-\infty) = u^l, \quad v(\infty) = u^r.$$

Set (10.70) into (10.67′); we get

$$(10.72) \qquad -sv' + f(v)' = v'',$$

where $v = v(p)$, $' = d/dp$. Integrate (10.72):

$$(10.72') \qquad -sv + f(v) = v' + c,$$

c a constant vector. Since $v'(p)$ tends to zero as $p \to \pm\infty$, c must be so chosen that

$$c = -su^l + f(u^l) = -su^r + f(u^r);$$

here we use the jump condition satisfied by u^l and u^r.

The solution of (10.72′) is a curve along the vector field $c - sv + f(v)$ in $\mathbb{R}^n$, connecting the point u^l to u^r; the field is zero at both points. The linear approximation to the fields at these points is

$$(A(u^l) - sI)(v - u^l) \quad \text{and} \quad (A(u^r) - sI)(v - u^r).$$

According to the admissibility condition (10.69), $A(u^l) - sI$ has $n - k + 1$ positive eigenvalues; so the unstable manifold issuing from u^l is $(n - k + 1)$–dimensional. Similarly, $A(u^r) - sI$ has k negative eigenvalues; so the stable manifold converging to u^r is k-dimensional. Foy has shown that for u^r close to u^l the unstable and stable manifold intersect in a smooth curve connecting u^l to u^r, which in the right parameterization satisfies equation (10.72′).

This completes the heuristic argument that limits of solutions of the equation with viscosity tending to zero have admissible discontinuities.

We turn now to the concept of entropy.

DEFINITION A function $S(u_1, \dots, u_n)$ is called an *entropy* for a system of conservation laws (10.46) if every smooth solution of (10.46) satisfies an additional conservation law

$$\partial_t S + \partial_x F = 0, \tag{10.73}$$

where $F(u_1, \dots, u_n)$ is called the *entropy flux*. In addition, S(u) is required to be a *convex* function of u.

To derive an equation satisfied by the pair S and F we write the system of conservation laws in the form (10.46′), and multiply this system on the left by grad S, denoted as S_u; we get

$$\partial_t S + S_u A(u) u_x = 0.$$

This implies the conservation law (10.73) if

$$S_u A(u) = F_u. \tag{10.74}$$

This is a system of n *linear* first-order equations for the pair of functions S and F. For $n = 1$ this system is under-determined, for $n = 2$ determined, and for $n > 2$ over-determined.

What if $u(x, t)$ is an admissible discontinuous solution of (10.46) in the distribution sense? Does $S(u(x,t))$, $F(u(x,t))$ satisfy (10.73) in the sense of distribution? To answer this question we regard admissible discontinuous solutions as limits of solutions of the viscous equation (10.67′). Multiply (10.67′) by S_u; if we use (10.74) we get

$$S_t + F_x = \varepsilon S_u u_{xx}. \tag{10.75}$$

By calculus

$$S_{xx} = S_u u_{xx} + S_{uu} u_x \cdot u_x;$$

here S_{uu} denotes the matrix of the second derivatives of S. Since $S(u)$ is assumed to be convex, S_{uu} is a positive matrix, and therefore $S_{uu}u_x \cdot u_x$ positive. It follows that

$$S_{xx} \le S_u u_{xx} .$$

Setting this into (10.75) we conclude that

(10.76) $$S_t + F_x \le \varepsilon S_{xx} .$$

Suppose u^ε converges in L^1 for all t, and that $u^\varepsilon(x,t)$ stays uniformly bounded. Then $S(u^\varepsilon)$ converges in L^1 to $S(u)$, and therefore the right side of (10.76) converges to zero in the sense of distributions. Since the limit of a negative distribution is negative, it follows that, in the sense of distributions,

(10.77) $$\partial_t S(u) + \partial_x F(u) \le 0$$

for any admissible distribution u and for any entropy–entropy flux pair S, F.

The *entropy inequality* (10.77) has some useful consequences:

THEOREM 10.16 *Let u be an admissible solution of a hyperbolic system of conservation laws. Suppose that the sup norm and the L^1 norm with respect to x are bounded for all t. Let S be an entropy for the system of conservation laws. Then*

(i)

(10.77′) $$\int S(u(x,t))dx$$

is a decreasing function of t.

(ii) *At a point of discontinuity of u*

(10.77″) $$F(u^r) - F(u^l) < s[S(u^r) - S(u^l)] .$$

We show now in the special case of conservation law $u_t + (u^2/2)_x = 0$ that (10.77″) is equivalent to the admissibility condition $u^r < u^l$. Take $S(u) = u^2/2$; by equation (10.74), $F(u) = 2u^3/3$. A simple calculation shows that $[F(u)] - s[S(u)] = \frac{2}{3}[(u^r)^3 - (u^l)^3] - \frac{u^r+u^l}{2}[(u^r)^2 - (u^l)^2] = (u^r - u^l)^3/6$. So (10.77″) is equivalent to $u^r < u^l$.

For scalar conservation laws every convex function is an entropy.

Systems of two conservation laws

(10.78) $$u_t + f_x = 0\,, \quad v_t + g_x = 0\,,$$

where f and g are functions of u and v, have many special properties not shared by systems of more than two equations. Here we shall discuss the construction of entropies.

Equations (10.78) can be rewritten as

(10.78′) $$\begin{pmatrix} u \\ v \end{pmatrix}_t + A \begin{pmatrix} u \\ v \end{pmatrix}_x = 0\,,$$

where

(10.79) $$A = \begin{pmatrix} f_u & f_v \\ g_u & g_v \end{pmatrix} .$$

Equation (10.78′) is hyperbolic if the matrix A has real, distinct eigenvalues. The equation (10.74) for entropy and entropy flux is

$$(S_u, S_v)A = (F_u, F_v).$$

Eliminate F; we get

$$aS_{uu} + bS_{uv} + cS_{vv} = 0, \tag{10.78''}$$

where

$$a = -f_v, \quad b = f_u - g_v, \quad c = g_u.$$

Clearly the second-order equation (10.78″) is hyperbolic if (10.78′) is. It is easy to see that it has solutions that are convex in the small. Under a simple additional condition it has convex solutions in the large; see Lax (1973).

For $n > 2$ there may be no entropy; but for many systems describing physical phenomena there is. And when there is, it leads to a symmetric form of the system. We rewrite equation (10.74) component-wise, using equation (10.47) to express the entries of A as

$$A_{ij} = \frac{\partial f_i}{\partial u_j} = f_{i,j}. \tag{10.79'}$$

So $S_u A = F_u$ becomes

$$\sum_i S_i f_{i,j} = F_j, \tag{10.80}$$

where subscript j denotes differentiation with respect to u_j. Differentiate (10.80) with respect to u_k, k any index whatsoever. We get

$$\sum_i S_{ik} f_{i,j} + S_i f_{i,jk} = F_{jk}.$$

Since the matrix of second derivatives is symmetric, the second term on the left, and the term on the right, are symmetric functions of j, k. Therefore so is the first term. We can, using the symmetry of S_{ik} rewrite that term as

$$\sum_i S_{ki} f_{i,j},$$

which can be rewritten without indices as the matrix $S_{uu}A$. So we have shown:

If a system of conservation laws has an entropy S, then $S_{uu}A$ is symmetric, where A is defined by (10.79′). Note that it follows that A has real eigenvalues, that is, that the system is hyperbolic.

The system is not only hyperbolic but symmetric hyperbolic, as discussed in Section 4.4, Chapter 4. To see this denote the Hessian S_{uu} as H, and multiply equation (10.46′) by H; we get

$$Hu_t + HAu_x = 0,$$

where H is symmetric and positive, HA symmetric.

10.5. Finite Difference Methods

In this section we shall present an overview of the numerical approximation of solutions of the initial value problem for hyperbolic systems of conservation laws by discretizing the differential equation. As in Chapter 8, we denote by u_h^n an approximation to the value of u at the point $x = h\delta$ at time $t = n\varepsilon$; δ and ε are the space and time scales of discretization. We denote $\varepsilon/\delta = \lambda$.

We shall only discuss explicit, two-level schemes, where u_h^{n+1} is expressed in terms of the values of u_k^n. It is of paramount importance to write the scheme in *conservation form*; by this we mean of the form

$$u_h^{n+1} = u_h^n - \lambda\left\{f_{h+1/2}^n - f_{h-1/2}^n\right\}, \tag{10.81}$$

where $f_{h+1/2}^n$, the *numerical flux*, is, in the simplest cases, of the form

$$f_{h+1/2}^n = f\left(u_h^n, u_{h+1}^n\right). \tag{10.82}$$

Here $f(u, v)$, the numerical flux function, is required to satisfy the consistency condition

$$f(u, u) = f(u). \tag{10.83}$$

More generally, the numerical flux function could be of form

$$f_{h+\frac{1}{2}} = f(u_{h-1}, u_h, u_{h+1}, u_{h+2}); \tag{10.82'}$$

$f(u, v, w, z)$ is required to satisfy the consistency condition

$$f(u, u, u, u) = f(u). \tag{10.83'}$$

We now give some examples.

EXAMPLE 9 The conservation form of the LF scheme, discussed in Section 8.1 of Chapter 8, is

$$u_h^{n+1} = \frac{1}{2}\left(u_{h-1}^n + u_{h+1}^n\right) - \frac{\lambda}{2}\left\{f\left(u_{h+1}^n\right) - f\left(u_{h-1}^n\right)\right\}.$$

This is of form (10.81), with

$$f(u, v) = \frac{f(u) + f(v)}{2} + \frac{u - v}{2\lambda}.$$

Clearly, the consistency condition (10.83) is satisfied.

EXAMPLE 10 The conservation form of the LW scheme is discussed in Section 8.4 of Chapter 8. This is a second-order scheme, based on the Taylor approximation

$$u(t + \varepsilon) = u + \varepsilon u_t + \frac{\varepsilon^2}{2} u_{tt} + O(\varepsilon^3). \tag{10.84}$$

We can approximate $u_t = -f_x$ with second-order accuracy by a symmetric difference quotient. u_{tt} can be expressed as follows:

$$u_{tt} = -f(u)_{xt} = -f(u)_{tx} = -(A(u)u_t)_x = (A(u)f(u)_x)_x.$$

So we set

$$u_t = -\frac{1}{2\delta}\{f(u(x + \delta)) - f(u(x - \delta))\} + 0(\delta^2),$$

and

$$u_{tt} = \frac{1}{\delta^2}\left\{A\left(x+\frac{\delta}{2}\right)[f(u(x)) - f(u(x-\delta))]\right.$$
$$\left. - A\left(x-\frac{\delta}{2}\right)[f(u(x)) - f(u(x-\delta))]\right\} + O(\delta).$$

The difference scheme obtained by inserting the approximations above into (10.84) is of the form (10.81), with the numerical flux

$$f(u,v) = \frac{f(u)+f(v)}{2} - \frac{\lambda}{2}\left[\frac{A(u)+A(v)}{2}(f(v)-f(u))\right]. \tag{10.85}$$

Clearly, the consistency condition (10.83) is satisfied.

EXAMPLE 11 The Richtmyer two-step method uses the LF scheme to construct an approximation to $u_{h+1/2}^{n+1/2}$ and $u_{h-1/2}^{n+1/2}$, and makes a "leap frog" to determine u_h^{n+1}:

$$u_h^{n+1} = u_h^n - \lambda\left\{f\left(u_{h+1/2}^{n+1/2}\right) - f\left(u_{h-1/2}^{n+1/2}\right)\right\}.$$

So in this case the numerical flux is

$$f(u,v) = f\left(\frac{u+v}{2} - \frac{\lambda}{2}[f(v)-f(u)]\right). \tag{10.85'}$$

Clearly, the consistency condition (10.83) is satisfied.

The reader is invited to verify that the Richtmyer two-step scheme is of second-order accuracy. Algorithmically it is more efficient then the LW scheme, for it avoids multiplication by the matrix $(A(u)+A(v))/2$.

EXAMPLE 12 Godunov's scheme described in Section 10.3 can be put in the form (10.81), (10.82). In this scheme the approximate solution is represented as a piecewise constant function:

$$u(x,t_n) = u_h^n \quad \text{in} \quad h\delta < x < (h+1)\delta. \tag{10.86}$$

$u(x,t_{n+1})$ is constructed by solving exactly the conservation laws with initial values (10.86). This is accomplished by solving the Riemann problems at the points of discontinuities of $u(x,t_n)$; $t_{n+1}-t_n$ has to be restricted to be so small that the waves issuing from the points of discontinuity don't interact. The resulting exact solution is then averaged over each interval $[h\delta,(h+1)\delta]$ to determine u_h^{n+1}:

$$u_h^{n+1} = \frac{1}{\delta}\int_{h\delta}^{(h+1)\delta} u(x,t)dx. \tag{10.87}$$

This algorithm can be described somewhat differently. Let $u(x,t)$ be any piecewise continuous solution of a system of conservation laws in the rectangle

$$h\delta \le x \le (h+1)\delta, \quad t_n \le t \le t_{n+1}.$$

Integrate the conservation law

$$u_t + f(u)_x = 0$$

over this rectangle; we get

$$\int_{h\delta}^{(h+1)\delta} u(x,t_{n+1})dx = \int_{h\delta}^{(h+1)\delta} u(x,t_n)dx - \int_{t_n}^{t_{n+1}} [f(u(h+1)\delta,t) - f(u(h\delta,t))]dt\,. \tag{10.88}$$

Choose $t_{n+1}-t_n$ so small that no wave issuing from $(h\delta, t_n)$ reaches the interval $((h+1)\delta, t)$, $t_n < t < t_{n+1}$, nor any wave issuing from $((h+1)\delta, t_n)$ reaches $(h\delta, t)$, $t_n < t < t_{n+1}$. Then $u(h\delta, t)$ and $u((h+1)\delta, t)$ are constant on $[t_n, t_{n+1}]$, and so the time integrals in (10.88) can be evaluated simply as

$$(t_{n+1}-t_n)[f(u(h+1)\delta, t_{n+1}) - f(u(h\delta, t_{n+1}))]\,. \tag{10.89}$$

The state $u(h\delta, t_{n+1})$ is uniquely determined by the two states u_{h-1}^n and u_h^n through the process of solving the Riemann problem. Therefore we can write

$$f(u(h\delta, t_{n+1})) = f\left(u_{h-1}^n, u_h^n\right); \tag{10.90}$$

similarly

$$f(u((h+1)\delta, t_{n+1})) = f\left(u_h^n, u_{h+1}^n\right). \tag{10.90$'$}$$

Now set (10.89) and (10.90), (10.90′) into (10.88), and divide by δ. In view of the definition of u_h^{n+1} as the average (10.87), we can write the resulting relation as

$$u_h^{n+1} = u_h^n - \lambda^n\left[f\left(u_h^n, u_{h+1}^n\right) - f(u_{h-1}^n, u_h^n)\right],$$

where $\lambda^n = (t_{n+1} - t_n)/\delta$. This is of the same form as (10.81), (10.82). The consistency condition

$$f(u,u) = f(u)$$

is satisfied, for when $u_{h-1}^n = u_h^n$, the solution of the Riemann problem is trivial, and $u(h\delta, t_{n+1}) = u_{h-1}^n = u_h^n$.

This analysis shows that $t_{n+1} - t_n$ may be chosen as $\delta/a_{\max}$, twice as large as permitted by the analysis in Section 10.3.

EXAMPLE 13 As in Example 12 the approximate solutions are piecewise constant as in (10.86). We integrate the conservation law over a shifted rectangle

$$\left(h-\frac{1}{2}\right)\delta \le x \le \left(h+\frac{1}{2}\right)\delta\,, \quad t_n \le t \le t_{n+1}\,. \tag{10.91}$$

We get a formula analogous to (10.88). If we choose $t_{n+1} - t_n < \delta/2a_{\max}$, the exact solution on the vertical sides $x = (h-1/2)\delta$ and $x = (h+1/2)\delta$ are just u_{h-1}^n and u_h^n. The resulting formula, after division by δ, is

$$u_{h-1/2}^{n+1} = \frac{u_{h-1}^n + u_h^n}{2} - \lambda\left[f\left(u_h^n\right) - f\left(u_{h-1}^n\right)\right].$$

This is just the LW scheme, with a shift.

For the basic result about difference schemes in conservation form, see Lax (1954) and Lax-Wendroff (1960).

THEOREM 10.17 *Let $u_{\delta,\varepsilon}$ be an approximate solution of the system of conservation laws*

$$u_t + f(u)_x = 0,$$

generated by a difference equation for form (10.81), (10.82), (10.83). *Regard $u_{\delta,\varepsilon}$ as equal to u_h^n in the rectangle* (10.91). *Suppose $u_{\delta,\varepsilon}$ converges boundedly and in L^1 to a limit u as a δ and ε tend to zero. Then u is a solution in the sense of distribution of the conservation laws.*

PROOF: The proof is simple. Choose any differentiable function $w(x, t)$ that is zero for $|x|$ and t large. Multiply (10.81) by $w(h\delta, n\varepsilon)$ and sum over all h and all $n \geq 0$. Then sum by parts; we get

$$\sum_{0<n,h} u_h^n\{w(h\delta, (n-1)\varepsilon) - w(h\delta, n\varepsilon)\}$$
$$+ \lambda f_{h+1/2}^n\{w(h\delta, n\varepsilon) - w(h+1)\delta, n\varepsilon)\} - \sum u_h^0 w(h\delta, 0) = 0.$$

Multiply by δ; using the relation $\lambda\delta = \varepsilon$ we can express this equation in terms of difference quotients of w with respect to t and x. Since w is differentiable, these tend to w_t and w_x. By assumption, u_h^n converges to u in L^1, and since u_h^n are assumed to be bounded, $f_{h+1/2} = f(u_h^n, u_{h+1}^n)$ converges to $f(u, u) = f(u)$. So in the limit we get

$$-\iint uw_t + f(u)w_x dx\, dt - \int u(x, 0)w(x, 0)dx = 0,$$

since this holds for all test functions w, u is a solution of

$$u_t + f(u)_x = 0$$

in the sense of distributions. □

Note that we haven't shown that the limit is an admissible solution. In fact there are examples where $u_{\delta,\varepsilon}$ converge to an inadmissible solution.

Note that if we only prove weak convergence of $u_{\delta,\varepsilon}$ to u, it doesn't follow that u satisfies the conservation laws.

We have left open the question of how one proves convergence of the solutions of a conservative finite difference scheme. We remark, as was already pointed out in Examples 12 and 13, that the Courant-Friedrichs-Lewy condition,

$$a_{\max}\varepsilon \leq \delta$$

must be satisfied, just as it must be satisfied in the linear case.

There is a big literature about convergence. In the scalar case, for special forms of $f(u)$ the difference equations can be linearized, solved explicitly, and the passage to the limit carried out; see Lax (1954, 1957). The general case was settled by Vvedenskaya. For isentropic compressible flows X. Ding, G. Q. Chen, and P. Luo have proved convergence of the LW scheme.

The convergence of second-order methods, such as LW and the Richtmyer two-step method is more delicate, just as in the linear case; see Section 8.4. In Section 8.5 we described the strategy of Harten and Zwass of switching from a high-order scheme such as LW, to a low-order scheme such as LF, in the neighborhood of discontinuities. This switch can be carried out for schemes in conservation form by applying the switch to the numerical fluxes themselves.

The concept of difference schemes in conservation form makes sense for approximating conservation laws in several space variables:

$$u_t + f(u)_x + g(u)_y = 0.$$

Many highly effective methods have been designed which work very well in resolving complicated flows. We mention in particular the methods designed by Alexander Choin, by Van Leer, the popular Colella-Woodland piecewise parabolic method, the ENO scheme of Harten and Osher, the method of Tadmor and Kurganov, the positive schemes of Zhu Dong Liu and Lax, by Yulian Radvogin, and many others.

10.6. The Flow of Compressible Fluids

The flow of inviscid, non-heat-conducting fluid that depends on a single space variable is described by a hyperbolic systems of first-order conservation laws of the type studied in Section 10.3; they go back to Euler. The conserved quantities are

$$\begin{aligned} \rho &= \text{mass per unit volume,} \\ m &= \text{momentum per unit volume,} \\ E &= \text{energy per unit volume.} \end{aligned}$$

ρ is called the *density* of the fluid; m can be expressed as ρv, where v is flow velocity. Energy E is the sum of *internal energy* per unit volume plus *kinetic energy* per unit volume. Denoting by e internal energy per unit mass, we can express E as

$$E = \rho e + \frac{1}{2}\rho v^2. \tag{10.92}$$

Mass flux is determined by the rate at which fluid is transposed out of the region; for one-dimensional flow it is

$$f_1 = v\rho.$$

Momentum flux is the sum of the rate at which the flow transports momentum of the fluid out of the region plus the rate at which the force of pressure imparts momentum:

$$f_2 = vm + p.$$

Energy flux is the sum of the rate at which the flow transports energy of the fluid out of the region plus the work alone by the force of pressure:

$$f_3 = vE + vp.$$

The three thermodynamic variables ρ, e, and p are related to each other by an *equation of state* which we put in the form

$$p = p(e, \rho). \tag{10.93}$$

The three fluxes have to be expressed as functions of the conserved quantities:

$$(10.94)\qquad \begin{gathered} f_1 = m\,, \quad f_2 = \frac{m^2}{2} + p\left(\frac{E}{\rho} - \frac{m^2}{2\rho^2}, \rho\right), \\ f_3 = \frac{m}{\rho}E + \frac{m}{\rho}p\left(\frac{E}{\rho} - \frac{m^2}{2\rho^2}, \rho\right). \end{gathered}$$

The equations of motion are

$$(10.95)\qquad \rho_t + f_{1x} = 0\,, \quad m_t + f_{2x} = 0\,, \quad E_t + f_{3x} = 0\,.$$

They can be written in the form

$$\begin{pmatrix} \rho \\ m \\ E \end{pmatrix}_t + A\begin{pmatrix} \rho \\ m \\ E \end{pmatrix}_x = 0\,,$$

where the matrix A is

$$(10.96)\qquad A = \begin{pmatrix} \operatorname{grad} f_1 \\ \operatorname{grad} f_2 \\ \operatorname{grad} f_3 \end{pmatrix}.$$

A is expressed in terms of the partial derivatives of the fluxes. A simple calculation will show that $\operatorname{tr} A = 3v$, and that v is an eigenvalue of A. It follows that the other two eigenvalues are of the form $v + c$ and $v - c$. A calculation of the determinant of A shows that c depends only on the thermodynamic variables. The requirement that c be real imposes some conditions on the equation of state.

Next we compute an entropy. We rewrite the conservation of energy equation by expressing E as $\rho e + \frac{1}{2}\rho v^2$:

$$(\rho e)_t + \frac{1}{2}(mv)_t + (me)_x + \frac{1}{2}(v^2 m)_x + (vp)_x = 0\,.$$

Using the other conservation laws

$$\begin{aligned} \rho_t + m_x &= 0 \\ m_t + (vm)_x + p_x &= 0 \end{aligned}$$

we deduce that

$$\rho(e_t + ve_x) + pv_x = 0\,.$$

We use the conservation of mass

$$\rho_t + v\rho_x + \rho v_x = 0$$

to obtain

$$(10.97)\qquad \rho^2(e_t + ve_x) - p(\rho_t + v\rho_x) = 0\,.$$

Recall that p is a function of e and ρ. Let $M(e, \rho)$ be a solution of

$$(10.98)\qquad pM_e + \rho^2 M_\rho = 0\,.$$

Multiply (10.97) by M_e; using (10.98) we get

$$\rho^2 M_e(e_t + ve_x) + \rho^2 M_\rho(\rho_t + v\rho_x) = 0\,.$$

This can be rewritten as

$$M_t + vM_x = 0. \tag{10.99}$$

It follows that for any smooth solution of the conservation laws, M is constant along particle paths.

Combining (10.97) with the conservation of mass equation we get

$$(\rho M)_t + (mM)_x = 0,$$

an additional conservation law.

Clearly we may replace M by any function of M. To obtain an entropy we have to choose this function, also denoted as M, so that ρM is a convex function of e and ρ. This is possible under the conditions imposed on the equation of state to make c real.

We shall now drastically specialize the equation of state to *polytropic* gases, defined as one whose interval energy is proportional to its temperature. The equation of state of a polytropic gas is of the form

$$p = (\gamma - 1)e\rho. \tag{10.100}$$

The constant γ, called the adiabatic exponent, lies between 1 and 3. Setting (10.94) into (10.96) we get after a brief calculation that

$$A = \begin{pmatrix} 0 & 1 & 0 \\ \left(\frac{\gamma-3}{2}\right)v^2 & (3-\gamma)v & \gamma - 1 \\ \left(\frac{\gamma-2}{2}\right)v^3 - \gamma ve & \left(\frac{3}{2} - \gamma\right)v^2 + \gamma e & \gamma v \end{pmatrix}, \tag{10.101}$$

The eigenvalues of A are v, $v - c$, and $v + c$, where $c = \sqrt{\gamma p/\rho}$. The eigenvalues $v - c$ and $v + c$ are genuinely nonlinear.

The gradient of the eigenvalue $v = m/\rho$ is $(-v/\rho, 1/\rho, 0)$; the corresponding right eigenvector of A is $(1, v, \frac{1}{2}v^2)'$, clearly orthogonal to grad v. This shows that the middle eigenvalue v is linearly degenerate. Discontinuities traveling with velocity v are called *contact discontinuities*.

Waves traveling with velocity v travel along the path of the fluid. Waves traveling with velocity $v + c$ and $v - c$ travel with speed c relative to the fluid; for this reason c is called the *speed of sound*.

EXERCISE Verify that $c = \sqrt{\gamma p/\rho}$ has the dimension of velocity.

We turn now to entropy. When $p = (\gamma - 1)e\rho$, equation (10.98) is

$$(\gamma - 1)eM_e + \rho M_\rho = 0.$$

A solution of this equation is

$$M = \frac{e}{\rho^{\gamma-1}},$$

and any function of it.

EXERCISE Find a function f such that

$$S = \rho f\left(\frac{e}{\rho^{\gamma-1}}\right)$$

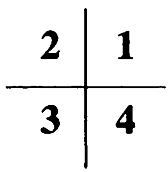
is a convex function of ρ, m, and E.

The equations of compressible flow in two (and three) dimensions are analogous. Denote the velocity vector in the x, y plane as (v, w), the momenta as $(m, n) = (\rho v, \rho w)$. The conservation laws are

$$\begin{aligned} &\rho_t + m_x + n_y = 0\,, \\ &m_t + (vm)_x + (wm)_y = 0\,, \\ &n_t(vn)_x + (wn)_y = 0\,, \\ &E_t + (v(E+p))_x + (w(E+p))_y = 0\,. \end{aligned}$$

The notion of solution in the sense of distributions is the same as in one dimension. So is the RRH jump condition across a discontinuous surface; the notion of an admissible discontinuity—shock or contact— can be defined similarly. There are useful theorems guaranteeing the existence of solutions for a finite range, see Majda, but nothing like the global existence theorems comparable to the one-dimensional case. What we do have is an impressive array of numerical methods capable of computing flows with very complicated structures. We close this chapter by presenting some calculational results.

The analogue of the Riemann problem in two dimensions is an initial value problem where the initial data are constant in each of the four quadrants 1, 2, 3, and 4. The states in the four quadrants are chosen so that the one-dimensional Riemann



problems between any two adjacent states are resolved by a *single* wave, either a shock, a rarefaction, or a contact discontinuity. For fluid flow in two dimensions a contact discontinuity is a *slip line*, across which the tangential velocity, as well as density, changes discontinuously.

Below we present the results of three numerical calculations, done by a method developed by Zhu-Dong Liu and the author. The figures represent the contour lines of density. Figure 10.1(a) pictures the interaction of four shockwaves, Figure 10.1(b) the interaction of four rarefaction waves, and Figure 10.3(c) the interaction of four contact discontinuities. The direction of propagation of each wave is clearly discernible from the figures.

What can we learn from these calculations? First of all that the interaction of four waves creates a complicated flow pattern, so that, unlike flows in one dimension, the Riemann problem in two dimensions is a not suitable building block to describe approximately general flows.

How much credence can we give to these numerical calculations? There is no proof, and there never will be, that these results approximate the exact solutions of the Riemann problem within some acceptable error bound. Our confidence is

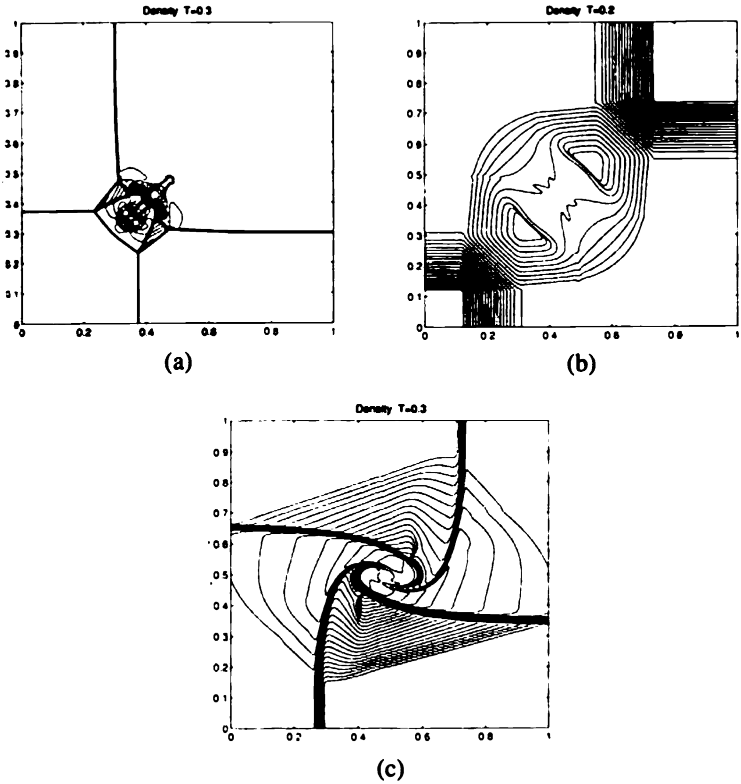


FIGURE 10.1

based on the remarkable agreement of calculations carried out by Colella and Glaz, and others, using entirely different numerical methods; see Lax, 2006.

Ami Harten has famously observed that for numerical analysts there are two kinds of truths: the truth that you prove and the truth you see when you compute.

References

The subject has a vast literature; here we list the main monographs and the articles referred to in the text.

Monographs.

Bressan, A. *Hyperbolic systems of conservation laws.* The one dimensional Cauchy problem. Oxford University Press, 2000.

Dafermos, C. M. *Hyperbolic conservation laws in continuum physics.* Grundlehren der Mathematischen Wissenschaften, 325. Springer, Berlin, 2000.

Hörmander, L. *Lectures on nonlinear hyperbolic differential equations.* Mathématiques & Applications (Berlin), 26. Springer, Berlin, 1997.

Lax, P. D. *Hyperbolic systems of conservation laws and the mathematical theory of shock waves.* Conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics, No. 11. Society for Industrial and Applied Mathematics, Philadelphia, 1973.

LeFloch, Philippe G. *Hyperbolic systems of conservation laws. The theory of classical and nonclassical shock waves.* Lectures in Mathematics ETH Zürich. Birkhäuser, Basel, 2002.

LeVeque, R. J. *Numerical methods for conservation laws.* Lectures in Mathematics ETH Zürich. Birkhäuser, Basel, 1990.

Liu, T. P. Admissible solutions of hyperbolic conservation laws. *Mem. Amer. Math. Soc.* 30 (1981): no. 240.

Majda, A. *Compressible fluid flow and systems of conservation laws in several space variables.* Applied Mathematical Sciences, 53. Springer, New York, 1984.

Serre, D. *Systems of conservation laws.* 2 vols. Cambridge University Press, Cambridge, 1999, 2000.

Smoller, J. *Shock waves and reaction-diffusion equations.* 2nd ed. Grundlehren der Mathematischen Wissenschaften, 258. Springer, New York, 1994.

Tadmor, E. Approximate solutions of nonlinear conservation laws. *Advanced numerical approximation of nonlinear hyperbolic equations (Cetraro, 1997),* 1–149. Lecture Notes in Mathematics, 1697. Springer, Berlin, 1998.

Taylor, M. *Partial differential equations.* III. *Nonlinear equations.* Applied Mathematical Sciences, 116. Springer, New York, 1996.

Articles.

Bianchini, S., and Bressan, A. Vanishing viscosity solutions of nonlinear hyperbolic systems. *Ann. of Math. (2)* 161, no. 1: 223–342, 2005.

Colella, P., and Glaz, H. M. Efficient solution algorithms for the Riemann problem for real gases. *J. Comp. Phys.* 59: 264–289, 1985.

Colella, P., and Woodward, P. The piecewise parabolic method (PPM) for gas-dynamical simulations. *J. Comp. Phys.* 54: 174–201, 1984.

Dafermos, C. M. The entropy rate admissibility criterion for solutions of hyperbolic conservation laws. *J. Differential Equations* 14: 202–212, 1973.

DeLellis, C., and Golse, F. A quantitative compactness estimate for scalar conservation laws. *Comm. Pure Appl. Math.* 58: 989–998, 2005.

Ding, X. X., Chen, G. Q., and Luo, P. Z. Convergence of the Lax-Friedrichs scheme for isentropic gas dynamics I, II. *Acta Math. Sci. (English Ed.).* 5: 415–432, 433–472.

Foy, R, L. Steady state solutions of hyperbolic systems of conservation laws with viscosity terms. *Comm. Pure Appl. Math.* 17: 177–188, 1964.

Friedrichs, K. O., and Lax, P. D. Systems of conservation equations with a convex extension. *Proc. Nat. Acad. Sci. U.S.A.* 68: 1686–1688, 1971.

Godunov, S. K. An interesting class of quasi-linear systems. *Dokl. Akad. Nauk SSSR* 139: 521–523, 1961.

Harten, A., and Osher, S. Uniformly high-order accurate nonoscillatory schemes. I. *SIAM J. Numer. Anal.* 24, no. 2: 279–309, 1987.

Hopf, E. The partial differential equation $u_t + uu_x = \mu u_{xx}$. *Comm. Pure Appl. Math.* 3: 201–230, 1950.

Hsiao, L. The entropy rate admissibility criterion in gas dynamics. *J. Differential Equations* 38: 226–238, 1980.

Iguchi, T., and LeFloch, P. G. Existence theory for hyperbolic systems of conservation laws with general flux-functions. *Arch. Ration. Mech. Anal.* 168: 165–244, 2003.

Kruzkov, S. First order quasilinear equations with several space variables. *Math USSR-Sb.* 10: 217–243, 1970.

Lax, P. D. Weak solutions of nonlinear hyperbolic equations and their numerical computation. *Comm. Pure Appl. Math.* 7: 159–193, 1954.

———. Hyperbolic systems of conservation laws. II. *Comm. Pure Appl. Math.* 10: 537–566, 1957.

———. Shock waves and entropy. *Contributions to nonlinear functional analysis (Proc. Sympos., Math. Res. Center, Univ. Wisconsin, Madison, Wis., 1971)*, pp. 603–634. Academic Press, New York, 1971.

———. The formation and decay of shock waves. *Amer. Math. Monthly* 79: 227–241, 1972.

———. Computational fluid dynamics. *J. Sci. Comput.*, Xu-Dong memorial issue, to appear.

Lax, P. D., and Liu, X-D. Solutions of two-dimensional Riemann problems of gas dynamics by positive schemes. *SIAM J. Sci. Comput.* 19: 319–340, 1998 (electronic).

Lax, P. D., and Wendroff, B. Systems of conservation laws. *Comm. Pure Appl. Math.* 13: 217–237, 1960.

LeVeque, R. J., and Temple, B. Stability of Godunov's method for a class of 2 × 2 systems of conservation laws. *Trans. Amer. Math. Soc.* 288: 115–123, 1985.

Lions, P.-L., Perthame, B., and Souganidis, P. E. Existence and stability of entropy solutions for the hyperbolic systems of isentropic gas dynamics in Eulerian and Lagrangian coordinates. *Comm. Pure Appl. Math.* 49: 599–638, 1996.

Liu, T-P. The deterministic version of the Glimm scheme. *Comm. Math. Phys.* 57: 135–148, 1977.

———. Decay to N-waves of solutions of general systems of nonlinear hyperbolic conservation laws. *Comm. Pure Appl. Math.* 30: 586–611, 1977.

———. Pointwise convergence to shock waves for viscous conservation laws. *Comm. Pure Appl. Math.* 50: 1113–1182, 1997.

Oleĭnik, O. A. On Cauchy's problem for nonlinear equations in a class of discontinuous functions. (Russian) *Doklady Akad. Nauk SSSR (N.S.)* 95: 451–454, 1954.

———. Discontinuous solutions of non-linear differential equations. (Russian) *Uspehi Mat. Nauk (N.S.)* 12: 3–73, 1957.

———. Uniqueness and stability of the generalized solution of the Cauchy problem for a quasi-linear equation. *Uspehi Mat. Nauk* 14, no. 2 (86), 165–170, 1959.

Quinn, B. K. Solutions with shocks: An example of an L_1-contractive semigroup. *Comm. Pure Appl. Math.* 24: 125–132, 1971.

Roždestvenskiĭ, B. L. A new method of solution of the Cauchy problem in the large for quasi-linear equations. (Russian) *Dokl. Akad. Nauk SSSR* 138: 309–312, 1961.

Schochet, S. Sufficient conditions for local existence via Glimm's scheme for large BV data. *J. Differential Equations* 89: 317–354, 1991.

Serre, D. Systemes hyperboliques riches de lois de conservation. (French) *Nonlinear partial differential equations and their applications. Collège de France Seminar, Vol. XI (Paris, 1989–1991)*, 248–281. Pitman Research Notes in Mathematics Series, 299. Longman, Harlow, 1994.

Tartar, L. C. The compensated compactness method applied to systems of conservation laws. *Systems of nonlinear partial differential equations (Oxford, 1982)*, 263–285. NATO Advanced Science Institutes Series C: Mathematical and Physical Sciences, 111. Reidel, Dordrecht, 1983.

Vvedenskaya, N. D. Solution of the Cauchy problem for a non-linear equation with discontinuous initial conditions by the method of finite differences. (Russian) *Dokl. Akad. Nauk SSSR (N.S.)* 111, 517–520, 1956.

Young, R. Sup-norm stability for Glimm's scheme. *Comm. Pure Appl. Math.* 46: 903–948, 1993.

APPENDIX A

Huygens' Principle for the Wave Equation on Odd-Dimensional Spheres

For simplicity we shall analyze the three-dimensional case.

EXERCISE Carry out the analysis in higher odd dimensions.

The four-dimensional Euclidean Laplace operator can be written in polar coordinates as follows:

$$\Delta_4 = \partial_r^2 + \frac{3}{r}\partial r + \frac{1}{r^2}\Delta_S ,$$

where Δ_S is the Laplace-Beltrami operator on the unit sphere S^3. The eigenfunctions of Δ_S are the spherical harmonics $h_j(\omega)$, where $H_j = r^3 h_j(\omega)$ is a harmonic polynomial of degree j, so

$$0 = \Delta_4 H_j = r^{j-2}(j(j-1) + 3j + \Delta_S)h_j .$$

It follows that $\Delta_S h_j = -(j^2 + 2j)h_j$, which we rewrite as

(A.1) $$(\Delta_S - 1)h_j = -(j+1)^2 h_j .$$

The spherical wave equation is defined to be

(A.2) $$u_{tt} - (\Delta_S - 1)u = 0 .$$

We expand solutions u in terms of the eigenfunctions of $\Delta_S - 1$:

(A.3) $$u(\omega, t) = \sum \left(a_j e^{\sqrt{\lambda_j}\, t} + b_j e^{-\sqrt{\lambda_j}\, t}\right) h_j(\omega) ,$$

where h_j are the eigenfunctions of $\Delta_S - 1$, $\lambda_j = -(j+1)^2$ its eigenvalues. The coefficients a_j and b_j are determined by the Cauchy data of u:

$$u(\omega, 0) = \sum (a_j + b_j)h_j ,$$
$$u_t(\omega, 0) = i \sum (a_j - b_j)(j+1)h_j .$$

Replace ω by $-\omega$, and set $t = \pi$ in (A.3). Since $r^j h_j(\omega)$ is homogeneous of degree j, $h_j(-\omega) = (-1)^j h_j(\omega)$ and $e^{i(j+1)\pi} = (-1)^{j+1}$; so we get

(A.4) $$\begin{aligned} u(-\omega, \pi) &= \sum \left(a_j(-1)^{j+1} + b_j(-1)^{-(j+1)}\right)(-1)^j h_j(\omega) \\ &= -\sum (a_j + b_j)h_j(\omega) = -u(\omega, 0) . \end{aligned}$$

This shows that the value of u at $-\omega$ and time π is determined by the value of u at the antipodal point ω at time 0. Antipodal points are connected by geodesics of length π; since geodesics are rays for the spherical wave equation, or for that matter on any Riemannian manifold, this is an instance of Huygens' principle.

To obtain Huygens' principle for any time t that is not a multiple of π we argue as follows:

Let t be any number between 0 and π. Take initial data whose support lies in a ball of radius ε around ω_0. Since, as is easy to show, signals propagate with speed ≤ 1 on the sphere, $u(\omega, t)$ is supported in a ball of radius $\varepsilon + t$ around ω_0.

It follows from (A.4) and a similar formula for $u_t(-\omega, \pi)$ that the data u and u_t at time π are supported in a ball of radius ε around the antipode $-\omega_0$. Since signals propagate backward in time with speed ≤ 1, it follows that $u(\omega, t)$ is supported in a ball of radius $\varepsilon + \pi - t$ around $-\omega_0$. The intersection of these two balls is the spherical strip consisting of points ω whose distance to ω_0 is $\leq t + \varepsilon$ but $\geq t - \varepsilon$. Since ε is an arbitrary positive number, it follows that the domain of influence at time t of the point ω_0 is the set of points whose distance from ω_0 is t. This is Huygens' principle for the spherical wave equation.

Huygens' principle holds of course on spheres of any radius. As the radius tends to ∞ the sphere tends to flat Euclidean space, and the spherical wave equation tends to the Euclidean one. So in the limit we obtain Huygens' principle in Euclidean 3-space.

The spherical and Euclidean wave equations appear to be quite different; nevertheless there is a mapping of a spherical cap less than half of S^3 onto a ball in Euclidean space that maps any solution u of the former into a solution fu of the latter. For details see Lax and Phillips and Ørsted.

An entirely analogous result holds for the wave equation in hyperbolic 3-space $\mathbb{H}_3$ defined as

$$u_{tt} - (\Delta_{\mathbb{H}} + 1)u = 0\,.$$

Here the reduction to the Euclidean case is much simpler; see the monograph by Lax and Phillips.

Semenov-Tian-Shansky has discovered a hyperbolic system, with many time variables, associated with any symmetric space, for which Huygens' principle holds. Further study of this system can be found in the papers of Shashahani, Phillips and Shashahani, and Helgason.

References

Helgason, S. *The Radon transform.* Second edition. Progress in Mathematics, 5. Birkhuser Boston, Boston, 1999.

Lax, P. D., and Phillips, R. S. An example of Huygens' principle. *Comm. Pure Appl. Math.* 31(4): 415–421, 1978.

Lax, P. D., and Phillips, R. S. *Scattering theory for automorphic functions.* Annals of Mathematics Studies, 87. Princeton, Princeton, N.J., 1976.

Ørsted, B. The conformal invariance of Huygens' principle. *J. Differential Geom.* 16(1): 1–9, 1981.

Phillips, R. S., and Shahshahani, M. M. Scattering theory for symmetric spaces of noncompact type. *Duke Math. J.* 72(1): 1–29, 1993.

Semenov-Tian-Shansky, M. Harmonic analysis on Riemannian symmetric spaces of negative curvature and scattering theory. *Math. USSR-Izv.* 10: 535–563, 1976.

Shahshahani, M. Invariant hyperbolic systems on symmetric spaces. *Differential geometry (College Park, Md., 1981/1982)*, 203–233. Progress in Mathematics, 32. Birkhäuser, Boston, 1983.

APPENDIX B

Hyperbolic Polynomials

We recall from Chapter 2 that a polynomial $p(\tau, \xi_1, \dots, \xi_k)$ is called *hyperbolic* in the τ direction if for all real choices of ξ the roots τ of the equation $p(\tau, \xi) = 0$ are real. In this appendix we look at hyperbolic polynomials $p(\tau, \xi, \eta)$ in three variables and homogeneous of degree n.

The prototype of such a polynomial, with $n = 2$, comes from the wave equation: $\tau^2 - \xi^2 - \eta^2$. It can be represented as a determinant:

(B.1) $$\tau^2 - \xi^2 - \eta^2 = \det \begin{pmatrix} \tau - \xi & \eta \\ \eta & \tau + \xi \end{pmatrix}.$$

In 1958 I surmised the following generalization of (B.1):

CONJECTURE Every homogeneous monic polynomial in three variables that is hyperbolic can be represented as a determinant:

(B.2) $$p(\tau, \xi, \eta) = \det(\tau I + \xi A + \eta B),$$

where A and B are real, symmetric matrices. *Monic* means that the coefficient of τ^n is 1.

Clearly every p of form (B.2) is homogeneous, monic, and hyperbolic. The heuristic argument for the converse is as follows:

If we subject the matrix, whose determinant appears on the right in (B.2), to an orthogonal transformation, that is, replace it by

(B.3) $$O(\tau I + \xi A + \eta B)O^{\mathrm{T}}$$

where O is an orthogonal matrix, we obtain

(B.3′) $$\tau I + \xi A' + \eta B',$$

where $A' = OAO^{\mathrm{T}}$ and $B' = OBO^{\mathrm{T}}$. Clearly, the determinant of (B.3) is equal to (B.2). Therefore, in seeking a representation of p in form (B.2) we might as well take A to be diagonal:

(B.2′) $$p(\tau, \xi, \eta) = \det(\tau I + \xi D + \eta B),$$

where D is a real, diagonal matrix.

The number of parameters on the right side of (B.2′) is n from D and $n(n + 1)/2$ from B, altogether $(n^2 + 3n)/2$. The number of powers of degree n in three variables is $\binom{n+2}{2} = (n + 2)(n + 1)/2$; therefore the number of homogeneous monic polynomials contains $(n^2 + 3n)/2$ variables, the same as on the right side of equation (B.2′).

Recently A. S. Lewis, Parrilo, and Ramana have succeeded in deducing this conjecture from a study of Helton and Vinnikov, based on a deep result of Vinnikov.

We now present an application of the determinantal representation (B.2) of hyperbolic polynomials. In Chapter 4, Section 4.5, we gave a derivation of energy inequalities for solutions of hyperbolic equations of order n with the aid of another hyperbolic operator whose characteristics separate those of the given hyperbolic operator. In terms of the associate symbol, a hyperbolic polynomial p of degree n, we seek another hyperbolic polynomial q of degree $n-1$ such that for all real ξ, the roots in τ of $q(\tau, \xi) = 0$ separate the roots of $p(\tau, \xi) = 0$.

We need the following result from the spectral theory of symmetric matrices:

LEMMA B.1 *Let M denote a real symmetric $n \times n$ matrix, and P a projection of $\mathbb{R}^{n-1}$ into $\mathbb{R}^n$; that is, $P^*P = I$, the identity map $\mathbb{R}^{n-1} \to \mathbb{R}^{n-1}$.*

CLAIM *The eigenvalues of P^*MP separate the eigenvalues of M.*

EXERCISE Use the variational characterization of the eigenvalues of a symmetric matrix to prove the lemma.

We apply the lemma to

$$M = \tau I + \xi A + \eta B \tag{B.4}$$

and define q as

$$q(\tau, \xi, \eta) = \det P^* M(\tau, \xi, \eta) P \,. \tag{B.5}$$

According to the lemma, the roots of q separate those of p.

The set of monic polynomials q of degree $n-1$ whose roots separate those of p form a convex set. It is tempting to conjecture that the convex hull of the polynomials q of form (B.5) is the set of all monic polynomials whose roots separate those of p. It is true for p given by (B.1).

References

Helton, J. W., and Vinnikov, V. Linear matrix inequality representation of sets. *Comm. Pure Appl. Math.*, to appear.

Lax, P. D. Differential equations, difference equations and matrix theory. *Comm. Pure Appl. Math.* 11: 175–194, 1958.

Lewis, A. S., Parrilo, P. A., and Ramana, M. V. The Lax conjecture is true. *Proc. Amer. Math. Soc.* 133(9): 2495–2499, 2005 (electronic).

Vinnikov, V. Selfadjoint determinantal representations of real plane curves. *Math. Ann.* 296(3): 453–479, 1993.

APPENDIX C

The Multiplicity of Eigenvalues

Strict hyperbolicity demands that the roots of the characteristic equation be real and distinct. Here we shall investigate symmetric hyperbolic systems of first order in three space variables, i.e., of the form

$$u_t + Au_x + Bu_y + Cu_z = 0, \tag{C.1}$$

where A, B, C are real, symmetric matrices of order $n \times n$.

THEOREM C.1 *If $n \equiv 2 \pmod 4$, a system of form* (C.1) *is not strictly hyperbolic; that is, there exist three real numbers ξ, η, ζ, $\xi^2 + \eta^2 + \zeta^2 = 1$, such that $\xi A + \eta B + \zeta C$ has a multiple eigenvalue.*

PROOF: Denote by $\mathcal{N}$ the set of $n \times n$ real symmetric matrices whose eigenvalues are distinct; this is an open set in the space of all real symmetric matrices. Every matrix N in $\mathcal{N}$ has distinct real eigenvalues that can be arranged in increasing order:

$$\lambda_1 < \lambda_2 < \cdots < \lambda_n .$$

We denote the corresponding real eigenvectors by $r_1, \ldots, r_n$, normalized so that $|r_j| = 1$; these eigenvectors are only determined up to a factor ± 1.

Let $N(\theta)$, $0 \le \theta \le 2\pi$, be a closed curve in $\mathcal{N}$. If we fix $r_j(0)$, the $r_j(\theta)$ is uniquely determined if we require it to depend continuously on θ. Since $N(2\pi) = N(0)$,

$$r_j(2\pi) = \tau_j r_j(0), \quad |\tau_j| = 1. \tag{C.2}$$

Clearly

(i) each τ_j is a homotopy invariant of the closed curve,
(ii) for a constant curve $N(\theta) = \text{const}$ each $\tau_j = 1$.

Combining (i) and (ii) we conclude that

(iii) if $N(\theta)$ is homotopic to a point, then each $\tau_j = 1$.

Consider a curve that is odd, i.e., it satisfies

$$N(\theta + \pi) = -N(\theta). \tag{C.3}$$

It follows that $\lambda_j(\theta + \pi) = -\lambda_{n-j+1}(\theta)$, and

$$r_j(\theta + \pi) = \mu_j r_{n-j+1}(\theta), \quad |\mu_j| = 1. \tag{C.4}$$

Using (C.4) for $\theta = \pi$ and 0 gives

$$r_j(2\pi) = \mu_j r_{n-j+1}(\pi) = \mu_j \mu_{n-j+1} r_j(0).$$

Comparing this with (C.2) we conclude that

$$\tau_j = \mu_j \mu_{n-j+1} . \tag{C.5}$$

For each θ, the eigenvectors

$$r_1(\theta), \dots, r_n(\theta)$$

form an ordered base. Since they depend continuously on θ, the orientation of this ordered base is the same for all θ; in particular, it is the same for $\theta = 0$ and $\theta = \pi$. Using (C.4) with $\theta = 0$, we get

$$\{r_1(\pi), \dots, r_n(\pi)\} = \{\mu_1 r_n(0), \dots, \mu_n r_1(0)\} .$$

In other words, the ordered base at π is obtained from the ordered base at 0 by reversing the order of the base vector and multiplying the j^{th} vector by μ_j. Interchanging the order amounts to $n/2$ transpositions. Since we have assumed $n \equiv 2 \pmod 4$, these would reverse the orientation; to preserve the orientation the product of μ_j must $= -1$:

$$\prod_1^n \mu_j = -1 . \tag{C.6}$$

We regroup the factors and write

$$-1 = \prod_1^n \mu_j = (\mu_1 \mu_n)(\mu_2 \mu_{n-1}) \cdots = \prod_1^{n/2} \tau_j ,$$

where in the last step we made use of (C.5). It follows that at least one of the $\tau_j = -1$; so we conclude from proposition (iii) that *an odd curve $N(\theta)$ is not homotopic to a point.*

The curve

$$N(\theta) = \cos\theta A + \sin\theta B \tag{C.7}$$

is odd; it is the equator of the unit sphere $\xi^2 + \eta^2 + \zeta^2 = 1$. If for all values of (ξ, η, ζ) the matrices $\xi A + \eta B + \zeta C$ had distinct eigenvalues, we could deform the equator on the sphere to a point. This contradicts the previously established fact that $N(\theta)$ given by (C.7) cannot be deformed to a point. □

Friedland, Robbin, and Sylvester extended this result and showed that the conclusion of the theorem holds for all $n \equiv 2, \dots, 6 \pmod 8$ but not for $n \equiv 0, \pm 1 \pmod 8$.

The condition of symmetry of the matrices A, B, C can be replaced by the requirement that all linear combinations $\xi A + \eta B + \zeta C$ have real eigenvalues.

Eigenvalues that have a constant multiplicity for all ξ, η, ζ do not affect the behavior of solutions. But a change in multiplicity is more than a technical issue; it alters the way singularities of solutions propagate. This is strikingly demonstrated in crystal optics, where rays of light propagating in special directions are refracted into a cone.

Fritz John obtained interesting results on the persistence of the multiplicity in eigenvalues in hyperbolic systems of second-order equations.

References

Friedland, S., Robbin, J. W., and Sylvester, J. H. On the crossing rule. *Comm. Pure Appl. Math.* 37(1): 19–37, 1984.

John, F. Restrictions on the coefficients of hyperbolic systems of partial differential equations. *Proc. Nat. Acad. Sci. U.S.A.* 74(10): 4150–4151, 1977.

Lax, P. D. The multiplicity of eigenvalues. *Bull. Amer. Math. Soc. (N.S.)* 6(2): 213–214, 1982.

APPENDIX D

Mixed Initial and Boundary Value Problems

We shall study mixed initial and boundary value problems for first-order symmetric hyperbolic systems of the form

(D.1) $$D_t u = Gu = \sum A_j D_j u + Bu\,,$$

A_j symmetric matrices of order $n \times n$ that are C^∞ functions of x and t, B not necessarily symmetric. The main tools for studying the initial value problem for such systems are the a priori estimate solutions in the H_0 and H_j norms. The basic estimate in the H_0 norm is obtained (see Section 4.3) by taking the scalar product of (D.1) with u and integrating with respect to x over $\mathbb{R}^k$. The left side can be written as

(D.2) $$D_t \int \frac{1}{2} u \cdot u\, dx\,;$$

integration by parts turns the right side $\int u \cdot Gu\, dx$ into $\int G^* u \cdot u\, dx$. Taking their average shows that (D.2) is equal to

(D.3) $$\frac{1}{2}\int \Big(B + B^* - \sum A_{j,j}\Big) u \cdot u\, dx\,, \quad \text{where } A_{j,j} = D_j A_j\,.$$

Recall that the change of variables $u = e^{\lambda t} v$ results in an equation for v similar to (D.1); the only difference is that the coefficient B is diminished by λI. For λ large enough the quadratic form in the integrand in (D.3) is negative; combined with (D.2) it leads to the conclusion that the L^2 norm of v is a decreasing function of t. In what follows we assume that such a transformation has already been performed, so that the integrand in (D.3) is negative.

In this appendix we shall investigate mixed initial boundary value problems, where initial values are prescribed on a domain S in x space, and linear boundary conditions are imposed on ∂S for all time ≥ 0. The boundary condition at a boundary point x restricts $u(x,t)$ to belong to a subspace U of $\mathbb{R}^n$; this subspace may vary smoothly from point to point. We are looking for boundary conditions that make the L^2 norm of u over S a decreasing function of time. To derive such conditions we proceed as before: take the scalar product of (D.1) with u, integrate over Ω, integrate by parts on the right, and average. We get

(D.4) $$D_t \int_S \frac{1}{2} u \cdot u\, dx = \frac{1}{2}\int_S Q(u,u)dx + \int_{\partial S} u \cdot A_\nu u\, dS\,,$$

where ν is the unit outward normal, and A_ν is the normal component of the $A_j - s$:

$$A_\nu = \sum A_j \nu_j \,. \tag{D.5}$$

Q is the quadratic form in (D.3). Since we made $Q \leq 0$, all that remains is to require that the boundary integral on the right in (D.4) be ≤ 0. That will be the case if the integrand in the boundary integral is ≤ 0. This leads to the following condition on the boundary space U:

$$u \cdot A_\nu u \leq 0 \quad \text{for all } u \in U \,. \tag{D.6}$$

We must avoid imposing too many boundary conditions. The number of boundary conditions imposed is the codimension of the subspace U. Therefore we require U to be maximal regarding property (D.6):

(D.7) There is no subspace U' properly containing U such that
$$u' \cdot A_\nu u' \leq 0 \quad \text{for all } u' \in U' \,.$$

For technical reasons we require that $A_\nu(x)$ be invertible at every point x of $\partial\Omega$.

THEOREM D.1 *The dimension of a boundary space U satisfying* (D.6) *and* (D.7) *is equal to the number of negative eigenvalues of A_ν.*

PROOF: If $\dim U$ were greater than the number of negative eigenvalues, there would be a nonzero vector v in U that is orthogonal to all eigenvectors corresponding to negative eigenvalues. Since zero is not an eigenvalue of A_ν, $v \cdot A_\nu v$ would be positive.

Conversely, if $\dim U$ were less than the number of negative eigenvalues, there would exist a nonzero vector w, a linear combination of eigenvectors corresponding to negative eigenvalues of A_ν that satisfies the following condition:

$$w \cdot A_\nu u = 0 \quad \text{for all } u \in U \,. \tag{D.8}$$

This vector w does not belong to U, for, if it did, it would satisfy $w \cdot A_\nu w = 0$. Since w is a linear combination of eigenvectors with negative eigenvalues, $w \cdot A_\nu w < 0$.

Now we define U' to consist of all vectors of the form $u + cw$, $u \in U$, U'properly contains U. Using (D.8), we get

$$u' \cdot A_\nu u' = (u + cw) \cdot A_\nu(u + cw) = u \cdot A_\nu u + c^2 w \cdot A_\nu w \leq 0 \,,$$

since both terms are ≤ 0. This contradicts requirement (D.7) of maximality, and thereby proves Theorem D.1. □

The adjoint of equation (D.1) is

$$D_t v = -G^* v \,, \tag{D.9}$$

where G^* is the adjoint of G. We define the adjoint of the boundary condition $u \in U$ to be $v \perp A_\nu U$; that is, the adjoint boundary space V is the orthogonal complement of $A_\nu U$.

The significance of adjoint boundary conditions lies in this: If u is a solution of (D.1) and v of the adjoint equation (D.9), and if u and v satisfy adjoint boundary conditions, then

$$\int u(x,t)\cdot v(x,t)dx \tag{D.10}$$

is independent of t.

PROOF: Differentiate (D.10) with respect to t, express the t derivatives of u and v as Gu and $-G^*$, and integrate by parts. □

THEOREM D.2 *Suppose the boundary condition $u \in U$ for solutions of equation* (D.1) *satisfies requirements* (D.6) *and* (D.7); *then the adjoint boundary condition $v \in V$, $V = (A_\nu U)^\perp$ satisfies*

$$v\cdot A_\nu v \geq 0 \quad \text{for all } v \in V\,, \tag{D.11}$$

and V is maximal with respect to this property.

PROOF: Suppose, on the contrary, that for some $v \in V$, $v \cdot A_\nu v < 0$. Such a v does not belong to U, since $v \cdot A_\nu u = 0$. We can then enlarge the space U by adjoining v to it. The elements $u + cv$ of this enlarged space satisfy

$$(u+cv)\cdot A_\nu(u+cv) = u\cdot A_\nu u + 2cv_1 A_\nu u + c^2 v\cdot A_\nu v \leq 0\,,$$

because the first term on the right is ≤ 0, the second term $= 0$, and the third term is negative except when $c = 0$. But since U is assumed to be maximal, such an extension is not possible; this proves (D.11).

To show that V is maximal, we appeal to Theorem D.1, which says that $\dim U =$ the number of negative eigenvalues of A_ν. Since A_ν is invertible, $A_\nu U$ has the same dimension; the orthogonal complement V of $A_\nu U$ has the complementary dimension $n - \dim A_\nu$. Since 0 is not an eigenvalue of A_ν, $\dim V =$ the number of positive eigenvalues of A_ν. Analogously to Theorem D.1, the dimension of a space of vectors that satisfies (D.11) and is maximal with respect to this property equals the number of positive eigenvalues of A_ν. Since this is the dimension of V, it follows that V cannot be enlarged and still retain property (D.11). □

Denote by I the time interval $I = [0, T]$. Let u and v be once-differentiable functions in the cylinder $S \times I$ that are 0 at $t = 0$ and $t = T$, respectively: $u(x,0) = 0$ and $v(x,T) = 0$. Suppose that u satisfies linear boundary conditions on ∂S and v the adjoint boundary conditions. Denote $D_t - Gu = f$ and $D_t v + G^* v = g$. Then integration by parts shows that

$$\iint_{S\times I} (f\cdot v - u\cdot g)dx\,dt = 0\,. \tag{D.12}$$

DEFINITION D.3 Let u and f be L^2 functions in $S \times I$; u is defined to be a *weak solution* of

$$D_t - Gu = f\,,\quad u(x,0) = 0,,\quad \text{and}\quad u(x,t) \in U(x)\quad \text{for } x \in \partial S\,, \tag{D.13}$$

if (D.12) holds for all once-differentiable functions v that are 0 at $t = T$, that satisfy the adjoint boundary conditions, and where $g = D_t v + G^* v$.

DEFINITION D.4 Let u and f be L^2 functions in $S \times I$. U is called a *strong solution* of (D.13) if u is the limit in the L^2 norm of a sequence of functions $\{u_k\}$ that have square integrable first derivatives, satisfy $u_k(x, 0) = 0$ and the boundary conditions, and $D_t u_k - G u_k = f_k$ tends to f in the L^2 norm.

NOTES.

(1) A function u_k that has square integrable first derivatives has initial and boundary values that belong to L^2.

(2) A strong solution is a weak solution.

The main existence theorem for mixed initial boundary value problems is the following:

THEOREM D.5 *Let f be an L^2 function in $S \times I$.*

(i) *The initial boundary value problem* (D.13) *has a weak solution in $S \times I$, provided that the boundary space satisfies conditions* (D.6) *and* (D.7).

(ii) *This weak solution is a strong solution.*

This result is due to Friedrichs; a somewhat simpler proof was given by Ralph Phillips and the author.

The boundary conditions (D.6), (D.7) are not the only ones for which the mixed initial boundary value problem has a unique solution. A general theory has been developed, independently, by Kreiss and Sakamoto.

Taylor and Melrose have shown, independently, that singularities of solutions of mixed initial boundary value problems propagate along rays, including rays reflected from the boundary.

References

Friedrichs, K. O. Symmetric positive linear differential equations. *Comm. Pure Appl. Math.* 11: 333–418, 1958.

Kreiss, H. Initial boundary value problems for hyperbolic systems. *Comm. Pure Appl. Math.* 23: 277–298, 1970.

Lax, P. D., and Phillips, R. S. Local boundary conditions for dissipative symmetric linear differential operators. *Comm. Pure Appl. Math.* 13: 427–455, 1960.

Melrose, R. B. *Geometric scattering theory*. Cambridge University Press, Cambridge, 1995.

Sakamoto, R. Mixed problems for hyperbolic equations I, II. *J. Math. Kyoto Univ.* 10: 349–373, 403–417, 1970.

Taylor, M. Reflection of singularities of solutions of systems of differential equations. *Comm. Pure Appl. Math.* 28: 457–478, 1975.

———. Grazing rays and reflection of singularities to wave equations. *Comm. Pure Appl. Math.* 29: 1–38, 1978.

APPENDIX E

Energy Decay for Star-Shaped Obstacles

by Cathleen S. Morawetz

The standard energy conservation law for U is found by multiplying the wave equation by U_T and noting that the resulting quadratic expression is a divergence:

$$U_T(U_{TT} - \Delta_X U) = \operatorname{div}_X P + Q_T \tag{E.1}$$

where

$$P = -U_T \nabla U, \quad Q = \frac{1}{2}\left(U_T^2 + |\nabla U|^2\right). \tag{E.2}$$

Integrated over any region $\mathcal{D}$ this expression therefore yields a surface integral in (X, T) space which vanishes whenever U is a solution of the wave equation; this is called the standard energy identity. It has the additional property that the integrand is a positive definite form on spacelike surfaces.

As is well known, the Kelvin transformation

$$X = \frac{x}{r^2 - t^2}, \quad T = \frac{t}{r^2 - t^2}, \quad RU = ru, \quad R = |X|, \quad r = |x|, \tag{E.3}$$

preserves the wave operator in the sense that

$$R^3(U_{TT} - \Delta_X U) = r^3(u_{tt} - \Delta_x u). \tag{E.4}$$

On the other hand,

$$RU_T = r[(r^2 + t^2)u_t + 2t(ru)_r] \tag{E.5}$$

and

$$\frac{dX\,dT}{R^4} = \frac{dx\,dt}{r^4} \tag{E.6}$$

Combining (E.4), (E.5), and (E.6) we get

$$\int U_T(U_{TT} - \Delta_X U)dX\,dT = \int Nu(u_{tt} - \Delta_x u)dx\,dt \tag{E.7}$$

with $Nu = (r^2 + t^2)u_t + 2t(ru)_r$.

Using (E.1), the left-hand side of (E.7) can be written as a surface integral and therefore so can the right side. Thus one obtains:

THEOREM E.1 *Suppose $u(x,t)$ is a solution of the wave equation that has square integrable derivatives. Then over any three-dimensional surface ∂ with the surface element dS,*

$$\int_{\partial} (pn + qn_t)dS = 0 \tag{E.8}$$

where n is the space component of the outward normal, n_t is the time component.

A tedious calculation gives

$$\begin{aligned} p = &-tu_t^2x - 2t(x\nabla u)\nabla u + t|\nabla u|^2 x - (r^2+t^2)u_t\nabla u - 2tu\nabla u \\ &- \frac{1}{2}r^{-2}((r^2+t^2)u^2)_t\, x \end{aligned} \tag{E.9}$$

$$\begin{aligned} q = &\, 2t(x\nabla u)u_t + \frac{1}{2}(r^2+t^2)\big(|\nabla u|^2 + u_t^2\big) + 2tuu_t \\ &+ r^{-2}(r^2+t^2)\left(u\nabla u x + \frac{1}{2}u^2\right). \end{aligned} \tag{E.10}$$

q is a definite form; in fact, q may be written as:

$$\begin{aligned} q = &\frac{1}{2}(r^2+t^2)\big(|\nabla u|^2 - u_r^2\big) \\ &+ \frac{1}{4r^2}\left\{(r+t)^2((ru)_r + (ru)_t)^2 + (r-t)^2((ru)_r - (ru)_t)^2\right\}. \end{aligned} \tag{E.11}$$

The positivity of q can also be deduced as follows: Under the Kelvin transformation, the inverse of (E.3), the surface t = constant is transformed into a spacelike surface in the (X, T) space. On this spacelike surface the integrand in the standard energy identity is a definite form. Hence on the transform of this surface, i.e., t = constant, the new integrand, q, is also definite.

THEOREM E.2 *Let u be a solution of the wave equation outside a star shaped body with boundary B and assume that $u = 0$ on B. Suppose further that the initial data f of u is zero for $|x| \geq k$. Then*

$$|u(\tau)|_h \leq \frac{2k}{\tau}|f| \tag{E.12}$$

for $\tau \geq 2h$. Here $|u(\tau)|_h^2$ is the energy, $\int \big(|\nabla u| + u_t^2\big)dx$, inside a sphere of radius h at time τ and $|f|^2$ is the total energy of the initial data.

PROOF: Choose the origin so that B is star shaped with respect to the origin, i.e., $xn \leq 0$, where n is the inward normal to B. We apply Theorem E.1 to a domain bounded by the planes $t = \tau$, $t = 0$, and the body cylinder $x \in B, 0 \leq t \leq \tau$. Then since the solution vanishes for r large enough,

$$\int_{t=\tau} q\,dx + \int_0^{\tau}\int_B pn\,ds\,dt = \int_{t=0} q\,dx. \tag{E.13}$$

Since u vanishes on B, $\nabla u = (\partial u/\partial n)n$ there and $u_t = 0$; thus from (E.9), it follows that $pn = -t(\partial u/\partial n)^2 xn$. Since B is star-shaped with respect to the origin, $xn \le 0$; thus $pn \ge 0$. Hence from (E.13),

$$\int_{t=\tau} q\,dx \le \int_{t=0} q\,dx\,. \tag{E.14}$$

From the expression (E.11) for q, we see that for $t = 0$, $q \le \frac{1}{2}r^2(|\nabla u|^2 + u_t^2)$. Therefore, since f has support in $|x| \le k$, we find

$$\int_{t=0} q\,dx \le \frac{1}{2}kk^2|f|^2\,. \tag{E.15}$$

Since the integrand q is positive, we get from (E.14) and (E.15) for any h,

$$\int_{\substack{r\le h\\ t=\tau}} q\,dx \le \int_{t=\tau} q\,dx \le \frac{1}{2}k^2|f|^2\,. \tag{E.16}$$

Using the expression (E.11) we can bound q from below for $r \le t/2$:

$$\frac{1}{4}t^2\left[\frac{1}{2}(|\nabla u|^2 - u_r^2) + \frac{1}{4r^2}((ru)_r + (ru)_t)^2 + \frac{1}{4r^2}((ru)_r - (ru)_t)^2\right] \le q \tag{E.17}$$

or

$$\frac{1}{8}t^2\left(|\nabla u|^2 + u_t^2 + \operatorname{div}\frac{1}{r^2}u^2x\right) \le q\,. \tag{E.18}$$

Using (E.18) in (E.16), we get for $r \ge 2h$

$$\int_{\substack{t=\tau\\ r\le h}} \left(|\nabla u|^2 + u_t^2 + \operatorname{div}\frac{1}{r^2}u^2x\right)dx \le \frac{4}{r^2}k^2|f|^2; \tag{E.19}$$

since $u = 0$ on B, integrating the divergence gives

$$\int_{\substack{t=\tau\\ r\le h}} (|\nabla u|^2 + u_t^2)dx + \int_{\substack{t=\tau\\ r=h}} \frac{1}{r}u^2\,dS \le \frac{4}{r^2}k^2|f|^2 \tag{E.20}$$

where dS is the surface element on the sphere $r = h$. Hence,

$$\int_{\substack{t=\tau\\ r\le h}} (|\nabla u|^2 + u_t^2)dx \le \frac{4}{r^2}k^2|f|^2, \tag{E.21}$$

which concludes the proof. □

Theorem E.2 shows that for solutions whose initial data have compact support, local energy decays as τ tends to ∞. Since solutions whose initial data have compact support are dense among all solutions with finite energy, it follows that local energy decays for all solutions with finite energy. This gives a much simpler proof of Lemma 9.12 for star-shaped obstacles than the one presented in Chapter 9.

Titles in This Series

Volume

14 **Peter D. Lax**
Hyperbolic partial differential equations
2006

13 **Oliver Bühler**
A brief introduction to classical, statistical, and quantum mechanics
2006

12 **Jürgen Moser and Eduard J. Zehnder**
Notes on dynamical systems
2005

11 **V. S. Varadarajan**
Supersymmetry for mathematicians: An introduction
2004

10 **Thierry Cazenave**
Semilinear Schrödinger equations
2003

9 **Andrew Majda**
Introduction to PDEs and waves for the atmosphere and ocean
2003

8 **Fedor Bogomolov and Tihomir Petrov**
Algebraic curves and one-dimensional fields
2003

7 **S. R. S. Varadhan**
Probability theory
2001

6 **Louis Nirenberg**
Topics in nonlinear functional analysis
2001

5 **Emmanuel Hebey**
Nonlinear analysis on manifolds: Sobolev spaces and inequalities
2000

3 **Percy Deift**
Orthogonal polynomials and random matrices: A Riemann-Hilbert approach
2000

TITLES IN THIS SERIES

2 **Jalal Shatah and Michael Struwe**
Geometric wave equations
2000

1 **Qing Han and Fanghua Lin**
Elliptic partial differential equations
2000

Hyperbolic Partial Differential Equations

PETER D. LAX

The theory of hyperbolic equations is a large subject, and its applications are many: fluid dynamics and aerodynamics, the theory of elasticity, optics, electromagnetic waves, direct and inverse scattering, and the general theory of relativity. This book is an introduction to most facets of the theory and is an ideal text for a second-year graduate course on the subject.

The first part deals with the basic theory: the relation of hyperbolicity to the finite propagation of signals, the concept and role of characteristic surfaces and rays, energy, and energy inequalities. The structure of solutions of equations with constant coefficients is explored with the help of the Fourier and Radon transforms. The existence of solutions of equations with variable coefficients with prescribed initial values is proved using energy inequalities. The propagation of singularities is studied with the help of progressing waves.

The second part describes finite difference approximations of hyperbolic equations, presents a streamlined version of the Lax-Phillips scattering theory, and covers basic concepts and results for hyperbolic systems of conservation laws, an active research area today.

Four brief appendices sketch topics that are important or amusing, such as Huygens' principle and a theory of mixed initial and boundary value problems. A fifth appendix by Cathleen Morawetz describes a nonstandard energy identity and its uses.

For additional information and updates on this book, visit

www.ams.org/bookpages/cln-14

ISBN 0-8218-3576-9

9 780821 835760

CLN/14

NEW YORK UNIVERSITY