STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

ZW 1961 - 012

Voordracht in de serie "Actualiteiten"

Dr. R.T. Seeley

25 november 1961

Uniqueness in the Cauchy Problem

1961

ZW

ZW 1961-012

STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

ZUIVERE WISKUNDE

Voordracht in de serie "Actualiteiten"

door

Dr. R.T. Seeley

25 november 1961

Uniqueness in the Cauchy Problem

Introduction.

This article gives a sketch of the solution of a problem of long standing, whose importance was indicated by Petrowsky in his well-known "Lectures on Partial Differential Equations" at a time when it was still unsolved. The problem is that of the uniqueness of solutions of the Cauchy problem ("initial value problem") in partial differential equations, uniqueness being sought in some class of continuously differentiable functions, not just in the class of analytic functions; uniqueness in this restricted class can be trivially established. A solution was given recently (A.P. Calderon, American Journal of Mathematics, 1958) relying on singular integral operators in several variables. The real aim of the present article is to indicate what these singular integral operators (s.i. ops) are, and how they help in handling partial differential equations, at least from a theoretical point of view. Singular integral operators are discussed first, followed by a sketch of the context of the Cauchy problem, and finally a sketch of the proof of uniqueness. Because of the scope of this undertaking, the most that can be acieved is to make the results seem credible - many details will be left out.

1. Background of singular integral operators in several variables, and relevant theorems.

S.i. ops in several variables arose primarily in potential theory, and this provides a good introduction. Let $x=(x_1,\ldots,x_k)$ and y be points in E_k, $x-y=(x_1-y_1,\ldots)$, $|x|^2=\sum x_j^2$, and

$\Delta=\sum \partial^2/\partial x_j^2$. Assume temporarily that $k>2$, and define $Pf(x)=c\int_{E_k} f(y)\,|x-y|^{2-k}dy$, where the constant c is chosen so that $-\Delta Pf=-P\Delta f=f$. Then $\partial Pf/\partial x_j=c\int f(y)\,\partial|x-y|^{2-k}/\partial x_j dy=(2-k)c\cdot \int f(y)(x_j-y_j)|x-y|^{-k}dy$; the singularity at y=x is still integrable, and the process is easily justified if f is a reasonably good function. A second differentiation, however, leads to a non-integrable singularity, one like $|x-y|^{-k}$, so that differentiating under the integral does not even make sense formally. Adopting the viewpoint of the theory of distributions, however, leads one to

$$1)\quad \partial^2 Pf/\partial x_m\partial x_n = -(\delta_{mn}/k)f(x)+\lim_{\epsilon\to 0}\int_{|x-y|>\epsilon} f(y)\,\partial^2|x-y|^{2-k}/\partial x_m\partial x_n dy = R_{mn}f(x),$$

at least if f is continuously differentiable and vanishes outside a compact set. The operator R_{mn} is a typical example of a s.i.op. It is easy to see how (1) leads to $\Delta Pf=-f$, since $\sum\partial^2|x-y|^{2-k}/\partial x_j^2=0$. If we set $f=-\Delta g$, we get $Pf=g$, and $\partial^2 g/\partial x_m\partial x_n = R_{mn}(-\Delta)g$. Thus $\partial^2/\partial x_m\partial x_n$ is represented by a s.i.op. times the canonical second order operator $-\Delta$. (The minus is used to make the operator positive: $\int(-\Delta f)\bar{f}\geq 0$.)

In order to obtain such a representation for operators of arbitrary order, it is easiest to consider the situation with f in L^2, and use Fourier transforms. Let $f(x)\longrightarrow\hat{f}(x) = \int_{E_k} e^{-ix\cdot y}f(y)dy$. For "$\longrightarrow$" in this context read "transform(s) into". Then it is well known that $\partial f/\partial x_m\longrightarrow ix_m\,\hat{f}(x)$, and consequently $-\Delta f(x)\longrightarrow |x|^2\,\hat{f}(x)$. Since $-P\Delta=$ the identity, we have $Pf(x)\longrightarrow|x|^{-2}\hat{f}(x)$, and $R_{mn}f(x)=\partial^2 Pf/\partial x_m\partial x_n\longrightarrow(ix_m)(ix_n)|x|^{-2}\hat{f}(x)$; i.e. the s.i.op. R_{mn} corresponds to multiplication of the Fourier transform by $(ix_m)(ix_n)|x|^{-2}$. Comparing this, namely that $-(\delta_{mn}/k)f(x)+\lim c\int f(y)\,\partial^2|x-y|^{2-k}/\partial x_m\partial x_n dy\longrightarrow(ix_m)(ix_n)|x|^{-2}\hat{f}(x)$, with the formula for the transform of a convolution $\int f(y)g(x-y)dy\longrightarrow \hat{f}(x)\hat{g}(x)$, it is clear that in some sense the kernel $H(z)=-(\delta_{mn}/k)\delta(z)+c\,\partial^2|z|^{2-k}/\partial z_m\partial z_n$ has the Fourier transform $\hat{H}(z)=(iz_m)(iz_n)|z|^{-2}$. Here $\delta(z)$ is the Dirac "delta function". This transform can in fact be obtained as $\hat{H}(z)=-(\delta_{mn}/k)+\lim_{\epsilon\to 0}\int_{\epsilon<|y|<\epsilon^{-1}} e^{-iy\cdot z}\,H(y)dy$.

Observe that this transform is a) homogeneous of degree 0, and is b) obtained on $|z|=1$ by replacing $\partial/\partial x_m$ by iz_m, i.e. on

$|z|=1$ it coincides with i^2 times the characteristic form of $\partial^2/\partial x_m \partial x_n$.

These transforms suggest how to extend such a representation to differential operators of any order. Define an operator Λ by $\Lambda f(x) \longrightarrow |x| \hat{f}(x)$; thus $\Lambda^2 = -\Delta$ .We define R_m by $R_m f(x) \longrightarrow i x_m |x|^{-1} \hat{f}(x)$. The same operator could be obtained in analogy with the previous R_{mn} by $R_m f = c' \lim_{\epsilon \to 0} \int_{|x-y|>\epsilon} f(y) \partial |x-y|^{1-k} / \partial x_m dy$, and is thus a s.i.op. (To see why the kernel $c' |x-y|^{1-k}$ yields a square root of the potential operator, consult e.g. Riesz: "L'Integrale de Riemann-Liouville" in Acta Math., 1949). For representing differentiation we have $\partial f/\partial x_m = R_m \Lambda f$, and $\partial^j f/\partial x_{m_1} \ldots \partial x_{m_j} = R_{m_1} \ldots R_{m_j} \Lambda^j f = H \Lambda^j f$. We know already that R_m and $R_{mn} = R_m R_n$ are s.i. ops, so it is not surprising that $R_{m_1} \ldots R_{m_j}$ is in general such an operator. In fact we have:

<u>Theorem 1.</u> There is a one-one correspondence between the functions $\varphi(z)$ with $\varphi(\lambda z) = \varphi(z)$ for $\lambda > 0$, $\varphi \in C^\infty$ in $|z| > 1$, and $\int_{|z|=1} \varphi = 0$; and the functions $\psi(z)$ with $\psi(\lambda z) = \lambda^{-k} \psi(z)$ for $\lambda > 0$, ψ in C^∞ in $|z| > 1$, and $\int_{|z|=1} \psi(z) = 0$; given by $\varphi(z) = \lim \int_{\epsilon < |y| < 1/\epsilon} e^{-iz.y} \psi(y) dy$. If Ψ is defined for f in L^2 by $\Psi f(x) \longrightarrow \varphi(x) \hat{f}(x)$, then

$$\Psi f(x) = \underset{\epsilon \to 0}{\text{l.i.m.}} \int_{|x-y|>\epsilon} f(y) \psi(x-y) dy.$$

There is a similar correspondence if we drop the condition of zero average. If φ is homogeneous of degree zero, i.e. $\varphi(\lambda z) = \varphi(z)$ for $\lambda > 0$, let $\varphi(z) = a + \hat{h}(z)$, where $\hat{h}$ has zero average on $|z| = 1$. Then if $\hat{h} \leftarrow h$, one finds that $\varphi \hat{f} \leftarrow af + \underset{\epsilon \to 0}{\text{l.i.m.}} \int_{|x-y|>\epsilon} h(x-y) f(y) dy$.

The significance of the mean value zero is easy to see: by virtue of this, $\int_{\epsilon < |x-y| < 1} h(x-y) dy = 0$, so we can write, for differentiable functions in L^2,

$$\text{l.i.m.} \int_{|x-y|>\epsilon} h(x-y) f(y) dy = \int_{|x-y|>1} h(x-y) f(y) dy + \int_{|x-y|<1} h(x-y) [f(y) - f(x)] dy.$$

The use of Fourier transforms yields immediately that, if $|a + \hat{h}| \geq \epsilon > 0$, then the operator H defined by $Hf(x) = af(x) + \text{l.i.m.} \int_{|x-y|>\epsilon} f(y) h(x-y) dy$ has a bounded inverse, whose norm is in fact $\leq 1/\epsilon$.

The operators considered so far might be called pure convolution s.i.ops. It is easy to see that their application in partial differential equations is restricted to equations with constant coefficients. However, an operator like $x_1\,\partial^2 f/\partial x_1^{\,2}+x_2\partial^2 f/\partial x_1\partial x_2$ can be represented by

$$[x_1R_{11}+x_2R_{12}]\Lambda^2 f(x) = k^{-1}x_1\Lambda^2 f(x)$$
$$+ \text{l.i.m.}\; c\int_{|x-y|>\epsilon}\Lambda^2 f(y)\left[x_1\partial^2/\partial x_1^{\,2}+x_2\partial^2/\partial x_1\partial x_2\right]|x-y|^{\,2-k}dy.$$

This leads us to the notion: a C_n^∞ singular integral operator is any of the form

$$(2)\qquad Hf(x) = a(x)\,f(x) + \underset{\epsilon\to 0}{\text{l.i.m.}}\int_{|x-y|>\epsilon} h(x,x-y)f(y)dy,$$

where a is in C_n, $h(x,\lambda z)=\lambda^{-k}h(x,z)$ for $\lambda>0$, and in $|z|>1$ h is in C_n with respect to x and C^∞ with respect to z, and $\int_{|z|=1} h(x,z) = 0.$

In applications to differential operators, the C^∞ is automatically satisfied, and the C_n refers to the differentiability of the coefficients of the equation.

The action of H can no longer be described completely by a Fourier transform, but there is a partial substitute which has come to be called the symbol: the symbol of the operator H in (2) is $\sigma(H) = a(x) + \hat{h}(x,z)$, where $\hat{h}(x,z) = \lim\int_{\epsilon<|y|<\epsilon^{-1}}e^{-iy.z}h(x,y)dy$ is the Fourier transform of $h(x,z)$ with respect to z.

Theorem 2. Every C_n^∞ function $F(x,z)$ with $F(x,\lambda z)=F(x,z)$ for $\lambda>0$, is the symbol of a unique C_n^∞ operator H: $F=\sigma(H)$. If N_H is the sup for $|z|=1$ of $\sigma(H)$ and its derivatives of order 2k with respect to z, then $\|H\|\le C_k N_H$; C_k is independent of H.

(The choice of the derivatives of order 2k is to give an estimate of the rate of convergence of the spherical harmonic series for h.)

On the basis of this theorem we define a special product between s.i.ops by $\sigma(H_1\circ H_2) = \sigma(H_1)\,\sigma(H_2)$. The point of this is, on the one hand, that $H_1\circ H_2 = H_2\circ H_1$, while $H_1\circ H_2$ is a fairly good approximation to H_1H_2, that is H_2 followed by H_1. The sense in which this holds is part of

Theorem 3. Let H_1 and H_2 be C_2^∞ s.i.ops. Then $H_1\Lambda-\Lambda H_1$ and $(H_1\circ H_2-H_1H_2)\Lambda$ are bounded operators on L^2.

The point here is that Λ is a canonical first order differential operator, and unbounded; but the difference between the unbounded operators $H_1\Lambda$ and ΛH_1 is bounded; while $H_1 \circ H_2 - H_1H_2$ is more than bounded, it is such a good operator that it removes the the unboundedness of Λ. Such an operator can be called "smoothing", and will often be denoted by S: thus $HG = H \circ G + S$.

In order to describe the Cauchy problem, the following symbolism is convenient. For $x=(x_0,\ldots,x_k)$ a point in k+1 space and $\alpha=(\alpha_0,\ldots,\alpha_k)$ a k+1 tuple of non-negative integers, set $|\alpha|=\alpha_0+\ldots+\alpha_k$, $x^\alpha = x_0^{\alpha_0}\ldots x_k^{\alpha_k}$, $(\partial/\partial x)^\alpha = (\partial/\partial x_1)^{\alpha_1}\ldots$

$$= \partial^{|\alpha|}/\partial x_1^{\alpha_1}\ldots\partial x_k^{\alpha_k}.$$

A linear partial differential equation of order N can then be written $Lu = \sum_{0\leq|\alpha|\leq N} a_\alpha(x)(\partial/\partial x)^\alpha u = f(x)$. The characteristic form of L is by definition $\sum_{|\alpha|=N} a_\alpha(x)z^\alpha$, i.e. the highest order derivative symbols in L are replaced by new variables $z=(z_0,\ldots,z_k)$. In connection with s.i.ops we have

Theorem 4. A differential operator $L=\sum_{|\alpha|=N} a_\alpha(x)(\partial/\partial x)^\alpha$, with a_α in C_N, can be represented as $L=H(i\Lambda)^N$, where H is the C_N^∞ s.i.op. whose symbol coincides on $|z|=1$ with the characteristic form of L.

Theorems 1 to 4 give the facts for s.i.ops. that are used in the uniqueness theorem.

2. The Cauchy problem, and a brief history of the uniqueness question.

Now the Cauchy problem for a linear equation can be easily stated. Given

i) an equation $\sum_{|\alpha|\leq N} a_\alpha(x)(\partial/\partial x)^\alpha u(x)=f(x)$

ii) a surface Σ: $\sigma(x) = 0$

iii) initial conditions: U and all its derivatives of order $<N$ are specified on Σ;

given this, find U.

Σ is called non-characteristic if this problem can be reduced to the following standard form in new variables $(t,y_1,\ldots,y_k)$:

i) $\partial\ U/\partial t^n + \sum_0^{N-1} \sum_{|\beta|+n\leq N} a_{n,\beta}(t,y)(\partial/\partial t)^n(\partial/\partial y)^\beta U$

$$+ F(t,y) = 0 \qquad (\ = (\ _1,\ldots,\ _k)),$$

or $\partial^N U/\partial t^N + \sum_{j=1}^{N} A_j(t,y)\partial^{N-j}U/\partial t^{N-j} + F=0$, with A_j a differential operator in the y variables of order j.

ii) S is t=0

iii) $\partial^n U/\partial t^n$ is specified for n=0,1,...,N-1, and t=0.

The uniqueness of solutions of CR in the class of analytic functions is easy to see. For suppose all the data in CR are analytic; then all the derivatives of U on Σ can be calculated inductively from (iii), (i), and repeated differentiation of these equations. On the other hand, if the original equation cannot be solved for $\partial^N U/\partial t^N$ in this way, it is easy to believe that $\partial^N U/\partial t^N$ might not be determined.

Thus the question of uniqueness is posed for data given on a non-characteristic surface by the given data.

It is easy to get the criterion for reducibility of C to CR. Suppose, to this end, that new coordinates (t,y) are introduced so that S becomes t=0. Then condition CR (iii) can be derived immediately from C (iii), and the only problem is to show that the coefficient of $\partial^N U/\partial t^N$ in the new version of C (i) does not vanish. This coefficient is $\sum_{|\alpha|=N} a_\alpha(x)(\nabla t)^\alpha$, where $t=(\partial t/\partial x_0,\ldots,\partial t/\partial x_k)$, and thus depends only on the normal to t=0, i.e. the normal to Σ. Thus the condition for reducibility is $0\neq\sum_{|\alpha|=N} a_\alpha(x)(\nabla\sigma)^\alpha$. Since reducibility depends only on the normal to the surface, it is easy to see that the form to which C is reduced need not give the data on t=0, but could just as well give them on $t=|y|^2$, for instance. We will, in the proof, suppose the data to be given on such a paraboloid.

Definition. A direction given by $z=(z_0,\ldots,z_k)\neq 0$ is characteristic for (c,i) if and only if $\sum_{|\alpha|=N} a_\alpha(x)z^\alpha=0$, i.e. if z is a zero of the characteristic form for (C,i). A surface Σ is characteristic at x if and only if its normal at x is characteristic.

In the case of second order equations, the characteristics are given by the zeroes of quadratic forms, giving rise to the terms elliptic, hyperbolic, and parabolic.

The question of uniqueness is complicated by the distinction between multiple and distinct characteristics. Consider the characteristic form of (CR,i), i.e.

$$3)\quad \lambda^N + \sum_{n=0}^{N-1} \sum_{|\beta|+n\leq N} a_{n,\beta}(t,y)\,\lambda^n z$$

$$= \lambda^N + \sum_{1}^{N} \lambda^{N-j} p_j(t,y;z),\ \text{where } p_j \text{ is the}$$

characteristic form of the A_j appearing in (CR,i). If for each $z \neq 0$ this has N distinct zeroes $\lambda_1, \ldots, \lambda_N$, then one could say that (CR,i) has distinct characteristics with respect to the coordinate system (t,y). In this concept of distinctness, the components of z are real, but the λ's may be complex. When the roots λ_j are all real, they yield N distinct characteristic directions in the plane through (t,y) containing the directions $(1,0,\ldots,0)$ and $(0,z_1,\ldots,z_k)$. We say that (C,i) has distinct characteristics if and only if for each coordinate system (t,y) in which the equation has the form (CR,i), the roots $\lambda_1,\ldots,\lambda_N$ are distinct.

The result we wish to prove can now be stated.

Uniqueness theorem (Calderon).

Let L(u)=f be a linear partial differential equation of order N in k+1 variables, with coefficients of the highest order terms real and in C_2, and the others measurable and bounded. If the characteristics of L are non-multiple, then the Cauchy problem for L with data given on a non-characteristic manifold has at most one solution of class C_N (unless k=2 and $N>3$). The same is true for non-linear equations, but uniqueness is in C_{n+2}.

To set this in context, we list some previous results on the same topic.

Cauchy-Kowalewski: analytic equation, data, and surface, equation not assumed linear, but surface non-characteristic for the given data; then a unique analytic solution exists.

Holmgren (1901): linear analytic equations; then solutions are unique in C_N.

Hadamard: the corresponding theorem for non-linear equations can be reduced to the case of a linear equation with smooth but not analytic coefficients.

Carleman (1939): Calderon's theorem for the special case of two independent variables.

Myshkis (1947), Plis (1954), deGiorgi (1955): examples showing lack of uniqueness with multiple characteristics.

Hartman and Wintner (1955), E. Heinz (1955), Aronszajn (1956): established Calderon's theorem for the case of elliptic equations; and similar results for hyperbolic equations are well-known in numerous papers and texts, e.g. Petrowsky.

3. Outline of the uniqueness proof.

In the present sketch all coefficients of L are assumed continuous. By considering the difference U between two proposed solutions of the Cauchy problem and introducing an appropriate change of independent variables, the problem can be posed as follows.

Suppose a function u in C_N is given and

i) $Lu = \partial^N u/\partial t^N + \sum_{j=1}^{N} A_j \partial u^{N-j}/\partial t^{N-j} + B(u) = 0,$

where $A_j = \sum_{|\alpha|=j} a_\alpha(t,x)(\partial/\partial x)^\alpha$, the a_α are in C_2, and B(u) involves derivates of order $<N$, with continuous coefficients; and

ii) u and all its derivatives of order $<N$ vanish on $t=|x|^2$.

Then u vanishes identically for $|x|^2 \leq t \leq h$, for some $h>0$.

From the given data, it follows that the N^{th} derivatives of U vanish on $t=|x|^2$, so we can assume u is zero for $t<|x|^2$. Thus the coefficients of the A_j may be modified at will there. We assume that this is done in such a way that they remain C_2, and that their oscillations in $0 \leq t \leq h$ (for all x) can be made arbitrarily small by making h small. We could, e.g. give them the values in $t \leq |x|^2 - \delta$ that they have for t=x=0, and interpolate in a C_2 way along the lines x=constant.

Now the A_j can be represented as $A_j = H_j(t)(i\Lambda)^j$, where Λ operates only on the x coordinates (thus $-\Lambda^2 = \sum_1^k \partial^2/\partial x_j{}^2$), and for each fixed τ $H_j(t)$ is a C_2^∞ s.i. op. of the form

$$[H_j(t)f](t,x) = a(t,x)f(t,x) + \text{l.i.m.}\int_{|x-y|>\varepsilon} f(t,y)h(t,x,x-y)dy.$$

Here h(t,x,z) is C_2 in t and x and C^∞ in z, for $|z|>1$.

Considering $\partial^{N-j}u/\partial t^{N-j}$ as a function in L^2 on the hyperplane t=constant, with $\|u\|^2 = \int_{E_k} |U(t,x)|^2 dx$, we can write (i) as

$$\partial^N u/\partial t^N + \sum_1^N H_j(i\Lambda)^j \partial^{N-j}u/\partial t^{N-j} = -B(u).$$

Written in this form the equation looks like an ordinary differential equation in t, and it can be reduced to à first order system in much the same way. Let $v_j(t) = (i\Lambda)^{N-j}\partial^{j-1}u/\partial t^{j-1}$, and $v = [v_1, \ldots, v_N]$. Then we have the system

$$\begin{aligned}
&\partial v_1/\partial t - i\Lambda v_2 && = 0\\
&\qquad \partial v_2/\partial t - i\Lambda v_3 && = 0\\
&\vdots\\
&H_N i\Lambda v_1 + \ldots \qquad + (\partial v_N/\partial t + H_1 i\Lambda v_n) && = -B(u).
\end{aligned}$$

We think of v in an L^2 space $\mathcal{L}$ with $\|v\|^2 = \sum_1^N \|v_j\|^2$.

Introducing $\mathcal{H}(t) = \begin{pmatrix} 0 & -I & 0 \ldots\ldots 0 \\ 0 & 0 & -I \quad 0 \ldots 0 \\ 0 \ldots\ldots\ldots\ldots 0 & -I \\ H_N \ldots\ldots\ldots\ldots H_1 \end{pmatrix}$ and

$w(t) = [0,0,\ldots,-B(u)]$, the equation becomes

4) $\partial v/\partial t + \mathcal{H}(t) i \Lambda v = w(t)$,

wherein H is a bounded operator on $\mathcal{L}$.

The algebraic side of the proof consists in reducing (4) as nearly as possible to a diagonal form. The method consists in diagonalizing the (matrix) symbol of the operator $\mathcal{H}$ exactly; this yields an approximate diagonalization of $\mathcal{H}$ itself. By $\sigma(\mathcal{H})$ we mean

$$\begin{pmatrix} 0 & -1 & 0 & \ldots & 0 \\ \vdots & & & & \\ 0 & & & & -1 \\ \sigma(H_N) & \ldots\ldots\ldots & & & \sigma(H_1) \end{pmatrix}$$

. Since $H_j(i\Lambda)^j = A_j$, we have

$\sigma(H_j) = p_j(t,x,z)|z|^{-j}$ = the characteristic form of A_j on $|z|=1$.

To diagonalize $\sigma(\mathcal{H})$ we look for the eigenvalues of $-\sigma(\mathcal{H})$, i.e. the roots of $\det[\lambda I + \sigma(\mathcal{H})] =$

$$\lambda^N + \lambda^{N-1} p_1(t,x,z)|z|^{-1} + \ldots + p_N(t,x,z)|z|^{-N}, \text{ which}$$

is seen to be the characteristic form of L, for $|z|=1$. Since the characteristics of L are assumed distinct, the roots λ_j are all distinct, and there is a diagonalizing matrix N such that $N\sigma(\mathcal{H})N^{-1}=D$, a diagonal matrix whose diagonal entries are the roots $\lambda_1,\ldots,\lambda_N$. An important question is whether N can be made a continuous function of t,x, and z for $0 \leq t \leq h$ and $|z|=1$. Topological difficulties in achieving this account for the exceptional case of three independent variables - in this case the set $|z| > 1$ is the outside of a circle in the plane, and it is often impossible to extend functions that are locally defined to all of this doubly connected region. We will take for granted that the diagonalizing matrix N(t,x,z) can be made a continuous function of t,x,z in the non-exceptional cases, where the space $|z|>1$ is simply connected. Let $N_0(z) = N(0,0,z)$. Then by considering Fourier transforms, as in Theorem 1, it is easy to see that the operator $\mathcal{N}_0$ with symbol N_0 has an inverse which we denote by $\mathcal{N}_0^{-1}$, and whose symbol we ˙enote by N_0^{-1}.

Now for $0 \leq t \leq h$ the entries in N(t,x,z), and their derivatives of order 2k with respect to z, differ by an arbitrarily small amount from those in $N_0(z)$. Hence, by Theorem 2, the operator $\mathcal{N}$ with symbol N differs in norm from $\mathcal{N}_0$ by very little, and is invertible.

To see to what extent $\mathcal{N}$ diagonalizes (4), consider

5) $\mathcal{N}(\partial v/\partial t) + \mathcal{N}\mathcal{H} i \Lambda v = \mathcal{N} w$,

and the expression we should approximate by diagonalizing the left side,

$$\partial(\mathcal{N}v)/\partial t + \mathcal{D} i \Lambda \mathcal{N} v,$$

where $\sigma(\mathcal{D}) = D$. We have

$$\partial(\mathcal{N}v)/\partial t = \mathcal{N}\partial v/\partial t + (\partial\mathcal{N}/\partial t)v,$$

where $\partial N/\partial t$ is the operator with symbol $\partial N(t,x,z)/\partial t$, thus a C_1^∞ operator. Further,

$$\begin{aligned} \mathcal{D} i \Lambda \mathcal{N} &= \mathcal{D}\mathcal{N} i \Lambda + \underbrace{i\mathcal{D}(\Lambda\mathcal{N} - \mathcal{N}\Lambda)} \\ &= (\mathcal{D}\circ\mathcal{N}) i \Lambda + \underbrace{\mathcal{S} i \Lambda + \mathcal{B}} \\ &= (\mathcal{N}\circ\mathcal{H}) i \Lambda + \mathcal{B}' \\ &= \mathcal{N}\mathcal{H} i \Lambda + \underbrace{\mathcal{S}' i \Lambda + \mathcal{B}'} \\ &= \mathcal{N}\mathcal{H} i \Lambda + \mathcal{B}'', \end{aligned}$$

where $\mathcal{S}$ denotes a smoothing and $\mathcal{B}$ a bounded operator on $\mathcal{L}$. This rests on Theorem 3, extended to matrices of s.i.ops. Thus (5) becomes

$$\begin{aligned} (6)\quad \partial\mathcal{N}v/\partial t + \mathcal{D} i \Lambda \mathcal{N} v &= \\ \mathcal{N}\partial v/\partial t + (\partial\mathcal{N}/\partial t)v + \mathcal{N}\mathcal{H} i \Lambda v + \mathcal{B}''v \\ &= \mathcal{N}w + \mathcal{B}'''\mathcal{N}v, \end{aligned}$$

where $\mathcal{B}''' = (\partial\mathcal{N}/\partial t + \mathcal{B}'')\mathcal{N}^{-1}$ is bounded.

Now it is time for concrete estimates. The desired conclusion ($u=0$ for $0 \leq t \leq h$) is based on the following two estimates, whose proof will not be considered.

<u>Lemma 1.</u> Let $u(t,x)$ be in C_1 in t and x, and let u and its first derivatives be in L^2 (X) for each fixed t in $0 \leq t \leq h$. Let $P(t)$ and $Q(t)$ be s.i.ops. such that $\sigma(P) = F_1(t,x,z)$ and $\sigma(Q) = F_2(t,x,z)$ are real, and in C_2^∞ in $|z| \geq 1$. (i.e. C_2 in t and x and C_∞ in z). Assume $P(t)$ has a 2-sided inverse for each t, or is zero for each t; and let $\phi_n(t) = (t+1/n)^{-n}$. If $u(0)=0$, and

$$\int_0^h \phi_n^2 \|\partial u/\partial t + (P+iQ)\Lambda u\|^2 dt \leq c \int_0^h \phi_n^2 \|u\|^2 dt$$

for some c and arbitrarily large n, then u(t)=0 in a neighbourhood $0 \le t \le \delta$ of t=0.

The role of ϕ_n here can be loosely described as that of a δ-function at the origin. If it were normalized so that $\int_0^h \phi_n = 1$, then we would have $\phi_n(0) \longrightarrow \infty$, $\phi_n(t) \longrightarrow 0$ for $t > 0$.

Lemma 2. Let u(t,x) be in C_1 in t and x, and u and its first derivatives in $L^2(X)$ for each t. Then

$$\int_0^h \phi_n^2 \|\partial u/\partial t\|^2 \ge n^2(1+h)^{-2} \int_0^h \phi_n^2 \|u\|^2 \, dt.$$

It is clear that the conclusion of Lemma 1 is of the right sort. To apply it to (6), one must first estimate $\|w\|^2 = \int_{E_k} |Bu|^2 \, dx$ in terms of $\|v\|^2 = \sum_1^N \int |v|^2 dx$. We have

$Bu = \sum_{l+j<N} A_{jl} \Lambda^j \partial^l u/\partial t^l$ with A_{jl} a C_0^∞ operator, and hence

$\|Bu\|^2 \le c \sum_{l+j<N} \|\partial^l(\Lambda^j u)/\partial t^l\|^2$. Since

$\int_0^h \phi_n^2 \|u\|^2 dt \le (1+h)^2 n^{-2} \int_0^h \phi_n^2 \|\partial u/\partial t\|^2 dt$, we get

$$\int_0^h \phi_n^2 \|Bu\|^2 dt \le c \sum_{l+j<N} \int_0^h \phi_n^2 \|\partial^l(\Lambda^j u)/\partial t^l\|^2 dt$$

$$\le c' \sum_{l+j<N} \int_0^h \phi_n^2 \|\partial^{N-j-1}(\Lambda^j u)/\partial t^{N-j-1}\|^2 dt$$

$$\le c' \sum_1^N \int_0^h \phi_n^2 \|v_j\|^2 dt = c' \int_0^h \phi_n^2 \|v\|^2 dt.$$

Since $\mathcal{N}^{-1}$ is bounded, we have also

$\int_0^h \phi_n^2 \|w\|^2 \le c' \int_0^h \phi_n^2 \|\mathcal{N}v\|^2 dt$. From (6) we get

$$(7) \quad \int_0^h \phi_n^2 \|\partial \mathcal{N}v/\partial t + \mathcal{D} i \Lambda \mathcal{N}v\|^2 dt$$

$$\le c'' \int_0^h \phi_n^2 (\|w\|^2 + \|\mathcal{N}v\|^2) dt$$

$$\le c''' \int_0^h \phi_n^2 \|\mathcal{N}v\|^2 dt.$$

Indicate the N components of $\mathcal{N}v$ by $[V_1, \ldots, V_N]$. Then there is an index j such that

$$(8) \quad \int_0^h \phi_n^2 \|v_j\|^2 dt \ge \frac{1}{N} \sum_{m=1}^N \int_0^h \phi_n^2 \|v_m\|^2 = \frac{1}{N} \int_0^h \phi_n^2 \|\mathcal{N}v\|^2 dt,$$

for infinitely many n.

Noting that $\mathcal{D} = \begin{pmatrix} D_1 & & 0 \\ & \ddots & \\ 0 & & D_N \end{pmatrix}$ is diagonal, and considering the j^{th} components, we get

$$\int_0^h \phi_n^2 \, \|\partial v_j/\partial t + D_j i \Lambda v_j\|^2 dt \leq \int_0^h \phi_n^2 \, \|\partial \mathcal{N} v/\partial t + \mathcal{D} i \Lambda \mathcal{N} v\|^2 dt$$
$$\leq c''' \int^h \phi_n^2 \, \|\mathcal{N} v\|^2 dt \leq N c''' \int_0^h \phi_n^2 \, \|v_j\|^2 dt,$$

for infinitely many n, in view of (7) and (8). This is quite close to the form of Lemma 1. To check the requirements for P and Q, note that $\sigma(D_j) = \lambda_j(t,x,z)$ is one of the characteristic roots of L. Since the coefficients of highest order terms of L are real, the complex λ_j come in conjugate pairs. Since they never coincide, the imaginary part of a complex $\lambda_j(t,x,z)$ can vanish for no (t,x,z). Thus the real part of $\sigma(iD_j)$ either vanishes identically, or vanishes never. From this we can derive that the P in $iD_j = P + iQ$ has the properties required in Lemma 2. The necessary invertibility follows as the invertibility of $\mathcal{N}$ did. Thus $V_j = 0$ for $0 \leq t \leq \delta$, and by (7) $\mathcal{N} v = 0$, and so $u = 0$, for $0 \leq t \leq \delta$.

This concludes the proof for a single equation. For a system, the topological details (construction of λ_j) are more difficult, and lead to other exceptional cases. The feeling is that these restrictions are inessential, while the distinctness of characteristics is essential.