Characteristic manifolds of linear partial differential equations*

Attila Máté
Brooklyn College of the City University of New York

October 6, 2011

Contents

1 The Cauchy problem for a partial differential equation **1**

2 The generalized Cauchy problem for a linear partial differential equation **2**
2.1 Transforming the generalized Cauchy problem 2

3 A note on noncommuting partial differential operators **4**

1 The Cauchy problem for a partial differential equation

Let n be a fixed positive integer. Let t, x_1, x_2, …, x_n be independent variables, where, intuitively, t stands for time, and x_i stand for position variables, and let u be the dependent variable. Consider the differential equation

$$\frac{\partial^m u}{\partial t^m} = F\left(t, x_1, x_2, \ldots, x_n, u, \ldots, \frac{\partial^k u}{\partial^{k_0} t\, \partial^{k_1} x_1 \ldots \partial^{k_n} x_n} \ldots\right) \tag{1}$$
$$k = k_0 + k_1 + \ldots + k_n \le m, \quad k_0 < m;$$

in other words, this equation has order m, and the mth partial derivative of u by t is explicitly expressed in a way that only involves lower derivatives with respect to t. This equation is to be solved for $t > 0$ if u, $\partial u/\partial t$,…, $\partial t^{m-1}/\partial t^{m-1}$ are given for $t = 0$. The problem of solving the above equations under these condition is called the *initial value problem* or *Cauchy problem.* There are a number of theorems asserting the existence and uniqueness of the solutions of these type of problems under various assumptions, the simplest being the Cauchy–Kowalevski Theorem.[1] Intuitively, such equations describe the time-evolution of a system, given the initial state of the system.

*Written for the course Mathematics 4211 at Brooklyn College of CUNY.

[1]The second name refers to Sophie Kowalevski (the transliteration she herself used). There are several other transliterations of her name. See `http://en.wikipedia.org/wiki/Sofia_Kovalevskaya` for the various transliterations.

2 The generalized Cauchy problem for a linear partial differential equation

Let n be a fixed positive integer. We will consider (possibly higher order) linear differential equations with dependent variables u and independent variables x_0, x_1, …, x_n. To simplify the notation, a sequence $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_n)$ of nonnegative integers will be called a multi-index. The length $|\boldsymbol{\alpha}|$ of $\boldsymbol{\alpha}$ will be defined as $|\boldsymbol{\alpha}| = \alpha_0 + \alpha_1 + \ldots + \alpha_n$. For a sequence $\mathbf{z} = (z_0, z_1, \ldots, z_n)$ we write $\mathbf{z}^{\boldsymbol{\alpha}} = (z_0^{\alpha_0}, z_1^{\alpha_1}, \ldots, z_n^{\alpha_n})$. For $0 \le i \le n$, we will write D_i for the differential operator $\partial^i/\partial x$, and we will write $\mathbf{D}^{\boldsymbol{\alpha}} = (D_0^{\alpha_0}, D_1^{\alpha_1}, \ldots, D_n^{\alpha_n})$. Let m be a positive integer. We will consider the differential equation

$$P(\mathbf{x}, \mathbf{D})u = f(\mathbf{x}), \tag{2}$$

where f is a given function,

$$P(\mathbf{x}, \mathbf{D}) = P_m(\mathbf{x}, \mathbf{D}) + \text{ lower order terms}, \tag{3}$$

and

$$P_m(\mathbf{x}, \mathbf{D}) = \sum_{|\boldsymbol{\alpha}|=m} a_{\boldsymbol{\alpha}}(\mathbf{x})\mathbf{D}^{\boldsymbol{\alpha}}; \tag{4}$$

here $\mathbf{x} = (x_0, x_1, \ldots, x_n)$, and $a_{\boldsymbol{\alpha}}(\mathbf{x})$ are given functions. $P_m(\mathbf{x}, \mathbf{D})$ is called the *principal part* of the differential operator $P_m(\mathbf{x}, \mathbf{D})$.

Let S be a smooth surface (i.e., and n dimensional submanifold of the space $\mathbb{R}^{n+1}$ of independent variables),[2] and let $\mathbf{n} = \mathbf{n}(\mathbf{x})$ be the normal vector (directed to a given side of the surface S) at the point $\mathbf{x}$ of S. Assume that at each point $\mathbf{x}$ of S, u and the directional derivatives of order 1, 2, …, $m-1$ in the direction of S are given.[3] Solving the differential equation $P(\mathbf{x}, \mathbf{D})u = f(\mathbf{x})$ under these conditions is called the *generalized Cauchy problem*.

2.1 Transforming the generalized Cauchy problem

We introduce new variables $t = \phi_0(x_0, x_1, \ldots, x_n)$, $y_1 = \phi_1(x_0, x_1, \ldots, x_n)$, …, $y_n = \phi_n(x_0, x_1, \ldots, x_n)$ in a neighborhood Σ of the point $(x_0^{(0)}, x_1^{(0)}, \ldots, x_n^{(0)})$ in such a way that in the new coordinates this point is $(0, 0, \ldots, 0)$, and, in this neighborhood, the equation of the surface S is $t = 0$ and the equation of a normal half-line to S in the direction of the normal vector is $t = s$, $x_1 = c_1$, …, $x_n = c_n$, for $s \ge 0$, where s is a parameter. According to the chain rule, for any i with $0 \le i \le n$ we have

$$D_i = \frac{\partial t}{\partial x_i}\frac{\partial}{\partial t} + \sum_{j=1}^{n} \frac{\partial y_i}{\partial x_i}\frac{\partial}{\partial y_i}. \tag{5}$$

[2] For easy visualization, it is best to assume that $n = 2$, i.e., that $\mathbb{R}^{n+1}$ is the ordinary three-dimensional space.

[3] By specifying u in S and all its directional derivatives up to order $m-1$ in the direction of the normal to S, all partial derivatives of S up to order $m-1$ at any point of S can be calculated, since these derivatives can be expressed in terms of the directional derivatives normal to S and the directional derivatives in directions inside S; the latter of course are known, because u in S is given. For all this, it is sufficient to assume the continuity of the partial derivatives of u up to order $m-1$. Instead of specifying the directional derivatives in the normal direction, one could specify the directional derivatives pointing outside the surface S – these would also be sufficient to determine all partial derivatives of u at any point of S.

Hence, for $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_n)$ with $|\boldsymbol{\alpha}| = m$ we have

$$\begin{aligned}\mathbf{D}^{\boldsymbol{\alpha}} u = D_0^{\alpha_0} D_1^{\alpha_1} \ldots D_n^{\alpha_n} u = \frac{\partial^m u}{\partial x_0^{\alpha_0}\, \partial x_1^{\alpha_1} \ldots \partial x_n^{\alpha_n}} &= \left(\prod_{i=0}^{n} \Big(\frac{\partial t}{\partial x_i} \frac{\partial}{\partial t} + \sum_{j=1}^{n} \frac{\partial y_j}{\partial x_i} \frac{\partial}{\partial y_j} \Big)^{\alpha_i} \right) u \\ &= \frac{\partial^m u}{\partial t^m} \left(\frac{\partial t}{\partial x_0} \right)^{\alpha_0} \left(\frac{\partial t}{\partial x_1} \right)^{\alpha_1} \cdots \left(\frac{\partial t}{\partial x_n} \right)^{\alpha_n} + \ldots,\end{aligned}$$

where terms that only involve derivatives lower than m with respect to t are represented by the dots at the end.[4] Thus, noting that

$$\boldsymbol{\nu} = \boldsymbol{\nu}(\mathbf{x}) = (\nu_0(\mathbf{x}), \nu_1(\mathbf{x}), \ldots \nu_n(\mathbf{x})) = \left(\frac{\partial t}{\partial x_0}, \frac{\partial t}{\partial x_1}, \ldots, \frac{\partial t}{\partial x_n} \right)$$

is the normal vector to S at a point $\mathbf{x}$, we have $P_m(\mathbf{x}, \mathbf{D})u = \tilde{P}_m(\mathbf{y}, \mathbf{n}) \partial^m u / \partial t^n + \ldots.$ where $\mathbf{y} = (t, y_1, y_2, \ldots, y_n)$, and, for any $n+1$-tuple $\boldsymbol{\xi} = (\xi_0, x_2, \ldots, \xi_n)$ we wrote $\tilde{a}_{\boldsymbol{\alpha}}(\mathbf{y}) = a_{\boldsymbol{\alpha}}(\mathbf{x})$ and

$$\tilde{P}_m(\mathbf{y}, \xi) = \sum_{|\boldsymbol{\alpha}|=m} \tilde{a}_{\boldsymbol{\alpha}}(\mathbf{y}) \boldsymbol{\xi}^{\boldsymbol{\alpha}}.$$

It is easy to see that

$$\tilde{P}_m(\mathbf{y}, \xi) = P_m(\mathbf{x}, \xi). \tag{6}$$

So, after transforming to the new variables, equation (2) can be written in the neighborhood Σ as

$$\tilde{P}_m(\mathbf{y}, \mathbf{n})\, \partial^m u / \partial t^n + \ldots = g(\mathbf{y}),$$

where $g(\mathbf{y}) = f(\mathbf{x})$ is the way the function f can be written in terms of the new variables. So, assuming $P_m(\mathbf{x}, \mathbf{n}) \neq 0$, in view of (6) we can solve this equation for $\partial^m u / \partial t^m$, to obtain an equation of form (1) in terms of the new variables.

That is, in case, the normal $\mathbf{n}$ of the surface S satisfies $P_m(\mathbf{x}, \mathbf{n}) \neq 0$, the generalized Cauchy problem for S behaves nicely (as in the original Cauchy problem). If $P_m(\mathbf{x}, \mathbf{n}) = 0$, the Cauchy problem is not expected to be solvable. The vectors $\mathbf{n}$ for which $P_m(\mathbf{x}, \mathbf{n}) = 0$ are called characteristic vectors. The surfaces (more precisely, n-dimensional submanifolds of $\mathbf{R}^{n+1}$) which are normal to characteristic vectors are called characteristic surfaces. It is important to note that the characteristic surfaces defined here are related, but not identical, to the characteristic curves defined for solving first order differential equations.

In the present discussion, we only considered linear equations, though the definition of characteristic surface can also be extended to nonlinear equations at the price of some complications; see [1, pp. 32–34]. In case $n = 1$ (i.e., when there are two independent variables) "surfaces" are curves, the characteristic surfaces are called characteristic curves, and for first order equations these are indeed identical to the characteristic curves used for solving these equations.

[4] Some care must be exercised in evaluating the power of the sum of differential operators in the above equation, since the *multinomial theorem* (the generalization of the binomial theorem to more than two terms) does not apply, because the proof of that theorem relies on the commutativity of product. See §3 below.

3 A note on noncommuting partial differential operators

Observe that the terms in the partial differential operator D in (5) do not commute. To illustrate what happens in a simpler example, consider the differential operator

$$D = fD_x + gD_y, \tag{7}$$

where $f = f(x, y)$ and $g = g(x, y)$ are given functions, and $D_x = \partial/\partial x$ and $D_y = \partial/\partial y$. Multiplication by the operators D, D_x, and D_y on the right acts as simple multiplication, and multiplication on the left acts as differentiation. For example, for functions u and v we have $uD_xv = uv_x$. Note that fD_x and gD_y do not commute. In fact

$$(fD_x)(gD_y)u = fD_xgD_yu = fD_x(gD_yu) = (fD_xg)(D_yu) + fgD_xD_yu = fg_xu_y + fgu_{yx},$$

where the first equation is to indicate that dropping the parentheses does not affect the meaning of the expression; the expression after the second equation is to indicate what exactly the operator D_x acts upon; the third equation involves the use of the product rule of differentiation; also note that in the first term after the third equation, the right parenthesis in (fD_xg) terminates the scope of D_x. On the other hand

$$(gD_y)(fD_x)u = (gD_yf)(D_xu) + gfD_yD_xu = gf_yu_x + gfu_{xy} = f_ygu_x + fgu_{yx},$$

where the last equation assumes that u is a nice enough function allowing the interchange of the order of partial derivatives; for example, the continuity of u_{xy} and u_{yx} suffices.[5] That is

$$(fD_x)(gD_y)u - (gD_y)(fD_x)u = fg_xu_y - f_ygu_x,$$

which can also be written as

$$(fD_x)(gD_y) - (gD_y)(fD_x) = fg_xD_y - f_ygD_x. \tag{8}$$

In calculating powers of D in (7) we can use the usual rules of algebra except that we need to take into account that fD_x and gD_y do not commute. Thus

$$\begin{aligned}
D^2 &= (fD_x + gD_y)^2 = (fD_x + gD_y)(fD_x + gD_y) \\
&= fD_xfD_x + fD_xgD_y + gD_yfD_x + gD_ygD_y \\
&= f(fD_xD_x + (D_xf)D_x) + f(gD_xD_y + (D_xg)D_y) \\
&\qquad + g(fD_yD_x + (D_yf)D_x) + g(gD_yD_y + (D_yg)D_y) \\
&= f^2D_x^2 + ff_xD_x + fgD_xD_y + fg_xD_y + gfD_yD_x + gf_yD_x + g^2D_y^2 + gg_yD_y \\
&= f^2D_x^2 + 2fgD_xD_y + g^2D_y^2 + (ff_x + f_yg)D_x + (fg_x + gg_y)D_y;
\end{aligned}$$

after the fourth equation, we used the product rule of differentiation, and in the last equation we assume that $D_yD_x = D_xD_y$, which is true, assuming that these operators are applied to a function u for which u_{xy} and u_{yx} are continuous, as we pointed out above. If f and g are constant functions, then the coefficients of D_x and D_y are zero on the right-hand side; thus, in this case, we have

$$D^m = (fD_x + gD_y)^m = \sum_{k=0}^{m} \binom{m}{k} f^k g^{m-k} D_x^k D_y^{m-k}$$

[5] See e.g. `http://www.sci.brooklyn.cuny.edu/~mate/misc/mixedpartial.pdf` for details and more subtle conditions.

for any positive integer m according to the binomial theorem. Indeed, for constant f and g the operators fD_x and gD_y do commute, so the binomial theorem is applicable,[6] assuming that D^m is applied to a function whose mth partial derivatives are continuous, so that in any mth derivative the order of D_x and D_y makes no difference. Even when f and g are not constant, we have

$$D^m = (fD_x + gD_y)^m = \sum_{k=0}^{m} \binom{m}{k} f^k g^{m-k} D_x^k D_y^{m-k} + \text{lower order terms}, \tag{9}$$

where the lower order terms are of form $c_{k,l} D_x^k D_y^l$ for $k + l < m$, with $c_{k,l} = c_{k,l}(x, y)$ being certain functions of x and y. This is because, in calculating the power $(fD_x + gD_y)^m$, when using the product rule of differentiation, if D_x or D_y is diverted to differentiate other parts of the product, the "unused" part $D_x^k D_y^l$ on the right of the term will have total power $k + l < m$.

One can say this in a different way. For a function H from the set $M = \{1, 2, \ldots, m\}$ into the set $\{fD_x, gD_y\}$ (i.e., $H(k) = fD_x$ or gD_y for k with $1 \le k \le m$), write $D(H) \overset{def}{=} H(1)H(2)\ldots H(m)$. That is, $D(H)$ is a product of m differential operators, each of which is either fD_x or gD_y. If σ is a permutation of the set M, i.e., if σ is a one-to-one function from M onto M, we can then write $D(H \circ \sigma) = H(\sigma(1))H(\sigma(2))\ldots H(\sigma(m))$. Then the operator $D(H) - D(H \circ \sigma)$ has order lower than m. More generally, if σ_1 and σ_2 are two permutations of M then

$$D(H \circ \sigma_1) - D(H \circ \sigma_2) \tag{10}$$

has order lower than m. This is clear from (8) if σ_1 and σ_2 are identical except for one interchange of adjacent values; that is, if there is an i with $1 \le i < m$ such that $\sigma_1(i) = \sigma_2(i+1)$, $\sigma_1(i+1) = \sigma_2(i)$, and $\sigma_1(j) = \sigma_2(j)$ for every j with $1 \le j \le m$ such that $j \ne i$, $i + 1$. Since for any two permutations σ_1 and σ_2 there is a sequence of permutations ρ_1, ρ_2, …, ρ_N such that $\rho_1 = \sigma_1$, $\rho_N = \sigma_2$, and for any k with $1 \le k < N$, the permutations ρ_k and ρ_{k+1} are identical except for one interchange of adjacent values. Hence (10) follows for any two permutations σ_1 and σ_2. Therefore, ignoring lower order terms in the proof of the binomial theorem, we obtain

$$D^m = (fD_x + gD_y)^m = \sum_{k=0}^{m} \binom{m}{k} (fD_x)^k (gD_y)^{m-k} + \text{lower order terms}$$

instead of (9), which is a more natural form of essentially the same equation.

That is, the principal part of the power $D^m = (fD_x + gD_y)^m$ is still given by the binomial theorem. This observation about the principal part of powers of partial differential operator in (7) can be extended to powers of sums containing more than two terms, such as the operator in (5), by using the *multinomial theorem* instead of the binomial theorem.[7]

References

[1] I. G. Petrovsky. *Lectures on Partial Differential Equations*. Dover Publications, New York, 1991. Translated from Russian by A. Shenitzer.

[6] The proof of the binomial theorem makes use of the commutativity of multiplication. Hence the binomial theorem is valid in any commutative ring with a unit element, while it need not be true in a noncommutative ring.

[7] See `http://en.wikipedia.org/wiki/Multinomial_theorem` for a statement of the multinomial theorem.