Characteristics and classification

Harald Hanche-Olsen
hanche@math.ntnu.no

Hypersurfaces.

A C^1 hypersurface in $\mathbb{R}^n$ locally looks like $\varphi = 0$ where $D\varphi \neq 0$. More precisely, let $S \subset \mathbb{R}^n$. We shall call S a C^1 *hypersurface* if every point $z \in S$ has a neighbourhood U on which there exists a C^1 function φ so that $D\varphi \neq 0$ in U and $S \cap U = \{x \in U : \varphi(x) = 0\}$.

Assume, without loss of generality,[1] that $\varphi_{x_n} \neq 0$ in U. Then the implicit function theorem guarantees the existence of a function w so that $\varphi(x_1, \dots, x_{n-1}, w(x_1, \dots, x_{n-1})) = 0$ for all $(x_1, \dots, x_{n-1})$ in a neighbourhood of $(z_1, \dots, z_{n-1})$. Thus S is locally the graph of a function: $x_n = w(x_1, \dots, x_{n-1})$.

We can locally change the coordinate system into one in which S has the simple form $y_n = 0$: Put $y = \psi(x) = (x_1, \dots, x_{n-1}, x_n - w(x_1, \dots, x_{n-1}))$. This defines y as a C^1 function of x, and this has a C^1 inverse which is trivially found to be given by $x = \psi^{-1}(y) = (y_1, \dots, y_{n-1}, y_n + w(y_1, \dots, y_{n-1}))$.

A C^1 hypersurface has a *unit normal vector* $\nu = \pm D\varphi/|D\varphi|$ at any point. A C^1 *curve* in S is just a C^1 function γ from a real interval into $\mathbb{R}^n$ so that $\gamma(t) \in S$ for every t. Differentiating the relation $\varphi(\gamma(t)) = 0$ we get $D\varphi(\gamma(t)) \cdot \dot\gamma(t) = 0$. In other words, $\dot\gamma \cdot \nu = 0$.

The set of all possible values for $\dot\gamma(t)$ where γ is a C^1 path in S with $\gamma(t) = z$ is called the *tangent space* of S at z. When S is locally represented as the graph $x_n = w(x_1, \dots, x_{n-1})$ any path in S is of the form $\gamma(t) = \bigl(\gamma_1(t), \dots, \gamma_{n-1}(t), w(\gamma_1(t), \dots, \gamma_{n-1}(t))\bigr)$ where $(\gamma_1, \dots, \gamma_{n-1})$ are arbitrary C^1 functions. It follows easily that the tangent space at z is really a subspace of $\mathbb{R}^n$ of dimension $n-1$. Since this space is orthogonal to the normal vector ν it is in fact equal to the orthogonal

[1] Since $D\varphi \neq 0$, there is at least one index i so $\varphi_{x_i}(z) \neq 0$. Then just renumber the independent variables $x = (x_1, \dots, x_n)$ to put x_i last. You may need to replace U by a smaller neighbourhood to make $\varphi_{x_n} \neq 0$ in all of U.

complement of ν (two subspaces of dimension $n-1$, one contained in the other, must be equal).

Cauchy problems for second order quasilinear PDEs.

A second order quasilinear PDE is a PDE of the form

$$\sum_{i=1}^n \sum_{j=1}^n a_{ij}(x, u(x), Du(x)) u_{x_i x_j}(x) = d(x, u(x), Du(x)).$$

We shall need some notation to deal with these equations more efficiently.

For two $n \times n$ matrices A and B, we write

$$A : B = \sum_{i,j} a_{ij} b_{ji} = \operatorname{tr}(AB) = \operatorname{tr}(BA).$$

Note that if B is symmetric we find $A : B = A^{\mathrm{T}} : B$, so we may as well replace A by its symmetric part $A = (A + A^{\mathrm{T}})/2$. We apply this in particular to the symmetric matrix

$$D^2 u = \begin{bmatrix} u_{x_1 x_1} & \dots & u_{x_1 x_n} \\ \vdots & \ddots & \vdots \\ u_{x_n x_1} & \dots & u_{x_n x_n} \end{bmatrix}$$

to rewrite our general second order quasilinear PDE as

$$A(x, u(x), Du(x)) : D^2 u(x) = d(x, u(x), Du(x))$$

where A is a symmetric matrix.

Cauchy data for this equation is given by the boundary conditions

$$u = g, \quad u_\nu = h \quad \text{on } S$$

where g and h are given functions on S. Here u_ν is short for $Du \cdot \nu$ (the normal derivative). The *Cauchy problem* is the problem of finding a solution to the PDE fitting given Cauchy data.

We can note that given Cauchy data on S immediately lets us deduce the full gradient Du on S: For if γ is a path in S we can differentiate the relation $u(\gamma(t)) = g(\gamma(t))$ to get $Du \cdot \dot\gamma(t) = \mathrm{d}g(\gamma(t))/\mathrm{d}t$ which is known; thus $Du \cdot \xi$ is known when ξ is a tangent vector, and also when $\xi = \nu$, and because these vectors together span $\mathbb{R}^n$, we know Du.

Can we also know D^2u on S? We shall see that the Cauchy data almost, but not quite, gives us this information. Sometimes the PDE will give us the missing information. In this case, the Cauchy data will be called noncharacteristic.

Now, recall that we know $u_{x_i}(\gamma(t))$ if γ is a path in S, so we will also know its derivative:

$$\frac{\mathrm{d}}{\mathrm{d}t} u_{x_i}(\gamma(t)) = \sum_j u_{x_i x_j}(\gamma(t)) \cdot \dot\gamma_j(t) \quad \text{so that } \frac{\mathrm{d}}{\mathrm{d}t} Du(\gamma(t)) = D^2u \cdot \dot\gamma(t)$$

and we know $D^2u \cdot \xi$ for every tangent vector ξ.

From the PDE we also know $A : D^2u$. [2]

The given Cauchy data is called *noncharacteristic* at $z \in S$ if the knowledge of $A : D^2d$ and $D^2u \cdot \xi$ for every tangent vector ξ is enough to determine D^2u.

This is a problem in linear algebra: We are given a symmetric matrix A and a unit vector ν, and we need to determine a symmetric matrix C given $A : C$ and $C \cdot \xi$ for all $\xi \perp \nu$.

The problem becomes easier to understand if we rotate the coordinate system so ν becomes parallel with a coordinate axis. So let Q be an orthogonal matrix with ν in its first column. Then the remaining $n-1$ columns span the orthogonal complement of ν, and our given data will let us compute CQy when $y = (0, y_2, \dots, y_n)$, i.e., when $y_1 = 0$. More interesting still, we can compute $C'y$ for all such vectors y, where $C' = Q^{\mathrm{T}}CQy$. Thus the last $n-1$ columns of C' are known. But C' is symmetric, and by symmetry then the first column is also known except for the first element. So the Cauchy data gives us all of the matrix $C' = (c'_{ij})$ except for c'_{11}, which is still unknown.

But from the PDE we also know $A : C = \operatorname{tr}(AC) = \operatorname{tr}(AQQ^{\mathrm{T}}CQQ^{\mathrm{T}}) = \operatorname{tr}(Q^{\mathrm{T}}AQQ^{\mathrm{T}}CQ) = A' : C'$ where $A' = Q^{\mathrm{T}}AQy$. That is, we know

[2] Note that because u is given and Du is known on S, we know the values of A and d on S.

$\sum_{ij} a'_{ij} c'_{ji}$. From this we can extract our missing piece of data c'_{11} if and only if $a'_{11} \neq 0$. But this is $a'_{11} = \nu^{\mathrm{T}} A \nu = \nu \cdot A \cdot \nu$ (use the second form if you do not wish to distinguish between row and column vectors), or if we write this out in component form, the *noncharacteristic condition* $a'_{11} \neq 0$ becomes

$$\sum_{i,j} a_{ij}(x, u, Du(x)) \nu_i \nu_j \neq 0,$$

which can be written in shorter form as $\nu^{\mathrm{T}} A \nu \neq 0$.

Though the derivation of this was somewhat lengthy, the result is easy to remember: In the original PDE, throw away the lower order terms, and consider the rest as the operator

$$\sum_{i,j} a_{ij}(x, u, Du(x)) \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j}$$

acting on u. Then formally replace $\partial/\partial x_i$ by ν_i. Finally, don't forget that the noncharacteristic condition is a condition on the *normal vector* of the initial surface.

The key to the Cauchy–Kowalevskaya theorem will be that, once we have cleared the initial hurdle of determining D^2u along S, we can in fact determine *all* higher order derivatives of u along S, and so we may be able to write up a power series representation for the solution.

The Cauchy–Kowalevskaya theorem.

Real analytic functions. A real function f defined on an open subset Ω of $\mathbb{R}^n$ is called *real analytic* if it can locally be expressed by a convergent power series; that is, around every point $x_0 \in \Omega$ there is a neighbourhood in which we can write $f(x) = \sum_\alpha c_\alpha (x - x_0)^\alpha$ where the sum is over all multi-indices α.

In the Cauchy–Kowalevskaya theorem, all given data is supposed to be real analytic. This includes the initial surface S, which means we locally can express S by the equation $\varphi(x) = 0$ where φ is a real analytic function with $D\varphi \neq 0$.

Straigthening out the initial surface. As described earlier, we can always change the coordinate system locally so that S is of the form $y_n = 0$. When φ is real analytic the change of coordinates is also real analytic, so when the Cauchy problem is expressed in the new coordinates, all functions involved remain real analytic. As we have seen, if the Cauchy data is noncharacteristic then this data and the PDE allows us to determine all the derivatives of u up to second order on the initial surface $y_n = 0$. We can in fact differentiate any number of times with respect to $y_1, \dots, y_{n-1}$ and thus compute $D^\alpha u$ on the initial surface for any multi-index α with $\alpha_n \le 2$.

To get the remaining derivatives we use the PDE. In our new coordinates the given equation has the form (using tildes to indicate values in the new coordinate system) $\sum_{i,j} \tilde{a}_{ij} u_{y_i y_j} = \tilde{d}$ and since $\tilde{\nu} = (0, \dots, 0, 1)$ the noncharacteristic condition is simply $\tilde{a}_{nn} \neq 0$. Thus we can solve the PDE for $u_{y_n y_n}$ and get

$$u_{y_n y_n} = \frac{1}{\tilde{a}_{nn}} \Big(\tilde{d} - \sum_{(i,j) \neq (n,n)} \tilde{a}_{ij} u_{y_i y_j} \Big)$$

The important part here is that nothing on the righthand side involves any derivatives of u in the y_n direction of any order higher than one. Differentiate this relation $\alpha_n - 2$ more times with respect to y_n and $(\alpha_1, \dots, \alpha_{n-1})$ times with respect to $(y_1, \dots, y_{n-1})$ to get a relation expressing $D^\alpha u$ in terms of derivatives $D^\beta u$ where $\beta_n < \alpha_n$. By induction on α_n, then, we can find an expression for $D^\alpha u(y_1, \dots, y_{n-1}, 0)$ for every multi-index α. This proves the uniqueness part of the Cauchy–Kowalevskaya theorem.

The Cauchy–Kowalevskaya theorem. *A Cauchy problem in which the coefficients of the PDE as well as the given Cauchy data and the initial surface are real analytic, and in which the Cauchy data is noncharacteristic for the PDE, has a unique real analytic solution in a neighbourhood of the initial surface.*

The existence proof for the Cauchy–Kowalevskaya theorem starts out simply enough: Since we know all the derivatives at the initial surface we know all the coefficients in a power series for the solution. The hard part is to show that this power series converges. We will not

go into this.

As a rather extreme example of Cauchy data which is *not* noncharacteristic, we mention the problem

$$u u_{x_n x_n} = 2u, \quad u(x_1, \dots, x_{n-1}, 0) = 0, \quad u_{x_n}(x_1, \dots, x_{n-1}, 0) = 0$$

which is characteristic because the initial data says the coefficient in front of the only second order derivative is zero at the initial hypersurface. And indeed this problem has several real analytic solutions: $u = 0$ and $u = x_n^2$ are two examples.

Holmgren's uniqueness theorem. *Under the hypotheses of the Cauchy–Kowalevskaya theorem, and if the PDE is in addition linear, that is, if it can be written as*

$$\sum_{i,j} a_{ij}(x) u_{x_i x_j}(x) + \sum_i b_i(x) u_{x_i}(x) + c(x) u(x) = d(x),$$

then a Cauchy problem with noncharacteristic data has at most one C^2 *solution* u.

Final remarks. The Cauchy–Kowalevskaya theorem and the Holmgren uniqueness theorem are much more general than you might think from this presentation. They work equally well for higher order PDEs as well as for systems of PDEs. In the more general cases, working out the noncharacteristic condition is a bit more involved. Do not forget that the Holmgren uniqueness theorem only holds for *linear* equations.

Classification in two variables

Here we consider the classification of quasilinear second order PDEs in two variables. These can be written

$$a u_{xx} + 2b u_{xy} + c u_{yy} = d$$

where a, b, c, and d are functions of x, y, u ,u_x, and u_y. This equation is of the general form we have considered with

$$A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}.$$

Writing $\nu = (\xi, \eta)$ we find the noncharacteristic condition: It is

$$a\xi^2 + 2b\xi\eta + c\eta^2 \neq 0.$$

Now the real symmetric matrix A has two real eigenvalues. Their product equals the determinant $ac - b^2$ of A. If $ac - b^2 > 0$ then the eigenvalues have the same sign. If they are positive then A is positive definite, and if they are negative $-A$ is positive definite. In both cases, the noncharacteristic condition always holds. The equation is called *elliptic*. The prototype elliptic equation is the Laplace (or Poisson) equation.

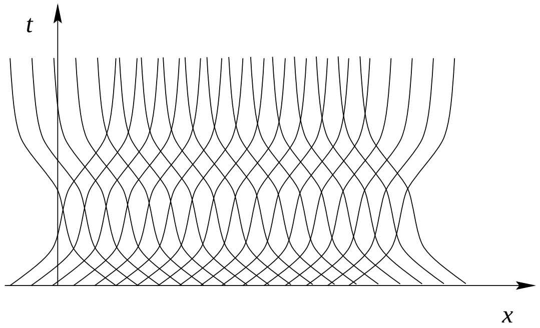
If, on the other hand, $ac - b^2 < 0$ then A has one positive and one negative eigenvalue. The equation is then called *hyperbolic*. If the eigenvalues are λ and $-\mu$ (with $\lambda > 0$ and $\mu > 0$) with eigenvectors p and q the characteristic condition with $\nu = vp + wq$ becomes $\lambda v^2 - \mu w^2 = 0$. The solution to this equation is the two lines given by $\sqrt{\lambda} v \pm \sqrt{\mu} w = 0$. Thus, there are two independent directions ν satisfying the characteristic condition. Do not, however, forget that ν is interpreted as a normal vector to a hypersurface (which is simply a line in two dimensions). The directions orthogonal to the vectors ν satisfying the characteristic condition are called characteristic directions, and curves in the plane which always follow a characteristic direction are called *characteristic curves*.

We should also consider the case $ac - b^2 = 0$. Then one or both eigenvalues of A are zero. If they are both zero then $A = 0$, and so our "second order" PDE is really first order. We shall disregard this case, and consider the case where just one eigenvalue is zero. Then the equation is called *parabolic*. The standard example is the heat equation $u_t = u_{xx}$. For a parabolic equation there is just one characteristic passing through each point. In the case of the heat equation, the characteristics are the lines $t =$ constant, so in particular the usual initial curve $t = 0$ is characteristic. And indeed, given Cauchy data (both u and u_t given for $t = 0$) do not seem to make much sense for the heat equation.

In a sense the heat equation is nontypical. If a general PDE is parabolic at a point, that point will typically be a boundary point between a region in which the equation is elliptic and one in which it is hyperbolic.

Example: A variable speed wave equation. Consider the equation

$$u_{tt} - c(x,t)^2 u_{xx} = 0.$$



The discriminant of this equation is $-c^2 < 0$, so the equation is hyperbolic. We expect to find two characteristics through each point, each with a direction orthogonal to (ξ, τ) given by the characteristic condition $\tau^2 - c^2\xi^2 = 0$. This becomes $\tau \pm c\xi = 0$, and if we put $\xi = 1$ we conclude that the characteristics will have normal vectors $(1, \mp c)$. The direction orhtogonal to this, and therefore parallel to the characteristic, will be $(\pm c, 1)$. The *characteristic curves* will therefore have the equation $\dot{x} = \pm c$ where we use the t variable to parametrize the characteristics.

For brevity, write ∂_t and ∂_x for the partial differential operators $\partial/\partial t$ and $\partial/\partial x$. When c is a constant, the wave operator can be factorized as $\partial_t^2 - c^2\partial_x^2 = (\partial_t - c\partial_x)(\partial_t + c\partial_x)$, which easily leads to the general solution $u(x,t) = \varphi(x - ct) + \psi(x + ct)$ where φ and ψ are arbitrary functions. For nonconstant c this is not quite so straightforward, but we can write

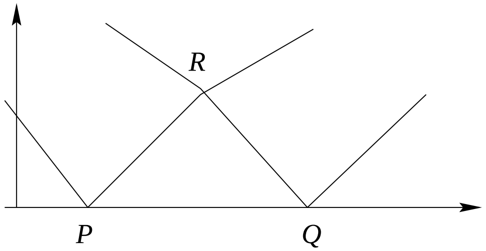
$$\begin{aligned}\partial_t^2 - c^2\partial_x^2 &= (\partial_t - c\partial_x)(\partial_t + c\partial_x) - (c_t - cc_x)\partial_x \\ &= (\partial_t + c\partial_x)(\partial_t - c\partial_x) - (c_t + cc_x)\partial_x\end{aligned}$$

Now if we put $u^{\mathsf{f}} = u_t - cu_x$ and $u^{\mathsf{b}} = u_t + cu_x$ we conclude

$$u_t^{\mathsf{f}} + cu_x^{\mathsf{f}} = (c_t + cc_x)u_x, \quad u_t^{\mathsf{b}} - cu_x^{\mathsf{b}} = (c_t - cc_x)u_x.$$

Thus we get a first order equation for u^{f} which can be integrated along the forward characteristics (since $u_t^{\mathsf{f}} + cu_x^{\mathsf{f}} = \dot{u}^{\mathsf{f}}$ when we put $x = x(t)$ in $u^{\mathsf{f}}(x,t)$ and $\dot{x} = c$) and one for u^{b} which can be integrated along the backward characteristics. Except, of course, the righthand sides of the two equations are not known until both equations are solved!

Still, this can be used as the basis for a numerical method known as the *method of characteristics.* (I am deliberately being a bit vague in the following description, since there are many different ways to fill in the details. In particular, the method outlined computes the solution on an irregular grid which is not known in advance. Variations of the method force this solution onto a regular grid, sacrificing some accuracy in order to gain convenience.)



Imagine we are given initial data $u(x,0) = g(x)$ and $u_t(x,0) = h(x)$. Then u_x is known at $t = 0$, and hence so are u^{f} and u^{b}. Pick a discrete set of points on the x axis and consider the characteristics starting from these points. It would seem reasonable to approximate each characteristic by a straight line between each collision with another characteristic. The forward characteristic starting at P has speed $c(P)$, and the backward characteristic starting at Q has speed $-c(Q)$. Approximate $\dot{u}^{\mathsf{f}}$ by the constant $(c_t + cc_x)u_x(P)$ along the former characteristic, and $\dot{u}^{\mathsf{b}}$ by $(c_t - cc_x)u_x$ along the latter. Then $u^{\mathsf{f}}(R)$ and $u^{\mathsf{b}}(R)$ can be computed approximately, and so can $u(R)$ since we know the value of u and its derivatives at nearby points. We continue in the same fashion to build the irregular grid consisting of intersections of characteristics, along with the solution on this grid.

To sum it up, we solve the given Cauchy problem by integrating a first order equation along each of the two families of characteristics. Since each solution depends on the other, this is not quite straightforward, but it is still a useful approach.

A more general hyperbolic equation. We return to the equation

$$au_{xx} + 2bu_{xy} + cu_{yy} = d$$

which we will rewrite as a system by putting $v = u_x$ and $w = u_y$:

$$v_y = w_x \tag{1}$$

$$av_x + 2bv_y + cw_y = d \tag{2}$$

Now consider a path in the xy plane. I.e., let x and y be functions of t and differentiate the unknown functions v and w with respect to t:

$$\dot{v} = \dot{x}v_x + \dot{y}v_y$$

$$\dot{w} = \dot{x}w_x + \dot{y}w_y$$

which we rewrite (using (1)) as

$$\begin{bmatrix} \dot{v} \\ \dot{w} \end{bmatrix} = \underbrace{\begin{bmatrix} \dot{x} & \dot{y} & 0 \\ 0 & \dot{x} & \dot{y} \end{bmatrix}}_{M} \begin{bmatrix} v_x \\ v_y \\ w_y \end{bmatrix}.$$

The matrix M in the above equation has rank 2 (unless $\dot{x} = \dot{y} = 0$, which we will assume not to be the case) and its nullspace is spanned by the vector $(\dot{y}^2, -\dot{x}\dot{y}, \dot{x}^2)$.

Because of equation (2) we are interested in the action of that matrix on the space $V = (a, 2b, c)^{\perp} \subset \mathbb{R}^3$. If $(\dot{y}^2, -\dot{x}\dot{y}, \dot{x}^2) \in V$ then multiplication by M maps V to a one-dimensional subspace of $\mathbb{R}^2$; otherwise it maps V to all of $\mathbb{R}^2$. We are interested in the former case, that is, in the case that

$$a\dot{y}^2 - 2b\dot{x}\dot{y} + c\dot{x}^2. \tag{3}$$

In this case, then, there exist α and β so that $M\zeta$ belongs to the subspace $\alpha z_1 + \beta z_2 = 0$ of $\mathbb{R}^2$ whenever $\zeta \in V$. Thus, since the general solution to (2), considered as an algebraic equation, is the sum of a particular solution and a general member of V, we find a constant γ so that

$$\alpha\dot{v} + \beta\dot{w} = \gamma \tag{4}$$

must hold. Moreover, it is *only* under the condition (3) that we can deduce such an equation. To recognize this condition we note that a normal vector to the parametrized curve $(x(t), y(t))$ is given by $(\xi, \eta) = (-\dot{y}, \dot{x})$. But then (3) is nothing but the characteristic condition $a\xi^2 + 2b\xi\eta + c\eta^2 = 0$.

To conclude, we have shown that the characteristic curves are precisely those curves for which the given equation implies a linear relationship of the form (4), which is seen to constrain the possible choices of (u_x, u_y) along the curve. Two such relations — one for each characteristic direction — is just what is needed for the method of characteristics to work. This is precisely the hyperbolic case. Moreover, a simple consideration of the geometry of characteristics show that the initial surface must not be tangential to any of the characteristics, or the method of characteristics might fail. We would expect the method of characteristics to yield a solution to a non-characteristic Cauchy problem for a hyperbolic equation. For equations of parabolic or elliptic type, we can expect no such thing.

Final remark. Hyperbolic equations often arise directly from conservation laws (e.g., the wave equation comes from conservation of mass and momentum together with the equation of state of the medium). A weakness of the method of characteristics is its failure to conserve quantities that should be conserved. To achieve this goal, it is better to leave the equation in conservation form, like $u_t + f(u)_x = 0$ where u may be a vector valued function, and to derive *conservative numerical schemes* directly from this form. But now we are far outside the scope of this course.