MATH 580 LECTURE NOTES 2: THE CAUCHY-KOVALEVSKAYA THEOREM

TSOGTGEREL GANTUMUR

Abstract. The Cauchy-Kovalevskaya theorem, characteristic surfaces, and the notion of well posedness are discussed. We review some basic facts about analytic functions of a single variable in Section 1, which can be skipped. I thank Ibrahim for making his class notes available to me.

Contents

1. Analytic functions of one variable

We understand by a *power series* an expression of the form

$$f(z) = \sum_{n=0}^{\infty} a_n (z-c)^n, \tag{1}$$

with the coefficients $a_n \in \mathbb{C}$, and the centre $c \in \mathbb{C}$. Assume that the above series converges. Then obviously $|a_n||z-c|^n \to 0$ as $n \to \infty$, so that $|a_n||z-c|^n \leq M$ for some constant $M < \infty$. In other words, we have

$$|z-c| \leq R := \sup\{r \geq 0 : \sup_n |a_n| r^n < \infty\}. \tag{2}$$

Put another way, the power series (1) diverges whenever $|z-c| > R$. The converse statement is almost true as seen from the theorem below, which justifies the fact that the above defined $R \in [0,\infty]$ is called the *convergence radius* of the power series (1). We use the notation $D_R(c) = \{z \in \mathbb{C} : |z-c| < R\}$.

Theorem 1. *Let $R \in [0,\infty]$ be defined by (2). Then the power series (1) converges absolutely uniformly on each compact subset of the open disk $D_R(c)$, and diverges at every $z \in \mathbb{C} \setminus \bar{D}_R(c)$. Moreover, R can be determined by the* Cauchy-Hadamard formula

$$\frac{1}{R} = \limsup_{n\to\infty} |a_n|^{1/n}, \tag{3}$$

Date: October 3, 2011.

with the conventions $1/\infty = 0$ *and* $1/0 = \infty$*, and furthermore, provided that* $a_n = 0$ *for only finitely many* n*, one can estimate* R *by the* ratio test

$$\liminf_{n\to\infty} \frac{|a_n|}{|a_{n+1}|} \leq R \leq \limsup_{n\to\infty} \frac{|a_n|}{|a_{n+1}|}. \tag{4}$$

Proof. Without loss of generality, we may take $c = 0$. Divergence at every $z \in \mathbb{C} \setminus \bar{D}_R(c)$ is demonstrated above. For convergence, let $z \in D_r$ with $r < R$. Then for any $\rho \in (r, R)$ we have $|a_n z^n| \leq |a_n|\rho^n \frac{r^n}{\rho^n} \leq M\frac{r^n}{\rho^n}$ for some constant $M < \infty$. Since $\frac{r}{\rho} < 1$, $\sum a_n z^n$ converges normally in D_r. Since any $z \in D_R$ is in some such D_r with $r < R$, the series converges absolutely uniformly on each compact subset of D_R.

To prove (3), let ϱ be defined by $1/\varrho = \limsup_{n\to\infty} |a_n|^{1/n}$ with the intention of showing that $\varrho = R$. By definition, for any $\varepsilon \in (0, 1)$, we have $|a_n|\varrho^n \geq (1-\varepsilon)^n$ for infinitely many n, and there is n_ε such that $|a_n|\varrho^n \leq (1+\varepsilon)^n$ for all $n > n_\varepsilon$. Thus if $|z| > \varrho$ then $|a_n z^n| > 1$ for infinitely many n, and the series $\sum a_n z^n$ diverges. This implies that $\varrho \geq R$. On the other hand, if $|z| < \varrho$, then for any $\varepsilon > 0$ we have $|a_n z^n| \leq |a_n|\varrho^n \frac{|z|^n}{\varrho^n} \leq (1+\varepsilon)^n \frac{|z|^n}{\varrho^n} =: k^n$ for all $n > n_\varepsilon$. By choosing $\varepsilon > 0$ small enough, one can ensure that $k \in [0, 1)$, and so $\sum a_n z^n$ converges. This implies that $\varrho \leq R$.

Now we shall prove the ratio test. Let α be the limit infimum in (4) and suppose that $|z| < \alpha$. By definition, for any $\varepsilon > 0$ we have $|a_n| \geq (\alpha - \varepsilon)|a_{n+1}|$ for all sufficiently large n. This gives $|a_n z^n| \leq C(\frac{|z|}{\alpha-\varepsilon})^n$ for all sufficiently large n, with some constant $C > 0$. By choosing ε small enough we show the convergence of $\sum a_n z^n$, which implies that $\alpha \leq R$.

For the upper bound on R, let β be the limit supremum in (4), and suppose that $|z| > \beta$ and $\varepsilon = |z| - \beta > 0$. Then by definition, we have $|a_n| \leq (\beta + \varepsilon)|a_{n+1}|$ for all sufficiently large n. So $|a_n z^n| \geq \frac{C|z|^n}{(\beta+\varepsilon)^n} \geq C$ for some constant $C > 0$, and the series diverges. This implies that $R \leq \beta$. □

In what follows, Ω will denote an open subset of $\mathbb{C}$.

Definition 2. A complex-valued function $f : \Omega \to \mathbb{C}$ is called *analytic at* $z \in \Omega$ if it is developable into a power series around z, i.e, if there are coefficients $a_n \in \mathbb{C}$ and a radius $r > 0$ such that the following equality holds for all $h \in D_r$

$$f(z + h) = \sum_{n=0}^{\infty} a_n h^n.$$

Moreover, f is said to be *analytic on* Ω if it is analytic at each $z \in \Omega$. The set of analytic functions on Ω is denoted by $C^\omega(\Omega)$.

Exercise 3. Show that the product of two analytic functions is analytic, and that their quotient is analytic wherever the denominator function is nonzero.

Suppose that f has a power series expansion at c with the convergence radius $R > 0$. Can we say that f is analytic on the disk $D_R(c)$? This is answered by the following.

Lemma 4. *Let* $R > 0$ *be the convergence radius of the power series* $f(z) = \sum a_n(z-c)^n$*, and let* $d \in D_R(c)$*. Then we have*

$$f(z) = \sum_{j=0}^{\infty} \left(\sum_{n=j}^{\infty} \binom{n}{j} a_n (d-c)^{n-j} \right) (z-d)^j,$$

where the convergence radius of the power series is at least $R - |d-c|$*. In particular, we have* $f \in C^\omega(D_R(c))$*, and the convergence radius of a rearranged power series depends continuously on its centre.*

Proof. We have

$$(z-c)^n = (z-d+d-c)^n = \sum_{j=0}^{n} \binom{n}{j} (z-d)^j (d-c)^{n-j},$$

so that the proof is established upon justifying

$$\sum_{n=0}^{\infty} a_n \sum_{j=0}^{n} \binom{n}{j} (z-d)^j (d-c)^{n-j} = \sum_{j=0}^{\infty} \sum_{n=j}^{\infty} \binom{n}{j} a_n (z-d)^j (d-c)^{n-j},$$

for $z \in D_r(d)$ with $r = R - |d-c|$. This can be done, for instance, if the left hand side is absolutely uniformly convergent on each compact subset of $D_r(d)$. To this end, let $|z-d| \leq \rho - |d-c|$ with $\rho < R$. Then we have

$$\sum_{j=0}^{n} \binom{n}{j} |z-d|^j |d-c|^{n-j} = (|z-d| + |d-c|)^n \leq \rho^n,$$

and since $a_n \rho^n = a_n R^n (\rho/R)^n$ we obtain the desired convergence.

The continuity of convergence radius can be shown as follows. Let R' denote the convergence radius of the rearranged series centred at d. We have $R' \geq R - |d-c|$, or put differently, $R - R' \leq |d-c|$. So if $|d-c| < R/2$ it is obvious that $c \in D_{R'}(d)$, which means that the above reasoning can be applied with the roles of the two power series interchanged, giving $R' - R \leq |c-d|$. □

Now we turn to the question of termwise differentiating and integrating power series.

Theorem 5. *Let R be the convergence radius of the power series (1). Then both*

$$g(z) = \sum_{n=0}^{\infty} n a_n (z-c)^{n-1}, \qquad \text{and} \qquad F(z) = \sum_{n=0}^{\infty} \frac{a_n}{n+1} (z-c)^{n+1},$$

have convergence radii equal to R, and there hold that

$$f' = g \qquad \text{and} \qquad F' = f, \qquad \text{in } D_R(c).$$

Proof. It is obvious that the convergence radius R' of the power series representing g is at most R, that is, $R' \leq R$. To prove the other direction, let $r < R$. Then for any $\varepsilon > 0$ there is a constant $C_\varepsilon > 0$ such that

$$n|a_n|r^n \leq C_\varepsilon (1+\varepsilon)^n |a_n| r^n \leq C_\varepsilon (1+\varepsilon)^n (r/R)^n |a_n| R^n,$$

and choosing ε small enough we see that $r \leq R'$, and so $R \leq R'$.

Now we will show that $f' = g$ in $D_R(c)$, i.e., that for each $z \in D_R(c)$ one has

$$f(z+h) = f(z) + g(z)h + o(|h|).$$

To this end, we write

$$f(z+h) - f(z) = \sum_{n=0}^{\infty} a_n \left((z+h)^n - z^n \right) = h \sum_{n=0}^{\infty} a_n \sum_{j=0}^{n-1} (z+h)^j z^{n-1-j} =: h\lambda_z(h).$$

Let $r < R$ be such that $|z| < r$, and consider all h satisfying $|z+h| \leq r$. Then

$$\sum_{n=0}^{\infty} |a_n| \sum_{j=0}^{n-1} |z+h|^j |z|^{n-1-j} \leq \sum_{n=0}^{\infty} |a_n| n r^{n-1} < \infty,$$

so the series for λ_z converges locally uniformly in a neighbourhood of the origin. Hence λ_z is continuous at 0, and moreover from $\lambda_z(0) = g(z)$, we infer

$$\lambda_z(h) = g(z) + o(1),$$

with $o(1) \to 0$ as $|h| \to 0$. The claim is proven since

$$f(z+h) - f(z) = h(g(z) + o(1)) = hg(z) + o(|h|).$$

The claims about F follow from the above if we start with F instead of f. □

By repeatedly applying the preceding theorem, we see that the coefficients of the power series of f about $c \in \Omega$ are given by $a_n = f^{(n)}(c)/n!$, or in other words, if $f \in C^\omega(\Omega)$ and $c \in \Omega$ then the following *Taylor series* converges in a neighbourhood of c.

$$f(z) = \sum_{n=0}^{\infty} \frac{f^{(n)}(c)}{n!}(z-c)^n. \tag{5}$$

Exercise 6. Prove the preceding statement.

Recall that an *accumulation point* of a set $D \subset \mathbb{C}$ is a point $z \in \mathbb{C}$ such that any neighbourhood of z contains a point $w \neq z$ from D. We say that $z \in D$ is an *isolated point* if it is not an accumulation point of D. If all points of D are isolated D is called *discrete*.

Theorem 7 (Identity theorem)**.** *Let $f \in C^\omega(\Omega)$ with Ω a connected open set, and let at least one of the followings hold.*
(a) There is $b \in \Omega$ such that $f^{(n)}(b) = 0$ for all n.
(b) The zero set of f has an accumulation point in Ω.
Then $f \equiv 0$ in Ω.

Proof. Each $\Sigma_n = \{z \in \Omega : f^{(n)}(z) = 0\}$ is relatively closed in Ω, so the intersection $\Sigma = \bigcap_n \Sigma_n$ is also closed. But Σ is also open, because $z \in \Sigma$ implies that $f \equiv 0$ in a small disk centred at z by a Taylor series argument. Part (a) is proven, since $b \in \Sigma$, meaning that Σ is nonempty, we get $\Sigma = \Omega$. For (b), we shall prove now that Σ is nonempty.

Let $c \in \Omega$ be an accumulation point of Σ_0. If $c \in \Sigma$, then $\Sigma = \Omega$. If $c \notin \Sigma$, then there is n such that $f^{(n)}(c) \neq 0$. So we have $f(z) = (z-c)^n g(z)$ for some continuous function g with $g(c) \neq 0$. This will imply the existence of a neighbourhood of c where f has at most one zero, contradicting that c is an accumulation point of the zero set of f. □

Exercise 8. Let $u, v \in C^\omega(\Omega)$ with Ω a connected open set, and let $u \equiv v$ in a non-discrete set $D \subset \Omega$. Then $u \equiv v$ in Ω.

By this exercise, an analytic function is completely determined by its restriction to any non-discrete subset of its domain of definition. In other words, if it is at all possible to extend an analytic function (defined on a non-discrete set) to a bigger domain, then there is only one way to do the extension.

2. Single autonomous ordinary differential equations

The Cauchy-Kovalevskaya theorem is a result on local existence of analytic solutions to a very general class of PDEs. However, it is best to start with the ODE case, which is simpler yet contains half the main ideas. Consider the problem

$$u' = f(u), \qquad u(0) = e, \tag{6}$$

where f is a given function analytic at e, and u is the unknown function. Cauchy's theorem, proved by him during 1831-35, guarantees that a *unique solution exists that is analytic at* 0.

First of all, note that we can take $e = 0$, by changing u by $u - e$ and f by $f(e + \cdot)$. With the intent of finding the Maclaurin series coefficients of u, we can repeatedly differentiate $u' = f(u)$ to get

$$u'' = [f(u)]' = f'(u)u', \qquad u''' = [f(u)]'' = f''(u)(u')^2 + f'(u)u'', \qquad \ldots. \tag{7}$$

The Faà di Bruno formula would give the precise expression for $[f(u)]^{(m)}$, but without having to look up or derive that formula, just from the considerations (7) it is clear that

$$u^{(k)} = [f(u)]^{(k-1)} = q_k(f(u), \dots, f^{(k-1)}(u), u', \dots, u^{(k-1)}), \tag{8}$$

where q_k is a multivariate polynomial with *nonnegative* integer coefficients. We evaluate this at $z = 0$, and use $u(0) = 0$, to get

$$u^{(k)}(0) = q_k(f(0), \dots, f^{(k-1)}(0), u'(0), \dots, u^{(k-1)}(0)). \tag{9}$$

Now we repeatedly apply the same formula (with k having values $k-1$, $k-2$, etc.) to eliminate all $u^{(m)}(0)$ from the right hand side, inferring

$$u^{(k)}(0) = Q_k(f(0), \dots, f^{(k-1)}(0)), \tag{10}$$

with another multivariate polynomial Q_k having *nonnegative* integer coefficients. This incidentally proves uniqueness of analytic solutions to (6), since (10) fixes their Maclaurin series coefficients at 0. Moreover, provided that the Maclaurin series

$$u(z) = \sum_{n=0}^{\infty} \frac{u^{(n)}(0)}{n!} z^n, \tag{11}$$

with $u^{(n)}(0)$ given by (10) converges in a neighbourhood of 0, the function $v = u' - f(u)$ is analytic at 0, and by construction, its Maclaurin series is identically zero. Hence, by the identity theorem v must vanish wherever it is defined, meaning that $u' = f(u)$ there. Now it remains only to show that the series above converges in a neighbourhood of 0.

The heart of Cauchy's proof is his *method of majorants*, which is an ingenious and a very peculiar way of exploiting the positivity of the coefficients of Q_k against the underlying analytic setting. For two functions g and G, both infinitely differentiable at $c \in \mathbb{C}$, we say that G *majorizes* g at c, if

$$|g^{(k)}(c)| \leq G^{(k)}(c), \qquad k = 0, 1, \dots. \tag{12}$$

In other words, the Taylor series coefficients of g at c is bounded in magnitude by the corresponding coefficients of G. Since our right hand side f is analytic at 0, there exist constants $M > 0$ and $r > 0$ such that

$$\frac{|f^{(k)}(0)|}{k!} \leq \frac{M}{r^k}, \qquad k = 0, 1, \dots. \tag{13}$$

Then certainly the function

$$F(z) = \frac{M}{1 - z/r} = M + \frac{M}{r} z + \dots + \frac{M}{r^k} x^k + \dots, \tag{14}$$

majorizes f at 0. Let us consider the initial value problem

$$U' = F(U), \qquad U(0) = 0. \tag{15}$$

Then by (10) we have

$$U^{(k)}(0) = Q_k(F(0), \dots, F^{(k-1)}(0)), \tag{16}$$

and

$$\begin{aligned} |u^{(k)}(0)| &= |Q_k(f(0), \dots, f^{(k-1)}(0))| \\ &\leq Q_k(|f(0)|, \dots, |f^{(k-1)}(0)|) \\ &\leq Q_k(F(0), \dots, F^{(k-1)}(0)) \\ &= U^{(k)}(0), \end{aligned} \tag{17}$$

where we have used the nonnegativity of the coefficients of Q_k in the second an third lines, and the majorant property of F in the third line. The conclusion is that the solution u of the original problem (6) is majorized by the solution U of (15) at 0. Hence, if (15) has an analytic solution, u is automatically analytic. But (15) is easily solvable, with

$$U(z) = r(1 - \sqrt{1 - Mz/r}) = Mz/2 + \ldots, \tag{18}$$

whose Taylor series around 0 has nonnegative coefficients. We have proved the following.

Theorem 9. *The initial value problem*

$$u' = f(u), \qquad u(0) = e, \tag{19}$$

with $f : \mathbb{C} \to \mathbb{C}$ analytic at 0, has a unique solution u that is analytic at 0.

From the majorant (18), the radius of convergence of the solution u can be estimated as $R \geq r/M$. Recalling that $M > 0$ and $r > 0$ are constants from the bounds (13), and recalling Cauchy's estimates (Cauchy 1831)

$$|f^{(n)}(0)| \leq \frac{n!}{r^n} \sup_{|z|=r} |f(z)|, \tag{20}$$

that is valid if f is analytic on the disk $|z| \leq r$, we can estimate $M \leq \sup_{|z|=r} |f(z)|$. But there are functions such as $r/(r - Mz)$ that saturate Cauchy's estimates, meaning that M is essentially the magnitude of f in its domain of analyticity. Now, the magnitude of f is equal to the "speed" $|u'|$, hence the "time" it takes for u to become of magnitude r is roughly r/M. If we assume that f ceases to be analytic outside the disk $|z| \leq r$, these considerations imply that the convergence radius of u is roughly of order r/M, which cannot be improved in general.

Exercise 10. By way of an explicit example, make the conclusion of the preceding paragraph precise.

3. Systems of ordinary differential equations

Our next step towards the Cauchy-Kovalevskaya theorem is Cauchy's existence theorem for the system:

$$u_j' = f_j(z, u_1, \ldots, u_m), \qquad u_j(0) = 0, \qquad j = 1, \ldots, m. \tag{21}$$

We could have eliminated the dependence of f on z by introducing the new variable u_{m+1} with the equation $u_{m+1}' = 1$, but we intentionally leave it there in anticipation of the PDE case that is considered in the next section. The above equation can be written compactly as

$$u' = f(z, u), \qquad u(0) = 0, \tag{22}$$

with u having values in $\mathbb{C}^m$. We want to assume analyticity on f, which is a multivariate function. Let us clarify this notion now.

In $\mathbb{C}^n$, a *power series* is an expression of the form

$$f(z) = \sum_{\alpha_1=0}^{\infty} \cdots \sum_{\alpha_n=0}^{\infty} a_{\alpha_1,\ldots,\alpha_n} (z_1 - c_1)^{\alpha_1} \cdots (z_n - c_n)^{\alpha_n}, \tag{23}$$

with the coefficients $a_{\alpha_1,\ldots,\alpha_n} \in \mathbb{C}$, and the centre $c \in \mathbb{C}^n$. Introducing the *multi-index* $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}_0^n$, and the conventions $|\alpha| = \alpha_1 + \ldots + \alpha_n$ and $z^\alpha = z_1^{\alpha_1} \cdots z_n^{\alpha_n}$ for $z \in \mathbb{C}^n$, the above series can also be written as

$$f(z) = \sum_{|\alpha| \geq 0} a_\alpha (z - c)^\alpha. \tag{24}$$

If the preceding series converges for some z, then obviously there is a constant $M < \infty$, such that $|a_\alpha||z_1 - c_1|^{\alpha_1} \cdots |z_n - c_n|^{\alpha_n} \leq M$ for all α. In particular, if this series converges in a neighbourhood of c, then there are constants $M < \infty$ and $r > 0$, such that $|a_\alpha| \leq M r^{-|\alpha|}$

for all α. On the other hand, if $r \in \mathbb{R}^n$ and $M < \infty$ satisfy $|a_\alpha| r_1^{\alpha_1} \cdots r_n^{\alpha_n} \leq M$ for all α, then the series converges absolutely for all $z \in \mathbb{C}^n$ satisfying $|z_i - c_i| < r_i$ for each i.

Exercise 11. Prove the statements in the previous paragraph.

Definition 12. Let Ω be an open subset of $\mathbb{C}^n$. A complex-valued function $f : \Omega \to \mathbb{C}$ is called *analytic at* $c \in \Omega$ if it is developable into a power series around c, i.e, if there are coefficients $a_\alpha \in \mathbb{C}$, $(\alpha \in \mathbb{N}_0^n)$, such that the power series (24) converges in a neighbourhood of c. Moreover, f is said to be *analytic on* Ω if it is analytic at each $c \in \Omega$. The set of analytic functions on Ω is denoted by $C^\omega(\Omega)$.

In parallel to the single variable case, one can show that if f is analytic at c, then the series (24) is its multivariate Taylor series, i.e., the coefficients are given by

$$a_\alpha = \frac{\partial^\alpha f(c)}{\alpha!} \equiv \frac{\partial_1^{\alpha_1} \ldots \partial_n^{\alpha_n} f(c)}{\alpha_1! \cdots \alpha_n!}, \tag{25}$$

where we have introduced the convention $\alpha! = \alpha_1! \cdots \alpha_n!$. We have the identity theorem for multivariate analytic functions, which is necessarily a bit weaker than its single variable counterpart. Namely, the zeros of a multivariate analytic function can form a non-discrete set. For example, the zero set of $f(z) = z_1$ in $\mathbb{C}^2$ is $\{0\} \times \mathbb{C}$.

Theorem 13 (Identity theorem)**.** *Let $f \in C^\omega(\Omega)$ with Ω a connected open set in $\mathbb{C}^n$, and with some $b \in \Omega$, let $\partial^\alpha f(b) = 0$ for all α. Then $f \equiv 0$ in Ω. In particular, the same conclusion holds if f vanishes on some open subset of Ω.*

Proof. Each $\Sigma_\alpha = \{z \in \Omega : \partial^\alpha f(z) = 0\}$ is relatively closed in Ω, so the intersection $\Sigma = \bigcap_{|\alpha| \geq 0} \Sigma_\alpha$ is also closed. But Σ is also open, because $z \in \Sigma$ implies that $f \equiv 0$ in a neighbourhood of z by a Taylor series argument. Since $b \in \Sigma$, Σ is nonempty, implying that $\Sigma = \Omega$. □

Now let us return to the problem (22). We assume that the right hand side f is analytic in its $m+1$ arguments. To determine the higher derivatives of u, we start differentiating the equation $u_j' = f_j(z, u)$ as

$$[f_j(z,u)]' = \partial_z f_j + \partial_{u_i} f_j \cdot u_i', \qquad [f_j(z,u)]'' = \partial_z^2 f_j + 2\partial_z \partial_{u_i} f_j \cdot u_i' + \partial_{u_i} \partial_{u_\ell} f_j \cdot u_i' u_\ell' + \partial_{u_i} f_j \cdot u_i'',$$

where summation is taken over repeated indices. From here it is clear that

$$u_j^{(k)} = [f_j(z,u)]^{(k-1)} = q_k(\partial^\beta f_j(u), u^{(\ell)}), \tag{26}$$

where q_k is a multivariate polynomial with nonnegative coefficients, and it is understood that the arguments of q_k are all $\partial^\beta f_j(u)$ with $|\beta| \leq k-1$, and all components of all $u^{(\ell)}$ with $\ell \leq k-1$. We evaluate this at $z = 0$, and use $u(0) = 0$, to get

$$u_j^{(k)}(0) = q_k(\partial^\beta f_j(0), u^{(\ell)}(0)) = Q_{j,k}(\partial^\beta f(0)), \tag{27}$$

with $Q_{j,k}$ a multivariate polynomial having nonnegative coefficients. Note that the arguments of $Q_{j,k}$ are all components of all $\partial^\beta f(0)$ with $|\beta| \leq k-1$.

Having found that the derivatives of u at 0 is given by a positive coefficient polynomial of the derivatives of f at 0, we would like to replace f by a simpler majorant of it. We say G *majorizes* g at $c \in \mathbb{C}^n$, if

$$|\partial^\alpha g(c)| \leq \partial^\alpha G(c), \qquad \text{for all } \alpha. \tag{28}$$

In other words, the Taylor series coefficients of g at c is bounded in magnitude by the corresponding coefficients of G. Since our right hand side f is componentwise analytic at 0, there exist constants $M > 0$ and $r > 0$ such that

$$\frac{|\partial^\alpha f_j(0)|}{\alpha!} \leq \frac{M}{r^{|\alpha|}}, \qquad \text{for all } \alpha, \quad \text{and all } j. \tag{29}$$

Exercise 14. Show that any of the functions

$$\begin{aligned}
F_j(z,v) &= \frac{M}{(1-z/r)(1-v_1/r)\cdots(1-v_m/r)},\\
F_j(z,v) &= \frac{M}{(1-z/r)(1-(v_1+\ldots+v_m)/r)},\\
F_j(z,v) &= \frac{M}{1-(z+v_1+\ldots+v_m)/r},\\
F_j(z,v) &= \frac{M}{1-(z/\rho+v_1+\ldots+v_m)/r}, \qquad \text{with a constant } \rho\in(0,1],
\end{aligned} \tag{30}$$

majorizes f_j at 0.

Let us consider the system

$$U_j' = F_j(z,U), \qquad U_j(0)=0, \qquad j=1,\ldots,m, \tag{31}$$

with F_j being a majorant of f_j. Using the positivity of the coefficients of $Q_{j,k}$, we get

$$|\partial^\alpha u_j(0)| = |Q_{j,k}(\partial^\beta f(0))| \leq Q_{j,k}(|\partial^\beta f(0)|) \leq Q_{j,k}(\partial^\beta F(0)) = \partial^\alpha U_j(0), \tag{32}$$

i.e., U_j majorizes u_j at 0. To establish analyticity of u_j, it only remains to solve (31) in analytic functions. Given the supply of majorants (30), it is not hard. For example, choosing F_j to be the second function in (30), and putting $U_1=\ldots=U_m$, we get

$$U_j(z) = \frac{r}{m}\left(1-\sqrt{1+2mM\log\left(1-\frac{z}{r}\right)}\right), \tag{33}$$

which is obviously analytic at 0. We have proved the following.

Theorem 15. *Consider the initial value problem*

$$u_j' = f_j(z,u), \qquad u_j(0)=e_j, \qquad j=1,\ldots,m. \tag{34}$$

Let $f_j:\mathbb{C}^{n+m}\to\mathbb{C}$ be analytic at 0, for each j. Then there exists a unique solution u that is analytic at 0.

This theorem can easily be generalized to higher order quasilinear equations, and in a certain sense to any ODE system that can be solved. The most general form of an ODE system for the unknown function $u:\mathbb{C}\to\mathbb{C}^m$ can be written as

$$F_i(z,u,u',u'',\ldots)=0, \qquad i=1,\ldots,m. \tag{35}$$

Suppose that this can be written in the form

$$u_i^{(q_i)} = f_i(z,u,u',\ldots), \qquad i=1,\ldots,m, \tag{36}$$

where for each i and j, the function f_i depends on the derivatives of u_j only up to order q_j-1. In other words, we solve for the highest over derivatives of each component of u. Note that $u_i^{(q_i)}$ in (36) is *not* necessarily the highest order derivative of u_i in (35). For example, consider the system

$$u''(v''+1)=0, \qquad u'=v'+z. \tag{37}$$

It looks like the system is second order in both u and v, so that the general solution involves 4 arbitrary constants. But one cannot be sure. Namely, differentiating $u'=v'+z$ gives $u''=v''+1$, and substituting this into $u''(v''+1)=0$ we get $u''=0$ and $v''=-1$. So $u(z)=az+b$ and $v(z)=-z^2/2+cz+d$, with constants a, b, c, and d. However, the equation $u'=v'+z$ fixes $a=c$, hence we have only 3 arbitrary constants. From what we have just discussed, the system can be written in the form (36), as

$$u''=0, \qquad v'=u'-z, \tag{38}$$

which makes it clear that we need 3 arbitrary constants. It was observed by Carl Gustav Jacob Jacobi that the key to bringing order to the general system (35) is to somehow write it in the form (36), which was called by him the *normal form*. General methods to do such transformations that often work include using the implicit function theorem, and differentiating the equation (35) with respect to the independent variable. As soon as one brings the system into the normal form, we have local existence.

Corollary 16. *Consider the system* (36) *with the initial conditions*

$$u_i^{(k)}(0) = e_{j,k}, \quad k = 0, \ldots, q_i - 1, \qquad i = 1, \ldots, m. \tag{39}$$

Suppose that for each i, the function f_i is analytic at 0 in all its arguments. Then there exists a unique solution u that is analytic at 0.

Exercise 17. Prove this corollary.

4. Partial differential equations

Along the same lines, one can establish the local existence of analytic solutions to a very general class of systems of partial differential equations. Such a result was proved by Augustin-Louis Cauchy in 1842 on first order quasilinear evolution equations, and formulated in its most general form by Sofia Vasilyevna Kovalevskaya in 1874. At about the same time, Gaston Darboux also reached similar results, although with less generality than Kovalevskaya's work. Both Kovalevskaya's and Darboux's papers were published in 1875, and the proof was later simplified by Édouard Jean-Baptiste Goursat in his influential calculus texts around 1900. Nowadays these results are collectively known as the *Cauchy-Kovalevskaya theorem.* The most basic form of such, from which all more general forms can be deduced, is as follows.

Theorem 18. *Consider the Cauchy (or initial value) problem*

$$\partial_n u_j = f_j(z, u, \partial_1 u, \ldots, \partial_{n-1} u), \qquad u_j(\zeta, 0) = 0, \quad \zeta \in \mathbb{C}^{n-1}, \qquad j = 1, \ldots, m. \tag{40}$$

Let $f_j : \mathbb{C}^{n+m+(n-1)\times m} \to \mathbb{C}$ be analytic at 0, for all j. Then there exists a unique solution u that is analytic at 0.

Proof. Without loss of generality, we can assume that $f_j(0) = 0$ for all j, by replacing $u(z)$ by $u(z) - z_n \partial_n u(0)$. Also, it will be convenient to label by p_{ik} the slot of f_j that takes $\partial_i u_k$ as its argument, i.e., $f_j = f_j(z, u, p)$ with $z \in \mathbb{C}^n$, $u \in \mathbb{C}^m$, and $p \in \mathbb{C}^{(n-1)\times m}$. Since the initial condition is identically zero, we have

$$\partial^\alpha u(0) = 0, \qquad \text{if} \quad \alpha_n = 0. \tag{41}$$

The derivatives $\partial^\alpha u$ with $\alpha_n > 0$ can be found by differentiating the equation (40). For example, we have

$$\partial_k \partial_n u_j = \frac{\partial f_j}{\partial z_k} + \frac{\partial f_j}{\partial u_q} \frac{\partial u_q}{\partial z_k} + \frac{\partial f_j}{\partial p_{iq}} \frac{\partial^2 u_q}{\partial z_i \partial z_k},$$

where in implicit summations over $q = 1, \ldots, m$, and $i = 1, \ldots, n-1$, are assumed in the terms they appear. In general, for α with $\alpha_n > 0$, we have

$$\partial^\alpha u_j = q_\alpha(\partial^\beta f_j, \partial^\gamma u), \tag{42}$$

where q_α is a polynomial with nonnegative coefficients, depending on $\partial^\beta f_j$ with $|\beta| \leq |\alpha| - 1$, and $\partial^\gamma u$ with $|\gamma| \leq |\alpha|$ and $\gamma_n \leq \alpha_n - 1$. Exactly as before, we can eliminate the terms $\partial^\gamma u$ and evaluate the result at 0 to get

$$\partial^\alpha u_j(0) = q_\alpha(\partial^\beta f_j(0), \partial^\gamma u(0)) = Q_{j,\alpha}(\partial^\beta f(0)), \tag{43}$$

where $Q_{j,\alpha}$ is a polynomial with nonnegative coefficients, depending on $\partial^\beta f_j$ with $|\beta| \leq |\alpha|-1$. Now it is time to consider the system

$$\partial_n U_j = F_j(z, U, \partial_1 U, \ldots, \partial_{n-1} U), \qquad j = 1, \ldots, m, \tag{44}$$

with F_j majorizing f_j at 0 for each j. Then for all multi-indices α with $\alpha_n > 0$, we have

$$|\partial^\alpha u_j(0)| = |Q_{j,\alpha}(\partial^\beta f(0))| \leq Q_{j,\alpha}(|\partial^\beta f(0)|) \leq Q_{j,\alpha}(\partial^\beta F(0)) = \partial^\alpha U_j(0). \tag{45}$$

If in addition, $U_j|_{z_n=0}$ majorizes 0 as a function of $(z_1, \ldots, z_{n-1}) \in \mathbb{C}^{n-1}$ at 0, then U_j majorizes u_j as a function of $z \in \mathbb{C}^n$ at 0.

Supposing that f_j satisfies the bound (29), let us try the following majorant of f_j.

$$F_j(z, u, p) = \frac{M}{\left(1 - \frac{z_1 + \ldots + z_{n-1} + z_n/\rho + u_1 + \ldots + u_m}{r}\right)\left(1 - \frac{1}{r}\sum_{i,k} p_{ik}\right)} - M, \tag{46}$$

where $\rho \in (0, 1]$ is a constant whose value is to be adjusted later. Put $s = z_1 + \ldots + z_{n-1}$, $t = z_n$, and $v := U_1 = \ldots = U_m$, to get

$$\partial_t v = \frac{M}{\left(1 - \frac{s + t/\rho + mv}{r}\right)\left(1 - \frac{(n-1)m}{r}\partial_s v\right)} - M. \tag{47}$$

Defining the new variable $\sigma = t + \rho s$, and assuming v depends only on σ, this becomes

$$\partial_\sigma v = \frac{M}{\left(1 - \frac{\sigma/\rho + mv}{r}\right)\left(1 - \frac{(n-1)m\rho}{r}\partial_\sigma v\right)} - M, \tag{48}$$

or, after rearranging

$$\left(1 - \frac{(n-1)mM\rho}{r}\right)\partial_\sigma v - \frac{(n-1)m\rho}{r}(\partial_\sigma v)^2 = \frac{M}{1 - \frac{\sigma/\rho + mv}{r}} - M. \tag{49}$$

We choose $\rho \in (0, 1]$ so small that $1 - \frac{(n-1)mM\rho}{r} > 0$. Then the preceding equation can be solved for $\partial_\sigma v$, in the power series

$$\partial_\sigma v = c_1(\sigma/\rho + mv) + c_2(\sigma/\rho + mv)^2 + \ldots, \tag{50}$$

convergent for some $\sigma/\rho + mv \neq 0$, with all coefficients nonnegative: $c_k \geq 0$. In other words, there is a function g analytic at 0, with nonnegative Maclaurin series coefficients and with $g(0) = 0$, such that

$$\partial_\sigma v = g(\sigma/\rho + mv). \tag{51}$$

Now we can apply Cauchy's theorem for analytic ODEs from the previous section to infer that the above equation has a solution v analytic at 0, satisfying $v(0) = 0$, whose Maclaurin series coefficients are nonnegative. Rewinding everything, the vector function U with components

$$U_j(z) = v(\rho(z_1 + \ldots + z_{n-1}) + z_n), \tag{52}$$

solves (44). Since the Maclaurin series coefficients of v are nonnegative, the same holds for U_j, implying that $U_j|_{z_n=0}$ majorizes 0 at 0. This establishes the proof. □

Remark 19. The choice of our majorant in this proof is one introduced by Goursat, that allows us to treat the right hand side f directly. An alternative proof can be obtained by differentiating the equation and transforming it into a quasilinear form.

It is easy to generalize the preceding theorem to higher order equations, and to solve them in a neighbourhood of an open subset of the hyperplane $\{z_n = 0\}$.

Corollary 20. *Consider the equations*

$$\partial_n^{q_i} u_i = f_i(z, u, u', \ldots), \qquad i = 1, \ldots, m, \tag{53}$$

where for each i and j, the function f_i depends on the derivatives of u_j only up to order q_j, is independent of $\partial_n^{q_i} u_i$, and analytic in all its arguments. Furthermore, consider the Cauchy problem of finding a solution to (53) with the prescribed initial values

$$\partial_n^k u_i(\zeta, 0) = \psi_{i,k}(\zeta), \quad \zeta \in \Omega, \quad k = 0, \ldots, q_i - 1, \qquad i = 1, \ldots, m, \tag{54}$$

where $\Omega \subseteq \mathbb{C}^{n-1}$ is open, and all $\psi_{i,k}$ are analytic on Ω. Then there exists a unique solution u to the Cauchy problem that is analytic on an open set of $\mathbb{C}^n$ containing $\Omega \times \{0\}$.

Proof. First, let us transform the system to the first order form amenable to Theorem 18. To this end we introduce the new variables $u_{\alpha,i} = \partial^\alpha u_i$ for $|\alpha| \leq q_i - 1$. Then the equations (53) become

$$\partial_n u_{(0,\ldots,0,q_i-1),i} = f_i(z, u, \partial_1 u, \ldots, \partial_{n-1} u), \qquad i = 1, \ldots, m, \tag{55}$$

where u now denotes the collection of all $u_{\alpha,i}$. We need equations for $\partial_n u_{\alpha,i}$ with $|\alpha| \leq q_i - 1$ and $\alpha_n < q_i - 1$. For $|\alpha| \leq q_i - 2$, we simply use the definitions

$$\partial_n u_{\alpha,i} = u_{(\alpha_1,\ldots,\alpha_{n-1},\alpha_n+1),i}. \tag{56}$$

For $|\alpha| = q_i - 1$, we necessarily have $\alpha_\ell > 0$ for some $\ell \neq n$, because of $\alpha_n < q_i - 1$. So we can use the equation

$$\partial_n u_{\alpha,i} = \partial_\ell u_{(\ldots,\alpha_\ell-1,\ldots,\alpha_n+1),i}. \tag{57}$$

The initial condition for $u_{\alpha,i}$ is obtained from that of $\partial_n^{\alpha_n} u_i$ by applying the "spatial" differential operator $\partial_1^{\alpha_1} \cdots \partial_{n-1}^{\alpha_{n-1}}$. So the system is reduced to the first order form, and for each $\zeta \in \Omega \times \{0\}$, Theorem 18 guarantees a nonempty ball B_ζ centered at ζ, such that a unique analytic solution u_ζ exists on B_ζ. If $B_\zeta \cap B_{\zeta'} \neq \varnothing$ for some ζ and ζ', then there is $\zeta'' \in B_\zeta \cap B_{\zeta'} \cap (\Omega \times \{0\})$, hence by uniqueness of the solution to the Cauchy problem, u_ζ and $u_{\zeta'}$ must coincide on $B_{\zeta''}$. We conclude that the collection of all u_ζ defines a single analytic function on $\bigcup_{\zeta \in \Omega \times \{0\}} B_\zeta$ that solves our equations. □

Each of the conditions in the preceding theorem can be shown to be necessary for the conclusion to be valid. At this point, the only item that needs explanation is the condition on the allowed derivatives of u appearing in the right hand side. That this condition is necessary is best illustrated by the following simple counterexample due to Kovalevskaya.

Example 21. Consider the heat equation

$$\partial_t u = \partial_x^2 u, \tag{58}$$

to be solved in a neighbourhood of the origin in $(x, t) \in \mathbb{C}^2$, with an analytic initial datum $u(x, 0)$ prescribed on the line $\{t = 0\}$. Differentiating the equation with respect to t gives

$$\partial_t \partial_t u = \partial_t \partial_x^2 u = \partial_x^2 \partial_t u = \partial_x^4 u, \tag{59}$$

and by repeated differentiations, we get

$$\partial_t^k u = \partial_x^{2k} u, \qquad \Rightarrow \qquad \partial_t^k u(0,0) = \partial_x^{2k} u(0,0). \tag{60}$$

The strongest bounds on $\partial_x^{2k} u(0,0)$ for general analytic initial data $u(x, 0)$ are of the form $M(2k)!/r^k$. On the other hand, in order for u to be analytic in the t-direction, the derivatives $\partial_t^k u(0,0)$ must necessarily have a bound of the form $Ck!/\rho^k$. The moral of the story is that by equating more spatial derivatives on the right hand side with less time derivatives on the left hand side, we generate faster growth in the right hand side than is allowed for the left hand side to be analytic.

Exercise 22. Cook up an initial datum $u(x,0)$ for the heat equation that is analytic for all x such that the function $u(0,t)$ is not analytic at $t=0$.

5. Characteristic surfaces

In this section we discuss how one can adapt the Cauchy-Kovalevskaya theorem if one were to specify Cauchy data on a general analytic surface. Since the theorem concerned is a local result, local considerations will suffice. So locally, an analytic surface is the zero level set of an analytic function. More precisely, $S\subset\mathbb{C}^n$ is an analytic surface if there is an analytic function $\varphi:U\to\mathbb{C}$ with U an open subset of $\mathbb{C}^n$, such that $S=\{z\in U:\varphi(z)=0\}$ and $\partial\varphi=(\partial_1\varphi,\ldots,\partial_n\varphi)$ is nonzero on S. In order to specify Cauchy data on S, we assume that there is an analytic injection $w:U\to\mathbb{C}^n$ with $w_n\equiv\varphi$, i.e., that there is an analytic coordinate system $(w_1,\ldots,w_n)$ in a neighbourhood of S, that makes $S=\{w_n=0\}$. This is always possible locally at any given point $z\in S$, by shrinking the neighbourhood U if necessary. For example, it suffices to take a rectilinear coordinate system with its n-th axis having the same direction as the normal of S at z, and then adjust the n-th coordinate so that S becomes $\{w_n=0\}$. The approach we take in this section is to specify the Cauchy data on S in the w-coordinate system. Then since in the w-coordinates S is just $\{w_n=0\}$, the Cauchy-Kovalevskaya theorem readily applies, provided that the equation can be solved for the term $\partial^q_{w_n}u$. Looking at what this tells us in the original z-coordinates, we will obtain an important insight on what type of initial surfaces the equation “prefers”.

For simplicity, let us consider the q-th order *semilinear equation*

$$\sum_{|\alpha|=q}A_\alpha(z)\partial^\alpha u+g\left(z,\{\partial^\alpha u\}_{|\alpha|<q}\right)=0. \tag{61}$$

Denote by B_α the coefficients of the q-th order derivatives in w-coordinates, i.e.,

$$\sum_{|\alpha|=q}A_\alpha(z)\partial_z^\alpha u=\sum_{|\alpha|=q}B_\alpha(w)\partial_w^\alpha u+\text{lower order terms}. \tag{62}$$

What is important for us is the particular coefficient B_{α^*} with $\alpha^*=(0,\ldots,0,q)$, because if, say, $B_{\alpha^*}(w)\neq0$, we can solve for the term $\partial^q_{w_n}u$ in a neighbourhood of w, and therefore can apply the Cauchy-Kovalevskaya theorem at w. Considerations such as

$$\frac{\partial u}{\partial z_i}=\frac{\partial u}{\partial w_k}\frac{\partial w_k}{\partial z_i},\qquad\frac{\partial^2u}{\partial z_i\partial z_j}=\frac{\partial^2u}{\partial w_k\partial w_l}\frac{\partial w_k}{\partial z_i}\frac{\partial w_l}{\partial z_j}+\frac{\partial u}{\partial w_k}\frac{\partial^2 w_k}{\partial z_i\partial z_j}, \tag{63}$$

imply that

$$B_{\alpha^*}=\sum_{|\alpha|=q}A_\alpha\left(\frac{\partial w_n}{\partial z_1}\right)^{\alpha_1}\cdots\left(\frac{\partial w_n}{\partial z_n}\right)^{\alpha_n}. \tag{64}$$

Shifting back to z-coordinates and using the notation $\varphi\equiv w_n$, we see that if

$$\sum_{|\alpha|=q}A_\alpha\left(\frac{\partial \varphi}{\partial z_1}\right)^{\alpha_1}\cdots\left(\frac{\partial \varphi}{\partial z_n}\right)^{\alpha_n}\neq0, \tag{65}$$

at some point $z\in S$, then the Cauchy problem with initial data on S is locally solvable at z.

It is a good time to introduce some terminologies. The function

$$C(z,\xi)=\sum_{|\alpha|=q}A_\alpha(z)\xi^\alpha\equiv\sum_{|\alpha|=q}A_\alpha(z)\xi_1^{\alpha_1}\cdots\xi_n^{\alpha_n}, \tag{66}$$

defined for $z\in U$ and $\xi\in\mathbb{C}^n$, is called the *characteristic form* of the equation (61). It is a homogeneous function of degree q in ξ, i.e.,

$$C(z,\lambda\xi)=\lambda^qC(z,\xi),\qquad\lambda\in\mathbb{C}. \tag{67}$$

In terms of the characteristic form, the condition (65) becomes

$$C(z, \partial\varphi(z)) \neq 0. \tag{68}$$

If $C(z, \partial\varphi(z)) = 0$ for $z \in S$, then S is said to be *characteristic at* z to the equation (61). If S is characteristic at each of its points, it is called a *characteristic surface* of (61). The Cauchy problem for (61) has a unique analytic solution near S, if S is nowhere characteristic.

We can also introduce the *characteristic cone at* $z \in U$ as

$$C_z(A) = \{\xi : \ C(z, \xi) = 0\}, \tag{69}$$

where A is to be understood as the collection of the coefficients A_α, ($|\alpha| = q$), or the differential operator associated to those coefficients. Then a surface is characteristic at a point if the normal to the surface at that point belongs to the characteristic cone at the same point.

Let us note that the discussions above can be extended without much difficulty to nonlinear equations, where whether or not a surface is characteristic now may depend on what function we plug into the differential operator. For more details we recommend Jeffrey Rauch's book.

Example 23. The characteristic form of the Laplace operator is

$$C(z, \xi) = \sum_{i=1}^{n} \xi_i^2.$$

There is no nonzero real vector $\xi \in \mathbb{R}^n$ that makes $C(z, \xi) = 0$, so the generators of the characteristic cones cannot be parallel to any real vector. Let us denote the characteristic cone by $C_z(\Delta_n)$, which is of course independent of z. To reiterate, we have $C_z(\Delta_n) \cap \mathbb{R}^n = \{0\}$, and any real surface cannot be characteristic to the Laplace equation. Equations without real characteristic surfaces are called *elliptic equations*. The cone $C_z(\Delta_n)$ can easily be described as a whole as an object in $\mathbb{C}^n$, but the most relevant to us is the behaviour of the cone on the hyperplanes $\mathbb{R}^n$ and $\mathbb{R}^{n-1} \times i\mathbb{R}$. The former is trivial and has just been discussed. For the latter, it is convenient to make the substitution $z_n \mapsto iz_n$, called the *Wick rotation*, under which the Laplace equation becomes the wave equation, and the set $\mathbb{R}^{n-1} \times i\mathbb{R}$ becomes $\mathbb{R}^n$. For the wave equation, we have

$$C(z, \xi) = -\xi_n^2 + \sum_{i=1}^{n-1} \xi_i^2.$$

Restricting every variable to the reals, the characteristic cone in this case is called the *light cone*, and any surface whose normal makes an angle $\pi/4$ with the direction of z_n is a characteristic surface.

The heat and Schrödinger equations transform into each other by Wick rotations. The both equations have

$$C(z, \xi) = \sum_{i=1}^{n-1} \xi_i^2,$$

as their characteristic form, and the characteristic cone is exactly $C_z(\Delta_{n-1}) \times \mathbb{C}$. Restricted to the reals, the characteristic cone is the vertical line $\{\xi : \xi_1 = \ldots = \xi_{n-1} = 0\}$, and so the characteristic surfaces are the horizontal planes $\{x : x_n = \text{const}\}$. ⊘

Exercise 24. For each of the following cases, determine the characteristic cones and characteristic surfaces, restricted to the reals.

a) Tricomi-type equation: $u_{xx} + yu_{yy} = 0$.
b) Wave equation with wave speed $c > 0$: $u_{xx} + u_{yy} + u_{zz} = c^{-2}u_{tt}$. How many regions does the characteristic cone divide $\mathbb{R}^4$ into?

c) Ultrahyperbolic "wave" equation: $u_{xx} + u_{yy} = u_{zz} + u_{tt}$. How many regions does the characteristic cone divide $\mathbb{R}^4$ into?
d) Linear transport equation: $\sum_{i=1}^{n} \alpha_i(x)\partial_i u = 0$.

As a prototypical example of what happens when one tries to prescribe initial data on a characteristic surface, let us look at the linear transport equation from part d) of the preceding exercise. There it is found that if S is characteristic at $x \in S$, then the vector $\alpha(x)$ is tangent to S. Let us assume that S is everywhere characteristic. Then all our transport equation tells us is the behaviour of u along S, and what u does in the transversal direction is completely "free". This means that the existence is lost unless the initial condition on S satisfies certain constraints, and if a solution exists, it will not be unique. The situation is entirely analogous to solving the linear system $Ax = b$ with a non-invertible square matrix A. Now let us forget about specifying initial conditions and take a slightly different point of view. Imagine that the graphs of several solutions to the transport equation are drawn in $\mathbb{R}^{n+1}$, and imagine also several surfaces in $\mathbb{R}^n$, which is to be understood as the base of the space $\mathbb{R}^{n+1}$ in which the graphs live. Then we see that the characteristic surfaces are the only surfaces along which two different solutions can touch each other, for if two solutions are the same on a non-characteristic surface, by uniqueness they must coincide in a neighbourhood of the surface.

6. Well posedness and basic classifications

The complex analytic setting is completely natural for the Cauchy-Kovalevskaya theorem. This is because any real analytic function uniquely extends to a complex analytic one in a neighbourhood of $\mathbb{R}^n$ considered as a subset of $\mathbb{C}^n$, and more importantly this point of view offers a better insight on the behaviour of analytic functions. Hence the complex analytic treatment contains the real analytic case as a special case. However, it is known that if we allowed only analytic solutions, we would be missing out on most of the interesting properties of partial differential equations. For instance, since analytic functions are completely determined by its values on any open set however small, it would be extremely cumbersome, if not impossible, to describe phenomena like wave propagation, in which initial data on a region of the initial surface are supposed to influence only a specific part of space-time. A much more natural setting for a differential equation would be to require its solutions to have just enough regularity for the equation to make sense. For example, the Laplace equation $\Delta u = 0$ already makes sense for twice differentiable functions. Actually, the solutions to the Laplace equation, i.e., harmonic functions, are automatically analytic, which has a deep mathematical reason that could not be revealed if we restricted ourselves to analytic solutions from the beginning. In fact, the solutions to the Cauchy-Riemann equations, i.e., holomorphic functions, are analytic by the same underlying reason, and complex analytic functions are nothing but functions satisfying the Cauchy-Riemann equations. From this point of view, looking for analytic solutions to a PDE in $\mathbb{R}^n$ would mean coupling the PDE with the Cauchy-Riemann equations and solving them simultaneously in $\mathbb{R}^{2n}$. In other words, if we are not assuming analyticity, $\mathbb{C}^n$ is better thought of as $\mathbb{R}^{2n}$ with an additional algebraic structure. Hence the real case is more general than the complex one, and from now on, we will be working explicitly in *real* spaces such as $\mathbb{R}^n$, unless indicated otherwise.

As soon as we allow non-analytic data and/or solutions, many interesting questions arise surrounding the Cauchy-Kovalevskaya theorem. First, assuming a setting to which the Cauchy-Kovalevskaya theorem can be applied, we can ask if there exists any (necessarily non-analytic) solution other than the solution given by the Cauchy-Kovalevskaya theorem. In other words, is the uniqueness part of the Cauchy-Kovalevskaya theorem still valid if we now allow non-analytic solutions? For linear equations an affirmative answer is given by *Holmgren's uniqueness theorem*. Moreover, uniqueness holds for first order equations, but fails in general for higher order equations and systems. Such a uniqueness result can also be thought of as a regularity theorem, in the sense that if u is a solution then it would be automatically analytic by uniqueness.

The second question is whether existence holds for non-analytic data, and again the answer is negative in general. A large class of counterexamples can be constructed, by using the fact that some equations, such as the Laplace and the Cauchy-Riemann equations, have only analytic solutions, therefore their initial data, as restrictions of the solutions to an analytic hypersurface, cannot be non-analytic. Hence such equations with non-analytic initial data do not have solutions. In some cases, this can be interpreted as one having "too many" initial conditions that make the problem overdetermined, since in those cases the situation can be remedied by removing some of the initial conditions. For example, with sufficiently regular closed surfaces as initial surfaces, one can remove either one of the two Cauchy data in the Laplace equation, arriving at the Dirichlet or Neumann problem. Starting with Hans Lewy's celebrated counterexample of 1957, more complicated constructions along similar lines have been made that ensure the inhomogeneous part of a linear equation to be analytic, thus exhibiting examples of linear equations with no solutions when the inhomogeneous part is non-analytic, regardless of initial data. The lesson to be learned from these examples is that the existence theory in a non-analytic setting is much more complicated than the corresponding analytic theory, and in particular one has to carefully decide on what would constitute the initial data for the particular equation.

Indeed, there is an illuminating way to detect the poor behaviour of some equations discussed in the previous paragraph with regard to the Cauchy problem, entirely from within the analytic setting, that runs as follows. Suppose that in the analytic setting, for a generic initial datum ψ it is associated the solution $u = S(\psi)$ of the equation under consideration, where $S : \psi \mapsto u$ is the solution map. Now suppose that the datum ψ is non-analytic, say, only continuous. Then by the Weierstrass approximation theorem, for any $\varepsilon > 0$ there is a polynomial ψ_ε that is within an ε distance from ψ. Taking some sequence $\varepsilon \to 0$, if the solutions $u_\varepsilon = S(\psi_\varepsilon)$ converge locally uniformly to a function u, we could reasonably argue that u is a solution (in a generalized sense) of our equation with the (non-analytic) datum ψ. The counterexamples from the preceding paragraph suggest that in those cases the sequence u_ε cannot converge. Actually, the situation is much worse, as the following example due to Jacques Hadamard shows.

Example 25. Consider the Cauchy problem for the Laplace equation

$$u_{tt} + u_{xx} = 0, \qquad u(x, 0) = a_\nu \sin \nu x, \qquad u_t(x, 0) = b_\nu \sin \nu x, \tag{70}$$

whose solution is given by

$$u(x, t) = (a_\nu \cosh \nu t + \frac{b_\nu}{\nu} \sinh \nu t) \sin \nu x. \tag{71}$$

Choosing, e.g., $a_\nu = 1/\nu$ and $b_\nu = 0$ with ν large, we see that the solution grows arbitrarily fast, although the initial data are arbitrarily small. In a certain sense, the relation between the solution and the Cauchy data becomes more and more difficult to invert as we add higher and higher frequencies. The initial data could be, for instance, the error of an approximation of non-analytic initial data in the uniform norm, with the approximation getting better as $\nu \to \infty$. Then the solutions with initial data given by the approximations diverge unless a_ν and b_ν decay faster than exponential. But functions that can be approximated by analytic functions with such small errors form a severely restricted class, being between the smooth functions C^∞ and the analytic functions C^ω.

Exercise 26. Consider the problem

$$u_{tt} + u_{xx} = 0, \qquad u(x,0) = \phi(x), \qquad u_t(x,0) = \psi(x).$$

For given $\varepsilon > 0$ and an integer $k > 0$, construct initial data ϕ and ψ such that

$$\|\phi\|_\infty + \ldots + \|\phi^{(k)}\|_\infty + \|\psi\|_\infty + \ldots + \|\psi^{(k)}\|_\infty < \varepsilon,$$

and

$$\|u(\cdot,\varepsilon)\|_\infty > 1/\varepsilon.$$

Repeat the exercise with the condition on the initial data replaced by $\|\phi^{(i)}\|_\infty + \|\psi^{(i)}\|_\infty \leq \varepsilon$ for all $i = 0, 1, \ldots$.

Let us contrast the preceding example with the following.

Example 27. Consider the Cauchy problem for the wave equation

$$u_{tt} - u_{xx} = 0, \qquad u(x,0) = \phi(x), \qquad u_t(x,0) = \psi(x), \tag{72}$$

whose solution is given by the d'Alambert formula

$$u(x,t) = \frac{\phi(x-t) + \phi(x+t)}{2} + \frac{1}{2}\int_{x-t}^{x+t} \psi(y)\,\mathrm{d}y, \tag{73}$$

where if $t < 0$, the integral over $(x-t, x+t)$ is understood to be the minus of the integral over $(x+t, x-t)$. From this, it is easy to deduce the bound

$$|u(x,t)| \leq \sup_{y\in[x-t,x+t]} |\phi(y)| + |t| \sup_{y\in[x-t,x+t]} |\psi(y)|, \tag{74}$$

making it clear that small initial data lead to small solutions. Moreover, one can show uniqueness by an energy argument, meaning that the solution given by the d'Alambert formula is the only one.

Triggered by considerations such as the preceding ones, Hadamard introduced the concept of *well-posedness* of a problem. To define this concept abstractly, we assume a set D, that represents all possible *data* in the problem, a second set S, that represents all possible *solutions*, and finally a *relation* $R(f,u) \in \{0,\ldots\}$, defined for $f \in D$ and $u \in S$. Then we consider the following *problem*: Given $f \in D$, find $u \in S$ such that $R(f,u) = 0$. This problem is said to be *well-posed* if

- For any $f \in D$ there exists a unique solution $u \in S$, and
- Varying f a bit results in a small variation of u, i.e., u depends on f continuously.

In order make the second point precise, we need to define what we mean by continuity of maps $\sigma : D \to S$, i.e., we need to choose topologies for the sets D and S. We can introduce some flexibility on the choice of the sets D and S too, leading to the meta-problem: Find "reasonable" topological spaces D and S such that the problem $R(f,u) = 0$ with $f \in D$ and $u \in S$ is well-posed. Usually, the "correct" topologies on D and S are suggested by the structure of the problem itself, or what is essentially the same, by the real world or mathematical phenomenon the problem is supposed to model. The concept of well-posedness has proved to be very useful in revealing the true nature of the equations, especially in identifying the "correct" initial and/or boundary conditions. Of course, one important motivation of the well-posedness concept is that in modelling of real world phenomena, the problem data always have some measurement or computational error in it, so without well-posedness, we cannot say that the solution corresponding to imprecise data is anywhere near the solution we are trying to capture. Thus, a necessary condition for a physics theory to have any predictive power is that it must produce well-posed problems. One might wonder if a counterexample to this statement can be exhibited by mentioning the fact that in practice people routinely solve what are normally considered as *ill-posed* problems, i.e., problems that are not well-posed.

However, in those situations "solving a problem" has a broader meaning, and as part of this process one replaces the original ill-posed problem by a well-posed one, with the aid of a *regularization* procedure. For example, from Hadamard's example we have seen that essentially the "trouble makers" are initial data that oscillate rapidly in space, and a bit more analysis shows that if the initial data has frequencies not exceeding ν, then the Cauchy problem can be solved without trouble for time of order $1/\nu$. This offers a good theory *provided* that in the particular situation under consideration, we know for sure there will not be frequencies higher than ν present in any realistic initial data, *and* we do not need to solve the Cauchy problem for time intervals much longer than $1/\nu$.

Finally, in this and the following paragraphs we say a few disconnected words on classifications of PDEs. Roughly speaking, the *hyperbolic*, *elliptic*, *parabolic*, and *dispersive* classes arise as one tries to identify the equations that are similar to, and therefore can be treated by extensions of techniques developed for, the wave (and transport), the Laplace (and the Cauchy-Riemann), the heat, and the Schrödinger equations, respectively. Indeed, the idea of *hyperbolicity* is an attempt to identify the class of PDE's for which the Cauchy-Kovalevskaya theorem can be rescued in some sense when we relax the analyticity assumption. The simplest examples of hyperbolic equations are the wave and transport equations. In contrast, trying to capture the essence of the poor behaviour of the Laplace and Cauchy-Riemann equations in relation to their Cauchy problems leads to the concept of *ellipticity*. Hallmarks of elliptic equations are having no real characteristic surfaces, smooth solutions for smooth data, overdeterminacy of the Cauchy data hence boundary value problems, and being associated to stationary phenomena.

The general form of q-th order *quasilinear* equation is

$$\sum_{|\alpha|=q} A_\alpha \partial^\alpha u + g(x, \{\partial^\beta u\}_{|\beta|<q}) = 0, \tag{75}$$

where $A_\alpha = A_\alpha(x, \{\partial^\beta u\}_{|\beta|<q})$, i.e., the coefficients A_α depend only on the lower order derivatives of u. More restricted classes are

- *Semilinear* equations, where $A_\alpha = A_\alpha(x)$, i.e., the coefficients A_α do not depend on the unknown solution u.
- *Linear* equations, where $A_\alpha = A_\alpha(x)$ and g is linear in the unknown solution u.
- *Constant coefficient linear* equations: $A_\alpha(x) = \text{const}$ and similarly all the coefficients of g (which is linear in u) are constant. The term *variable coefficient* is used to explicitly indicate that the coefficients are possibly not constant.