From Tangent Lines to Characteristic Curves

An expository account for students of Mathematics

V. P. Anoop

Department of Mathematical Sciences
Kannur University
anoopvp@kannuruniv.ac.in

These notes were prepared as part of the Meenakshisundaram–Patodi Lecture, delivered at the Department of Mathematics, MES Mampad College, on 14 August 2026.

Abstract

This article is an invitation, not a lecture note. We begin with something every student meets in the first semester — the tangent line to a graph — and follow a single idea as far as it will take us. That idea is: *a derivative is a linear map, and a linear map is completely determined by what it does in a few directions.* Following it faithfully leads us out of one-variable calculus, through the total derivative and directional derivatives, into ordinary differential equations, and finally to partial differential equations, where the same idea reappears in geometric dress as the *method of characteristics*. Along the way, almost every statement is accompanied by a picture, because the whole story is really a story about geometry.

Contents

1 Prologue: one idea, told four times

A student learning calculus is told that $f'(a)$ is "the slope of the tangent line". A student learning several–variable calculus is told that the derivative is "the Jacobian matrix". A student learning differential equations is told to "solve along characteristics". These sound like three unrelated pieces of folklore. They are not. They are one statement, repeated at increasing levels of generality:

> *To differentiate is to replace a complicated object by the best linear object near a point; and a linear object is known once you know it along enough directions.*

The plan of the article is to make that sentence precise four times over.

(i) In §2 we rewrite the familiar definition of $f'(a)$ so that the derivative appears not as a *number* but as a *linear map* $\mathbb{R} \to \mathbb{R}$. Nothing is gained computationally; a great deal is gained conceptually.

(ii) In §3 we copy that rewritten definition word for word into $\mathbb{R}^n \to \mathbb{R}^m$. This is Rudin's definition of the total derivative.

(iii) In §§4–5 we ask what the linear map does along a single direction. This produces directional derivatives, and then the special ones — partial derivatives — which turn out to carry all the information.

(iv) In §§6–9 we put derivatives inside equations. In one variable we get ODEs, with a beautiful and complete existence–uniqueness theory. In several variables we get PDEs, where no such complete theory exists — but where, for first order equations, we can *convert the PDE back into an ODE* along special curves. Those curves are the characteristic curves of the title.

We will state theorems, but we will not prove them. The proofs are in the books listed at the end. What we want here is the shape of the subject.

2 The tangent line, revisited

2.1 The definition we all know

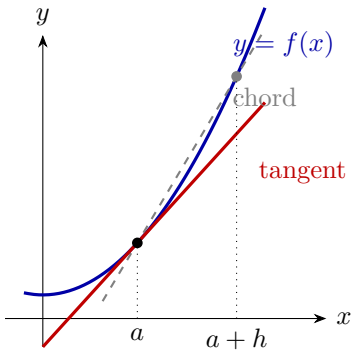
Let $U \subseteq \mathbb{R}$ be an *open* set and let $f : U \to \mathbb{R}$. Openness is not a technicality to be skipped: it guarantees that for each $a \in U$ there is some $\delta > 0$ with $(a - \delta, a + \delta) \subseteq U$, so that we may approach a from *both* sides and $f(a + h)$ makes sense for all small h. Every definition below silently uses this.

Definition 2.1 (The usual one)**.** f is differentiable at $a \in U$ if the limit

$$f'(a) \;=\; \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$

exists in $\mathbb{R}$.

The picture attached to this definition is the familiar one. The chord joining $(a, f(a))$ to $(a + h, f(a + h))$ has slope $\frac{f(a+h)-f(a)}{h}$; as $h \to 0$ the chord pivots about $(a, f(a))$ and settles down to a limiting line. That limiting line is the tangent line, and its slope is $f'(a)$.



2.2 The same definition, rearranged

Now we do something that looks like a trivial algebraic manipulation but is in fact the key to the whole article. Say that f is differentiable at a with derivative $f'(a) = c$. Then

$$\lim_{h\to 0} \frac{f(a+h) - f(a)}{h} = c \qquad \Longleftrightarrow \qquad \lim_{h\to 0} \frac{f(a+h) - f(a) - c\,h}{h} = 0.$$

Since the limit of a quotient is 0 exactly when the limit of its absolute value is 0, this is the same as

$$\lim_{h\to 0} \frac{|\, f(a+h) - f(a) - c\,h\,|}{|h|} \;=\; 0. \tag{1}$$

Look carefully at the expression $c\,h$ appearing in (1). Define

$$L : \mathbb{R} \longrightarrow \mathbb{R}, \qquad L(h) = c\,h.$$

This L is a *linear transformation* from $\mathbb{R}$ to $\mathbb{R}$: it satisfies $L(h_1 + h_2) = L(h_1) + L(h_2)$ and $L(\lambda h) = \lambda L(h)$. So (1) reads:

$$\lim_{h\to 0} \frac{|f(a+h) - f(a) - L(h)|}{|h|} = 0.$$

Definition 2.2 (The usual one, rewritten)**.** Let $U \subseteq \mathbb{R}$ be open, $f : U \to \mathbb{R}$, $a \in U$. We say f is *differentiable at a* if there exists a linear transformation $L : \mathbb{R} \to \mathbb{R}$ such that

$$\lim_{h\to 0} \frac{|\, f(a+h) - f(a) - L(h)\,|}{|h|} \;=\; 0.$$

In that case we call L *the derivative of f at a* and write $f'(a) = L$.

Definitions 2.1 and 2.2 are logically equivalent, and the equivalence is the two-line computation above. But they *say* different things.

- Definition 2.1 says: the difference quotients converge.
- Definition 2.2 says: the increment $f(a+h) - f(a)$ is a linear function of h, *up to an error that is negligible compared with h.*

The second formulation is the one that generalises, for the simple reason that "difference quotient" requires division, and one cannot divide by a vector, whereas "linear map" makes perfect sense in any dimension.

2.3 Why the distinction between c and L is not pedantry

A student may reasonably object: the space $\mathcal{L}(\mathbb{R}, \mathbb{R})$ of linear maps $\mathbb{R} \to \mathbb{R}$ is one dimensional, and the correspondence $c \leftrightarrow L_c$, where $L_c(h) = ch$, is an isomorphism $\mathbb{R} \cong \mathcal{L}(\mathbb{R}, \mathbb{R})$. So why insist on the difference?

Because in one dimension the isomorphism is so natural that it hides a distinction which becomes essential later. Note that

$$c \;=\; L(1),$$

that is, the number c is nothing but *the matrix of L with respect to the standard basis $\{1\}$ of* $\mathbb{R}$. In $\mathbb{R}^n \to \mathbb{R}^m$ the derivative will again be a linear map, and its matrix in the standard bases will be an $m \times n$ array of partial derivatives — the Jacobian matrix. Keeping "the linear map" and "its matrix" separate in one's mind from the start prevents a very common confusion later (see Remark 5.4).

2.4 The geometry: the tangent line is a graph

Rewriting the definition also clarifies what the tangent line *is*. Consider the affine map

$$T : \mathbb{R} \to \mathbb{R}, \qquad T(x) = f(a) + L(x-a) = f(a) + f'(a)(x-a).$$

Its graph is precisely the tangent line to the graph of f at $(a, f(a))$.

So the picture behind Definition 2.2 is: *translate the origin to the point $(a, f(a))$; in these new coordinates the graph of f hugs the graph of a linear map, and L is that linear map.* The tangent line is the graph of the derivative, shifted into position.

Moreover L is the *unique* linear map with this property, and it is the *best* linear approximation in a strong sense: if $M(h) = c'h$ is any other linear map, then $f(a+h) - f(a) - M(h)$ does *not* go to zero faster than h; it is comparable to h. Differentiability is the statement that a first-order error can be arranged.

Example 2.3. Let $f(x) = x^2$ on $U = \mathbb{R}$, and take $a = 1$. For $h \in \mathbb{R}$,

$$f(1+h) - f(1) = (1+h)^2 - 1 = 2h + h^2.$$

Take $L(h) = 2h$. Then

$$\frac{|f(1+h) - f(1) - L(h)|}{|h|} = \frac{|h^2|}{|h|} = |h| \ \xrightarrow[h\to 0]{} \ 0.$$

So f is differentiable at 1 and $f'(1) = L$, whose matrix is the number 2. The tangent line is $y = 1 + 2(x-1)$.

Notice how transparent the computation is: the increment $2h + h^2$ splits into a *linear part* $2h$ and a *quadratic remainder* h^2, and differentiability is exactly the statement that the remainder is small compared to h. All of differential calculus is this splitting.

Example 2.4 (A linear map is its own derivative)**.** Let $f(x) = cx$. Then $f(a+h) - f(a) = ch = L(h)$ exactly, with zero remainder. So $f'(a) = L = f$ for every a. A linear function does not need to be approximated: it is already linear. Keep this example in mind; it will reappear verbatim in $\mathbb{R}^n$.

3 The total derivative

We now take Definition 2.2 and change $\mathbb{R}$ to $\mathbb{R}^n$ on the left and $\mathbb{R}^m$ on the right. Absolute values become norms. That is the entire change.

Throughout, $|\cdot|$ denotes the Euclidean norm on $\mathbb{R}^n$ or $\mathbb{R}^m$, as context demands.

Definition 3.1 (Rudin, *Principles*, 9.11)**.** Suppose E is an open set in $\mathbb{R}^n$, $\mathbf{f}$ maps E into $\mathbb{R}^m$, and $\mathbf{x} \in E$. If there exists a linear transformation $A : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{|\,\mathbf{f}(\mathbf{x}+\mathbf{h}) - \mathbf{f}(\mathbf{x}) - A\mathbf{h}\,|}{|\mathbf{h}|} \ = \ 0, \tag{2}$$

then we say that $\mathbf{f}$ is *differentiable at* $\mathbf{x}$, and we write

$$\mathbf{f}'(\mathbf{x}) \ = \ A.$$

If $\mathbf{f}$ is differentiable at every $\mathbf{x} \in E$, we say $\mathbf{f}$ is differentiable in E. The map A is called the *total derivative* (or *Fréchet derivative*) of $\mathbf{f}$ at $\mathbf{x}$.

Some comments, each worth pausing over.

(a) **The numerator is a vector in $\mathbb{R}^m$; the denominator is a scalar.** We never divide by $\mathbf{h}$. This is exactly why Definition 2.2 rather than 2.1 was the right thing to generalise.

(b) **The derivative is a linear map, not a number and not a matrix.** $\mathbf{f}'(\mathbf{x}) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$. It is an object that eats a vector $\mathbf{h} \in \mathbb{R}^n$ (a displacement from $\mathbf{x}$) and returns a vector in $\mathbb{R}^m$ (the corresponding first-order change in $\mathbf{f}$).

(c) **The derivative varies with the point.** The map $\mathbf{x} \mapsto \mathbf{f}'(\mathbf{x})$ is a function from E into the vector space $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) \cong \mathbb{R}^{mn}$. Even in one variable this was true — f' is a function — but here the values live in a space of linear maps.

(d) **The limit is a genuine multivariable limit: $\mathbf{h} \to \mathbf{0}$** in *every possible way*, along every path, from every direction. This is a strong requirement, and it is precisely what directional derivatives alone will fail to capture (§4).

Theorem 3.2 (Uniqueness; Rudin 9.12)**.** *The linear map A in Definition 3.1, if it exists, is unique.*

This is what allows us to write $\mathbf{f}'(\mathbf{x})$ and speak of *the* derivative. (Sketch of the idea, for the curious: if A_1 and A_2 both work, put $B = A_1 - A_2$; the two limits force $|B\mathbf{h}| / |\mathbf{h}| \to 0$, and taking $\mathbf{h} = t\mathbf{u}$ with $t \to 0$ and $\mathbf{u}$ fixed gives $|B\mathbf{u}| = 0$ for every $\mathbf{u}$.)

3.1 First examples

Example 3.3 (Constants)**.** If $\mathbf{f} \equiv \mathbf{c}$ is constant, then the numerator in (2) is $|-A\mathbf{h}|$, and $A = 0$ (the zero linear map) works. So $\mathbf{f}'(\mathbf{x}) = 0$ for all $\mathbf{x}$.

Example 3.4 (Linear maps are their own derivatives)**.** Let $A : \mathbb{R}^n \to \mathbb{R}^m$ be linear and put $\mathbf{f}(\mathbf{x}) = A\mathbf{x}$. Then for every $\mathbf{x}$ and every $\mathbf{h}$,

$$\mathbf{f}(\mathbf{x} + \mathbf{h}) - \mathbf{f}(\mathbf{x}) = A(\mathbf{x} + \mathbf{h}) - A\mathbf{x} = A\mathbf{h},$$

so the numerator in (2) is exactly 0 — not merely small, but zero. Hence $\mathbf{f}$ is differentiable everywhere and

$$\mathbf{f}'(\mathbf{x}) = A \qquad \text{for every } \mathbf{x} \in \mathbb{R}^n.$$

This example deserves emphasis, because it is the sanity check for the whole definition. The derivative is supposed to be "the best linear approximation". If the function is *already* linear, the best linear approximation had better be the function itself — and it is. Note the pleasant subtlety: the map $\mathbf{x} \mapsto \mathbf{f}'(\mathbf{x})$ is a *constant* map (its value is always A), even though $\mathbf{f}$ itself is not constant. This is the exact analogue of "the derivative of cx is the constant c".

More generally, an affine map $\mathbf{f}(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$ has $\mathbf{f}'(\mathbf{x}) = A$ everywhere: translation does not affect derivatives.

Example 3.5 (A paraboloid)**.** Let $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x, y) = x^2 + y^2$, and fix $\mathbf{p} = (a, b)$. With $\mathbf{h} = (h_1, h_2)$,

$$f(\mathbf{p} + \mathbf{h}) - f(\mathbf{p}) = \underbrace{2ah_1 + 2bh_2}_{\text{linear in } \mathbf{h}} + \underbrace{h_1^2 + h_2^2}_{= |\mathbf{h}|^2}.$$

Take $A\mathbf{h} = 2ah_1 + 2bh_2$, a linear map $\mathbb{R}^2 \to \mathbb{R}$. The error is $|\mathbf{h}|^2$, and $|\mathbf{h}|^2 / |\mathbf{h}| = |\mathbf{h}| \to 0$. So f is differentiable at $\mathbf{p}$ with $f'(\mathbf{p}) = A$, whose matrix is the row vector $\begin{pmatrix} 2a & 2b \end{pmatrix}$.

Observe that this is Example 2.3 with one more variable, and the structure is identical: *increment = linear part + higher-order remainder.*

Example 3.6 (A curve in space)**.** Let $\gamma : \mathbb{R} \to \mathbb{R}^3$, $\gamma(t) = (\cos t, \sin t, t)$ (a helix). Here $n = 1$, $m = 3$, so $\gamma'(t)$ is a linear map $\mathbb{R} \to \mathbb{R}^3$. Its matrix is the column

$$\begin{pmatrix} -\sin t \\ \cos t \\ 1 \end{pmatrix},$$

and the linear map itself is $s \mapsto s\,(-\sin t, \cos t, 1)$.

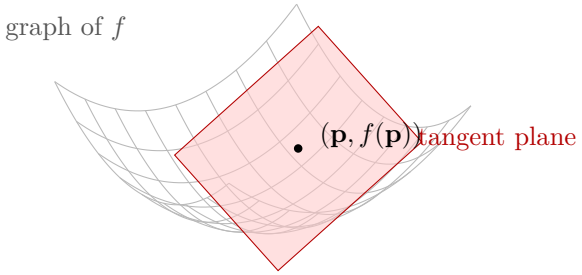
Its *image* is a line through the origin in $\mathbb{R}^3$: the direction of the tangent line to the helix. Translating that line to the point $\gamma(t)$ gives the tangent line in the classical sense. Once again: *the derivative is the linear object; the tangent line is that linear object moved into position.*

3.2 The geometry: from tangent lines to tangent planes

For $f : E \subseteq \mathbb{R}^2 \to \mathbb{R}$ the graph $\{(x, y, f(x, y)) : (x, y) \in E\}$ is a surface in $\mathbb{R}^3$. Differentiability at $\mathbf{p} = (a, b)$ says the surface is well approximated near $(\mathbf{p}, f(\mathbf{p}))$ by the graph of the affine map

$$T(\mathbf{x}) = f(\mathbf{p}) + f'(\mathbf{p})(\mathbf{x} - \mathbf{p}),$$

whose graph is a plane — the *tangent plane*. In general, for $\mathbf{f} : E \subseteq \mathbb{R}^n \to \mathbb{R}^m$, the graph lives in $\mathbb{R}^{n+m}$ and the tangent object is an n-dimensional affine subspace.



The moral of §2 and §3 together: nothing conceptually new happened when we passed from one variable to many. We only had to say the old thing in the right words first.

4 Looking in one direction: directional derivatives

4.1 The idea

Definition 3.1 is elegant but, taken at face value, hard to use: to verify it one must control $\mathbf{f}(\mathbf{x} + \mathbf{h})$ for *all* small $\mathbf{h}$ at once, and to compute A one must somehow guess it.

Here is a very natural strategy, and it is the strategy a physicist or an engineer would try first. We know how to differentiate functions of one variable. So: *let us walk along a straight line and use one-variable calculus.*

Fix $\mathbf{x} \in E$ and a direction $\mathbf{u} \in \mathbb{R}^n$, $\mathbf{u} \neq \mathbf{0}$. Because E is open, there is $\delta > 0$ such that $\mathbf{x} + t\mathbf{u} \in E$ for all $|t| < \delta$. Define

$$\varphi : (-\delta, \delta) \to \mathbb{R}^m, \qquad \varphi(t) \;=\; \mathbf{f}(\mathbf{x} + t\mathbf{u}).$$

This φ is a function of *one real variable*. We know exactly what to do with such functions.

Definition 4.1 (Directional derivative)**.** With the notation above, if φ is differentiable at $t = 0$, we call

$$\mathrm{D}_{\mathbf{u}}\mathbf{f}(\mathbf{x}) \;:=\; \varphi'(0) \;=\; \lim_{t \to 0} \frac{\mathbf{f}(\mathbf{x} + t\mathbf{u}) - \mathbf{f}(\mathbf{x})}{t} \;\in\; \mathbb{R}^m$$

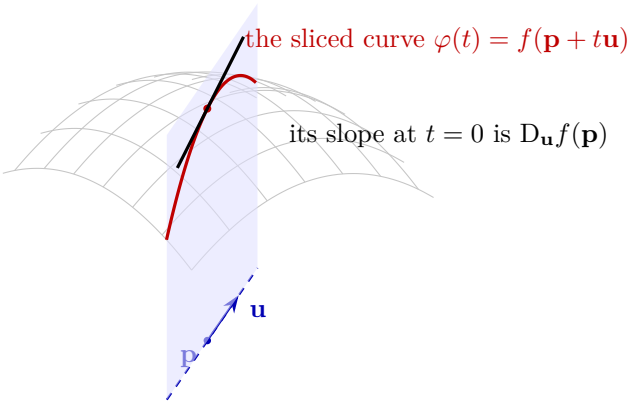
the *directional derivative of* $\mathbf{f}$ *at* $\mathbf{x}$ *in the direction* $\mathbf{u}$.

Note that here t is a scalar, so dividing by t is legitimate: we are back in the world of difference quotients. That is precisely the point — restricting to a line has reduced an n-variable problem to a 1-variable one.

4.2 The geometry: slicing the graph

Take $m = 1$ and $n = 2$, so that the graph of f is a surface in $\mathbb{R}^3$. Fix a point $\mathbf{p}$ in the domain and a direction $\mathbf{u}$. Consider the *vertical plane* containing the line $\{\mathbf{p} + t\mathbf{u}\}$ in the xy-plane. This plane cuts the surface in a curve. That curve, read in the coordinates (t along the line, height z), is exactly the graph of $\varphi(t) = f(\mathbf{p} + t\mathbf{u})$.

The directional derivative $\mathrm{D}_{\mathbf{u}} f(\mathbf{p})$ *is the slope of that sliced curve at* $t = 0$. Different $\mathbf{u}$ means slicing with a different vertical plane, and generally a different slope.



This picture is worth internalising: *a function of several variables is a whole family of functions of one variable, one for each direction.* Directional derivatives are the derivatives of that family.

4.3 How the total derivative contains all directional derivatives

Suppose now that $\mathbf{f}$ is differentiable at $\mathbf{x}$ in the strong sense of Definition 3.1. Put $\mathbf{h} = t\mathbf{u}$ in (2) with $t \to 0$. Then $|\mathbf{h}| = |t|\,|\mathbf{u}|$, and the limit gives

$$\frac{|\mathbf{f}(\mathbf{x} + t\mathbf{u}) - \mathbf{f}(\mathbf{x}) - t\,\mathbf{f}'(\mathbf{x})\mathbf{u}|}{|t|\;|\mathbf{u}|} \xrightarrow[t\to 0]{} 0,$$

which says exactly that $\frac{\mathbf{f}(\mathbf{x}+t\mathbf{u}) - \mathbf{f}(\mathbf{x})}{t} \to \mathbf{f}'(\mathbf{x})\mathbf{u}$.

Theorem 4.2 (Rudin 9.17, first half)**.** *If* $\mathbf{f}$ *is differentiable at* $\mathbf{x}$*, then* $\mathrm{D}_{\mathbf{u}}\mathbf{f}(\mathbf{x})$ *exists for* every *direction* $\mathbf{u}$*, and*

$$\boxed{\mathrm{D}_{\mathbf{u}}\mathbf{f}(\mathbf{x}) \;=\; \mathbf{f}'(\mathbf{x})\,\mathbf{u}}$$

Read that boxed formula slowly. On the left is a limit of difference quotients along a line. On the right is a linear map applied to a vector. The theorem says: *once you know the linear map* $\mathbf{f}'(\mathbf{x})$*, you know the slope in every one of the infinitely many directions, and the dependence on the direction is linear.*

In particular $\mathbf{u} \mapsto \mathrm{D}_{\mathbf{u}}\mathbf{f}(\mathbf{x})$ is a *linear* function of $\mathbf{u}$, so $\mathrm{D}_{\lambda\mathbf{u}}\mathbf{f} = \lambda\,\mathrm{D}_{\mathbf{u}}\mathbf{f}$ and $\mathrm{D}_{\mathbf{u}+\mathbf{v}}\mathbf{f} = \mathrm{D}_{\mathbf{u}}\mathbf{f} + \mathrm{D}_{\mathbf{v}}\mathbf{f}$.

4.4 The converse fails — and it fails for a reason worth seeing

The natural hope is that the implication reverses: if all directional derivatives exist, surely the function is differentiable? It does not.

Example 4.3 (All directions, but no derivative)**.** Define $f : \mathbb{R}^2 \to \mathbb{R}$ by

$$f(x,y) = \begin{cases} \dfrac{x^3}{x^2+y^2}, & (x,y) \neq (0,0), \\ 0, & (x,y) = (0,0). \end{cases}$$

Fix a direction $\mathbf{u} = (\alpha, \beta) \neq (0,0)$. For $t \neq 0$,

$$\frac{f(t\alpha, t\beta) - f(0,0)}{t} = \frac{1}{t} \cdot \frac{t^3\alpha^3}{t^2(\alpha^2+\beta^2)} = \frac{\alpha^3}{\alpha^2+\beta^2}.$$

This is independent of t, so the limit exists for every direction:

$$\mathrm{D}_{\mathbf{u}} f(0,0) = \frac{\alpha^3}{\alpha^2+\beta^2}.$$

Every directional derivative exists at the origin. Yet f is *not* differentiable at the origin — because if it were, Theorem 4.2 would force $\mathbf{u} \mapsto \mathrm{D}_{\mathbf{u}} f(0,0)$ to be linear in $\mathbf{u}$, and $\frac{\alpha^3}{\alpha^2+\beta^2}$ is manifestly not linear (take $\mathbf{u} = (1,0)$ and $\mathbf{v} = (0,1)$: the values are 1 and 0, but at $\mathbf{u}+\mathbf{v} = (1,1)$ the value is $\frac{1}{2}$, not $1+0$).

Remark 4.4 (Why this had to be possible)**.** Geometrically, knowing all the directional derivatives means knowing the slope of every *straight-line* slice through the point. Differentiability requires $\mathbf{h} \to \mathbf{0}$ along *all* paths — including spirals and parabolic approaches — and requires the approximation to be good *uniformly* in the direction. Straight lines simply do not see enough. Directional derivatives are a family of one-dimensional facts; differentiability is a genuinely n-dimensional fact.

So directional derivatives are a good idea, but by themselves they are not the derivative. What saves the situation is the next section.

5 Special directions: partial derivatives

5.1 Definition

Among all the directions $\mathbf{u} \in \mathbb{R}^n$, n of them are distinguished: the standard basis vectors

$$\mathbf{e}_1 = (1,0,\ldots,0), \quad \ldots, \quad \mathbf{e}_n = (0,\ldots,0,1).$$

Definition 5.1 (Partial derivative)**.** The directional derivative in the direction $\mathbf{e}_j$ is called the *j-th partial derivative*:

$$\mathrm{D}_j \mathbf{f}(\mathbf{x}) \;:=\; \mathrm{D}_{\mathbf{e}_j} \mathbf{f}(\mathbf{x}) \;=\; \lim_{t\to 0} \frac{\mathbf{f}(\mathbf{x}+t\mathbf{e}_j) - \mathbf{f}(\mathbf{x})}{t} \;=\; \frac{\partial \mathbf{f}}{\partial x_j}(\mathbf{x}).$$

Concretely: freeze all variables except the j-th and differentiate in the ordinary one-variable sense. This is the operation students actually perform. Geometrically, we are slicing the graph by a vertical plane *parallel to a coordinate axis* — the same picture as before, in a special position.

5.2 Why partial derivatives are enough

Here is the fact that makes multivariable calculus computable.

Theorem 5.2 (Rudin 9.17)**.** *Let $E \subseteq \mathbb{R}^n$ be open and $\mathbf{f} = (f_1, \dots, f_m) : E \to \mathbb{R}^m$ be differentiable at $\mathbf{x} \in E$. Then each partial derivative $\mathrm{D}_j f_i(\mathbf{x})$ exists, and the matrix of the linear map $\mathbf{f}'(\mathbf{x})$ with respect to the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$ is the $m \times n$ matrix*

$$[\mathbf{f}'(\mathbf{x})] \;=\; \begin{pmatrix} \mathrm{D}_1 f_1 & \mathrm{D}_2 f_1 & \cdots & \mathrm{D}_n f_1 \\ \mathrm{D}_1 f_2 & \mathrm{D}_2 f_2 & \cdots & \mathrm{D}_n f_2 \\ \vdots & \vdots & & \vdots \\ \mathrm{D}_1 f_m & \mathrm{D}_2 f_m & \cdots & \mathrm{D}_n f_m \end{pmatrix} (\mathbf{x}),$$

called the Jacobian matrix *of $\mathbf{f}$ at $\mathbf{x}$.*

The reason is now transparent. By Theorem 4.2, $\mathbf{f}'(\mathbf{x})\mathbf{e}_j = \mathrm{D}_j\mathbf{f}(\mathbf{x})$; and the columns of the matrix of a linear map are precisely its values on the basis vectors. The Jacobian matrix is not a new idea — it is bookkeeping for the values of $\mathbf{f}'(\mathbf{x})$ on a basis.

Combining with Theorem 4.2 we get the formula that the whole of §4 was heading towards. Write $\mathbf{u} = \sum_{j=1}^{n} u_j \mathbf{e}_j$. Then, by linearity of $\mathbf{f}'(\mathbf{x})$,

$$\boxed{\mathrm{D}_{\mathbf{u}}\mathbf{f}(\mathbf{x}) \;=\; \mathbf{f}'(\mathbf{x})\mathbf{u} \;=\; \sum_{j=1}^{n} u_j \, \mathrm{D}_j\mathbf{f}(\mathbf{x}).} \tag{3}$$

> *There are infinitely many directions in $\mathbb{R}^n$, and hence infinitely many directional derivatives. Formula* (3) *says that all of them are determined by just n of them — the partial derivatives. This is the single most important structural fact in differential calculus of several variables.*

It is worth seeing why this is not a miracle but a consequence of *linearity*. Directions form an infinite set; but the map $\mathbf{u} \mapsto \mathrm{D}_{\mathbf{u}}\mathbf{f}(\mathbf{x})$ is linear, and a linear map is determined by its values on a basis. Differentiability is what buys us the linearity; linear algebra does the rest.

5.3 The scalar case: the gradient

When $m = 1$, the matrix of $f'(\mathbf{x})$ is a single row, and it is convenient to regard it as a vector.

Definition 5.3. For $f : E \subseteq \mathbb{R}^n \to \mathbb{R}$ differentiable at $\mathbf{x}$, the *gradient* is

$$\nabla f(\mathbf{x}) = \big(\mathrm{D}_1 f(\mathbf{x}), \dots, \mathrm{D}_n f(\mathbf{x})\big) \in \mathbb{R}^n.$$

Then (3) becomes

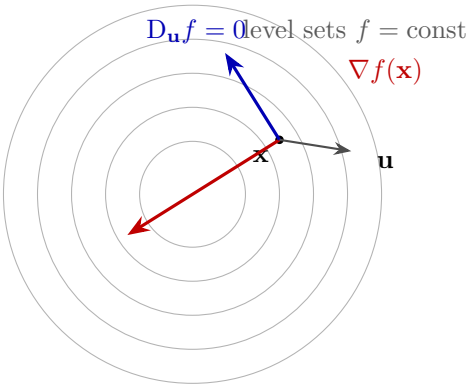
$$\mathrm{D}_{\mathbf{u}} f(\mathbf{x}) \;=\; \langle \nabla f(\mathbf{x}), \mathbf{u} \rangle. \tag{4}$$

This one formula carries a lot of geometry. Take $|\mathbf{u}| = 1$. By Cauchy–Schwarz,

$$|\mathrm{D}_{\mathbf{u}} f(\mathbf{x})| \le |\nabla f(\mathbf{x})|,$$

with equality iff $\mathbf{u}$ is parallel to $\nabla f(\mathbf{x})$. Hence:

- $\nabla f(\mathbf{x})$ points in the direction of *steepest increase* of f, and $|\nabla f(\mathbf{x})|$ is that steepest rate;
- $\mathrm{D}_{\mathbf{u}} f(\mathbf{x}) = 0$ exactly when $\mathbf{u} \perp \nabla f(\mathbf{x})$; these are the directions in which f does not change to first order, i.e. the directions *tangent to the level set* through $\mathbf{x}$. So ∇f is normal to level sets.



Remark 5.4 (The Jacobian matrix is not the derivative)**.** It is very common — and wrong — to say "the derivative *is* the Jacobian matrix". Three reasons to resist:

(1) **Logical direction.** Theorem 5.2 says *differentiable* $\Rightarrow$ *partials exist and assemble into the matrix*. The converse is false: in Example 4.3 the partials at the origin exist ($\mathrm{D}_1 f(0,0) = 1$, $\mathrm{D}_2 f(0,0) = 0$), so the "Jacobian" $\begin{pmatrix} 1 & 0 \end{pmatrix}$ exists as an array of numbers — but there is no derivative. *A matrix of partials can exist without being the matrix of anything.*

(2) **Basis dependence.** The derivative is a linear map, an intrinsic object. Its matrix depends on a choice of bases. Change coordinates and the matrix changes; the linear map does not.

(3) **Conceptual.** The derivative is the object that answers "how does $\mathbf{f}$ change if I move by $\mathbf{h}$?"; the Jacobian is a convenient table of numbers for computing that answer.

The correct slogan is: *the Jacobian matrix is the matrix of the derivative, in the standard bases, when the derivative exists.*

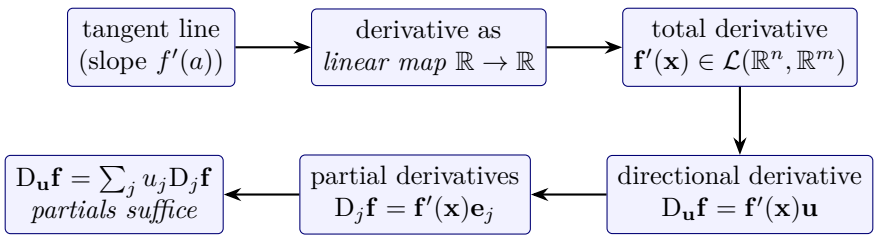
Finally, the theorem that rescues practice: it says that the pathology of Example 4.3 disappears under a mild hypothesis.

Theorem 5.5 (Rudin 9.21)**.** *Let $E \subseteq \mathbb{R}^n$ be open and $\mathbf{f} : E \to \mathbb{R}^m$. Then $\mathbf{f}$ is continuously differentiable on E (that is, $\mathbf{f}' : E \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ exists and is continuous) if and only if all the partial derivatives $\mathrm{D}_j f_i$ exist and are continuous on E.*

So: *existence of partials is not enough; continuity of partials is.* In every situation we shall meet from now on, the functions are as smooth as we like, and we may pass freely between "the derivative" and "the collection of partial derivatives". That licence is what makes the rest of the story possible.

5.4 Summary of the calculus half

Let us record the chain of ideas before putting derivatives into equations.



6 Putting derivatives into equations: ODEs

Having understood what a derivative is, the natural next move is to *impose a condition on it* and ask which functions satisfy that condition. An equation relating an unknown function of *one* variable and its derivatives is an *ordinary differential equation*.

6.1 The first order initial value problem

The basic object is

$$\frac{dy}{dt} = F(t, y), \qquad y(t_0) = y_0, \tag{5}$$

where F is defined on some open set of the (t, y)-plane containing (t_0, y_0).

Theorem 6.1 (Picard–Lindelöf)**.** *Let* $R = \{(t, y) : |t - t_0| \le a,\ |y - y_0| \le b\}$ *and suppose* F *is continuous on* R *and* Lipschitz in y*: there is* $L > 0$ *with*

$$|F(t, y_1) - F(t, y_2)| \le L\,|y_1 - y_2| \qquad \text{for all } (t, y_1), (t, y_2) \in R.$$

Then there is $h > 0$ *such that the initial value problem* (5) *has a* unique *solution on* $(t_0 - h,\ t_0 + h)$.

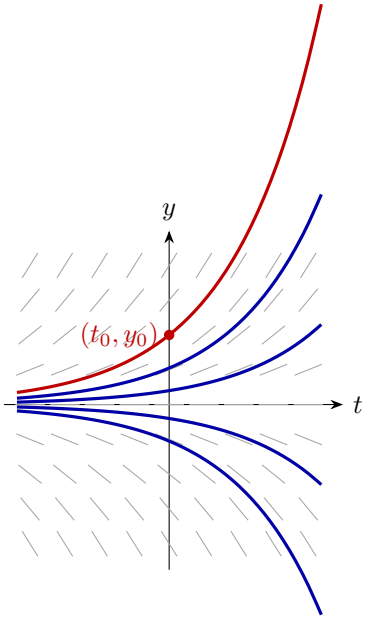
Two remarks on the hypotheses, since students often meet them as unmotivated technicalities.

- Continuity alone gives existence (Peano's theorem) but *not* uniqueness.
- The standard illustration: $y' = 3y^{2/3}$, $y(0) = 0$. The right-hand side is continuous but not Lipschitz in y at $y = 0$. Both $y(t) \equiv 0$ and $y(t) = t^3$ solve it — and so does the function that stays at 0 until time c and then takes off as $(t - c)^3$. Uniqueness genuinely fails. Whenever we later say "the solution", a Lipschitz-type hypothesis is lurking in the background.

6.2 The geometry: direction fields and integral curves

This is the picture to carry into the PDE sections. Equation (5) prescribes, at every point (t, y) of the plane, a *slope* $F(t, y)$. Draw at each point a short segment of that slope. The result is a *direction field* (equivalently, the vector field $(1, F(t, y))$).

A solution of (5) is a curve which is *everywhere tangent to this field*: at each of its points, its slope agrees with the prescribed slope. Such a curve is called an *integral curve*.



In this language, Theorem 6.1 says something visually striking:

Through every point of the region there passes exactly one integral curve.

Consequently distinct integral curves *never cross*. The region is filled up by a family of non-intersecting curves — the plane is *foliated* by solutions. Solving the ODE means picking out the one leaf of this foliation that passes through the prescribed initial point. Remember this: in §9 we will meet a situation where the analogous curves *do* cross, and something dramatic will happen.

6.3 Everything reduces to the first order system

One might worry that Theorem 6.1 is too narrow — what about higher order equations, or several coupled unknowns? It is not narrow at all; both cases reduce to it.

Higher order. Consider

$$y^{(n)} = G(t, y, y', \dots, y^{(n-1)}).$$

Introduce new unknowns $x_1 = y,\ x_2 = y',\ \dots,\ x_n = y^{(n-1)}$. Then

$$x_1' = x_2, \quad x_2' = x_3, \quad \dots, \quad x_{n-1}' = x_n, \quad x_n' = G(t, x_1, \dots, x_n),$$

which is a first order *system* for $\mathbf{x} = (x_1, \dots, x_n)$.

Systems. A first order system is exactly an equation

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(t, \mathbf{x}), \qquad \mathbf{x}(t_0) = \mathbf{x}_0, \qquad \mathbf{F} : \text{open set of } \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n.$$

And here the language of §3 pays off: $\mathbf{x} : I \to \mathbb{R}^n$ is a function of one variable with values in $\mathbb{R}^n$, its derivative at t is a linear map $\mathbb{R} \to \mathbb{R}^n$ (Example 3.6), and we are prescribing that linear map — equivalently, the tangent vector $\mathbf{x}'(t)$ — at each point.

Theorem 6.2 (Picard–Lindelöf, vector form)**.** *If* $\mathbf{F}$ *is continuous on a closed box around* $(t_0, \mathbf{x}_0)$ *and Lipschitz in* $\mathbf{x}$ *there, then the system has a unique solution on some interval about* t_0*.*

Geometrically, $\mathbf{F}$ is a (time-dependent) *vector field* on $\mathbb{R}^n$, and a solution is a curve in $\mathbb{R}^n$ whose velocity vector at every instant equals the field at the point where it happens to be: a *trajectory* of the flow. This is the picture of a particle carried by a fluid.

> *Summary: in one independent variable we have a complete and satisfying theory. Given a vector field, curves tangent to it exist, are unique, and fill the space without crossing. The rest of this article is about exploiting this theory in a setting where no such general theory exists.*

7 From ODE to PDE: why partial derivatives appear

Now let the unknown be a function u of *several* variables, say $u : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$. What should "a differential equation" mean?

Follow the logic of §§4–5. A derivative of u at a point is now a linear map, and what it does is measured direction by direction. So the most natural thing one can impose is a condition on the *directional derivatives* of u: for instance,

$$\mathrm{D}_{\mathbf{a}(\mathbf{x})} u(\mathbf{x}) \ = \ c(\mathbf{x}, u(\mathbf{x})), \tag{6}$$

where at each point $\mathbf{x}$ a direction $\mathbf{a}(\mathbf{x})$ is specified and the rate of change of u in *that* direction is prescribed.

But by (3), if $\mathbf{a} = (a_1, \ldots, a_n)$ then

$$\mathrm{D}_{\mathbf{a}(\mathbf{x})} u(\mathbf{x}) = \sum_{j=1}^{n} a_j(\mathbf{x}) \, \frac{\partial u}{\partial x_j}(\mathbf{x}),$$

so (6) is

$$a_1(\mathbf{x}) \, u_{x_1} + a_2(\mathbf{x}) \, u_{x_2} + \cdots + a_n(\mathbf{x}) \, u_{x_n} \;=\; c(\mathbf{x}, u). \tag{7}$$

> *An equation about directional derivatives* is *an equation about partial derivatives. That is a* partial differential equation*. The reason PDE theory is phrased in terms of $\partial u/\partial x_j$ and not in terms of directional derivatives is not that partial derivatives are more natural — they are not — but that they are* sufficient*, by* (3)*.*

This deserves to be said in the other direction too, because it is the key that unlocks §8:

> *Conversely, a first order linear PDE such as* (7) *should always be* read back *as a statement about a directional derivative: "the rate of change of u, along the direction field $\mathbf{a}$, equals c."*

In general a PDE of order k is a relation

$$F(\mathbf{x},\ u,\ \mathrm{D}u,\ \mathrm{D}^2 u,\ \ldots,\ \mathrm{D}^k u) = 0$$

among $\mathbf{x}$, the unknown u, and its partial derivatives up to order k.

7.1 The bad news

For ODEs we had Theorem 6.1: one theorem, covering essentially everything, giving existence and uniqueness. *There is no such theorem for PDEs.* There is no general algorithm, and no general existence theory; different equations behave in radically different ways, and the same equation can be well behaved with one kind of data and hopeless with another. (The Cauchy–Kovalevskaya theorem is a general local result, but it needs analyticity and, as Hadamard's example shows, does not give the stability one wants.)

The response of the subject has been to *classify* and to study important representative equations:

- by **order**: first order, second order, higher;
- by **linearity**, for first order equations in two variables:

linear:	$a(x,y)u_x + b(x,y)u_y + c(x,y)u = d(x,y)$
semilinear:	$a(x,y)u_x + b(x,y)u_y = c(x,y,u)$
quasilinear:	$a(x,y,u)u_x + b(x,y,u)u_y = c(x,y,u)$
fully nonlinear:	$F(x,y,u,u_x,u_y) = 0$

  the distinguishing question being: *do the coefficients of the highest derivatives depend on u?*
- by **type**, for second order equations: elliptic (Laplace $u_{xx} + u_{yy} = 0$), parabolic (heat $u_t = u_{xx}$), hyperbolic (wave $u_{tt} = c^2 u_{xx}$).

7.2 The good news

For *first order* equations there is a beautiful and complete geometric method, and it works by the oldest trick in mathematics: *reduce the new problem to the one you have already solved.* We shall convert the PDE into a family of ODEs. The curves along which this conversion happens are the *characteristic curves.*

8 The method of characteristics

8.1 The transport equation

We start with the simplest possible example, and we will get everything out of it. Let $c \in \mathbb{R}$ be a constant and consider

$$u_t + c\,u_x = 0 \quad \text{in } \mathbb{R} \times (0,\infty), \qquad u(x,0) = u_0(x) \quad \text{for } x \in \mathbb{R}, \tag{8}$$

where $u_0 : \mathbb{R} \to \mathbb{R}$ is a given (C^1, say) function. So we are working on the upper half plane $\{(x,t) : t > 0\}$, and the data is prescribed on the x-axis.

Step 1: read the equation as a directional derivative. The left-hand side of (8) is

$$c\,u_x + 1 \cdot u_t = \langle (c,1), (u_x, u_t) \rangle = \langle (c,1), \nabla u \rangle = \mathrm{D}_{(c,1)} u$$

by (4). So the PDE says, in plain words:

> *The directional derivative of u in the fixed direction $(c,1)$ of the (x,t)-plane is zero. That is:* u does not change as you walk in the direction $(c,1)$.

Everything else is bookkeeping.

Step 2: walk along that direction. Define the curves in the (x,t)-plane obtained by moving in the direction $(c,1)$:

$$\frac{dx}{ds} = c, \qquad \frac{dt}{ds} = 1. \tag{9}$$

Note that (9) is an *ODE system* — a very easy one. Starting from a point $(x_0, 0)$ on the initial line at $s = 0$:

$$t(s) = s, \qquad x(s) = x_0 + cs.$$

These are straight lines of slope $1/c$ in the (x,t)-plane; equivalently, they are the lines

$$x - ct = x_0 = \text{constant}.$$

They are called the *characteristic curves* (here: characteristic lines) of (8).

Step 3: restrict u to a characteristic. Define $z(s) = u(x(s), t(s))$: this is the value of the unknown as seen by a traveller moving along the characteristic. Then, by the chain rule,

$$\frac{dz}{ds} = u_x \frac{dx}{ds} + u_t \frac{dt}{ds} = c\,u_x + u_t = 0,$$

the last step by the PDE itself. So z is *constant* along each characteristic.

We have converted a PDE into an ODE. The ODE was $dz/ds = 0$, whose solution we know.

Step 4: assemble. Since z is constant along the characteristic through $(x_0, 0)$, and $z(0) = u(x_0, 0) = u_0(x_0)$, we get $u(x,t) = u_0(x_0)$ whenever $x - ct = x_0$:

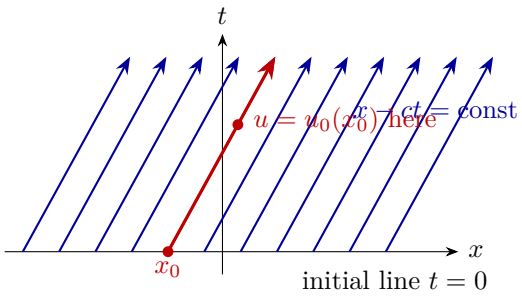
$$\boxed{u(x,t) \;=\; u_0(x - ct).} \tag{10}$$

One checks directly that this solves (8). Moreover, exactly one characteristic passes through each point of the upper half plane, and each one meets the initial line exactly once, so the solution is determined everywhere and is unique.

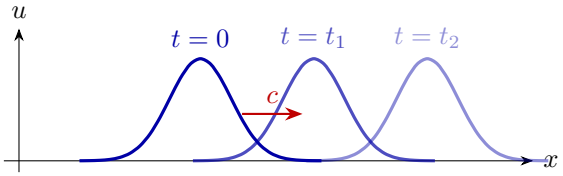
8.2 The geometry of the transport equation

Two pictures tell the whole story.

(a) In the (x,t)-plane. The plane is filled by the parallel lines $x - ct = \text{const}$. The function u is constant on each line. The initial data is carried, unchanged, up along each line.



(b) As a moving profile. Fix t and plot $x \mapsto u(x,t)$. By (10) this is the graph of u_0 shifted to the right by ct. So the solution is the initial profile *translating rigidly with speed c*, without changing shape at all. This is why (8) is called the transport (or advection) equation: it models a substance carried along by a stream moving at constant speed c.



A remark on information. Notice what (10) says about *where the solution at a point comes from*: the value $u(x,t)$ depends on the initial data at the *single* point $x - ct$, and on nothing else. Characteristics are the routes along which information propagates. Change u_0 far from $x - ct$ and $u(x,t)$ does not notice. This finite, directed propagation of information is characteristic (in both senses) of hyperbolic problems, and it contrasts sharply with, say, the heat equation, where the value at any point instantly feels all the initial data.

8.3 The general first order linear equation

Nothing above used the constancy of c in an essential way. Consider

$$a(x,y)\,u_x + b(x,y)\,u_y = c(x,y,u), \tag{11}$$

with u prescribed on some curve Γ in the (x,y)-plane. Reading it as before:

$$\mathrm{D}_{(a,b)}u = c,$$

that is, *the rate of change of u along the direction field $(a(x,y), b(x,y))$ is c*. So:

- Solve the ODE system (*the characteristic equations for the base curves*)

$$\frac{dx}{ds} = a(x,y), \qquad \frac{dy}{ds} = b(x,y).$$

  Its solution curves are the *characteristic curves* in the plane. By §6 these exist, are unique, and do not cross (under a Lipschitz hypothesis).

- Along such a curve set $z(s) = u(x(s), y(s))$. The chain rule and (11) give

$$\frac{dz}{ds} = u_x \frac{dx}{ds} + u_y \frac{dy}{ds} = a\,u_x + b\,u_y = c(x,y,z),$$

  a scalar ODE for z.

- Start each characteristic on the data curve Γ, where z is known; solve the ODEs; the values of z sweep out the solution.

The three-line system

$$\frac{dx}{ds} = a, \qquad \frac{dy}{ds} = b, \qquad \frac{dz}{ds} = c \tag{12}$$

is the *characteristic system* (Lagrange's equations). Everything about first order equations flows from it.

8.4 The picture that explains the method: the solution surface

Here is the geometric statement that makes the whole method inevitable rather than clever.

A solution u of (11) is the same thing as a surface

$$S = \{(x,y,z) \in \mathbb{R}^3 : z = u(x,y)\}$$

— the *integral surface* or *solution surface*. Now, the surface $z - u(x,y) = 0$ is a level set of $\Phi(x,y,z) = z - u(x,y)$, so a normal vector to S at a point is

$$\mathbf{n} = \nabla\Phi = (-u_x,\, -u_y,\, 1).$$

Consider the vector field on $\mathbb{R}^3$

$$\mathbf{V}(x,y,z) = \big(a(x,y),\ b(x,y),\ c(x,y,z)\big).$$

Then

$$\langle \mathbf{V}, \mathbf{n} \rangle = -a\,u_x - b\,u_y + c,$$

and therefore

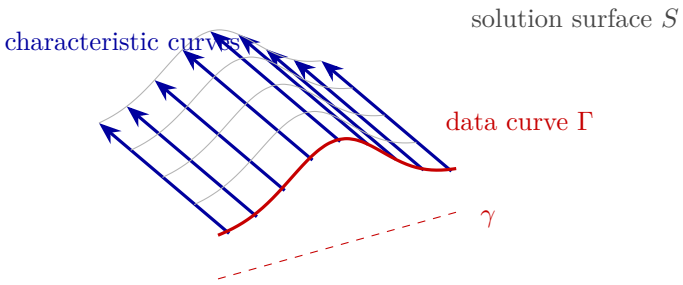
u solves the PDE $\iff$ $\langle \mathbf{V}, \mathbf{n} \rangle = 0$ at every point of S $\iff$ $\mathbf{V}$ is *tangent* to S everywhere.

> *A first order PDE is the statement that a given vector field in $\mathbb{R}^3$ is tangent to the unknown surface.*

And now the method writes itself. The integral curves of $\mathbf{V}$ — the solutions of (12), called *characteristic curves* (or Monge curves) — exist and are unique through each point by ODE theory. If such a curve touches S at one point, then, because $\mathbf{V}$ is tangent to S everywhere, the curve has no way to leave S: it must stay inside S. Hence:

> *Every solution surface is a union of characteristic curves. Conversely, to build a solution surface, take a curve Γ on which the data is prescribed and sweep it along the characteristics.*

We are literally *weaving* the surface out of curves — and each curve is found by solving an ODE.



Remark 8.1 (The transversality condition)**.** For this weaving to produce a genuine surface (a graph $z = u(x, y)$) near Γ, the characteristics must move *away* from the data curve, not run along it. If we parametrise Γ's projection γ by $\tau \mapsto (x_0(\tau), y_0(\tau))$, the required condition at each point is

$$\det \begin{pmatrix} a & \dfrac{dx_0}{d\tau} \\ b & \dfrac{dy_0}{d\tau} \end{pmatrix} \neq 0,$$

i.e. the characteristic direction (a, b) is not tangent to γ. If the data curve *is* characteristic, then either the data is incompatible (no solution) or it is compatible and then infinitely many solutions exist. For (8), γ is the x-axis with tangent $(1, 0)$ and the characteristic direction is $(c, 1)$; the determinant is $-1 \neq 0$, so all is well.

Example 8.2 (A variable-coefficient example)**.** Solve $x\,u_x + y\,u_y = 0$ away from the origin, with $u = f(x)$ on the line $y = 1$.

The characteristic ODEs are $x' = x$, $y' = y$, $z' = 0$. Starting from $(x_0, 1)$: $x = x_0 e^s$, $y = e^s$, $z = f(x_0)$. Eliminating: $x/y = x_0$, so

$$u(x, y) = f\left(\frac{x}{y}\right).$$

The characteristics are the rays through the origin (indeed x/y is constant along them), u is constant on each ray, and the data on $y = 1$ is propagated outward along the rays. Note that the method fails at the origin, where the vector field (x, y) vanishes — no characteristic direction there.

9 When the direction depends on the solution: the quasilinear case

We now change one letter and watch the subject change character completely. Compare

$$u_t + c\,u_x = 0 \qquad \text{with} \qquad u_t + u\,u_x = 0.$$

The second is the *inviscid Burgers equation*. The constant speed c has been replaced by the unknown u itself: the equation is *quasilinear* — the coefficient of a first derivative depends on u.

Consider

$$u_t + u\,u_x = 0 \quad \text{in } \mathbb{R} \times (0, \infty), \qquad u(x, 0) = u_0(x). \tag{13}$$

9.1 The same method, applied

Reading the equation as a directional derivative:

$$u_t + u\,u_x = \mathrm{D}_{(u,\,1)}u = 0.$$

So u is constant along the direction $(u, 1)$ — but that direction now depends on the value of u there. The characteristic system (12) becomes, with (x, t) in place of (x, y),

$$\frac{dx}{ds} = z, \qquad \frac{dt}{ds} = 1, \qquad \frac{dz}{ds} = 0.$$

The third equation says z is constant along the characteristic; if we start at $(x_0, 0)$ then $z \equiv u_0(x_0)$. The second gives $t = s$. Feeding the constant z into the first gives $dx/ds = u_0(x_0)$, hence

$$x = x_0 + u_0(x_0)\,t.$$

> *The characteristics are still* straight lines*, but their slopes are no longer all equal: the line emanating from x_0 carries the value $u_0(x_0)$ and travels with exactly that speed.*

So we obtain the solution in implicit form:

$$\boxed{u(x,t) = u_0\big(x - u(x,t)\,t\big).} \tag{14}$$

Compare with (10): the shape is the same, but u now appears on both sides. The solution is defined implicitly, and that is not a technical inconvenience — it is a signal that something can go wrong.

9.2 Example 1: spreading

Take $u_0(x) = x$. Then $x = x_0 + x_0 t = x_0(1 + t)$, so $x_0 = x/(1 + t)$ and

$$u(x,t) = \frac{x}{1+t}, \qquad t > -1.$$

Here the characteristics fan *outwards* (larger x_0 carries larger speed), they never meet for $t > 0$, and the solution is smooth for all positive time. The profile flattens as time increases.

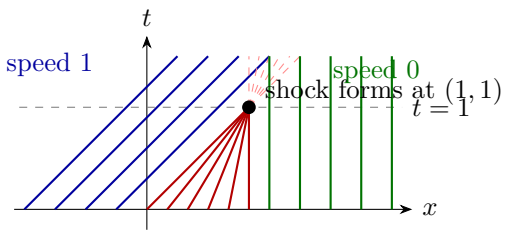
9.3 Example 2: steepening and shock formation

Now take a *decreasing* initial profile, for instance

$$u_0(x) = \begin{cases} 1, & x \le 0, \\ 1 - x, & 0 < x < 1, \\ 0, & x \ge 1. \end{cases}$$

The fluid on the left moves at speed 1; the fluid on the right stands still. The fast part catches up with the slow part.

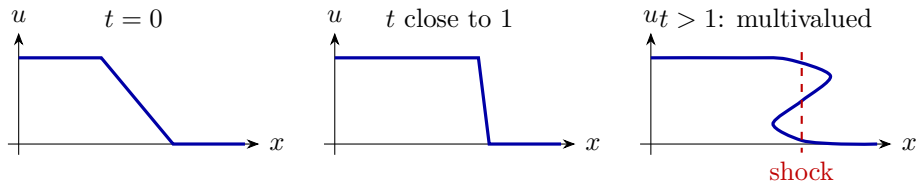
Quantitatively, the characteristic from $x_0 \in [0, 1]$ is $x = x_0 + (1 - x_0)t$, and at $t = 1$ every one of them is at $x = 1$. All these characteristics meet at the single point $(1, 1)$.



Now look at what this means. Beyond $t = 1$, a point (x, t) can lie on *several* characteristics carrying *different* values of $u_0(x_0)$. The recipe "u is constant along characteristics" would assign several values at once. There is no single-valued smooth solution past $t = 1$.

Contrast this with §6, where the integral curves of an ODE never crossed. The crossing is possible here precisely because the field $(u, 1)$ depends on the unknown: two different characteristics belong to two different values of u, so there is no contradiction with ODE uniqueness — the characteristics live in the (x, t, z)-space, where they do *not* cross; it is their *projections* to the (x, t)-plane that cross.

Equivalently, in terms of the profile: since faster parts overtake slower ones, the graph of $x \mapsto u(x, t)$ steepens; at the critical time the tangent becomes vertical ($u_x \to -\infty$), and afterwards the "solution" would have to be multivalued — a wave that has broken.



9.4 The breaking time

In general, differentiating (14) shows that the smooth solution exists exactly as long as the map $x_0 \mapsto x_0 + u_0(x_0)t$ is injective, i.e. as long as $1 + u_0'(x_0)t > 0$ for all x_0. Hence:

- if $u_0' \geq 0$ everywhere (non-decreasing data), characteristics never cross and the smooth solution exists for all $t > 0$ (Example 1);
- if $u_0'(x_0) < 0$ somewhere, the smooth solution breaks down at the finite time

$$t^* \;=\; \frac{-1}{\min_{x_0} u_0'(x_0)}.$$

For the profile above, $\min u_0' = -1$, so $t^* = 1$, matching the picture.

Remark 9.1 (What comes next)**.** This is not a failure of the method — it is a discovery made *by* the method. The characteristics told us, from the initial data alone and without ever solving anything, exactly when and where the solution would cease to exist. The subject's response is to widen the notion of solution: one allows u to be discontinuous and asks it to satisfy the equation in an integral (weak) sense. The discontinuity is a *shock*, its speed is governed by the Rankine–Hugoniot condition, and an extra *entropy condition* is imposed to restore uniqueness. This is the beginning of the theory of conservation laws — the mathematics of traffic jams and of shock waves in gases. It is a fine place for an interested student to read next.

10 Epilogue: the thread

Let us look back at the road.

1. We rewrote $f'(a)$ as a *linear map*, and saw the tangent line as the graph of that map moved into position.
2. The rewriting transplanted verbatim to $\mathbb{R}^n \to \mathbb{R}^m$, giving the total derivative — again a linear map, again with a tangent affine space as its geometric shadow.

3. Asking what that linear map does along a single line gave *directional derivatives*: the several-variable function seen as a family of one-variable functions, one per direction.

4. Because the dependence on the direction is *linear*, n special directions suffice: the *partial derivatives* determine everything.

5. Constraining derivatives gives differential equations. In one variable the geometry is: a vector field, and curves tangent to it. Existence and uniqueness say those curves fill space without crossing.

6. In several variables, an equation on directional derivatives is — by step 4 — an equation on partial derivatives: a PDE.

7. For first order PDEs, reading the equation *backwards* as a statement about a directional derivative recovers a vector field in one dimension higher, and the PDE becomes: *find a surface tangent to this field*. Such a surface is woven out of the field's integral curves — the *characteristic curves* — and those are found by solving ODEs. We are back at step 5.

Notice that a single word runs through the entire story: *tangency*. The tangent line is tangent to a graph. The tangent plane is tangent to a surface. An integral curve is tangent to a direction field. A solution surface of a first order PDE is tangent to a vector field in $\mathbb{R}^3$. Differential calculus is the study of what is tangent to what; differential equations are the art of prescribing tangency and asking what object is left.

And there is a second thread, quieter but just as important: *reduce the new to the known*. Multivariable differentiation was made computable by restricting to lines. First order PDEs are made solvable by restricting to characteristics. Both times, the trick is to find the right one-dimensional slice — and both times, the right slice is dictated by the geometry of the problem, not chosen by us.

Finally, a word on where this stops. Restricting to characteristics works for *first order* equations. For second order equations the notion of characteristic survives (and gives the classification into elliptic, parabolic and hyperbolic) but no longer reduces the problem to ODEs. That is where a genuinely new set of ideas — energy estimates, Fourier analysis, function spaces, weak solutions — has to begin. But the geometric instinct built here is exactly the one that makes those ideas feel natural when you meet them.

Suggestions for further reading

1. W. Rudin, *Principles of Mathematical Analysis*, 3rd ed., McGraw–Hill. Chapter 9 is the source of our Definition 3.1 and of Theorems 4.2, 5.2, 5.5. Terse, but exactly right.

2. M. Spivak, *Calculus on Manifolds*. The same material with more geometric commentary.

3. A. K. Nandakumaran and P. S. Datti, *Partial Differential Equations: Classical Theory with a Modern Touch*, Cambridge University Press, 2020. Chapter 2 develops the method of characteristics for first order equations carefully and geometrically; our treatment in §§8, 9 follows its spirit.

4. L. C. Evans, *Partial Differential Equations*, 2nd ed., AMS. §3.2 for the full characteristic system including the fully nonlinear case; Chapter 11 for conservation laws, shocks and entropy.

5. F. John, *Partial Differential Equations*, Springer. A classic, very strong on the geometry of characteristics.

6. V. I. Arnold, *Ordinary Differential Equations*. For the geometric view of ODEs as vector fields and flows, unmatched.