An Introduction to First-Order Partial Differential Equations and the Method of Characteristics

A Foundation Course in Classical PDE Theory

Contents of the Course

Preface

A partial differential equation is an equation that an unknown function must satisfy involving its rates of change in several independent directions simultaneously. The challenge it poses is to find the function—or, given additional conditions, the unique function—that satisfies the relation at every point of a region. First-order partial differential equations, in which the highest derivatives appearing are of order one, are the natural entry point into this theory: they are simple enough to be fully tractable, yet rich enough to contain, in their purest and most visible form, the geometry that governs the whole subject.

The central question this course sets out to answer is how data prescribed on some surface in the domain propagates into the domain when a first-order partial differential equation is given. The answer is along curves determined by the equation itself—curves along each of which the partial differential equation reduces to an ordinary differential equation. These are the characteristics. The form they take, and the system of equations needed to find them, depends on how deeply the solution is coupled into the structure of the equation. A first-order equation in which the coefficients of the derivatives depend only on position is a different object from one in which they depend on the solution itself, and both are different from an equation that depends nonlinearly on the entire gradient. Each of these levels demands its own version of the method, and understanding why is the same as understanding the classification.

The course develops the method of characteristics across the full hierarchy of first-order equations—linear, semilinear, quasilinear, and fully nonlinear—treating each level not as a variation on the one before it but as a genuinely new setting in which a new structural feature of the method becomes necessary. The reader it is written for has met ordinary differential equations and the calculus of several variables; no prior knowledge of partial differential equations is assumed. The aim throughout is to equip the reader to take a first-order equation, identify its type, construct its characteristics, and carry the initial data along them to an explicit solution.

1. The Emergence of Partial Differential Equations

1.1. Functions of Several Variables and Partial Differential Equations

A function of a single variable $f(x)$ has, at each point x, a single rate of change: the derivative $f'(x)$, which measures how f responds to a small change in x. When a function $u(x,t)$ depends on two independent variables—say a spatial position x and a time t—it has, at each point (x,t), not one but two natural rates of change. The *partial derivative* $\partial u/\partial x$, written u_x, measures how u changes as x is varied while t is held fixed. The partial derivative $\partial u/\partial t$, written u_t, measures how u changes as t is varied while x is held fixed. These are computed exactly as ordinary derivatives, treating the other variable as a constant. For example, if $u(x,t) = x^2 t + \sin(xt)$, then

$$u_x = 2xt + t\cos(xt), \qquad u_t = x^2 + x\cos(xt).$$

A *partial differential equation* is an equation that an unknown function must satisfy involving one or more of its partial derivatives. The function is not known in advance; it is what is to be found. The equation imposes a constraint on how the function and its rates of change relate to one another at every point of its domain, and the problem is to identify all functions—or, given additional conditions, the unique function—that satisfy it.

This is structurally different from an ordinary differential equation, where the unknown depends on a single variable and the equation involves only ordinary derivatives. The presence of more than one independent variable means that the equation does not simply prescribe a single rate of change at each point; it prescribes a relationship among several rates of change, and their richness and their difficulty flows directly from this. A function satisfying a PDE has many degrees of freedom even after the equation is imposed, and singling out the right function requires additional conditions: data prescribed on some curve or surface in the domain.

1.2. The Wave Equation and the Question It Raised

The story of partial differential equations begins not in abstraction but in sound. In 1746, the French mathematician Jean le Rond d'Alembert set out to give a precise mathematical account of the motion of a vibrating string. If $u(x,t)$ denotes the transverse displacement of the string at position x and time t, then d'Alembert showed that u satisfies the relation

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2}, \tag{1}$$

where c is a constant related to the tension and density of the string. This is the one-dimensional wave equation, and its appearance marked the beginning of a new branch of mathematics. Equations in ordinary derivatives—in which the unknown depended on a single variable—had been the central object of analysis since Newton and Leibniz. Equation (1) was of a different kind: its unknown u depended on two independent variables and the equation related their partial derivatives.

D'Alembert found that the general solution of (1) takes the form $u(x,t) = f(x + ct) + g(x - ct)$ for two arbitrary functions f and g. The physical meaning is immediate: $f(x+ct)$ is a wave profile travelling to the left at speed c, and $g(x-ct)$ is a wave profile travelling to the right. The string's motion is a superposition of two counter-propagating waves, each maintaining its shape. But this result raised a question that would occupy mathematicians for a generation: what additional data—what values of u or its derivatives prescribed at some initial moment or on some curve—was needed to select a unique, physically meaningful solution from the infinity the equation allowed?

The debate that followed d'Alembert's work exposed a gap at the foundation of eighteenth-century analysis: the concept of a function had not been made sufficiently precise, and the theory of PDEs was demanding that precision. What the disagreement among d'Alembert, Euler, and Bernoulli ultimately clarified was the right question to ask: given a PDE, what additional conditions—what *data* prescribed on some curve or surface—are needed to select a unique solution, and what is the relationship between the structure of the equation and the geometry of the surface on which that data is given? These are the questions that Augustin-Louis Cauchy eventually formulated precisely in the nineteenth century, and they are precisely the questions the method of characteristics is designed to answer.

1.3. First-Order Equations and the Scope of This Course

By the nineteenth century, the study of PDEs had broadened considerably. Joseph Fourier's 1822 treatise on heat conduction introduced the heat equation $u_t = \kappa\, u_{xx}$, governing the diffusion of temperature; Laplace contributed the equation $u_{xx} + u_{yy} = 0$, governing gravitational and electrostatic potential in source-free regions. The central question that Cauchy and Riemann identified—given a PDE and data prescribed on a curve, when can the solution be determined in a neighbourhood of that curve?—would come to organise the whole theory.

First-order PDEs—equations involving only first-order partial derivatives—may appear simpler than the wave or heat equations, but they contain the geometry of characteristics in its purest and most visible form. They arise independently in a wide range of settings: transport of a substance carried by a fluid, the Hamilton–Jacobi equations of classical mechanics, geometric optics, and conservation laws all lead naturally to first-order equations.

The method of characteristics, developed here in the first-order setting, is the proper foundation for the treatment of second-order equations and everything beyond, making the theory developed in this course both complete in itself and the entry point into the broader subject.

2. General First-Order Equations and Their Classification

2.1. Partial Differential Equations and Their Classification

A partial differential equation of order k in n independent variables $x = (x_1, \dots, x_n)$ is defined as follows.

Definition 2.1. A *partial differential equation* (PDE) of order k in n independent variables $x = (x_1, \dots, x_n)$ is an equation of the form

$$F\big(x,\, u(x),\, Du(x),\, D^2u(x),\, \dots,\, D^k u(x)\big) = 0,$$

where $u\colon \Omega \subset \mathbb{R}^n \to \mathbb{R}$ is the unknown function, $Du = (\partial_{x_1} u, \dots, \partial_{x_n} u)$ is its gradient, and $D^j u$ denotes the collection of all partial derivatives of u of order exactly j.

The order of the PDE is the order of the highest derivative actually appearing. A first-order PDE in two variables x and t takes the form $F(x, t, u, u_x, u_t) = 0$. A *classical solution* of such an equation on an open set Ω is a function u that is continuously differentiable up to the order of the equation and satisfies the equation at every point of Ω. Throughout this course, all solutions are classical.

First-order PDEs fall into four types according to how deeply the solution u is coupled into the structure of the equation. This classification predicts, with precision, what the method of characteristics must do at each level and what can happen to solutions. Each type is defined below, followed by a remark on what distinguishes it from the one preceding it.

Definition 2.2. A first-order PDE in n variables is called *linear* if it has the form

$$\sum_{i=1}^{n} a_i(x)\, \frac{\partial u}{\partial x_i} + b(x)\, u = f(x),$$

where the coefficients a_i, b, and the right-hand side f depend only on the independent variables x.

In the linear case, the coefficients of the derivatives depend only on position. The directions along which the equation propagates information are therefore entirely fixed before any solution is sought, and the equation for u along each of those directions is linear. This is the most tractable level of the hierarchy.

Definition 2.3. A first-order PDE in n variables is called *semilinear* if it has the form

$$\sum_{i=1}^{n} a_i(x)\,\frac{\partial u}{\partial x_i} = f(x,u),$$

where the coefficients a_i of the derivatives depend only on x, but the right-hand side f may depend nonlinearly on u.

The semilinear case preserves the linear case's propagation directions: since the coefficients $a_i(x)$ still depend only on position, the geometry of how the equation carries information through the domain is unchanged. What changes is the behaviour of u along each direction, which is now governed by a nonlinear term. The characteristics can still be determined independently of the solution.

Definition 2.4. A first-order PDE in n variables is called *quasilinear* if it has the form

$$\sum_{i=1}^{n} a_i(x,u)\,\frac{\partial u}{\partial x_i} = f(x,u),$$

where the coefficients a_i are allowed to depend on u as well as on x.

The quasilinear case introduces a genuine coupling between the solution and the structure of the equation. Since the coefficients $a_i(x,u)$ now depend on u, the directions along which the equation propagates information can no longer be determined before the solution is known. The characteristics and the solution must be found together. This mutual dependence is the source of qualitatively new behaviour that the semilinear theory cannot exhibit.

Definition 2.5. A first-order PDE in n variables is called *fully nonlinear* if it takes the general form

$$F(x,u,Du) = 0,$$

where F depends nonlinearly on the gradient Du itself.

The fully nonlinear case is the most general. The equation depends nonlinearly on the entire gradient Du, which means that neither the propagation directions nor the solution can be extracted from the equation without simultaneously tracking how the gradient evolves. A new and larger space must be introduced to carry the method through, as will be seen in Section 6.

2.2. Hypersurfaces, Hyperplanes, and the Cauchy Problem

Two geometric terms must be in place before the central problem can be stated precisely.

Definition 2.6. A *hypersurface* in $\mathbb{R}^n$ is a smooth surface of dimension $n-1$. It is the natural generalisation, to n dimensions, of a curve in the plane ($n = 2$) and a surface in three-dimensional space ($n = 3$). Formally, a hypersurface is a set of the form $\Gamma = \{ x \in \Omega : \phi(x) = 0 \}$ where $\phi\colon \Omega \to \mathbb{R}$ is a smooth function with $\nabla\phi \neq 0$ on Γ.

Definition 2.7. A *hyperplane* in $\mathbb{R}^n$ is a flat hypersurface—one that can be written as $\{x : a \cdot x = c\}$ for some vector $a \in \mathbb{R}^n$ and constant $c \in \mathbb{R}$. A hyperplane is the special case of a hypersurface in which ϕ is a linear function. The surfaces $\{t = 0\}$ and $\{x_n = 0\}$ that appear throughout this course as initial surfaces are hyperplanes.

The distinction matters because the general theory applies to any smooth hypersurface, while the examples are typically stated with hyperplanes for concreteness. Whenever the course refers to a general initial surface, it means a hypersurface; whenever an example sets $t = 0$ or $x_n = 0$, it is working with the simplest possible case of a hyperplane.

Definition 2.8. The *Cauchy problem* for a first-order PDE consists of finding a function u satisfying the equation in a neighbourhood of a hypersurface $\Gamma \subset \Omega$, subject to the condition $u\big|_{\Gamma} = u_0$, where u_0 is a given function on Γ. The surface Γ is called the *initial surface* and u_0 the *initial data* or *Cauchy data*.

This formulation encompasses the classical initial-value problem, where $\Gamma = \{t = 0\}$ and the data is the profile of u at time zero, as well as more general geometric settings where the surface on which data is prescribed is curved or inclined. The question the method of characteristics answers is how the data on Γ propagates into the domain, and under what conditions on Γ the method succeeds.

3. Linear Equations and the Method of Characteristics

3.1. Why the Characteristics Are the Right Object to Study

Content not available in this sample version.

3.2. The Noncharacteristic Condition

Content not available in this sample version.

3.3. Worked Examples in the Linear Setting

Content not available in this sample version.

4. Semilinear Equations and Fixed Characteristic Propagation

4.1. How the Semilinear Case Differs from the Linear

Content not available in this sample version.

4.2. Worked Examples of Semilinear Equations

Content not available in this sample version.

5. Quasilinear Equations and Coupled Characteristics

5.1. How the Quasilinear Case Differs from the Semilinear

Content not available in this sample version.

5.2. Worked Examples of Quasilinear Equations

Content not available in this sample version.

6. Fully Nonlinear Equations and Extended Characteristics

6.1. Why the Characteristics Must Be Extended into a Larger Space

Content not available in this sample version.

6.2. The Extended Characteristic System

Content not available in this sample version.

6.3. Worked Examples of Fully Nonlinear Equations

Content not available in this sample version.

7. What the Method of Characteristics Has Achieved

Content not available in this sample version.