

Mathematical Methods in the Physical Sciences II
Georgia Tech
MATH 6702 Spring 2023

John McCuan

April 4, 2023

Contents

1	Introduction	9
1.1	Ordinary Derivatives and Partial Derivatives	10
1.2	Ordinary differential equations and PDE	35
1.3	Wave equation	54
1.4	Heat Equation	65
1.5	Laplace's equation	66
2	Integration	69
2.1	Evaluation of integrals	72
2.2	Divergence and the divergence theorem	75
3	PDE: A Quick and Dirty Intro	83
3.1	Laplace's equation	85
3.1.1	Derivation	87
3.1.2	Separated variables solutions	94
3.1.3	Integral identities and properties	104
3.2	The heat equation	112
3.2.1	Derivation	112
3.2.2	separated variables solutions	117
3.2.3	integral identities	118
3.3	The wave equation	120
3.3.1	derivation	120
3.3.2	separated variables solutions	122
3.3.3	integral identities	126
4	Appendix 1: Partial Derivatives	129

Preface to the Student

These notes from Spring semester 2023 go along with the course MATH 6702 (Georgia Tech) entitled Math Methods of Applied Sciences II. The course is supposed to consist of a review of multivariable calculus (Calc III) and an introduction to partial differential equations. I also usually include a couple additional topics and (I expect) this semester will be no exception. Those topics are some elementary aspects of modeling physical systems and an introduction to the calculus of variations. A relatively large number of students have reported to me that these additional topics were among the most interesting and, later, most valuable topics they encountered in coursework, and I hope the same holds for you...as long as you don't use them for anything too destructive.

I will also refer to a few texts. The only ones you might want to purchase for the course are *Mathematical Methods in the Physical Sciences* by Mary Boas and *Applied Partial Differential Equations with Fourier Series and Boundary Value Problems* by Richard Haberman. Technically, neither is "required," and any edition should do. (Probably the oldest edition you can find of Haberman is the best one.) I will type up homework assignments, and the problems I suggest you work on (and turn in) will be stated in relatively complete form, though many are adapted from the text(s). The books will be for reference (along with these notes). Other good books are *Partial Differential Equations* by L.C. Evans and *Variational Calculus and Optimal Control* by John Troutman. I may refer to those occasionally.

Start with the homework assignment problems and try to connect them to something you know. Figure out what you understand (mathematically, physically, etc.) and try to build on that. I may not know some things you understand, so I'll ask you to explain those things. If something I say does not connect to anything you know, you should try to ask a question to make it connect, and I will try to help you make a connection. **Ask questions.**

Mathematics is about communication, and you should think of yourself as responsible for about half of the communication. If you have no good question to ask, one

wonders why you are taking the course. If you are taking the course because some administrator decided it was a “requirement,” then it would probably be better for you to find a different course or a different instructor, or wait until you accrue some better motivation.

Preface

These are notes for MATH 6702 (Math Methods in Applied Science II) offered in the Spring semester of 2023 at Georgia Institute of Technology. The text for the course was *Mathematical Methods in the Physical Sciences* by Mary Boas.

Chapter 1

Introduction

A **physical system** is some phenomenon which you and I can both look at, discuss, and “compare notes on” so to speak. In particular, we can use words to describe what we “see,” and especially measure. Hopefully, this results in some kind of common understanding of the phenomenon (the physical system). Among the choices of language, or words, we can use for our discussion, **mathematics** is a possible choice. It is the language of choice for this course. Often this choice of language leads to a kind of parallel entity called a **mathematical model**. The mathematical model is not the physical system, but hopefully some aspects of the model can be compared to (especially) measurements associated with the physical system. When two humans have a discussion about the **comparison** of a mathematical model to a physical system, the result, i.e., the resulting discussion, is what we call **science**.

Note that the identification of the physical system with the mathematical model, or the identification of the mathematical model with the physical system, is not science. That is where science ends and deception begins. Science is about understanding why and how the model is different from the physical system, rather than the assumption (or presumption) that the two are the same.

If we want to have a meaningful scientific discussion, that is involve ourselves with meaningful science, then we should make an attempt to use precise language. There are limitations, but many have felt some of the options we can (will?) discuss are compelling.

1.1 Ordinary Derivatives and Partial Derivatives

A basic object of study in this course is the **real valued function** $f : A \rightarrow \mathbb{R}$ where A is some set we consider as the **domain** of the function f . If $A = (a, b)$ is an open interval in the real line \mathbb{R} and $x \in (a, b)$, then for $h \in \mathbb{R}$ with $|h|$ small enough we can consider the **difference quotient(s)**

$$\frac{f(x+h) - f(x)}{h}. \quad (1.1)$$

Exercise 1.1 Draw the **graph**

$$\mathcal{G} = \{(x, f(x)) \in \mathbb{R}^2 : x \in (-2, 1)\}$$

of the function $f : (-2, 1) \rightarrow \mathbb{R}$ with values given by

$$f(x) = \frac{1}{(x+2)(1-x)}.$$

Do a good job: Label the axes, make sure the minimum value is in the correct location and has the correct value; get the monotonicity and concavity correct. Perhaps you will wish to use a mathematical software package/program like Mathematica, Maple, or Matlab (or octave).

On the graph you have drawn, illustrate the difference quotient at the point $x = 0$ with $h = 1/2$.

For functions like this it makes sense to consider the limit (or at least try to take the limit) of the difference quotient(s) as h tends to zero:

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}. \quad (1.2)$$

If the limit exists, we call the value $f'(x)$ the **ordinary derivative** of f at x and say f is **differentiable at the point** $x \in (a, b)$. For many functions we consider, the ordinary derivative will exist and be well-defined at every point x in the domain interval (a, b) , or at most points in (a, b) .

Exercise 1.2 (absolute value) Consider the function $g : \mathbb{R} \rightarrow \mathbb{R}$ given by

$$g(x) = \begin{cases} -x, & \text{if } x \leq 0 \\ x, & \text{if } x \geq 0. \end{cases} \quad (1.3)$$

- (a) Find the value of the derivative $g'(x)$ for $x \in (-\infty, 0) \cup (0, \infty)$.
- (b) Show the derivative of g is not well-defined at $x = 0$.
- (c) Compute the **left and right derivatives**

$$\lim_{h \nearrow 0} \frac{g(x+h) - g(x)}{h} = \lim_{h \rightarrow 0^-} \frac{g(x+h) - g(x)}{h}$$

and

$$\lim_{h \searrow 0} \frac{g(x+h) - g(x)}{h} = \lim_{h \rightarrow 0^+} \frac{g(x+h) - g(x)}{h}$$

for all $x \in \mathbb{R}$.

If we let $g' : \mathbb{R} \rightarrow \{-1, 1\}$ denote the right derivative of g , then another important function $h : \mathbb{R} \rightarrow \{0, 1\}$ has values given by

$$h(x) = \frac{1}{2}(g'(x) + 1).$$

Continuity

A more elementary property of (some) functions $f : (a, b) \rightarrow \mathbb{R}$ is **continuity**. Such a function is said to be **continuous at a point** $x \in (a, b)$ if

$$\lim_{h \rightarrow 0} f(x+h) = f(x). \quad (1.4)$$

Note: When the symbol h is used to represent a quantity assumed to be small in absolute value, or tending to zero, as in (1.1), (1.2), or (1.4) we call the quantity an **increment**. Sometimes it is convenient to use different symbols for an increment.

Exercise 1.3 (Heaviside function) Consider the function $h : (-\infty, \infty) \rightarrow \mathbb{R}$ by

$$h(x) = \begin{cases} 0, & \text{if } x < 0 \\ 1, & \text{if } x \geq 0. \end{cases} \quad (1.5)$$

- (a) Find the value of the derivative $h'(x)$ for $x \in (-\infty, 0) \cup (0, \infty)$.
- (b) Show the derivative of g is not well-defined at $x = 0$.
- (c) Attempt to compute the left and right derivatives of h .

(d) Show h is not continuous at $x = 0$.

(e) Compute the **left and right limits**

$$\lim_{k \nearrow 0} h(x+k) = \lim_{k \rightarrow 0^-} h(x+k) \quad \text{and} \quad \lim_{k \searrow 0} h(x+k) = \lim_{k \rightarrow 0^+} h(x+k)$$

for all $x \in \mathbb{R}$.

If we let $h' : \mathbb{R} \rightarrow \{0, \infty\}$ denote the left derivative of h , then h' is not a real valued function, but it is a (sort of) interesting **extended real valued function**.

A function $f : (a, b) \rightarrow \mathbb{R}$ is said to be **continuous on the interval** (a, b) if f is continuous at each point $x \in (a, b)$. In this case, we write $f \in C^0(a, b)$ and f is said to be C^0 . When you read this it should sound like “ f is ‘see zero’.”

Similarly, a function is said to be **differentiable on the interval** (a, b) if f is differentiable at each point $x \in (a, b)$. In this case, the values $f'(x)$ define for us a (new, second) function to consider (and talk about). If the function $f' : (a, b) \rightarrow \mathbb{R}$ is well-defined **and continuous**, we say f is **continuously differentiable** and write

$$f \in C^1(a, b).$$

Can you guess how to read this?

Exercise 1.4 (ordinary differentiability) Find an example of a function $f : (-\infty, \infty) \rightarrow \mathbb{R}$ which is differentiable (at every point $x \in \mathbb{R}$) but is **not** continuously differentiable.

Any function f giving a solution to Exercise 1.4 **must** be continuous:

Theorem 1 (ordinary differentiability) If $f : (a, b) \rightarrow \mathbb{R}$ is differentiable on (a, b) , then $f \in C^0(a, b)$.

Proof: Let $x \in (a, b)$. Then for any $\epsilon > 0$, there is some $\delta > 0$ so that

$$\left| f'(x) - \frac{f(x+h) - f(x)}{h} \right| < \epsilon \quad \text{if} \quad 0 < |h| < \delta. \quad (1.6)$$

Note carefully that the condition given in (1.6) is precisely what

$$\lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} = f'(x)$$

means. Now we estimate:

$$\begin{aligned} |f(x+h) - f(x)| &= |f(x+h) - f(x) - hf'(x) + hf'(x)| \\ &\leq |f(x+h) - f(x) - hf'(x)| + |hf'(x)| \end{aligned} \quad (1.7)$$

$$\begin{aligned} &= |h| \left(\left| f'(x) - \frac{f(x+h) - f(x)}{h} \right| + |f'(x)| \right) \\ &< \delta(\epsilon + |f'(x)|). \end{aligned} \quad (1.8)$$

Given any $\eta > 0$, we can take $\epsilon < |f'(x)| + 1$ in (1.6) and we can also assume the tolerance δ , for which (1.6) holds, satisfies

$$\delta < \frac{\eta}{2|f'(x)| + 1}.$$

Then following along from (1.8) we have that $0 < |h| < \delta$ implies

$$\begin{aligned} |f(x+h) - f(x)| &< \frac{\eta}{2|f'(x)| + 1} (|f'(x)| + 1 + |f'(x)|) \\ &= \eta. \end{aligned}$$

This is precisely what it means to say

$$\lim_{h \rightarrow 0} f(x+h) = f(x).$$

This means f is continuous at x for every $x \in (a, b)$ and, hence, $f \in C^0(a, b)$.

Note: We used something called the **triangle inequality** in (1.7). You can't get very far with the precise mathematical analysis of most physical systems without the triangle inequality.

A Physical System

I have a coil of spring steel, commonly called a Slinky[™]. I also have a 5 foot long vertical “yard stick” to the top of which I have attached a bracket to accommodate hanging the Slinky so that it dangles down from the top end. Here are some starting assertions (assumptions) with which you are free to agree or disagree.

- (i) My hanging Slinky constitutes a physical system which can be compared to a mathematical model.

- (ii) My hanging Slinky constitutes a **nice example** of a physical system which can be compared to a mathematical model, and it is an **interesting example**.
- (iii) A natural element of a mathematical model for the hanging Slinky is a real valued function of one real variable $f : (a, b) \rightarrow \mathbb{R}$ as discussed above.

If you disagree with the first assertion, then probably this is not the best course for you to take. If you disagree with the second assertion, then we can certainly discuss that, so at least we have some place to start. If, on the other hand, you agree that there is some interest in this hanging slinky, then that should make the following exercise easier:

Exercise 1.5 Determine an interval $(a, b) \subset \mathbb{R}$ and a measurement function $f : (a, b) \rightarrow \mathbb{R}$ which you think are appropriate to describe the shape/position of a hanging (or otherwise extended) Slinky. Describe any properties of this real valued function of one variable, e.g, boundary values $f(a)$, $f(b)$, monotonicity, differentiability, concavity, you would expect this function to exhibit if any reasonable comparison between a model involving your measurement function f and the physical system (Slinky) is possible.

Functions of several variables and partial derivatives

If $f : A \rightarrow \mathbb{R}$ is a function of several real variables x_1, x_2, \dots, x_n instead of just one real variable x , then the discussion surrounding continuity and differentiability gets a little bit more complicated. We can start with **partial derivatives** which should be familiar from calculus. Let's assume a point $(x_1, x_2, \dots, x_n) \in A \subset \mathbb{R}^n$ has the property that there is some $\delta > 0$ for which

$$\{(x_1, x_2, \dots, x_j + h, \dots, x_n) \in \mathbb{R}^n : |h| < \delta\} \subset A. \quad (1.9)$$

Exercise 1.6 Draw the set

$$\{(x_1, x_2, \dots, x_j + h, \dots, x_n) \in \mathbb{R}^n : |h| < \delta\}$$

appearing in (1.9) when

- (a) $n = 2$, $(x_1, x_2) = (2, 1)$, and $\delta = 1/3$.
- (b) $n = 3$, $(x_1, x_2, x_3) = (1, 0, 1/2)$, and $\delta = 1/4$.

Under the assumption (1.9) we can (again) consider a difference quotient

$$\frac{f(x_1, \dots, x_j + h, \dots, x_n) - f(x_1, \dots, x_n)}{h}.$$

Some notation (from linear algebra)

Writing (and typing) points like

$$(x_1, \dots, x_j + h, \dots, x_n) - (x_1, \dots, x_n) = (0, \dots, h, \dots, 0) \quad (1.10)$$

is a little cumbersome (and also a little annoying) sometimes. Recall the **standard unit basis vectors** in \mathbb{R}^n :

$$\mathbf{e}_1 = (1, 0) \text{ and } \mathbf{e}_2 = (0, 1) \quad \text{when } n = 2$$

$$\mathbf{e}_1 = (1, 0, 0), \mathbf{e}_2 = (0, 1, 0) \text{ and } \mathbf{e}_3 = (0, 0, 1) \quad \text{when } n = 3$$

$$\mathbf{e}_1 = (1, 0, 0, 0), \mathbf{e}_2 = (0, 1, 0, 0), \mathbf{e}_3 = (0, 0, 1, 0) \text{ and } \mathbf{e}_4 = (0, 0, 0, 1) \quad \text{when } n = 4$$

⋮

Using these let us write

$$\mathbf{x} = (x_1, x_2, \dots, x_n) = \sum_{j=1}^n x_j \mathbf{e}_j \in \mathbb{R}^n$$

so that (1.10) becomes

$$(\mathbf{x} + h\mathbf{e}_j) - \mathbf{x} = h\mathbf{e}_j.$$

You may recall that in linear algebra one usually uses the transposes

$$\left\{ \mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\} \subset \mathbb{R}^2$$

$$\left\{ \mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\} \subset \mathbb{R}^3$$

$$\vdots$$

of these points/vectors because these are more convenient for matrix multiplication. We will use the transposes sometimes too; if we want to be careful, we should make it clear by indicating the transposes explicitly like this:

$$\text{If } \mathbf{e}_1 = (1, 0, 0), \quad \text{then} \quad \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = (1, 0, 0)^T = \mathbf{e}_1^T,$$

but

$$\text{if } \mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \text{ then } (1, 0, 0) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}^T = \mathbf{e}_1^T.$$

(Sometimes it just doesn't make much difference, but sometimes it does.)

Given $f : A \rightarrow \mathbb{R}$ with $\mathbf{x} \in A \subset \mathbb{R}^n$, $j \in \{1, 2, \dots, n\}$, and $\delta > 0$ for which

$$\{\mathbf{x} + h\mathbf{e}_j : |h| < \delta\} \subset A, \quad (1.11)$$

the j -th difference quotient of f at \mathbf{x} is

$$\frac{f(\mathbf{x} + h\mathbf{e}_j) - f(\mathbf{x})}{h}$$

and the j -th partial derivative of f at \mathbf{x} (if it exists) is

$$\frac{\partial f}{\partial x_j} = \frac{\partial f}{\partial x_j}(\mathbf{x}) = \lim_{h \rightarrow 0} \frac{f(\mathbf{x} + h\mathbf{e}_j) - f(\mathbf{x})}{h}.$$

Exercise 1.7 Draw the graph

$$\mathcal{G} = \{(x, y, f(x, y)) \in \mathbb{R}^3 : (x, y) \in A\}$$

of $f : A \rightarrow \mathbb{R}$ where

$$A = (0, 1) \times (0, 2) \rightarrow \{(x, y) : 0 < x < 1, 0 < y < 2\}$$

is an open rectangle in \mathbb{R}^2 and

$$f(x, y) = 1 + x^2 + \frac{1}{10}y^2.$$

Exercise 1.8 Draw the graph of $u : A \rightarrow \mathbb{R}$ where $A = (0, 1) \times (0, 2)$ and

$$u(x, y) = 1 + x^2.$$

and illustrate (in your drawing) the difference quotients for u at $(1/2, 1)$ with $h = 1/4$.

As in the case of ordinary derivatives, one convenient way to ensure all sets of the form of the set considered in Exercise 1.6 (for small $|h|$) are in the domain of the function is to require the domain set to be **open**.

Definition 1 A set $U \subset \mathbb{R}^n$ is **open** if for each

$$\mathbf{p} = \sum_{j=1}^n p_j \mathbf{e}_j \in U,$$

there is some $\delta > 0$ for which

$$Q_\delta(\mathbf{p}) = \left\{ \mathbf{x} = \sum_{j=1}^n x_j \mathbf{e}_j : |x_j - p_j| < \frac{\delta}{2} \right\} \subset U.$$

The set $Q_\delta(\mathbf{p})$ is called the **open cube of side length δ centered at \mathbf{p}** .

Exercise 1.9 Draw $Q_1(1/2, 1/2) \subset \mathbb{R}^2$ and $Q_{1/2}(1, 1, 1) \subset \mathbb{R}^3$.

Exercise 1.10 Show that for $\delta > 0$ and $\mathbf{p} \in \mathbb{R}^n$, the cube $Q_\delta(\mathbf{p})$ is open.

Exercise 1.11 Show that if $-\infty \leq a < b \leq \infty$, then the interval (a, b) is open in \mathbb{R} .

Exercise 1.12 Find an open set in \mathbb{R} which is not an open interval.

Exercise 1.13 Given $r > 0$ and $\mathbf{p} \in \mathbb{R}^n$, the set

$$B_r(\mathbf{p}) = \{ \mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{p}\| < r \}$$

where

$$\|\mathbf{x} - \mathbf{p}\| = \sqrt{\sum_{j=1}^n (x_j - p_j)^2}$$

is called the **Euclidean ball**. Show a Euclidean ball is open.

Definition 2 Given an open set $U \subset \mathbb{R}^n$, a function $f : U \rightarrow \mathbb{R}$ is said to be **partially differentiable** (or partial differentiable) **at the point $\mathbf{p} \in U$** if

$$\frac{\partial f}{\partial x_j}(\mathbf{p}) = \lim_{h \rightarrow 0} \frac{f(\mathbf{p} + h\mathbf{e}_j) - f(\mathbf{p})}{h} \quad \text{exists for } j = 1, 2, \dots, n.$$

The same function is said to be **partially differentiable on the open set U** if f is partially differentiable at each point $\mathbf{p} \in U$.

Partial differentiability does not imply continuity.¹

Definition 3 (continuity) Given $f : U \rightarrow \mathbb{R}$ with U an open set in \mathbb{R}^n , we say f is **continuous at the point** $\mathbf{p} \in U$ if

$$\lim_{\mathbf{q} \rightarrow \mathbf{p}} f(\mathbf{q}) = f(\mathbf{p}). \quad (1.12)$$

The limit condition (1.12) means (precisely) this:

For any $\epsilon > 0$, there exists some $\delta > 0$ such that

$$\|\mathbf{q} - \mathbf{p}\| < \delta \quad \text{implies} \quad |f(\mathbf{q}) - f(\mathbf{p})| < \epsilon.$$

If $f : U \rightarrow \mathbb{R}$ is continuous at each point \mathbf{p} in an open set $U \subset \mathbb{R}^n$, we say f is **continuous on the open set** U , and we write

$$f \in C^0(U).$$

Note: While the notion of (partial) differentiability mostly makes sense (and is almost always applied to) functions $f : U \rightarrow \mathbb{R}$ with U open in \mathbb{R}^n , the notion of continuity can be (quite easily) generalized to arbitrary domains:

Definition 4 (continuity) Given **any set** $A \subset \mathbb{R}^n$ and a function $f : A \rightarrow \mathbb{R}$, we say f is **continuous at the point** $\mathbf{p} \in A$ if

$$\lim_{A \ni \mathbf{q} \rightarrow \mathbf{p}} f(\mathbf{q}) = f(\mathbf{p}),$$

that is, for any $\epsilon > 0$, there is some $\delta > 0$ for which

$$\left. \begin{array}{l} \mathbf{q} \in A \\ \|\mathbf{q} - \mathbf{p}\| < \delta \end{array} \right\} \quad \text{implies} \quad |f(\mathbf{q}) - f(\mathbf{p})| < \epsilon.$$

Exercise 1.14 State precisely (the definition of) what it means to have $f \in C^0(A)$ for $f : A \rightarrow \mathbb{R}$ and A any subset of \mathbb{R}^n .

I wrote above that partial differentiability does not imply $f \in C^0(U)$. Here is an example: Consider $f : (-1, 1) \times (-1, 1) \rightarrow \mathbb{R}$ by

$$f(x, y) = \begin{cases} 1, & x = 0 \\ 1, & y = 0 \\ 0, & xy \neq 0. \end{cases} \quad (1.13)$$

¹And this may be unexpected in view of Theorem 1.

Exercise 1.15 Show $f : (-1, 1) \times (-1, 1) \rightarrow \mathbb{R}$ with values given by (1.13) is partially differentiable at every point in $U = (-1, 1) \times (-1, 1)$ but not continuous at the points in

$$X = \{(x, y) \in U : xy = 0\}.$$

Exercise 1.16 Using condition (1.12) in the case $n = 1$, show the definition of continuity you get in this special case is precisely the same as the notion of continuity defined by (1.4).

Exercise 1.17 Generalize Theorem 1 as follows: If U is an open subset of \mathbb{R} (see Exercise 1.12) and $f : U \rightarrow \mathbb{R}$ is differentiable at each $x \in U$, then (show) f is continuous on U , i.e., $f \in C^0(U)$.

Definition 5 Given an open set $U \subset \mathbb{R}^n$ and a partially differentiable function $u : U \rightarrow \mathbb{R}$, we say u is **continuously differentiable** if

$$\frac{\partial u}{\partial x_j} \in C^0(U) \quad \text{for} \quad j = 1, 2, \dots, n.$$

In this case, we write $u \in C^1(U)$.

Theorem 2 $C^1(U) \subset C^0(U)$.

In order to prove this result, which one should note gives a weaker conclusion than Theorem 1 in the special case when $n = 1$ and $U = (a, b)$ is an open interval, we will need to recall a little more about functions of a single variable. In order to motivate the discussion, let us consider first the quantity in (1.12) which we will need to estimate in order to show a function $u \in C^1(U)$ is continuous, namely, $|u(\mathbf{q}) - u(\mathbf{p})|$. In one spatial dimension $n = 1$ it is convenient to write $\mathbf{p} = x$ and $\mathbf{q} = x + h$ with $|h|$ small enough so $x + h \in U$. More precisely, we can (and more or less should) take $\delta > 0$ small enough so that $x + h \in U$ whenever $|h| < \delta$. With this minor change of notation, the quantity we want to estimate takes the form

$$|u(\mathbf{q}) - u(\mathbf{p})| = |u(x + h) - u(x)|.$$

We showed in the proof of Theorem 1, as you should have reviewed in Exercise 1.17, that this quantity must become small when $|h|$ is small in order for the limit of the difference quotient

$$\lim_{h \rightarrow 0} \frac{u(x + h) - u(x)}{h} = u'(x) \tag{1.14}$$

to be well-defined. We have more information here, namely, that $u'(\xi)$ exists for all ξ with $x - h \leq \xi \leq x + h$ or even with $x - \delta < \xi < x + \delta$, and we want to make a more precise conclusion for reasons I will explain in a moment. That more precise conclusion is the conclusion of the **mean value theorem** and it tells us, for example, that if $0 < h < \delta$, then not only does the limit equality (1.14) hold, but that for some ξ with $x < \xi < x + h$, there holds

$$\frac{u(x+h) - u(x)}{h} = u'(\xi).$$

Here is a more general precise statement:

Theorem 3 (1-D mean value theorem) If $a, b \in \mathbb{R}$ with $a < b$ and $f : [a, b] \rightarrow \mathbb{R}$ satisfies $f \in C^1(a, b) \cap C^0[a, b]$, then there is some $\xi \in (a, b)$ such that the **difference quotient**

$$\frac{f(b) - f(a)}{b - a}$$

satisfies

$$\frac{f(b) - f(a)}{b - a} = f'(\xi).$$

Exercise 1.18 Let $a, b \in \mathbb{R}$ with $a < b$. Draw pictures to illustrate the following assertions:

(a) If $f \in C^1(a, b)$ and $\delta > 0$ is a positive number for which $x + h \in (a, b)$ whenever $|h| < \delta$, the given h with $-\delta < h < 0$, there is some $\xi \in (a, b)$ with

$$f'(\xi) = \frac{f(x+h) - f(x)}{h}. \quad (1.15)$$

Identify and indicate in your picture the largest possible value of δ and a particular subinterval of length $|h|$ on which you can be sure to find a point ξ for which (1.15) holds.

(b) If $f \in C^1(a, b) \cap C^0[a, b]$, then there is some $\xi \in (a, b)$ such that

$$f'(\xi) = \frac{f(b) - f(a)}{b - a}.$$

Let us now return to the quantity

$$|u(\mathbf{q}) - u(\mathbf{p})|$$

when $n > 1$. First, let us assume $\delta_0 > 0$ is small enough so that the cube $Q_{\delta_0}(\mathbf{p}) \subset U$. Let us also assume $\|\mathbf{q} - \mathbf{p}\| < \delta_0/2$ so that $\mathbf{q} \in Q_{\delta_0}(\mathbf{p})$.

Exercise 1.19 Show that for any $\mathbf{p} \in \mathbb{R}^n$ there holds

$$B_r(\mathbf{p}) \subset Q_{2r}(\mathbf{p}) \subset B_{r\sqrt{n}}(\mathbf{p}).$$

For $\mathbf{q} \in Q_{\delta_0}(\mathbf{p})$, we can write all the points on the segment between \mathbf{p} and \mathbf{q} as $(1-t)\mathbf{p} + t\mathbf{q} = \mathbf{p} + t(\mathbf{q} - \mathbf{p})$ for $0 \leq t \leq 1$. In particular, the quantity we wish to estimate can be written as

$$|u(\mathbf{q}) - u(\mathbf{p})| = \left| u(\mathbf{p} + t(\mathbf{q} - \mathbf{p})) \Big|_{t=0}^1 \right|.$$

That is, we can express $|u(\mathbf{q}) - u(\mathbf{p})|$ in terms of evaluation of the function $f : [0, 1] \rightarrow \mathbb{R}$ by $f(t) = u(\mathbf{p} + t(\mathbf{q} - \mathbf{p}))$ at the endpoints $t = 0$ and $t = 1$:

$$|u(\mathbf{q}) - u(\mathbf{p})| = \left| f(t) \Big|_{t=0}^1 \right|.$$

Here then is the main result we need:

Lemma 1 Given $u \in C^1(U)$ with points $\mathbf{q}, \mathbf{p} \in Q_{\delta_0}(\mathbf{p}) \subset U$, there is some $\delta_1 > 0$ such that the function $f : [-\delta_1, 1 + \delta_1] \rightarrow \mathbb{R}$ with values given by

$$f(t) = u(\mathbf{p} + t(\mathbf{q} - \mathbf{p}))$$

is well-defined and satisfies $f \in C^1(-\delta_1, 1 + \delta_1)$ and

$$f'(t) = \sum_{j=1}^n (q_j - p_j) \frac{\partial u}{\partial x_j}(\mathbf{p} + t(\mathbf{q} - \mathbf{p})).$$

Proof: In order to see what is going on here, let us consider the case $n = 2$. In that case, we can write $\mathbf{p} = (x, y)$ and $\mathbf{q} = (x + h, y + k)$ for some fixed h and k . The values of f take the form

$$f(t) = u(x + th, y + tk),$$

and a difference quotient for f looks like

$$\frac{f(t + \eta) - f(t)}{\eta} = \frac{u(x + (t + \eta)h, y + (t + \eta)k) - u(x + th, y + tk)}{\eta}. \quad (1.16)$$

Notice that if $|\eta|$ is small enough, then $(x + (t + \eta)h, y + (t + \eta)k)$ is still in $Q_{\delta_0}(\mathbf{p}) \subset U$ and the point $R = (x + th, y + (t + \eta)k)$ will also be in U along with all the points on the segments from $\mathbf{p} + t(h, k) = (x + th, y + tk)$ to R and from R to $\mathbf{p} + (t + \eta)(h, k) = (x + (t + \eta)h, y + (t + \eta)k)$. This is illustrated in Figure 1.1.

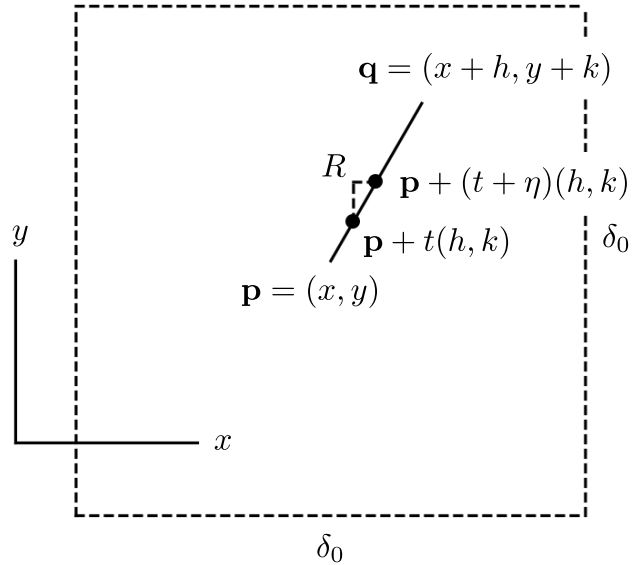


Figure 1.1: Points involved in the domain of u when taking a difference quotient with increment η of the function f .

Exercise 1.20 The claim was made above that for $|\eta|$ small enough the point $R = \mathbf{p} + t(h, k) + \eta(h, k)$ satisfies $R \in Q_{\delta_0}(\mathbf{p})$. Illustrate the relation between $\eta > 0$ and the tolerance δ_1 using the drawing in Figure 1.1. Use the triangle inequality to verify the claim concerning the point R and the assertion in the statement of Lemma 1 concerning the tolerance δ_1 and the domain of definition of the function f .

Introducing the point R into the difference quotient (1.16) for f we can write

$$\begin{aligned} & \frac{u(x + (t + \eta)h, y + (t + \eta)k) - u(x + th, y + tk)}{\eta} \\ &= \frac{u(x + (t + \eta)h, y + (t + \eta)k) - u(x + th, y + (t + \eta)k)}{\eta} \\ & \quad + \frac{u(x + th, y + (t + \eta)k) - u(x + th, y + tk)}{\eta}. \end{aligned}$$

Notice the second quotient is a standard partial difference quotient based at the point

$\mathbf{p} + t(h, k) = (x + th, y + tk)$ for the function u . In particular,

$$\begin{aligned} \lim_{\eta \rightarrow 0} \frac{u(x + th, y + (t + \eta)k) - u(x + th, y + tk)}{\eta} \\ &= \lim_{\eta \rightarrow 0} k \frac{u(x + th, y + tk + \eta k) - u(x + th, y + tk)}{\eta k} \\ &= k \frac{\partial u}{\partial x_2}(x + th, y + tk) \end{aligned}$$

because ηk tends to zero when η tends to zero. The first quotient

$$\frac{u(x + (t + \eta)h, y + (t + \eta)k) - u(x + th, y + (t + \eta)k)}{\eta}$$

is, on the face of it, more problematic. It looks something like a difference quotient for the x_1 partial derivative of u with increment η , but when η changes, the common second argument $y + (t + \eta)k$ changes as well, so there is no fixed base point. This is where the mean value theorem comes in: Let $g(\xi) = u(x + \xi h, y + (t + \eta)k)$. The domain interval for g depends on whether $\eta < 0$ or $\eta > 0$. Let us take the case $\eta > 0$ as illustrated in Figure 1.1. Then $g : [t, t + \eta] \rightarrow \mathbb{R}$. The fact that $u \in C^1(Q_{\delta_0}(\mathbf{p}))$ tells us $g \in C^1(t, t + \eta) \cap C^0[t, t + \eta]$ with derivative

$$g'(\xi) = h \frac{\partial u}{\partial x_1}(x + \xi h, y + (t + \eta)k) \quad (1.17)$$

Exercise 1.21 Write a difference quotient for g and carefully compute an appropriate limit to verify (1.17).

The mean value theorem applied to g tells us that ξ can be found with $t < \xi < t + \eta$ and

$$\frac{u(x + (t + \eta)h, y + (t + \eta)k) - u(x + th, y + (t + \eta)k)}{\eta} = h \frac{\partial u}{\partial x_1}(x + \xi h, y + (t + \eta)k).$$

Thus, the original difference quotient for f takes the form

$$\begin{aligned} \frac{f(t + \eta) - f(t)}{\eta} &= h \frac{\partial u}{\partial x_1}(x + \xi h, y + (t + \eta)k) \\ &\quad + \frac{u(x + th, y + (t + \eta)k) - u(x + th, y + tk)}{\eta}. \end{aligned}$$

When η tends to zero ξ tends to t and $\partial u / \partial x_1$ is continuous, so

$$f'(t) = \lim_{\eta \rightarrow 0} \frac{f(t + \eta) - f(t)}{\eta} = h \frac{\partial u}{\partial x_1}(x + th, y + tk) + k \frac{\partial u}{\partial x_2}(x + th, y + tk).$$

Exercise 1.22 Verify that this is the conclusion of Lemma 1 when $n = 2$.

Hopefully, the discussion above makes one proof of Lemma 1 (in the general case) more or less clear. We can move coordinates one at a time and use the mean value theorem. To be specific, if we take a difference quotient

$$\frac{f(t+h) - f(t)}{h} = \frac{u(\mathbf{p} + (t+h)(\mathbf{q} - \mathbf{p})) - u(\mathbf{p} + t(\mathbf{q} - \mathbf{p}))}{h},$$

we can write this as a sum of quotients

$$\frac{f(t+h) - f(t)}{h} = \frac{u(\mathbf{p} + (t+h)(\mathbf{q} - \mathbf{p})) - u(\mathbf{p} + t(\mathbf{q} - \mathbf{p}))}{h} = \sum_{j=1}^n Q_j.$$

In the last quotient only the last coordinate moves:

$$Q_n = \frac{u(\mathbf{p} + t(\mathbf{q} - \mathbf{p}) + h(q_n - p_n)\mathbf{e}_n) - u(\mathbf{p} + t(\mathbf{q} - \mathbf{p}))}{h}.$$

This is a standard difference quotient with

$$\lim_{h \rightarrow 0} \frac{u(\mathbf{p} + t(\mathbf{q} - \mathbf{p}) + h(q_n - p_n)\mathbf{e}_n) - u(\mathbf{p} + t(\mathbf{q} - \mathbf{p}))}{h} = \frac{\partial u}{\partial n}(\mathbf{p} + t(\mathbf{q} - \mathbf{p})) (q_n - p_n).$$

The second to the last quotient moves the second to the last coordinate:

$$Q_{n-1} = \frac{1}{h} [u(\mathbf{p} + t(\mathbf{q} - \mathbf{p}) + h(q_{n-1} - p_{n-1})\mathbf{e}_{n-1} + h(q_n - p_n)\mathbf{e}_n) - u(\mathbf{p} + t(\mathbf{q} - \mathbf{p}) + h(q_n - p_n)\mathbf{e}_n)]. \quad (1.18)$$

Technically, the last two coordinates of the “moving” value $u(\mathbf{p} + t(\mathbf{q} - \mathbf{p}) + h(q_{n-1} - p_{n-1})\mathbf{e}_{n-1} + h(q_n - p_n)\mathbf{e}_n)$ are both moving and the last coordinate of the “fixed” value $u(\mathbf{p} + t(\mathbf{q} - \mathbf{p}) + h(q_n - p_n)\mathbf{e}_n)$ is moving as well. This is why we need to use the mean value theorem for quotients like this. Here are the details² for this second to last term: We set

$$g(\xi) = u(\mathbf{p} + t(\mathbf{q} - \mathbf{p}) + \xi(q_{n-1} - p_{n-1})\mathbf{e}_{n-1} + h(q_n - p_n)\mathbf{e}_n).$$

The expression in (1.18) is

$$\frac{g(h) - g(0)}{h}$$

²Exercise 1.22.

and

$$\begin{aligned} g'(\xi) &= \lim_{k \rightarrow 0} \frac{1}{k} [u(\mathbf{p} + t(\mathbf{q} - \mathbf{p}) + (\xi + k)(q_{n-1} - p_{n-1})\mathbf{e}_{n-1} + h(q_n - p_n)\mathbf{e}_n) \\ &\quad - u(\mathbf{p} + t(\mathbf{q} - \mathbf{p}) + \xi(q_{n-1} - p_{n-1})\mathbf{e}_{n-1} + h(q_n - p_n)\mathbf{e}_n)] \\ &= \frac{\partial u}{\partial x_{n-1}}(\mathbf{p} + t(\mathbf{q} - \mathbf{p}) + \xi(q_{n-1} - p_{n-1})\mathbf{e}_{n-1} + h(q_n - p_n)\mathbf{e}_n) (q_{n-1} - p_{n-1}). \end{aligned}$$

The mean value theorem tells us we can find ξ between 0 and h such that

$$\frac{g(h) - g(0)}{h} = g'(\xi).$$

That is, we can replace the quotient (1.18) with

$$Q_{n-1} = \frac{\partial u}{\partial x_{n-1}}(\mathbf{p} + t(\mathbf{q} - \mathbf{p}) + \xi(q_{n-1} - p_{n-1})\mathbf{e}_{n-1} + h(q_n - p_n)\mathbf{e}_n) (q_{n-1} - p_{n-1}),$$

and by continuity (because $u \in C^1$)

$$\lim_{h \rightarrow 0} Q_{n-1} = \frac{\partial u}{\partial x_{n-1}}(\mathbf{p} + t(\mathbf{q} - \mathbf{p})) (q_{n-1} - p_{n-1}).$$

In general,

$$\begin{aligned} Q_j &= \frac{1}{h} \left[u \left(\mathbf{p} + t(\mathbf{q} - \mathbf{p}) + \sum_{k=j}^n h(q_k - p_k)\mathbf{e}_k \right) \right. \\ &\quad \left. - u \left(\mathbf{p} + t(\mathbf{q} - \mathbf{p}) + \sum_{k=j+1}^n h(q_k - p_k)\mathbf{e}_k \right) \right] \\ &= \frac{g_j(h) - g_j(0)}{h} \end{aligned}$$

where g_j is defined on some open interval containing the closed interval with endpoints³ 0 and h by

$$g_j(\xi) = u \left(\mathbf{p} + t(\mathbf{q} - \mathbf{p}) + \xi(q_j - p_j)\mathbf{e}_j + \sum_{k=j+1}^n h(q_k - p_k)\mathbf{e}_k \right).$$

³This somewhat complicated wording is the result of the fact that I don't know if $h < 0$ or $h > 0$.

And finally by the mean value theorem

$$Q_j = g'(\xi) = \frac{\partial u}{\partial x_j} \left(\mathbf{p} + t(\mathbf{q} - \mathbf{p}) + \xi(q_j - p_j)\mathbf{e}_j + \sum_{k=j+1}^n h(q_k - p_k)\mathbf{e}_k \right) (q_j - p_j)$$

for some ξ between 0 and h , so that by continuity

$$\lim_{h \rightarrow 0} Q_j = \frac{\partial u}{\partial x_j}(\mathbf{p} + t(\mathbf{q} - \mathbf{p})) (q_j - p_j).$$

We conclude

$$\begin{aligned} f'(t) &= \lim_{h \rightarrow 0} \frac{f(t+h) - f(t)}{h} \\ &= \lim_{h \rightarrow 0} \sum_{j=1}^n Q_j \\ &= \sum_{j=1}^n \frac{\partial u}{\partial x_j}(\mathbf{p} + t(\mathbf{q} - \mathbf{p})) (q_j - p_j). \end{aligned}$$

This was the essential assertion of the lemma. From this expression, we can see f' is continuous because the partial derivatives of u are continuous. \square

Proof of Theorem 2: Recall that we are wanting to show $C^1(U) \subset C^0(U)$, that is if each of the partial derivatives of a function $u : U \rightarrow \mathbb{R}$ is continuous on the open set $U \subset \mathbb{R}^n$, then the function u itself is continuous (at each point of U).

Letting $\mathbf{p} \in U$, we can take an open cube $C_\delta(\mathbf{p})$ with

$$\overline{Q_\delta(\mathbf{p})} \subset U. \tag{1.19}$$

This is standard notation for what should be said here, but I realize you may not know or understand this notation, so let me explain it (at some length). You may recall the notion of an open set. The set U , the cube $Q_\delta(\mathbf{p})$, and the ball $B_r(\mathbf{p})$ are all open sets: Around each point in such a set, there is some “room” in the sense that there is a ball or a cube centered at the point and lying entirely within the open set. Such a set does not contain “edge” points or what we call mathematically **boundary points**. Precisely, a point \mathbf{x} is a **boundary point** of the set A if every open ball $B_r(\mathbf{x})$ contains points in both A and the complement A^c of A . The collection of all boundary points of A is called the **boundary of A** and is denoted by ∂A . An open set contains within itself no boundary points.

Exercise 1.23 Show that a set $U \subset \mathbb{R}^n$ is open if and only if

$$U \cap (\partial U) = \phi.$$

The extreme in the other direction is called a **closed set**:

Exercise 1.24 Show the following are equivalent for a set $A \subset \mathbb{R}^n$:

- (i) $\partial A \subset A$.
- (ii) A^c is open.

A set A satisfying the conditions of Exercise 1.24 is said to be **closed**. A closed set A may contain a point \mathbf{x} for which there is a ball with $B_r(\mathbf{x}) \subset A$, but usually there are other points in ∂A as well. If $\partial A = \phi$ is empty, then the set A is both open and closed. The set \mathbb{R}^n is both open and closed. Any point \mathbf{x} in a set A having the property that there is some ball $B_r(\mathbf{x})$ with $B_r(\mathbf{x}) \subset A$ is called an **interior point**. Remember that open sets U are characterized by the condition that every point $\mathbf{x} \in U$ is an interior point. In general the set of all interior points of a set A is denoted by $\text{int}(A)$ (or by A° in some contexts).

I think we are ready for the explanation of the notation in (1.19). Given any set $A \subset \mathbb{R}^n$, there is a unique smallest closed set containing A . That set is called the **closure** of A and is denoted by \bar{A} .

Exercise 1.25 Show the closure of a set A is always given by

$$\bar{A} = A \cup (\partial A).$$

Hint: Show that $A \cup (\partial A)$ is closed and any closed set F with $A \subset F$ satisfies $F \subset A \cup (\partial A)$. (This makes $A \cup (\partial A)$ the smallest.)

Show in particular,

$$\partial Q_\delta(\mathbf{p}) = \cup_{j=1}^n \left\{ \mathbf{x} = \sum_{k=1}^n x_k \mathbf{e}_k \in \mathbb{R}^n : |x_j - p_j| = \frac{\delta}{2} \text{ and } |x_k - p_k| \leq \frac{\delta}{2} \text{ for all } k. \right\},$$

so that the closure of the cube is the closed cube

$$\overline{Q_\delta(\mathbf{p})} = \{ \mathbf{x} : |x_j - p_j| \leq \delta/2 \text{ for } j = 1, 2, \dots, n \}.$$

The case we have here where $\overline{Q_\delta(\mathbf{p})}$ is a subset of an open set U as in (1.19) is special because $\overline{Q_\delta(\mathbf{p})}$ is **bounded**, that is to say the entire set $\overline{Q_\delta(\mathbf{p})}$ will fit inside some ball $B_R(\mathbf{0})$. A set that is not bounded in this way is called... (you guessed it!) unbounded. Why is this situation special? One good reason (and the main reason of interest right at the moment) is because of the following important result:

Theorem 4 (extreme value theorem) If $f : A \rightarrow \mathbb{R}$ is a continuous function defined on a set $A \subset \mathbb{R}^n$ and the set A is closed and bounded, then there are values m and M and corresponding points $\mathbf{x}_m \in A$ and $\mathbf{x}_M \in A$ such that

$$m = f(\mathbf{x}_m) \leq f(\mathbf{x}) \leq f(\mathbf{x}_M) = M \quad \text{for all } \mathbf{x} \in A.$$

The values m and M are called the extreme values of the function f on A ; the value m is called the minimum, and the value M is called the maximum.

Returning to the proof of Theorem 2, we know the partial derivatives

$$\frac{\partial u}{\partial x_j} \in C^0(\overline{Q_\delta(\mathbf{p})}) \quad \text{for } j = 1, 2, \dots, n.$$

Since the absolute value of a continuous function is also continuous, we know that for each $j = 1, 2, \dots, n$

$$\left| \frac{\partial u}{\partial x_j} \right|$$

takes a nonnegative maximum value on $\overline{Q_\delta(\mathbf{p})}$. It follows that there is some $M > 0$ with

$$\sum_{j=1}^n \left| \frac{\partial u}{\partial x_j}((1-t)\mathbf{p} + t\mathbf{q}) \right| \leq M \quad (1.20)$$

for all $\mathbf{q} \in \overline{Q_\delta(\mathbf{p})}$ and t with $0 \leq t \leq 1$.

Recall that we want to estimate $|u(\mathbf{q}) - u(\mathbf{p})|$ and show this is small when \mathbf{q} is close to \mathbf{p} . In view of Lemma 1 we can write

$$\begin{aligned} |u(\mathbf{q}) - u(\mathbf{p})| &= \left| u((1-t)\mathbf{p} + t\mathbf{q}) \Big|_{t=0}^1 \right| \\ &= \left| \int_0^1 \frac{d}{dt} u((1-t)\mathbf{p} + t\mathbf{q}) dt \right|. \end{aligned}$$

Because

$$\frac{d}{dt}u((1-t)\mathbf{p} + t\mathbf{q}) = \sum_{j=1}^n (q_j - p_j) \frac{\partial u}{\partial x_j}((1-t)\mathbf{p} + t\mathbf{q})$$

we can estimate as follows:

$$\begin{aligned} |u(\mathbf{q}) - u(\mathbf{p})| &\leq \int_0^1 \left| \frac{d}{dt}u((1-t)\mathbf{p} + t\mathbf{q}) \right| dt \\ &= \int_0^1 \left| \sum_{j=1}^n (q_j - p_j) \frac{\partial u}{\partial x_j}((1-t)\mathbf{p} + t\mathbf{q}) \right| dt \\ &\leq \int_0^1 \sum_{j=1}^n |q_j - p_j| \left| \frac{\partial u}{\partial x_j}((1-t)\mathbf{p} + t\mathbf{q}) \right| dt \\ &\leq M \max_j |q_j - p_j| \end{aligned}$$

where $M > 0$ is the (upper) bound for

$$\sum_{j=1}^n \left| \frac{\partial u}{\partial x_j}((1-t)\mathbf{p} + t\mathbf{q}) \right|$$

obtained from the extreme value theorem. Now, for any tolerance $\epsilon > 0$ we can be sure $|u(\mathbf{q}) - u(\mathbf{p})| < \epsilon$ as long as we take $\mathbf{q} \in Q_\delta(\mathbf{p})$ and so that

$$\max_j |q_j - p_j| < \frac{\epsilon}{M + 1}.$$

This means u is continuous. \square

We have established (at least) a couple things:

1. Partial differentiability does not imply continuity when $n > 1$. (Though differentiability is the same as partial differentiability when $n = 1$, and differentiability does imply continuity when $n = 1$.)
2. Continuity of the first partial derivatives, i.e., $u \in C^1(U)$, does imply continuity for all n .

There is also something interesting in Lemma 1. This result is a special case of one form of what is known as the **multidimensional chain rule**. You may remember

the chain rule from calculus, at least in one dimension, as it applies to compositions: If $a, b, c, d \in \mathbb{R}$ with $a < b$ and $c < d$ and $f \in C^1(a, b)$ with $f(x) \in (c, d)$ for all x and $g \in C^1(c, d)$, then the composition $g \circ f : (a, b) \rightarrow \mathbb{R}$ by $g \circ f(x) = g(f(x))$ satisfies $g \circ f \in C^1(a, b)$ with

$$(g \circ f)'(x) = g'(f(x)) f'(x).$$

Roughly speaking, the multidimensional chain rule⁴ says that if a function $u \in C^1(U)$ with $U \subset \mathbb{R}^n$ so that $u = u(x_1, x_2, \dots, x_n)$ depends on multiple variables, all of which may depend on other (multiple) variables

$$x_k = f_k(\xi_1, \xi_2, \dots, \xi_m) \quad \text{for} \quad k = 1, 2, \dots, n$$

(and the composition $u \circ \mathbf{f}$ makes sense where

$$\mathbf{f} = (f_1, f_2, \dots, f_n) : V \rightarrow \mathbb{R}^n,$$

V is open in \mathbb{R}^m and $f_1, f_2, \dots, f_m \in C^1(V)$ satisfy $\mathbf{f}(\xi) \in U$ for all $\xi = (\xi_1, \xi_2, \dots, \xi_m) \in V$, then $u \circ \mathbf{f} \in C^1(V)$ and each partial derivative is given by the inner product

$$\frac{\partial(u \circ \mathbf{f})}{\partial \xi_j}(\xi_1, \xi_2, \dots, \xi_m) = \sum_{k=1}^n \frac{\partial u}{\partial x_k}(\mathbf{f}(\xi)) \frac{\partial f_k}{\partial \xi_j}(\xi).$$

Lemma 1 is the special case when $\mathbf{f} : (0, 1) \rightarrow \mathbb{R}^n$ is an **affine function** given by $\mathbf{f} = \mathbf{f}(t) = \mathbf{p} + t(\mathbf{q} - \mathbf{p})$. Notice that in this case

$$\frac{\partial f_k}{\partial t} = \frac{df_k}{dt} = f'(t) = q_k - p_k.$$

Boas has two interesting sections on this topic in Chapter 4 (sections 5.1 Chain Rule and 5.7 More Chain Rule). This is also, more or less, the topic of sections 5.11 and 5.12 in Boas, and generally, understanding the chain rule (or how to differentiate to determine the partial derivatives (or rates of change) of quantities depending on multiple variables in complicated ways is rather important for the analysis of many physical systems. she also talks about power series in two variables, but I'll tell you soon about power series in any (finite) number of variables—no extra charge.

Let's finish up with basic partial differentiation (theory).

There is another condition on a function $u : U \rightarrow \mathbb{R}$ which is weaker than C^1 but which also implies the continuity of u . It is also fairly important. This is called (full or total) **differentiability**. Here is the definition:

⁴Or at least one multidimensional chain rule.

Definition 6 (differentiability for a function of several variables) Given an open set $U \subset \mathbb{R}^n$ and $u : U \rightarrow \mathbb{R}$, we say u is **differentiable** at $\mathbf{p} \in U$ if there exists a linear function $L : \mathbb{R}^n \rightarrow \mathbb{R}$ such that

$$u(\mathbf{x}) - u(\mathbf{p}) - L(\mathbf{x} - \mathbf{p}) = o(\|\mathbf{x} - \mathbf{p}\|)$$

as $\mathbf{x} \rightarrow \mathbf{p}$. The notation on the right here is read “little-o of $\|\mathbf{x} - \mathbf{p}\|$.” It means the **limit of the quotient** of $u(\mathbf{x}) - u(\mathbf{p}) - L(\mathbf{x} - \mathbf{p})$ and $\|\mathbf{x} - \mathbf{p}\|$ is zero as \mathbf{x} tends to \mathbf{p} , or more properly for any $\epsilon > 0$, there is some $\delta > 0$ such that

$$0 < \|\mathbf{x} - \mathbf{p}\| < \delta \quad \text{implies} \quad \left| \frac{u(\mathbf{x}) - u(\mathbf{p}) - L(\mathbf{x} - \mathbf{p})}{\|\mathbf{x} - \mathbf{p}\|} \right| < \epsilon.$$

Exercise 1.26 Show differentiability implies partial differentiability.

The linear function $L : \mathbb{R}^n \rightarrow \mathbb{R}$ featured in the definition of differentiability is called the **differential map** determined by u at \mathbf{p} and is denoted by

$$du_{\mathbf{p}} : \mathbb{R}^n \rightarrow \mathbb{R}.$$

This linear function is determined by a single vector (as all real valued linear functions): Recall that owing to the linearity of L we have for

$$\mathbf{x} = \sum_{j=1}^n x_j \mathbf{e}_j \in \mathbb{R}^n$$

$$L(\mathbf{x}) = L\left(\sum_{j=1}^n x_j \mathbf{e}_j\right) = \sum_{j=1}^n x_j L(\mathbf{e}_j) = \langle \mathbf{x}, Du(\mathbf{p}) \rangle$$

where the vector $Du(\mathbf{p}) = (du_{\mathbf{p}}(\mathbf{e}_1), du_{\mathbf{p}}(\mathbf{e}_2), \dots, du_{\mathbf{p}}(\mathbf{e}_n))$ with the real images of the standard unit basis vectors, i.e., the real numbers $du_{\mathbf{p}}(\mathbf{e}_1), du_{\mathbf{p}}(\mathbf{e}_2), \dots, du_{\mathbf{p}}(\mathbf{e}_n)$, is called the **gradient** or **total derivative** of u .

Exercise 1.27 According to Exercise 1.26, a differentiable function $u : U \rightarrow \mathbb{R}$ is partially differentiable, meaning that the partial derivatives of u exist. Determine the gradient vector associated with the differential $du_{\mathbf{p}}$ in terms of the partial derivatives.

Theorem 5 Differentiability implies continuity.

Proof: The proof is not so different from the proof that differentiability implies continuity for $n = 1$. Which may be a little surprising since C^1 implies differentiability (see Theorem 6 below) and Theorem 2 was not so easy to prove. On the other hand, Theorem 6 is not so easy to prove. In any case, let $\epsilon > 0$. If $\mathbf{x} = \mathbf{p}$, then $|u(\mathbf{x}) - u(\mathbf{p})| = 0 < \epsilon$. If, on the other hand, $\|\mathbf{x} - \mathbf{p}\| > 0$, then we can write

$$\begin{aligned} |u(\mathbf{x}) - u(\mathbf{p})| &= \|\mathbf{x} - \mathbf{p}\| \left| \frac{u(\mathbf{x}) - u(\mathbf{p})}{\|\mathbf{x} - \mathbf{p}\|} \right| \\ &= \|\mathbf{x} - \mathbf{p}\| \left| \frac{u(\mathbf{x}) - u(\mathbf{p}) - L(\mathbf{x} - \mathbf{p})}{\|\mathbf{x} - \mathbf{p}\|} + \frac{L(\mathbf{x} - \mathbf{p})}{\|\mathbf{x} - \mathbf{p}\|} \right| \\ &\leq \|\mathbf{x} - \mathbf{p}\| \left| \frac{u(\mathbf{x}) - u(\mathbf{p}) - L(\mathbf{x} - \mathbf{p})}{\|\mathbf{x} - \mathbf{p}\|} \right| + |L(\mathbf{x} - \mathbf{p})|. \end{aligned}$$

The first term tends to zero by differentiability. More precisely, there is some $\delta_1 > 0$ for which $0 < \|\mathbf{x} - \mathbf{p}\| < \delta_1$ implies

$$\|\mathbf{x} - \mathbf{p}\| \left| \frac{u(\mathbf{x}) - u(\mathbf{p}) - L(\mathbf{x} - \mathbf{p})}{\|\mathbf{x} - \mathbf{p}\|} \right| < \delta_1 \frac{\epsilon}{2}.$$

Taking, without loss of generality $\delta_1 < 1$, we have

$$\|\mathbf{x} - \mathbf{p}\| \left| \frac{u(\mathbf{x}) - u(\mathbf{p}) - L(\mathbf{x} - \mathbf{p})}{\|\mathbf{x} - \mathbf{p}\|} \right| < \frac{\epsilon}{2}$$

for $0 < \|\mathbf{x} - \mathbf{p}\| < \delta_1$.

The second term $|L(\mathbf{x} - \mathbf{p})| = |du_{\mathbf{p}}(\mathbf{x} - \mathbf{p})|$ also tends to zero as \mathbf{x} tends to \mathbf{p} . There are various ways to see this. One good fact to know is that every linear function (on a finite dimensional vector space) is continuous. More explicitly, we can use a famous inequality, which may be thought of as a second inequality that can be very important for modeling physical systems—like the triangle inequality. This inequality is called the **Cauchy-Schwarz inequality** and it says that given two vectors \mathbf{v} and \mathbf{w} in \mathbb{R}^n the dot product satisfies

$$|\mathbf{v} \cdot \mathbf{w}| = |\langle \mathbf{v}, \mathbf{w} \rangle| \leq \|\mathbf{v}\| \|\mathbf{w}\|$$

where $\|\mathbf{v}\|$ and $\|\mathbf{w}\|$ are the Euclidean norms of \mathbf{v} and \mathbf{w} . The gradient vector $Du(\mathbf{p})$ for which $Lu(\mathbf{v}) = du_{\mathbf{p}}(\mathbf{p}) = \langle Du(\mathbf{p}), \mathbf{v} \rangle$, as you should have noted when you did Exercise 1.27 above, has a pretty simple form in terms of the partial derivatives of u at \mathbf{p} . In particular, $Du(\mathbf{p})$ is a fixed vector independent of \mathbf{x} , so that

$$|L(\mathbf{x} - \mathbf{p})| = |du_{\mathbf{p}}(\mathbf{x} - \mathbf{p})| = |\langle Du(\mathbf{p}), \mathbf{x} - \mathbf{p} \rangle| \leq \|Du(\mathbf{p})\| \|\mathbf{x} - \mathbf{p}\|$$

by the Cauchy-Schwarz inequality. Thus, we can take

$$\|\mathbf{x} - \mathbf{p}\| < \delta_2 \leq \frac{\epsilon}{2(\|Du(\mathbf{p})\| + 1)}$$

and conclude that for $\|\mathbf{x} - \mathbf{p}\| < \min\{\delta_1, \delta_2\}$ there holds

$$\|u(\mathbf{x}) - u(\mathbf{p})\| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Thus, $u \in C^0(U)$. \square

Theorem 6 If $u \in C^1(U)$, then u is differentiable on U .

As mentioned, this theorem is a little more difficult to prove, but it is not too difficult using some kind of multidimensional chain rule like Lemma 1 which ultimately depends on the mean value theorem.

Exercise 1.28 Prove Theorem 6. Hint: Consider the vector containing the partial derivatives of $u \in C^1(U)$:

$$\left(\frac{\partial u}{\partial x_1}(\mathbf{p}), \frac{\partial u}{\partial x_2}(\mathbf{p}), \dots, \frac{\partial u}{\partial x_n}(\mathbf{p}) \right) = \sum_{j=1}^n \frac{\partial u}{\partial x_j}(\mathbf{p}) \mathbf{e}_j$$

which we can still call $Du(\mathbf{p})$, though perhaps it is not quite proper to call this a gradient vector in all circumstances.⁵ Define a linear function $L : \mathbb{R}^n \rightarrow \mathbb{R}$ by

$$L\mathbf{v} = Du(\mathbf{p}) \cdot \mathbf{v}.$$

Use this function in the definition, and prove the limiting statement using Lemma 1.

Exercise 1.29 Prove a multidimensional mean value theorem: If $u \in C^1(U)$ and the segment

$$\Gamma = \{(1-t)\mathbf{p} + t\mathbf{q} : 0 \leq t \leq 1\}$$

is a subset of U , then there is some \mathbf{x} along the segment Γ for which

$$u(\mathbf{q}) - u(\mathbf{p}) = Du(\mathbf{x}) \cdot (\mathbf{q} - \mathbf{p}).$$

⁵For example when the partial derivatives exist but the function u is not even continuous.

We have given a pretty careful discussion of derivatives (partial derivatives, differentiability, differential approximation, and so forth) including estimates and tolerances. In many instances, engineers use the differential approximation results above in a kind of informal manner without worrying really about how good the approximation they are using actually is. One such approximation formula is the following:

$$u(\mathbf{q}) \approx u(\mathbf{p}) + du_{\mathbf{p}}(\mathbf{q} - \mathbf{p})$$

which can also take the form(s)

$$u(\mathbf{q}) \approx u(\mathbf{p}) + Du(\mathbf{p}) \cdot (\mathbf{q} - \mathbf{p})$$

or

$$u(\mathbf{q}) \approx u(\mathbf{p}) + \langle Du(\mathbf{p}), \mathbf{q} - \mathbf{p} \rangle.$$

Notice there is no explicit indication of **how good** such an approximation may or may not be. If you can get uniform bounds on higher derivatives, then sometimes you can get some information about tolerances, but engineers (and even applied mathematicians) sometimes do not have too much tolerance for that. The important thing I would like for you to note is that the symbol “ \approx ” here is not an equality. Even if you know the term $du_{\mathbf{p}}(\mathbf{p} - \mathbf{q})$ completely, you still don’t know exactly how much you might be off. You only know something vague, like “If $\|\mathbf{q} - \mathbf{p}\|$ is small (?), then the approximation should (?) be good.

Editorial note: Generally for the flow of these notes as I’m composing them this semester, I would like to now proceed to the next section “ordinary differential equations and partial differential equations.” This is, however, a natural place to discuss multivariable power series which probably should be done at some point (and was promised above). I found a discussion of power series in some notes on “differentiation” from spring 2020 (at least that seemed to be the last edit, perhaps they were even older notes). Reading over that presentation, I don’t think I could have said it better myself—actually I did say it myself—but I don’t think I will produce substantially better notes. Consequently, I’ve cut the entire section 6 from those notes on “partial differentiation” and included it below as Appendix 1. I suggest you read it as a review of the material above paying especially close attention to the treatment of multi-index notation (which is not above) and the discussion of multivariable power series. I will lecture on both topics in class (using Appendix 1 below as reference). This material will also be addressed in Assignment 3 as it was in Assignment 2 in the spring semesters of 2020 and 2021.

Before I proceed to Section 1.2, I offer something of a suggestive aside:

Affirmative action for the absolute value function

You may recall that $g : \mathbb{R} \rightarrow [0, \infty)$ by $g(x) = |x|$ was considered above and was put forward as an example of a function which is not differentiable. There is, however, a good theory of derivatives for functions like this. It is called the theory of **weak derivatives**. Leaving out some details, a function $v : \mathbb{R} \rightarrow \mathbb{R}$ is the **weak derivative** of a function $u : \mathbb{R} \rightarrow \mathbb{R}$ if

$$-\int u\phi' = \int v\phi$$

where ϕ is a function which is differentiable in the usual (classical) sense involving difference quotients and satisfies $\phi \in C^1(\mathbb{R})$ but for which there is some $R > 0$ for which $\phi(x) \equiv 0$ for $|x| \geq R$. Those integrals probably look a little strange to you. Once we know the value of R we can, at least in some circumstances, write them in the more familiar form

$$-\int_{-R}^R u(x)\phi'(x) dx = \int_{-R}^R v(x)\phi(x) dx. \quad (1.21)$$

Exercise 1.30 Guess what is the weak derivative of the absolute value function and check that it satisfies the condition (1.21) defining weak derivatives.

1.2 Ordinary differential equations and PDE

There is supposed to be a review of ordinary differential equations in MATH 6701 (Math Methods of Applied Science I and a prerequisite for this course). Let's start with a review of that review—which may not quite be a review at all because perhaps you have not considered ODE from the point(s) of view I will now adopt/present.

The main object of study in ODE is the **first order system** in the form

$$\mathbf{x}' = \mathbf{F}(\mathbf{x}, t) \quad (1.22)$$

and in particular the **initial value problem**

$$\begin{cases} \mathbf{x}' = \mathbf{F}(\mathbf{x}, t), & a < t < b \\ \mathbf{x}(t_0) = \mathbf{x}_0 \end{cases} \quad (1.23)$$

for some $t_0 \in (a, b)$. Two main themes should be kept in mind:

1. Any ordinary differential equation (in particular any ODE of any order) is essentially equivalent to a first order system of the form (1.22).

2. There is a sweeping theory of existence and uniqueness for the initial value problem (IVP) which essentially says that there exists a unique solution for the problem (1.23) in some open interval containing t_0 under extremely mild structural assumptions on the equation and that existence holds under even milder assumptions.

Let us briefly review some details of these two points. One might start by saying an ODE of **order** n is an equation of the form

$$f^{(n)} = G(f^{(n-1)}, f^{(n-2)}, \dots, f', f, t) \quad (1.24)$$

where

$$f^{(j)} = \frac{d^j f}{dt^j} \quad (1.25)$$

represents the j -th order derivative of a function $f : (a, b) \rightarrow \mathbb{R}$ on some interval (a, b) of the real line. One can also use the Lagrange notation⁶ f', f'', f''' for the first few derivatives or, alternatively, Newton's notation $\dot{f}, \ddot{f}, \dddot{f}$ which is often used to suggest the use of a (model) time variable these days. Though it was not mentioned in section 1.1 above, it should almost go without saying that a function $f : (a, b) \rightarrow \mathbb{R}$ which is differentiable on (a, b) determines a function $f' : (a, b) \rightarrow \mathbb{R}$, that function f' may have a derivative itself $f'' = f^{(2)}$ and this process may be repeated as often as regularity allows. There is much more to say about this, and I should probably say at least some of it at some point, but for now let me simply mention the (essentially obvious) observation that a similar situation persists with regard to partial derivatives of functions of several variables, and we have already encountered the use of such repeated differentiation in the context of power series expansions.

Returning to the ODE (1.24), the function G is assumed to be a function of n variables usually defined on some open subset of \mathbb{R}^n but, in practice and to avoid complications, very often is assumed to be defined on all of \mathbb{R}^n . One might imagine the following slightly more general form for an n -th order ODE:

$$F(f^{(n)}, f^{(n-1)}, f^{(n-2)}, \dots, f', f, t) = 0 \quad (1.26)$$

where $F : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$. This leaves open the possibility, however, that F depends on its first argument in some “singular” manner, for example **not at all**, which would

⁶This notation was paradoxically first introduced by Euler, but traditionally “Euler notation” for derivatives (which paradoxically was introduced by Arbogast, was probably never used by Euler, and will be used in this course and these notes for something else) is something different yet: $D^j f$. What can you do? At least Leibniz' notation using a quotient of differentials as in (1.25) was introduced and used by Leibniz, and Newton's notation was introduced and used by Newton.

mean you really somehow do not have a proper n -th order ODE but maybe an ODE of some lower order. The usual **nonsingularity condition** for such an equation is

$$\frac{\partial F}{\partial p_n} \neq 0 \quad (1.27)$$

where we think of F as a function of the $n + 2$ variables $p_n, p_{n-1}, \dots, p_1, p_0, t$ so

$$F = F(p_n, p_{n-1}, \dots, p_1, p_0, t).$$

Exercise 1.31 Given a nonsingular n -th order ODE in the form (1.24) above, write the equation in the form (1.26) and compute

$$\frac{\partial F}{\partial p_n}.$$

The set on which the nonsingularity condition (1.27) should hold is, like the domain of F , largely dependent on circumstances, but most safely assumed to be holding most everywhere, e.g., on \mathbb{R}^{n+2} . When (1.27) does hold, there is a technical theorem called the **implicit function theorem** which asserts that the ODE (1.26) really does depend on $f^{(n)}$ and, not only that but, (1.26) can be “solved” for $f^{(n)}$ and written in the form (1.24).

This is not at all to say the nonsingularity condition (1.27) always holds. A good example where this condition most definitively does not hold is for **Bessel’s ODE**

$$x^2 y'' + xy' + (x^2 - \alpha^2)y = 0. \quad (1.28)$$

Bessel’s equation (1.28) is manifestly **singular** at the value $x = 0$ of the independent variable. It is an important ODE, the singularity is one that should be studied carefully and understood, and this is an ODE we should even talk about in this course before the semester is over. In the big scheme of things with regard to the subject of ODEs however, we can think of such singular points as details which can be (and will be for the moment) swept under the rug.

Theorem 7 The nonsingular n -th order ODE (1.24) is equivalent to a system of n first order equations as introduced in (1.22) in the sense that there is an n -th order system $\mathbf{x}' = \mathbf{F}(\mathbf{x}, t)$ determined by (1.24) and the solutions $\mathbf{x} \in C^1$ of this system are in one to one correspondence with the solutions $f \in C^n$ of (1.24).

On the one hand, perhaps we are a little ahead of ourselves. The question should be asked: What is a solution of an ODE? Here are some answers:

- (i) A solution of $\mathbf{x}' = \mathbf{F}(\mathbf{x}, t)$ is a differentiable function $\mathbf{u} : (a, b) \rightarrow \mathbb{R}^n$ defined on some open interval (a, b) of the real line for which the equation holds, i.e., when you substitute $\mathbf{u}(t)$ for \mathbf{x} on the right in (1.22) you get $\mathbf{u}'(t)$ for every $t \in (a, b)$.
- (ii) A solution of $f^{(n)} = G(f^{(n-1)}, f^{(n-2)}, \dots, f', t)$ is an n times differentiable function $u : (a, b) \rightarrow \mathbb{R}$ defined on some open interval (a, b) for which the equation holds, i.e., $u^{(n)}(t) = G(u^{(n-1)}(t), \dots, u'(t), u(t), t)$ for every $t \in (a, b)$.

This question and answer are almost not worth mentioning, but it is probably worth emphasizing that whenever one considers a differential equation, and an ODE in particular, one is implicitly **asking a question**, and it is a question you should (always) keep in mind. That question is:

Can you find a function satisfying this equation? If there is more than one, can you find all of them?

There are also some new kinds of players involved in the system of ODEs (1.22) and (i) which are worth it for me to mention and for you to contemplate. First of all the solution itself $\mathbf{x} : (a, b) \rightarrow \mathbb{R}^n$ is a **vector valued function of one real variable** which is a basic object of study in ODEs, but has not been mentioned too much in this course so far. Such a function consists of n real valued **coordinate functions** so that $\mathbf{x} = (x_1, x_2, \dots, x_n)$ where for each $j = 1, 2, \dots, n$, the function $x_j : (a, b) \rightarrow \mathbb{R}$ is differentiable. If we want to make things simple, we can imagine $x_j \in C^1(a, b)$ for $j = 1, 2, \dots, n$. Geometrically, one does not usually talk about the “graph” of such a function,⁷ but the **image**

$$\{(x_1(t), x_2(t), \dots, x_n(t)) \in \mathbb{R}^n : t \in (a, b)\}$$

is usually the geometric object of interest. As t moves long the domain interval (a, b) , the point $\mathbf{x}(t)$ traces out some curve in \mathbb{R}^n . The derivative

$$\mathbf{x}' = (x'_1, x'_2, \dots, x'_n),$$

which is a vector consisting of the derivatives of the coordinate functions, may be visualized as a tangent or velocity vector along the image curve at $\mathbf{x}(t)$.

⁷If you happen to be interested, the graph of a vector valued function of one variable $\mathbf{x} : (a, b) \rightarrow \mathbb{R}^n$ is $\{(t, x_1(t), x_2(t), \dots, x_n(t)) \in \mathbb{R}^{n+1} : t \in (a, b)\}$. When $n = 2$, this set can be visualized in \mathbb{R}^3 . You can decide if the graph in this case gives more information than the image; take for example $\mathbf{x}(t) = (\cos t, \sin t)$.

The other entity worth mentioning is the function $\mathbf{F} : \mathbb{R}^{n+1} \rightarrow \mathbb{R}^n$ which is a **vector valued function of a vector variable** and determines the structure of the ODE. For now simply note that $\mathbf{F} = (F_1, F_2, \dots, F_n)$ where each function F_j is a function of $n + 1$ variables $p_{n-1}, p_{n-2}, \dots, p_0, t$.

Here are two theorems about ODEs to know/recall:

Theorem 8 (existence and uniqueness) If $\mathbf{F} \in C^1(\mathcal{N} \rightarrow \mathbb{R}^n)$ where \mathcal{N} is an open subset of \mathbb{R}^{n+1} , by which we mean $F_j \in C^1(\mathcal{N})$ for $j = 1, 2, \dots, n$, and

$$(p_1, p_2, \dots, p_n, t_0) \in \mathcal{N},$$

then there exists some $\delta > 0$ such that the initial value problem

$$\begin{cases} \mathbf{x}' = \mathbf{F}(\mathbf{x}, t), & |t - t_0| < \delta \\ \mathbf{x}(t_0) = \mathbf{p} \end{cases}$$

has a unique solution. The structural assumption on \mathbf{F} in this theorem can be weakened⁸ to the assumption that \mathbf{F} is **Lipschitz continuous** on \mathcal{N} , that is, there is some constant $\lambda > 0$ such that

$$\|\mathbf{F}(\mathbf{x}, t) - \mathbf{F}(\xi, \tau)\| \leq \lambda \|(\mathbf{x}, t) - (\xi, \tau)\| \quad \text{for } (\mathbf{x}, t), (\xi, \tau) \in \mathcal{N}.$$

Existence, but not necessarily uniqueness, will follow if it is only assumed that $\mathbf{F} \in C^0(\mathcal{N} \rightarrow \mathbb{R}^n)$.

Here is the example you should know concerning the relation between the size of \mathcal{N} and the size of δ .

Exercise 1.32 Explain how Theorem 8 applies to the ODE $y' = y^2$. Choose \mathcal{N} as large as possible. Solve the ODE to determine δ .

Here is the/an example you should know concerning uniqueness.

Exercise 1.33 Explain how Theorem 8 applies to the ODE $y' = \sqrt{|y|}$. Solve the ODE to show how and where uniqueness fails.

There is a second theorem from ODEs everyone should know. It applies to a class of ODEs with a special structure. A system of first order ODEs is said to be **linear** if there is an $n \times n$ matrix $A = (a_{ij})$ with each entry $a_{ij} \in C^0(a, b)$ and a vector valued function $\mathbf{b} \in C^0((a, b) \rightarrow \mathbb{R}^n)$ such that the system can be written as

$$\mathbf{x}' = A\mathbf{x} + \mathbf{b} = A(t)\mathbf{x} + \mathbf{b}(t).$$

⁸Actually, only continuity is required in t and Lipschitz continuity in \mathbf{x} ; look up the Picard-Lindelöf theorem for details. In any case, the $C^1(\mathcal{N})$ version is adequate in many instances.

Theorem 9 (existence and uniqueness for linear ODEs) Given $t_0 \in (a, b)$ and any $\mathbf{p} \in \mathbb{R}^n$, the initial value problem

$$\begin{cases} \mathbf{x}' = A\mathbf{x} + \mathbf{b}, & t \in (a, b) \\ \mathbf{x}(t_0) = \mathbf{p} \end{cases}$$

has a unique solution on all of (a, b) .

The big difference here is that you get, so called, “long time” existence rather than unspecified local existence for $|t - t_0| < \delta$. For example, if the domain interval is $(a, b) = \mathbb{R}$, then you get existence for all $t \in \mathbb{R}$.

The other big thing

To summarize our bi-review of ODE:

1. The subject of “ordinary differential equations” **is** the study of **first order systems** of ordinary differential equations (though there are a lot of other topics in the subject, e.g., linear second order equations and Laplace transforms, that may obscure this assertion.
2. The existence and uniqueness theorems (a local one for nonlinear ODE and a global one for linear ODE) should be kept in mind.

And there is one more thing you should realize when thinking about ODE which I did not mention:

For **most** ODEs, you can find out **most** of what you want to know by using the ODE (adaptive fourth and fifth order Runge-Kutta) **numerical “solvers”** in standard mathematical software, e.g., Mathematica, Maple, Matlab, Octave, etc..

In this regard, I note the following:

- (i) You often need to understand the existence and uniqueness theorems both to interpret the output of a numerical solver as well as to configure/use them correctly and/or to your advantage.
- (ii) Probably the most useful/powerful thing you can pick up in a course on ODEs is familiarity with and competence in using one of these solvers, e.g., `NDSolve` in Mathematica or `ODE45` in Matlab.
- (iii) You can write your own Runge-Kutta solver as well, and this is probably a good exercise, but it will be hard to beat the “canned” solvers.

An exercise in geometric ODEs

Here is an exercise that involves most of the ODE topics mentioned above and constitutes a pretty good review of most of what you should have learned in ODEs.

If a curve in the plane is given by the graph of a function $f \in C^2(a, b)$ where $a, b \in \mathbb{R}$ with $a < b$ as usual, then the **signed curvature of the graph of f** is given by

$$k = \frac{f''}{(1 + f'^2)^{3/2}}.$$

Exercise 1.34 Consider the functions $f_j \in C^2(-r, r)$ for $j = 1, 2$ where $r > 0$ and

$$f_j(x) = (-1)^j \sqrt{r^2 - x^2}.$$

- (a) Plot/draw the graph of f_j for $j = 1, 2$.
- (b) Compute the signed curvature of the graph of f_j for $j = 1, 2$.

Exercise 1.35 Let $a, b \in \mathbb{R}$ with $a < b$ and consider a function $f \in C^2(a, b)$. Given a point $x_0 \in (a, b)$,

- (a) Find the formula for the tangent line to $\mathcal{G} = \{(x, f(x)) : x \in (a, b)\}$ at the point $(x_0, f(x_0))$ and plot/draw a picture illustrating this tangent line.
- (b) Assuming $f'(x_0) = 0$ and $f''(x_0) \neq 0$, find the equation of the **osculating circle** to the graph of f at $(x_0, f(x_0))$. This circle should pass through $(x_0, f(x_0))$ and be given locally near this point by the graph of a function g satisfying
 - (i) $g \in C^2(x_0 - \epsilon, x_0 + \epsilon)$ for some $\epsilon > 0$,
 - (ii) $g'(x_0) = 0$, and
 - (iii) $g''(x_0) = f''(x_0)$.
- (c) What is the maximum possible value of the tolerance ϵ in part (b)(i) above?

Exercise 1.36 What is the relation between the radius of the osculating circle you found in Exercise 1.35 above and the signed curvature of the graph of f at $(x_0, f(x_0))$?

Now I'm going to focus this discussion on a certain specific function $f : (-a, a) \rightarrow \mathbb{R}$ on a symmetric interval having graph with signed curvature equal to the signed

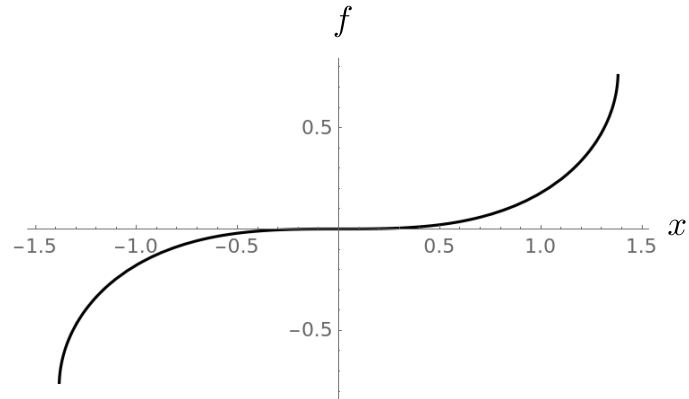


Figure 1.2: The graph of a function f whose signed curvature is equal to the signed arclength.

arclength along the graph. The graph of the function I have in mind is illustrated in Figure 1.2.

The **signed arclength** along the graph of f is defined by $s : (-a, a) \rightarrow \mathbb{R}$ by

$$s(x) = \int_0^x \sqrt{1 + [f'(\xi)]^2} d\xi. \quad (1.29)$$

This function is monotone increasing with

$$s'(x) = \sqrt{1 + [f'(x)]^2} \geq 1$$

and therefore is invertible with inverse we can denote by $\tau : (-L, L) \rightarrow \mathbb{R}$ where

$$L = \lim_{x \rightarrow a} s(x) \in [0, \infty].$$

In fact, for the graph we will construct $L < \infty$. Notice that there holds

$$\sigma = \int_0^{\tau(\sigma)} \sqrt{1 + [f'(\xi)]^2} d\xi,$$

so

$$\frac{d\tau}{d\sigma} = \frac{1}{\sqrt{1 + [f'(\xi)]^2}}.$$

Consequently, $\gamma : (-L, L) \rightarrow \mathbb{R}^2$ by $\gamma(\sigma) = (\tau(\sigma), f(\tau(\sigma)))$ is a parameterization of the graph of f by arclength. The speed of this parameterization is a unit vector, namely,

$$\frac{d\gamma}{d\sigma} = \frac{(1, f'(\tau))}{\sqrt{1 + [f'(\tau)]^2}}.$$

Most curves can be parameterized by arclength, though we have only carried out the details for a particular curve given as the graph of a function. It is also customary to use the symbol s both for the **arclength function** defined in (1.29) and for the **arclength parameter**, so that one writes

$$s = \int_0^\tau \sqrt{1 + [f'(\xi)]^2} d\xi \quad \text{and} \quad \frac{d\gamma}{ds} = \frac{(1, f')}{\sqrt{1 + [f']^2}}$$

where it is understood that the argument of f' in the latter formula is $\tau = \tau(s) = s^{-1}$. Furthermore, it is customary to denote derivatives with respect to arclength s , that is with respect to an arclength parameter, using Newton's notation for derivatives:

$$\dot{\gamma} = \frac{(1, f')}{\sqrt{1 + [f']^2}}.$$

Definition 7 (curvature vector) Given a curve parameterized by arclength $\gamma = \gamma(s) : (s_0 - \ell, s_0 + \ell) \rightarrow \mathbb{R}^n$ for some $\ell > 0$, meaning

$$\|\dot{\gamma}\| = 1,$$

the **curvature vector** of $\Gamma = \{\gamma(s) : s_0 - \ell < s < s_0 + \ell\}$ at the point $\gamma(s) \in \Gamma$ is defined to be

$$\vec{k} = \ddot{\gamma}(s).$$

The norm of the curvature vector $\|\vec{k}\|$ is called the (unsigned) curvature.

Exercise 1.37 Show the curvature vector is normal to the unit tangent vector $\dot{\gamma}$.

Definition 8 (signed curvature of a planar curve) Given a planar curve Γ parameterized by γ with respect to an arclength parameter s on an interval $(s_0 - \ell, s_0 + \ell)$, the **signed curvature** of Γ with respect to the parameterization (direction) γ is defined to be

$$\ddot{\gamma} \cdot \dot{\gamma}^\perp$$

where $\dot{\gamma}^\perp = (-\dot{\gamma}_2, \dot{\gamma}_1)$ when $\dot{\gamma} = (\dot{\gamma}_1, \dot{\gamma}_2)$.

Exercise 1.38 Show the definition of signed curvature we have given for a general plane curve agrees with the signed curvature we have defined for the graph of f above when the (forward) parameterization introduced above is used.

Exercise 1.39 Show that in general for a plane curve one has

$$\|\vec{k}\| = |k|.$$

That is, the unsigned curvature is the absolute value of the signed curvature.

If one thinks carefully about the assertion of Exercise 1.38, according to which

$$\ddot{\gamma} \cdot \dot{\gamma}^\perp = \frac{f''}{(1 + f'^2)^{3/2}}$$

where γ is the parameterization by arclength of the parameterized graph $\alpha(x) = (x, f(x))$ for $-a < x < a$, there is no reason to restrict attention to a graph. That is, it is very natural to consider the initial value problem

$$\begin{cases} \ddot{\gamma} \cdot \dot{\gamma}^\perp = s, & s \in \mathbb{R} \\ \gamma(0) = (0, 0) & \dot{\gamma}(0) = (1, 0) \end{cases} \quad (1.30)$$

for a curve parameterized by arclength. In this situation, one can do something very clever.

Exercise 1.40 What happens if you attempt to write down the ordinary differential equation in (1.30) as an equivalent system of first order ordinary differential equations? Hint: Writing $\ddot{\gamma} = (\ddot{\gamma}_1, \ddot{\gamma}_2)$, you should be able to write down a regular system for $-L_1 < s < L_1$ where $L = L_1$ is the positive arclength

$$L = \lim_{x \nearrow a} s(x)$$

discussed above. For $|s| \geq L_1$, this system (you have written down) is singular, but (1.30) as a whole is not singular.

The unpleasantness encountered in Exercise 1.40 may be avoided through the introduction of a certain “geometric” parameter having nice compatibility with the arclength parameter. This new parameter is **inclination angle**. Let Γ denote a curve parameterized by $\gamma : I \rightarrow \mathbb{R}^2$ with

$$\gamma \in C^2(I \rightarrow \mathbb{R}^2) \quad (1.31)$$

with respect to an arclength parameter where I is some (perhaps open) interval in \mathbb{R} and (1.31) means the coordinate functions in $\gamma = (\gamma_1, \gamma_2)$ satisfy $\gamma_j \in C^2(I)$ for $j = 1, 2$. Recall the assumption that we have an arclength parameterization means $\dot{\gamma} = \dot{\gamma}(s)$ is a unit (length) vector for all $s \in I$. Given an **initial inclination** $\dot{\gamma}(s_0)$ for some $s_0 \in I$, by which we mean an angle $\theta_0 \in \mathbb{R}$ for which

$$\dot{\gamma}(s_0) = \theta_0,$$

i.e.,

$$\cos \theta_0 = \dot{\gamma}_1(s_0) \quad \text{and} \quad \sin \theta_0 = \dot{\gamma}_2(s_0),$$

there exists a unique **inclination angle** $\psi \in C^1(I)$ satisfying

$$\begin{cases} \cos \psi & = \dot{\gamma}_1(s), & s \in I \\ \sin \psi & = \dot{\gamma}_2(s), & s \in I \\ \psi(s_0) & = \theta_0 \end{cases} \quad (1.32)$$

and the initial value problem

$$\begin{cases} \sin \psi \dot{\psi} & = -\ddot{\gamma}_1(s), & s \in I \\ \cos \psi \dot{\psi} & = \ddot{\gamma}_2(s), & s \in I \\ \psi(s_0) & = \theta_0. \end{cases} \quad (1.33)$$

I have used the existence of an inclination angle for a parameterized curve as asserted above in my research many times. It occurs to me now that I do not know a reference for this “fact.” It might be expected that (1.33) might be used as a kind of system of ordinary differential equations to determine ψ . It is, in fact, a system of ordinary differential equations for ψ , and what we are asserting is that this system has a unique solution. This approach is problematic for a couple reasons. Both differential equations in the system are potentially singular, though it will be noted that they cannot both be singular at the same time. Beyond that, taking one of the equations as locally non-singular, say $\sin \theta_0 \neq 0$, that equation can be solved for ψ locally. After that, it is not immediately obvious that the other (potentially singular) equation will be satisfied.

Everything asserted above concerning the existence of an inclination angle does turn out to be true. I’ll try to write up an explanation for why that is the case and make it available later. For now, let us assume that is the case. Most of the time when I use inclination angle it is, to a certain extent, in the “reverse direction” with regard to the differential equations, which is what I wanted to show you (and use as

a nice review of ODEs). Specifically, if we believe a certain curve should exist and we believe that curve should have a well-defined inclination angle, then we can attempt to construct the curve by starting with the ordinary differential equations

$$\begin{cases} \dot{\gamma}_1 = \cos \psi \\ \dot{\gamma}_2 = \sin \psi \end{cases}$$

for the coordinate functions $\gamma = (\gamma_1, \gamma_2)$ assumed to be parameterized by arclength. It (usually or at least often) only remains to find a differential equation satisfied by the inclination angle function $\psi = \psi(s)$ given as a function of arclength. This is often very convenient, as the derivative ψ has an extremely useful interpretation as the signed curvature of the curve.

Let us begin by noting that the signed curvature of the graph of f may be expressed as

$$k = \frac{d}{dx} \left(\frac{f'}{\sqrt{1+f'^2}} \right) = \frac{d}{dx} \sin \theta = \cos \theta \frac{d\theta}{dx}.$$

On the other hand, we can note that

$$\cos \theta = \frac{1}{\sqrt{1+f'^2}} \quad \text{and} \quad \frac{d\theta}{dx} = \dot{\theta} \frac{ds}{dx} = \dot{\theta} \sqrt{1+f'^2}.$$

In this way, we see the signed curvature satisfies

$$k = \dot{\theta}.$$

This final expression is not limited by the restriction that the curve is the graph of a function f , and in fact we may assume (up to some questions about orientation) that

$$k = \dot{\theta} = \dot{\gamma}^\perp \ddot{\gamma}.$$

Thus, we see that in practical terms the initial value problem (1.30) may be expressed as

$$\begin{cases} \dot{\gamma}_1 = \cos \theta, & \gamma_1(0) = 0 \\ \dot{\gamma}_2 = \sin \theta, & \gamma_2(0) = 0 \\ \dot{\theta} = s, & \theta(0) = 0. \end{cases}$$

The solution of this system is given explicitly in terms of indefinite integrals with

$$\gamma(s) = \left(\int_0^s \cos \left(\frac{\sigma^2}{2} \right) d\sigma, \int_0^s \sin \left(\frac{\sigma^2}{2} \right) d\sigma \right)$$

and these coordinate functions may also be expressed in terms of standard special functions called **Fresnel integrals**. Practically speaking, however, it is probably preferable to simply put the system directly into a numerical ODE solver/approximator. The resulting parametric curve is plotted in Figure 1.3.

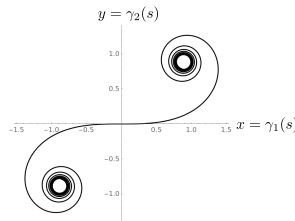


Figure 1.3: A curve whose signed curvature is equal to the signed arclength starting from the origin/horizontal point.

Notice that this curve coincides with the graph of f illustrated in Figure 1.2 on some interval $[-a, a]$. The actual value of a is difficult to discern from the non-parametric system because a corresponds to the singular value where $f'(a) = +\infty$. This singularity is not present in the parametric representation and corresponds simply to $\theta = \pi/2$. More precisely, we have $a = \gamma_1(s_1)$ where s_1 is the first positive arclength for which $\theta(s_1) = \pi/2$. Plotting $\theta = \theta(s)$ as on the right in Figure 1.4, we see

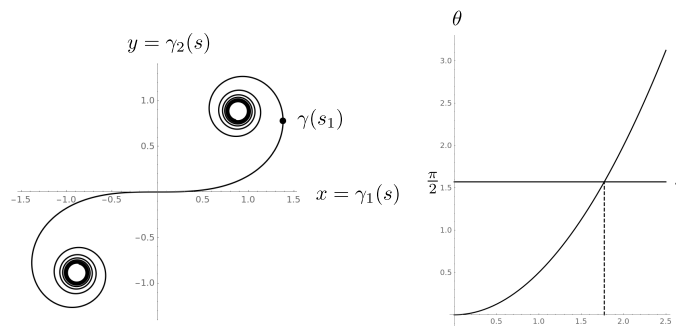


Figure 1.4: A curve whose signed curvature is equal to the signed arclength starting from the origin/horizontal point and this curve's first vertical point.

there is a well-defined intersection which is easily approximated numerically using Newton's method or with, for example, a more sophisticated root find algorithm

like Mathematica's `FindRoot`. Using such software, we find the $s_1 \doteq 1.77245$ and $a = \gamma_1(s_1) \doteq 1.38232$.

Partial Differential Equations

Assuming you know a thing or two about ordinary differential equations, there is no reason to avoid consideration of partial differential equations (PDE). The first thing to know about PDE is that there is no general theory of existence and uniqueness. There is no theorem like Theorem 8.

Actually, there is a PDE theorem called the Cauchy-Kowalevski theorem that looks superficially something like Theorem 8, and it can be a pretty useful theorem sometimes. At least as a mathematician studying partial differential equations, I've used the Cauchy-Kowalevski theorem a time or two, but I do not know of any really common application of the result in either applied math or engineering. Furthermore, the theorem is really not very much like Theorem 8 in that it requires everything involved with the structure of the PDE to be real analytic. And if you give up the real analytic structure, then **you can find**⁹ not so bad looking **PDE for which there is no solution** anywhere to be found. The upshot of this somewhat surprising situation is that it makes a lot more sense to just start with some very special classes of PDE and see what one can say about them.

In this course, we will mostly stick to four special classes of PDE, and it might be even more accurate to say we will consider one special class and three additional specific PDE. The PDE we will consider are

1. First order linear (and maybe quasilinear) PDE.
2. Laplace's PDE.
3. The heat equation (or Fourier's PDE).
4. The wave equation.

In defence of the three specific PDE, this trio is, on the one hand, known as the three **partial differential equations of classical physics**, which is a pretty grandiose title. On the other hand, the properties of these three are representative of three

⁹Actually, a mathematician named Hans Lewy found these examples of PDE in 1957. They were not so easy to find, and many other mathematicians were really surprised that they were out there, i.e., that this was the situation with existence (or more properly non-existence) of solutions for PDE.

special classes of linear second order PDE, namely the elliptic PDE (of which Laplace's equation is an example), parabolic PDE (of which the heat equation is an example), and (you guessed it) hyperbolic PDE of which the wave equation is an example.

Let's start. Here is a first order linear PDE:

$$x \frac{\partial u}{\partial x} + y \frac{\partial u}{\partial y} = 0. \quad (1.34)$$

It is linear because the (classical partial differential) operator $L : C^1(\mathbb{R}) \rightarrow C^0(\mathbb{R})$ by

$$Lu = x \frac{\partial u}{\partial x} + y \frac{\partial u}{\partial y}$$

is linear. These first order linear PDE (it has been discovered) are intimately related to special curves called **characteristic curves**. Here's the idea: Imagine you have a parameterized curve $\gamma : (a, b) \rightarrow \mathbb{R}^2$ given by $\gamma(t) = (\xi(t), \eta(t))$ for some functions $\xi, \eta \in C^1(a, b)$. Then if $u \in C^1(U)$ for some open set U containing the image $\{\gamma(t) : t \in (a, b)\}$, then it makes sense to consider (and differentiate) the composition

$$u \circ \gamma(t) = u(\xi(t), \eta(t)).$$

This leads to

$$\frac{d}{dt} u(\xi(t), \eta(t)) = \frac{\partial u}{\partial x} \xi' + \frac{\partial u}{\partial y} \eta' \quad (1.35)$$

by the chain rule. Comparing (1.35) to our PDE (1.34) strongly suggests the consideration of the following **characteristic equations**:

$$\xi' = \xi \quad \text{and} \quad \eta' = \eta.$$

This is, of course, a system of two first order ODEs. They happen to be completely decoupled¹⁰ so we can consider them independently as two single ODEs. In fact, we know the general solutions of these ODEs:

$$\xi(t) = c_1 e^t \quad \text{and} \quad \eta(t) = c_2 e^t$$

where c_1 and c_2 are some real numbers. This seems to be going pretty well, but what does it tell us about a solution for the original PDE (1.34)? The idea is the following: Given a points $(x_0, y_0), (x, y) \in U$ and a particular initial value $u(x_0, y_0)$ if we can find one of these characteristic curves γ **connecting** (x_0, y_0) and (x, y) with

¹⁰As you might guess, this decoupling doesn't always happen.

$\gamma(t_0) = (x_0, y_0)$ and $\gamma(t) = (x, y)$, then we can be sure $u(x, y) = u(x_0, y_0)$ at all points $(x, y) = \gamma(t)$ along the characteristic curve. The claim is that the characteristic curve must “transport” the value $u(x_0, y_0)$ to all the points along the curve.

More precisely, if $u \in C^1(U)$ is a solution of (1.34) on U and $\gamma(t) \in U$ for $t_0 \leq t \leq T$, then

$$\frac{d}{dt}u \circ \gamma(t) = \xi \frac{\partial u}{\partial x}(\xi, \eta) + \eta \frac{\partial u}{\partial y}(\xi, \eta) = 0.$$

In particular, $u \circ \gamma(t)$ is constant for $t_0 \leq t \leq T$, so if $(x, y) = \gamma(T) = (\xi(T), \eta(T))$, then

$$u(x, y) = u(x_0, y_0).$$

The questions then become:

1. Where do we get the starting points $(x_0, y_0) \in U$? What is the nature of the initial values?
2. Which points in U can be “reached,” i.e., connected to initial points, by a characteristic curve (in U)?

In two dimensions, the answer to the first question is what is called a **non-characteristic curve**. In higher dimensions the non-characteristic manifold will typically also be higher dimensional, e.g., a non-characteristic surface when $n = 3$. Let’s go ahead and deal with $n = 2$. Remember (or note) that for the characteristic curve of the first order quasilinear PDE

$$a_1 \frac{\partial u}{\partial x_1} + a_2 \frac{\partial u}{\partial x_2} = f$$

we wanted $\gamma = (\xi, \eta)$ to satisfy $\xi' = a_1$ and $\eta' = a_2$. What we want for a non-characteristic curve is that this doesn’t happen, and nothing remotely like this happens. Precisely, a curve $\nu : (\alpha, \beta) \rightarrow U$ is **non-characteristic** for the first order quasilinear PDE

$$a_1 \frac{\partial u}{\partial x_1} + a_2 \frac{\partial u}{\partial x_2} = f$$

if

$$(\alpha', \beta') \cdot (-a_2, a_1) = a_1\beta' - a_2\alpha' \neq 0. \tag{1.36}$$

Geometrically, this means the characteristic curve should be **transverse**, i.e., not tangent, to the non-characteristic curve. Let’s illustrate this point with a specific example. For the PDE (1.34) above, take U to be the punctured plane $U = \mathbb{R}^2 \setminus \{(0, 0)\}$

and consider $\nu(t) = (\cos t, \sin t)$ parameterizing the unit circle. Recall that the field $(a_1, a_2) = (x, y)$ from the PDE. Along the image of ν , this field is $(a_1, a_2) = (\cos t, \sin t)$. Therefore,

$$\nu' \cdot (a_1(\nu), a_2(\nu))^\perp = (-\sin t, \cos t) \cdot (-\sin t, \cos t) = 1 \neq 0.$$

So our chosen initial curve $\nu(t) = (\cos t, \sin t)$ is non-characteristic. You will often have the non-characteristic initial curve chosen for you—or be stuck with a characteristic initial curve (which can be a problem). With practice, however, you should get a feel for good initial non-characteristic manifolds for a given PDE.

Now, let's take a point $(x, y) \in \mathbb{R}^2 \setminus \{(0, 0)\}$, and see if we can address the second question above, at least for this particular example. We imagine a point $(x_0, y_0) = (\cos t_0, \sin t_0)$ along the non-characteristic circle, and we look for a characteristic $\gamma = (\xi, \eta)$ starting at (x_0, y_0) . This means

$$\gamma(t) = (x_0 e^t, y_0 e^t).$$

Can we choose x_0 , y_0 , and t so that $\gamma(t) = (x, y)$? Sure we can. Taking $x_0 = x/\sqrt{x^2 + y^2}$ and $y_0 = y/\sqrt{x^2 + y^2}$ we see that when $t = \ln \sqrt{x^2 + y^2}$ $\gamma(t) = (x, y)$. Thus, if we take any function $u_0 : \mathbb{S}^1 \rightarrow \mathbb{R}$ where $\mathbb{S}^1 = \{(x, y) : x^2 + y^2 = 1\}$ is the unit circle and $u_0 \in C^1(\mathbb{S}^1)$ (whatever that means), then

$$u(x, y) = u_0 \left(\frac{x}{\sqrt{x^2 + y^2}}, \frac{y}{\sqrt{x^2 + y^2}} \right)$$

should be a/the solution of our PDE. One way to interpret the regularity condition $u_0 \in C^1(\mathbb{S}^1)$ is to say there exists an open subset V of \mathbb{R}^2 and a function $v \in C^1(V)$ so that $\mathbb{S}^1 \subset V$ and the restriction

$$v|_{\mathbb{S}^1} \quad \text{satisfies} \quad v|_{\mathbb{S}^1} = u_0.$$

In this case, we can write the proposed solution as

$$u(x, y) = v \left(\frac{x}{\sqrt{x^2 + y^2}}, \frac{y}{\sqrt{x^2 + y^2}} \right).$$

Then we can compute.

$$\begin{aligned} \frac{\partial u}{\partial x} &= \frac{\partial v}{\partial x} \left(\frac{1}{\sqrt{x^2 + y^2}} - \frac{x^2}{(x^2 + y^2)^{3/2}} \right) + \frac{\partial v}{\partial y} \left(-\frac{xy}{(x^2 + y^2)^{3/2}} \right) \\ &= \frac{1}{(x^2 + y^2)^{3/2}} \left(y^2 \frac{\partial v}{\partial x} - xy \frac{\partial v}{\partial y} \right) \end{aligned}$$

and

$$\begin{aligned}\frac{\partial u}{\partial y} &= \frac{\partial v}{\partial x} \left(-\frac{xy}{(x^2 + y^2)^{3/2}} \right) + \frac{\partial v}{\partial y} \left(\frac{1}{\sqrt{x^2 + y^2}} - \frac{y^2}{(x^2 + y^2)^{3/2}} \right) \\ &= \frac{1}{(x^2 + y^2)^{3/2}} \left(x^2 \frac{\partial v}{\partial y} - xy \frac{\partial v}{\partial x} \right).\end{aligned}$$

This means

$$x \frac{\partial u}{\partial x} + y \frac{\partial u}{\partial y} = 0.$$

And this means we have solved our first PDE.

Let's think a little bit about what happened here. For one thing, you probably should remember that when you solve ODEs (in general), you expect to get one or more “arbitrary” constants. These are the constants you usually try to choose to fit some initial condition. Maybe you didn't think about it like this, but what this means is that the solution set of an ODE, though it may include infinitely many different functions, e.g. $x'' = -x$ has solutions $x(t) = c_1 \cos t + c_2 \sin t$ for arbitrary c_1 and c_2 , but this solution set is essentially (somehow) **finite dimensional** or at least it is parameterized by a finite dimensional set $\{(c_1, c_2) : c_1, c_2 \in \mathbb{R}\} = \mathbb{R}^2$.

In contrast, for the PDE we can choose an **arbitrary function** $u_0 \in C^1(\mathbb{S}^1)$, and $C^1(\mathbb{S}^1)$, the collection of all continuously differentiable functions on the circle, no matter how you cut it, is an infinite dimensional space. What this indicates is that (at least sometimes) the solution spaces of PDEs are much bigger than the solution spaces of ODEs. Remember, however, that I mentioned some PDE have no solutions whatsoever, so this observation, if adopted as intuition, should be taken with a grain of salt.

At any rate, there are new and interesting things to consider here. Among them is the set $C^1(\mathbb{S}^1)$. Continuity, i.e., the set of function $C^0(\mathbb{S}^1)$ wouldn't be so bad. You can define the set of real valued continuous functions on any set with a notion of distance (or more generally on any set with a notion of open sets). Since these topics will probably be of more direct interest soon, I will postpone that discussion until later. For now, let me point out that there is a natural identification of $C^1(\mathbb{S}^1)$ with a certain subspace of $C^1(\mathbb{R})$, namely the subspace of 2π periodic functions traditionally denoted¹¹ by

$$C^1(\mathbb{T}) = \{f \in C^1(\mathbb{R}) : f(x + 2\pi) = f(x) \text{ for all } x \in \mathbb{R}\}.$$

¹¹Generally, \mathbb{T} is just another name for the circle \mathbb{S}^1 , but it is used especially in the context of Fourier series where one focuses on functions defined on \mathbb{S}^1 and specifically on the correspondence with 2π periodic functions on \mathbb{R} or $[0, 2\pi)$. More generally, \mathbb{T}^k for $k > 1$ is the Cartesian product of

Exercise 1.41 Given a real valued function $f : \mathbb{S}^1 \rightarrow \mathbb{R}$, define a function $g : \mathbb{R} \rightarrow \mathbb{R}$ by $g(\theta) = f(\cos \theta, \sin \theta)$. Show that given a function f for which there exists an open set $\mathcal{N} \subset \mathbb{R}^2$ with $\mathbb{S}^1 \subset \mathcal{N}$ and an extension $F \in C^1(\mathcal{N})$ with

$$F|_{\mathbb{S}^1} = f,$$

the function g given by the correspondence above satisfies $g \in C^1(\mathbb{T})$. Hint: Use F to compute derivatives of g and show they are continuous.

Exercise 1.42 Given $g \in C^1(\mathbb{T})$, show the following:

(a) There exists a unique function $f : \mathbb{S}^1 \rightarrow \mathbb{R}$ given by

$$f(x, y) = g(\theta)$$

where

$$\cos \theta = x \quad \text{and} \quad \sin \theta = y.$$

(b) There exists an open set $\mathcal{N} \subset \mathbb{R}^2$ and a function $F \in C^1(\mathcal{N})$ such that

$$F|_{\mathbb{S}^1} = f.$$

(c) Show the function F given/found in part (b) above is not unique.

Exercise 1.43 Find a natural non-characteristic curve on which to consider an initial function for the PDE

$$y \frac{\partial u}{\partial x} = x \frac{\partial u}{\partial y},$$

and solve the PDE.

\mathbb{S}^1 with itself k times. This is called the k -**torus** and is the natural setting for the consideration of Fourier series in several variables, which is in a certain sense analogous to our discussion of power series in several variables. For $k \geq 1$ the set \mathbb{S}^k denotes the boundary of the unit ball in \mathbb{R}^{k+1} or the k -**sphere**. When $k = 1$ the spaces \mathbb{S}^1 and $\mathbb{T} = \mathbb{T}^1$ coincide.

1.3 Wave equation

First order linear PDE are most closely related to the (one dimensional) wave equation

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2}$$

for the simple reason that the **wave operator** or d'Alembertian factors as two linear first order partial differential operators.

When we talk about a **one-dimensional wave equation** we are referring to the **spatial dimension**. Of course, if we only have one dimension altogether for (the domain of a solution of) a differential equation, then that should be an ordinary differential equation. It is customary for equations like the wave equation and the heat equation to designate one particular variable t as the time variable. Such equations are called **evolution equations** and, given a solution $u = u(x, t)$ of such an equation, we can think of the function f given by

$$f(x) = u(x, t_0)$$

for a fixed time $t = t_0$ as a particular function (in some function space) which changes in time as a function of t . The domain Ω of such a “snap-shot” function $f : \Omega \rightarrow \mathbb{R}$ is called the **spatial domain**. The variable $x \in \Omega$ in this context is called the **spatial variable** and might be a vector variable, i.e., $x = (x_1, x_2, \dots, x_n)$ might itself actually represent a dependence on **several variables**.

The simplest possibility for an evolution equation might be

$$\frac{\partial u}{\partial t} = f_t = \Phi[u]$$

where Φ is a partial differential operator only depending on partial derivatives with respect to the spatial variables. The **heat equation** would fall into this category, and perhaps this is why the heat equation is usually discussed before the wave equation, cf. Chapter 13 of Boas. But we're aiming to get a (preliminary, small) taste of the wave equation first. Before we really get started with that, let me introduce a little more general terminology and notation. You know the notation for the **gradient**

$$Du = \left(\frac{\partial u}{\partial x_1}, \frac{\partial u}{\partial x_2}, \dots, \frac{\partial u}{\partial x_n} \right) \tag{1.37}$$

of a function of several variables $u : U \rightarrow \mathbb{R}$ where U is an open subset of \mathbb{R}^n . We use the same notation in the context of evolution equations (PDE) with the small

difference that the function $u : \Omega \times [0, T)$ (and its partial derivatives as well) are assumed to also depend on the time variable t , usually found in some time interval $[0, T) \subset \mathbb{R}$ as I've indicated. So then technically, according to our previous discussion of the gradient, we should really have

$$Du = \left(\frac{\partial u}{\partial x_1}, \frac{\partial u}{\partial x_2}, \dots, \frac{\partial u}{\partial x_n}, \frac{\partial u}{\partial t} \right). \quad (1.38)$$

We do not do this. When we write Du with $u = u(x, t)$ a solution of a PDE, specifically an evolution equation, we use the notation suggested in (1.37). If we want to emphasize this usage, we refer to the gradient in (1.37) as the **spatial gradient**, and sometimes we might write something like

$$D^x u = \left(\frac{\partial u}{\partial x_1}, \frac{\partial u}{\partial x_2}, \dots, \frac{\partial u}{\partial x_n} \right)$$

for the spatial gradient of $u = u(x, t)$, but using such notation for the spatial gradient is a bit unusual. On the other hand, if we really do want to refer to the full gradient with respect to space and time given in (1.38), then we will definitely denote it by $D^{x,t}u$. This also doesn't happen very often, but it does happen.

In more than one space dimension, the wave equation looks like

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2} + \dots + \frac{\partial^2 u}{\partial x_n^2}. \quad (1.39)$$

The operator on the right has natural domain $C^2(\Omega \times (0, T))$ where Ω is an open subset of \mathbb{R}^n and is denoted by

$$\Delta u = \frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2} + \dots + \frac{\partial^2 u}{\partial x_n^2}.$$

Again, these homogeneous second partials may (and probably will) depend on the time variable t . Following the convention we used for the gradient, this operator is called the **Laplacian** in this context, but should perhaps more properly be called the **spatial Laplacian**.

Exercise 1.44 Consider the function $u \in C^\infty(\mathbb{R}^{n+1})$ given by

$$u(x, t) = t^2 + \sum_{j=1}^n x_j^2.$$

Show u satisfies Poisson's equation

$$\frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2} + \cdots + \frac{\partial^2 u}{\partial x_n^2} + \frac{\partial^2 u}{\partial t^2} = 2n + 2.$$

I do not anticipate saying much about the n -dimensional wave equation given in (1.39). We will stick mostly to one space dimension when it comes to the wave equation. For the moment, I would like to draw your attention to some simple solutions and a technique for finding some more general solutions of the one-dimensional wave equation by factoring the operator as mentioned above.

Simple wave forms

Notice that in one dimension

$$\frac{\partial^2}{\partial x^2} \cos(ax) = -a^2 \cos(ax).$$

Consequently, if we consider $u : \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}$ by

$$u(x, t) = g(t) \prod_{j=1}^n \cos(a_j x_j)$$

for any constants $a_1, a_2, \dots, a_n \in \mathbb{R}$, then

$$\Delta u = - \left(\sum_{j=1}^n a_j^2 \right) u. \tag{1.40}$$

Also,

$$u_t = g'(t) \prod_{j=1}^n \cos(a_j x_j).$$

Therefore, if we take $g(t) = \cos(bt)$ where

$$b = \sqrt{\sum_{j=1}^n a_j^2},$$

then we get

$$u(x, t) = \cos \left(\sqrt{\sum_{j=1}^n a_j^2} t \right) \prod_{j=1}^n \cos(a_j x_j)$$

is a solution of the n -dimensional wave equation. This observation/construction simplifies in a pretty nice way in one space dimension.

Exercise 1.45 Consider

$$u(x, t) = \cos(at) \sin(ax)$$

for some $a > 0$.

- (a) Plot u as a function of t and x on the strip $[0, \pi/a] \times [0, \infty)$.
- (b) Plot the evolution of $u_0(x) = \sin(ax)$ for $0 \leq x \leq \pi/a$ and $t > 0$.
- (c) Use mathematical software to **animate** the evolution of u_0 .

So far, I've said what the wave equation is as a PDE, introduced some related notation and terminology, and given a few rather simple explicit solutions. When I return to the wave equation next time, I should start by discussing the initial and boundary conditions properly associated with this PDE. It is natural at this point, however, to discuss one more elementary topic, namely the construction of solutions obtained by factoring the operator. This uses directly the material above on first order linear PDE and the method of characteristics. One consequence of this discussion is a nice derivation of a general solution of the (one-dimensional) wave equation with spatial domain the entire real line. This is called d'Alembert's solution, and I guess I will basically leave that derivation to you as an exercise.

Say we have Cauchy data given by $u(x, 0) = x^2$ and $u_t(x, 0) = -3x$ on the spatial interval $(-1, 2)$ in the line. Notice that the wave operator or d'Alembertian

$$\square u = u_{tt} - u_{xx} = (u_t - u_x)_t + (u_t - u_x)_x.$$

That is,

$$\square = \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial x} \right) \left(\frac{\partial}{\partial t} - \frac{\partial}{\partial x} \right)$$

factors as a composition of two constant coefficient linear first order partial differential operators. Furthermore, if we set $w = u_t - u_x$, then a solution

$$u \in C^2(W) \cap C^1(\underline{W}) \tag{1.41}$$

of the wave equation¹² with the Cauchy data given above provides us with a Cauchy problem satisfied by w . To see this, note first of all, that

$$w_t + w_x = 0, \quad (1.42)$$

so that is a first order PDE for w . Furthermore when $t = 0$, we should have

$$w(x, 0) = u_t(x, 0) - u_x(x, 0) = -3x - 2x = -5x \quad \text{for} \quad -1 < x < 2. \quad (1.43)$$

Thus, we can consider a characteristic propagating into the half plane of positive time. More precisely, let $x_0 \in (-1, 2)$ and for $t \geq 0$ let $\xi = \xi(t)$ satisfy the IVP

$$\begin{cases} \xi' = 1, & t > 0 \\ \xi(0) = x_0. \end{cases} \quad (1.44)$$

If we compose w on the parameterized curve $\alpha(t) = (\xi(t), t)$, then we find

$$\frac{d}{dt}w \circ \alpha(t) = \frac{\partial w}{\partial x} \xi'(t) + \frac{\partial w}{\partial t} = 0$$

by (1.42). Thus, w is constant along the characteristic. Now, let (x, t) be any point with $t > 0$, we seek a starting position x_0 so that the characteristic determined by (1.44) passes through (x, t) . This is not difficult. Since $\alpha(t) = (x_0 + t, t)$, we see $x_0 = x - t$ is the unique choice. There is of course an additional subtlety to consider. Our initial data for w given in (1.43) above required the argument x to lie in the interval $(-1, 2) \subset \mathbb{R}$. This means that if our choice $x_0 = x - t$ is to give us a valid initial point $(x_0, 0)$ we must have

$$-1 < x - t < 2.$$

This tells us, perhaps not surprisingly in retrospect, a natural condition to impose on the domain W . Namely,

$$W \subset \{(x, t) : t > 0, t - 1 < x < t + 2\}.$$

The superset here, it will be noted, is a diagonal strip tending spatially to the right as illustrated in Figure 1.5. Thus, assuming (x, t) is in the strip $\{(x, t) : t > 0, t - 1 <$

¹²You will presumably have no idea what the regularity assumption on u specified in (1.41) actually means. I haven't even said what W might be, much less \underline{W} . You might guess what these conditions might mean, however, and I can either tell you in a moment below, or I can tell you if you ask, or I can cover it when we come back to the wave equation in more detail later. I guess we'll see what happens.

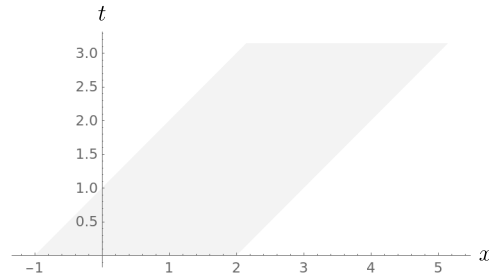


Figure 1.5: An initial region of exclusion: Because the characteristics associated with the PDE for w propagate to the right, the interior of the domain of definition W should lie within the strip $\{(x, t) : t > 0, t - 1 < x < t + 2\}$ in which the spatial domain propagates to the right with increasing t .

$x < t + 2\}$ we have

$$w(x, t) = w(x_0, 0) = w(x - t, 0) = -5(x - t).$$

This, then becomes a forcing term in a first order PDE for the solution $u = u(x, t)$ of the wave equation:

$$u_t - u_x = w = -5(x - t).$$

We couple this with the initial condition/Cauchy data $u(x, 0) = x^2$ for $-1 < x < 2$. Again we use the method of characteristics seeking a characteristic $\beta(t) = (\eta(t), t)$ with $\beta'(t) = (-1, 1)$ and $\eta(0) = x_0$. The characteristic is given by $\beta(t) = (x_0 - t, t)$ and propagates left in space so that

$$\frac{d}{dt}u \circ \beta(t) = w \circ \beta(t) = -5(x_0 - 2t).$$

We conclude

$$u \circ \beta(t) = u(x_0, 0) + \frac{5}{4}(x_0 - 2t)^2 - \frac{5}{4}x_0^2,$$

as long as there is some x_0 with $-1 < x_0 < 2$ and $\beta(t) = (x, t) \in W$. For this latter condition, we need W in the strip in space-time determined by propagation of the spatial domain $(-1, 2)$ left with unit speed. Consequently, our solution method applies to the wave domain

$$W = \{(x, t) : 0 < t < 3/2, t - 1 < x < -t + 2\}$$

and gives the solution $u : W \rightarrow \mathbb{R}$ by

$$\begin{aligned} u(x, t) &= u(x_0, 0) + \frac{5}{4}(x_0 - 2t)^2 - \frac{5}{4}x_0^2 \\ &= (x + t)^2 + \frac{5}{4}(x - t)^2 - \frac{5}{4}(x + t)^2 \\ &= \frac{5}{4}(x - t)^2 - \frac{1}{4}(x + t)^2. \end{aligned}$$

This function is easily seen to satisfy all conditions of the problem with $u \in C^2(W) \cap C^1(\underline{W})$ where $\underline{W} = \{(x, t) : 0 \leq t < 3/2, t - 1 < x < -t + 2\}$. In fact, u extends by the same formula to a function in

$$C^2(W) \cap C^1(\underline{W}) \cap C^0(|\underline{W}|) \quad \text{where} \quad |\overline{W}| = \{(x, t) : 0 \leq t < 3/2, t - 1 \leq x \leq -t + 2\},$$

and this represents the “natural classical regularity” for the problem. It will not escape your notice that u extends by the same formula to be real analytic on all of \mathbb{R}^2 . In particular, $u \in C^\omega(\overline{W}) \cap C^\omega(\mathbb{R}^2)$ where we recall

$$\overline{W} = \{(x, t) : 0 \leq t \leq 3/2, t - 1 \leq x \leq -t + 2\}.$$

This last extension is a solution of the wave equation on all of \mathbb{R}^2 , but it is interesting to consider to what extent the values of a solution of the wave equation on sets larger than W must agree with the values of this natural extension. Note, for example that the extension formula

$$u(x, t) = \frac{5}{4}(x - t)^2 - \frac{1}{4}(x + t)^2$$

satisfies

$$u(-1, t) = \frac{3t^2 + 4t + 3}{2}$$

which is not constant in time.

Let us attempt to find a solution¹³ $v : [-1, 2] \times [0, 3/2] \rightarrow \mathbb{R}$ of the wave equation satisfying $v(-1, t) \equiv 1 = u_1(-1, 0)$ and $v(2, t) \equiv 4 = u_1(2, 0)$. We again write $w = v_t - v_x$ so that $w_t + w_x = 0$ and $w(-1, t) = v_t(-1, t) - v_x(-1, t) = -v_x(-1, t)$.

¹³There is some natural interest in having an evolution on all of the initial spatial domain $(-1, 2)$. Of course, we have such a solution in u , but it is also natural (perhaps) to impose different boundary conditions at the spatial endpoints $x = -1$ and $x = 2$.

The function $g(t) = -v_x(-1, t)$ we may take as arbitrary but to be specified later. That is, we assume for now there is some function $g : (0, \infty) \rightarrow \mathbb{R}$ for which

$$w(-1, t) = g(t) \quad \text{for} \quad t > 0. \quad (1.45)$$

Notice the characteristics associated with the equation

$$w_t + w_x = 0$$

impinge on the boundary line $x = -1$. Accordingly, we start with a point $(-1, t_0)$ on this line, and attempt to propagate the value $g(t_0)$ into the region

$$\{(x, t) : t > 0, -1 < x < -1 + t\}$$

as indicated in Figure 1.6. Accordingly we have

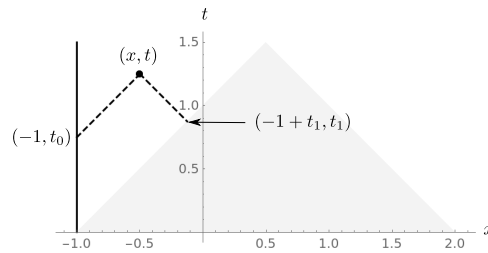


Figure 1.6: Given a solution u of the wave equation determined by Cauchy data at time $t = 0$ in the region $W = \{(x, t) : 0 < t < 3/2, -1 + t < x < 2 - x\}$, we attempt to find an extension v satisfying the wave equation classically in $W_- = \{(x, t) : t > 0, -1 < x < -1 + t\}$ and satisfying $v(-1, t) \equiv u(-1, t) = 1$.

$$\frac{d}{d\tau} w(-1 + \tau, t_0 + \tau) = w_x + w_t = 0.$$

From this and (1.45) we conclude

$$w(-1 + \tau, t_0 + \tau) \equiv w(-1, t_0) = g(t_0).$$

It remains to determine the argument of the function g . For this, we need $x = -1 + \tau$ and $t = t_0 + \tau$. That is,

$$\tau = x + 1 \quad \text{and} \quad t_0 = t - x - 1.$$

Hence,

$$w(x, t) = g(t - x - 1) = G(x, t).$$

Having specified w in this tentative form, we turn to the PDE for v . The characteristic of $v_t - v_x = G(x, t)$ terminating at $(x, t) \in W_-$ impinges, as indicated in Figure 1.6, on the boundary of W at a point $(-1 + t_1, t_1)$. Again, we compute using the chain rule/directional derivatives:

$$\frac{d}{d\tau} v(-1 + t_1 - \tau, t_1 + \tau) = -v_x + v_t = G(-1 + t_1 - \tau, t_1 + \tau) = g(2\tau).$$

Integrating, we find for a solution v with values continuously matching those of u along ∂W there must hold

$$v(-1 + t_1 - \tau, t_1 + \tau) = v(-1 + t_1, t_1) + \int_0^\tau g(2\sigma) d\sigma = u(-1 + t_1, t_1) + \int_0^\tau g(2\sigma) d\sigma.$$

The starting point for this characteristic should be determined by the terminal relations $-1 + t_1 - \tau = x$ and $t_1 + \tau = t$, or

$$\begin{cases} t_1 - \tau = x + 1 \\ t_1 + \tau = t. \end{cases}$$

We conclude

$$\begin{aligned} v(x, t) &= u\left(\frac{x+t-1}{2}, \frac{x+t+1}{2}\right) + \int_0^{(t-x-1)/2} g(2\sigma) d\sigma \\ &= \frac{5 - (x+t)^2}{4} + \int_0^{(t-x-1)/2} g(2\sigma) d\sigma. \end{aligned}$$

We recall finally that we wished to have $v(-1, t) \equiv 1$. For this we need

$$\int_0^{t/2} g(2\sigma) d\sigma = 1 - \frac{5 - (t-1)^2}{4}$$

or

$$\frac{1}{2} g(t) = \frac{t-1}{2}.$$

That is,

$$g(t) = t - 1.$$

We obtain then

$$\begin{aligned}
 v(x, t) &= \frac{5 - (x + t)^2}{4} + \int_0^{(t-x-1)/2} (2\sigma - 1) d\sigma \\
 &= \frac{5 - (x + t)^2}{4} + \frac{1}{4}(t - x - 1)^2 - \frac{t - x - 1}{2} \\
 &= 2 + x - xt - t.
 \end{aligned}$$

Weak solutions

Say we take a function $u \in C^2(\mathbb{R} \times [0, \infty))$ meaning there is some open set $U \subset \mathbb{R}^2$ with $\mathbb{R} \times [0, \infty) \subset U$ and $u \in C^2(U)$. Let us also assume this function u satisfies the wave equation classically:

$$u_{tt} = u_{xx} \quad \text{on} \quad (x, t) \in \mathbb{R} \times (0, \infty)$$

with initial values/Cauchy data $u(x, 0) = f(x)$ and $u_t(x, 0) = g(x)$ taken classically. Multiplying the PDE by a test function $\phi \in C_c^\infty(\mathbb{R} \times [0, \infty))$ and integrating we have

$$\int_{\mathbb{R} \times (0, \infty)} [u_{tt} - u_{xx}] \phi = 0. \quad (1.46)$$

Now, take the first term and write the area integral as an iterated integral:

$$\int_{\mathbb{R} \times (0, \infty)} u_{tt} = \int_{\mathbb{R}} \left(\int_{(0, \infty)} u_{tt} \phi \right).$$

Writing the inner integral in the usual notation and integrating by parts we find

$$\begin{aligned}
 \int_0^\infty u_{tt} \phi dt &= (u_t \phi) \Big|_{t=0}^\infty - \int_0^\infty u_t \phi_t dt \\
 &= -u_t(x, 0) \phi(x, 0) - \left((u \phi_t) \Big|_{t=0}^\infty - \int_0^\infty u \phi_{tt} dt \right) \\
 &= -u_t(x, 0) \phi(x, 0) + u(x, 0) \phi_t(x, 0) + \int_0^\infty u \phi_{tt} dt \\
 &= -g(x) \phi(x, 0) + f(x) \phi_t(x, 0) + \int_0^\infty u \phi_{tt} dt.
 \end{aligned}$$

We consider also the term

$$\int_{\mathbb{R} \times (0, \infty)} u_{xx} = \int_{(0, \infty)} \left(\int_{\mathbb{R}} u_{xx} \phi \right).$$

For this inner integral we have

$$\begin{aligned} \int_{-\infty}^{\infty} u_{xx} \phi \, dx &= (u_x \phi) \Big|_{x=-\infty}^{\infty} - \int_{-\infty}^{\infty} u_x \phi_x \, dx \\ &= - \left((u \phi_x) \Big|_{x=-\infty}^{\infty} - \int_{-\infty}^{\infty} u \phi_{xx} \, dx \right) \\ &= \int_{-\infty}^{\infty} u \phi_{xx} \, dx. \end{aligned}$$

Substituting these expressions for the inner integrals (1.46) we conclude

$$- \int_{x \in \mathbb{R}} g(x) \phi(x, 0) + \int_{x \in \mathbb{R}} f(x) \phi_t(x, 0) + \int_{\mathbb{R} \times (0, \infty)} u[\phi_{tt} - \phi_{xx}] = 0.$$

Notice that this expression formally makes sense whenever $u \in C^0(\mathbb{R} \times [0, \infty))$. Thus, we arrive at the following definition:

Definition 9 (continuous weak solution of the wave equation) A function $w \in C^0(\mathbb{R} \times [0, \infty))$ is said to be a **weak solution** of the initial value problem

$$\begin{cases} u_{tt} = u_{xx}, & (x, t) \in \mathbb{R} \times (0, \infty) \\ u(x, 0) = f(x), & x \in \mathbb{R} \\ u_t(x, 0) = g(x), & x \in \mathbb{R} \end{cases} \quad (1.47)$$

if

$$\int_{\mathbb{R} \times (0, \infty)} w[\phi_{tt} - \phi_{xx}] = \int_{x \in \mathbb{R}} g(x) \phi(x, 0) - \int_{x \in \mathbb{R}} f(x) \phi_t(x, 0)$$

for every $\phi \in C_c^\infty(\mathbb{R} \times [0, \infty))$.

There are a couple nice theorems that go along with this definition. One of them states that every weak solution of (1.47) is unique and is given by d'Alembert's formula:

$$w(x, t) = \frac{1}{2}[f(x+t) + f(x-t)] + \frac{1}{2} \int_{(x-t, x+t)} g. \quad (1.48)$$

This is a little bit tricky to prove, and we probably won't cover it, though when we have done a little integration theory, you should have most of the tools to go through the proof.

It is also true that given any $f, g \in C^0(\mathbb{R})$, the formula (1.48) defines a weak solution of (1.47). Notice, however, that the function g is not in general required to be continuous either for the definition of a weak solution to hold nor for (1.48) to give a weak solution.

Of course, a classical solution is a weak solution and a classically differentiable weak solution is a classical solution. These assertions are included in Assignment 9.

1.4 Heat Equation

The heat equation has the form

$$u_t = \Delta u,$$

and many of the comments above concerning the wave equation as an evolution equation apply, though the behavior of solutions is strikingly different. We will again consider some simple solutions in one space dimension (or perhaps sometimes more) both on a bounded domain and on all of the spatial domain \mathbb{R} (or perhaps \mathbb{R}^n). Returning to (1.40) we observe that

$$u(\mathbf{x}, t) = e^{-c^2 t} \prod_{j=1}^n \sin\left(\frac{\pi x_j}{\ell_j}\right)$$

where $\ell_1, \ell_2, \dots, \ell_n$ are positive lengths and

$$c^2 = \sum_{j=1}^n \frac{\pi^2}{\ell_j^2}$$

is a solution $u \in C^\omega(\overline{\Omega \times (0, \infty)})$ where

$$\Omega = \prod_{j=1}^n (0, \ell_j) = (0, \ell_1) \times (0, \ell_2) \times \cdots \times (0, \ell_n)$$

of the initial/boundary value problem

$$\begin{cases} u_t = \Delta u, & (\mathbf{x}, t) \in \Omega \times (0, \infty) \\ u(\mathbf{x}, 0) = \prod_{j=1}^n \sin(\pi x_j / \ell_j) & \mathbf{x} \in \Omega \\ u(\mathbf{x}, t) = 0, & \mathbf{x} \cdot \mathbf{e}_j = 0, (\mathbf{x} - \ell_j \mathbf{e}_j) \cdot \mathbf{e}_j = 0, t > 0. \end{cases} \quad (1.49)$$

Again, we specialize to one space dimension, so the problem becomes

$$\begin{cases} u_t = u_{xx}, & (x, t) \in (0, \ell) \times (0, \infty) \\ u(x, 0) = \sin(\pi x / \ell) & x \in (0, \ell) \\ u(0, t) = 0 = u(\ell, t), & t > 0 \end{cases}$$

with solution

$$u(x, t) = e^{-\pi^2 t / \ell^2} \sin\left(\frac{\pi x}{\ell}\right).$$

Exercise 1.46 Consider

$$u(x, t) = e^{-a^2 t} \sin(ax)$$

for some $a > 0$.

- (a) Plot u as a function of t and x on the strip $[0, \pi/a] \times [0, \infty)$.
- (b) Plot the evolution of $u_0(x) = \sin(ax)$ for $0 \leq x \leq \pi/a$ and $t > 0$.
- (c) Use mathematical software to **animate** the evolution of u_0 .

1.5 Laplace's equation

All the solutions u of the heat equation considered above have the property

$$\lim_{t \nearrow \infty} u(x, t) = 0.$$

One should not get the idea from this that all solutions of the heat equation have this property, though admittedly in one space dimension the limit

$$\lim_{t \nearrow \infty} u(x, t)$$

of a solution of the heat equation, i.e., the (model) long time heat distribution, is not very exciting. Note that $v(x, t) = x$ is a solution of the heat equation.

Exercise 1.47 Consider

$$u(x, t) = x + e^{-a^2 t} \sin(ax)$$

for some $a > 0$.

- (a) Plot u as a function of t and x on the strip $[0, \pi/a] \times [0, \infty)$.
- (b) Plot the evolution of $u_0(x) = x + \sin(ax)$ for $0 \leq x \leq \pi/a$ and $t > 0$.
- (c) Use mathematical software to **animate** the evolution of u_0 .

Functions in the kernel of the Laplace operator $\Delta : C^2(U) \rightarrow C^0(U)$ are called **harmonic functions**.

Exercise 1.48 Find the general solution of Laplace's equation (which becomes an ODE) in one space dimension.

Chapter 2

Integration

As with most topics presented in this course, I attempt to present the topic in a relatively general form with some key points for guidance. There are many kinds of integrals in the world. Among the most useful are the **abstract Riemann integrals on sets** presented below, in fact presented now:

Given a set A on which integration is possible¹ the abstract Riemann integral of a real valued function $f : A \rightarrow \mathbb{R}$ is defined as the limit

$$\int_A f = \lim_{\|\mathcal{P}\| \rightarrow 0} \sum_{j=1}^n f(p_j^*) \mu(A_j). \quad (2.1)$$

The sum on the right is called a **Riemann sum**. The sets A_1, A_2, \dots, A_n form a **partition** of the set A and are called the “pieces” of A . This simply means

$$A = \bigcup_{j=1}^n A_j$$

with “small” pairwise overlaps, i.e., $\mu(A_i \cap A_j) = 0$ if $i \neq j$. The value $\mu(A_j)$ is the **measure**. One has to define a measure to integrate, and there are many different measures in the world. I won’t go into the details of what it means for a function $\mu : \mathcal{M} \rightarrow [0, \infty)$ (or more often $\mu : \mathcal{M} \rightarrow [0, \infty]$) to be a measure on a collection of subsets of a given set, but you already know quite a few measures, or can get to know them easily:

¹I’ll discuss what is required later.

1. **Counting measure** is the number of points in a set. (This may not be a measure you have thought of before, but it's a measure.) The counting measure of a set A is often denoted by $\#(A)$. Notice $\#(A) = \infty$ if A has infinitely many elements.
2. **Length** is a measure both for many subsets you know in the real line and for subsets of curves and other sets in higher dimensional Euclidean spaces as well. Length is sometimes denoted by $m(A)$ or $\mu(A)$. Sometimes we can also write $\text{length}(A)$. Especially, if we are measuring the length of a subset $A \subset \mathbb{R}^n$, then we usually denote the length measure by $\mathcal{H}^1(A)$. This is called length measure in \mathbb{R}^n or **one-dimensional Hausdorff measure**. Notice that $\mathcal{H}^1(\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1\}) = \infty$.
3. **Area** is a nice measure. As with length, there are various flavors of area. If you measure area of subsets of \mathbb{R}^1 or in any line, you'll always get zero. If you measure area of subsets of the plane, then the area measure is often denoted (again) by m or μ . These symbols are used simply because the first letter of the word measure is "m." We can also write $\text{area}(A)$ in many cases. But if we measure the area of a subset of \mathbb{R}^3 , for example a surface or a piece of a surface—assuming say that we wanted to integrate on a surface—then we should use **two-dimensional Hausdorff measure** in \mathbb{R}^3 : \mathcal{H}^2 .

Hopefully, this is an adequate description for you to understand the basics of what it means to **measure** sets. There is **volume** measure and **hypervolume measure** of all dimensions (and in all ambient dimensions). There are lots of less familiar measures too, but these will get us a long way.

Conclusion/important take-away: You need a measure to integrate.

What else is up there in the Riemann sum and the definition of the integral (2.1)? Working from right to left I see the **evaluation points** p_j^* . It is assumed that $p_j^* \in A_j$ and $f(p_j^*)$ is (of course) the value of f (evaluated) at the point p_j^* .

In this definition, the partition is given a name. That name is \mathcal{P} . Thus,

$$\mathcal{P} = \{A_1, A_2, \dots, A_n\} = \{A_j\}_{j=1}^n.$$

Partitions can take various forms, and the indexing can vary a bit too, but this gives the basic idea. One notable variation is when one partitions an interval $[a, b]$. In this

case, one often refers to the **partition points** rather than the **partition sets**. To be precise, many calculus books will introduce the partition

$$a = x_0 < x_1 < x_2 < \dots < x_n = b.$$

This means there are subintervals which might be designated by something like $A_0 = [x_0, x_1]$, $A_1 = [x_1, x_2]$, \dots , $A_{n-1} = [x_{n-1}, x_n]$. And the Riemann sum might look like

$$\int_a^b f(x) dx = \lim_{\max(x_{j+1}-x_j) \rightarrow 0} \sum_{j=0}^{n-1} f(x_j^*)(x_{j+1} - x_j).$$

This explains, for example, why the evaluation points are called p_j^* instead of just p_j .

Returning to the general situation, the **norm of the partition** is taken (abstractly) to be

$$\|\mathcal{P}\| = \max_j \text{diam}(A_j)$$

where the maximum is taken over all the partition pieces, and diam is some kind of measure of the extent of the piece. The measure diam is often not really a measure in the technical sense, but it is another function on sets which usually dominates the measure in the sense that

$$\lim \text{diam}(A) = 0 \quad \text{implies} \quad \lim \mu(A) = 0.$$

Very often there are constants α and β so that $\mu(A) \leq \alpha \text{diam}(A)^\beta$. Also, in Euclidean spaces we very often take

$$\text{diam}(A) = \sup\{\|p - q\| : p, q \in A\}.$$

What (2.1) means then is the following: *There is some number $I \in \mathbb{R}$ so that for any $\epsilon > 0$, there is some $\delta > 0$ for which*

$$\|\mathcal{P}\| < \delta \quad \text{implies} \quad \left| I - \sum_{j=1}^n f(p_j^*)\mu(A_j) \right| < \epsilon$$

no matter which evaluation points and which partition pieces are used. If this happens, we call the number I the integral and denote it by the left side of (2.1).

The definition I have just given you agrees in most respects with all the other integrals you've used and run across before. There is one significant difference: **This**

kind of integral does not have orientation. For example, there is no general version of the identity

$$\int_b^a f(x) dx = - \int_a^b f(x) dx$$

which depends on a fixed convention concerning orientation/direction of the integration on \mathbb{R} . Our abstract integration doesn't have that, and this must be taken into account occasionally. But there is a big advantage in that we can use this kind of integral to integrate on a much broader collection of objects. All that is required is to be able to (1) cut the objects up in “small” pieces, (2) calculate the measures of those pieces, and (3) find the required limit—or show it exists.

Concerning the last point, there is a separate general structure with which you are familiar and which, when in place, can be used to get the existence of integrals. It is this:

Generally speaking, when it can be said that $f \in C^0(A)$, i.e., the real valued function $f : A \rightarrow \mathbb{R}$ is **continuous**, then abstract set integration will very often work without a hitch.

This is not to say there may still be various special considerations associated with things like improper integrals that have to be obtained as secondary limits, and there may be other “pathological” situations where an integral or two may not exist. But continuity is generally sufficient for Riemann integration.

I've now presented the definition. It's time for the general guiding principles.

2.1 Evaluation of integrals

Most every integral we encounter must eventually be reduced to the basic integrals from 1-D calculus, i.e., Calculus I. There are basically two tools used to do this:

1. change of variables
2. iterated integrals

The first tool is somewhat more general and the second one is somewhat more special, but they are both necessary, and you should master them.

The basic principle of change of variables, is that if you have a nice parameterization of your set A given by some function $\psi : B \rightarrow A$, which we assume is invertible,

then you can integrate f on A by integrating the composition $f \circ \psi$ on B . The formula looks like this:

$$\int_A f = \int_B f \circ \psi \sigma.$$

The factor σ is called the **scaling factor** and it is included in the integrand of the second integral, so it's important to remember, you can't just integrate the composition $f \circ \psi$. You know this.

Say I want to integrate the function x^2 on the interval $[0, 2]$. Of course, I can just integrate, but say for whatever reason, I want to change variables using $\psi : [0, 1] \rightarrow [0, 2]$ by $\psi(\xi) = 2\xi$. The set A here is $[0, 2]$ and the set B is $[0, 1]$. The function $f : [0, 2] \rightarrow \mathbb{R}$ is given by $f(x) = x^2$. If I were to write

$$\int_{x \in [0, 2]} x^2 = \int_{\xi \in [0, 1]} (2\xi)^2 \tag{2.2}$$

that would be wrong. The composition $f \circ \psi(\xi) = (2\xi)^2$ is basically okay giving the function to integrate, but this only gives the function values. I have to take account of the fact that the domain variable is getting stretched by ψ . Put another way, the measure $\mu(B_j)$ of a small piece of $B = [0, 1]$ corresponding to a piece $\psi(B_j) = A_j$ does not have the correct measure $\mu(A_j)$ to give the Riemann sum for the integral on the left in (2.2). A scaling factor is required, and that factor is $\sigma = 2$, because the pieces in $B = [0, 1]$ are half the length of the corresponding pieces in $A = [0, 2]$. Thus, the correct change of variables is

$$\int_{x \in [0, 2]} x^2 = \int_{\xi \in [0, 1]} (2\xi)^2 (2).$$

Back in Calculus I this would have been expressed as a "u"-substitution as $x = 2u$ with $dx = 2 du$ so that

$$\int_0^2 x^2 dx = \int_0^1 (2u)^2 (2) du = \int_0^1 (2\xi)^2 \psi'(\xi) d\xi.$$

For integration on \mathbb{R}^1 we see the orientation difference: In Calculus I

$$\int_a^b f(x) dx = \int_{\psi^{-1}(a)}^{\psi^{-1}(b)} f \circ \psi(\xi) \psi'(\xi) d\xi,$$

but with our set integrals, the scaling factor is $\sigma = |\psi'(\xi)|$ in this case, so

$$\int_{(a,b)} f = \int_{\psi^{-1}(a,b)} f \circ \psi |\psi'|$$

with $\psi^{-1}(a, b) = \{\psi^{-1}(x) : x \in (a, b)\}$. You see, we have no way to determine if ψ reversed the orientation of the endpoints. And of course, for most higher dimensional integrals, it's convenient to not keep track of such questions of orientation. We just integrate on sets.

Here are two really useful rules for scaling factors you'll want to memorize/remember:

1. If $\psi : B \rightarrow A$ with $B, A \subset \mathbb{R}^n$ and you're integrating with respect to the full dimension n -volume measure, then

$$\sigma = |\det D\psi|$$

where

$$D\psi = \left(\frac{\partial \psi_i}{\partial \xi_j} \right)_{ij}$$

is the $n \times n$ matrix containing all the first partial derivatives of ψ .

2. If $\psi : B \rightarrow A$ with $B \subset \mathbb{R}^k$ and $A \subset \mathbb{R}^n$ with $n > k$, then

$$\sigma = \sqrt{\det[D\psi^T D\psi]}.$$

In this case, $D\psi$ will not be a square matrix, but the product $D\psi^T D\psi$ will be a $k \times k$ positive definite matrix, so the determinant will be positive.

Those two scaling rules will get you a long way.

When $A \subset \mathbb{R}^n$ has a special form, then an integral $\int_A f$ can sometimes be expressed as iterated integrals over sets of lower dimension. You should have some familiarity with how this works. If for example $A \subset \mathbb{R}^2$ is a domain of the form

$$A = \{(x, y) \in \mathbb{R}^2 : a < x < b, \psi_1(x) < y < \psi_2(x)\}$$

for some function $\psi_1, \psi_2 : (a, b) \rightarrow \mathbb{R}$ with $\psi_1 < \psi_2$, then

$$\int_A f = \int_{x \in (a, b)} \left(\int_{y \in (\psi_1(x), \psi_2(x))} f(x, y) \right).$$

In the "inner" integral

$$\int_{y \in (\psi_1(x), \psi_2(x))} f(x, y)$$

the value of x is considered a constant. Generally, such relations between general integrals on sets and iterated integrals are considered applications of Fubini's theorem.

Let me conclude this section with the basic physical modeling principle of integration:

The real valued function f appearing in an integral may always be interpreted to model a **density**, so that

$$\int_A f$$

models the amount of the physical quantity having units matching the physical dimension

$$[f]L^k$$

where k is the dimension of the domain of integration A .

Perhaps the simplest example is when $f = \rho \in C^0(A)$ is a **mass density** with physical dimensions

$$[\rho] = \frac{M}{L^k}.$$

In this case

$$\int_A \rho$$

models the mass contained in the (model) region A .

If $n = 1$ and $[\rho] = M/L$, then ρ is called a **lineal mass density** or a **linear mass density**. If $n = 2$ and $[\rho] = M/L^2$, then ρ is called an **areal mass density** or an **area mass density** or a **laminar mass density**. These terms often are used to distinguish the context from that where $n = 3$ and $[\rho] = M/L^3$ and ρ is simply referred to as a **mass density**.

2.2 Divergence and the divergence theorem

Let us restrict attention to integration on a “full dimension” domain of integration with “smooth boundary.” This means, for example, that we have an open set $U \subset \mathbb{R}^n$ and ∂U is a C^1 hypersurface. As an even more specific example, consider $B_r(\mathbf{p}) \subset \mathbb{R}^3$ where \mathbf{p} is a fixed center in \mathbb{R}^2 and $U = B_r(\mathbf{p}) = \{\mathbf{x} \in \mathbb{R}^3 : \|\mathbf{x} - \mathbf{p}\| < r\}$ is the open ball of radius r centered at \mathbf{p} . The boundary of U in this case is a sphere, which is a two-dimensional surface in \mathbb{R}^3 . That spherical surface is C^1 in the sense that the surface admits a **continuous ambient outward unit normal field** N given by

$$N(\mathbf{x}) = \frac{\mathbf{x} - \mathbf{p}}{r}.$$

The detail may vary among domains of integration U , but this should give the basic idea. On such domains, we can form the **flux integral**

$$\int_{\partial U} \mathbf{v} \cdot N \quad (2.3)$$

associated with a vector field $\mathbf{v} \in C^0(\bar{U} \rightarrow \mathbb{R}^n)$. If the vector field has $n - 1$ spatial-rate-density dimension, that is

$$[\mathbf{v}] = \frac{V}{TL^{n-1}}$$

where V is some/any physical dimension, then the flux integral (2.3) gives the “amount” measured in units of V passing through ∂U per unit of time. For example, if \mathbf{v} is a simple velocity field with

$$[\mathbf{v}] = \frac{L}{T} = \frac{L^n}{TL^{n-1}},$$

measured in, say meters per second, then

$$\int_{\partial U} \mathbf{v} \cdot N$$

measures the n -dimensional volume measure passing out of U over ∂U . If a simple velocity field \mathbf{v} is weighted by a **mass density** function $\rho \in C^0(\bar{U})$ of full dimension so that

$$[\rho] = \frac{M}{L^n}$$

then

$$\int_{\partial U} \rho \mathbf{v} \cdot N$$

measures the rate of mass exiting U along ∂U .

Returning to a general field \mathbf{v} , it is interesting² to consider subdomains $Q \subset U$ shrinking “nicely” to a point $\mathbf{p} \in U$. I won’t go into the full generality of what it means to shrink “nicely,” but I’ll give you two examples:

$$Q = B_r(\mathbf{p}) = \{\mathbf{x} \in U : \|\mathbf{x} - \mathbf{p}\| < r\} \rightarrow \{\mathbf{p}\} \quad \text{as} \quad r \searrow 0$$

and

$$Q = Q_\epsilon(\mathbf{p}) = \left\{ \mathbf{x} \in U : |x_j - p_j| < \frac{\epsilon}{2}, j = 1, 2, \dots, n \right\} \rightarrow \{\mathbf{p}\} \quad \text{as} \quad \epsilon \searrow 0.$$

²It turns out to be interesting.

In words, a ball of radius $r > 0$ centered at \mathbf{p} shrinks nicely to \mathbf{p} as $r \searrow 0$, and a coordinate cube of side length ϵ centered at \mathbf{p} shrinks nicely to \mathbf{p} as $\epsilon \searrow 0$.

One can take as a goal to extract some real valued measure of the **divergence** of the field \mathbf{v} at the single point \mathbf{p} by considering the quantity

$$\int_{\partial Q} \mathbf{v} \cdot N$$

and the limit

$$\lim_{Q \rightarrow \{\mathbf{p}\}} \int_{\partial Q} \mathbf{v} \cdot N \quad (2.4)$$

in particular. Unfortunately,³ elementary estimation gives

$$\left| \int_{\partial Q} \mathbf{v} \cdot N \right| \leq \max_{\mathbf{x} \in U} \|\mathbf{v}\| \mathcal{H}^{n-1}(\partial Q)$$

and this quantity tends to zero as $Q \rightarrow \{\mathbf{p}\}$. Thus, the limit (2.4) does not recover a meaningful number. It is natural to consider next the average value

$$\frac{1}{\mathcal{H}^{n-1}(\partial Q)} \int_{\partial Q} \mathbf{v} \cdot N.$$

You should recall from the discussion above that \mathcal{H}^{n-1} , the $n - 1$ dimensional Hausdorff measure, is the natural measure for the domain of integration $\partial Q \subset \mathbb{R}^n$. Again, it can be shown that

$$\lim_{Q \rightarrow \{\mathbf{p}\}} \frac{1}{\mathcal{H}^{n-1}(\partial Q)} \int_{\partial Q} \mathbf{v} \cdot N = 0,$$

so this does not work. In this instance, the third time is the charm, and the strange quantity

$$\operatorname{div} \mathbf{v}(\mathbf{p}) = \lim_{Q \rightarrow \{\mathbf{p}\}} \frac{1}{\mu(Q)} \int_{\partial Q} \mathbf{v} \cdot N = 0, \quad (2.5)$$

where $\mu(Q)$ is the full dimension (Lebesgue) measure of the set Q gives a real number which can, at least in some instances capture the divergence of a vector field. In fact, this quantity $\operatorname{div} \mathbf{v} : U \rightarrow \mathbb{R}$ is called the **divergence** of the vector field \mathbf{v} .

Let me mention in passing that it is also quite possible (and perhaps natural) to define the divergence on other domains of integration where the concepts of “nicely

³Technically, this is not really unfortunate; it’s just the way it is.

shrinking” domains with boundaries providing domains of integration carry over. Such a definition might look like this

$$\operatorname{div}^A \mathbf{v}(\mathbf{p}) = \lim_{Q \rightarrow \mathbf{p}} \frac{1}{\mathcal{H}^k(Q)} \int_{\partial Q} \mathbf{v} \cdot \nu$$

where A is, for example, a k -dimensional submanifold of \mathbb{R}^n , e.g., a surface in \mathbb{R}^3 and Q is a k dimensional subset of A nicely shrinking to $\mathbf{p} \in A$. The normal ν here is typically called an **outward conormal** to ∂Q , and is nominally a bit more complicated: ν should be a tangent vector to A (or Q) that is normal to ∂Q . Of course, ν should point out of Q along ∂Q .

We won't have much practical use for this last “intrinsic” divergence except in the case of full dimension open subsets of \mathbb{R}^n considered first.

In rectangular coordinates, the divergence takes the familiar form

$$\operatorname{div} \mathbf{v} = \sum_{j=1}^n \frac{\partial v_j}{\partial x_j}.$$

It is now (perhaps) natural to consider the integral

$$\int_U \operatorname{div} \mathbf{v}. \tag{2.6}$$

If we partition U into small pieces, then the quantity in (2.6) is approximately

$$\sum_j \operatorname{div} \mathbf{v}(\mathbf{p}_j^*) \mu(U_j).$$

In fact, the integral in (2.6) is the limit of Riemann sums of this form. On the other hand, the quantity

$$\operatorname{div} \mathbf{v}(\mathbf{p}_j^*)$$

appearing in the Riemann sum is approximately

$$\frac{1}{\mu(U_j)} \int_{\partial U_j} \mathbf{v} \cdot N_j.$$

This requires that we imagine each of the pieces U_j as part of some sequence of subdomains Q shrinking nicely to \mathbf{p}_j^* , and that U_j is far enough along in that sequence so that the approximation is a good one. Technically, this procedure requires some

estimates which we will not do (at the moment at least). If these approximations are valid, i.e., if they can be backed up with estimates, then we are looking at a quantity like

$$\sum_j \operatorname{div} \mathbf{v}(\mathbf{p}_j^*) \mu(U_j) \approx \sum_j \int_{\partial U_j} \mathbf{v} \cdot N_j$$

where N_j is the outward unit cornormal along ∂U_j . In particular, setting $\mathcal{P} = \{U_j\}_{j=1}^k$, we suggest/claim

$$\begin{aligned} \int_U \operatorname{div} \mathbf{v} &= \lim_{\|\mathcal{P}\| \rightarrow 0} \sum_j \int_{\partial U_j} \mathbf{v} \cdot N_j \\ &= \lim_{\|\mathcal{P}\| \rightarrow 0} \int_{\partial U} \mathbf{v} \cdot N \\ &= \int_{\partial U} \mathbf{v} \cdot N. \end{aligned}$$

The interior boundary integrals cancel in pairs of adjacent regions U_i and U_j because $N_i = -N_j$ on the common boundary manifold. The last expression

$$\int_{\partial U} \mathbf{v} \cdot N$$

is independent of the partition, and this is how/why we obtain the limit. The assertion we are suggesting is called the **divergence theorem**:

$$\int_U \operatorname{div} \mathbf{v} = \int_{\partial U} \mathbf{v} \cdot N.$$

In two dimensions for $\Gamma = \partial U$ a simple closed (smooth) curve bounding an open set U , and even more generally for $\Gamma \subset \mathbb{R}^n$ a simple closed curve in any Euclidean/Cartesian space, the **circulation integral**

$$\int_{\Gamma} \mathbf{v} \cdot T,$$

where \mathbf{v} is a continuous vector field along Γ and T is a unit tangent field along Γ , makes sense. In the special case $\Gamma = \partial U$ and U a full dimension set in \mathbb{R}^2 , there is a natural relation between the circulation integral

$$\int_{\partial U} \mathbf{v} \cdot T$$

and the divergence theorem. It is precisely this: If T is taken as the counterclockwise normal on ∂U , then the unit outward normal satisfies $N^\perp = T$ or $-T^\perp = N$ where $\mathbf{v}^\perp = (-v_2, v_1)$ is the counterclockwise rotation by an angle $\pi/2$ of any vector \mathbf{v} . Thus we find

$$\int_{\partial U} \mathbf{v} \cdot T = \int_{\partial U} (-\mathbf{v}^\perp) \cdot N = \int_U \operatorname{div}(v_2, -v_1)$$

by the divergence theorem. That is, (in rectangular coordinates)

$$\int_{\partial U} \mathbf{v} \cdot T = \int_U \left(\frac{\partial v_2}{\partial x} - \frac{\partial v_1}{\partial y} \right).$$

This is called **Green's theorem**.

Stokes' theorem

$$\int_{\mathcal{S}} \operatorname{curl} \mathbf{v} \cdot N = \int_{\partial \mathcal{S}} \mathbf{v} \cdot T$$

where $\mathcal{S} \subset \mathbb{R}^3$ is a (two-dimensional) surface, T is the unit tangent vector field (counterclockwise with respect to the ambient surface normal N to \mathcal{S}) and

$$\operatorname{curl} \mathbf{v} = \left(\frac{\partial v_3}{\partial x_2} - \frac{\partial v_2}{\partial x_3}, - \left(\frac{\partial v_3}{\partial x_1} - \frac{\partial v_1}{\partial x_3} \right), \frac{\partial v_2}{\partial x_1} - \frac{\partial v_1}{\partial x_2} \right)$$

is also a manifestation of the divergence theorem taking the (usual) form

$$\int_{\mathcal{S}} \operatorname{div} \mathbf{v} = \int_{\partial \mathcal{S}} \mathbf{v} \cdot \nu. \quad (2.7)$$

It will be noted that the divergence $\operatorname{div} \mathbf{v} = \operatorname{div}^{\mathcal{S}} \mathbf{v}$ appearing in (2.7) is an **intrinsic surface divergence** which is not necessarily familiar from a formula, but has a good definition in the form

$$\operatorname{div}^{\mathcal{S}} \mathbf{v}(\mathbf{p}) = \lim_{Q \rightarrow \{\mathbf{p}\}} \frac{1}{\operatorname{area}(Q)} \int_{\partial Q} \mathbf{v} \cdot \nu$$

where ν denotes the unit outward conormal field (tangent to \mathcal{S}) along ∂Q , just as ν denoted the unit outward conormal field to $\partial \mathcal{S}$ in (2.7).

If you have trouble remembering the formula for the curl, it may be helpful (if you can remember how to take a cross-product of vectors in \mathbb{R}^3 to remember the “operator

cross-product form”

$$\operatorname{curl} \mathbf{v} = \begin{pmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \\ \frac{\partial}{\partial x_3} \end{pmatrix} \times \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}.$$

Incidentally, you’ll notice that the vector field \mathbf{v} considered on the surface \mathcal{S} in Stokes’ theorem is assumed to be defined in some open set of \mathbb{R}^3 containing \mathcal{S} . For such vector fields, there is also an intrinsic/coordinate free definition of the curl. It is the following: Given any unit vector $N \in \mathbb{R}^3$, let Q be a planar domain in a plane Π_N through \mathbf{p} with normal N . Assuming also that Q shrinks nicely to \mathbf{p} in Π_N , we define

$$\operatorname{curl} \mathbf{v}(\mathbf{p}) \cdot N = \lim_{Q \rightarrow \{\mathbf{p}\}} \frac{1}{\operatorname{area}(Q)} \int_{\partial Q} \mathbf{v} \cdot T.$$

It turns out that if you determine the dot product $\mathbf{w} \cdot N$ of a given vector $\mathbf{w} \in \mathbb{R}^3$ with every unit vector $N \in \mathbb{R}^3$, then you have determined \mathbf{w} uniquely. We are applying this fact to the vector field $\mathbf{w} = \operatorname{curl} \mathbf{v}$.

Chapter 3

PDE: A Quick and Dirty Intro

Let me recall the three fundamental partial differential equations of nineteenth century mathematical physics. These are **Laplace's equation**

$$\Delta u = \sum_{j=1}^n \frac{\partial^2 u}{\partial x_j^2} = 0,$$

The **heat equation**

$$u_t = \Delta u$$

and the **wave equation**

$$u_{tt} = \Delta u.$$

It has been said that the great discovery of nineteenth century physics¹ is that the fundamental laws of nature are linear...² Indeed, each of these three introductory PDE are linear. This means associated with each PDE there is an operator $L : V \rightarrow W$ with a natural (classical) domain V and codomain W of functions; these collections of functions are vector spaces, and

$$L(av + bw) = aLv + bLw \quad \text{for} \quad a, b \in \mathbb{R}, \quad u, v \in V. \quad (3.1)$$

The easiest is the equilibrium equation associated with the Laplace operator

$$Lu = \Delta u = \sum_{j=1}^n \frac{\partial^2 u}{\partial x_j^2}$$

¹In my view “physics” cannot “discover” anything. Only an individual human can discover something, but I'm just quoting, perhaps T.W. Körner, here to produce some kind of psychological feeling about the subject.

²I'll give the other half of this quote at the end of this chapter.

or oftentimes

$$Lu = -\Delta u = -\sum_{j=1}^n \frac{\partial^2 u}{\partial x_j^2}.$$

Either way we can take the Laplace operator $L : C^2(U) \rightarrow C^0(U)$ with $V = C^2(U)$ and $W = C^0(U)$, and observe that the **linearity condition** (3.1) holds. Sometimes it may be convenient to look for solutions of a boundary value problem for Laplace's equation $Lu = 0$ in the smaller vector subspace $V_0 = C^2(U) \cap C^0(\bar{U})$ or for weak solutions in a much larger vector space like $H^1(U)$, the collection of all square integrable measurable functions $u \in L^2(U)$ with first order weak derivatives in $L^2(U)$, but the classical domain $C^2(U)$ is convenient at least for a start.

Briefly, a **heat domain** is typically of the form $U \times [0, \infty)$ or more generally $U \times [0, T)$ for some $T > 0$ where U is an open subset of \mathbb{R}^n . Thus, the simplest (classical) domain and codomain for the heat operator $L : V \rightarrow W$ are $V = C^2(U \times (0, T))$ and $W = C^0(U \times (0, T))$ with

$$Lu = u_t - \Delta u.$$

More often it is natural to only require one continuous partial derivative with respect to time t and the existence and continuity of all second partial derivatives with respect to the spatial variables. This space has no standard notation, but is sometimes denoted by $C^{2,1}(U \times (0, T))$.

Similarly, the d'Alembertian or wave operator $\square : C^2((U \times (0, T))) \rightarrow C^0(U \times (0, T))$ by

$$\square u = \frac{\partial^2 u}{\partial t^2} - \Delta u.$$

The primary consequence of linearity for these operators is called the **principle of superposition** which is discussed in each case below. For each operator, and hence each equation, we will discuss natural boundary and/or initial conditions along with some comments which may be viewed as falling roughly into three categories

1. Derivation (or at least some comments about some kind of derivation)
2. Separated variables solutions (or Fourier series solutions on some special domains)
3. Integral identities and related properties, especially energy considerations.

3.1 Laplace's equation

The natural boundary value problem(s) for Laplacian involve some combination of prescribing the value of a solution $u \in C^2(U) \cap C^0(\bar{U})$ on the boundary ∂U of the domain U and prescribing the (outward unit) normal derivative of u along ∂U . Separately,

$$\begin{cases} \Delta u = 0, & \text{on } U \\ u|_{\partial U} = g \end{cases} \quad (3.2)$$

is called the **Dirichlet problem** for Laplace's equation and

$$\begin{cases} \Delta u = 0, & \text{on } U \\ (Du \cdot \nu)|_{\partial U} = h \end{cases} \quad (3.3)$$

is called the **Neumann problem** for Laplace's equation. More generally, the boundary ∂U may be partitioned with $\partial U = A \cup B$ and a Dirichlet condition

$$u|_A = g$$

can be prescribed on A while a Neumann condition

$$(Du \cdot \nu)|_B = h$$

can be prescribed on B . Roughly speaking, one would first like to assert the existence and uniqueness of a solution of the Dirichlet problem (3.2) when, for example, U is a bounded open subset of \mathbb{R}^n , ∂U is a C^1 $n - 1$ dimensional submanifold, and $g \in C^0(\partial U)$. (Think when $n = 2$ of the situation when ∂U is a simple closed curve.) This turns out to be a relatively difficult proposition, but it turns out to be true. Perhaps we can (and will) make some simple observations related to this assertion. One is that the problem is quite closely related to the boundary value problem for **Poisson's equation**

$$\begin{cases} \Delta v = f, & \text{on } U \\ v|_{\partial U} = 0 \end{cases} \quad (3.4)$$

with **homogeneous boundary values**. The reason is quite simple: Let us assume we can find a function $G \in C^2(U) \cap C^0(\bar{U})$ satisfying

$$G|_{\partial U} \equiv g.$$

Then given a solution $u \in C^2(U) \cap C^0(\overline{U})$ of (3.2) the function $v = u - G$ satisfies

$$\begin{cases} \Delta v = -\Delta G, & \text{on } U \\ v|_{\partial U} = 0 \end{cases}$$

which is clearly a version of (3.4).

Exercise 3.1 Assume $g \in C^2(\partial B_1(0,0))$ where $B_1(0,0)$ is the open unit disk in \mathbb{R}^2 . Find an extension $G \in C^2(\overline{B_1(0,0)})$ on the entire closed unit disk with restriction to $\partial B_1(0,0)$ given by g . Hint: You should think carefully about what it means to have $g \in C^2(\partial B_1(0,0))$.

Exercise 3.2 Assume $g \in C^2(\partial B_1(0,0))$ as in Exercise 3.1 above. Formulate a Dirichlet problem for Poisson's equation with homogeneous boundary values which, if you can solve it, allows you to solve the boundary value problem (3.2) for Laplace's equation.

The conclusion you should observe here is that existence and uniqueness for the Dirichlet problem (3.2) for Laplace's is equivalent to existence and uniqueness for the Dirichlet problem (3.4) for Poisson's equation.

Two final comments about existence and uniqueness: The first is that we are not really going to address much in detail about existence and uniqueness for the Dirichlet problem for Laplace's equation, but this question is a very important one, and it is a difficult one. It would be very easy to spend an entire semester discussing the existence and uniqueness of solutions for the Dirichlet problem for Laplace's equation. The second observation is that there is no uniqueness for the Neumann problem—actually, that's not quite true, but if $u_1 \in C^1(\overline{U})$ is a solution of (3.3), then every function $u \in C^1(\overline{U})$ with $u(\mathbf{x}) = u_1(\mathbf{x}) + c$ and $c \in \mathbb{R}$ constant is also a solution of (3.3). It does turn out that solutions (on a smooth bounded domain) are unique up to a constant.

Exercise 3.3 Let U be a bounded domain with C^1 boundary. Prove a solution $u \in C^1(\overline{U})$ of (3.3) with $h \equiv 0$ satisfies $u \equiv c$ is constant. Hint: Integrate $\operatorname{div}(uDu)$.

Exercise 3.4 Let U be a bounded domain with C^1 boundary. Prove a solution $u \in C^1(\overline{U})$ of (3.3) is unique up to an additive constant.

3.1.1 Derivation

I will discuss two topics/observations concerning Laplace's equation, each of which might be considered a "derivation." Each has its drawbacks when viewed as such.

The first observation is that solutions of the Dirichlet problem for Laplace's equation may be viewed as the equilibrium (i.e., long time steady state solutions) associated with a certain heat evolution problem. It is interesting to note that this was put forward by Dirichlet as a proof of existence and uniqueness for the Dirichlet problem for Laplace's equation, though such an approach is not considered particularly insightful, or at least not very mathematically rigorous/satisfying these days.

Consider the initial/boundary value problem

$$\begin{cases} v_t = \Delta v, & \text{on } U \times [0, \infty) \\ v(\mathbf{x}, 0) = v_0(\mathbf{x}), & \mathbf{x} \in U \\ v(\mathbf{x}, t) = g(\mathbf{x}), & \mathbf{x} \in \partial U, t > 0. \end{cases}$$

"Obviously," said Dirichlet, "this problem has a solution and that solution will reach a long time steady state"

$$u(\mathbf{x}) = \lim_{t \nearrow \infty} v(\mathbf{x}, t).$$

That limiting function will be independent of the initial temperature distribution v_0 , and the function u will be independent of time

$$\frac{\partial u}{\partial t} = 0.$$

Thus, commuting the derivatives and assuming the second partial derivatives with respect to the spatial variables converge we conclude

$$\begin{aligned} 0 &= \lim_{t \nearrow \infty} \frac{\partial v}{\partial t} \\ &= \lim_{t \nearrow \infty} (-\Delta v) \\ &= -\Delta \left(\lim_{t \nearrow \infty} v \right) \\ &= -\Delta u \end{aligned}$$

while for $\mathbf{x} \in \partial U$,

$$\begin{aligned} g(\mathbf{x}) &= v(\mathbf{x}, t) \\ &= \lim_{t \nearrow \infty} v(\mathbf{x}, t) \\ &= u(\mathbf{x}). \end{aligned}$$

Obviously, there is a lot to justify mathematically, or more properly a lot we have not mathematically justified, chiefly the existence (and uniqueness?) of solutions for the heat evolution problem. Nevertheless, Dirichlet has suggested a heuristic it is perhaps useful to keep in mind: Physically, solutions of the Dirichlet problem for Laplace's equation model equilibrium solutions for an initial/boundary value problem for the heat equation, and the long time steady state of such a problem "should" be independent of the initial temperature distribution.

Differentiating in infinite dimensions

Our second "derivation" was provided by Lagrange. It is much more interesting, in some sense. Lagrange starts with the assumption that the real valued **functional**³ $\mathcal{E} : C^2(U) \cap C^1(\bar{U}) \rightarrow \mathbb{R}$ given by

$$\mathcal{E}[u] = \int_U |Du|^2$$

gives a kind of "energy" measuring the irregularity,⁴ in the vague sense of "variability," of the temperature distribution u on U . More precisely, $\mathcal{E} : C^2(U) \cap C^1(\bar{U}) \rightarrow [0, \infty)$ is a non-negative functional and, if restricted to the **admissible class**

$$\mathcal{A} = \left\{ u \in C^2(U) \cap C^1(\bar{U}) : u|_{\partial U} = g \right\},$$

we can look for a **minimizer**, that is a function $u_0 \in \mathcal{A}$ satisfying

$$\mathcal{E}[u_0] \leq \mathcal{E}[u] \quad \text{for every } u \in \mathcal{A}.$$

If the domain of a real valued function \mathcal{E} happened to be an open subset A of a (finite dimensional) Euclidean space \mathbb{R}^n , then we would know a systematic method to find (at least interior) minimizers: We would take the derivative $\mathcal{E}' : A \rightarrow \mathbb{R}$ and try to find a point $a \in A$ with $\mathcal{E}'(a) = 0$. Of course, not so much of this makes sense for $\mathcal{E} : \mathcal{A} \rightarrow \mathbb{R}$ where \mathcal{A} is a subset (open?, does open even make sense?) of an infinite dimensional vector space of functions like $C^2(U) \cap C^1(\bar{U})$. Nevertheless, Lagrange realized that something (interesting) can be done:

³A functional is simply a real valued function defined on a collection, very often as in this case a vector space, of functions.

⁴Incidentally, this functional \mathcal{E} is called the **Dirichlet energy** of the function u . Lagrange's insight was that, roughly speaking, solutions of Laplace's equation are minimizers of the Dirichlet energy.

If $u_0 \in \mathcal{A}$, and $\phi \in C_c^\infty(U)$, then $u_0 + t\phi \in \mathcal{A}$ for every $t \in \mathbb{R}$. This means $f : \mathbb{R} \rightarrow \mathbb{R}$ by

$$f(t) = \mathcal{E}[u_0 + t\phi],$$

for u_0 and ϕ (temporarily) fixed, is a real valued function of one real variable. This function is not so unlike the functions considered in first semester calculus, and if u_0 is a minimizer of \mathcal{E} , then $t = 0$ is a minimum point of f . This means, from Calc I, that if f has a derivative, then

$$f'(0) = 0.$$

In this way was born what is called the **first variation** of a functional \mathcal{E} :

$$f'(0) = \left(\frac{d}{dt} \mathcal{E}[u_0 + t\phi] \right) \Big|_{t=0}. \quad (3.5)$$

Naturally, there are lots of interesting and important questions we should ask about what exactly is the quantity defined in (3.5). At the top of the list: What kind of derivative is this? And indeed, we should ask some of those questions sooner or later, but let me suggest (with Lagrange that) we make a computation first:

$$f'(0) = \left(\frac{d}{dt} \int_U |D(u_0 + t\phi)|^2 \right) \Big|_{t=0}.$$

Note that the integral and the gradient in the integrand are taken with respect to the spatial variables, but this parameter t is not among those spatial variables. Therefore, first of all,

$$D(u_0 + t\phi) = Du_0 + tD\phi.$$

Furthermore, we know

$$|D(u_0 + t\phi)|^2 = D(u_0 + t\phi) \cdot D(u_0 + t\phi)$$

is given by a dot product, so

$$|D(u_0 + t\phi)|^2 = |Du_0|^2 + 2(Du_0 \cdot D\phi) t + |D\phi|^2 t^2.$$

The integrand is quadratic in t . If we differentiate under the integral sign, which can be mathematically justified if the dependence on t is relatively nice, which it is, then

$$\begin{aligned} f'(0) &= \left(\frac{d}{dt} \int_U [2(Du_0 \cdot D\phi) + 2|D\phi|^2 t] \right) \Big|_{t=0} \\ &= 2 \int_U Du_0 \cdot D\phi. \end{aligned}$$

Now we need to “integrate by parts” in several variables, which amounts to using the divergence theorem along with a product rule for the divergence operator. Specifically,

$$\operatorname{div}(\phi Du_0) = D\phi \cdot Du_0 + \phi \Delta u_0.$$

Therefore,

$$\begin{aligned} \int_U Du_0 \cdot D\phi &= \int_U \operatorname{div}(\phi Du_0) - \int_U \phi \Delta u_0 \\ &= \int_{\partial U} \phi Du_0 \cdot n - \int_U \phi \Delta u_0 \\ &= - \int_U \phi \Delta u_0 \end{aligned}$$

because $\phi \in C_c^\infty(U)$ and

$$\phi|_{\partial U} \equiv 0.$$

Putting this together with our computation of the derivative $f'(0)$ and the necessary condition $f'(0) = 0$ for a minimizer, we conclude:

$$\int_U \phi \Delta u_0 = 0 \quad \text{for every } \phi \in C_c^\infty(U). \quad (3.6)$$

So, we should ask ourselves at this point, what does this tell us about the minimizer u_0 ?

Two important principles of integration

We will use the following result in the derivation of the heat equation below:

Theorem 10 *If U is an open subset of \mathbb{R}^n and $f \in C^0(U)$ satisfies*

$$\int_V f = 0 \quad \text{for every open set } V \subset U,$$

then $f(\mathbf{x}) = 0$ for $\mathbf{x} \in U$.

Proof: Assume $f(\mathbf{x}_0) \neq 0$ for some $\mathbf{x}_0 \in U$. Then either $f(\mathbf{x}_0) > 0$ or $f(\mathbf{x}_0) < 0$. Take the case $f(\mathbf{x}_0) > 0$. The other case is similar. By continuity, there is some $\delta > 0$

for which $V = B_\delta(\mathbf{x}_0) \subset U$ and $|f(\mathbf{x}) - f(\mathbf{x}_0)| < f(\mathbf{x}_0)/2$ for $\mathbf{x} \in B_\delta(\mathbf{x}_0)$. It follows that

$$f(\mathbf{x}) > \frac{f(\mathbf{x}_0)}{2} > 0 \quad \text{for } \mathbf{x} \in B_\delta(\mathbf{x}_0).$$

In fact,

$$\begin{aligned} \frac{f(\mathbf{x}_0)}{2} &= f(\mathbf{x}_0) - \frac{f(\mathbf{x}_0)}{2} + f(\mathbf{x}) - f(\mathbf{x}) \\ &\leq |f(\mathbf{x}) - f(\mathbf{x}_0)| + f(\mathbf{x}) - \frac{f(\mathbf{x}_0)}{2} \\ &< \frac{f(\mathbf{x}_0)}{2} + |f(\mathbf{x})| - \frac{f(\mathbf{x}_0)}{2} \\ &= f(\mathbf{x}). \end{aligned}$$

Therefore, taking $V = B_\delta(\mathbf{x}_0)$ we find

$$\begin{aligned} \int_V f &\geq \int_V \frac{f(\mathbf{x}_0)}{2} \\ &= \frac{f(\mathbf{x}_0)}{2} \int_V 1 \\ &= \frac{f(\mathbf{x}_0)}{2} \mu(B_\delta(\mathbf{x}_0)) \\ &> 0. \end{aligned}$$

This is a contradiction of the hypothesis

$$\int_V f = 0. \quad \square$$

Our second result, which is proved much the same way, is called the **fundamental lemma of the calculus of variations**, though perhaps a better name is the **fundamental lemma of vanishing integrals**.

Theorem 11 *If U is an open subset of \mathbb{R}^n and $f \in C^0(U)$ satisfies*

$$\int f \phi = 0 \quad \text{for every } \phi \in C_c^\infty(U),$$

then $f(\mathbf{x}) = 0$ for $\mathbf{x} \in U$.

Proof: Assume $f(\mathbf{x}_0) \neq 0$. Then either $f(\mathbf{x}_0) > 0$ or $f(\mathbf{x}_0) < 0$. This time let's take the case $f(\mathbf{x}_0) < 0$. By continuity, there is some $\delta > 0$ with $B_\delta(\mathbf{x}_0) \subset U$ and

$$u(\mathbf{x}) < \frac{u(\mathbf{x}_0)}{2} < 0 \quad \text{for } \mathbf{x} \in B_\delta(\mathbf{x}_0).$$

Now we take a function $\phi \in C_c^\infty(U)$ with $\phi \geq 0$ and $\text{supp}(\phi) = \overline{B_\delta(\mathbf{x}_0)}$. For example, we can take the standard test function with value

$$\mu_0(\mathbf{x}) = \begin{cases} e^{-1/|\mathbf{x}|^2}, & \mathbf{x} \in B_1(\mathbf{0}) \\ 0, & \mathbf{x} \notin B_1(\mathbf{0}) \end{cases}$$

and then take ϕ given by

$$\phi(\mathbf{x}) = \mu_0\left(\frac{\mathbf{x} - \mathbf{x}_0}{\delta}\right).$$

For this choice of ϕ we find

$$\begin{aligned} \int u\phi &= \int_{B_\delta(\mathbf{x}_0)} u\phi \\ &\leq \frac{u(\mathbf{x}_0)}{2} \int_{B_\delta(\mathbf{x}_0)} \phi \\ &< 0. \end{aligned}$$

This is a contradiction of the hypotheses of the lemma. \square

Applying the fundamental lemma (the second theorem above) to (3.6) we find it is a necessary condition for $u \in C^2(U) \cap C^1(\overline{U})$ to minimize the Dirichlet energy that u satisfy Laplace's equation on U :

$$\Delta u = 0. \tag{3.7}$$

We will give a third, and perhaps somewhat more compelling, derivation of Laplace's equation (3.7) in connection with our derivation of the heat equation below. There are several other mathematical "derivations" of Laplace's equation, that is to say, there are problems of mathematical (and often practical) interest to which solutions of Laplace's equation are intimately related. We briefly mention one topic falling into this category.

Conformal mapping

Given an open subset U of the complex plane \mathbb{C} , we say U is a conformal image of the open unit disk $D = \{z \in \mathbb{C} : |z| < 1\}$ if there exists a complex differentiable function $f : D \rightarrow U$ such that f is one-to-one and onto. Such functions are very often used on the making of maps depicting the geographic features of the earth. Such maps preserve angles, though they may scale by different amounts at different points. This is why, for example, Greenland appears to be much larger than South America in a Mercator projection of the land masses on the surface of the earth. Given a complex differentiable function $f : D \rightarrow U$, we can write $f = u + iv$ where $u = u(x, y) = \operatorname{Re}(f)$ and $v = v(x, y) = \operatorname{Im}(f)$ and $z = x + iy$ is the complex independent variable. In this case, u and v satisfy the Cauchy-Riemann equations, which is a system of partial differential equations:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \quad \text{and} \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}.$$

It is easy to check that u and v are solutions of Laplace's equation. Given a harmonic function $u = u(x, y)$, satisfying $\Delta u = 0$, a function v is said to be a **harmonic conjugate** for u if the complex valued function $f = u + iv$ is complex differentiable, or in other words conformal.

Note: When we say a function $f : D \rightarrow U$ is **complex differentiable** here we mean the limit

$$f'(z) = \lim_{\zeta \rightarrow z} \frac{f(\zeta) - f(z)}{\zeta - z} \tag{3.8}$$

exists at every point $z \in D$. Such a function has many interesting properties:

1. If the first complex derivative⁵ exists, then the second complex derivative f'' and all subsequent derivatives exist.
2. The function f is locally representable at each $z_0 \in D$ by a power series

$$f(z) = \sum_{j=0}^{\infty} \frac{f^{(j)}(z_0)}{j!} (z - z_0)^j.$$

⁵This limit of a difference quotient (3.8) looks superficially like some kind of real derivative, but this is very much not the case. The fact that the limit is taken as the variable ζ takes values in the complex plane \mathbb{C} makes this kind of derivative fundamentally different from any real derivative.

This condition (local power series representation) is equivalent to complex differentiability on an open set, and when this property is emphasized one says the function f is **analytic** or **complex analytic** in contrast to a function of several real variables which may (or may not be) real analytic, i.e., locally represented by a real power series.

3. The function f considered as geometrically mapping the domain D to the domain U preserves angles at points where $f'(z) \neq 0$. The value of the derivative $f'(z_0)$ is called the **conformal factor** at z_0 , and given a curve $\gamma : (-1, 1) \rightarrow D$ with $\gamma(0) = z_0$ there holds, by the chain rule,

$$\frac{d}{dt}(f \circ \gamma)(0) = f'(z_0) \frac{d}{dt} \gamma(0).$$

This corresponds to a local scaling by length, in every direction, by a factor $|f'(z_0)|$ and a rotation determined by the argument θ of $f'(z_0) = |f'(z_0)|e^{i\theta}$. To emphasize this property one can say the function is **conformal** or is a **conformal mapping**. Conversely, if a function $f : D \rightarrow U$ preserves angles, then the function will be complex differentiable.

3.1.2 Separated variables solutions

Let U be a bounded open connected subset of \mathbb{R}^n with ∂U a smooth, say C^2 hypersurface, i.e., $n - 1$ -dimensional submanifold. Then we have the following:

Theorem 12 (existence and uniqueness for the Dirichlet problem for Laplace's equation) Given any $g \in C^2(\partial U)$, there exists a unique solution $u \in C^2(U) \cap C^0(\partial U)$ of the Dirichlet boundary value problem

$$\begin{cases} \Delta u = 0, & \text{on } U \\ u|_{\partial U} = g, \end{cases}$$

We will not prove this theorem here, but it is good to know about it. Another result that is similarly important to know is the following:

Theorem 13 (interior regularity for solutions of Laplace's equation) If $u \in C^2(U)$ and $\Delta u = 0$, then $u \in C^\infty(U)$. In fact, u is real analytic, i.e., $u \in C^\omega(U)$.

The real analyticity, local power series representation, of a harmonic function is closely related to the complex analyticity of a complex differentiable function. See the discussion in the "derivation" section above. We will not prove the regularity result

either, though the integration techniques, the mean value property in particular, discussed in the third section below will take us quite close to a proof. In certain cases, the integration techniques may also be extended to prove existence and uniqueness, though we will not discuss this directly. If you are interested, look up the **Poisson integral formula for harmonic functions**.

Here we will consider certain special domains. In particular, we will consider a rectangle in \mathbb{R}^2 . This is not a domain with C^2 boundary, but in certain cases existence and uniqueness does hold. Our interest is in computing formulas for solutions and formulas involving (Fourier) series in particular. If we take finitely many terms in a series solution, then we can think of our objective as approximating a solution.

Fourier series representation

A Fourier series for a periodic function $f : \mathbb{R} \rightarrow \mathbb{R}$ with period 2π , i.e., $f(x + 2\pi) = f(x)$ for all $x \in \mathbb{R}$ has the form

$$\sum_{j=0}^{\infty} a_j \cos(jx) + \sum_{j=1}^{\infty} \sin(jx).$$

Notice $j = 0$ gives a constant term a_0 . You can think of such a series as roughly comparable to a power series. As we have written things so far, there is no intimate connection between the function f and the series. The only connection is that they are both 2π periodic functions. This is the only thing we have used about the function f to write down the “form” of the series, but we did use the fact that f is 2π periodic in this way. The idea is to find the coefficients, so that the series represents a given function f . The first sum (if convergent) is even, and the second sum (if convergent) is odd. We will not concern ourselves greatly with convergence, but if $f \in L^2(-\pi, \pi)$, i.e., f is measurable and

$$\int_{(-\pi, \pi)} |f|^2 < \infty,$$

then the series will converge at some set of points G satisfying

$$\mu(G \cap (-\pi, \pi)) = 2\pi,$$

and it makes sense (if one explains what one means) to write

$$f(x) = \sum_{j=0}^{\infty} a_j \cos(jx) + \sum_{j=1}^{\infty} \sin(jx). \quad (3.9)$$

In general what this means is the following: There exists a well-defined function $g : G \rightarrow \mathbb{R}$ given by

$$g(x) = \sum_{j=0}^{\infty} a_j \cos(jx) + \sum_{j=1}^{\infty} \sin(jx),$$

and setting

$$g_0(x) = \begin{cases} g(x), & x \in G \\ 0, & x \notin G, \end{cases}$$

there holds

$$\int_{(-\pi, \pi)} |g_0 - f|^2 = 0.$$

If $f \in C^1(\mathbb{R})$, then one can say the series converges at every point and (3.9) holds classically, i.e., pointwise at every $x \in \mathbb{R}$. Of primary initial interest to us is the determination of the coefficients a_j and b_j .

Fourier series can also sometimes be discussed for 2π periodic function $f : \mathbb{R} \rightarrow \mathbb{C}$ in the (unified) form

$$f(t) \sim \sum_{j=0}^{\infty} \alpha_j e^{it}$$

where the coefficients are also complex.

Fourier sine series

For simplicity, we will restrict attention to a slightly different framework than the one just described, but the approach can be applied more generally. Instead of focusing on the symmetric interval $(-\pi, \pi)$ directly, we will consider the interval $(0, \pi)$ and the situation where $a_j = 0$ for all j . In this way, we are effectively assuming f is an odd function. As we shall see, Fourier series are quite robust in their behavior, and this turns out to be less of a restriction than one might think.

Here is the main point: If we assume a representation

$$f(x) = \sum_{j=1}^{\infty} b_j \sin(jx)$$

and whatever convergence might be convenient for us, then multiplying both sides of the representation formula by $\sin(kx)$, which is called a single Fourier sine **mode**, we

have

$$f(x) \sin(kx) = \sum_{j=1}^{\infty} b_j \sin(jx) \sin(kx).$$

Integrating both sides

$$\int_{(0,\pi)} f(x) \sin(kx) = \sum_{j=1}^{\infty} b_j \int_{(0,\pi)} \sin(jx) \sin(kx). \quad (3.10)$$

There holds

$$\int_{(0,\pi)} \sin(jx) \sin(kx) = \int_0^{\pi} \sin(jx) \sin(kx) dx = \begin{cases} 0, & j \neq k \\ \pi/2, & j = k. \end{cases}$$

Thus, rearranging the relation (3.10) we find

$$\frac{\pi}{2} b_k = \int_{(0,\pi)} f(x) \sin(kx),$$

or

$$b_k = \frac{2}{\pi} \int_{(0,\pi)} f(x) \sin(kx). \quad (3.11)$$

This is a formula for the coefficients because the integral(s) on the right are simply numbers determined by the function f . In specific cases, these numbers can be computed. In this case the coefficients b_k are called the **Fourier sine coefficients** of f .

As an example, let's consider the function

$$f(x) = \frac{\pi^2}{4} - \left(x - \frac{\pi}{2}\right)^2.$$

According to (3.11)

$$\begin{aligned} b_k &= \frac{2}{\pi} \int_0^{\pi} \left[\frac{\pi^2}{4} - \left(x - \frac{\pi}{2}\right)^2 \right] \sin(kx) dx \\ &= \frac{\pi}{2} \int_0^{\pi} \sin(kx) dx - \frac{2}{\pi} \int_0^{\pi} \left(x - \frac{\pi}{2}\right)^2 \sin(kx) dx. \end{aligned}$$

First we note that

$$\int_0^{\pi} \sin(kx) dx = -\frac{1}{k} \cos(kx) \Big|_{x=0}^{\pi} = \begin{cases} 2/k, & k \text{ odd} \\ 0, & k \text{ even.} \end{cases}$$

Next we integrate by parts:

$$\begin{aligned}
 \int_0^\pi \left(x - \frac{\pi}{2}\right)^2 \sin(kx) dx &= -\frac{1}{k} \left(x - \frac{\pi}{2}\right)^2 \cos(kx) \Big|_{x=0}^\pi + \frac{2}{k} \int_0^\pi \left(x - \frac{\pi}{2}\right) \cos(kx) dx \\
 &= \frac{2}{k} \left[\frac{\pi^2}{8} [1 - (-1)^k] - \frac{1}{k} \sin(kx) \Big|_{x=0}^\pi + \frac{1}{k} \int_0^\pi \sin(kx) dx \right] \\
 &= \frac{2}{k} \left[\frac{\pi^2}{8} [1 - (-1)^k] - \frac{1}{k^2} \cos(kx) \Big|_{x=0}^\pi \right] \\
 &= \frac{2}{k} \left[\frac{\pi^2}{8} - \frac{1}{k^2} \right] [1 - (-1)^k].
 \end{aligned}$$

Thus,

$$\int_0^\pi \left(x - \frac{\pi}{2}\right)^2 \sin(kx) dx = \begin{cases} (1/k)[\pi^2/2 - 4/k^2], & k \text{ odd} \\ 0, & k \text{ even,} \end{cases}$$

and

$$b_k = \begin{cases} 8/(\pi k^3), & k \text{ odd} \\ 0, & k \text{ even.} \end{cases}$$

We have not shown it, but in fact it is true, that for each $x \in [0, \pi]$ there holds

$$\frac{\pi^2}{4} - \left(x - \frac{\pi}{2}\right)^2 = \sum_{k=0}^{\infty} \frac{8}{\pi(2k+1)^3} \sin[(2k+1)x]. \quad (3.12)$$

Perhaps a more practically interesting consideration for you might be the convergence of the partial sums implied by (3.12). The first term for example, corresponding to $k = 0$ in the sum, is $f_1(x) = (8/\pi) \sin x$. I have plotted the functions f , f_1 and P_1 and Q_1 the first order Taylor expansions of f at $x = 0$ and $x = \pi/2$ respectively; see Figure 3.1.

It should be absolutely no surprise that the first term of the Fourier expansion does not match the quadratic function f exactly, simply because the Fourier approximation is sinusoidal. What may be slightly surprising is how well the single Fourier term matches the quadratic function globally of the interval of interested $0 \leq x \leq \pi$. We can get a quantitative measure of how well f_1 approximates f by plotting the difference $f_1 - f$. In fact, adding the second term of the Fourier expansion to obtain $f_3 : [0, \pi] \rightarrow \mathbb{R}$ by

$$f_3(x) = \frac{8}{\pi} \left[\sin x - \frac{1}{27} \sin(3x) \right]$$

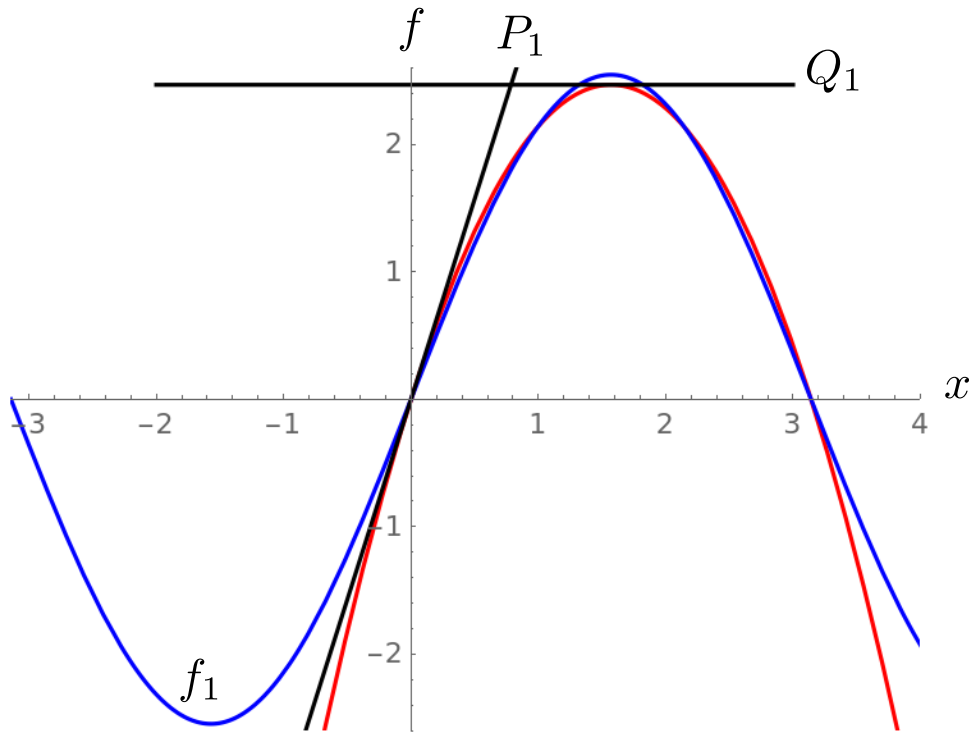


Figure 3.1: The function $f : [0, \pi] \rightarrow \mathbb{R}$ by $f(x) = \pi^2/4 - (x - \pi/2)^2$ is plotted in red along with the natural extension outside the interval of interest. Various approximations are plotted as well: The first term of the Fourier sine expansion f_1 is plotted in blue. Two Taylor polynomial approximations of first order are plotted in black. It will be noted that the Fourier series is a kind of “global” approximation as opposed to the Taylor approximation which is inherently local. (Note well that the Fourier coefficients are determined by an integral while the Taylor coefficients are determined by a derivative (at a point)). This can have some significant advantages.

gives an approximation globally/uniformly within about 0.03 of f on an interval of length about 3, that is, the approximation is within about 1/100 of the length of the domain of the function. (If the functions f and f_3 are plotted at a reasonable scale, the two are essentially indistinguishable on the interval of interest.) In Figure 3.2 we have plotted the differences $\delta_j = f_j - f$ for f_j corresponding to $k = 0, \dots, 3$.

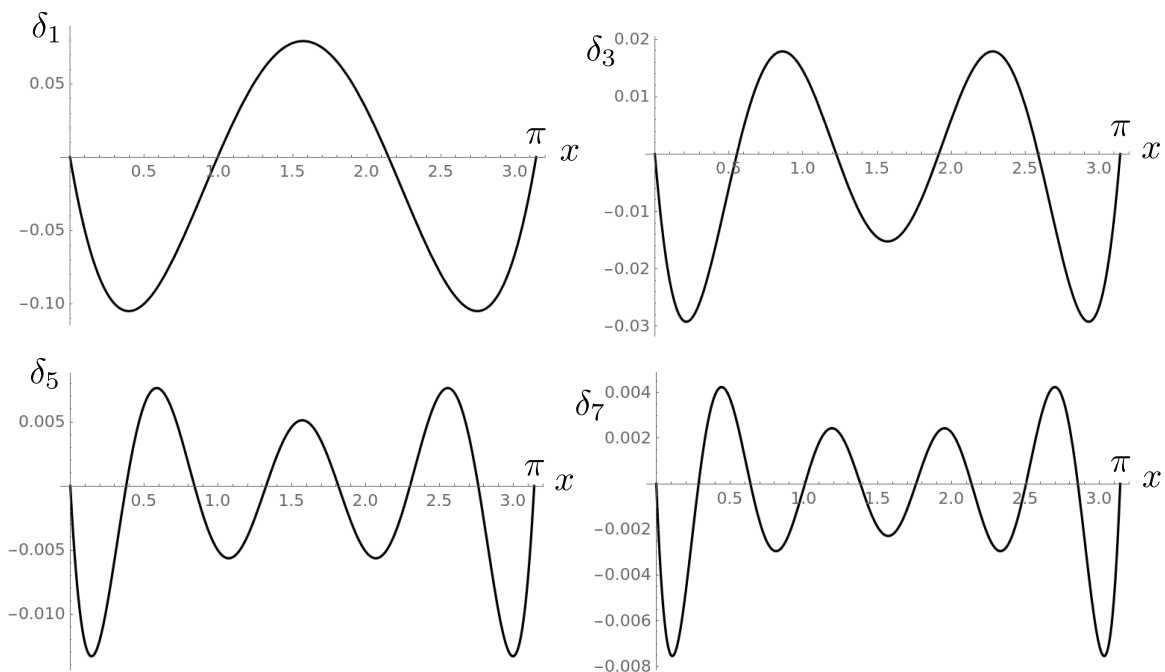


Figure 3.2: The signed pointwise error $\delta_j = f_j - f$ of the Fourier expansion with $k+1$ terms and $j = 2k + 1$ for $k = 0, 1, 2, 3$.

Solving the PDE

Now we apply the discussion of the previous section to solve the boundary value problem

$$\begin{cases} \Delta u = 0, & \text{on } U \\ u(x, \pi) = g(x), & x \in (0, \pi) \\ u(\mathbf{x}) = 0, & \mathbf{x} \in \partial U \setminus \Gamma \end{cases} \quad (3.13)$$

where $U = (0, \pi) \times (0, \pi)$ is an open domain in \mathbb{R}^2 with boundary a square and boundary component $\Gamma = \{(x, \pi) : x \in (0, \pi)\}$, and $g : [0, \pi] \rightarrow \mathbb{R}$ by

$$g(x) = \pi^2/4 - (x - \pi/2)^2$$

is the example function considered (and called f) above. We seek first separated variables solutions having the form $u(x, y) = A(x)B(y)$. Substituting into Laplace's equation we find

$$A''B + AB'' = 0$$

where the derivatives on A are with respect to x and the derivatives on B are with respect to y of course. At points (x, y) where A and B are nonzero we can write

$$\frac{A''}{A} = -\frac{B''}{B}. \quad (3.14)$$

Furthermore, at points y where $B(y) \neq 0$ the boundary condition gives

$$A(0) = A(\pi) = 0.$$

Differentiating (3.14) with respect to x , we find

$$\frac{d}{dx} \left(\frac{A''}{A} \right) = 0.$$

This means A''/A is constant. Let us call the constant λ and focus on the “problem”

$$\begin{cases} A'' = \lambda A, & x \in (0, \pi) \\ A(0) = 0 = A(\pi). \end{cases} \quad (3.15)$$

This is a somewhat different kind of mathematical problem than you may have encountered before. It has a special name: This kind of problem is called a **Sturm-Liouville problem**. Superficially, the Sturm-Liouville problem looks a little bit like an ordinary differential equation. It has an ordinary derivative of second order A'' in it. On the other hand, there are two things that make it different. First the boundary condition is rather unlike the standard initial condition encountered in the elementary study of second order ordinary differential equations, namely $A(0) = a_0$ and $A'(0) = b_0$. The condition we actually have here $A(0) = 0$ and $A'(\pi) = 0$ is called a **two point boundary condition**, and if λ is a given constant, then the problem (3.15) would be a **two point boundary value problem** which is usually studied in, for example, a second course on ODEs or a course in applied mathematics like this one.⁶ The really big difference, however, is that λ is not a given constant, but finding the value(s) of λ for which a solution A of the corresponding two point boundary value problem exists is included as a part of the Sturm-Liouville problem. As a consequence, we must consider cases.

Let us consider the two point boundary value problem (3.15) for a given, fixed, constant $\lambda < 0$. Again, note this may be a slightly different kind of ODE problem than you have considered before, but you can do it, and maybe you have solved such

⁶How appropriate!

a problem before. At any rate what you should know is that the general solution of $A'' = \lambda A$ when $\lambda < 0$ is $A(x) = a \cos(\mu x) + b \sin(\mu x)$ where a and b are some constants, $\mu^2 = -\lambda$, and we can take $\mu > 0$ for simplicity. Having made this observation we may turn to the boundary conditions which require

$$a = 0 \quad \text{and} \quad a \cos(\mu\pi) + b \sin(\mu\pi) = 0.$$

In view of the fact that $a = 0$, we find $b \sin(\mu\pi) = 0$. Now there are two possibilities: Either $a = 0$ or $\sin(\mu\pi) = 0$. The first possibility is not interesting because we get from it only the trivial solution $A(x) \equiv 0$. Turning to the second possibility, we get interesting solutions for $\mu = \mu_j = j$ and $j = 1, 2, 3, \dots$. These correspond to the sequence

$$\lambda = \lambda_j = -j^2 \quad \text{for} \quad j = 1, 2, 3, \dots$$

You may note/recall that the symbol λ was used extensively in linear algebra to denote **eigenvalues**. In that case, there was invariably a linear function $L : \mathbb{R}^n \rightarrow \mathbb{R}^n$ defined on a finite dimensional vector space for which there was a nonzero vector $\mathbf{v} \in \mathbb{R}^n$ with $L(\mathbf{v}) = \lambda \mathbf{v}$. The same dynamic is at work here, but the operator is the ordinary differential operator

$$\frac{d^2}{dx^2} : C^\infty(0, \pi) \rightarrow C^\infty(0, \pi)$$

defined on the infinite dimensional vector space $C^\infty(0, \pi)$. Nevertheless, it is the case for the (nonzero) functions

$$A_j(x) = \sin(jx),$$

that

$$\frac{d^2}{dx^2} A_j = A_j'' = \lambda_j A_j.$$

Thus, A_j is the **eigenvector**. (Think about that.) Also, the sequence of values $\{\lambda_j\}_{j=1}^\infty$ with $\lambda_j = -j^2$ are called the **Sturm-Liouville eigenvalues**. Remember that in linear algebra there were only finitely many eigenvalues, but here, on an infinite dimensional space, there are infinitely many.

To be thorough, we should check to see if there are any others. If $\lambda = 0$ and $A'' = \lambda A$, then $A(x) = ax + b$ for some $a, b \in \mathbb{R}$. The solution has quite a different qualitative behavior. The boundary conditions give $b = 0$ and $\pi a + b = 0$. It follows that $a = b = 0$, and we only get the trivial solution (and no eigenvalue).

If $\lambda > 0$, then the general solution of $A'' = \lambda A$ is

$$A(x) = ae^{\mu x} + be^{-\mu x}$$

or alternatively

$$A(x) = a \cosh(\mu x) + b \sinh(\mu x)$$

where $\mu^2 = \lambda$. Taking the second form and applying the boundary values gives $a = 0$ and $a \cos(\mu\pi) + b \sinh(\mu\pi) = 0$. Since there is no positive value $\mu\pi$ with $\sinh(\mu\pi) = 0$, we only get the trivial solution $a = b = 0$.

At this point, we can be sure we have found every number λ and every function A so that the **Sturm-Liouville pair** (λ, A) gives a solution of (3.15). In fact, there is a sequence of eigenvalues $\lambda_j = -j^2$ and corresponding solutions $a \sin(jx)$ for $j = 1, 2, 3, \dots$. We can also take $a = a_j = 1$ for every j if we like.

In the end, the separated variables solutions must all look like

$$u_j(x, y) = a_j \sin(jx) B_j(y)$$

with $B_j'' = -\lambda_j B_j = -(-j^2) B_j$. It follows that

$$B_j(y) = \alpha_j \cosh(jy) + \beta_j \sinh(jy).$$

We also have a boundary condition $B_j(0) = 0$. This means $\alpha_j = 0$, and we can take, again for simplicity, $\beta_j = 1$ for all $j = 1, 2, 3, \dots$

Finally, then we attempt to solve the original problem (3.13) with a solution given by a **superposition**

$$u(x, y) = \sum_{j=1}^{\infty} a_j A_j(x) B_j(y) = \sum_{j=1}^{\infty} a_j \sin(jx) \sinh(jy)$$

of separated variables solutions. A superposition is rather like a linear combination of solutions, except that infinitely many terms are allowed, and (we at least hope) we find coefficients a_1, a_2, a_3, \dots so that we have convergence of the series. Assuming the convergence the PDE should be satisfied and the boundary condition(s) on $\partial U \setminus \Gamma$ should be satisfied. For the final step, we consider the boundary condition on Γ , i.e., where $y = \pi$. For this, we need

$$u(x, \pi) = \sum_{j=1}^{\infty} a_j \sinh(j\pi) \sin(jx) = g(x) = \frac{\pi^2}{4} - \left(x - \frac{\pi}{2}\right)^2.$$

We know how the coefficients must be chosen to accomplish this. Namely, we must have

$$a_j \sinh(j\pi) = \frac{8}{\pi(2k+1)^3} \quad \text{for} \quad j = 2k+1$$

and $a_j = 0$ otherwise, i.e., for the even index terms. Thus, we obtain a solution

$$u(x, y) = \frac{8}{\pi} \sum_{k=0}^{\infty} \frac{1}{(2k+1)^2 \sinh[(2k+1)\pi]} \sin[(2k+1)x] \sinh[(2k+1)y]. \quad (3.16)$$

The partial sum with two terms is illustrated in Figure 3.3.

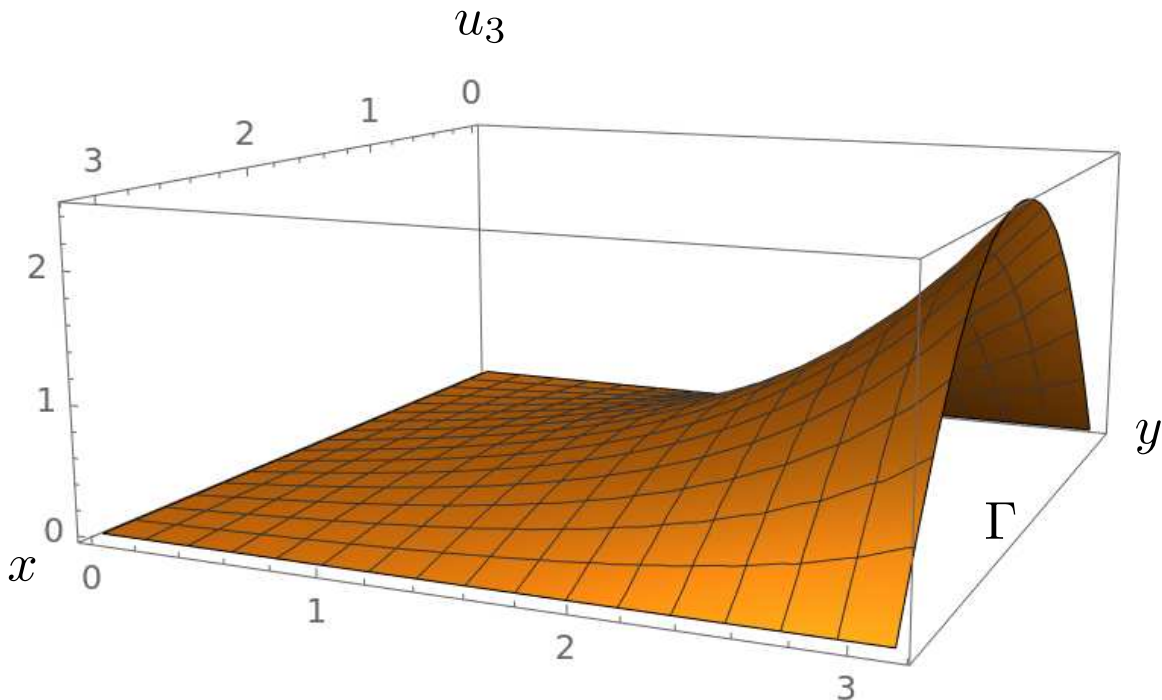


Figure 3.3: The partial sum u_3 of the solution given in (3.16) consisting of two terms corresponding to $k = 0$ and $k = 1$. The function u_3 is actually harmonic, satisfies $u_3(x, y) = 0$ for $(x, y) \in \partial U \setminus \Gamma$, and matches the boundary values along Γ to within about $\delta = 0.03$. The error in the interior can be no larger: You can prove that using the maximum principle.

3.1.3 Integral identities and properties

Perhaps the two most interesting properties of solutions of Laplace's equation are the **mean value property** and the **maximum principle**. Remember that a solution of Laplace's equation $\Delta u = 0$ is called a **harmonic function**.

integral identities

Owing to the divergence structure of the Laplace operator, that is because

$$\Delta u = \operatorname{div} Du,$$

one can obtain some kind of integral identities simply by integrating each of the fundamental equations and applying the divergence theorem. For example, if V is a nice domain of integration contained in an open set U on which $\Delta u = 0$, then

$$0 = \int_V \Delta u = \int_V \operatorname{div} Du = \int_{\partial V} Du \cdot n$$

where n is the unit outward conormal to ∂V . It may be recalled from our discussion of the divergence operator that

$$\int_{\partial V} Du \cdot n$$

is the flux integral of the gradient with respect to the domain V . The quantity

$$D_n u = Du \cdot n$$

is also the **outward normal derivative**. In particular $Du \cdot n$ is a directional derivative with respect to the outward normal and gives the rate of change (with respect to length in the domain) of the value of u along a normal line leaving the domain u . Thus, the integral identity

$$\int_{\partial V} Du \cdot n = 0$$

may be interpreted to say the average value of the normal derivative around the boundary of V is zero.

A second integral identity may be obtained using the product formula

$$\operatorname{div}(uD_u) = |Du|^2 + u \operatorname{div} Du.$$

Integrating both sides of the product rule and noting/assuming $\operatorname{div} Du = \Delta u = 0$, we have

$$\int_V |Du|^2 = \int_{\partial V} u Du \cdot n = \int_{\partial V} u D_n u.$$

Let us assume this discussion may be applied on the entire domain U , so that

$$\int_{\partial U} u D_n u = \int_U |Du|^2 \geq 0.$$

If this is the case, we can consider the Dirichlet boundary value problem and prove a uniqueness result:

Theorem 14 If $u, v \in C^2(\overline{U})$ are two solutions of

$$\begin{cases} \Delta u = 0, & \text{on } U \\ u|_{\partial U} = g, \end{cases}$$

where U is a bounded open subset of \mathbb{R}^n with C^1 boundary, then $u \equiv v$.

Proof: The difference $w = u - v$ is harmonic and vanishes on ∂U . Thus,

$$\int_U |D(u - v)|^2 = \int_{\partial U} (u - v) D_n(u - v) = 0.$$

Since the integrand is non-negative and continuous, it must be the case that $D(u - v) = Du - Dv$ is identically the zero vector. This means $u - v = c$ is constant on each connected component C of U . Given any $\mathbf{p} \in C$, we can take a segment $\{(1 - t)\mathbf{p} + t\mathbf{q} : 0 \leq t \leq 1\}$ with $(1 - t)\mathbf{p} + t\mathbf{q} \in C$ for $0 \leq t < 1$ but $\mathbf{q} \in \partial C$. By continuity

$$0 = g(\mathbf{q}) - g(\mathbf{p}) = u(\mathbf{q}) - v(\mathbf{q}) = \lim_{t \nearrow 1} [u((1 - t)\mathbf{p} + t\mathbf{q}) - v((1 - t)\mathbf{p} + t\mathbf{q})] = c.$$

Therefore $u(\mathbf{p}) = v(\mathbf{p})$ for every $\mathbf{p} \in U$. \square

The mean value property

Given an open set $U \subset \mathbb{R}^n$, the value of a harmonic function $u : U \rightarrow \mathbf{r}$ at a point $\mathbf{p} \in U$ is given by the average over any (hyper)sphere $B_r(\mathbf{p}) \subset\subset U$ centered at \mathbf{p} :

$$u(\mathbf{p}) = \frac{1}{n\omega_n r^{n-1}} \int_{\partial B_r(\mathbf{p})} u. \quad (3.17)$$

The $(n - 1)$ -dimensional measure of $\partial B_r(\mathbf{p})$ is $n\omega_n r^{n-1}$ where ω_n is the n -dimensional measure of the unit ball $B_1(\mathbf{0})$. It is a nice exercise to explore these measures. You know, for example, that the area measure of the unit disk in \mathbb{R}^2 is given by $\omega_2 = \pi$, and the area of the disk of radius r is obtained by scaling by the square of the radius: πr^2 . It is interesting, on the one hand, that the length of the boundary circle is the derivative of the area with respect to the radius:

$$2\omega_2 r = \frac{d}{dr} \pi r^2.$$

The same thing holds in all dimensions. You are probably familiar with the case $n = 3$ in which the volume of the unit ball is

$$\omega_3 = \int_{B_1(\mathbf{0})} 1 = \frac{4\pi}{3} \quad \text{and} \quad \int_{B_r(\mathbf{0})} 1 = \frac{4\pi}{3} r^3.$$

You may not have noticed that

$$\int_{\partial B_r(\mathbf{0})} 1 = \frac{d}{dr} \omega_3 r^3 = 3\omega_3 r^2.$$

What becomes quite interesting is to find a formula for the n -dimensional measure

$$\omega_n = \int_{B_1(\mathbf{0})} 1$$

when $B_1(\mathbf{0}) = \{\mathbf{x} \in \mathbb{R}^n : |\mathbf{x}| < 1\} \subset \mathbb{R}^n$. A formula involving the gamma function $\Gamma : [0, \infty) \rightarrow \mathbb{R}$ by

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$$

can be obtained by induction, namely,

$$\omega_n = \frac{\pi^{n/2}}{\Gamma(n/2 + 1)}. \quad (3.18)$$

Then

$$\int_{B_r} 1 = \omega_n r^n \quad \text{and} \quad \int_{\partial B_r} 1 = n\omega_n r^{n-1}.$$

Incidentally, for each positive integer m there holds $\Gamma(m) = (m-1)!$, and these values play a role in the formula for ω_n when n is even. For example, when $n = 2$ formula (3.18) becomes $\omega_2 = \pi$ because $\Gamma(2) = 1$. More generally, when $n = 2k$

$$\omega_{2k} = \frac{\pi^k}{k!} \quad \text{and} \quad n\omega_n = 2k \frac{\pi^k}{k!},$$

so these are all rational multiples of an integer power of π , namely $\pi^{n/2}$. The situation is somewhat different when $n = 2k + 1$ is odd. When $n = 1$, we know $\omega_1 = 2$ and the measure of the boundary, which in this case is counting measure, has the same

numerical value because the boundary of the interval $B_1(0) = (-1, 1)$ consists of two points. According to (3.18) we should have

$$2 = \omega_1 = \frac{\sqrt{\pi}}{\Gamma(3/2)},$$

so

$$\Gamma(3/2) = \int_0^\infty \sqrt{t}e^{-t} dt = \frac{\sqrt{\pi}}{2}$$

which is indeed the case. A similar comparison of (3.18) in the case $n = 3$ to the known value of ω_3 gives

$$\Gamma(5/2) = \frac{3}{4} \pi^{n/2-1} = \frac{3}{4} \sqrt{\pi}.$$

It is not difficult to show $\Gamma(x+1) = x\Gamma(x)$. It follows from this recursion that $\Gamma(n/2+1)$ is a rational multiple of $\sqrt{\pi}$ for each n and ω_{2k+1} is a rational multiple of an integer power of π as well.

Let us return to the mean value property. When $B_r(\mathbf{p}) \subset\subset U$, it is also possible to average over the entire ball:

$$u(\mathbf{p}) = \frac{1}{\omega_n r^n} \int_{B_r(\mathbf{p})} u. \quad (3.19)$$

Both mean value formulas (3.17) and (3.19) relate the value of u at a single point \mathbf{p} in a very rigid/structured way with many values of u at points different from \mathbf{p} . The proof(s) of these formulas follow from various change of variable formulas for integrals and the divergence theorem.

$$\begin{aligned} \frac{d}{dt} \frac{1}{t^{n-1}} \int_{\partial B_t(\mathbf{p})} u &= \frac{d}{dt} \frac{1}{t^{n-1}} \int_{\mathbf{x} \in \partial B_1(\mathbf{0})} u(\mathbf{p} + t\mathbf{x}) t^{n-1} \\ &= \frac{d}{dt} \int_{\mathbf{x} \in \partial B_1(\mathbf{0})} u(\mathbf{p} + t\mathbf{x}) \\ &= \int_{\mathbf{x} \in \partial B_1(\mathbf{0})} Du(\mathbf{p} + t\mathbf{x}) \cdot \mathbf{x} \\ &= \int_{B_1(\mathbf{0})} \operatorname{div}[Du(\mathbf{p} + t\mathbf{x})]. \end{aligned}$$

The second expression follows from a change of variables in the integral. Notice the scaling factor t^{n-1} . The expression in the third line is obtained by differentiating

under the integral sign. The last expression comes from the divergence theorem, and we have to be a little careful because the divergence is with respect to \mathbf{x} rather than the original variable $\mathbf{p} + t\mathbf{x}$ of integration. It follows from the chain rule, however, that

$$\operatorname{div}[Du(\mathbf{p} + t\mathbf{x})] = t\Delta u(\mathbf{p} + t\mathbf{x}).$$

Thus, we can continue:

$$\begin{aligned} \frac{d}{dt} \frac{1}{t^{n-1}} \int_{\partial B_t(\mathbf{p})} u &= t \int_{\mathbf{x} \in B_1(\mathbf{0})} \Delta u(\mathbf{p} + t\mathbf{x}) \\ &= t \int_{\partial B_1(\mathbf{p})} \Delta u \frac{1}{t^n} \\ &= \frac{1}{t^{n-1}} \int_{\partial B_1(\mathbf{p})} \Delta u \\ &= 0. \end{aligned}$$

It follows that the average

$$\frac{1}{n\omega_n t^{n-1}} \int_{\partial B_t(\mathbf{p})} u$$

over $\partial B_t(\mathbf{p})$ is a constant independent of t . On the other hand, by continuity we can, for any $\epsilon > 0$, take $r > 0$ small enough so that

$$|u(\mathbf{x}) - u(\mathbf{p})| < \epsilon \quad \text{whenever} \quad t < r.$$

Thus, for $t < r$ we have

$$\begin{aligned} \left| \frac{1}{n\omega_n t^{n-1}} \int_{\partial B_t(\mathbf{p})} u - u(\mathbf{p}) \right| &= \left| \frac{1}{n\omega_n t^{n-1}} \int_{\partial B_t(\mathbf{p})} [u - u(\mathbf{p})] \right| \\ &\leq \frac{1}{n\omega_n t^{n-1}} \int_{\partial B_t(\mathbf{p})} |u - u(\mathbf{p})| \\ &\leq \epsilon \frac{1}{n\omega_n t^{n-1}} \int_{\partial B_t(\mathbf{p})} 1 \\ &= \epsilon. \end{aligned}$$

This means the constant value of the average is the same as the limiting value and

$$\lim_{t \searrow 0} \frac{1}{n\omega_n t^{n-1}} \int_{\partial B_t(\mathbf{p})} u = u(\mathbf{p}). \quad \square$$

We have established the mean value property for averages over spheres. To get the averages over entire balls, we use a version of Fubini's theorem which sometimes goes by the names **generalized polar coordinates** or the **coarea formula**. Specifically,

$$\begin{aligned}
 \frac{1}{\omega_n r^n} \int_{B_r(\mathbf{p})} u &= \frac{1}{\omega_n r^n} \int_0^r \left(\int_{\partial B_t(\mathbf{p})} u \right) dt \\
 &= \frac{1}{\omega_n r^n} \int_0^r n\omega_n t^{n-1} \left(\frac{1}{n\omega_n t^{n-1}} \int_{\partial B_t(\mathbf{p})} u \right) dt \\
 &= \frac{1}{\omega_n r^n} \int_0^r n\omega_n t^{n-1} (u(\mathbf{p})) dt \\
 &= \frac{nu(\mathbf{p})}{r^n} \int_0^r t^{n-1} dt \\
 &= \frac{nu(\mathbf{p})}{r^n} \frac{r^n}{n} \\
 &= u(\mathbf{p}). \quad \square
 \end{aligned}$$

The application of Fubini's theorem/the coarea formula is in the first line. You can think about whether you believe it or not.

The maximum principle

The place to start in understanding the maximum principle is with the **weak maximum principle** which is easy to state, understand, and prove.

Theorem 15 (weak maximum principle) If $u : U \rightarrow \mathbb{R}$ is harmonic on a bounded open subset U of \mathbb{R}^n and $u \in C^0(\overline{U})$, then

$$u(\mathbf{p}) \leq \max_{\mathbf{x} \in \partial U} u(\mathbf{x}) \quad \text{for every } \mathbf{p} \in U. \quad (3.20)$$

Proof: If it were the case that $u(\mathbf{p}) > \max_{\mathbf{x} \in \partial U} u(\mathbf{x})$ for some $\mathbf{p} \in U$, then we can begin by considering a function of the form

$$w_0(\mathbf{x}) = -\epsilon |\mathbf{x} - \mathbf{p}|^2.$$

In this case, there is certainly a maximum point $\mathbf{p}_0 \in U$ at which

$$m_0 = u(\mathbf{p}_0) = \max_{\mathbf{x} \in \overline{U}} u(\mathbf{x}) > \max_{\mathbf{x} \in \partial U} u(\mathbf{x}).$$

Let us start with the function $v = m_0 + w_0$. Since U is bounded, it is possible to take $\epsilon > 0$ small enough so that

$$v|_{\partial U} > u|_{\partial U}.$$

In fact, if every point $\mathbf{x} \in U$ satisfies $|\mathbf{x} - \mathbf{p}_0| < R$, then taking $\epsilon > 0$ with

$$\epsilon < \frac{m_0 - \max_{\mathbf{x} \in \partial U} u(\mathbf{x})}{R^2}$$

we have

$$\begin{aligned} v(\mathbf{x}) &= m_0 - \epsilon|\mathbf{x} - \mathbf{p}_0|^2 \\ &\geq m_0 - \epsilon R^2 \\ &> m_0 - [m_0 - \max_{\mathbf{x} \in \partial U} u(\mathbf{x})] \\ &= \max_{\mathbf{x} \in \partial U} u(\mathbf{x}). \end{aligned}$$

Consequently, there is some $\mathbf{p}_1 \in U$ with $v(\mathbf{p}_1) \leq u(\mathbf{p}_1)$ and

$$c_1 = v(\mathbf{p}_1) - u(\mathbf{p}_1) = \min_{\mathbf{x} \in \bar{U}} [v(\mathbf{x}) - u(\mathbf{x})] \leq 0.$$

Thus, if we let $w = v - c_1$, then the following hold:

1. $w(\mathbf{x}) \geq u(\mathbf{x})$ for $\mathbf{x} \in \bar{U}$,
2. $w(\mathbf{p}_1) = u(\mathbf{p}_1)$.

From these conditions, it follows that $Du(\mathbf{p}_1) = Dv(\mathbf{p}_1)$ and

$$u_{xx}(\mathbf{p}_1) \leq v_{xx}(\mathbf{p}_1) \quad \text{and} \quad u_{yy}(\mathbf{p}_1) \leq v_{yy}(\mathbf{p}_1).$$

Therefore,

$$\Delta v(\mathbf{p}_1) \geq \Delta u(\mathbf{p}_1) = 0.$$

On the other hand, direct calculation gives $\Delta v \equiv -2n < 0$. Therefore, we have obtained a contradiction, meaning the assumption $u(\mathbf{p}) > \max_{\mathbf{x} \in \partial U} u(\mathbf{x})$ is impossible, and the weak maximum principle is proved. \square

The **strong maximum principle** is most easily understood by taking the weak maximum principle as a starting point. The strong maximum principle asserts that strict inequality holds in (3.20) in most cases. The exact statement is as follows:

Theorem 16 (E. Hopf strong maximum principle⁷) If $u : U \rightarrow \mathbb{R}$ is harmonic on a bounded open subset U of \mathbb{R}^n and $u \in C^0(\overline{U})$, then

$$u(\mathbf{p}) < \max_{\mathbf{x} \in \partial U} u(\mathbf{x}) \quad \text{for every } \mathbf{p} \in U \quad (3.21)$$

unless equality holds at every point and u is a constant function.

I will not prove the strong maximum principle here, but it is most often derived from a nice result called the E. Hopf boundary point lemma.

3.2 The heat equation

$$u_t = \Delta u$$

3.2.1 Derivation

We have gone over this in class. I'll try to include some more detail here.

We start with the assumption that heat energy within subregions V of a given solid⁸ U is expressible by a **thermal energy density** Θ $\mathbf{x} \in U$ and time t and having units/physical dimensions

$$[\Theta] = \frac{[\text{energy}]}{L^n} = \frac{[\text{force}]}{L^{n-1}} = \frac{M}{L^{n-2}T^2}$$

so that

$$E = E(t, V) = \int_V \Theta$$

models the thermal energy within the region V . It is assumed the thermal energy density $\Theta : U \times [0, T) \rightarrow \mathbb{R}$ is dependent on spatial position $\mathbf{x} \in U$ and time $t \in [0, T)$ for some positive final time T . Thus, while $\Theta = \Theta(\mathbf{x}, t)$, the energy within V is time dependent. We assume sufficient regularity so that differentiation under the integral sign with respect to t is justified and

$$\frac{d}{dt} E = \int_V \frac{\partial \Theta}{\partial t}.$$

⁷Eberhard Hopf, German mathematician (1902-1983), not to be confused with Heinz Hopf, German mathematician (1894-1971) known for the Hopf fibration.

⁸The mathematical assumption that the media is solid means we do not (need to) consider **transport** of heat energy via motion in the medium. This could also be included, but the model necessarily becomes substantially more complicated.

This quantity models/gives the rate of change of the energy within the specified region V . Notice this, when positive, is the rate thermal energy (globally in a net sense) enters V . When $E_t < 0$, this model quantity gives the negative rate/speed at which thermal energy leaves V .

We introduce next a **thermal energy density flux field** $\Phi : U \times [0, T) \rightarrow \mathbb{R}^n$. The physical dimensions of this model function are

$$[\Phi] = \frac{[\text{energy}]}{L^{n-1}T}$$

so that the flux integral

$$B(t) = \int_{\partial V} \Phi \cdot n,$$

with n the outward unit conormal, i.e., tangent vector to V normal to ∂V , along ∂V , models also the rate at which thermal energy passes through ∂V with, in a certain sense, the opposite sign convention: $B > 0$ gives the net rate at which thermal energy leaves V through ∂V , and $B < 0$ gives the negative of the rate at which thermal energy enters V .

There is another natural contribution to E_t we can easily model. Say, for example, an electrical current is passed through U creating heat **internally** or some cooling device is externally applied draining thermal energy from U , not through ∂U , but into some space ambient to the model space $\mathbb{R}^n \supset U$. The latter can be imagined if one is modeling a lamina region $U \subset \mathbb{R}^2$ so that $n = 2$, but a heat sink like an ice cube is applied to the interior of the lamina U from a larger physical space \mathbb{R}^3 containing both U and the nominal ambient space \mathbb{R}^2 containing U and (defining) ∂U . In any case, we introduce a second energy-rate density term $\rho : U \times [0, T) \rightarrow \mathbb{R}$ with physical units

$$[\rho] = \frac{[\text{energy}]}{L^n T}$$

so that

$$F = \int_V \rho$$

gives the rate of change of energy in V due to **interior sources and/or sinks**.

For our model, then, we begin with the assumed relation

$$E_t = -B + F;$$

the net rate of energy entering V is the negative of the rate at which energy exits through the boundary ∂V plus the rate at which energy is generated internally. That

is,

$$\int_V \Theta_t = - \int_{\partial V} \Phi \cdot n + \int_V \rho \quad \text{for every domain of integration } V \subset U. \quad (3.22)$$

This may be taken as a first (weak) form of the heat equation involving (usually) two unknown functions Θ and Φ , the latter of which constitutes actually n unknown real valued functions, and one structural, or driving, function f , often referred to as the “forcing.” Applying the divergence theorem to the flux integral, when the regularity of Φ allows it, we obtain a second weak form for the same functions

$$\int_V [\Theta_t + \operatorname{div} \Phi - \rho] = 0. \quad (3.23)$$

Again, assuming the regularity of the integrand in (3.23) is adequate, we may apply some version of Theorem 10 to conclude

$$\Theta_t = - \operatorname{div} \Phi + \rho. \quad (3.24)$$

This third form of the heat equation may, at this point, be simply taken as a classical model equation for the evolution of heat energy within U , involving a partial differential equation, and augmented with appropriate initial and boundary value conditions. For example, one may seek, for $\rho = \rho(\mathbf{x}, t)$ given, thermal energy densities Θ and flux fields Φ satisfying (3.24) for $(\mathbf{x}, t) \in U \times [0, T)$ and

$$\begin{aligned} \Theta(\mathbf{x}, 0) &= \theta(\mathbf{x}), & \mathbf{x} \in U & \quad \text{(an initial thermal energy density on } U) \\ \Phi(\mathbf{x}, t) &= \phi(\mathbf{x}, t), & \mathbf{x} \in \partial U & \quad \text{(a prescribed flux field on } \partial U). \end{aligned}$$

A common alternative to the general boundary condition here is the prescription of the normal flux along ∂U :

$$\Phi(\mathbf{x}, t) \cdot n(\mathbf{x}) = v(\mathbf{x}, t), \quad \mathbf{x} \in \partial U. \quad (3.25)$$

In general, the evolution of thermal energy in such a model may be considered as **underdetermined** due to lack of additional constitutive relations. Nevertheless, consideration of aspects of heat flow can sometimes be clarified by consideration of these conditions/relations.

The standard constitutive relations are obtained by the introduction of a further quantity assumed to have its own fundamental physical dimension: temperature. Precisely, we assume there exists a real valued function $u : U \times [0, T) \rightarrow \mathbb{R}$ for which the following model assumptions/relations hold:

1. (the law of specific heat) The thermal energy density is proportional to the temperature

$$\Theta = \sigma u.$$

The “constant” σ having physical dimensions

$$[\sigma] = \frac{[\text{energy}]}{\text{temp}}$$

is called the **specific heat capacity**, and generally may be spatially and time dependent within U .⁹

2. (Fourier’s law of heat conduction) The flux field Φ is proportional to the negative of the gradient

$$\Phi = -kDu$$

where the “constant” of proportionality is called the thermal **conductivity**. Generally, $k = k(\mathbf{x}, t, u)$. Again, the assumption that k depends on Du is a direct affront to the modeling hypothesis.

Applying the temperature relations to (3.22) we obtain a fourth form of the heat equation

$$\int_V [(\sigma u)_t - \rho] = \int_{\partial V} kDu \cdot n \quad (3.26)$$

assumed to hold for every suitable subdomain $V \subset U$. From the divergence theorem we derive a fifth form involving the single unknown function u :

$$\int_V [(\sigma u)_t - \text{div}(kDu) - \rho] = 0. \quad (3.27)$$

From this weak form we obtain the classical (sixth) form

$$(\sigma u)_t = \text{div}(kDu) + \rho. \quad (3.28)$$

⁹If you look up specific heat capacity, you may (and should) be troubled to find this “constant” of proportionality depends on the temperature u —and perhaps even derivatives of u . Such dependence obviously throws into question the entire modeling assumption. In fact, what is really being pointed out is that the model does not, strictly speaking, accord with experimental observations. On the other hand, assuming a given relation $\sigma = \sigma(\mathbf{x}, t, u)$ does lead to a nominally reasonable model equation, though one that has not received as much attention as the standard heat equation. Similar comments apply to the “constant” of conductivity in Fourier’s law.

All of these equations are called **forced heat equations** due to the inhomogeneity f , assuming f is nonzero. Note that the heat operator $L : C^2(U \times (0, T)) \rightarrow C_0(U \times (0, T))$ by

$$Lu = (\sigma u)_t - \operatorname{div}(kDu)$$

is linear. Assuming $\sigma = \sigma(\mathbf{x}, t)$ is independent of u and differentiable, we may write

$$(\sigma u)_t = \sigma u_t + \sigma_t u.$$

Similarly, if $k = k(\mathbf{x}, t)$ is independent of u and differentiable, we can write

$$\operatorname{div}(kDu) = k\Delta u + Dk \cdot Du.$$

These assumptions lead to various expanded forms of (3.28):

$$\begin{aligned} u_t &= \frac{1}{\sigma} \operatorname{div}(kDu) + \tau u + f && (\tau = -\sigma_t/\sigma, \text{ and } f = \rho/\sigma) \\ (\sigma u)_t &= k\Delta u + Dk \cdot Du + \rho \\ u_t &= \kappa \Delta u + \frac{1}{\sigma} Dk \cdot Du + \tau u + f && (\kappa = k/\sigma) \end{aligned}$$

The assumption that σ and k are strictly constant gives the standard forced heat equation

$$u_t = \kappa \Delta u + f.$$

Either scaling in space, scaling in time, or simply the assumption $\kappa = 1$ leads to consideration of the PDE

$$u_t = \Delta u + f$$

and the homogeneous equation $u_t = \Delta u$ we have called “the” heat equation above.

Assuming σ is given and the law of specific heat, the general initial condition $\Theta(\mathbf{x}, 0) = \theta(\mathbf{x})$ gives the standard initial condition

$$u(\mathbf{x}, 0) = u_0(\mathbf{x}) \quad \text{for } \mathbf{x} \in U$$

where $u_0 = \theta/\sigma$. Similarly, assuming the conductivity k is given and Fourier’s law, the general boundary condition $\Phi(\mathbf{x}, t) = \phi(\mathbf{x})$ becomes

$$Du(\mathbf{x}, t) = \mathbf{v}(\mathbf{x}) \quad \text{for } \mathbf{x} \in \partial U$$

where $\mathbf{v} = -\phi/k$. It is more common to use the boundary flux condition (3.25) so that

$$Du \cdot n = h \quad \text{for } \mathbf{x} \in \partial U \text{ and } t > 0.$$

In this case we have of course $h = -\phi \cdot n/k$. The case $h \equiv 0$ is of special note: It is a homogeneous boundary condition, and it models an **insulated boundary**. It will be noted that this is, essentially, a Neumann boundary condition much like that encountered in connection with Laplace's equation.

With the temperature u in hand as part of the model, it is also natural to introduce a **fixed temperature** boundary condition

$$u|_{\partial U} = g$$

where "fixed" means possibly dependent on both space and time so that $g : \partial U \times (0, T) \rightarrow \mathbb{R}$ and $g = g(\mathbf{x}, t)$, but it is (simply) the temperature that is fixed. This is a Dirichlet condition for the heat equation.

3.2.2 separated variables solutions

Sticking with the one-dimensional heat equation which models heat conduction along a thin rod or wire, let's consider a relatively simple but somewhat interesting problem: Say we have some initial temperature distribution $u_0 : (0, L) \rightarrow \mathbb{R}$ and for positive times we impose the fixed temperature boundary conditions

$$u(0, t) = 0 = u(L, t), \quad t > 0.$$

What is the evolution $u = u(x, t)$ satisfying these conditions and $u_t = \Delta u$? We already know how the sinusoidal separated variables solutions decay to zero:

$$u_j(x, t) = e^{-j^2 L^2 t / \pi^2} \sin\left(\frac{jL}{\pi} x\right).$$

You should be able to derive these solutions from the Sturm-Liouville problem resulting from assuming $u_j = A_j(x)B_j(t)$. We seek a superposition

$$u = \sum_{j=1}^{\infty} a_j e^{-j^2 L^2 t / \pi^2} \sin\left(\frac{jL}{\pi} x\right).$$

Given good convergence of the series, the heat equation and the boundary conditions should be satisfied. The only thing we need to (be able to) do is choose the coefficients so that the initial condition

$$\sum_{j=1}^{\infty} a_j \sin\left(\frac{jL}{\pi} x\right) = u_0(x)$$

holds. But this is just expanding $u_0(x)$ as a Fourier sine series. In Chapter 13 Section 3 Boas gives the example where $u_0(x) = ax$ which illustrates an interesting property of the heat equation, namely **instantaneous or infinite speed propagation**. The initial condition has $u_0(L) = aL \neq 0$, but for all positive times $u(L, t) \equiv 0$. That is to say, immediately after time $t = 0$ the temperature drops instantaneously to $u(L, t) = 0$. This is a characteristic property of solutions of the heat equation and manifests in various different ways.

Another interesting manifestation of infinite speed propagation is that even if u_0 satisfies $u_0 \in C^0[0, L] \setminus C^1(0, L)$, it can be shown that the corresponding evolution satisfies $u \in C^\infty((0, L) \times (0, \infty))$. Thus the initial temperature distribution is smoothed infinitely quickly.

3.2.3 integral identities

The first natural quantity is the (spatially) **total temperature** by which we mean the integral

$$\int_U u.$$

This is a function of time, and if the specific heat capacity is assumed constant, the value is proportional to the total thermal energy within U as a function of time:

$$\text{int}_U u = \frac{1}{\sigma} \int_U \Theta.$$

We have already differentiated $\int_U \Theta$ with respect to t , and now, in terms of the temperature we find

$$\frac{d}{dt} \int_U u = \int_U u_t = \int_U \Delta u = \int_U \text{div}(Du) = \int_{\partial U} Du \cdot n.$$

In particular, if the zero flux condition $Du \cdot n = 0$ holds on ∂U , i.e., we have an insulated boundary condition, then the total temperature (and the total thermal energy) remain constant:

$$\int_U u \equiv \int_U u_0.$$

This is particularly interesting in cases where the initial temperature profile u_0 is not a smooth function, and so some kind of discontinuous jump takes place in the evolution of u_0 at time $t = 0$. The value of the integral $\int_U u$ should not jump discontinuously.

A second quantity of interest is the L^2 norm (squared)

$$\int_U u^2$$

of the solution (temperature) u . Again, this is a function of time.

$$\frac{d}{dt} \int_U u^2 = 2 \int_U uu_t = 2 \int_U u\Delta u.$$

The product rule $\operatorname{div}(uD u) = |D u|^2 + u\Delta u$ gives

$$\frac{1}{2} \frac{d}{dt} \int_U u^2 = \int_U \operatorname{div}(uD u) - \int_U |D u|^2 = \int_{\partial U} uD u \cdot n - \int_U |D u|^2.$$

This identity has several interesting consequences. First of all, if for any reason the boundary integral

$$\int_{\partial U} uD u \cdot n$$

vanishes, then $\|u\|_{L^2}^2 = \int_U u^2$ must be non-increasing, and strictly decreasing unless u is constant. Vanishing of the boundary integral follows from any combination of zero temperature $u = 0$ and zero flux $D u \cdot n = 0$, i.e., points with an insulated boundary condition, at all points on the boundary.

A striking consequence of this non-decreasing conclusion is the uniqueness for boundary value problems of prescribed temperature or flux. Specifically, if u and v are solutions of $u_t = \Delta u + f$ for a given forcing function $f = f(\mathbf{x}, t)$ and satisfying the same initial

$$u(\mathbf{x}, 0) = v(\mathbf{x}, 0), \quad \mathbf{x} \in U$$

and boundary conditions (either Dirichlet or Neumann) along ∂U , then the computation above applied to $w = u - v$, which satisfies $w_t = \Delta w$ and

$$(wD w \cdot n)|_{\partial U} \equiv 0,$$

gives

$$\frac{d}{dt} \int_U w^2 = - \int_U |D w|^2 \leq 0.$$

Since

$$m(t) = \int_U w^2 \geq 0$$

with $m(0) = 0$ and $m'(0) \leq 0$, we conclude $m(t) \equiv 0$, and $w \equiv 0$. Thus, $u \equiv v$.

3.3 The wave equation

$$u_{tt} = \Delta u$$

3.3.1 derivation

Here we consider an idealized model for the dynamics (time dependent motion) of a compressible/extendable continuum (like a slinky) subject to a elastic constitutive relation¹⁰

$$T = \epsilon(u_x - 1)$$

where $u : [0, L_0] \times (0, T) \rightarrow [0, L]$ models a deformation of the continuum assumed to satisfy $u_x \geq 1$ and a fixed endpoint boundary condition

$$u(0, t) = 0 \quad \text{and} \quad u(L_0, t) = L \quad \text{for all } t \geq 0.$$

We assume further a constant linear density ρ , so the mass corresponding to the interval $I = (a, b) \subset [0, L_0]$ is

$$(b - a)\rho = \int_I \rho = \int_{u(I, t)} \frac{\rho}{u_x}. \quad (3.29)$$

Note that $u_x = u_x(\xi, t)$ on the right in (3.29) where $\xi = \xi(v, t)$ is the time dependent spatial inverse of u , that is, $u(\xi(v, t), t) = v$ for each $v \in [0, L]$ and v is the variable of integration. Crucial to this derivation is the **continuum acceleration assumption** that

$$\frac{d}{dt} \int_{u(I)} u_t \frac{\rho}{u_x} = T(b) - T(a) = \epsilon[u_x(b, t) - u_x(a, t)]. \quad (3.30)$$

The quantity

$$\int_{u(I)} u_t \frac{\rho}{u_x} = \rho \int_I u_t$$

may be considered as the **total momentum** corresponding to the interval I , and the relation above may be compared to Newton's second law

$$\frac{d}{dt} \left(m \frac{dx}{dt} \right) = F$$

¹⁰This assumption may be compared to the assumption $F = -\epsilon(u' - 1)$ in the time-independent model for an inhomogeneously deformed slinky.

for the motion of a point mass, but it is a continuum modeling generalization of this momentum/acceleration law. Changing variables on the left in (3.30) and differentiating under the integral sign, we get

$$\frac{d}{dt} \int_{u(I)} u_t \frac{\rho}{u_x} = \frac{d}{dt} \int_I \rho u_t = \int_I \rho u_{tt}.$$

Applying the fundamental theorem of calculus on the right in (3.30) we have

$$u_x(b, t) - u_x(a, t) = \int_I u_{xx}.$$

Putting these manipulations together we obtain

$$\int_I [\rho u_{tt} - \epsilon u_{xx}] = 0$$

which is assumed to hold for every $I \subset [0, L_0]$. It follows as usual that

$$\rho u_{tt} = \epsilon u_{xx}. \quad (3.31)$$

This is essentially the wave equation.

Note: It may seem that the changes of variables and the constant density ρ are unnecessary in this derivation. However, the presentation is made in this way for easy adaptation for a linear density $\rho = \rho(x, t)$ which is non-constant or even time dependent (and for other natural generalizations as well).

Exercise 3.5 As mentioned in the first chapter above, it is possible to apply the method of characteristics to find particular solutions in the form of d'Alembert's solution to the **initial value problem**

$$\begin{cases} u_{tt} = u_{xx} & \text{for } (x, t) \in \mathbb{R} \times [0, \infty) \\ u(x, 0) = u_0(x), & x \in \mathbb{R} \\ u_t(x, 0) = v_0(x), & x \in \mathbb{R}, \end{cases} \quad (3.32)$$

and the situation is more difficult/complicated if one considers the one-dimensional wave equation on a finite spatial interval $[a, b] \subset \mathbb{R}$ with additional boundary conditions at $x = a$ and $x = b$.

Starting with the initial value problem

$$\begin{cases} w_{tt} = \alpha w_{xx} & \text{for } (x, t) \in \mathbb{R} \times [0, \infty) \\ w(x, 0) = w_0(x), & x \in \mathbb{R} \\ w_t(x, 0) = p_0(x), & x \in \mathbb{R}, \end{cases} \quad (3.33)$$

for the equation in (3.31) where $\alpha > 0$,

- (a) Scale the spatial independent variable to determine specific values of the initial functions $u_0 : \mathbb{R} \rightarrow \mathbb{R}$ and $v_0 : \mathbb{R} \rightarrow \mathbb{R}$ in (3.32) so that the solution w of (3.33) can be expressed in terms of the solution u of (3.32).
- (b) Use a scaling in the time variable t to determine specific values of the initial functions $u_0 : \mathbb{R} \rightarrow \mathbb{R}$ and $v_0 : \mathbb{R} \rightarrow \mathbb{R}$ in (3.32) so that the solution w of (3.33) can be expressed in terms of the solution u of (3.32).
- (c) Given a scaling $\sigma(x) = \beta x$ in the spatial variable so that the solution of (3.33) is assumed to be given/obtained in the form

$$w(x, t) = u(\beta x, \gamma t)$$

where u is a solution of (3.32) and β and γ are positive real numbers, determine the required values of the initial functions $u_0 : \mathbb{R} \rightarrow \mathbb{R}$ and $v_0 : \mathbb{R} \rightarrow \mathbb{R}$ in (3.32).

3.3.2 separated variables solutions

Here we observe that the technique of separation of variables and Fourier series expansion may be used to solve a simple **initial/boundary value problem** for the one-dimensional wave equation on the entire real line. This is (essentially) Problem 5 in Chapter 13 Section 4 of Boas. I will leave to you as an exercise the problem Boas considers as an example in Chapter 13 Section 4 itself.

Taking the model for the (internal) deformation¹¹ of a one-dimensional continuum derived in the previous section with elasticity $\epsilon = 1$ and linear mass density $\rho = 1$, natural boundary conditions on the spatial interval $[0, L]$ for $L > 0$ are given by

$$u(0, t) = 0 \quad \text{and} \quad u(L, t) = L.$$

Thus, we consider for $h > 0$ the initial/boundary value problem

$$\begin{cases} u_{tt} = u_{xx} & \text{for } (x, t) \in (0, L) \times [0, \infty) \\ u(x, 0) = x, & x \in [0, L] \\ u_t(x, 0) = h - 2h|x - L/2|/L, & x \in [0, L] \\ u(0, t) = 0, & t \geq 0 \\ u(L, t) = L, & t \geq 0. \end{cases} \quad (3.34)$$

¹¹Along with many authors Boas presents the wave equation as a model for the transverse motion of a stretched vibrating elastic string—like a guitar string. We are working with the wave equation as a model for a fundamentally different physical system.

As a preliminary step, which may be viewed simply as a natural normalization or an attempt to reduce the initial and boundary values to homogeneous conditions, we introduce the function

$$w(x, t) = u(x, t) - x$$

and observe that w should satisfy

$$\begin{cases} w_{tt} = w_{xx} & \text{for } (x, t) \in (0, L) \times [0, \infty) \\ w(x, 0) = 0, & x \in [0, L] \\ w_t(x, 0) = h - 2h|x - L/2|/L, & x \in [0, L] \\ w(0, t) = 0, & t \geq 0 \\ w(L, t) = 0, & t \geq 0. \end{cases} \quad (3.35)$$

We seek first separated variables solutions of the PDE: $w(x, t) = A(x)B(t)$. This leads to

$$A(x) B''(t) = A''(x) B(t) \quad \text{or} \quad \frac{B''(t)}{B(t)} = \frac{A''(x)}{A(x)}$$

at points where $A(x)B(t) \neq 0$. Since the left side quotient only depends on time t and the right side quotient only depends on the spatial variable x , we obtain

$$\frac{B''(t)}{B(t)} = \frac{A''(x)}{A(x)} = \lambda$$

for some separation constant $\lambda \in \mathbb{R}$. (To formally prove this, differentiate both sides with respect to either x or t .)

We consider next the boundary conditions for $w(x, t) = A(x)B(t)$:

$$A(0)B(t) = 0 = A(L)B(t) \quad \text{for } t \geq 0.$$

These conditions suggest, assuming $B \equiv 0$, $A(0) = 0 = A(L)$ and direct our attention to the ODE $A'' = \lambda A$. We should be familiar with the cases for such a Sturm-Liouville problem by now, but I review: If $\lambda < 0$, we obtain a sequence of separation constants

$$\lambda_j = -\frac{j^2 \pi^2}{L^2}$$

and corresponding (basis) solutions

$$A_j(x) = \sin\left(\frac{j\pi}{L} x\right) \quad \text{for } j = 1, 2, 3, \dots$$

From the time dependent ODE we obtain

$$B_j(t) = a_j \cos\left(\frac{j\pi}{L}t\right) + b_j \sin\left(\frac{j\pi}{L}t\right)$$

so that we have a sequence of separated variables solutions

$$w_j(x, t) = A_j(x)B_j(t)$$

satisfying the incomplete initial/boundary valued problem

$$\begin{cases} w_{tt} = w_{xx} & \text{for } (x, t) \in (0, L) \times [0, \infty) \\ w(0, t) = 0, & t \geq 0 \\ w(L, t) = 0, & t \geq 0 \end{cases} \quad (3.36)$$

for the wave equation where the initial conditions are omitted. Naturally, for each of these separated variables solutions, the initial conditions

$$w_j(x, 0) = A_j(x)B_j(0) = a_j \sin\left(\frac{j\pi}{L}x\right) \quad \text{and} \quad \frac{\partial w_j}{\partial t}(x, 0) = \frac{j\pi}{L}b_j \sin\left(\frac{j\pi}{L}x\right)$$

may be appended to the problem (3.36), though these do not match the initial conditions for the original problem.

We briefly turn to the vacuous cases: If $\lambda = 0$, the equation for A becomes $A'' = 0$, and hence $A(x) = ax + b$. Taking the boundary values into account

$$b = 0 \quad \text{and} \quad aL + b = 0,$$

so that $a = 0$ as well, and we obtain no nontrivial separated variables solution. Similarly, if $\lambda > 0$, the general solution of $A'' = \lambda A$ is

$$A(x) = a \cosh(\sqrt{\lambda}x) + b \sinh(\sqrt{\lambda}x).$$

Taking the boundary conditions into account gives

$$a = 0 \quad \text{and} \quad a \cosh(\sqrt{\lambda}L) + b \sinh(\sqrt{\lambda}L) = 0,$$

so that $a = b = 0$.

With this (review) out of the way, we seek a solution as a **superposition**

$$w(x, y) = \sum_{j=1}^{\infty} A_j(x)B_j(t) = \sum_{j=1}^{\infty} \sin\left(\frac{j\pi}{L}x\right) \left[a_j \cos\left(\frac{j\pi}{L}t\right) + b_j \sin\left(\frac{j\pi}{L}t\right) \right]$$

of separated variables solutions. Assuming adequate convergence of such a series, we should have that w satisfies the incomplete boundary value problem (3.36) with the initial conditions

$$w(x, 0) = \sum_{j=1}^{\infty} a_j \sin\left(\frac{j\pi}{L} x\right)$$

and

$$w_t(x, 0) = \sum_{j=1}^{\infty} b_j \frac{j\pi}{L} \sin\left(\frac{j\pi}{L} x\right).$$

In view of the desired initial condition $w(x, 0) = 0$ of (3.35), we take $a_j = 0$ for $j = 1, 2, 3, \dots$. Similarly, we want

$$\sum_{j=1}^{\infty} b_j \frac{j\pi}{L} \sin\left(\frac{j\pi}{L} x\right) = h - 2h|x - L/2|/L$$

for which we find a Fourier sine series determined by

$$b_j \frac{j\pi}{L} \frac{1}{2L} = h \int_0^L \left(1 - \frac{2}{L} \left|x - \frac{L}{2}\right|\right) \sin\left(\frac{j\pi}{L} x\right) dx. \quad (3.37)$$

We compute:

$$\int_0^L \sin\left(\frac{j\pi}{L} x\right) dx = -\frac{L}{j\pi} \cos\left(\frac{j\pi}{L} x\right) \Big|_{x=0}^L = \frac{L}{j\pi} [1 - (-1)^j].$$

$$\begin{aligned} \int_0^L \left|1 - \frac{L}{2}\right| \sin\left(\frac{j\pi}{L} x\right) dx &= \int_0^{L/2} \left(\frac{L}{2} - x\right) \sin\left(\frac{j\pi}{L} x\right) dx \\ &\quad + \int_{L/2}^L \left(x - \frac{L}{2}\right) \sin\left(\frac{j\pi}{L} x\right) dx. \end{aligned}$$

Thus, when $j = 2k + 1$ is odd,

$$\int_0^L \left|1 - \frac{L}{2}\right| \sin\left(\frac{j\pi}{L} x\right) dx = \frac{L^2}{j\pi} - (-1)^k \frac{2L^2}{j^2\pi^2},$$

and

$$\int_0^L \left|1 - \frac{L}{2}\right| \sin\left(\frac{j\pi}{L} x\right) dx = 0$$

when j is even. Thus, we find

$$h - 2h|x - L/2|/L = \sum_{k=0}^{\infty} b_{2k+1} \frac{(2k+1)\pi}{L} \sin\left(\frac{(2k+1)\pi}{L} x\right)$$

with

$$\begin{aligned} b_{2k+1} &= \frac{2L^2h}{(2k+1)\pi} \left[\frac{2L}{(2k+1)\pi} - \frac{2}{L} \left(\frac{L^2}{(2k+1)\pi} - (-1)^k \frac{2L^2}{(2k+1)^2\pi^2} \right) \right] \\ &= (-1)^k \frac{8L^3h}{(2k+1)^3\pi^3}. \end{aligned}$$

This gives the solution

$$w(x, t) = \sum_{k=0}^{\infty} (-1)^k \frac{8L^3h}{(2k+1)^3\pi^3} \sin\left(\frac{(2k+1)\pi}{L} t\right) \sin\left(\frac{(2k+1)\pi}{L} x\right). \quad (3.38)$$

From this, one can plot, animate, and otherwise graphically illustrate the solution $u(x, t) = x + w(x, t)$.

Exercise 3.6 Determine the frequency and maximum displacement of the solution given in (3.38). Hint: Use the method of characteristics.

3.3.3 integral identities

Let us consider the quantity

$$E(t) = \frac{1}{2} \int_U [w_t^2 + |Du|^2]$$

associated with a function $u \in C^2(\bar{U} \times [0, \infty))$ with respect to an open bounded set $U \subset \mathbb{R}^n$ with C^1 boundary and satisfying

$$u_{tt} = \Delta u.$$

Differentiating with respect to time t we have

$$\frac{dE}{dt} = \int_U \left[w_t w_{tt} + \sum_{j=1}^n \frac{\partial u}{\partial x_j} \frac{\partial^2 u}{\partial x_j \partial t} \right].$$

The first term/integral is clearly

$$\int_U w_t \Delta u$$

by the equation. Each of the other n terms/integrals can be written as

$$\int_U \frac{\partial u}{\partial x_j} \frac{\partial u_t}{\partial x_j}.$$

For each such term and for fixed t , we consider the field

$$\mathbf{v} = u_t \frac{\partial u}{\partial x_j} \mathbf{e}_j.$$

The divergence of this field is

$$\operatorname{div} \mathbf{v} = \frac{\partial u_t}{\partial x_j} \frac{\partial u}{\partial x_j} + u_t \frac{\partial^2 u}{\partial x_j^2}.$$

Thus, integrating and applying the divergence theorem, we obtain

$$\int_{\partial U} u_t \frac{\partial u}{\partial x_j} \mathbf{e}_j \cdot \mathbf{n} = \int_U \frac{\partial u_t}{\partial x_j} \frac{\partial u}{\partial x_j} + \int_U u_t \frac{\partial^2 u}{\partial x_j^2}$$

where \mathbf{n} is the outward unit conormal to ∂U . Summing from $j = 1$ to $j = n$ and using these substitutions we see

$$\frac{dE}{dt} = \int_U u_t \Delta u + \int_{\partial U} u_t Du \cdot \mathbf{n} - \int_U u_t \Delta u = \int_{\partial U} u_t Du \cdot \mathbf{n}.$$

This identity has a number of interesting consequences. First of all, if the function $u_t Du \cdot \mathbf{n}$ vanishes on ∂U , then this **energy is conserved**. In particular for the fixed boundary condition

$$u(\mathbf{x}, t) = u_0(\mathbf{x}) \quad \text{for } \mathbf{x} \in \partial U \text{ and } t > 0$$

this is the case because in this case

$$u_t(\mathbf{x}, t) = \frac{d}{dt}[u_0(\mathbf{x})] \equiv 0 \quad \text{for } \mathbf{x} \in \partial U.$$

Evidently, roughly the same integral technique used above for showing uniqueness for the heat equation (or for Laplace's equation) may be used here to obtain uniqueness of solutions for certain boundary/initial value problems for the wave equation.

Chapter 4

Appendix 1: Partial Derivatives

Note: The following exposition was extracted from my MATH 6702 course notes on differentiation from spring 2020. These notes may be found in their entirety on the course page for MATH 6702 in spring 2023 (along with associated second assignments from 2020 and 2021 mentioned below). This material appears in some form in the third assignment from 2023—I guess I’m getting old and slowing down. Generally, there is a lot of good material in the notes from spring 2020 and on the assignments; you might want to have a look if you have time and interest. What is below can serve you, I think, as a pretty good review of a lot of the material in Chapter 1 above. There is a touch of additional material included as well beyond multi-index notation and multivariable power series.

The simplest case in which partial derivatives arise is that in which one has a function of two variables. Say $u = u(x, y)$ is a function of two variables defined in an open set U with $\mathbf{p} = (p_1, p_2) \in U$. We defined the **first partial derivative** of u in the x -direction at \mathbf{p} to be

$$\frac{\partial u}{\partial x}(\mathbf{p}) = \lim_{h \rightarrow 0} \frac{u(p_1 + h, p_2) - u(p_1, p_2)}{h}$$

when this limit exists. We can also call this the first partial of u in the \mathbf{e}_1 direction in honor of the standard basis vector $\mathbf{e}_1 = (1, 0)$. Notice that the limit above may also be written

$$\lim_{h \rightarrow 0} \frac{u(\mathbf{p} + h\mathbf{e}_1) - u(\mathbf{p})}{h}. \tag{4.1}$$

Exercise 4.1 *Sketch the graph of $u(x, y) = x^2 + y^2$. Let \mathbf{p} be a point in the first quadrant of the domain of u .*

1. Along with your graph, sketch the curve

$$\alpha(t) = \mathbf{p} + t\mathbf{e}_1 \quad \text{for } t \in \mathbb{R}$$

where where we interpret \mathbf{p} as $(p_1, p_2, 0) \in \mathbb{R}^3$ and $\mathbf{e}_1 = (1, 0, 0)$.

2. Also sketch the curve

$$\gamma(t) = (\mathbf{p} + t\mathbf{e}_1, u(\mathbf{p} + t\mathbf{e}_1))$$

where here we interpret \mathbf{p} and \mathbf{e}_1 as points in the domain of u .

3. Realize the second curve as the intersection of a vertical plane through the first curve.

Your solution should look something like what you see in Figure 4.1.

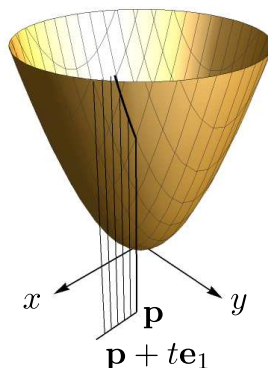


Figure 4.1: A curve in the plane $y = p_2$ on the graph of u .

The difference quotient appearing in (4.1) is the slope of a secant line in the plane $y = p_2$. Thus, the limit (the x -partial derivative of u) is the slope, in this plane, of the line tangent to the curve of intersection of the graph of u

$$\mathcal{G} = \{(x, y, u(x, y)) : (x, y) \in U\}$$

with the vertical plane $y = p_2$. A similar interpretation applies to the other partial derivative

$$\frac{\partial u}{\partial y}(\mathbf{p}) = \lim_{h \rightarrow 0} \frac{u(\mathbf{p} + h\mathbf{e}_2) - u(\mathbf{p})}{h}.$$

More generally, we can take a curve $\alpha(t) = \mathbf{p} + t\mathbf{v}$ in any direction \mathbf{v} , and attempt to compute the derivative:

$$D_{\mathbf{v}}u(\mathbf{p}) = \lim_{h \rightarrow 0} \frac{u(\mathbf{p} + h\mathbf{v}) - u(\mathbf{p})}{h}. \quad (4.2)$$

If this limit exists, it is called the **directional derivative** of u at \mathbf{p} in the direction \mathbf{v} .

Exercise 4.2 *If \mathbf{v} is a unit vector in (4.2), then the value of $D_{\mathbf{v}}u(\mathbf{p})$ is the rate of change of u in the direction \mathbf{v} . What does $D_{\mathbf{v}}u(\mathbf{p})$ measure when \mathbf{v} is a nonzero vector but does not have unit length?*

Returning to the special case where the direction of differentiation is $\mathbf{v} = \mathbf{e}_1$ or \mathbf{e}_2 , the values

$$\frac{\partial u}{\partial x}(\mathbf{p}) \quad \text{and} \quad \frac{\partial u}{\partial y}(\mathbf{p})$$

defined above are called the **first partial derivatives** of u at \mathbf{p} . There are various notations for these partial derivatives. Among them are

$$\begin{array}{ccc} u_x & \text{and} & u_y, \\ D_1u & \text{and} & D_2u, \\ u_1 & \text{and} & u_2, \end{array}$$

and

$$D^{(1,0)}u \quad \text{and} \quad D^{(0,1)}u. \quad (4.3)$$

And of course, we can interpret these partials as special directional derivatives

$$D_{\mathbf{e}_1}u \quad \text{and} \quad D_{\mathbf{e}_2}u.$$

The notation in (4.3) can be especially useful in certain contexts and we will discuss it carefully below, though it is rather cumbersome in this simple case.

Notice that the construction of taking a path in the direction of any standard unit basis vector and attempting the form the limit of a difference quotient works the same way in any dimension. More precisely, given $u : U \rightarrow \mathbb{R}$ with U an open subset of \mathbb{R}^n and $\mathbf{p} \in U$, the **partial derivative** of u at $\mathbf{p} = (p_1, p_2, \dots, p_n)$ is given by

$$\frac{\partial u}{\partial x_j}(\mathbf{p}) = \lim_{h \rightarrow 0} \frac{u(\mathbf{p} + h\mathbf{e}_j) - u(\mathbf{p})}{h}$$

when this limit exists. The picture is harder to draw in higher dimensions. The construction for general **directional derivatives** (4.2) may also be applied with only a dimensional change.

Exercise 4.3 How do each of the notations for the first partial derivatives above change when the dimension is greater than two?

If first partials exist at every point in an open set U on which the function u is defined, then they also define functions on U . Here is an important point:

For functions of one variable, if the derivative of f exists on an open interval (a, b) , then we say the function f is **differentiable** and write $f \in \text{Diff}(a, b)$, but...

In more than one dimension, we do not (usually) identify the existence of partial derivatives with differentiability.

Let us say that if the first partial derivatives of a function of several variables all exist on an open set U , then u is **partially differentiable** on U . The following results illustrate why partially differentiable and differentiable are different things in higher dimensions.

Theorem 17 If $f : (a, b) \rightarrow \mathbb{R}$ is differentiable at a point $x_0 \in (a, b)$, then f is continuous at x_0 . But the first partial derivatives of $u : \mathbb{R}^2 \rightarrow \mathbb{R}$ by

$$u(x, y) = \begin{cases} 0, & xy \neq 0 \\ 1, & xy = 0 \end{cases} \quad (4.4)$$

both exist at the point $(0, 0) \in \mathbb{R}^2$, but u is not continuous at $(0, 0)$. Therefore, partial differentiability at a point does not imply continuity.

Exercise 4.4 Compute the first partials u_x and u_y at $(x, y) = (0, 0)$ where $u : \mathbb{R}^2 \rightarrow \mathbb{R}$ is given by (4.4), and show that u is not continuous at $(0, 0)$.

Theorem 18 If $f : (a, b) \rightarrow \mathbb{R}$ is differentiable at every point in (a, b) , then $f \in C^0(a, b)$. But the first partial derivatives of $u : \mathbb{R}^2 \rightarrow \mathbb{R}$ by

$$u(x, y) = \begin{cases} xy/(x^2 + y^2), & (x, y) \neq (0, 0) \\ 0, & (x, y) = (0, 0) \end{cases} \quad (4.5)$$

both exist at all points of \mathbb{R}^2 , but $u \notin C^0(\mathbb{R}^2)$. Therefore, partial differentiability at all points does not imply continuity.

Exercise 4.5 Compute the first partials u_x and u_y at all points in the plane where $u : \mathbb{R}^2 \rightarrow \mathbb{R}$ is given by (4.5), and show that u is not continuous at $(0, 0)$.

The following result is about differential approximation:

Theorem 19 *If $f : (a, b) \rightarrow \mathbb{R}$ is differentiable at $x_0 \in (a, b)$, then*

$$\lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0) - hf'(x_0)}{h} = 0,$$

but for either of the functions $u : \mathbb{R}^2 \rightarrow \mathbb{R}$ given by (4.4) or (4.5), one does not have

$$\lim_{\mathbf{w} \rightarrow \mathbf{0}} \frac{u(\mathbf{p} + \mathbf{w}) - u(\mathbf{p}) - (u_x(\mathbf{p}), u_y(\mathbf{p})) \cdot \mathbf{w}}{|\mathbf{w}|} = 0 \quad (4.6)$$

where $\mathbf{p} = (0, 0)$ and

$$\mathbf{v} \cdot \mathbf{w} = (v_1, v_2) \cdot (w_1, w_2) = v_1w_1 + v_2w_2$$

is the **dot product**.

Exercise 4.6 *Verify the assertion of this theorem.*

Notice that the limit in (4.6) is a somewhat different kind of limit than we have encountered before. The vector valued increment \mathbf{w} can approach \mathbf{p} in a much greater variety of ways than a number $x_0 + h$ can approach x_0 as $h \rightarrow 0$. And this makes a significant difference.

Definition 10 *A function $u : U \rightarrow \mathbb{R}$ with U an open subset of \mathbb{R}^n and $\mathbf{p} \in U$ is **differentiable** at \mathbf{p} if there is a linear function $L : \mathbb{R}^n \rightarrow \mathbb{R}$ such that*

$$\lim_{\mathbf{w} \rightarrow \mathbf{0}} \frac{u(\mathbf{p} + \mathbf{w}) - u(\mathbf{p}) - L(\mathbf{w})}{|\mathbf{w}|} = 0. \quad (4.7)$$

The function u is **differentiable** on U if u is differentiable at every point $\mathbf{p} \in U$.

Theorem 20 *If $u : U \rightarrow \mathbb{R}$ with U an open subset of \mathbb{R}^n has first order partial derivatives defined at all points in some open ball $B_r(\mathbf{p}) \subset U$, and $D_j u$ is continuous at \mathbf{p} for $j = 1, \dots, n$, then u is differentiable at \mathbf{p} .*

Exercise 4.7 *If $u : U \rightarrow \mathbb{R}$ is differentiable at $\mathbf{p} \in U$, then show the first partial derivatives $D_j u(\mathbf{p})$ exist for $j = 1, 2, \dots, n$, and express the linear function $L : \mathbb{R}^n \rightarrow \mathbb{R}$ for which (4.7) holds in terms of the **gradient vector***

$$Du(\mathbf{p}) = (D_1 u(\mathbf{p}), D_2 u(\mathbf{p}), \dots, D_n u(\mathbf{p})).$$

Hint: Compare (4.7) to (4.6).

Theorem 21 *If $u : U \rightarrow \mathbb{R}$ with U an open subset of \mathbb{R}^n has first order partial derivatives which are continuous, i.e., $u \in C^1(U)$, then u is differentiable and u is continuous.*

We are now in a position to define the differentiability classes $C^k(U)$ for $U \subset \mathbb{R}^n$. Some aspects of these definitions formalize and extend our discussion even in \mathbb{R}^1 . First of all, if U is open,

$$C^1(U) = \{u \in C^0(U) : D_j u(\mathbf{p}) \text{ exists for every } \mathbf{p} \in U \text{ and} \\ D_j u \in C^0(U) \text{ for } j = 1, \dots, n \}.$$

For $k = 2, 3, 4, \dots$,

$$C^k(U) = \{u : D^\beta u \in C^1(U) \text{ for } |\beta| = k - 1\}.$$

Notice that this definition implies the partial derivatives $D^\beta u(\mathbf{p})$ with $|\beta| = k - 1$ and $\mathbf{p} \in U$ exist as well as the derivatives $D^\beta u$ with $|\beta| = k$.

For a general set $E \subset \mathbb{R}^n$,

$$C^k(E) = \{u \in C^0(E) : \text{there exists an open set } U \subset \mathbb{R}^n \text{ such that } E \subset U \\ \text{and there exists } \bar{u} \in C^k(U) \text{ with } \bar{u}|_E = u\}.$$

The spaces $C^k(U)$ and $C^k(E)$ as we have defined them, will not be normed spaces in general. Following the constructions above, we define

$$[u]_{C^1} = \sup_j \sup_{\mathbf{x}} |D_j u(\mathbf{x})|$$

which is a seminorm, called the C^1 seminorm, on the set of all C^1 functions for which the value is finite. We also define for an open set $U \subset \mathbb{R}^n$

$$C_b^1(U) = \{u \in C^1(U) : \|u\|_{C^0(U)} < \infty \text{ and } [u]_{C^1(U)} < \infty\}$$

which is a normed space with

$$\|u\|_{C^1(U)} = \|u\|_{C^0(U)} + [u]_{C^1(U)}.$$

For $k = 2, 3, \dots$, we define the seminorm

$$[u]_{C^k} = \sup_{|\beta|=k} \sup_{\mathbf{x}} |D^\beta u(\mathbf{x})|$$

and

$$C_b^k(U) = \{u \in C_b^{k-1}(U) : [u]_{C^k(U)} < \infty\}.$$

The C^k norm is given in general by

$$\|u\|_{C^k} = \sum_{j=0}^k [u]_{C^j}$$

where we take $[u]_{C^0} = \|u\|_{C^0}$.

Taylor Expansion and Power Series

As an application of the multi-index notation for partial derivatives, we mention some useful facts about multivariable Taylor expansions and power series. This material is also covered in Assignment 2 (spring 2020 and spring 2021) and relates to Boas section 4.2. As in the assignment, we start with recalling how Taylor expansion works in one dimension. The **Taylor expansion** of a function

$$f \in C^\infty(\mathbb{R}) = \bigcap_{k=0}^\infty C^k(\mathbb{R})$$

at $x_0 \in \mathbb{R}$ is given by

$$\sum_{j=0}^{\infty} \frac{f^{(j)}(x_0)}{j!} (x - x_0)^j. \quad (4.8)$$

Here $f^{(j)}$ denotes the j -th (ordinary) derivative of f as usual:

$$f^{(j)} = \frac{d^j f}{dx^j}.$$

A function $f \in C^\infty(\mathbb{R})$ is said to be **real analytic** in the interval $I = (x_0 - r, x_0 + r)$ if the series in (4.8) converges for each $x \in I$ and

$$f(x) = \sum_{j=0}^{\infty} \frac{f^{(j)}(x_0)}{j!} (x - x_0)^j.$$

The set of real analytic functions is denoted by C^ω . Even for a function which is only in $C^k(a, b)$, the **Taylor approximation theorem** always holds. It is very useful and powerful.

Theorem 22 *If $f \in C^{k+1}(a, b)$ and $x_0 \in (a, b)$, then for any $x \in (a, b)$*

$$f(x) = \sum_{j=0}^k \frac{f^{(j)}(x_0)}{j!} (x - x_0)^j + R_k(x)$$

where $R_k = R_k(x, x^*)$ is the k -th order **Taylor remainder** given by

$$R_k(x) = \frac{f^{(k+1)}(x^*)}{(k+1)!} (x - x_0)^{k+1}$$

and x^* is some (unknown) point between x and x_0 . Generally, one can say x^* depends (in some complicated unknown way) on x , but it is often useful to consider R_n as a function of both x and x^* (and x_0 as well) since the dependence on each of these “variables” is explicitly known.

The question of whether a C^∞ function is real analytic is precisely the question of whether or not one has

$$\lim_{k \rightarrow \infty} R_k(x) = 0.$$

This will generally be true in some open interval $B_r(x_0) = (x_0 - r, x_0 + r)$, though for actual convergence, one can only ensure $r \geq 0$. If $r = 0$, then this (open) interval is empty. You still get trivial convergence for $x = x_0$, but of course that tells you nothing. More generally, using the Taylor approximation formula effectively boils down to getting an estimate (a uniform estimate on some interval) for the $k + 1$ order derivative.

The basic outline of these results holds in any dimension, though the set of convergence becomes a general ball $B_r(\mathbf{x}_0)$ and you need estimates for all partial derivatives of order $k + 1$. I will not state the results in detail.

The **Taylor expansion** of a function $u \in C^\infty(U)$ at $\mathbf{x}_0 \in U \subset \mathbb{R}^n$ is given by

$$\sum_{j=0}^{\infty} \sum_{|\beta|=j} \frac{D^\beta u(\mathbf{x}_0)}{\beta!} (\mathbf{x} - \mathbf{x}_0)^\beta. \quad (4.9)$$

In this expansion formula $\beta = (\beta_1, \beta_2, \dots, \beta_n)$ is a **multi-index**, which simply means

$$\beta \in \mathbb{N}^n = \{(m_1, \dots, m_n) : m_1, \dots, m_n \in \mathbb{N}\} \quad \text{where} \quad \mathbb{N} = \{0, 1, 2, 3, \dots\}.$$

The derivative $D^\beta u$ denotes the partial derivative taken β_j times with respect to x_j for each $j = 1, 2, \dots, n$:

$$D^\beta u = \frac{\partial^{|\beta|} u}{\partial x_1^{\beta_1} \partial x_2^{\beta_2} \dots \partial x_n^{\beta_n}}.$$

The “length” of a multi-index β is defined by

$$|\beta| = \sum_{j=1}^n \beta_j.$$

The **factorial** of a multi-index β is given by

$$\beta! = \beta_1! \beta_2! \cdots \beta_n!.$$

The **multi-index power** of a vector variable $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ is

$$\mathbf{x}^\beta = x_1^{\beta_1} x_2^{\beta_2} \cdots x_n^{\beta_n}.$$

A function $u \in C^\infty(\mathbb{R}^n)$ is said to be **real analytic** in the ball $B_r(\mathbf{p}) = \{\mathbf{x} \in \mathbb{R}^n : |\mathbf{x} - \mathbf{p}| < r\}$ if the series in (4.9) converges for each $\mathbf{x} \in B_r(\mathbf{p})$ and

$$u(\mathbf{x}) = \sum_{j=0}^{\infty} \sum_{|\beta|=j} \frac{D^\beta u(\mathbf{x}_0)}{\beta!} (\mathbf{x} - \mathbf{x}_0)^\beta.$$

The set of real analytic functions denoted by $C^\omega(U)$ where U is an open subset of \mathbb{R}^n consists of those functions u for which u is real analytic on some ball centered at each point $\mathbf{x}_0 \in U$.