MA 35 (Honors Calculus) Section 2 Lecture Notes

Jonathan Ben-Artzi

Fall 2008

The following notes are *by no means* complete, and cannot be used instead of lectures and the course textbook. Please let me know of any errors (which probably exist) and any diskrepancies in relation to the lectures and/or the textbook.

0 Introduction and Syllabus

Homework 0.1 (Reading due Friday 9/5; Problems due Monday 9/8)**.**
Read *§1.1 and* ***solve*** *1, 3, 5, 7, 9, 12, 15, 16, 25*
Read *§1.2 and* ***solve*** *2, 8, 14, 18*

0.1 Introduction

This is a beautiful yet difficult course. We will get a chance to see some theorems that are the corner stones of modern mathematics, physics and engineering. This course will require hard work - probably more hours outside the classroom than those spent in doors, each week. In addition, in this course you will be introduced to techniques of modern, rigorous mathematics, which will require some mathematical maturity.

If you find that this is more than you had signed up for, MA 18 and MA 20 cover (roughly) the same topics, but with less rigor.

0.1.1 What you have seen before

In previous calculus courses, whether in high school or in college, you have seen 1-dimensional theory. The main fields of calculus you have seen can be summarized as follows:

Differential calculus. If $y = f(x)$ is a certain quantity (function) that depends on some variable x (that could be time, position, etc...), we ask the question *"how rapidly does y change as x changes?"*

This brings us to the definition of the derivative,

$$f'(x) = \frac{dy}{dx} = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}.$$

Integral calculus. Given a certain function $y = f(x)$, we look for the area under the graph of f between two specific values of x, say a and b. This, for example, could be useful when trying to answer questions like *"what distance did the car cover over the last two hours?"*

To calculate this area, we introduced the notion of *Riemann sums*: We partition the segment $[a, b]$ into N segments $[a, x_1], [x_1, x_2], [x_2, x_3], ..., [x_{N-1}, b]$, denote $x_0 = a, x_N = b$, and only require $a = x_0 < x_1 < x_2 < x_3 < \cdots < x_{N-1} < x_N = b$. We then pick (randomly) a point $\xi_i \in [x_{i-1}, x_i]$. The N^{th} Riemann sum is then defined to be

$$S_{\{x_i\}}(N, \xi_i) = \sum_{i=1}^{N} f(\xi_i)(x_i - x_{i-1}),$$

which depends on N, on the partition $\{x_i\}$ and on the choice of ξ_i.

If for any partition $\{x_i\}$, whenever the partition *parameter* (i.e. the length of the longest segment $x_{k+1} - x_k$) tends to 0 as $N \to \infty$ we find that $S_{\{x_i\}}(N, \xi_i)$ tends to the same number $I(a, b)$, we say that this number $I(a, b)$ is the *integral* of the function $f(x)$ between a and b, and we denote:

$$\int_a^b f(x)dx := I(a, b) = \lim_{N\to\infty} \sum_{i=1}^{N} f(\xi_i)(x_i - x_{i-1}).$$

Relationship between the two. One of the most fundamental diskoveries of modern mathematics (by Newton and Leibniz, around 1700) was that these two problems are actually the inverses of one another. This is summarized in the following theorem:

Theorem 0.2 (Fundamental Theorem of Calculus)**.** *If $F'(x) = f(x)$, then*

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

0.1.2 What we will see in this course

In this course, we will expand these ideas beyond 1-dimension. Remember, that the world that surrounds us is a 3-dimensional world; we'd like to be able to say something about it. Furthermore, the theory of relativity looks at the world as *4-dimensional* (there's another dimension for time), and more modern theories (string thoery for example) regard the world as *26-dimensional.*

0.2 Syllabus

Topics will include:

1. *Differentiation*
   - Brief survey of Euclidean geometry, scalar and vector products.
   - Multivariate functions: graphical representation (surfaces), continuity.
   - Differentiation in two and three dimensions: partial derivatives, directional derivatives.
   - Gradients, tangent lines and planes.
   - Extremal problems.
   - Lagrange multipliers and constraints.
   - Higher order derivatives and Taylor's theorem.
   - The implicit function theorem, the inverse function theorem.
2. *Integration*
   - Brief survey of one dimensional integration.
   - Integration in two dimensions: Cartesian, polar.
   - Fubini's theorem.
   - Integration in three dimensions: Cartesian, cylindrical, spherical.
   - Change of variables: the Jacobian.
   - Geometrical applications: solid volumes, surface area, center of mass.
3. *Vector analysis*
   - Vector valued functions.
   - The divergence and the curl of a vector field.
   - Line integrals in two and three dimensions.
   - Green's theorem (in two dimensions).
   - Surface integrals.
   - Divergence theorem (Gauss' theorem).
   - Stokes' theorem.

Course website	`http://math.brown.edu/~yonib/Fall2008.MA35/index.html`
Book	J. Marsden and A. Tromba, *Vector Calculus*, Fifth edition.
Course hours	MWF 11:00-11:50 am.
Room	Barus and Holley 163.
Office hours	Thursdays & Fridays, 4:30-5:30 pm.
Grading	20% Mid term 1.
	25% Mid term 2.
	40% Final.
	15% Assignments and participation.
Mid term 1	October 06, 2008.
Mid term 2	November 10, 2008.
Final	December 16, 2008.

Homework rules:

It is crucial to give ample time to prepare your homework assignments. Assignments will be long and demanding, so start them early. Homework will be assigned each class, and the accumulated assignment is to be handed in each Monday. If, for some reason, you cannot hand in your homework on Monday, let me know the preceding Friday at the latest.

The assignments will be declared each class, but will also be clearly posted on the website. Be sure to always consult the site for the correct, and up-to-date information by clicking on the `course timeline` tab.

Getting help:

If you are having trouble with any of the course topics, you are encouraged to go to the Math Resource Center (MRC). There, you will find graduate students and advanced undergraduate students who can help you with notions from the course and homework.

MRC hours are 8 - 10 pm, Monday through Thursday, at Kassar 105. Directions and info can be found at: `http://www.math.brown.edu/mrc/`

Lecture notes:

I will try to prepare lecture notes throughout the semester. This is a service to you, and I cannot promise to keep it up. In any case, the notes can only be considered as *study aids*, and cannot replace the lectures and the book. You will be able to find them under the `course timeline` tab of the website.

My contact information:

My contact information can always be found on my website. The best way to contact me is by email, at `yonib@math.brown.edu`. My office is Kassar 012, and the office phone number is (401) 863 7956.

1 Brief survey of Euclidean geometry

Homework 1.1 (Reading due Monday 9/8).
__Read__ §1.3
__Read__ §1.5
We are skipping §1.4 for the moment, but you may want to familiarize yourself with it.

1.1 Vectors, Lines and Planes

1.1.1 Notation

Scalars, i.e. fixed numbers, will be written in *italics*

Points in the plane or in space will be written in *ITALIC CAPITALS*

Vectors will be written in **bold**

The origin will be denoted by O

The vector representing the point P will be written **OP**

Note that in class, since I can't write in bold-face on the board, those quantities that need to be in bold-face will have an arrow above them.

1.1.2 Vectors

Definition 1.2 (Vector). *A* vector *is a length and a direction.*

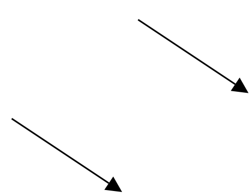


Figure 1: These two arrows represent the same vector

The best way to think of a vector, is to think of an arrow, which points is a certain direction, and has a certain length.

We normally do not care about the point at which the vector is based.

In the plane (which is denoted $\mathbb{R}^2 = \{(a,b) \mid a,b \in \mathbb{R}\}$), we can represent a vector as a pair of numbers - or **components**: the first representing the x-component, and the second representing the y-component. For example we can write $(3,-5)$ for the vector that has a length of 3 units along the x-axis, and -5 along the y-axis.

Similarly, in the "physical" space $\mathbb{R}^3 = \{(a,b,c) \mid a,b,c \in \mathbb{R}\}$, we represent a vector by 3 numbers. For example, $(3,-5,\mathbf{7})$.

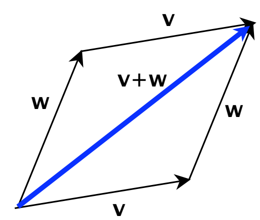


Figure 2: Vector addition

Vector operations. We can add vectors: $(a,b,c)+(a',b',c') = (a+a', b+b', c+c')$, and we can multiply a vector by a scalar: $t(a,b,c) = (ta,tb,tc)$. (By saying that t is a scalar, we simply mean that t is some *fixed* number)

If t is negative, the meaning is that we look at a vector that points in the opposite direction, and is of length $|t|$ times the original length of the vector.

Question 1.3. *How would the diagram for* $\mathbf{v}-\mathbf{w}$ *look like?*

Example 1.4. *In physics, we want to find the* net force *acting on an object. If the various forces acting are* $\mathbf{f_1}, \mathbf{f_2}, ..., \mathbf{f_n}$ *(note: force is a vector), then the net force would be* $\mathbf{f_{tot}} = \mathbf{f_1} + \mathbf{f_2} + \cdots + \mathbf{f_n}$.

A common notation (and an intuitive one) is to represent vectors by a **standard basis**.

Definition 1.5 (Standard Basis of $\mathbb{R}^3$)**.** *The* standard basis *of* $\mathbb{R}^3$ *is comprised of the three vectors*

$$\begin{aligned}\mathbf{i} &= (1,0,0)\\ \mathbf{j} &= (0,1,0)\\ \mathbf{k} &= (0,0,1).\end{aligned}$$

These three vectors are convenient choices; they are very intuitive. However, it is important to understand that there is nothing special about them. We could have chosen any three vectors that are *linearly independent*, i.e. that do not "point in the same direction". You will learn more about this term in a linear algebra course. For our purposes, the above definition suffices.

Representing points as vectors. There's a very natural duality between the notion of a point $P \in \mathbb{R}^3$ and a certain vector. Which vector? Well, when we sketch the x, y, z axes in the usual way (i.e. with the standard basis), we can put a dot that will represent P. This dot will be removed from the origin by some distance P_1 along the x-axis, P_2 along the y-axis and P_3 along the z-axis. We call these the x, y and z *coordinates* of the point P, respectively.

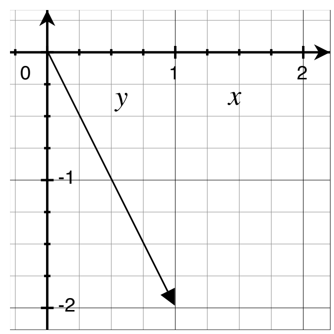


Figure 3: The vector **OP** that corresponds to the point $P = (1, -2)$

So, the vector $\mathbf{OP} = (P_1, P_2, P_3)$ can be viewed as representing the point $P \in \mathbb{R}^3$. The result is that the x-component of **OP** is the same as the x-coordinate of P, and similarly with y and with z.

It is important to note, however, that in this case it *is* important where this vector is based; it must be based at the origin for the "tip of the arrow" to be at the point P.

Example 1.6. *We may write our vector* $(3, -5, 7)$ *from above in the following way:*

$$\mathbf{v} = (3, -5, 7) = 3\mathbf{i} - 5\mathbf{j} + 7\mathbf{k}.$$

Now, we also know that this vector may also represent the point $P = (3, -5, 7) \in \mathbb{R}^3$.

Example 1.7 (For some vectors the basepoint matters, while for others it doesn't)**.** *Think of an airplane flying over your head. If you fix the origin where you are standing, the position of the airplane* $P = (P_1, P_2, P_3)$ *can be thought of as a point in space, but one could also think of the vector* $\mathbf{OP} = (P_1, P_2, P_3)$ *joining you and the airplane. In contrast, the vector representing the airplane's velocity,* $\mathbf{v} = (v_1, v_2, v_3)$ *is not really attached to any specific basepoint in space.*

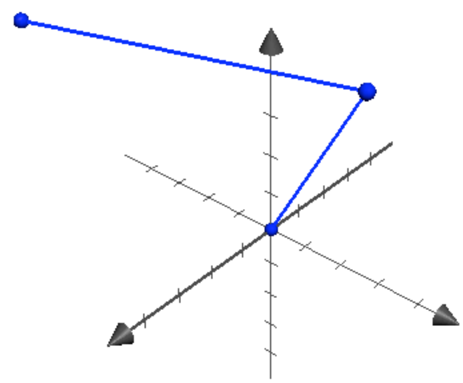


Figure 4: The airplane and its velocity

1.1.3 Lines

Definition 1.8 (Span)**.** *The* span *of a (nonzero) vector* $\mathbf{v}$*, denoted Span*$(\mathbf{v})$*, is the set* $C = \{t\mathbf{v} \quad | \quad t \in \mathbb{R}\}$ *of all vectors that are scalar multiples of* $\mathbf{v}$*.*

Definition 1.9 (Line)**.** *A* line *is a point together with a direction.*

The *point* can be represented as a vector. For example $\mathbf{OA} = (A_1, A_2, A_3)$. How should we represent the *direction*? We know that a vector is a direction and a length. If we lost the *'length'* part, we'd be in shape. This is indeed what we shall do!

To eliminate the *length* part of a vector $\mathbf{v} = (v_1, v_2, v_3)$ (here we assume that $\mathbf{v}$ is not the trivial 0-vector $\mathbf{0} = (0, 0, 0)$), we will look at the set of all vectors that are multiples of $\mathbf{v}$ by a scalar. I.e., we look at vectors that are twice as long as $\mathbf{v}$ (but point the same way), three times as long, 0.4577 as long, etc. Also, we look at vectors that are $-2, -0.98,$ and -10000 times as long. In other words, we look at $Span(\mathbf{v})$!

Note that the span of a (nonzero) vector is, in fact, a line that passes through the origin, when we think of the vector as being based at the origin.

We write the **parametric line equation** in the following way:

$$l(t) = \mathbf{OA} + t\mathbf{v}.$$

Taking the collection of all possible values of $t \in \mathbb{R}$ will give us the line, which we will usually denote by L.

1.1.4 Planes

Definition 1.10 (Plane). *A* plane *is a point together with two distinct directions.*

Similarly to the line representation, we can represent a plane as:

$$P(t, s) = \mathbf{OA} + t\mathbf{v} + s\mathbf{w},$$

again, assuming that $\mathbf{v}, \mathbf{w} \neq \mathbf{0}$, pointing in two different directions (i.e. $\mathbf{v} \notin Span(\mathbf{w})$ and $\mathbf{w} \notin Span(\mathbf{v})$; these two statements are actually equivalent. Think about it) and taking all possible $t, s \in \mathbb{R}$.

1.2 The Inner Product, Length and Distance

1.2.1 Inner Products

Definition 1.11 (Inner Product). *Let* $\mathbf{a} = (a_1, a_2, a_3)$ *and* $\mathbf{b} = (b_1, b_2, b_3)$ *be two vectors in* $\mathbb{R}^3$. *The* inner product *of* $\mathbf{a}$ *and* $\mathbf{b}$ *is defined to be*

$$\mathbf{a} \cdot \mathbf{b} := a_1 b_1 + a_2 b_2 + a_3 b_3.$$

It is **very important** to note that the result of the inner product is a number.

1.2.2 Length

Definition 1.12 (Length). *Let* $\mathbf{a} = (a_1, a_2, a_3)$ *be a vector in* $\mathbb{R}^3$. *The* length *of* $\mathbf{a}$ *is defined to be*

$$a = \|\mathbf{a}\| := \sqrt{\mathbf{a} \cdot \mathbf{a}} = \sqrt{a_1^2 + a_2^2 + a_3^2}$$

which is no more than the Pythagorean theorem*.*

Definition 1.13 (Normalizing a vector). *For any vector* $\mathbf{a} \neq 0$*, we define the* normalized *vector* $\hat{\mathbf{a}}$ *to be a vector of length* 1 *parallel to* $\mathbf{a}$. *It is given by the formula:*

$$\hat{\mathbf{a}} = \frac{\mathbf{a}}{\|\mathbf{a}\|}.$$

1.2.3 Distance

Definition 1.14 (Distance). *Let* $\mathbf{a} = (a_1, a_2, a_3)$ *and* $\mathbf{b} = (b_1, b_2, b_3)$ *be two vectors in* $\mathbb{R}^3$. *The* distance *between the endpoints of* $\mathbf{a}$ *and of* $\mathbf{b}$ *(when both have the same base point) is defined to be*

$$\|\mathbf{a} - \mathbf{b}\|.$$

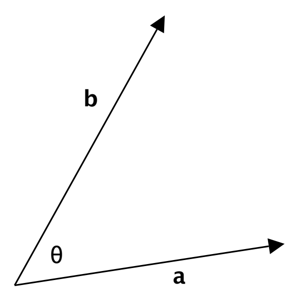
Similarly, the distance between two points P *and* Q *will be*

$$\|\mathbf{OP} - \mathbf{OQ}\|.$$

1.2.4 Results

Proposition 1.15.

$$\mathbf{a} \cdot \mathbf{b} = \|\mathbf{a}\|\|\mathbf{b}\| \cos\theta$$



Proof. In the book, pages 27-28. □

Definition 1.16. *Two vectors* $\mathbf{a}$ *and* $\mathbf{b}$ *are said to be* perpendicular (orthogonal) *if they form a right angle:* $\theta = \pi/2$.

Since $\cos\theta = 0$ if and only if $\theta = \pi/2$, we have the **important result**:

$$\mathbf{a} \text{ and } \mathbf{b} \text{ are perpendicular} \iff \mathbf{a} \cdot \mathbf{b} = 0$$

Projection of one vector onto another. The *orthogonal projection* of the vector $\mathbf{b}$ onto the (nonzero) vector $\mathbf{a}$ is given by the formula:

$$\mathbf{p} = \left(\frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\|^2}\right) \mathbf{a}$$

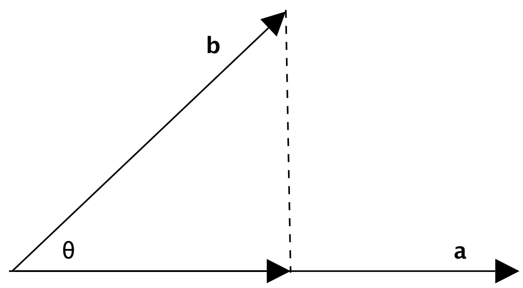


Figure 5: The projection $\mathbf{p}$ of $\mathbf{b}$ onto $\mathbf{a}$

Proof. First, for simplicity, assume that $-\pi/2 \leq \theta \leq \pi/2$.

Denote the projected vector by $\mathbf{p}$. Then, on the one hand, since $\mathbf{p}$ and $\mathbf{a}$ point in the same direction, there exists some number $t \in \mathbb{R}$ such that $\mathbf{p} = t\mathbf{a}$, and, thus $\|\mathbf{p}\| = |t|\|\mathbf{a}\|$. (Actually, because of our assumption on θ, $t \geq 0$, so that we can drop the absolute value).

On the other hand, $\|\mathbf{p}\|$ equals exactly $\|\mathbf{b}\| \cos\theta$ which is no more than $\mathbf{a} \cdot \mathbf{b}/\|\mathbf{a}\|$.

Summarizing, we have that

$$t\|\mathbf{a}\| = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\|}$$
$$t = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\|^2}$$

□

Question 1.17. *What happens if $\theta > \pi/2$?*

Some intuition. What we have just seen gives us a suggestion as to why define an inner product in the first place: *the inner product measures, in a sense, to what extent vectors fail to be orthogonal.* If the result of the inner product is 0, we know that the vectors are indeed orthogonal. Proposition 1.15 tells us the converse: the dot product is maximal when the two vectors point in the same direction (or most *negative* when they point in opposing directions).

Another way to represent a plane in $\mathbb{R}^3$. If $\mathbf{n} = (a, b, c)$ is a vector that is perpendicular to the plane (we usually say ***normal to the plane***) $M(t, s)$, and $P_0 = (p_1, p_2, p_3)$ is a fixed point on the plane, then any other point $P = (x, y, z)$ on the plane would satisfy:

$$0 = \mathbf{n} \cdot \mathbf{PP_0} = a(p_1 - x) + b(p_2 - y) + c(p_3 - z),$$

where the first equality holds since $\mathbf{PP_0}$ is a vector lying *inside* the plane, and $\mathbf{n}$ is perpendicular to the plane, i.e. to *all* vectors *in* the plane.

So we get that all points $P = (x, y, z)$ on the plane satisfy:

$$ax + by + cz = ap_1 + bp_2 + cp_3 =: d$$

(we usually just call the quantity on the right hand side d).

Conversely if $P \in \mathbb{R}^3$ is a point such that $\mathbf{n} \cdot \mathbf{PP_0} = 0$, then $\mathbf{PP_0}$ is perpendicular to $\mathbf{n}$ and therefore $\mathbf{PP_0}$ must lie in the plane, so that P is a point in the plane.

Conclusion: the equation $\mathbf{n} \cdot \mathbf{PP_0} = 0$ gives the set of all points in the plane and can therefore be called ***the equation of the plane***.

Remark 1.18. *A common convention which we will be using is that the vector* $\mathbf{n}$ *is of length* 1*. If it is not, we normalize it by dividing it by its length.*

Finding the distance between a point and a plane. Now we have the tools to also calculate the distance between a point and a plane. The *distance* between a plane and a point is defined to be the length of the shortest possible line joining them. Considering Figure 5, this distance will be given by the length of the vector joining the point and the plane, that *has no projection onto the plane*, or, in other words, *parallel to the vector* $\mathbf{n}$ *that is normal to the plane.*

Proposition 1.19. *The distance between the point* $E = (x_1, y_1, z_1)$ *and the plane* $ax + by + cz = d$ *is given by the equation:*

$$\textit{Distance} = \frac{|ax_1 + by_1 + cz_1 + d|}{\sqrt{a^2 + b^2 + c^2}}.$$

Proof. Read the proof in the book, page 53. □

1.3 Matrices, Determinants, and the Cross Product

Homework 1.20 (Reading due Wednesday 9/10; Problems due Monday 9/15).
Read *§2.1*
Solve *§1.3: 2, 3, 4, 13, 15, 16, 33* ***and bonus****: §1.5: 15*
Also *prove the properties of the cross product below.*

1.3.1 Matrices and Determinants

In general, matrices are no more than just arrays of numbers, that can have any number of rows and columns. However, usually the numbers within a matrix can have some meaning. We will begin seeing this today, but we will see this to its full extent when we learn about the *Jacobian* later in the course.

The matrices we will deal with right now will have the special property of having the same number of rows and columns. In general, however, there is no reason for those numbers to be equal.

2×2 **Matrices.** Probably the simplest matrix (that is not just one row by one column) is a matrix with two rows and two columns.

We usually write such a matrix as follows:

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

where $a_{11}, a_{12}, a_{21}, a_{22} \in \mathbb{R}$.

The **determinant** of A, denoted $\det A$ is defined to be

$$\det A = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} := a_{11}a_{22} - a_{12}a_{21}.$$

3×3 **Matrices.** In a similar manner, we may write the matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

where $a_{ij} \in \mathbb{R}$ for all $i, j \in \{1, 2, 3\}$.

The definition of the **determinant** is a bit more involved in this case, but not very hard to remember. We find the determinant by reducing to the 2×2 case, using a special case of what is usually called the *Laplace expansion*:

$$\begin{aligned}\det A &= \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \\ &:= a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ &= a_{11}\left(a_{22}a_{33} - a_{23}a_{32}\right) - a_{12}\left(a_{21}a_{33} - a_{23}a_{31}\right) + a_{13}\left(a_{21}a_{32} - a_{22}a_{31}\right).\end{aligned}$$

The three determinants $\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}$, $\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix}$ and $\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$ are called **minors** of A. We denote a minor according to the row and column missing in it. For example, we write

$$A_{32} = \begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix}.$$

Thus, we can now write the equation above as:

$$\det A = a_{11}A_{11} - a_{12}A_{12} + a_{13}A_{13}.$$

The basic rule for taking the determinant of a 3×3 matrix. In the equation above we have +'s and -'s interchanging before each summand. This comes from the general formula for taking the determinant of a 3×3 matrix:

1. Pick a single row or column.
2. Single out each element of your row/column.
3. Multiply each element with its corresponding minor.
4. Add up the three expressions according to these signs:

$$\begin{pmatrix} + & - & + \\ - & + & - \\ + & - & + \end{pmatrix}$$

Remark 1.21. *The notion of a determinant exists* only *for matrices that have the same number of rows and columns.*

1.3.2 The Cross Product

All the technical work we have invested defining the determinant, is going to help us calculate cross products. The cross product is an important tool, that will be used throughout this course, and is a basic notion both in mathematics and in physics.

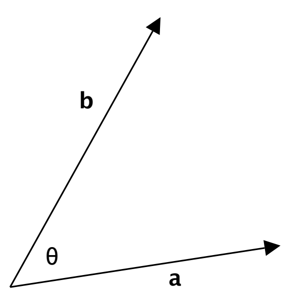
The cross takes two vectors in $\mathbb{R}^3$ and generates a third one, that has two important properties:

- It is perpendicular to the plane spanned by the two vectors.
- Its magnitude equals the area of the parallelogram the two vectors generate.

Definition 1.22 (Cross Product)**.** *If* $\mathbf{a} = (a_1, a_2, a_3), \mathbf{b} = (b_1, b_2, b_3)$*, then we define*

$$\begin{aligned}\mathbf{a} \times \mathbf{b} \;&:= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} \\ &= (a_2 b_3 - a_3 b_2)\,\mathbf{i} - (a_1 b_3 - a_3 b_1)\,\mathbf{j} + (a_1 b_2 - a_2 b_1)\,\mathbf{k}\end{aligned}$$

This is a formal definition, as we have defined the determinant to involve arrays of real numbers.



Important properties of the cross product. ***Prove these!***

1. Anti-commutativity: $\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a}$
2. $\mathbf{a} \times \mathbf{b} = 0 \iff \mathbf{a} = 0$ or $\mathbf{b} = 0$, or if they are parallel. In particular $\mathbf{a} \times \mathbf{a} = \mathbf{0}$
3. Distributivity over addition: $\mathbf{a} \times (\mathbf{b} + \mathbf{c}) = (\mathbf{a} \times \mathbf{b}) + (\mathbf{a} \times \mathbf{c})$
4. $\mathbf{a} \times \mathbf{b}$ is perpendicular to both $\mathbf{a}$ and $\mathbf{b}$.
5. $\|\mathbf{a} \times \mathbf{b}\| = \|\mathbf{a}\|\|\mathbf{b}\||\sin\theta|$, which is the area of the parallelogram spanned by $\mathbf{a}$ and $\mathbf{b}$ (prove both of these statements).

6. The *triple product*: $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}$, which is the volume of the $3-d$ parallelogram spanned by $\mathbf{a}, \mathbf{b}$ and $\mathbf{c}$.
7. The *Jacobi identity*: $\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) + \mathbf{b} \times (\mathbf{c} \times \mathbf{a}) + \mathbf{c} \times (\mathbf{a} \times \mathbf{b}) = 0$.

At this point we skip §1.4.

1.5 n-Dimensional Euclidean Space

We will cover this chapter very rapidly, as it will not come up much in the remainder of the course. This is, however, and important section on the *intellectual* level: understanding that it is possible to talk of dimensions greater than 3, and utilizing these spaces was quite a big leap in the development of modern mathematics.

Important things to remember

- Vectors and matrices can be of arbitrary sizes. We would usually denote

$$\begin{aligned} \mathbf{v} &= (v_1, v_2, ..., v_n) \in \mathbb{R}^n \\ A &= \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \in \mathbb{R}^{m \times n} \end{aligned}$$

  vectors with an array of n numbers, and matrices with an array of $m \times n$ numbers, respectively.

- Vector addition remains the same:

$$\begin{aligned} \mathbf{v} + \mathbf{w} &= (v_1, v_2, ..., v_n) + (w_1, w_2, ..., w_n) \\ &= (v_1 + w_1, v_2 + w_2, ..., v_n + w_n). \end{aligned}$$

- The *dot product* of two vectors $\mathbf{v}$ and $\mathbf{w}$ with an array of size n each, is given by

$$\mathbf{v} \cdot \mathbf{w} = v_1 w_1 + v_2 w_2 + \cdots + v_n w_n = \sum_{i=1}^{n} v_i w_i$$

- There is *no* cross product in higher dimensions.

Please refer to the book for more information. This section has a lot of basic notions which are very important for mathematicians, but will be of less importance for us in this course.

2 Differentiation

Homework 2.1 (Reading due Friday 9/12; Problems due Monday 9/15)**.**
Read *§2.2*
Solve *§2.1: 1, 2, 4, 9, 13, 15, 19, 25, 26*

This chapter, the second of four, will take us about one month to cover. Here we will extend the notion of a *derivative* to many dimensions. We will see that the new dimensions give rise to new properties of derivatives, which do not exist in one-dimensional differentiation, such as the **gradient**.

2.1 Introduction to Functions of Several Variables

2.1.1 Real functions of a single variable

From previous calculus courses, we know how to plot what we call a ***scalar valued, single variable*** function. We denote such a function by

$$f : \mathbb{R} \longrightarrow \mathbb{R},$$

a function from the *reals* to the *reals*.

Domain. For some functions, we have only a limited domain; not all numbers are assigned a value by the function. For example, the square root is defined only for non-negative numbers. We would write that fact in the following way:

$$\sqrt{\ } : \mathbb{R}_+ \subseteq \mathbb{R} \longrightarrow \mathbb{R}.$$

Range. The range of a function is the set of all possible values it achieves. For example, we can write

$$Range(\sin) = [-1, 1]$$
$$Range(x^2) = \mathbb{R}_+ = [0, \infty).$$

Continuity. We say that the function $f(x)$ is continuous at the point x_0 in its domain, if $f(x) \to f(x_0)$ as $x \to x_0$.

Intermediate value property. This important property, states something that intuitively is very clear: if $f(x)$ is a continuous function on the interval $[a, b]$, and if $f(a) < f(b)$, then for any number $f(a) < d < f(b)$ there exists a point $a < c < b$ such that $f(c) = d$. This basically means that f does not skip over any value (otherwise it wouldn't be continuous).

Derivatives. The formal definition of the derivative appears earlier in the notes, and I will not repeat it at the moment. Intuitively, the derivative gives us the value of the slope of a function at each point (if the function is differentiable, of course).

Applications. All these properties combined, enable us to plot functions, find their minima, maxima, intersection with the axes, etc.

2.1.2 Real functions of many variables

Now we attempt to extend our understanding to functions that depend on several variables. Some examples may be:

- Temperature.
- Elevation.
- The average grade for the last homework assignment.
- The highest grade for the last homework assignment.
- The lowest grade for the last homework assignment.

How many variables does each of these have?

Example 2.2. *Consider a mountain with water flowing over it. The mountain's contour specifies the elevation at each point. It is a real valued function of* two *variables: x and y of the plane on which the mountain is lying. However, the water flow is a different*

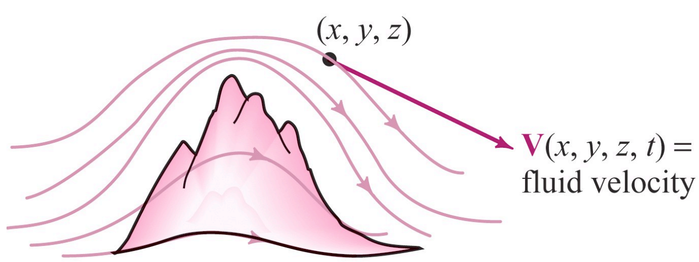


Figure 6: The mountain's contour is a scalar-valued function while the velocity of water flowing off the mountain is a vector-valued function

type of function which we will diskuss more later: instead of having a scalar value at each point of the mountain, the water flow is a vector: it has a direction and intensity at each point. Thus, the water flow is a function of the two variables x and y, that gives three *numbers. We call this a* **vector-valued function**.

For the time being, however, we focus on ***scalar-valued functions of several variables***.

In general, we will denote a scalar (real) valued function of n different variables by

$$f : \mathbb{R}^n \longrightarrow \mathbb{R}.$$

Domain. If the function has a certain domain $U \subseteq \mathbb{R}^n$, we write

$$f : U \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}.$$

Example 2.3. *For example, we may have some domain U in the plane.*

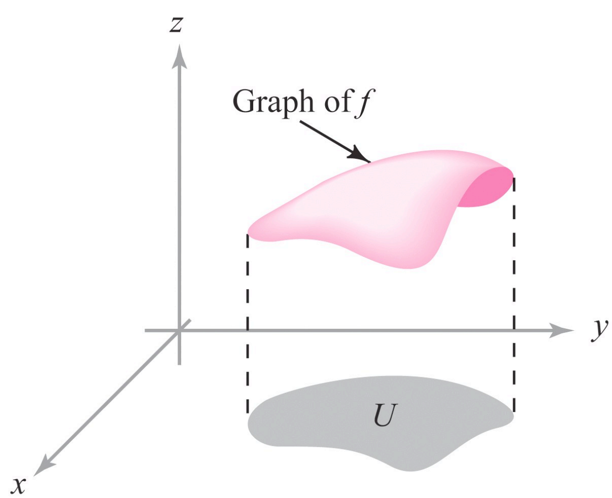


Figure 7: Domain $U \subseteq \mathbb{R}^2$

Range. We can also talk of the range of such functions.

Example 2.4. *For example, the range of $g(x) = x^2$ is $[0, \infty)$, the range of $h(y) = -y^2$ is $(-\infty, 0]$, but the range of $f(x, y) = g(x) + h(y) = x^2 - y^2$ is all of $\mathbb{R}$.*

Continuity. The question of continuity becomes tough in the multivariate case. While when diskussing functions of *one* variable we had two possible directions to approach any point x_0 - from the left or from the right, in *several* dimensions we have many more ways to approach x_0. In fact, we have infinitely many ways, even in the two-variable case.

For this reason, we will have to define the notion of an ***open set*** when we get to talk about continuity.

Intermediate value property. Here too, the definition will become more sophisticated, buy very intuitive: if $f(x, y)$ is a continuous function of two variables, our two selected points are (x_0, y_0) and (x_1, y_1) and suppose $f(x_0, y_0) < f(x_1, y_1)$, then we can apply the original intermediate value property for any path connecting the two points.

Example 2.5 (Intermediate value property on the plane $\mathbb{R}^2$)**.** *Suppose at a given moment the temperature in San Francisco is 60F and the temperature in Los Angeles is 90F. Then* no matter how *we drive from SF to LA, we will drive through a point where the temperature is 73F.*

Here we assume that temperature changes continuously, and that our car leaves a continuous trail as it travels.

Derivatives. This is the main goal of the current chapter, and we will not dive into this topic just yet. Try to think, however, how one could characterize change of a function of several variables. To visualize, imagine yourself standing on the slope of a mountain: what do you see when you look in different directions? Is there a direction where the descent/ascent is steepest?

Level sets. Although the properties mentioned thus far are enough to enable us to characterize a given function of several variables, we have some more tools that help us, that do not exist in functions of a single variable.

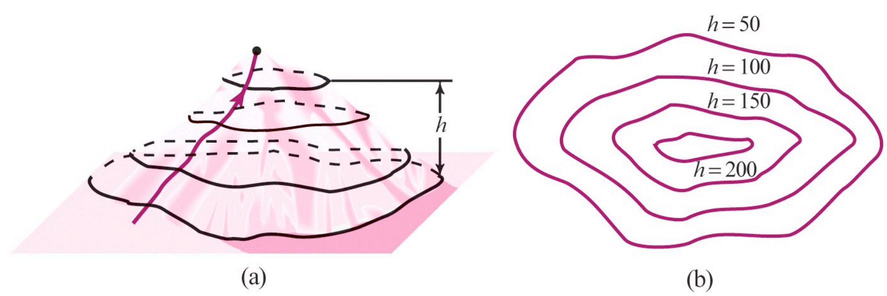


Figure 8: Level sets

We run into level sets in everyday life: when we read a weather map and when we look at a topographic map.

Informally, a level set of a function $f : \mathbb{R}^2 \to \mathbb{R}$ is the set of all points *in the plane* for which the function has a certain value.

Formally, we define:

Definition 2.6 (Level set)**.** *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be a function. The* level set of f with value c *is the set*

$$L_c = \{\mathbf{x} \in \mathbb{R}^n \mid f(\mathbf{x}) = c\} \subseteq \mathbb{R}^n.$$

2.1.3 Extending the definition of a function

Now that we have a fairly good understanding of what real valued functions are, we may generalize them to ***vector-valued functions***. A general vector-valued function will be written as

$$f : U \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}^m,$$

where $U \subseteq \mathbb{R}^n$ is the domain. Thus, for every point $\mathbf{x} \in U$, the function gives us an array of m numbers, say $f(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), ..., f_m(\mathbf{x}))$. Note that the point $\mathbf{x}$ itself is an array of n numbers, say $\mathbf{x} = (x_1, x_2, ..., x_n)$.

Example 2.7. *The function that describes wind in a room with a fan on would be a function $v : U \subseteq \mathbb{R}^3 \to \mathbb{R}^3$, where U is the subset of $\mathbb{R}^3$ which the room occupies. The function v assigns* to each point $\mathbf{p} = (x, y, z)$ in the room *three numbers: $v(\mathbf{p}) = (v_x(\mathbf{p}), v_y(\mathbf{p}), v_z(\mathbf{p}))$ which make up a* vector *that specifies how strong the wind is at that point, and in which direction it is pointing.*

Example 2.8. *When describing water flow down a mountain, at each point of the mountain we have a vector specifying the direction and intensity of the water flow at that point. The point itself is only determined by its x and y coordinates. The z coordinate does not matter, because for each point in the x, y plane there is a unique corresponding point that is exactly on the mountain's surface.*

Thus, our function in this case would be $v : U \subseteq \mathbb{R}^2 \to \mathbb{R}^3$.

2.2 Limits and Continuity

In this section we will really start to get a taste of what modern, rigorous math is.

Homework 2.9 (Reading and Problems due Monday 9/15).
Read *§2.2 again*
Solve *§2.2: 1, 4*

2.2.1 Open sets in $\mathbb{R}^n$

Our goal is to extend the notion of continuity from real functions of a single variable, to vector valued functions of many variables.

When diskussing real functions of a single variable, the question of "continuity" of a function $f(x)$ at a given point x_0, reduced to checking the limits of f as we approach x_0 from the left and from the right.

In higher dimension, however, we no longer have just *two* ways of approaching a point; we can approach any point in *infinitely* many fashions.

For this reason we introduce the notion of an open set.

First, though, we introduce the following definition: if $\mathbf{x_0} \in \mathbb{R}^n$ is a point, and $r > 0$ is a positive number, then we denote by $D_r(\mathbf{x_0})$ the set

$$D_r(\mathbf{x_0}) = \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x} - \mathbf{x_0}\| < r\}.$$

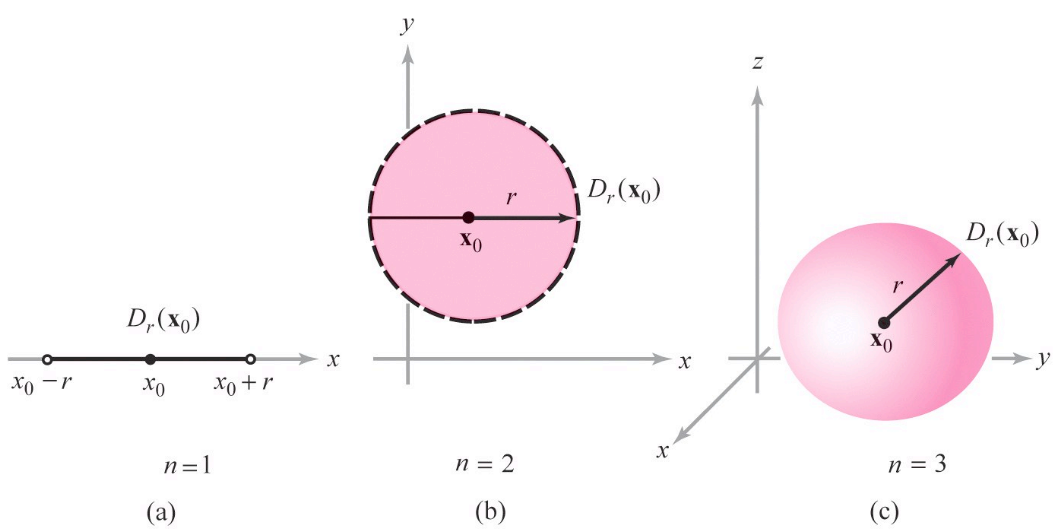


Figure 9: The set $D_r(\mathbf{x_0})$ in dimensions 1, 2 and 3

Definition 2.10 (Open set)**.** *We say that* $U \subseteq \mathbb{R}^n$ *is an* **open** *subset of* $\mathbb{R}^n$*, if for each* $\mathbf{x_0} \in U$*, there exists an* $r > 0$ *such that* $D_r(\mathbf{x_0}) \subseteq U$*.*

Proposition 2.11. *The set* $D_r(\mathbf{x_0})$ *is open for all* $r > 0$ *and* $\mathbf{x_0} \in \mathbb{R}^n$*. We call it the* **open ball** *of radius* r *about* $\mathbf{x_0}$*.*

Proof. We want to prove that $D_r(\mathbf{x_0})$ is open. For this, we have to go back to the definition to understand what it is that we exactly need to prove. The set that is called U in the definition above, is $D_r(\mathbf{x_0})$ in this case. By definition of an open set, we want to show that for any choice of a point $\mathbf{x} \in D_r(\mathbf{x_0})$, we can find *another* ball, with possibly another radius s, that will be centered at $\mathbf{x}$, and completely contained in $D_r(\mathbf{x_0})$.

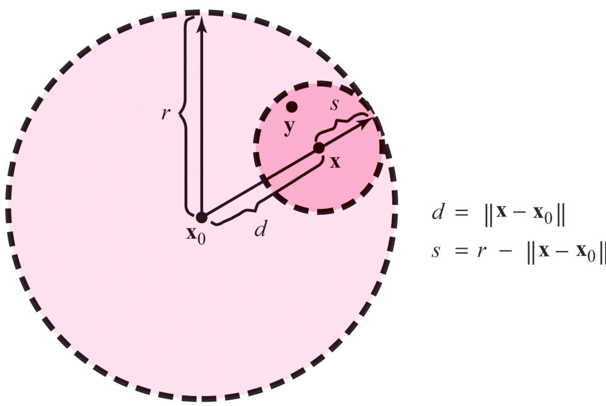


Figure 10: $D_r(\mathbf{x_0})$ is open

Thus we proceed as follows:

We know that $\|\mathbf{x} - \mathbf{x_0}\| < r$ by definition of $D_r(\mathbf{x_0})$. Denote that distance by d.

Since $d < r$, we define $s = r - d > 0$. This s will end up being the radius we are looking for (see Figure 10).

Now that we have $\mathbf{x}$ and we have s that we think will work, we consider the ball $D_s(\mathbf{x})$ of radius s centered at $\mathbf{x}$. We want to show that this ball is, in fact, contained in the original ball $D_r(\mathbf{x_0})$.

To do this, we take any point $\mathbf{y} \in D_s(\mathbf{x})$, and we show that $\mathbf{y} \in D_r(\mathbf{x_0})$ as well. This implies (by definition of *containment*) that $D_s(\mathbf{x}) \subseteq D_r(\mathbf{x_0})$, which will finish the proof.

How can we show that $\mathbf{y} \in D_r(\mathbf{x_0})$? Well, $D_r(\mathbf{x_0})$ is defined to be the set of all points distanced up to (and not including) distance r from $\mathbf{x_0}$. So we need to show that $\|\mathbf{y} - \mathbf{x_0}\| < r$.

This follows directly from the *triangle inequality*:

$$\|\mathbf{y} - \mathbf{x_0}\| = \|(\mathbf{y} - \mathbf{x}) + (\mathbf{x} - \mathbf{x_0})\| \leq \|\mathbf{y} - \mathbf{x}\| + \|\mathbf{x} - \mathbf{x_0}\| < s + d = r.$$

So we indeed found that $\|\mathbf{y} - \mathbf{x_0}\| < r$, which finished the proof. $\square$

Some intuition. The intuition for what an open set is, is simple. Just think of the one-dimensional case and recall the definition of an *open interval*, (a, b). Any $x \in (a, b)$ satisfies $a < x < b$, i.e. x is between a and b, but doesn't equal either of them. So, x always has some distance from both a and b.

The same intuition applies in higher dimensions: $U \subseteq \mathbb{R}^n$ is open, if for any point $\mathbf{x_0} \in U$ there is some positive (*nonzero*) distance from the boundary of U.

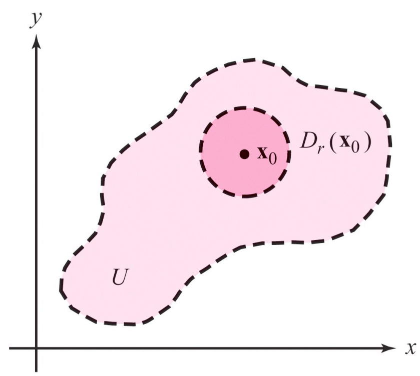


Figure 11: An open set

Neighborhoods. A neighborhood of a point $\mathbf{x} \in \mathbb{R}^n$ is an extremely important concept, and is used in all fields of mathematics all the time. The definition itself is very simple: $U \subseteq \mathbb{R}^n$ is a ***neighborhood*** of a given point $\mathbf{x} \in \mathbb{R}^n$ if U is an open set containing $\mathbf{x}$.

Boundary. Let us formally define the boundary of a set $A \subseteq \mathbb{R}^n$ (A is not necessarily open):

Definition 2.12 (Boundary)**.** *A point* $\mathbf{x} \in A$ *is called a* **boundary point** *of* A*, if every neighborhood of* $\mathbf{x}$ *contains at least one point in* A *and at least one point not in* A*. The* **boundary** *of* A*, denoted* ∂A*, is the set of all such points.*

Homework 2.13 (Reading due Wednesday 9/17; Problems due Monday 9/22)**.**
Read *§2.3*
Solve *§2.2: 5(b), 12(a,b), 15, and bonus: 9(a,b), 26*
Remember l'Hôpital!

2.2.2 Limits of vector-valued functions with domain $= \mathbb{R}^n$

Now that we have set up our basic definitions, we are ready to define the notion of a limit, and later of continuity.

With both notions we begin by defining only for functions with domain being the whole space $\mathbb{R}^n$. Later, we will extend the definitions to take into account different domains. The difference in the definitions will only lie in the special care we will have to give to boundary points.

Definition 2.14 (Limit of a vector-valued function defined on all of $\mathbb{R}^n$)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}^m$, and let $\mathbf{x_0} \in \mathbb{R}^n$. We say that $f(\mathbf{x})$* **tends to b as x tends to $\mathbf{x_0}$** *if for any choice of a neighborhood $N \subseteq \mathbb{R}^m$ of $\mathbf{b}$, there exists a neighborhood $U \subseteq \mathbb{R}^n$ of $\mathbf{x_0}$ such that for any $\mathbf{x_0} \neq \mathbf{x} \in U$, $f(\mathbf{x}) \in N$. We write*

$$\mathbf{b} = \lim_{\mathbf{x} \to \mathbf{x_0}} f(\mathbf{x}).$$

If $f(\mathbf{x})$ does not approach any particular value as $\mathbf{x} \to \mathbf{x_0}$, we say that **the limit $\lim_{\mathbf{x} \to \mathbf{x_0}} f(\mathbf{x})$ does not exist***.*

Remark 2.15. *We only consider points $\mathbf{x} \neq \mathbf{x_0}$. It may be the case that $f(\mathbf{x_0}) \neq \mathbf{b}$.*

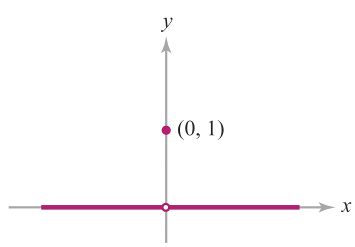


Figure 12: The limit doesn't have to equal the value at the limiting point

Properties of limits. The properties we know from functions of a single variable hold for vector-valued functions of several variables as well.

Theorem 2.16 (Uniqueness of Limits)**.** *If $\lim_{\mathbf{x} \to \mathbf{x_0}} f(\mathbf{x}) = \mathbf{b_1}$ and $\lim_{\mathbf{x} \to \mathbf{x_0}} f(\mathbf{x}) = \mathbf{b_2}$, then $\mathbf{b_1} = \mathbf{b_2}$.*

Theorem 2.17 (Properties of Limits)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}^m,\ g : \mathbb{R}^n \to \mathbb{R}^m,\ \mathbf{x_0} \in \mathbb{R}^n,\ \mathbf{b} \in \mathbb{R}^m,\ c \in \mathbb{R}$. Then the following hold:*

1. $\lim_{\mathbf{x}\to\mathbf{x_0}} f(\mathbf{x}) = \mathbf{b}$ *implies* $\lim_{\mathbf{x}\to\mathbf{x_0}} cf(\mathbf{x}) = c\mathbf{b}$.
2. $\lim_{\mathbf{x}\to\mathbf{x_0}} f(\mathbf{x}) = \mathbf{b_1}$ *and* $\lim_{\mathbf{x}\to\mathbf{x_0}} g(\mathbf{x}) = \mathbf{b_2}$ *imply* $\lim_{\mathbf{x}\to\mathbf{x_0}} (f+g)(\mathbf{x}) = \mathbf{b_1} + \mathbf{b_2}$.
3. $m = 1$, $\lim_{\mathbf{x}\to\mathbf{x_0}} f(\mathbf{x}) = b_1$ *and* $\lim_{\mathbf{x}\to\mathbf{x_0}} g(\mathbf{x}) = b_2$ *imply* $\lim_{\mathbf{x}\to\mathbf{x_0}} (fg)(\mathbf{x}) = b_1 b_2$.
4. $m = 1$, $\lim_{\mathbf{x}\to\mathbf{x_0}} f(\mathbf{x}) = b \neq 0$ *and* $f(\mathbf{x}) \neq 0$ *in a neighborhood of* $\mathbf{x_0}$, *imply* $\lim_{\mathbf{x}\to\mathbf{x_0}} (1/f)(\mathbf{x}) = 1/b$.
5. *If* $f(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), ..., f_m(\mathbf{x}))$ *are the scalar-valued components of* f, *then* $\lim_{\mathbf{x}\to\mathbf{x_0}} f(\mathbf{x}) = \mathbf{b} = (b_1, b_2, ..., b_m)$ *if and only if* $\lim_{\mathbf{x}\to\mathbf{x_0}} f_i(\mathbf{x}) = b_i$ *for* $i = 1, 2, ..., m$.

2.2.3 Continuity of vector-valued functions with domain $= \mathbb{R}^n$

Remark 2.15 and Figure 12 naturally bring us to the definition of continuity. As intuition suggests, the definition of continuity will simply state that indeed $f(\mathbf{x_0}) = \mathbf{b}$. It is always useful to think of functions $f : \mathbb{R} \to \mathbb{R}$ which we know from before, while keeping in mind that new phenomena may occur in higher dimensions.

Definition 2.18 (Continuous vector-valued function defined on all of $\mathbb{R}^n$)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}^m$, and let $\mathbf{x_0} \in \mathbb{R}^n$. We say that f* **is continuous at $\mathbf{x_0}$** *if and only if*

$$\lim_{\mathbf{x}\to\mathbf{x_0}} f(\mathbf{x}) = f(\mathbf{x_0}).$$

We say that f is **continuous** *if it is continuous at any point $\mathbf{x_0} \in \mathbb{R}^n$.*

Properties of continuous functions. The same properties we know from our first calculus course for functions of a single variable, apply here, for vector-valued functions of many variables. For the diskussion below we let $f : \mathbb{R}^n \to \mathbb{R}^m$, $g : \mathbb{R}^n \to \mathbb{R}^m$ and c a real number. Be aware that sometimes we will require $m = 1$.

1. If f is continuous at $\mathbf{x_0}$, so is cf.
2. If f and g are continuous at $\mathbf{x_0}$, so is $f + g$.
3. If f and g are continuous at $\mathbf{x_0}$ and $m = 1$, then fg is continuous at $\mathbf{x_0}$.
4. If $m = 1$ and f is continuous at $\mathbf{x_0}$ with $f(\mathbf{x_0}) \neq 0$, then $1/f$ is defined on a neighborhood of $\mathbf{x_0}$, and f is continuous at $\mathbf{x_0}$.
5. If $f(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), ..., f_m(\mathbf{x}))$, then f is continuous at $\mathbf{x_0}$ if and only if each of the f_i is continuous at $\mathbf{x_0}$ for $i = 1, 2, ..., m$.

The last property is perhaps the most important, since it tells us that vector-valued functions can be reduced to scalar-valued functions when it comes to continuity.

Composition of continuous functions. What we know for functions from $\mathbb{R}$ to $\mathbb{R}$, applies here too, only that we need to adjust the dimensions of the spaces:
Let $g : \mathbb{R}^n \to \mathbb{R}^m$, $f : \mathbb{R}^m \to \mathbb{R}^p$, and suppose that f is defined on $g(\mathbf{x_0})$. Then, if g is continuous at $\mathbf{x_0}$ and f is continuous at $g(\mathbf{x_0})$, then $f \circ g$ is continuous at $\mathbf{x_0}$.

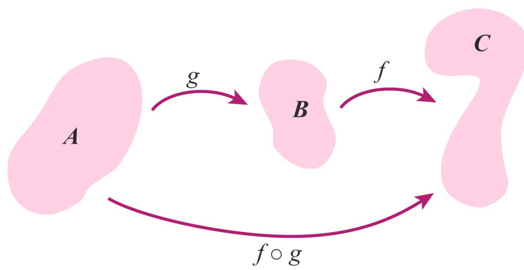


Figure 13: Composition of functions

2.2.4 Extending the definitions to general domains

After getting a pretty good understanding of how limits are defined for vector-valued functions, and what it means for such a function to be continuous at a given point, we extend the definition to include functions defined only on a *subset* of the space $\mathbb{R}^n$.

The issue we need to address in this case is the issue of *boundary points:* we can talk about the limit of a function as we approach a boundary point, even if that point is not in the domain. However, for a function to be continuous at a point, it needs to be defined at that point, so that the point must belong to the domain.

In the diskussion to follow, we let $f : A \subseteq \mathbb{R}^n \to \mathbb{R}^m$ be a function with domain A.

Interior points. Interior points are points $\mathbf{x} \in A$ that are not on the boundary. For such points, all of our previous analysis holds, since each such point has a neighborhood $U_\mathbf{x}$ contained in A: $\mathbf{x} \in U_\mathbf{x} \subseteq A$. Think why this is enough (hint: read the next paragraph).

Boundary points. Boundary points are a bit more problematic. We need to tweak all of our definitions for limits and continuity when dealing with such points, to make sure that we look at the sides of the neighborhoods that are *in* the domain. I.e., in Definition 2.14, we need to make sure that the point $\mathbf{x} \in U$ that is chosen is also in the domain so that the value $f(\mathbf{x})$ makes sense.

Read the definitions in the book, and compare to the definitions above.

2.3 Differentiation

Homework 2.19 (Reading due Friday 9/19; Problems due Monday 9/22).

Read *§2.4*

Solve *§2.3: 1(c), 2(a), 3(a), 6(c), 8(d), 18*

The notions we define in this section are very *natural* extensions of what we already know for functions $f : \mathbb{R} \to \mathbb{R}$.

As a first step, we think of functions $f : \mathbb{R}^2 \to \mathbb{R}$; an example of such a function is the surface of a mountain, where the elevation of each point is given by $z = f(x, y)$. Standing at a point $(x, y, f(x, y))$ on the mountain, the slope varies as we look around us in different directions.

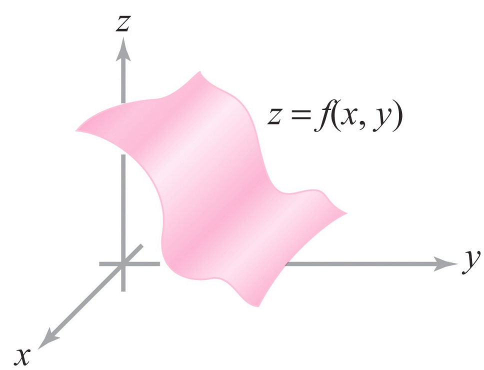


Figure 14: Surface of a mountain

Calculating the slope in any given direction, is achieved by calculating the *directional derivative* in that direction.

2.3.1 Partial Derivatives

We begin by a baby version of this *directional derivative* notion: the derivatives in the x and y directions.

Standing on the mountain, these so-called *partial derivatives* give us the slopes in two specific directions: east-west and north-south.

Given $f : \mathbb{R}^2 \to \mathbb{R}$, the ***partial derivative of*** f ***with respect to the*** x***-axis at the point*** $\mathbf{p} = (x_0, y_0)$, denoted $f_x(\mathbf{p})$ or $f_x(x_0, y_0)$ or $\frac{\partial f}{\partial x}(x_0, y_0)$, is defined as the infinitesimal change of f at (x_0, y_0) along the x-axis:

$$f_x(\mathbf{p}) = f_x(x_0, y_0) = \frac{\partial f}{\partial x}(x_0, y_0) = \lim_{h\to 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h},$$

if the limit exists.

Now, since $x_0 + h$ is a small increment in the $\mathbf{x}$ direction, we recall that $\mathbf{i}$ is our notation for a unit vector in that direction, so that the last expression may be written as

$$\lim_{h\to 0} \frac{f(\mathbf{p}+h\mathbf{i}) - f(\mathbf{p})}{h}.$$

The exact same process may be applied in the y-direction, so we can write

$$f_y(\mathbf{p}) = f_y(x_0, y_0) = \frac{\partial f}{\partial y}(x_0, y_0) = \lim_{h\to 0} \frac{f(x_0, y_0 + h) - f(x_0, y_0)}{h} = \lim_{h\to 0} \frac{f(\mathbf{p}+h\mathbf{j}) - f(\mathbf{p})}{h}.$$

Generalization to general scalar-valued functions. There is, of course, no reason for this definition not to hold for functions of more variables.

Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ be a scalar-valued function, with U open. Let $\mathbf{e_1}, \mathbf{e_2}, ..., \mathbf{e_n}$ denote the standard basis vectors in $\mathbb{R}^n$ (these are the n-dimensional analogues of the $\mathbf{i}, \mathbf{j}, \mathbf{k}$ we defined in 3-dimensions), i.e. $\mathbf{e_j}$ is a vector of length 1 pointing in the j^{th} direction.

Let $\mathbf{x} = (x_1, x_2, ..., x_n) \in \mathbb{R}^n$ be a point in $\mathbb{R}^n$. The ***partial derivative of*** f ***in the*** j^{th} ***direction at the point*** $\mathbf{p} = (p_1, p_2, ..., p_n)$ is defined by

$$\frac{\partial f}{\partial x_j}(p_1, p_2, ..., p_n) = \lim_{h\to 0} \frac{f(p_1, ..., p_j + h, ..., p_n) - f(p_1, ..., p_j, ..., p_n)}{h} = \lim_{h\to 0} \frac{f(\mathbf{p}+h\mathbf{e_j}) - f(\mathbf{p})}{h}$$

if the limit exists.

2.3.2 The Linear Approximation and the Tangent Plane

From our previous calculus course, we know that for a function $f : \mathbb{R} \to \mathbb{R}$ we have several ways to approximate it around a given point $x_0 \in \mathbb{R}$:

0. The so-called *0th approximation* is the best approximation if we only know the value of f at x_0: we approximate $f(x) = f(x_0)$ for all x.

1. The next type of approximation, the so-called *1st approximation*: if we also know $f'(x_0)$ we approximate $f(x)$ by a the linear function $f(x_0) + f'(x_0)(x - x_0)$. **This is the basis of what we do in this section**.

2. Next, we have the so-called *2nd approximation*: if we also know $f''(x_0)$ we approximate $f(x)$ by the parabola $f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2$.

And so on... In the end we get a *Taylor series* (if the series converges). The concept of a Taylor series in higher dimensions will be diskussed later.

In this section, we diskuss the multivariate version of the first (linear) approximation. In higher dimensions this approximation is also given by a linear function.

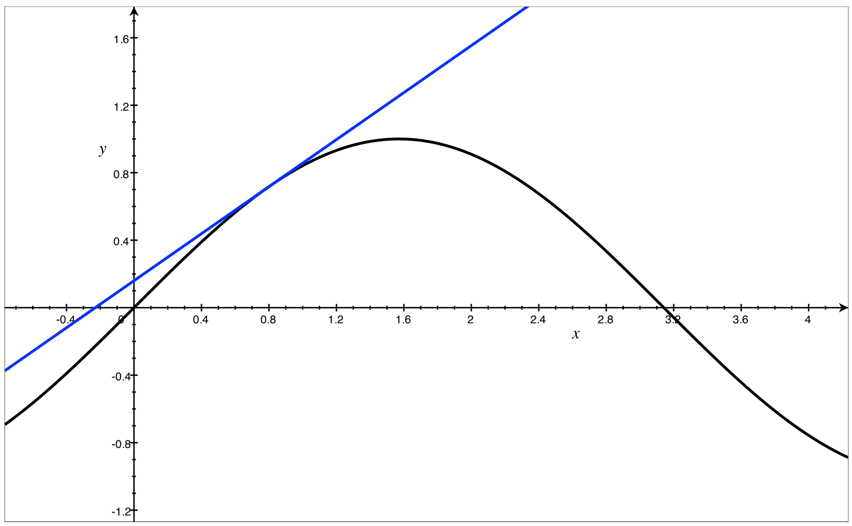


Figure 15: Tangent line: linear approximation for the function $\sin x : \mathbb{R} \to \mathbb{R}$ at $x = 0.8$

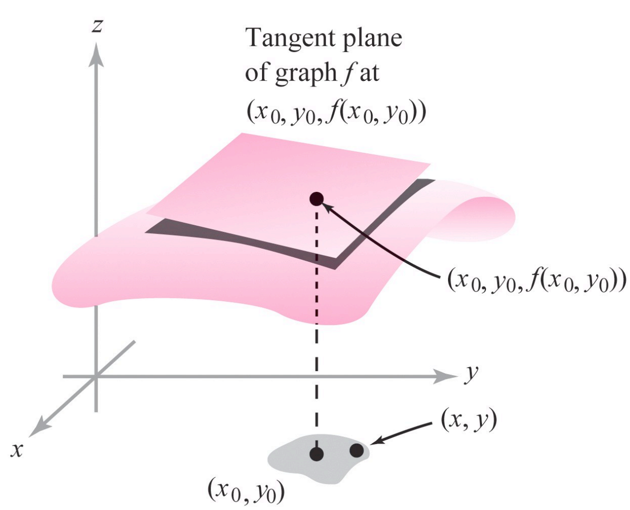


Figure 16: Tangent plane: linear approximation in higher dimensions

Thus, we find the formula for the ***linear approximation*** of a scalar-valued function of *two* variables $f : \mathbb{R}^2 \to \mathbb{R}$ around (x_0, y_0):

$$z = f(x_0, y_0) + \left[\frac{\partial f}{\partial x}(x_0, y_0)\right](x - x_0) + \left[\frac{\partial f}{\partial y}(x_0, y_0)\right](y - y_0).$$

We have seen that such an equation represents a plane in the $x - y - z$ space. It is, in fact, the equation of the ***tangent plane*** to the surface $z = f(x, y)$ at the point $(x_0, y_0, f(x_0, y_0))$.

Remark 2.20. *For the preceding diskussion to hold, we require that f is* **differentiable** *at (x_0, y_0). This is a more technical definition, the explanation for which can be found in the book on page 132. Here I just give the definition:*

We say that $f : \mathbb{R}^2 \to \mathbb{R}$ is differentiable at the point (x_0, y_0) if $\partial f/\partial x$ and $\partial f/\partial y$ exist at (x_0, y_0) and if

$$\frac{f(x,y) - f(x_0, y_0) - \left[\frac{\partial f}{\partial x}(x_0, y_0)\right](x - x_0) + \left[\frac{\partial f}{\partial y}(x_0, y_0)\right](y - y_0)}{\|(x, y) - (x_0, y_0)\|} \to 0$$

as $(x, y) \to (x_0, y_0)$.

Basically, this definition says that a function is differentiable if the approximation by a plane is "pretty good" around (x_0, y_0), even when divided by $\|(x, y) - (x_0, y_0)\|$.

2.3.3 The Matrix of Partial Derivatives

Now we begin extending our observations from functions $f : \mathbb{R}^2 \to \mathbb{R}$ to general functions $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$. Before doing that, we need to set our notation straight. One important element we will be using frequently is the matrix of partial derivatives of f:

1. As we have seen, for a function $f : \mathbb{R}^2 \to \mathbb{R}$ one can speak of the derivatives with respect to the x-axis and with respect to the y-axis:
$$\frac{\partial f}{\partial x}(x, y),\ \frac{\partial f}{\partial y}(x, y)$$
2. For a function $f : \mathbb{R}^3 \to \mathbb{R}$ we have
$$\frac{\partial f}{\partial x}(x, y, z),\ \frac{\partial f}{\partial y}(x, y, z),\ \frac{\partial f}{\partial z}(x, y, z).$$
3. For the most general scalar-valued function $f : \mathbb{R}^n \to \mathbb{R}$, we have
$$\frac{\partial f}{\partial x_1}(x_1, ..., x_n),\ \frac{\partial f}{\partial x_2}(x_1, ..., x_n), ...,\ \frac{\partial f}{\partial x_n}(x_1, ..., x_n).$$

4. Now comes the "leap": we consider a vector-valued function $f : \mathbb{R}^n \to \mathbb{R}^m$. Recall that such a function is actually an array of m *scalar*-valued functions: $f = (f_1, f_2, ..., f_m)$. Each of these f_i is a scalar-valued function of n variables, as in the case above. So we end up having

$$\frac{\partial f_1}{\partial x_1}(x_1, ..., x_n),\ \frac{\partial f_1}{\partial x_2}(x_1, ..., x_n), ..., \frac{\partial f_1}{\partial x_n}(x_1, ..., x_n)$$

$$\frac{\partial f_2}{\partial x_1}(x_1, ..., x_n),\ \frac{\partial f_2}{\partial x_2}(x_1, ..., x_n), ..., \frac{\partial f_2}{\partial x_n}(x_1, ..., x_n)$$

$$\vdots$$

$$\frac{\partial f_m}{\partial x_1}(x_1, ..., x_n),\ \frac{\partial f_m}{\partial x_2}(x_1, ..., x_n), ..., \frac{\partial f_m}{\partial x_n}(x_1, ..., x_n).$$

This way of writing leads us to write these partial fractions in a matrix, naturally called the ***matrix of partial derivatives of*** f ***at*** $\mathbf{x}$:

$$\mathbf{D}f(\mathbf{x}) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}$$

where all the partial derivatives are evaluated at $\mathbf{x} = (x_1, x_2, ..., x_n)$.

2.3.4 Generalizing the Notion of Differentiability

We can extend the preceding definition to a wide class of functions in a natural way:

Definition 2.21. *Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$, with U open. We say that f is* **differentiable** *at $\mathbf{x_0} \in U$ if the partial derivatives $\partial f / \partial x_j$ of f at $\mathbf{x_0}$ exist and if*

$$\lim_{\mathbf{x} \to \mathbf{x_0}} \frac{\| f(\mathbf{x}) - f(\mathbf{x_0}) - \mathbf{T}(\mathbf{x} - \mathbf{x_0}) \|}{\| \mathbf{x} - \mathbf{x_0} \|} = 0,$$

where $\mathbf{T} = \mathbf{D}f(x_0)$, and $\mathbf{T}(\mathbf{x} - \mathbf{x_0})$ is the multiplication of an $m \times n$ matrix, by a column vector with n components.

2.3.5 Gradients

The gradient, which we will encounter again later, and which is one of the most important notions studied in this course is the matrix of partial derivatives in the special case that $f : \mathbb{R}^n \to \mathbb{R}$. It is usually denoted by ∇f:

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, ..., \frac{\partial f}{\partial x_n} \right).$$

2.3.6 Theorems

There are some very important theorems in this context. In this course we do not dive deep enough into the material to actually prove them, but they have fully rigorous proofs.

Theorem 2.22. *Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ be differentiable at $\mathbf{x_0} \in U$. Then f is continuous at $\mathbf{x_0}$.*

Theorem 2.23. *Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$. If the the partial derivatives $\partial f_i/\partial x_j$, $1 \leq i \leq m$, $1 \leq j \leq n$ all exist and are continuous at a neighborhood of a point $\mathbf{x_0} \in U$, then f is differentiable at $\mathbf{x_0}$.*

For both theorems, the reverse implication is incorrect. There are examples of functions that are continuous at some point $\mathbf{x_0}$, but not differentiable there (for example, the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = |x|$ at the point 0) as there are examples of functions that are differentiable at a point $\mathbf{x_0}$, but their partial derivatives are not necessarily continuous in a neighborhood of $\mathbf{x_0}$ (an example for such a function is more sophisticated).

2.4 Introduction to Paths and Curves

Homework 2.24 (Reading due Wednesday 9/24; Problems due Monday 9/29).
__Read__ §2.5
__Solve__ §2.4: 2, 8, 11, 13, 14, 15

So far we have diskussed functions $f:\mathbb{R}^n\to\mathbb{R}^m$, we looked at surfaces given by functions $f:\mathbb{R}^2\to\mathbb{R}$ and the partial derivatives on them. We extended the notion of a partial derivative for general functions $f:\mathbb{R}^n\to\mathbb{R}^m$.

In this section, we want to build a crucial tool that will allow us to calculate the rate of change of a function $f:\mathbb{R}^2\to\mathbb{R}$ in an arbitrary direction. To get some intuition, our goal is to be able to calculate the slope of a mountain in *any* direction. Right now, we can only calculate the slope along the east-west and the north-south directions.

Our method of calculating the slope in a general direction of our choice is as follows: we construct a curve that "sits" on the surface (just like a trail we may take on a mountain we're climbing), and we calculate the derivative (single variable derivative!) along this curve. But, before doing that, we define what a curve exactly is, how is differs from a path, and we consider some examples.

In our convention, a ***curve*** C is a 2 or 3-dimensional line in the plane or in space. A corresponding ***path*** $\mathbf{c}(t)$ is a parameterization of that curve. That is, a path $\mathbf{c}(t)$ is a function $\mathbf{c}:[a,b]\subseteq\mathbb{R}\to\mathbb{R}^n$ (with $n=2$ or 3), where t is thought of as time: we trace the curve C starting from time a up to time b.

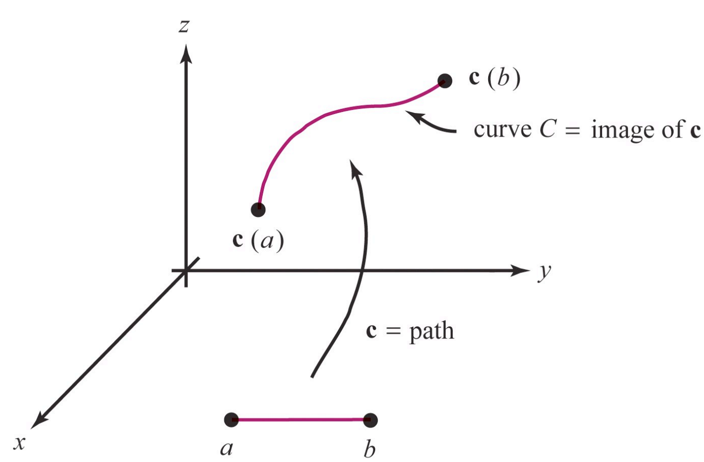


Figure 17: The curve C and a parameterization $\mathbf{c}(t)$

One curve, many paths. Our distinction between a curve and a path, means that many (in fact, infinitely many) different paths can correspond to the same curve. For example, think of the curve in space joining San Francisco and New York. Any airplane tracing this curve, traces it in a slightly different path (this is since no two airplanes fly in exactly the same velocity the entire flight).

For simplicity, we do not allow a path to "reverse" and trace a curve backwards, although, in general, there is no mathematical reason to not allow it.

Velocity. For a path $\mathbf{c}(t)$ that is differentiable, we define its ***velocity*** at time t to be the limit

$$\mathbf{c}'(t) = \lim_{h \to 0} \frac{\mathbf{c}(t+h) - \mathbf{c}(t)}{h}.$$

(if the limit exists).

Tangent line. An immediate result of the previous definition, is that of the ***tangent line***. The tangent line to the path $\mathbf{c}(t)$ at the point $\mathbf{c}(t_0)$ is

$$\mathbf{l}(t) = \mathbf{c}(t_0) + (t - t_0)\mathbf{c}'(t_0).$$

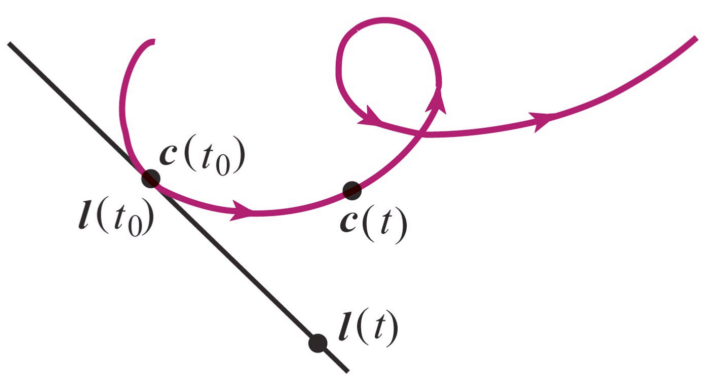


Figure 18: The tangent line $\mathbf{l}(t)$ to $\mathbf{c}(t)$ at $\mathbf{c}(t_0)$

2.5 Properties of the Derivative

Homework 2.25 (Reading due Friday 9/26; Problems due Monday 9/29)**.**
Read *§2.6*
Solve *§2.5: 3, 5d, 7, 12, 16, 19*

2.5.1 Reminder: functions of a single variable

Below we will see that all the properties we know for scalar-functions of a single variable (i.e. functions $f : \mathbb{R} \to \mathbb{R}$), holds for vector-valued functions of more variables.

Theorem 2.26. *[Properties of scalar-valued functions of a single variable] Let $f : \mathbb{R} \to \mathbb{R}$, $g : \mathbb{R} \to \mathbb{R}$ both be differentiable at x_0 and let $c \in \mathbb{R}$ be real number.*

1. ***Constant Multiple Rule.*** *$h(x) = (cf)(x)$ is also differentiable at x_0, and*
$$h'(x_0) = cf'(x_0).$$
2. ***Sum Rule.*** *$h(x) = f(x) + g(x)$ is also differentiable at x_0, and*
$$h'(x_0) = f'(x_0) + g'(x_0).$$
3. ***Product Rule.*** *$h(x) = f(x)g(x)$ is also differentiable at x_0, and*
$$h'(x_0) = f'(x_0)g(x_0) + f(x_0)g'(x_0).$$
4. ***Quotient Rule.*** *If g is nonzero on a neighborhood of x_0 (i.e. on an open interval containing x_0), then $h(x) = f(x)/g(x)$ is also differentiable at x_0, and*
$$h'(x_0) = \frac{f'(x_0)g(x_0) - f(x_0)g'(x_0)}{(g(x_0))^2}.$$

Theorem 2.27. *[Chain rule for scalar-valued functions of a single variable] Let $U \subseteq \mathbb{R}$ and $V \subseteq \mathbb{R}$ be open sets. Let $g : U \subseteq \mathbb{R} \to \mathbb{R}$ and $f : V \subseteq \mathbb{R} \to \mathbb{R}$ be functions such that g maps U into V, so that $h = f \circ g$ is defined. Suppose g is differentiable at x_0 and f is differentiable at $y_0 = g(x_0)$. Then h is differentiable at x_0 and*
$$h'(x_0) = f'(y_0)g'(x_0).$$

(see Figure 19)

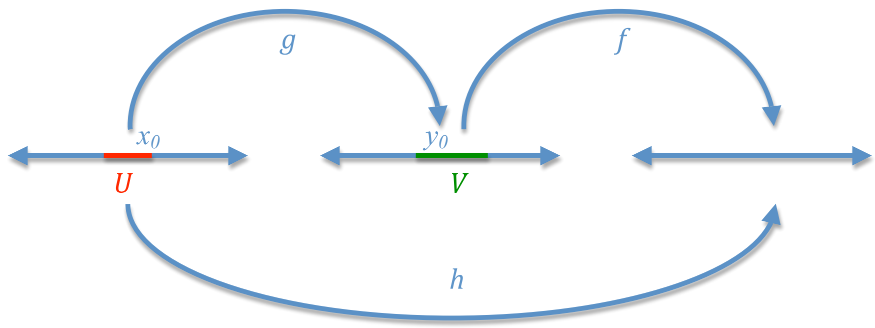


Figure 19: Composition of two scalar-valued single variable functions

2.5.2 Extending to general functions

Theorems 2.26 and 2.27 extend to general vector-valued functions of several variables. Theorem 2.26 works almost as-is in the general case, where the simple derivative is always replaced by the *matrix of partial derivatives*. **Please compare to the general statement in the book.**

The chain rule in higher dimensions. We will spend more time on the second theorem, Theorem 2.27. The book addresses composition of functions in great generality, and I recommend reading it. You will need to refresh §1.5, §2.3 to fully understand.

For the purpose of this course, however, we are only interested in two special cases.

Scalar-valued function along a curve in space. In this first case, you can think of temperature along a mountain path: as you hike along the path, you keep track of the temperature at each point.

The path is a function $\mathbf{c}(t) : \mathbb{R} \to \mathbb{R}^3$, and the function that keeps track of the temperature is a function $f : \mathbb{R}^3 \to \mathbb{R}$ that associates to each point $\mathbf{c}(t_0)$ on the path its temperature. Thus, when describing the temperature as a function of time (forgetting for a moment about the path), you would consider the function $h(t) = f(\mathbf{c}(t))$.

The change in temperature with respect to time is given by the formula

$$\frac{dh}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt},$$

or, in vector notation

$$\begin{aligned}\frac{dh}{dt} &= \nabla f(\mathbf{c}(t)) \cdot \mathbf{c}'(t) \\ &= \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} & \frac{\partial f}{\partial y} \end{pmatrix} \cdot \begin{pmatrix} x'(t) \\ y'(t) \\ z'(t) \end{pmatrix}.\end{aligned}$$

To summarize, we have the following:

Given the functions

$$\mathbb{R} \xrightarrow{\mathbf{c}} \mathbb{R}^3 \xrightarrow{f} \mathbb{R}$$

with the composition $f \circ \mathbf{c}$ denoted by $h(t)$ we have

$$\frac{dh}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt}.$$

Scalar-valued function after change of coordinates. In this case, we change coordinates, then apply a scalar function. This means that our usual x, y, z are changed into some other coordinates u, v, w, and we then apply the function f to these new coordinates.

What we have is the following:

Given the functions

$$\mathbb{R}^3_{xyz} \xrightarrow{g} \mathbb{R}^3_{uvw} \xrightarrow{f} \mathbb{R}$$

with the composition $f \circ g$ denoted by h, where g is a change of coordinates

$$g(x, y, z) = (u(x, y, z), v(x, y, z), w(x, y, z))$$

we have

$$\begin{aligned}\frac{\partial h}{\partial x} &= \frac{\partial f}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial f}{\partial v}\frac{\partial v}{\partial x} + \frac{\partial f}{\partial w}\frac{\partial w}{\partial x} \\ \frac{\partial h}{\partial y} &= \frac{\partial f}{\partial u}\frac{\partial u}{\partial y} + \frac{\partial f}{\partial v}\frac{\partial v}{\partial y} + \frac{\partial f}{\partial w}\frac{\partial w}{\partial y} \\ \frac{\partial h}{\partial z} &= \frac{\partial f}{\partial u}\frac{\partial u}{\partial z} + \frac{\partial f}{\partial v}\frac{\partial v}{\partial z} + \frac{\partial f}{\partial w}\frac{\partial w}{\partial z}.\end{aligned}$$

In matrix notation, this set of equations may be written as:

$$\begin{pmatrix} \frac{\partial h}{\partial x} & \frac{\partial h}{\partial y} & \frac{\partial h}{\partial z} \end{pmatrix} = \begin{pmatrix} \frac{\partial f}{\partial u} & \frac{\partial f}{\partial v} & \frac{\partial f}{\partial w} \end{pmatrix} \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} & \frac{\partial u}{\partial z} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} & \frac{\partial v}{\partial z} \\ \frac{\partial w}{\partial x} & \frac{\partial w}{\partial y} & \frac{\partial w}{\partial z} \end{pmatrix}.$$

Example 2.28. *Let $x = \cos\theta, y = \sin\theta$, and consider the function $h(\theta) = \exp[x(\theta)y(\theta)]$.*

Our situation is

$$\mathbb{R} \xrightarrow{(x(\theta),y(\theta))} \mathbb{R}^2 \xrightarrow{exp(xy)} \mathbb{R}$$

so we invoke the first formula:

$$\begin{aligned} \frac{dh}{d\theta} &= \frac{\partial}{\partial x}(e^{xy})\frac{dx}{d\theta} + \frac{\partial}{\partial y}(e^{xy})\frac{dy}{d\theta} \\ &= ye^{xy}(-\sin\theta) + xe^{xy}(\cos\theta) \\ &= e^{\sin\theta\cos\theta}\left(-\sin^2\theta + \cos^2\theta\right). \end{aligned}$$

Example 2.29. *Let $x = r\cos\theta, y = r\sin\theta$, and consider the function $h(r,\theta) = \exp[x(r,\theta)y(r,\theta)]$.*

Our situation is

$$\mathbb{R}^2 \xrightarrow{(x(r,\theta),y(r,\theta))} \mathbb{R}^2 \xrightarrow{exp(xy)} \mathbb{R}$$

so we invoke the second formula:

$$\begin{aligned} \frac{\partial h}{\partial \theta} &= \frac{\partial}{\partial x}(e^{xy})\frac{\partial x}{\partial \theta} + \frac{\partial}{\partial y}(e^{xy})\frac{\partial y}{\partial \theta} \\ &= ye^{xy}(-r\sin\theta) + xe^{xy}(r\cos\theta) \\ &= re^{\sin\theta\cos\theta}\left(-\sin^2\theta + \cos^2\theta\right). \end{aligned}$$

and

$$\begin{aligned} \frac{\partial h}{\partial r} &= \frac{\partial}{\partial x}(e^{xy})\frac{\partial x}{\partial r} + \frac{\partial}{\partial y}(e^{xy})\frac{\partial y}{\partial r} \\ &= ye^{xy}(\cos\theta) + xe^{xy}(\sin\theta) \\ &= 2e^{\sin\theta\cos\theta}\sin\theta\cos\theta. \end{aligned}$$

Example 2.30. *Let $x = r\cos\theta, y = r\sin\theta$, and consider the function*

$$\mathbf{h}(r,\theta) = (\exp[x(r,\theta)y(r,\theta)], \sin[x(r,\theta)y(r,\theta)]).$$

Our situation is

$$\mathbb{R}^2 \xrightarrow{(x(r,\theta),y(r,\theta))} \mathbb{R}^2 \xrightarrow{(exp(xy),\sin xy)} \mathbb{R}^2$$

Our method, in such a case, is to separate $\mathbf{h}$ *into* $\mathbf{h} = (h_1, h_2)$*, where* $h_1 = \exp\left[x(r,\theta)y(r,\theta)\right]$ *and* $h_2 = \sin\left[x(r,\theta)y(r,\theta)\right]$ *and then treat* h_1 *and* h_2 *as in the previous example.*

Implicit differentiation. There are cases in which we want to find the dependence of one variable on another, but cannot extract an explicit formula. For example, from the theory of functions of a single variable, we know how to use this method to find dy/dx in the case $x^2 + y^2 = 4$. Recall, that implicit differentiation gives us

$$2x + 2yy' = 0$$

so that

$$y' = -\frac{x}{y}.$$

Example 2.31. *Let us modify the single variable example we saw above, to a multivariate example: consider the relation*

$$x^2 + y^5 + z^3 = 4$$

and suppose we treat z *as a function of* x *and* y*. Now let us calculate* $\partial z/\partial x$ *and* $\partial z/\partial y$*:*

First, differentiating implicitly with respect to x*, we get*

$$2x + 0 + 3z^2\frac{\partial z}{\partial x} = 0$$

so that, at any point where $z \neq 0$*,*

$$\frac{\partial z}{\partial x} = -\frac{2x}{3z^2}.$$

Similarly, we find

$$\frac{\partial z}{\partial y} = -\frac{5y^4}{3z^2}.$$

Generally. When we have a relation

$$F(x, y, z(x, y)) = 0$$

and we want to calculate $\partial z/\partial x$, we proceed as follows:

We first want to prevent confusion, so we label the entries of F as u, v, w, where $u = x, v = y, w = z(x, y)$. Now we can differentiate implicitly:

$$\begin{aligned} 0 &= \frac{\partial F}{\partial x} \\ &= \frac{\partial F}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial F}{\partial v}\frac{\partial v}{\partial x} + \frac{\partial F}{\partial w}\frac{\partial w}{\partial x} \\ &= \frac{\partial F}{\partial u} \cdot 1 + \frac{\partial F}{\partial v} \cdot 0 + \frac{\partial F}{\partial w}\frac{\partial w}{\partial x}. \end{aligned}$$

So we find (at any point where $\frac{\partial F}{\partial w} \neq 0$)

$$\frac{\partial z}{\partial x} = -\frac{\frac{\partial F}{\partial u}}{\frac{\partial F}{\partial w}}.$$

Remark 2.32 (Confusing notation)**.** *Relabeling the variables above served as an important step to avoid confusion: labeling the entries of F as u, v, w, enabled us to have an intermediate step in our chain rule. We transformed the situation into*

$$\mathbb{R}^2 \xrightarrow{(u(x),v(y),w(x,y))} \mathbb{R}^3 \xrightarrow{F(u,v,w)} \mathbb{R}.$$

Without this relabeling, we'd have confusion arising with regard to treatment of F: Is F a function of the two variables (x, y) or of the three variables (x, y, z)? As we saw in class, when we don't address this issue, it is unclear what $\partial F/\partial x$ means: Is it the change of F with respect to the first entry or is it the change of F with respect to the variable x, which also appears implicitly in the third entry?

Relabeling solves this issue.

Remark 2.33 (When to use d and when to use ∂)**.** *Another confusing point in this subject is the usage of the ∂ symbol. Our convention, unless otherwise stated, is that we use a regular d only when dealing with a function of one variable. As an example, if $z = f(x, y)$ and both x and y are functions of t, we write*

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}.$$

The ∂ notation is reserved for functions of several variables.

2.6 Gradients and Directional Derivatives

Homework 2.34 (Reading due Monday 9/29; Problems due Monday 9/29)**.**

Read *§3.1, §3.2*

Solve *§2.6: 1, 2d, 3c, 5a, 20, 22, 24a; §3.1: 3; §3.3: 1 (just find critical points using first partial derivatives)*

2.6.1 The Gradient

Recall our definition and notation for the gradient: Given a differentiable function $f : \mathbb{R}^3 \to \mathbb{R}$, we define its ***gradient vector field*** to be

$$\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right).$$

Example 2.35. *Let* $f(x, y, z) = \sin(xy) + z^2$*. Then*

$$\nabla f = (y \cos(xy), x \cos(xy), 2z).$$

2.6.2 Directional Derivatives

Now we finally can define the *most general notion of differentiation:* ***calculating the rate of change of a function in any direction***. Recall, that the partial derivatives give us the rate of change only along the *axes*.

Two Dimensions. The example of the mountain slope comes in handy here: the contour of the mountain can be thought of as a function $h(x, y)$ that assigns to each point (x, y) in the plane, the elevation of the mountain above it. Now, let $\mathbf{c}(t)$ be a path in the xy plane, as in Figure 20. This path maps to a path $h(\mathbf{c}(t))$ on the mountain, as in Figure 21.

Now, assume that you trace the curve at speed 1, i.e. that $\|\mathbf{c}'(t)\| = 1$ for all t. Also, suppose that $\mathbf{c}(0) = \mathbf{p}$ and that $\mathbf{c}'(0) = \mathbf{v}$. Then we define the ***directional derivative*** of h at the point $\mathbf{p}$ in the direction $\mathbf{v} = \mathbf{c}'(0)$ to be

$$\frac{d}{dt} h(\mathbf{c}(t))\Big|_{t=0} = \nabla h(\mathbf{c}(t)) \cdot \mathbf{c}'(t)\Big|_{t=0} = \nabla h(\mathbf{p}) \cdot \mathbf{v}$$

(the expression on the left is the definition, and the other expressions are obtained by applying the chain rule and then plugging in $t = 0$)

Three Dimensions. The exact same rationalization works in three-dimensions. See page 164 in the book for an analogous explanation of the one just given here. The formula we have above remains true (see page 165 of the book).

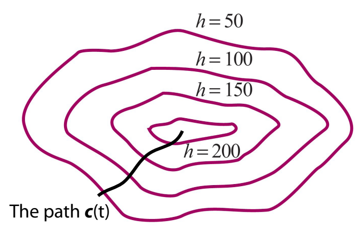


Figure 20: The path $\mathbf{c}(t)$ and the level sets of the mountain

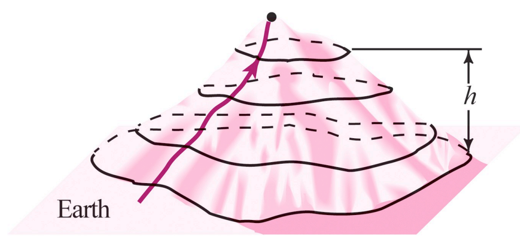


Figure 21: The path along the mountain slope is given by $h(\mathbf{c}(t))$

Example 2.36. *Let $f(x,y,z) = x\sin(yz) + z^2$. Compute the rate of change of f in the direction of the unit vector $\mathbf{v} = (0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ at the point $\mathbf{p} = (3,1,8)$.*

Solution. *Let us first make a preliminary calculation:*

$$\nabla f = (\sin(yz), xz\cos(yz), xy\cos(yz) + 2z).$$

So we have

$$\begin{aligned}\nabla f \cdot \mathbf{v} &= (\sin(yz), xz\cos(yz), xy\cos(yz) + 2z) \cdot (0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}) \\ &= \frac{1}{\sqrt{2}} xz\cos(yz) + \frac{1}{\sqrt{2}}(xy\cos(yz) + 2z)\end{aligned}$$

Plugging in $x = 3, y = 1, z = 8$ will give us our desired result.

2.6.3 The Gradient is the direction of fastest increase

We saw that the directional derivative of h at the point $\mathbf{p}$ in the direction $\mathbf{v}$ (with $\|\mathbf{v}\| = 1$) is given by $\nabla h(\mathbf{p}) \cdot \mathbf{v}$. But recall that if θ is the angle between $\nabla h(\mathbf{p})$ and $\mathbf{v}$, then

$$\nabla h(\mathbf{p}) \cdot \mathbf{v} = \|\nabla h(\mathbf{p})\|\|\mathbf{v}\| \cos\theta = \|\nabla h(\mathbf{p})\| \cos\theta.$$

This number is maximized when $\theta = 0$, i.e. when $\mathbf{v}$ and $\nabla h(\mathbf{p})$ point in the same direction.

What we have just found is that the gradient points in the direction of steepest ascent.

2.6.4 The Gradient is perpendicular to level sets

It turns out that the direction of steepest ascent (the direction of the gradient) is perpendicular to directions of constant value (level sets). We see this by choosing $\mathbf{c}(t)$ to lie *inside* the level set. By doing this, we have $h(\mathbf{c}(t)) = const$ for all times t. Thus $\frac{d}{dt}h(\mathbf{c}(t)) = 0$ for all t. But recall that

$$\frac{d}{dt}h(\mathbf{c}(t)) = \nabla h(\mathbf{c}(t)) \cdot \mathbf{c}'(t).$$

Since $\mathbf{c}'(t) \neq 0$, and $\nabla h(\mathbf{c}(t)) \neq 0$ by assumption, the two must be perpendicular.

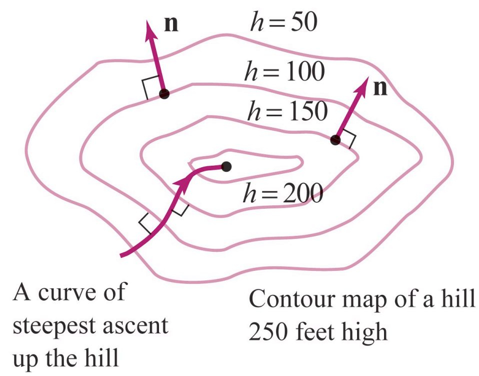


Figure 22: The path with steepest ascent is perpendicular to level sets (and vice-versa)

This enables us to define the ***tangent plane*** to the level set L_c of the function $h(x, y, z)$ at the point (x_0, y_0, z_0): It is the set of all points (x, y, z) such that

$$\nabla h(x_0, y_0, z_0) \cdot (x - x_0, y - y_0, z - z_0) = 0.$$

3 Higher-Order Derivatives

Homework 3.1 (Reading due Wednesday 10/1; Problems not to hand in)**.**
Read *§3.3, §3.4*
Solve *§3.1: 8, 19; §3.3: 17, 23, 31, 34*

This chapter explores applications of results we saw in the previous chapters. The first two sections don't contain anything that is *essentially* new, so we will survey them briefly. Sections §3.3 and §3.4 are important (and very interesting). Section §3.5 is *extremely* important in mathematics and physics. However, we will skip it in this course.

3.1 Iterated Partial Derivatives

In the one-dimensional case, for a function $g : \mathbb{R} \to \mathbb{R}$, we already know that we can take several derivatives (if the function is "nice" enough). Thus, we get functions $g(x), g'(x), g''(x)$, etc. All of these are functions of x.

Similarly, for a function $f : \mathbb{R}^2 \to \mathbb{R}$, one can continue to differentiate the functions $\frac{\partial f}{\partial x}(x, y), \frac{\partial f}{\partial y}(x, y)$ in both the x and the y variables:

$$\begin{aligned} f_{xx}(x,y) &= \frac{\partial^2 f}{\partial x^2}(x,y) = \frac{\partial}{\partial x}\frac{\partial f}{\partial x}(x,y) \\ f_{yy}(x,y) &= \frac{\partial^2 f}{\partial y^2}(x,y) = \frac{\partial}{\partial y}\frac{\partial f}{\partial y}(x,y) \\ f_{xy}(x,y) &= \frac{\partial}{\partial y}\frac{\partial f}{\partial x}(x,y) \\ f_{yx}(x,y) &= \frac{\partial}{\partial x}\frac{\partial f}{\partial y}(x,y), \end{aligned}$$

and so on... Usually, in practice, we stop at the second partial derivative stage, i.e. we do not continue beyond what is listed above. In addition, all sufficiently "nice" functions we encounter satisfy the famous ***equality of mixed partials*** due to Euler:

$$f_{xy}(x,y) = f_{yx}(x,y).$$

Please refer to the book for a proof of this statement.

3.2 Taylor's Theorem

3.2.1 The Single Variable Case

Recall, that for a function $f : \mathbb{R} \to \mathbb{R}$ we have the following hierarchy of approximations around some point x_0:

0. The so-called *0th approximation* is the best approximation if we only know the value of f at x_0: we approximate $f(x)$ by $f(x_0)$ for all x.

1. The next type of approximation, the so-called *1st (linear) approximation*: if we also know $f'(x_0)$ we approximate $f(x)$ by a the linear function $f(x_0) + f'(x_0)(x - x_0)$.

2. Next, we have the so-called *2nd approximation*: if we also know $f''(x_0)$ we approximate $f(x)$ by the parabola $f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2$.

Taylor's Theorem. It is Taylor's Theorem that makes these approximations rigorous, by also giving an estimate of the error. The results of the theorem, in accordance with the 3 cases listed above are as follows:

0.
$$f(x_0 + h) = f(x_0) + R_1(x_0, h)$$
where $R_1(x_0, h)$ is a function of x_0 and of h, that is small as $h \to 0$ in the sense that
$$\lim_{h\to 0} \frac{|R_1(x_0, h)|}{h} = 0,$$
i.e. even when divided by the *small* number h, R_1 tends to 0. (Remember that dividing by a *small* number h is the same as multiplying by the *large* number $1/h$).

1.
$$f(x_0 + h) = f(x_0) + f'(x_0) \cdot h + R_2(x_0, h)$$
where $R_2(x_0, h)$ is a function of x_0 and of h, that is small as $h \to 0$ in the sense that
$$\lim_{h\to 0} \frac{|R_2(x_0, h)|}{h^2} = 0,$$
i.e. even when divided by the *very small* number h^2, R_2 tends to 0.

2.
$$f(x_0 + h) = f(x_0) + f'(x_0) \cdot h + \frac{1}{2!}f''(x_0) \cdot h^2 + R_3(x_0, h)$$
where $R_3(x_0, h)$ is a function of x_0 and of h, that is small as $h \to 0$ in the sense that
$$\lim_{h\to 0} \frac{|R_3(x_0, h)|}{h^3} = 0,$$
i.e. even when divided by the *extremely small* number h^3, R_3 tends to 0.

3.2.2 The Two Variable Case

When considering a function $f : \mathbb{R}^2 \to \mathbb{R}$ of two variables x and y we find that the result is very similar. The only major difference, is the appearance of terms of mixed derivatives.

In §2.3 we learned how to find a linear approximation to the surface $z = f(x, y)$ at some point (x_0, y_0). This linear approximation is, in fact, the tangent plane to the surface at (x_0, y_0). We are now seeking an improvement to this approximation.

Example 3.2. *Investigate the function* $f(x, y) = 3x^2 + 5xy + 2y^2$ *around* $(0, 0)$.

Solution. *First, consider the section* $x = 0$*: When we set* $x = 0$*, we are left with the function* $g(y) = f(0, y) = 2y^2$*. We can pick up this "2" coefficient by considering* $f_{yy}/2 = 4/2 = 2$*. Next, we consider the section* $y = 0$*: In this case, we see the function* $h(x) = f(x, 0) = 3x^2$*. This "3" can be picked up by looking at* $f_{xx}/2 = 6/2 = 3$.

The only coefficient we are unable to "pick up" using one-dimensional methods is the "5" in front of the xy*. To pick it up, we must consider* $f_{xy} = 5$.

We see that because of the extra dimension, we need to gather more information to be able to approximate the function to a higher degree (beyond the linear). This shows us the importance of the mixed partials. Now let us give the statement of Taylor's Theorem for the two variable case.

Theorem 3.3 (Second Order Taylor's Formula for a Function of Two Variables)**.** *Let* $f : \mathbb{R}^2 \to \mathbb{R}$ *be a function of two variables* x *and* y*, that is continuously differentiable, and so are its partial derivatives, and their partial derivatives. Let* $\mathbf{h} = (h_1, h_2)$*. Then*

$$f(\mathbf{p}+\mathbf{h}) = f(\mathbf{p}) + \nabla f(\mathbf{p}) \cdot \mathbf{h} + \frac{1}{2}\left[f_{xx}(\mathbf{p})h_1^2 + f_{yy}(\mathbf{p})h_2^2 + 2f_{xy}(\mathbf{p})h_1h_2\right] + R_2(\mathbf{p}, \mathbf{h}).$$

Reorganizing the sum, we get:

$$\begin{aligned} f(\mathbf{p}+\mathbf{h}) &= f(\mathbf{p}) \\ &+ f_x(\mathbf{p})h_1 + f_y(\mathbf{p})h_2 \\ &+ \frac{1}{2}\left[f_{xx}(\mathbf{p})h_1^2 + f_{yy}(\mathbf{p})h_2^2\right] \\ &+ f_{xy}(\mathbf{p})h_1h_2 \\ &+ R_2(\mathbf{p}, \mathbf{h}), \end{aligned}$$

where the first row represents the 0th approximation, the second row represents the 1st approximation, the third row represents what we would *expect* to be the second approximation, but the fourth row is the additional mixed term that comes into play because of the extra dimension, and the last row is the usual error term that we see in the single variable case as well.

Proof. Our goal is to reduce the problem to a one-dimensional problem, by defining the function $g(t) = f(\mathbf{p}+t\mathbf{h})$ of the single variable t. Applying the one-dimensional theorem to g (with $x_0 = 0, h = 1$), we get

$$g(1) = g(0) + g'(0)\cdot 1 + \frac{1}{2!}g''(0)\cdot 1^2 + R_2(0,1).$$

Using the chain rule, we obtain:

$$\begin{aligned} g'(t) &= \nabla f(\mathbf{p}+t\mathbf{h})\cdot\mathbf{h} = f_x(\mathbf{p}+t\mathbf{h})h_1 + f_y(\mathbf{p}+t\mathbf{h})h_2 \\ g''(t) &= f_{xx}(\mathbf{p}+t\mathbf{h})h_1^2 + f_{yy}(\mathbf{p}+t\mathbf{h})h_2^2 + f_{xy}(\mathbf{p}+t\mathbf{h})h_1h_2 + f_{yx}(\mathbf{p}+t\mathbf{h})h_2h_1, \end{aligned}$$

which, if we plug back into the one-dimensional formula for g, proves the theorem (except for the error term R_2, which is left for you to read about in the book). □

Intuition. What we have just seen can be summarized as following: We want to approximate the function at a given point $\mathbf{p}+\mathbf{h} = (p_1+h_1, p_2+h_2)$. We know the value of the function, as well as its partial derivatives, and their own partial derivatives at the point $\mathbf{p} = (p_1, p_2)$. Plugging into Taylor's Formula, if $\mathbf{h} = (h_1, h_2)$ is small enough, we expect to get a pretty good approximation of what $f(\mathbf{p}+\mathbf{h}) = f(p_1+h_1, p_2+h_2)$ is.

3.2.3 The General Case

Please read the general case in the book, p. 196.

The big "leap" is when going from understanding functions of a single variable to understanding functions of two variables. Going beyond two variables doesn't contain any deep new ideas.

3.3 Extrema

3.3.1 Functions of a Single Variable

In the single variable case, we know that the first indicator of a "suspicious" (critical) point is when the derivative is 0 there. This implies that that point can be either a local minimum, or a local maximum, *or* a point of inflection.

It is the *second derivative test* that exposes the local minima and maxima: Let our function $f:\mathbb{R}\to\mathbb{R}$ be such that $f'(x_0) = 0$. Then:

1. $f''(x_0) > 0$ implies that x_0 is a local minimum (f is concave up at x_0).

2. $f''(x_0) < 0$ implies that x_0 is a local maximum (f is concave down at x_0).
3. $f''(x_0) = 0$ implies that x_0 is a point of inflection or a local minimum or a local maximum.

3.3.2 Functions of Two Variables

In the two variable case the situation is very similar, except that first derivatives are replaced by the gradient, and second derivatives are replaced by the matrix of second derivatives, called the ***Hessian***:

$$H(f) = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix}$$

We now define the determinant of H,

$$D = \det H(f) = f_{xx} f_{yy} - f_{xy}^2.$$

(Note that both H and D are functions of the point at which the derivatives are evaluated)

The ***second derivative test for a function of two variables*** is as follows:

Let $\mathbf{p} = (p_1, p_2)$ be a critical point, i.e. satisfy

$$\nabla f(\mathbf{p}) = (f_x(p_1, p_2), f_y(p_1, p_2)) = (0, 0).$$

(This implies that the tangent plane to the surface $z = f(x, y)$ at $\mathbf{p}$ is parallel to the xy plane). Then:

1. $D > 0$ and $f_{xx} > 0$ imply that $\mathbf{p}$ is a local minimum.
2. $D > 0$ and $f_{xx} < 0$ imply that $\mathbf{p}$ is a local maximum.
3. $D < 0$ implies that $\mathbf{p}$ is a saddle point.
4. $D = 0$ implies that the nature of $\mathbf{p}$ cannot be determined.

Example 3.4 (This example is similar to the one-dimensional example $g(x) = x^2$)**.** *Consider the function $f(x, y) = x^2 + y^2$. First we have $f_x = 2x$, $f_y = 2y$. Let us find the critical points:*

We have to solve

$$\begin{aligned} 0 &= f_x = 2x \\ 0 &= f_y = 2y \end{aligned}$$

which has solution $(0,0)$*. Let us check what type of critical point* $(0,0)$ *is:*

$$H(f) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

so that $D = 2 \cdot 2 - 0 \cdot 0 = 4$*, and* $f_{xx} = 2 > 0$*, i.e. case (1) holds, and* $(0,0)$ *is a local minimum.*

Example 3.5 (This example is similar to the one-dimensional example $h(x) = -x^2$)**.** *Consider the function* $f(x,y) = -x^2 - y^2$*. First we have* $f_x = -2x, f_y = -2y$*. Let us find the critical points:*

We have to solve

$$\begin{aligned} 0 &= f_x = -2x \\ 0 &= f_y = -2y \end{aligned}$$

which has solution $(0,0)$*. Let us check what type of critical point* $(0,0)$ *is:*

$$H(f) = \begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix}$$

so that $D = (-2) \cdot (-2) - 0 \cdot 0 = 4$*, and* $f_{xx} = -2 < 0$*, i.e. case (2) holds, and* $(0,0)$ *is a local maximum.*

Example 3.6 (This is an example of a new situation which we get in higher dimensions)**.** *Consider the function* $f(x,y) = x^2 - y^2$*. First we have* $f_x = 2x, f_y = -2y$*. Let us find the critical points:*

We have to solve

$$\begin{aligned} 0 &= f_x = 2x \\ 0 &= f_y = -2y \end{aligned}$$

which has solution $(0,0)$*. Let us check what type of critical point* $(0,0)$ *is:*

$$H(f) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$$

so that $D = 2 \cdot (-2) - 0 \cdot 0 = -4$*, i.e. case (3) holds, and* $(0,0)$ *is a saddle point.*

3.3.3 Global Extrema

Finding global minima and maxima is done on a problem-to-problem basis. We can find, however, a few ground rules to work by. After listing the ground rules, we implement these rules in examples.

Closed sets. We have already seen what an *open* set is: It is (intuitively) a set that does not contain any of its boundary points. A closed set, for our purposes, is a set that *does* include *all* of its boundary points.

Compare, for example, $U = (0, 1) \subseteq \mathbb{R}$ that is open, and $D = [0, 1] \subseteq \mathbb{R}$ that is closed. ∂U (the boundary of U) that consists of the points $0, 1$ is included in D but not in U. (Note that $\partial U = \partial D$).

Finding global minima/maxima for functions of two variables. We consider the surface $z = f(x, y)$, where f is restricted to some closed (bounded) region $D \subseteq \mathbb{R}^2$. Suppose that D is the union of some open set U and its boundary ∂U, as is the case in the example above: $D = [0, 1]$ is the union of $U = (0, 1)$ and ∂U which consists of the two points 0 and 1. To find the global minima/maxima of f on D, we follow the following ground rules:

1. Find the local minima/maxima of f on U.
2. Find the minima/maxima of f along ∂U:
   (a) Parametrize the curve ∂U.
   (b) Find the critical points along the curve.
   (c) Do not forget the end-points of the curve! (that is, the end-points of the associated path. Look at the example below to see what this means).
3. Find the largest/smallest of all these numbers: those are the global maxima/minima on D.

The simplest way to explain these rules is through an example:

Example 3.7 (§3.3 #31). *Find the absolute maximum and minimum values of the function $f(x, y) = x^2 + xy + y^2$ on the closed disk $D = \{x, y \mid x^2 + y^2 \leq 1\}$*

Solution. *We follow our three steps:*

1. *We find extrema on the open disk $U = \{x, y \mid x^2 + y^2 < 1\}$:*

First, we find all points $\mathbf{p} = (p_1, p_2)$ *for which* $\nabla f(\mathbf{p}) = (0,0)$*: This means, we want to find points that satisfy*

$$\begin{aligned} 0 &= \frac{\partial f}{\partial x} = 2x + y \\ 0 &= \frac{\partial f}{\partial y} = 2y + x. \end{aligned}$$

We have only one critical point: $\mathbf{p} = (p_1, p_2) = (0,0)$.

Let us see what kind of a point this is (local maximum, minimum or saddle):
We have $f_{xx} = 2, f_{xy} = f_{yx} = 1, f_{yy} = 2$, *so that*

$$H(f) = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

and thus

$$\det H = 2 \cdot 2 - 1 \cdot 1 = 3 > 0.$$

Since $f_{xx} > 0$, $(0,0)$ *is a local minimum, where* $f = 0$.

2. *We find extrema along* $\partial U = \{x, y \mid x^2 + y^2 = 1\}$*:*

   (a) *The boundary of* U *is the unit circle (all points having distance 1 from the origin). We parametrize this curve by the path* $\mathbf{c}(t) = (\cos t, \sin t)$, $t \in [0, 2\pi]$.

   (b) *Considering the values of the function* f *along* $\mathbf{c}(t)$, *we want to find extrema, i.e. when*

   $$\frac{d}{dt}\left(f(\mathbf{c}(t))\right) = 0.$$

   But by the chain rule we know that $\frac{d}{dt}\left(f(\mathbf{c}(t))\right) = \nabla f(\mathbf{c}(t)) \cdot \mathbf{c}'(t)$.

   As we've seen $\nabla f = (2x + y, 2y + x)$. *Thus*

   $$\nabla f(\mathbf{c}(t)) = (2\cos t + \sin t, 2\sin t + \cos t).$$

   In addition

$$\mathbf{c}'(t) = (-\sin t, \cos t).$$

So, we finally have

$$\begin{aligned}
\frac{d}{dt}\left(f(\mathbf{c}(t))\right) &= \nabla f(\mathbf{c}(t)) \cdot \mathbf{c}'(t) \\
&= (2\cos t + \sin t, 2\sin t + \cos t) \cdot (-\sin t, \cos t) \\
&= (2\cos t + \sin t) \cdot (-\sin t) + (2\sin t + \cos t) \cdot \cos t) \\
&= \cos^2 t - \sin^2 t.
\end{aligned}$$

Remark 3.8. *We could have also done this step by defining*

$$g(t) = f(\mathbf{c}(t)) = \cos^2 t + \cos t \sin t + \sin^2 t = 1 + \cos t \sin t$$

so that

$$g'(t) = \cos^2 t - \sin^2 t.$$

Now we want to check when this expression vanishes, i.e. when $\cos^2 t = \sin^2 t$. *It is easily checked that this happens for* $t = \pi/4, 3\pi/4, 5\pi/4, 7\pi/4$. *We note that*

- $t = \pi/4$ *corresponds to the point* $(\sqrt{2}/2, \sqrt{2}/2)$.
- $t = 3\pi/4$ *corresponds to the point* $(-\sqrt{2}/2, \sqrt{2}/2)$.
- $t = 5\pi/4$ *corresponds to the point* $(-\sqrt{2}/2, -\sqrt{2}/2)$.
- $t = 7\pi/4$ *corresponds to the point* $(\sqrt{2}/2, -\sqrt{2}/2)$.

Let us compute the values of f *at these points:*

- *At the point* $(\sqrt{2}/2, \sqrt{2}/2)$ $f = 1.5$.
- *At the point* $(-\sqrt{2}/2, \sqrt{2}/2)$ $f = 0.5$.
- *At the point* $(-\sqrt{2}/2, -\sqrt{2}/2)$ $f = 1.5$.
- *At the point* $(\sqrt{2}/2, -\sqrt{2}/2)$ $f = 0.5$.

(c) *We inspect the end-points of our curve* $\mathbf{c}(t)$*:*

In fact, the two end-points (corresponding to $t = 0$ *and to* $t = 2\pi$*) are the same point* $x = 1, y = 0$*, where* f *achieves the value* $1^2 + 1 \cdot 0 + 0^2 = 1$.

3. *We found:*

- $(0, 0)$ *is a global minimum with value 0.*
- $(\sqrt{2}/2, \sqrt{2}/2), (-\sqrt{2}/2, -\sqrt{2}/2)$ *are global maxima with value 1.5.*

3.4 Lagrange Multipliers

What would be the shortest way for Tom to get from point A to point B, if he needs to pick up some water at the nearby river on the way? From our basic high-school class, we know that we take the mirror image of the point B, call it B', draw the straight line AB', and reflect back the part of the line that is beyond the river (see Figure 23).

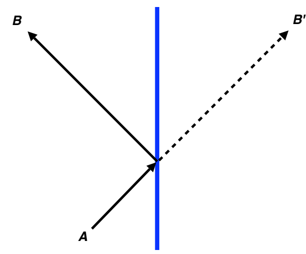


Figure 23: The shortest path from A to B passing through the river

But now, what would Tom do if the river twisted and turned? How would he decide which is the optimal path?

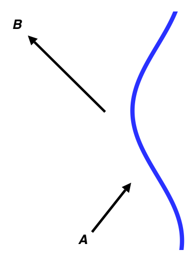


Figure 24: What's the shortest path?

Well, if he looked at Figure 24 (and thought for a while) he would come to the following conclusion: Stick a nail at each of the points A and B, and connect the two nails with a loose string. Now, trace an ellipse around the points with a pen by holding the string as stretched out as possible with it. If the string's length is c, this method gives us an ellipse that describes all points in the plane, for which the sum of their distances from A and from B equals c.

If we define

$$f(x, y) = \|(x, y) - \mathbf{OA}\| + \|(x, y) - \mathbf{OB}\|$$

then this ellipse corresponds to the level set $L_c = \{(x, y) \mid f(x, y) = c\}$.

Now, if Tom continued to draw such ellipses with strings of different lengths, in the end he would find a string length c_1 such that the ellipse just *touches* the river for the first time. The point (or points) where the ellipse touches the river is the point through which traveling would be optimal.

This point may be found analytically by defining, in addition to the function f, another function, $g(x, y)$, that has the curve defined by the river as its level set. For example, in the diagram above, the river is the curve $x = \sin y$, so that we'd define g to be $g(x, y) = x - \sin y$, and the river corresponds to the *zero* level set.

Now, recall that we have shown that gradients are perpendicular to level sets (see §2.6.4). In our example, we want both level sets to *just* "touch" each other, or, in other words, have the same tangent line. This, obviously, implies that both have the same perpendicular direction at that point, i.e. that the gradients are a scalar multiple of each other (see Figure 25).

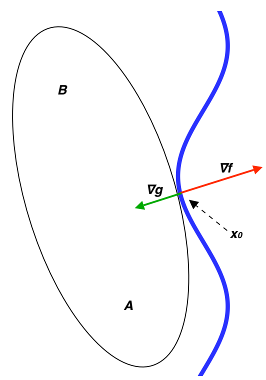


Figure 25: Gradients are on the same line

Let us now state the theorem:

Theorem 3.9 (Method of Lagrange Multipliers). *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ be differentiable functions, with continuous partial derivatives. Consider the level set L_c of the function g (i.e. all points (x, y) where $g(x, y) = c$), and suppose $\mathbf{x_0} \in L_c$. Assume $\nabla g(\mathbf{x_0}) \neq \mathbf{0}$.*

If $f\big|_{L_c}$ (i.e. "the value of f on the set L_c") has a local minimum or maximum at the point $\mathbf{x_0}$, then there is a real number λ (called the **Lagrange Multiplier***) such that*

$$\nabla f(\mathbf{x_0}) = \lambda \nabla g(\mathbf{x_0}).$$

3.A Some Applications

Before proceeding to our next section where we discuss vector-valued functions in detail (as opposed to real-valued functions), let us briefly see some examples of how the notions we studied can be applied.

3.A.1 Partial Differential Equations

Partial differential equations (PDE) is *huge* field of mathematics, greatly linked to physics. The main concern of this field is the investigation and understanding of interesting equations involving partial derivatives, and then the understanding of the nature of their solutions (if those exist).

Some important examples include:

Laplace's Equation. Let $u : \mathbb{R}^3 \to \mathbb{R}$ be a function of three variables x, y, z, then *Laplace's equation* is

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = 0.$$

Poisson's Equation. Let $u : \mathbb{R}^3 \to \mathbb{R}$ and $f : \mathbb{R}^3 \to \mathbb{R}$ be two functions of three variables x, y, z, then *Poisson's equation* is

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = f(x, y, z).$$

The Heat Equation. Let $u : \mathbb{R}^4 \to \mathbb{R}$ be a function of four variables x, y, z and t, then the *Heat equation* is

$$\frac{1}{\alpha^2}\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}.$$

The Wave Equation. Let $u : \mathbb{R}^4 \to \mathbb{R}$ be a function of four variables x, y, z and t, then the *Wave equation* is

$$\frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}.$$

Schrödinger's Equation. Let $u : \mathbb{R}^4 \to \mathbb{R}$ be a function of four variables x, y, z and t, then *Schrödinger's equation* is

$$i\hbar\frac{\partial u}{\partial t} = -\frac{\hbar^2}{2m}\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}\right).$$

The first two examples are so-called ***steady-state*** problems, where there is no time dependence, while the last three examples are time-dependent problems (there's a time variable t).

3.A.2 Calculus of Variations

Calculus of variations is an extensive field of mathematics that deals with finding *functions* that minimize/maximize a certain quantity, rather than *points* in space. For example, given two points A and B in space (suppose A is above B), which is the curve joining them such that if a bead slid down it, it would reach reach B in the shortest time? This problem, called the ***Brachistochrone problem***, posed in the late 1600's by Johan Bernoulli, initiated this subject. The answer to the problem was given by Bernoulli himself, as well as his brother Jacob, Newton, l'Hôpital and Leibniz. It turns out that the ***cycloid*** is the (unique) solution to this problem.

Calculus of variations is the central tool used in classical mechanics, and in mathematics is studied under the framework of *functional analysis* - the analysis of functions.

Homework 3.10 (Reading due Friday 10/10; Problems due Wednesday 10/15).
Read *§4.1, §4.2*
Solve *§4.1: 13, 18; §4.2: 6, 12*

4 Vector-Valued Functions

Homework 4.1 (Reading due Wednesday 10/15; Problems due Wednesday 10/15).
Read *§4.3, §4.4*
Solve *§4.2: 10, 13ab, 13c*d*, 18*, 19*; §4.3: 4, 8, 18*

This chapter is more about intuition than mathematics, and getting a good grasp of the notions discussed here (on an intuitive level) will help with understanding the more precise mathematical definitions that will be introduced later. §4.1 is left for reading alone, as it is more descriptive, discusses more physical aspects, and tells interesting historical stories and anecdotes. However, I do ask you to

Read the differentiation rules for paths in $\mathbb{R}^3$ on page 262!

4.2 Arc Length

We begin by studying the important tool of measuring the length of a curve. This is a preliminary application of *integration along a path* which we will see later.

Initially it is unclear how to measure the length of an arbitrary curve in space. If it is not a a straight line, how would we proceed? Our best method would be to attach a wire along the curve, and then stretch it on the floor and measure its length there.

This is basically what we do. We use the relation

$$speed = \frac{distance}{time}$$

(on purpose we talk of *speed*, which is the magnitude of the *velocity*, and thus is just a number, not a vector).

Instead of measuring the length of the stretched wire with a ruler, we advance along it in a set speed, and measure how long it takes. Then, the total distance is given by

$$distance = speed \times time.$$

If we parametrize the curve by a path $\mathbf{c}(t)$, the speed at each point is given by $\|\mathbf{c}'(t)\|$. Thus, after a short time dt we advance a distance of $\|\mathbf{c}'(t)\|dt$. The full distance is

$$L(\mathbf{c}) = \int_{t_0}^{t_1} \|\mathbf{c}'(t)\|dt,$$

if we travel along the curve from time t_0 up to time t_1.

Example 4.2. *Let us calculate the length of the helix* $(\cos t, \sin t, t)$ *for* $0 \leq t \leq \pi$*:*

$\mathbf{c}(t) = (\cos t, \sin t, t)$*, so that* $\mathbf{c}'(t) = (-\sin t, \cos t, 1)$*, and* $\|\mathbf{c}'(t)\| = \sqrt{\sin^2 t + \cos^2 t + 1^2} = \sqrt{2}$*. Thus we get*

$$L = \int_0^\pi \sqrt{2} dt = \sqrt{2}\pi.$$

4.3 Vector Fields

4.3.1 Introduction

As we have seen in §2.1.2, we can define the notion of vector-valued functions. A ***vector field*** is the special case when a vector-valued function has the same dimension as the domain on which it is defined. For example, two-dimensional vectors on the plane, or three-dimensional vectors in space. In short, a function $f : \mathbb{R}^n \to \mathbb{R}^n$ is a vector field on $\mathbb{R}^n$. We can visualize the $2D, 3D$ cases by attaching a $2D$ or $3D$ arrow (respectively) to each point.

A vector field on $\mathbb{R}^3$ is usually denoted

$$\mathbf{F}(x, y, z) = (F_1(x, y, z), F_2(x, y, z), F_3(x, y, z)).$$

4.3.2 The gradient vector field

We have already seen that the gradient of a function is a vector-valued function which gives us the direction of greatest increase, and how big that increase is. This is what we call a ***gradient vector field***, and, as an example, one can think of the velocity arrows traced by water flowing down a mountain, as we have seen and discussed before. Given a gradient vector field $\mathbf{F}$, a function V whose gradient (or minus the gradient) is $\mathbf{F}$ (i.e. $\nabla V = \pm\mathbf{F}$), is usually called the ***potential*** associated to the field.

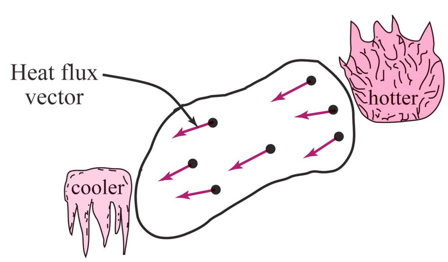


Figure 26: A temperature *gradient* generates a gradient vector field

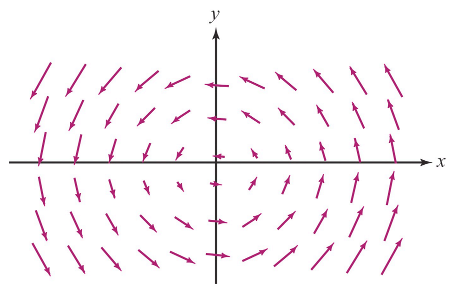


Figure 27: Is this a gradient vector field?

In Figures 26 and 27 we see two examples of vector fields. The first one is, indeed, a gradient vector field, as it is the result of taking minus the gradient of the temperature function in the region.

Is the second figure a gradient field? If it is, the arrows are supposed to represent the direction and magnitude of the steepest ascent at any given point. If we start from anywhere along the positive x-axis, going along the arrows, we are forced to turn around in a circle, and return to the same point. Is this possible if we keep going up? The answer is of course *NO*.

Later we will see methods of determining whether a vector field is a gradient vector field or not.

Example 4.3 (Gradient of the distance function)**.** *Consider the distance function*

$$r = \sqrt{x^2 + y^2 + z^2}.$$

The gradient vector field it generates is

$$\begin{aligned} \nabla r &= \left(\frac{\partial r}{\partial x}, \frac{\partial r}{\partial y}, \frac{\partial r}{\partial z}\right) \\ &= \left(\frac{x}{r}, \frac{y}{r}, \frac{z}{r}\right) \\ &= \frac{\mathbf{r}}{r} \end{aligned}$$

which is the unit radial vector field pointing outwards. Note that this calculation fails at $r = 0$.

Example 4.4 (Gravitational Vector Field)**.** *Newton's law of gravity, claims that the gravitational force any object feels is*

$$\mathbf{F} = -\frac{mMG}{r^3}\mathbf{r},$$

where M is the Earth's mass, m is the object's mass, G is some constant number, and $r = \|\mathbf{r}\|$, where $\mathbf{r}$ is the position of the object relative to the Earth's center.

Let us show that this force $\mathbf{F}$ is given by the **gravitational potential**

$$V = -\frac{mMG}{r}.$$

We need to show that $\mathbf{F} = -\nabla V$. Let us compute $\frac{\partial V}{\partial x}$:

$$\begin{aligned} -\frac{\partial V}{\partial x} &= mMG\frac{\partial}{\partial x}\left(\left(x^2+y^2+z^2\right)^{-1/2}\right) \\ &= -\frac{mMG}{r^3}x \end{aligned}$$

so that

$$\begin{aligned} -\nabla V &= -\frac{mMG}{r^3}(x,y,z) \\ &= -\frac{mMG}{r^3}\mathbf{r} \end{aligned}$$

as desired.

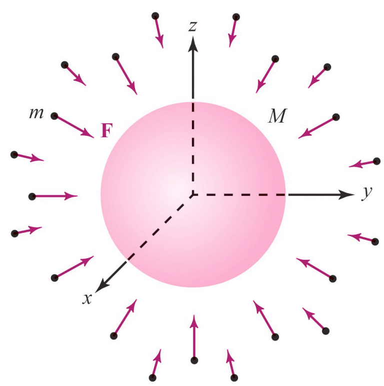


Figure 28: The gravitational force field

4.4 The Divergence and the Curl

Homework 4.5 (Reading due Friday 10/17; Problems due Monday 10/20).

Read *§4.4 again*

Solve *§4.3: 5, 6, 15, 20*; §4.4: 2, 5, 7, 9, 13, 32**

4.4.1 The "Del" Operator

We begin by making an abstract definition: We have thoroughly investigated the gradient of a function $f : \mathbb{R}^3 \to \mathbb{R}$, which is defined to be $\nabla f(x, y, z) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right)$. We now extend the definition of this symbol ∇: Abstractly, define ∇ to be the 'vector' (operator)

$$\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right).$$

The Gradient. In this setting, the gradient (that we have seen before) is simply a multiplication of the del operator and a function:

Definition 4.6 (Gradient)**.** *The* **gradient** *of a scalar function $f : \mathbb{R}^3 \to \mathbb{R}$ is defined*

$$\text{grad } f = \nabla f \quad := \quad \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right)$$

4.4.2 Divergence, curl

Two important operations associated with the del operator are the following operations on vector fields $\mathbf{F} = (F_1, F_2, F_3) : \mathbb{R}^3 \to \mathbb{R}^3$.

The Divergence.

Definition 4.7 (Divergence)**.** *The* **divergence** *of a vector field $\mathbf{F} : \mathbb{R}^3 \to \mathbb{R}^3$ is defined*

$$\text{div } \mathbf{F} = \nabla \cdot \mathbf{F} \quad := \quad \frac{\partial F_1}{\partial x} + \frac{\partial F_2}{\partial y} + \frac{\partial F_3}{\partial z}$$

The Curl.

Definition 4.8 (Curl)**.** *The* **curl** *of a vector field $\mathbf{F} : \mathbb{R}^3 \to \mathbb{R}^3$ is defined*

$$\begin{aligned} \text{curl } \mathbf{F} = \nabla \times \mathbf{F} \quad &:= \quad \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ F_1 & F_2 & F_3 \end{vmatrix} \\ &= \quad \left(\frac{\partial F_3}{\partial y} - \frac{\partial F_2}{\partial z}\right)\mathbf{i} + \left(\frac{\partial F_1}{\partial z} - \frac{\partial F_3}{\partial x}\right)\mathbf{j} + \left(\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y}\right)\mathbf{k}. \end{aligned}$$

Remark 4.9. *Notice that the three different operations the ∇ 'vector' induces are:*

	Acting on	***Resulting in***
grad	*Function* $\mathbb{R}^3 \to \mathbb{R}$	*Vector field*
div	*Vector field* $\mathbb{R}^3 \to \mathbb{R}^3$	*Function*
curl	*Vector field* $\mathbb{R}^3 \to \mathbb{R}^3$	*Vector field*

That is, the gradient *takes scalar-valued functions to vector-valued functions, the* divergence *takes vector-valued functions to scalar-valued functions and the* curl *takes vector-valued functions to vector-valued functions.*

4.4.3 The geometric interpretation of the divergence and the curl

Flow lines of a vector field. The idea of a flow line is a very fundamental idea in vector calculus, and, in fact, is very intuitive: Given a vector field, a ***flow line*** is simply a line that always follows the direction dictated by the arrow it encounters at each point. The formal definition is as follows:

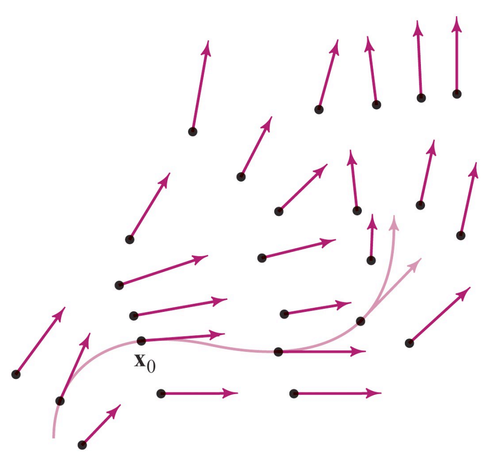


Figure 29: An example of a flow line

Definition 4.10 (Flow line)**.** *Let* $\mathbf{F}$ *be a vector field. Then a* **flow line** *for* $\mathbf{F}$ *is a path* $\mathbf{c}(t)$ *such that*

$$\mathbf{c}'(t) = \mathbf{F}(\mathbf{c}(t)).$$

In Section 4.3.2 we defined the gradient vector field $\mathbf{F}$ of a potential V. Now we see, that the flow lines of this $\mathbf{F}$ are exactly the paths along which V has the greatest change, and, equivalently, are always perpendicular to level sets of V. The concept of the *flow line* will help us understand what the divergence and the curl mean geometrically.

Geometric meaning of the divergence. The divergence of the vector field $\mathbf{F}$ at the point $\mathbf{x}_0$ measures to what extent the flow lines around $\mathbf{x}_0$ converge or diverge. If the flow lines *converge*, then the divergence is *negative*, while if the flow lines *diverge*, the divergence is *positive*. If the flow lines do neither, then the divergence is *zero*. Thinking of a gas (or a fluid), we have the following rules:

Positive divergence	Gas expanding
Negative divergence	Gas condensing
Zero divergence	Gas volume preserved

Slightly more precisely, although not fully rigorously, if we let W be a small neighborhood about the point $\mathbf{x}_0$, then we may consider the set $W(t)$, which is the result of letting the gas in W evolve up to time t. Then $W = W(0)$. Let $\mathcal{V}(t)$ be the volume of $W(t)$. Then we have the following approximation

$$\frac{1}{\mathcal{V}(0)}\frac{d}{dt}\mathcal{V}(t)\Big|_{t=0} \approx \text{div } \mathbf{F}(\mathbf{x}_0).$$

This approximation becomes more exact as W 'shrinks' to the point $\mathbf{x}_0$. We shall have

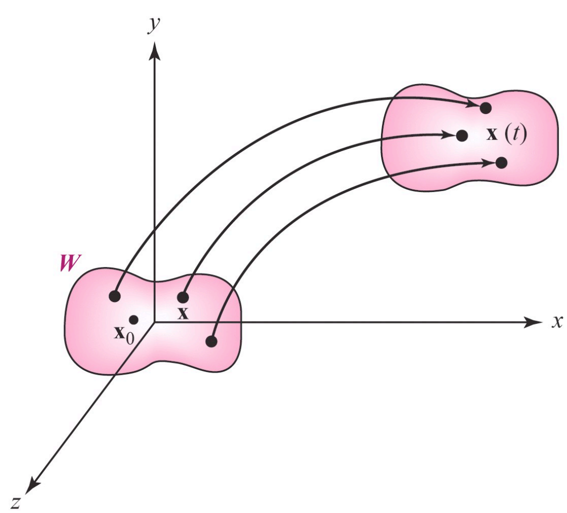


Figure 30: The divergence measures the infinitesimal change in volume

a better understanding of this formula once we study the change-of-variable formula in integration theory.

Homework 4.11 (Reading due Monday 10/20; Problems due Monday 10/20)**.**
Read *§5.1*
Solve *§4.3: 9, 11; §4.4: 11, 15, 26 (hint: use Proposition 4.14), 28, 31 and* ***prove*** *Proposition 4.16 below*

Example 4.12 (§4.4 Problem 8)**.** *Let* $\mathbf{F}(x, y) = (-3x, -y)$*. We begin by drawing some arrows representing the vector field, as well as some flow lines. (The arrows are all normalized, thus only indicate a direction, and all point inwards).*

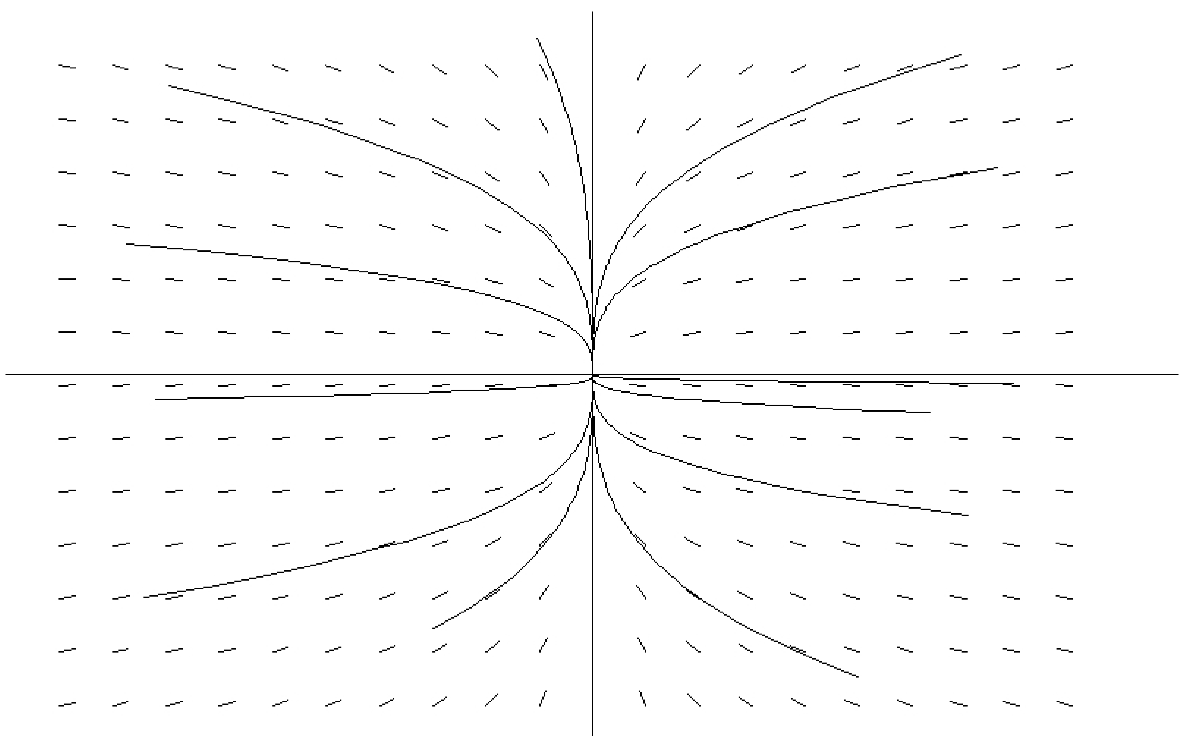


Figure 31: $\mathbf{F}(x, y) = (-3x, -y)$: The flow is inwards

Now, let us compute the divergence of $\mathbf{F}$*:*

$$\nabla \cdot \mathbf{F} = -3 - 1 = -4,$$

which expresses the fact that (thinking of this as a gas flow) the gas is condensing everywhere (not only in the origin). From this calculation, we also learn that the rate of condensation is the same everywhere (the divergence is -4 *for all* x, y*), contrary, maybe, to intuition gained from the figure.*

Geometric meaning of the curl. The curl of the vector field $\mathbf{F}$ at the point $\mathbf{x}_0$ measures to what extent the flow lines around $\mathbf{x}_0$ are rotating.

Example 4.13. *Let $\mathbf{\Omega} = (0,0,\omega)$ be a given vector pointing along the z-axis. Let $\mathbf{r} = (x,y,z)$ be the standard radial vector. Then $\mathbf{F} = \mathbf{\Omega} \times \mathbf{r}$ is a circular vector field in any plane parallel to the xy plane.*

$\mathbf{\Omega}$ expresses the axis of rotation around which the vector field $\mathbf{F}$ rotates. This is similar to Example 9 on page 300 of the book - you may want to compare the two.

A direct calculation shows us that $\nabla \times \mathbf{F} = 2\mathbf{\Omega}$ (verify this!). Thus, $\nabla \times \mathbf{F}$ captures the rotational character of the vector field $\mathbf{F}$.

4.4.4 Basic vector analysis identities

The gradient is curl free. This means that the curl of a gradient of a scalar valued function is always zero. Intuitively, this means that a gradient vector field has no rotations in it. Considering once again Figures 26 and 27 we see exactly that: A gradient vector field (Figure 26) has no rotational effects, while a vector field that *does* have rotational effects (Figure 27) cannot be a gradient vector field.

We write this as follows:

Proposition 4.14. *For any 'nice' function $f : \mathbb{R}^3 \to \mathbb{R}$*

$$\text{curl (grad } f) = \nabla \times (\nabla f) = (0,0,0) = \mathbf{0}.$$

Remark 4.15. *When we say 'nice', we mean that f is at least twice continuously differentiable, so that the equality of mixed partials holds.*

Proof.

$$\nabla f = (f_x, f_y, f_z),$$

so that

$$\begin{aligned} \nabla \times (\nabla f) &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \partial_x & \partial_y & \partial_z \\ f_x & f_y & f_z \end{vmatrix} \\ &= (\partial_y f_z - \partial_z f_y)\mathbf{i} - (\partial_x f_z - \partial_z f_x)\mathbf{j} + (\partial_x f_y - \partial_y f_x)\mathbf{k} \\ &= (f_{zy} - f_{yz})\mathbf{i} - (f_{zx} - f_{xz})\mathbf{j} + (f_{yx} - f_{xy})\mathbf{k} \\ &= \mathbf{0}. \end{aligned}$$

□

The curl is divergence free. A similar intuition applies to the converse of the previous statement: The curl of a vector field, which is a measurement of how much the vector field rotates, has no divergence, or no *sources* as a physicist would say.

Proposition 4.16. *For any 'nice' vector field* $\mathbf{F} : \mathbb{R}^3 \to \mathbb{R}^3$

$$\mathrm{div}\ (\mathrm{curl}\ \mathbf{F}) = \nabla \cdot (\nabla \times \mathbf{F}) = 0.$$

Proof. Homework! (similar to the above proof) □

Proposition 4.17 (Vector Analysis Identities)**.** *Given 'nice'* $f : \mathbb{R}^3 \to \mathbb{R}$, $g : \mathbb{R}^3 \to \mathbb{R}$ *and* $\mathbf{F} : \mathbb{R}^3 \to \mathbb{R}^3$, $\mathbf{G} : \mathbb{R}^3 \to \mathbb{R}^3$*, we have*

1. $\nabla(f+g) = \nabla f + \nabla g$
2. $\nabla(cf) = c\nabla f$*, for a constant* c
3. $\nabla(fg) = f\nabla g + g\nabla f$
4. $\nabla(f/g) = (g\nabla f - f\nabla g)/g^2$*, at points* $\mathbf{x}$ *where* $g(\mathbf{x}) \neq 0$
5. $\mathrm{div}\ (\mathbf{F}+\mathbf{G}) = \mathrm{div}\ \mathbf{F} + \mathrm{div}\ \mathbf{G}$
6. $\mathrm{curl}\ (\mathbf{F}+\mathbf{G}) = \mathrm{curl}\ \mathbf{F} + \mathrm{curl}\ \mathbf{G}$
7. $\mathrm{div}\ (f\mathbf{F}) = f\mathrm{div}\ \mathbf{F} + \mathbf{F} \cdot \nabla f$
8. $\mathrm{div}\ (\mathbf{F} \times \mathbf{G}) = \mathbf{G} \cdot \mathrm{curl}\ \mathbf{F} - \mathbf{F} \cdot \mathrm{curl}\ \mathbf{G}$
9. $\mathrm{div}\ (\mathrm{curl}\ \mathbf{F}) = 0$
10. $\mathrm{curl}\ (f\mathbf{F}) = f\mathrm{curl}\ \mathbf{F} + \nabla f \times \mathbf{F}$
11. $\mathrm{curl}\ (\mathrm{grad}\ f) = \mathbf{0}$
12. $\mathrm{div}\ (\nabla f \times \nabla g) = 0$

Notice that we have repeated the two previous propositions (which are items 9 and 11) for completeness of this list.

Proof. Let us prove some of these identities. The proof for 11 can be found above, while the proof of 9 is left as homework. The book has the proof for 7.

Let us prove 8:

On the one hand

$$
\begin{aligned}
\operatorname{div}(\mathbf{F}\times\mathbf{G}) &= \operatorname{div}\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ F_1 & F_2 & F_3 \\ G_1 & G_2 & G_3 \end{vmatrix} \\
&= \operatorname{div}\left[(F_2G_3 - F_3G_2)\mathbf{i} - (F_1G_3 - F_3G_1)\mathbf{j} + (F_1G_2 - F_2G_1)\mathbf{k}\right] \\
&= \frac{\partial}{\partial x}(F_2G_3 - F_3G_2) - \frac{\partial}{\partial y}(F_1G_3 - F_3G_1) + \frac{\partial}{\partial z}(F_1G_2 - F_2G_1) \\
&= G_3\partial_x F_2 - G_2\partial_x F_3 - G_3\partial_y F_1 + G_1\partial_y F_3 + G_2\partial_z F_1 - G_1\partial_z F_2 \\
&\quad + F_2\partial_x G_3 - F_3\partial_x G_2 - F_1\partial_y G_3 + F_3\partial_y G_1 + F_1\partial_z G_2 - F_2\partial_z G_1 \\
&= \mathbf{G}\cdot\left[(\partial_y F_3 - \partial_z F_2)\mathbf{i} - (\partial_z F_1 - \partial_x F_1)\mathbf{j} + (\partial_x F_2 - \partial_y F_1)\mathbf{k}\right] \\
&\quad - \mathbf{F}\cdot\left[(\partial_y G_3 - \partial_z G_2)\mathbf{i} - (\partial_z G_1 - \partial_x G_1)\mathbf{j} + (\partial_x G_2 - \partial_y G_1)\mathbf{k}\right]
\end{aligned}
$$

On the other hand

$$
\begin{aligned}
\mathbf{G}\cdot\operatorname{curl}\mathbf{F} - \mathbf{F}\cdot\operatorname{curl}\mathbf{G} &= \mathbf{G}\cdot\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \partial_x & \partial_y & \partial_z \\ F_1 & F_2 & F_3 \end{vmatrix} - \mathbf{F}\cdot\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \partial_x & \partial_y & \partial_z \\ G_1 & G_2 & G_3 \end{vmatrix} \\
&= \mathbf{G}\cdot\left[(\partial_y F_3 - \partial_z F_2)\mathbf{i} - (\partial_z F_1 - \partial_x F_1)\mathbf{j} + (\partial_x F_2 - \partial_y F_1)\mathbf{k}\right] \\
&\quad - \mathbf{F}\cdot\left[(\partial_y G_3 - \partial_z G_2)\mathbf{i} - (\partial_z G_1 - \partial_x G_1)\mathbf{j} + (\partial_x G_2 - \partial_y G_1)\mathbf{k}\right]
\end{aligned}
$$

so we are done. $\square$

5 Double and Triple Integrals

Homework 5.1 (Reading due Wednesday 10/22; Problems due Monday 10/27).
Read *§5.2*
Solve *§5.1: 1a, 3, 4; §5.2: 1b, 2b, 3*

5.1 Introduction

We begin by extending the single-variable definite integral that we already know, to a double-variable definite integral by intuitive means.

Similar to the one-dimensional definite integral which measure *area*, the two-dimensional definite integral measures *volume*. Let us compare these two integrals:

	Single Integral	**Double Integral**
Function involved	$f(x)$	$f(x,y)$
Domain	Interval $[a,b]$	Rectangle $R = [a,b] \times [c,d]$
Result	Area under the graph of f between a and b	Volume under the graph of f over R
Riemann sum	Split $[a,b]$ into small intervals and approximate f by rectangles to get approximate area	Split R into small slices parallel to yz plane according to partitions on $[a,b]$, then treat each slice as a function of y on $[c,d]$ and find area according to method described to left. Multiplying this by width of slice gives volume. Adding up slices gives an approximate volume

We discuss the definition of the double integral by Riemann sums extensively in the next section.

Cavalieri's Principle. In simple cases, we are able to exactly mimic the one-dimensional Riemann sum approach in a basic way: Given some object whose volume we want to calculate, suppose it is lying between a and b along the x-axis, and suppose the area of each cross section of the object at each given x value is $A(x)$. Then its volume is simply given by

$$Volume = \int_a^b A(x)dx$$

The Riemann sum explanation for this, would be given by viewing the object as made up of thin slices, like a loaf of bread.

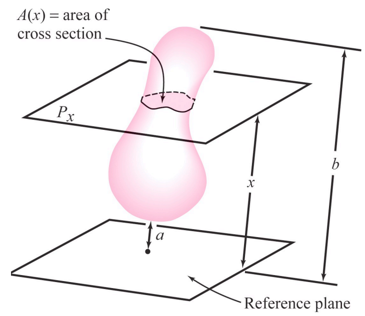


Figure 32: Cavalieri's Principle: $Volume = \int_a^b A(x)dx$

5.2 The Double Integral Over a Rectangle

Here we define the double integral in a more rigorous way. We first present the statement of our main theorem, as well as some results and important properties of the (double) integral. We then present some results to set the ground, and we finally prove our theorem.

5.2.1 Basic definitions

The Riemann Sum. The definition of the Riemann Sum for a function $f(x, y)$ of two variables is completely equivalent to the definition already known for the single-variable case. We recall, that in the **single variable** case, the n^{th} Riemann sum for a function

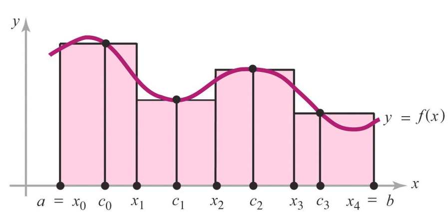


Figure 33: A Riemann sum in the single-variable case ($n = 4$ here)

$f(x)$ on the interval $[a, b]$ is given by:

1. Subdividing the interval into n identical segments, of length $\Delta x = (b - a)/n$ each, labeling the division points by $a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$.

2. Choosing *random* points $c_j \in [x_j, x_{j+1}]$.

3. Approximating f by rectangles with $[x_j, x_{j+1}]$ as bases, and with heights $f(c_j)$.

4. Adding up the areas of all these rectangles, we call the sum S_n:

$$S_n = \sum_{i=0}^{n-1} f(c_i)\Delta x.$$

Now the integral is defined to be the limit of S_n as n tends to ∞, if this limit exists for *any* choice of the points c_j:

$$\int_a^b f(x)dx = \lim_{n\to\infty} S_n.$$

If the integral of f exists, we say that f is ***integrable***.

The **two variable** case is *precisely* the same, with the only difference being in the dimensionality of all objects we deal with. The n^{th} Riemann sum for a function $f(x, y)$ on the rectangle $R = [a, b] \times [c, d]$ is given by:

1. Subdividing the rectangle into n^2 identical sub-rectangles, of size $\Delta x \times \Delta y = \frac{b-a}{n} \times \frac{c-d}{n}$ each, labeling the division points by $a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$ and $c = y_0 < y_1 < \cdots < y_{n-1} < y_n = d$ along the x and y axes respectively, and indexing the rectangles as $R_{ij} = [x_i, x_{i+1}] \times [y_j, y_{j+1}]$ where $0 \le i \le n-1$ and $0 \le j \le n-1$.
2. Choosing *random* points $\mathbf{c}_{ij} \in R_{ij}$.
3. Approximating f by rectangular boxes with bases R_{ij}, and heights $f(\mathbf{c}_{ij})$.
4. Adding up the volumes of all these boxes, we call the sum S_n.

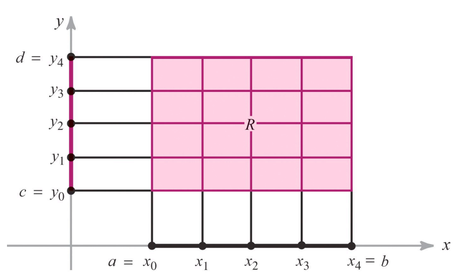


Figure 34: A partition in the double-variable case ($n = 4$ here)

$$S_n = \sum_{i,j=0}^{n-1} f(\mathbf{c}_{ij})\Delta x\Delta y.$$

Definition of the Double Integral. Now the integral is defined to be the limit of S_n as n tends to ∞, if this limit exists for *any* choice of the points $\mathbf{c}_{ij}$:

$$\int\int_R f(x,y)dA = \lim_{n\to\infty} S_n.$$

If the double-integral of f exists, we say that f is ***integrable***.

Example 5.2 (When everything fails!)**.** *This is an example of a function of a single-variable, for which the Riemann sums do not converge, and thus this is an example of a function that is* not *integrable.*

Let $f(x)$ be defined as follows, with domain $[0,1]$:

$$f(x) = \begin{cases} 1 & \textit{if } x \textit{ is rational} \\ 0 & \textit{if } x \textit{ is irrational} \end{cases}$$

Then, given a partition $\left\{0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1\right\}$ of the interval $[0,1]$, if we choose the points c_j to be rational, the corresponding Riemann sum is always 1, whereas if we choose the points c_j to be irrational, the corresponding Riemann sum is always 0.

This implies that the limit $\lim_{n\to\infty} S_n$ does not exist in this case, so that this function is **not integrable**.

Remark 5.3. *Those of you who will learn more advanced analysis courses (e.g. if you are a math, applied math or physics major) will see that one can define a more general type of integral (called a* **Lebesgue integral***), for which this function is integrable. This topic is well outside the scope of our course.*

Homework 5.4 (Reading due Friday 10/24; Problems due Monday 10/27)**.**
Read *§5.3*
Solve *§5.2: 6, 9, 12**

5.2.2 Iterated integration

Consider a function $f(x, y)$ defined on a rectangle $R = [a, b] \times [b, c]$. Let $x_0 \in [a, b]$ be a fixed number between a and b. Consider the cross section parallel to the yz plane passing through x_0. Suppose we are able to calculate the area of this cross section. Call it

$$A(x_0) = \int_c^d f(x_0, y)dy.$$

Now we can apply Cavalieri's principle to calculate the volume under the graph of f over R:

$$\begin{aligned} Volume &= \int_a^b A(x)dx \\ &= \int_a^b \left[\int_c^d f(x, y)dy\right] dx \end{aligned}$$

Similarly, if we are able to calculate the area of a cross section at any y_0, we get

$$Volume = \int_c^d \left[\int_a^b f(x, y)dx\right] dy.$$

Remark 5.5 (***Intuition vs. Rigorous mathematics***)**.** *Our intuition tells us that the double integral measures volume under a graph, in a similar way to the (intuitive) understanding that an integral of a single variable measures area under a graph. But we*

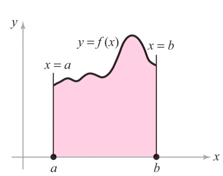


Figure 35: $\int_a^b f(x)dx$ =Area under the graph between a and b

must understand that we have only one method to measure length/area/volume: Using **rectangles***, which are our basic building blocks, and we know how to define length/area/ volume for them.*

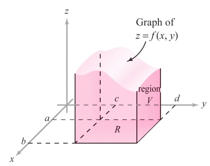


Figure 36: $\int\int_R f(x,y)dA$ =Volume under the graph over R

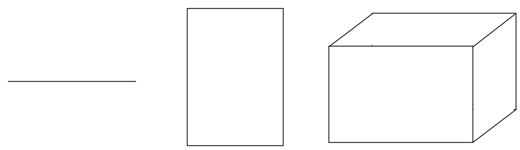


Figure 37: One-, two- and three-dimensional rectangles (boxes)

For example, to measure the length of a string, we must take a ruler and see how many centimeters fit along the string. To measure the area of a piece of paper, we must check how many rectangles (whose area we know) we can fit on it. To measure the volume of a loaf of bread, we must check how many boxes (three-dimensional rectangles) we can fit in it.

While our intuition is usually correct, we must put it aside (at times), and do the mathematics in a rigorous way.

This explains why we begin by denoting the double integral as $\int \cdots dA$ *and not* $\int \cdots dxdy$*:* dA *is the correct way to denote a small bit of area, not* $dxdy$*, which is a notation that implies that there's some preferred order -* x *first, then* y*. Fubini's theorem is going to set this issue straight.*

5.2.3 Important results

We begin with a theorem, which we prove in almost full detail and generality below. We then present some important, basic properties of the double-integral (and any integral, as it turns out).

Theorem 5.6 (Integrability of continuous functions on closed rectangles)**.** *Any continuous function defined on a closed rectangle* R *is integrable.*

Before proving this theorem, we turn to some important properties of the double integral:

1. **Linearity**:
$$\int\int_R (f+g)dA = \int\int_R f\ dA + \int\int_R g\ dA.$$

2. **Homogeneity**:
$$\int\int_R (cf)dA = c\int\int_R f\ dA.$$
3. **Monotonicity**: If $f(x,y) \geq g(x,y)$, then
$$\int\int_R f\ dA \geq \int\int_R g\ dA.$$
4. **Additivity**: If f is integrable on a rectangle R_1, and is also integrable on another rectangle R_2 (R_1 and R_2 disjoint), and if $R = R_1 \cup R_2$ is a rectangle, then
$$\int\int_R f\ dA = \int\int_{R_1} f\ dA + \int\int_{R_2} f\ dA.$$

Proposition 5.7. *f is integrable implies that $|f|$ is integrable too.*

Remark 5.8. *The converse to this proposition is not true.*

Corollary 5.9. *If f is integrable on some rectangle R, then*
$$\left|\int_R f\ dA\right| \leq \int_R |f|\,dA$$

Fubini's Theorem. In a nutshell, Fubini's theorem is a special case of Cavalieri's principle which we saw above: It clarifies the fact that it doesn't (usually) matter if we measure volume by first integrating along the x-axis and then along the y-axis, or the other way around.

Theorem 5.10 (Fubini's Theorem)**.** *Let f be a continuous function over some domain $R = [a,b] \times [c,d]$. Then*
$$\int_a^b \int_c^d f(x,y)dydx = \int_c^d \int_a^b f(x,y)dxdy = \iint_R f(x,y)dA$$

The book contains a more general version of this theorem as well (page 336), which we are not covering. You may wish to read the statement of that theorem, and try to see if you are able to find the subtle issues it addresses.

5.2.4 Proof of Theorem 5.6

We now turn to prove Theorem 5.6, the first main proof we present in this course. To make it simpler, we prove the theorem for functions of a single-variable, over a closed interval, and note that the *exact* same proof works in higher dimensions. Thus, let us restate the theorem for this case:

Theorem 5.11. *Any continuous function $f(x)$ defined on a closed interval $[a, b]$ is integrable there.*

We first introduce some definitions and preliminary results, to set the ground for the proof of the above theorem. We give an alternative definition of the *integral*, which turns out to be equivalent to the one we already know (we won't prove the equivalence of the two definitions here). The definitions we make here are in the spirit of Modern Mathematics, and are a good introduction to how *real analysis* looks like.

Definition 5.12 (Step function)**.** *A* **step function** *on $[a, b] \subseteq \mathbb{R}$ is a function $s : [a, b] \to \mathbb{R}$ such that*

$$s(x) = \sum_{n=1}^{M} c_n \chi_{[a_n, b_n]}(x)$$

where $c_n \in \mathbb{R}$ $(n = 1, \ldots, M)$ are scalars and $\chi_{[a_n, b_n]}$ $(n = 1, \ldots, M)$ are functions defined by

$$\chi_{[a_n, b_n]}(x) = \begin{cases} 1 & \text{if } x \in [a_n, b_n] \\ 0 & \text{otherwise} \end{cases}$$

with $(a_n, b_n) \subseteq (a, b)$ pairwise disjoint intervals (so that two intervals $[a_n, b_n]$ and $[a_k, b_k]$ can have at most one point in common), and $\cup_{n=1}^{M} [a_n, b_n] = [a, b]$.

(There is some issue with the endpoints of the subintervals - don't worry about it. We skip investing the extra time to take care of this issue. You may want to think what is the issue exactly?)

Definition 5.13 (The integral of a step function)**.** *Given a step function $s : [a, b] \to \mathbb{R}$ defined as above, we define its* **integral** *to be*

$$\int_a^b s(x)\, dx = \sum_{n=1}^{M} c_n \cdot (b_n - a_n)$$

Definition 5.14 (Alternative definition of integrability)**.** *We say that a function $f : [a, b] \to \mathbb{R}$ is* **integrable** *if for any $\epsilon > 0$ there exist step functions s_1 and s_2 such that $s_1(x) \leq f(x) \leq s_2(x)$ for all $x \in [a, b]$, and*

$$\int_a^b [s_2(x) - s_1(x)] dx < \epsilon.$$

(Note that $s_2 - s_1$ is also a step function)

Definition 5.15 (Uniform continuity)**.** *A function $f : [a, b] \to \mathbb{R}$ is said to be* **uniformly continuous** *if for any $\epsilon > 0$ there exists $\delta > 0$ such that $|x_1 - x_2| < \delta$ implies that*

$$|f(x_1) - f(x_2)| < \epsilon$$

for all $x_1, x_2 \in [a, b]$.

Theorem 5.16 (Minimum-maximum principle)**.** *A continuous function $f(x)$ on a closed interval $[a, b]$ attains its minimum and maximum values, i.e. there exist points c_{min} and c_{max} in $[a, b]$, such that for every $x \in [a, b]$,*

$$f(c_{min}) \leq f(x) \leq f(c_{max}).$$

Theorem 5.17 (Uniform continuity principle)**.** *A continuous function $f(x)$ on a closed interval $[a, b]$ is uniformly continuous.*

We postpone the proofs of these two theorems, and first prove Theorem 5.6 (or, rather, its one-dimensional version, Theorem 5.11) given these theorems.

Proof of Theorem 5.11. Fix $\epsilon > 0$. Let $\bar{\epsilon} = \epsilon/(b - a)$. Let $\delta > 0$ be the one attained by applying Theorem 5.17 to f with $\bar{\epsilon}$.

Partition the interval $[a, b]$ into $a = x_0 < x_1 < \cdots < x_{N-1} < x_N = b$ with $x_{i+1} - x_i < \delta$ for all $i = 0, 1, \ldots, N - 1$.

Consider one of these subintervals $[x_k, x_{k+1}]$: By Theorem 5.16 there exist points $c_k^m, c_k^M \in [x_k, x_{k+1}]$ on which f attains its minimum and maximum respectively on $[x_k, x_{k+1}]$.

By construction, $f(c_k^M) - f(c_k^m) < \bar{\epsilon}$.

Define two step functions s_1, s_2 as follows:

$$\begin{aligned} s_1(x) &= \sum_{i=0}^{N-1} f(c_i^m)\chi_{[x_i, x_{i+1}]}(x) \\ s_2(x) &= \sum_{i=0}^{N-1} f(c_i^M)\chi_{[x_i, x_{i+1}]}(x). \end{aligned}$$

These two step functions bound f from above and from below very closely: $s_1(x) \leq f(x) \leq s_2(x)$. We are only left with making sure that f satisfies the definition of integrability given above.

$$\begin{aligned}
\int_a^b [s_2(x) - s_1(x)]dx &= \sum_{i=0}^{N-1} [(f(c_i^M) - f(c_i^m)) \cdot (x_{i+1} - x_i)] \\
&< \sum_{i=0}^{N-1} [\bar{\epsilon} \cdot (x_{i+1} - x_i)] \\
&= \bar{\epsilon} \cdot (x_1 - x_0 + x_2 - x_1 + \cdots + x_N - x_{N-1}) \\
&= \bar{\epsilon} \cdot (x_N - x_0) \\
&= \bar{\epsilon} \cdot (b - a) \\
&= \epsilon.
\end{aligned}$$

□

Proof of Theorem 5.16. We want to find $c_{min}, c_{max} \in [a, b]$ such that

$$f(c_{min}) \leq f(x) \leq f(c_{max})$$

for all $x \in [a, b]$.

Let us show how to find c_{max}; finding c_{min} is done in precisely the same fashion. We quote a crucial axiom for the real numbers:

Axiom 5.18 (Bolzano-Weierstrass Axiom for the Real Numbers)**.** *Any bounded sequence x_n of real numbers, has a convergent subsequence x_{n_k}.*

We begin by showing that f is bounded on $[a, b]$**.** By contradiction, suppose it is not bounded from above. Then there exist points $x_n \in [a, b]$, $n = 1, 2, 3 \ldots$, such that $f(x_n) > n$, by the contradiction assumption. Now, $x_n \in [a, b]$ is a bounded sequence, hence we can extract a convergent subsequence x_{n_k}. I.e., there exists some point $x^* \in [a, b]$ such that as $k \to \infty$, $x_{n_k} \to x^*$. Since f is continuous on $[a, b]$, $f(x_{n_k}) \to f(x^*)$ as $k \to \infty$. But $f(x_{n_k}) > n_k \to \infty$ as $k \to \infty$ while $f(x^*) < \infty$, a contradiction.

Now we can finish the proof. We know that f is bounded on $[a, b]$. Let M_1 be a number so large, that $f(x) \leq M_1$ for all $x \in [a, b]$. If there exists x for which $f(x) = M_1$ we are done - this x is a maximum for f on $[a, b]$.

Otherwise, let x_1 be any number in $[a, b]$. Define $a = M_1 - f(x_1)$.

Step S for x_1 and M_1
Consider the midpoint M' between $f(x_1)$ and M_1.
If there exists $x_2 \in [a, b]$ for which $f(x_2) > M'$ define $M_2 = M_1$.
Otherwise, pick any $x_2 \in [a, b]$ with $f(x_1) < f(x_2) \leq M'$, and define $M_2 = M'$.

Thus

$$|M_2 - f(x_2)| < \frac{a}{2}.$$

Repeat **Step S** with x_2 and M_2, to get x_3, M_3, then repeat **Step S** once again to get x_4, M_4, et cetera, to finally get two sequences x_k and M_k satisfying

1. $f(x_k)$ an increasing sequence
2. M_k a decreasing sequence
3. $f(x) \leq M_k$ for all k and all $x \in [a, b]$
4. $|M_k - f(x_k)| < \dfrac{a}{2^{k-1}}$

Thus, $f(x_k)$ is an increasing sequence, bounded from above, M_k is a decreasing sequence, bounded from below, so that both have limits. Due to (4) above, those limits are, in fact, one limit:

$$M_k \text{ and } f(x_k) \longrightarrow M^* \qquad \text{as } k \to \infty$$

By (3) above, we have $f(x) \leq M^*$ for all $x \in [a, b]$.

x_k is a bounded sequence (in $[a, b]$), and so it has a convergent subsequence $x_{k_j} \to x^* \in [a, b]$ as $j \to \infty$. But, since f is continuous, $f(x_{k_j}) \to f(x^*)$. Since a sequence can have at most one limit, we conclude $f(x^*) = M^*$, so that, by construction, x^* is a point where the maximum is achieved. □

Proof of Theorem 5.17. This proof is left as a **serious** bonus assignment. Hint: Prove by contradiction. □

5.3 The Double Integral Oven More General Regions

Homework 5.19 (Reading due Wednesday 10/29; Problems due Monday 11/3).
Read *§5.4*
Solve *§5.3: 1cd, 2abd, 3, 15**

5.3.1 Semi-rigorous treatment

If the domain of our function $f(x,y)$ is some set $D \subseteq \mathbb{R}^2$ which is not necessarily a rectangle, we run into trouble, since our definition of the integral as the limit of Riemann sums (if those converge) used the fact that the domain was some rectangle R which we could split into a 'nice' grid. For each small rectangle in the grid the treatment is easy - we multiply the area of the base and the value of the function at a random point $\mathbf{c}_{jk}$ there (see the discussion of Riemann sums for functions of two-variables above).

Unfortunately, most sets D which are *not* rectangles, cannot be 'nicely' split into a grid of rectangles, and we have a problem. We solve this problem by enclosing D in a big rectangle R. This rectangle *can* be split into a grid.

Now we only need to take care of the function f which is only defined on D. It is a theorem (which we don't prove here), that if we extend f to a function f^* on R, so that f and f^* coincide on D, and f^* is 0 outside D, f^* is integrable on R. Since f^* is 0 outside D, we can intuitively understand that our desired result is

$$\iint_D f\, dA = \iint_R f^*\, dA.$$

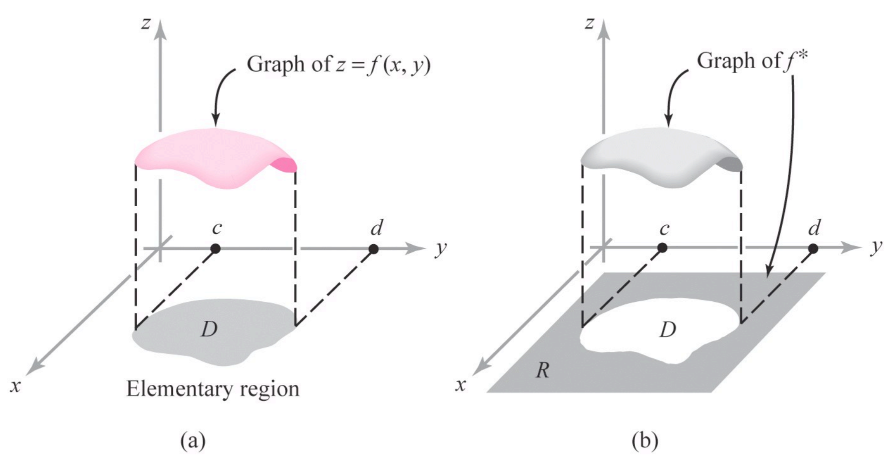


Figure 38: The function f and its extension f^*

Homework 5.20 (Reading due Friday 10/31; Problems due Monday 11/3).
Read *§5.5*
Solve *§5.3: 4, 6, 7, 11*

5.3.2 Practical computing methods

Elementary regions. When seeking actual practical computing methods, one needs to look at the set D, and see if its shape may facilitate reduction to iterated integrals. For this purpose, we introduce the notion of an ***elementary region***: a region that is bounded by graphs of functions, as follows:

y-simple regions. The first type of elementary region, is one that is bounded from 'below' and from 'above' by functions of x, say $\phi_1(x)$ and $\phi_2(x)$.

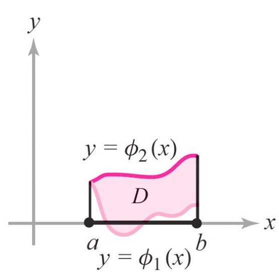


Figure 39: A y-simple region

x-simple regions. The second type of elementary region, is one that is bounded from the 'left' and from the 'right' by functions of y, say $\psi_1(y)$ and $\psi_2(y)$.

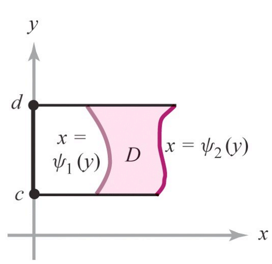


Figure 40: An x-simple region

We usually call a region D that is both x and y -simple, a ***simple*** region. The unit disk (ball) $x^2 + y^2 < 1$ is a good example of such a region.

Reduction to iterated integrals. We are now ready to state the main result concerning the actual calculation of double-integrals over general domains: Because of the way we have set up our integral over the region D as an integral over a bigger rectangle R with the extension function f^* as defined above, we can reduce to iterated integrals, as discussed in §5.2.

We have the results:

1. If D is y-simple, then

$$\iint_D f(x,y)\,dA = \int_a^b \left[\int_{\phi_1(x)}^{\phi_2(x)} f(x,y)\,dy\right] dx$$

2. If D is x-simple, then

$$\iint_D f(x,y)\,dA = \int_c^d \left[\int_{\psi_1(y)}^{\psi_2(y)} f(x,y)\,dx\right] dy$$

In our first calculus course we usually practice integrating tough expressions. Here, our purpose is not to calculate tough-to-compute expressions in two-variables; we want to focus on being able to *set up* the double-integral correctly. Once that is done, performing the actual integration is usually not the problem. Thus, here are some **tips for setting up the double integral:**

1. Draw the domain of integration D, in the most accurate and scaled way possible.
2. Determine if D is y-simple, x-simple, or both.
3. Determine which way the integration will be the simplest.

The last point will be discussed in greater detail in the next section.

An important thing to remember when reducing to iterated integrals, is that the lower and upper bounds of the inner integral *cannot* contain the variable of that integration: I.e., consider the expression $\int_c^d \left[\int_{\psi_1(y)}^{\psi_2(y)} f(x,y)\,dx\right] dy$: The inner integral, which is with respect to x, has as lower and upper bounds functions of y, *never* x!

Calculating area. We calculate the area of the domain D by integrating the function $f(x,y) = 1$ on D. For example, the area of the unit disk is given by the integral

$$\int_{-1}^1 \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} 1\,dy\,dx.$$

5.4 Changing the Order of Integration

f we consider the very last example given on the last page - that of the unit disk - we see that the disk in a *simple* region (i.e. both x and y -simple), and thus the integral written above may be written in two ways:

$$\begin{aligned} \iint_{D_1(\mathbf{0})} 1\, dA &= \int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} 1\, dy\, dx \\ &= \int_{-1}^{1}\int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} 1\, dx\, dy \end{aligned}$$

where $D_1(\mathbf{0})$ is the unit disk. In the integral in the first row (on the right), we think of the disk as a y-simple region, and in the integral in the second row we think of the disk as an x-simple region.

Sometimes, integrals over simple regions are very hard to calculate when the region is considered as x-simple, but easy when it is considered as y-simple, or vice-versa. For this purpose, it is important to be able to figure out how to change the order of integration. **The most crucial step is to be able to draw the domain D in an accurate way**.

Example 5.21. *Compute*

$$\int_0^1\int_y^1 \frac{\sin x}{x}\, dx\, dy.$$

Solution. *We first note the our domain D consists of all pairs (x, y) such that*

$$0 \le y \le 1 \qquad\qquad y \le x \le 1$$

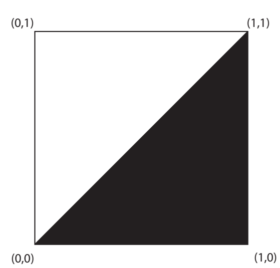


Figure 41: The domain D

Considering the figure, we may write the domain D as

$$0 \le y \le x \qquad\qquad 0 \le x \le 1$$

and we get the integral

$$\begin{aligned}
\int_0^1 \int_y^1 \frac{\sin x}{x}\, dx\, dy &= \int_0^1 \int_0^x \frac{\sin x}{x}\, dy\, dx \\
&= \int_0^1 \left[\frac{\sin x}{x}\, y \right]_0^x dx \\
&= \int_0^1 \left[\frac{\sin x}{x}\, x - \frac{\sin x}{x}\, 0 \right] dx \\
&= \int_0^1 \sin x\, dx \\
&= 1 - \cos 1.
\end{aligned}$$

Homework 5.22 (Reading due Monday 11/3; Problems due Monday 11/3).
Read *§1.4, §6.1*
Solve *§5.4: 1a, 2ab, 7; §5.5: 3, 7, 9, 13, 15, 23*

5.5 Triple Integrals

There is *nothing new* in the treatment of triple integrals (or any higher dimensional integrals, for that matter) compared to what we have already seen. All the tools and definitions we have developed, hold for the three-dimensional case. The only difference is that the extra dimension adds another integral. The properties that still hold are:

1. Linearity
2. Homogeneity
3. Monotonicity
4. Additivity

As well as

- Fubini's Theorem
- Integration over non-rectangular domains
- Iterated integrals
- Changing the order of integration

As for elementary regions - those are slightly more complicated, and are of the forms

$$\begin{aligned} a &\le x \le && b \\ \phi_1(x) &\le y \le && \phi_2(x) \\ \gamma_1(x,y) &\le z \le && \gamma_2(x,y) \end{aligned}$$

or

$$\begin{aligned} c &\le y \le && d \\ \psi_1(y) &\le x \le && \psi_2(y) \\ \gamma_1(x,y) &\le z \le && \gamma_2(x,y) \end{aligned}$$

where in both cases the third condition captures the fact that the domain at hand is three-dimensional. (We could describe a three-dimensional region in more than just these two ways. Think of an example. See page 360 of the book).

Drawing the regions is the toughest part of triple-integration. It is important to be able to produce fairly accurate sketches, so that it is clear how to perform the integration.

Homework 5.23 (Reading due Wednesday 11/5; Problems due Friday 11/7).
Read *§6.2*
Solve *§1.4: 1*

6 Change of Variables and Applications

6.1 The Geometry of Maps $\mathbb{R}^2 \to \mathbb{R}^2$

6.1.1 Reminder: One-to-one and onto maps $\mathbb{R} \to \mathbb{R}$

Our purpose is the investigation of deformation of domains in the plane by mappings. Before we perform this more complicated investigation, let us begin by the simpler investigation of mappings of the real number line to itself.

Definition 6.1 (One-to-one)**.** *A mapping $T : D^* \subseteq \mathbb{R} \to \mathbb{R}$ is said to be* **one-to-one** *on D^* if for any $u, u' \in D^*$, $T(u) = T(u')$ implies that $u = u'$.*

That is, T doesn't map two different points in D^* to one point in the image.

Definition 6.2 (Onto)**.** *A mapping $T : D^* \subseteq \mathbb{R} \to D$ is said to be* **onto** *D if for any $x \in D$, there exists some $u \in D^*$ such that $T(u) = x$.*

That is, D is completely "covered" by T.

Example 6.3. *The map $T(u) = u^2$* ***is one-to-one*** *on the domain $D_1^* = [0, 10]$. To see this, we assume $u, u' \in [0, 10]$ are two points such that $u^2 = u'^2$. Thus $|u| = |u'|$. But since $u, u' \geq 0$, we may drop the absolute value signs, to get $u = u'$.*

T ***is onto*** *the set $D_1 = [0, 100]$. How do we show this? Let $x \in D_1$. Does there exist $u \in D^*$ such that $T(u) = x$? We know that $0 \leq x \leq 100$, so that $0 \leq \sqrt{x} \leq 10$. $u = \sqrt{u}$ is such that $T(u) = x$.*

However, T is ***not onto*** *the set $D_2 = [0, 110]$, since for any $100 < x \leq 110$, $\sqrt{x} > 10$ and thus $\sqrt{x} \notin D_1^*$.*

Also, T is ***not one-to-one*** *on the domain $D_2^* = [-1, 10]$, since there exist two points $u, u' \in D_2^*$ that are different, but $T(u) = T(u')$, for example 1 and -1, or 0.5 and -0.5, etc...*

6.1.2 One-to-one and onto maps $\mathbb{R}^2 \to \mathbb{R}^2$

The exact same definitions carry on to maps $\mathbb{R}^2 \to \mathbb{R}^2$:

Definition 6.4 (One-to-one)**.** *A mapping $T : D^* \subseteq \mathbb{R}^2 \to \mathbb{R}^2$ is said to be* **one-to-one** *on D^* if for any $(u, v), (u', v') \in D^*$, $T(u, v) = T(u', v')$ implies that $(u, v) = (u', v')$.*

Definition 6.5 (Onto)**.** *A mapping $T : D^* \subseteq \mathbb{R}^2 \to D$ is said to be* **onto** *D if for any $(x, y) \in D$, there exists some $(u, v) \in D^*$ such that $T(u, v) = (x, y)$.*

6.1.3 Linear maps $\mathbb{R}^2 \to \mathbb{R}^2$

Definition 6.6 (Linear map)**.** *We say that a mapping $T : \mathbb{R}^2 \to \mathbb{R}^2$ is* **linear***, if for any $(u, v) \in \mathbb{R}^2$, $T(u, v) = uT(1, 0) + vT(0, 1)$.*

That is, if $T(1, 0) = (a, c)$ and $T(0, 1) = (b, d)$, then the mapping T can be thought of as multiplication by the matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

since then

$$T(u, v) = A\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} au + bv \\ cu + dv \end{pmatrix} = u\begin{pmatrix} a \\ c \end{pmatrix} + v\begin{pmatrix} b \\ d \end{pmatrix}.$$

Intuitively, T is linear if it *distorts* the plane $\mathbb{R}^2$ "uniformly", i.e. the distortion of the plane is governed by how T "twists and turns" the axes.

What is most important for us, however, is the fact that **linear maps $\mathbb{R}^2 \to \mathbb{R}^2$ take parallelograms to parallelograms**, mapping sides to sides and vertices to vertices.

6.2 Change of Variables

6.2.1 Maps $\mathbb{R} \to \mathbb{R}$

Example 6.7. *Consider the integral*

$$\int_0^1 e^{2x}\, dx.$$

Let us solve this integral using the substitution

$$\begin{aligned} u &= 2x \\ du &= 2\, dx, \end{aligned}$$

to get

$$\int_0^1 e^{2x}\, dx = \int_0^1 \frac{1}{2} e^{2x}\, 2dx = \int_0^2 \frac{1}{2} e^{u}\, du = \frac{1}{2}\left(e^2 - 1\right).$$

Example 6.8. *Consider the integral*

$$\int_0^1 \sqrt{1 - x^2}\, dx.$$

Let us solve this integral using the substitution

$$\begin{aligned} x &= \sin u \\ dx &= \cos u \, du, \end{aligned}$$

to get

$$\int_0^1 \sqrt{1-x^2}\, dx = \int_0^{\pi/2} \sqrt{1-\sin^2 u}\ \cos u\, du = \int_0^{\pi/2} \cos^2 u\, du$$

which can now be solved using the identity $\cos^2 u = \frac{1}{2}(1+\cos 2u)$ *to get* $\pi/4$.

In both cases, we used (maybe without knowing) the general formula for ***change of variables (for functions of a single-variable)***:

$$\int_a^b f(x(u))\, \frac{dx}{du}\, du = \int_{x(a)}^{x(b)} f(x)\, dx$$

where f is continuous, and the function that takes u to $x(u)$ is continuously differentiable on $[a,b]$ (i.e. it is continuous, as is its derivative).

In the first case, we had $f(x) = e^{2x}, x(u) = \frac{1}{2}u$ and $a = 0, b = 2$, so we got

$$\int_0^2 e^{2\frac{1}{2}u}\, \frac{1}{2}\, du = \int_a^b f(x(u))\, \frac{dx}{du}\, du = \int_{x(a)}^{x(b)} f(x)\, dx = \int_0^1 e^{2x}\, dx$$

while in the second case, we had $f(x) = \sqrt{1-x^2}, x(u) = \sin u$ and $a = 0, b = \pi/2$, so we got

$$\int_0^{\pi/2} \sqrt{1-\sin^2 u}\ \cos u\, du = \int_a^b f(x(u))\, \frac{dx}{du}\, du = \int_{x(a)}^{x(b)} f(x)\, dx = \int_0^1 \sqrt{1-x^2}\, dx$$

6.2.2 Maps $\mathbb{R}^2 \to \mathbb{R}^2$

1. Reminder: Polar coordinates. We begin by recalling the main alternative description of a point on the plane $\mathbb{R}^2$: Instead of describing a point in terms of its displacement along the x and y axes, we describe it in terms of its displacement from the origin, and the angle from the x axis.

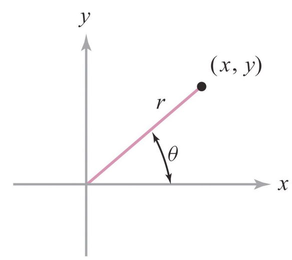


Figure 42: Polar coordinates

We have the relations:

$$\begin{array}{ll} x = r\cos\theta & r = \sqrt{x^2+y^2} \\ y = r\sin\theta & \theta = \arctan\frac{y}{x} \end{array}$$

where θ varies between 0 and 2π.

2. Integration in polar coordinates. Since the area of a small "rectangle" in polar coordinates is given by $(rd\theta)\cdot dr$, we get the formula

$$\iint f(x,y)\,dx\,dy = \iint f(r\cos\theta, r\sin\theta)\,r\,d\theta\,dr$$

where the integration is over the appropriate domains.

6.2.3 Maps $\mathbb{R}^3 \to \mathbb{R}^3$

1. Reminder: Cylindrical coordinates. Cylindrical coordinates are the three-dimensional extension of the two-dimensional polar coordinates. We simply add the "z"-coordinate to the existing r, θ-coordinates.

We have the relations:

$$\begin{array}{ll} x = r\cos\theta & r = \sqrt{x^2+y^2} \\ y = r\sin\theta & \theta = \arctan\frac{y}{x} \\ z = z & \end{array}$$

where θ varies between 0 and 2π.

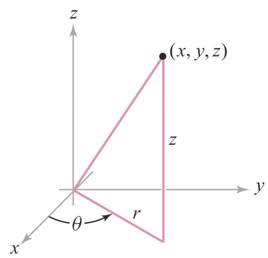


Figure 43: Cylindrical coordinates

2. Integration in cylindrical coordinates.

$$\iiint f(x,y,z)\;dx\;dy\;dz = \iiint f(r\cos\theta, r\sin\theta, z)\;r\;d\theta\;dr\;dz$$

where the integration is over the appropriate domains.

3. Reminder: Spherical coordinates. In spherical coordinates, we describe a point in $\mathbb{R}^3$ by its distance from the origin (ρ), its angle from the x-axis (θ) and its angle from the z-axis (ϕ). These two angles are similar to what we know as longitude and latitude respectively. The main difference is that latitude is measured from the xy plane instead of the z-axis.

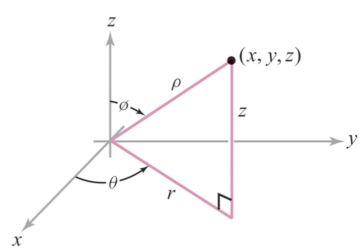


Figure 44: Spherical coordinates

We have the relations:

$$\begin{array}{ll} x = \rho\sin\phi\cos\theta & \rho = \sqrt{x^2+y^2+z^2} \\ y = \rho\sin\phi\sin\theta & \theta = \arctan\frac{y}{x} \\ z = \rho\cos\phi & \end{array}$$

where θ varies between 0 and 2π, and ϕ varies between 0 and π.

4. Integration in spherical coordinates.

$$\iiint f(x,y,z)\,dx\,dy\,dz = \iiint f(\rho\sin\phi\cos\theta, \rho\sin\phi\sin\theta, \rho\cos\phi)\,\rho^2\sin\phi\,d\theta\,d\rho\,d\phi$$

where the integration is over the appropriate domains.

6.2.4 Examples

Example 6.9. *Calculate the area of a disk of radius R.*

Solution 1. *We start by calculating in Cartesian coordinates:*

$$\int_{-R}^{R}\int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} 1\,dy\,dx = 2\int_{-R}^{R}\sqrt{R^2-x^2}\,dx$$

which we calculate using the substitution $x = R\sin u$.

Solution 2. *We calculate using polar coordinates:*

$$\begin{aligned}\int_0^R\int_0^{2\pi} 1\cdot r\,d\theta\,dr &= 2\pi\int_0^R r\,dr\\ &= \pi R^2.\end{aligned}$$

Notice that since the function which we were integrating on had no θ dependence, the integration with respect to θ simply gave us a 2π factor.

Example 6.10. *Calculate the volume of a sphere of radius R.*

Solution 1. *We start by calculating in Cartesian coordinates:*

$$\int_{-R}^{R}\int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}}\int_{-\sqrt{R^2-x^2-y^2}}^{\sqrt{R^2-x^2-y^2}} 1\,dz\,dy\,dx = 2\int_{-R}^{R}\int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}}\sqrt{R^2-x^2-y^2}\,dy\,dx$$

which becomes nasty.

Solution 2. *We calculate using spherical coordinates:*

$$\begin{aligned}\int_0^R\int_0^{2\pi}\int_0^{\pi} 1\cdot\rho^2\sin\phi\,d\phi\,d\theta\,d\rho &= 2\pi\int_0^R\int_0^{\pi}\rho^2\sin\phi\,d\phi\,d\rho\\ &= 2\pi\int_0^R \rho^2\left[-\cos\phi\right]_0^{\pi}\,d\rho\\ &= 4\pi\left[\frac{\rho^3}{3}\right]_0^R\\ &= \frac{4}{3}\pi R^3.\end{aligned}$$

Solution 3. *We may also calculate using cylindrical coordinates:*

$$\begin{aligned}\int_{-R}^{R}\int_{0}^{2\pi}\int_{0}^{\sqrt{R^2-z^2}} 1\cdot r\,dr\,d\theta\,dz &= 2\pi\int_{-R}^{R}\left[\frac{r^2}{2}\right]_0^{\sqrt{R^2-z^2}} dz\\ &= \pi\int_{-R}^{R}(R^2-z^2)\,dz\\ &= \pi\left[R^2 z-\frac{z^3}{3}\right]_{-R}^{R}\\ &= \pi\left(R^3-\frac{R^3}{3}-R^2(-R)+\frac{(-R)^3}{3}\right)\\ &= \frac{4}{3}\pi R^3.\end{aligned}$$

Example 6.11. *Calculate the volume of a cylinder with radius R and height h.*

Solution 1. *Cartesian coordinates:*

$$\int_0^h\int_{-R}^{R}\int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} 1\,dy\,dx\,dz = \int_0^h A(z)\,dz$$

where $A(z)=\pi R^2$ is the area of the disk of radius R which we calculated above. The resulting integral is exactly the one we would have gotten by applying Cavalieri's principle.

Solution 2. *Cylindrical coordinates:*

$$\begin{aligned}\int_0^h\int_0^{2\pi}\int_0^R 1\cdot r\,dr\,d\theta\,dz &= 2\pi h\int_0^R r\,dr\\ &= \pi R^2 h.\end{aligned}$$

Example 6.12. *Evaluate*

$$\iint_D \sin\left(x^2+y^2\right)\,dx\,dy$$

where D is the region in the first quadrant lying between the arcs of the circles

$$x^2+y^2=\pi/2 \qquad\qquad x^2+y^2=\pi.$$

Solution. *We calculate in polar coordinates. We notice that our ranges are $0\le\theta\le\pi/2$ and $\sqrt{\pi/2}\le r\le\sqrt{\pi}$ and use the fact that $x^2+y^2=r^2$ to get*

$$\int_{\sqrt{\pi/2}}^{\sqrt{\pi}}\int_0^{\pi/2}\sin\left(r^2\right)\cdot r\,d\theta\,dr = \frac{\pi}{2}\int_{\sqrt{\pi/2}}^{\sqrt{\pi}}\sin\left(r^2\right)\cdot r\,dr,$$

which we solve by substituting $u=r^2$ and $du=2r\,dr$ to get $\pi/4$.

6.2.5 Generalization: the Jacobian determinant

We saw that in the plane,

$$dx\ dx \rightsquigarrow r\ dr\ d\theta$$

and that in space

$$\begin{aligned} dx\ dy\ dz &\rightsquigarrow r\ dr\ d\theta\ dz \\ dx\ dy\ dz &\rightsquigarrow \rho^2 \sin\phi\ d\phi\ d\theta\ d\rho. \end{aligned}$$

These three cases, are merely examples of a general theorem, called ***Change of Variables theorem***. If we recall the general one-dimensional expression

$$\int_a^b f(x(u))\, \frac{dx}{du}\, du = \int_{x(a)}^{x(b)} f(x)\ dx,$$

then we discover that the factors r and $\rho^2 \sin\phi$ are in fact equivalent to the term $\frac{dx}{du}$ above: they all describe the change in the size of a length/area/volume element.

Let us define the ***Jacobian matrix*** of the transformation $u, v, w \to x, y, z$ (i.e. $x = x(u, v, w), y = y(u, v, w), z = z(u, v, w)$ are new coordinates that depend on u, v, w) as the matrix

$$\frac{\partial(x, y, z)}{\partial(u, v, w)} = \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} & \frac{\partial z}{\partial w} \end{pmatrix}$$

The **determinant** of this matrix turns out to be the ratio of the volume elements, measured in the x, y, z coordinates, vs. measured in the u, v, w coordinates.

Let us check this for the two-dimensional case of *polar coordinates*:

$$\begin{aligned} \left|\frac{\partial(x, y)}{\partial(r, \theta)}\right| &= \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{vmatrix} \\ &= \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} \\ &= r\cos^2\theta + r\sin^2\theta \\ &= r \end{aligned}$$

as required! The general formula is

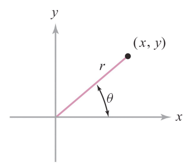
$$\iiint f(x, y, z)\ dx\ dy\ dz = \iiint f\left(x(u, v, w), y(u, v, w), z(u, v, w)\right) \left|\frac{\partial(x, y, z)}{\partial(u, v, w)}\right|\ du\ dv\ dw$$

where the integration is performed over the appropriate corresponding domains.

6.2.6 Summary

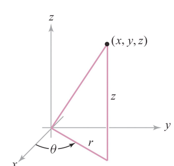
1. Integration in polar coordinates.

$$\iint f(x,y)\;dx\;dy=\iint f(r\cos\theta,r\sin\theta)\;r\;d\theta\;dr$$



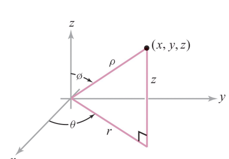
2. Integration in cylindrical coordinates.

$$\iiint f(x,y,z)\;dx\;dy\;dz=\iiint f(r\cos\theta,r\sin\theta,z)\;r\;d\theta\;dr\;dz$$



3. Integration in spherical coordinates.

$$\iiint f(x,y,z)\;dx\;dy\;dz=\iiint f(\rho\sin\phi\cos\theta,\rho\sin\phi\sin\theta,\rho\cos\phi)\;\rho^2\sin\phi\;d\theta\;d\rho\;d\phi$$



7 Path and Surface Integrals

7.1 Path Integrals

Homework 7.1 (Reading due Friday 11/14; Problems due Monday 11/17)**.**
__Read__ §7.1, 7.2
*__Solve__ §7.1: 1, 2, 6, 9, 12, 16**

We begin our final push towards the theorems that will culminate this course by extending our understanding of the integral of a single variable to integration of *scalar functions* ***along a path in space***. An immediate application of this would be to compute the total mass of a wire, given its density at each point.

The formula that will be specified below can be justified either using *Riemann sums*, or using the *change-of-variables* methods of the previous section. The former is motivated in the book (p. 423), thus let us motivate the latter.

Motivation. Suppose a scalar valued function $f(x, y, z)$ is defined along a path in $\mathbb{R}^3$ given by $\mathbf{c}(t)$, where $t \in [a, b]$. We wish to define the integral of f along this path. Well, we can first picture a particle traveling along the path, as time varies from a to b. For each time t, the particle on the path "sees" a specific value of the function – $f(\mathbf{c}(t))$. We can draw this function over the interval $[a, b]$, and we have thus achieved a function on the line: $t \mapsto f(\mathbf{c}(t))$.

BUT! What happens if we travel *along the same curve* in different speeds? Well, over the time axis, we get the same graph, but it will be stretched or contracted. This implies, as was the case in the previous section, that some correction term is needed, that will take into consideration the fact the we may trace the same curve in space in different speeds. This correction term *is* the speed $\|\mathbf{c}'(t)\|$! We get the formula

$$\int_{\mathbf{c}} f \, ds = \int_a^b f(\mathbf{c}(t)) \|\mathbf{c}'(t)\| \, dt$$

where the integral on the left is simply the notation for the line integral: the ds being the equivalent of the good old dx – while dx is an increment along the x axis, ds is an increment along the path.

Planar curves. In the special case that $\mathbf{c}(t)$ is in fact a path in the plane $\mathbb{R}^2$, it is easy to get a more intuitive understanding of what the path integral is: It is the integration of a function over the two-dimensional path $\mathbf{c}$ – the *"area of a fence"* as the book describes it. Please read p. 424 for this intuitive explanation.

7.2 Line Integrals

Homework 7.2 (Reading due Monday 11/17; Problems due Monday 11/17)**.**
Read *§7.2 again*
Solve *§7.2: 1, 2, 4, 5, 6*

7.2.1 The Arc-Length Differential (p. 278)

We now try to make more rigorous the definition of the path integral, and add to it the definition of the line integral. To do so, we discuss the concept of *arc length*: When

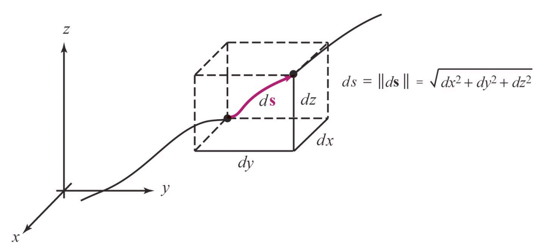


Figure 45: The arc length differential

integrating a function f of a single variable x, we use the symbol dx to denote a small increment along the x-axis. The integral is defined as the limit of the appropriate Riemann sums, and this dx differential is the limit of the small increments Δx. ds plays the same role, only along a curve: It is defined intrinsically, and is a property of the curve. If we could stretch the curve onto the axis and therefore turn it into a straight line segment, ds would identify with dx. But since this is not the case, we must define ds extrinsically, using the coordinate system xyz which we impose. Later we will see why it is important to keep this in mind.

Using the Pythagorean theorem we define

$$\begin{aligned} ds &= \sqrt{dx^2 + dy^2 + dz^2} \\ &= \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dy}\right)^2 + \left(\frac{dz}{dt}\right)^2}\, dt \\ &= \|\mathbf{c}'(t)\|\, dt \end{aligned}$$

where the last expressions are valid when the curve C is parametrized by a path $\mathbf{c}(t) = (x(t), y(t), z(t))$. ds is called the ***arc-length differential***. In the case of the path, we may also define

$$\begin{aligned} d\mathbf{s} &= \left(\frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt}\right) dt \\ &= \mathbf{c}'(t)\; dt \end{aligned}$$

$d\mathbf{s}$ is the ***infinitesimal displacement*** (and is a *vector*!).

The definition of the arc-length differential enables us to understand more clearly the definition of the path integral, by plugging in our expression for ds:

$$\int_{\mathbf{c}} f\; ds = \int_a^b f(\mathbf{c}(t))\|\mathbf{c}'(t)\|\; dt$$

where $\mathbf{c}(t) : [a, b] \to \mathbb{R}^3$.

More rigorously, we may also see this using Riemann sums: Consider a partitioning $a = t_0 < t_1 < t_2 < \cdots < t_{N-1} < t_N = b$ of the interval $[a, b]$. This partitioning maps to a partitioning (not necessarily equal partitioning) on the path (see Figure 46).

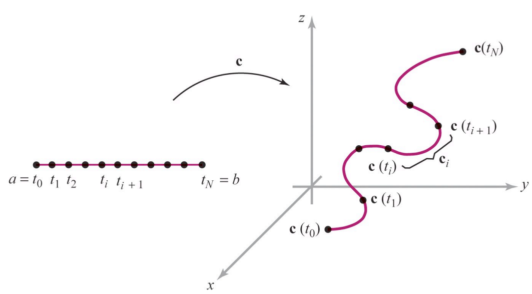


Figure 46: A partitioning of the interval $[a, b]$ translates to a partitioning of the path

Now, consider the i^{th} small element of the partitioning on the path $\mathbf{c}_i$. By our definition of arc-length, the length Δs_i of this element is

$$\Delta s_i = \int_{t_i}^{t_{i+1}} \|\mathbf{c}'(t)\|\; dt.$$

Let (x_i, y_i, z_i) be any point on the small element $\mathbf{c}_i$. As we let $N \to \infty$, the small elements on the path will get smaller and smaller, so that f won't vary too much on each of them. This suggests we should define the Riemann sum

$$S_N = \sum_{i=0}^{N-1} f(x_i, y_i, z_i)\Delta s_i.$$

By the mean value theorem, $\Delta s_i = \|\mathbf{c}'(t_i^*)\|\Delta t_i$, so that we may write

$$\begin{aligned}
\lim_{N\to\infty} S_N &= \lim_{N\to\infty}\sum_{i=0}^{N-1} f(x_i,y_i,z_i)\Delta s_i \\
&= \lim_{N\to\infty}\sum_{i=0}^{N-1} f(x_i,y_i,z_i)\|\mathbf{c}'(t_i^*)\|\Delta t_i \\
&= \int_a^b f(x(t),y(t),z(t))\|\mathbf{c}'(t)\|\,dt \\
&= \int_{\mathbf{c}} f(x,y,z)\,ds.
\end{aligned}$$

7.2.2 Intuition

We now want to define a path integral for vector fields; we call it a ***line integral***. The result we get will again be a scalar value. This is a very important integral in physics, as the following example shows:

Example 7.3. *Consider a bead that is attached to a straight wire given by the vector* **d***. If there's a constant force* **F** *that is pushing it along the wire, we say that the* work *done by the force in moving the bead along the wire is*

$$Work = \mathbf{F}\cdot\mathbf{d} = \|\mathbf{F}\|\ \|\mathbf{d}\|\ \cos\theta$$

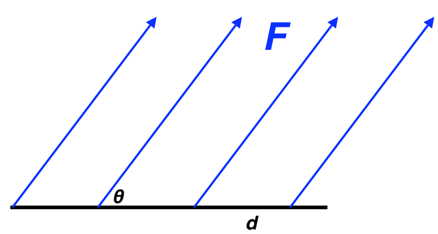


Figure 47: The force **F** is pushing the bead along the wire

That is, the work is the component of the force that points in the direction of movement, multiplied by the displacement.

Our goal is to generalize this concept, to instances where (1) the force field isn't necessarily constant and (2) the displacement doesn't occur along a straight line.

7.2.3 Definition

Homework 7.4 (Reading due Wednesday 11/19; Problems due Monday 11/24).
Read *§7.3*
Solve *§7.2: 10, 11, 14, 15, 17*, 18**

If $\mathbf{c}(t)$ is some curved path in space, and $\mathbf{T}(t)$ is the unit tangent vector to $\mathbf{c}(t)$ at each time $t \in [a, b]$, and if $\mathbf{F}(x, y, z)$ is a vector field in space, then $\mathbf{F} \cdot \mathbf{T}(t)$ is the component of $\mathbf{F}$ which is parallel to the path at $\mathbf{c}(t)$ (and it is a number, not a vector). Thus, it seems reasonable to consider the integral

$$\int_{\mathbf{c}} \mathbf{F} \cdot \mathbf{T} \, ds.$$

Indeed, our definition of the ***line integral*** will be given by the one above: If we note that $d\mathbf{s} = \mathbf{T} \, ds$, and we notice that (if $\|\mathbf{c}'(t)\| \neq 0$)

$$\mathbf{T}(t) = \frac{\mathbf{c}'(t)}{\|\mathbf{c}'(t)\|}$$

we get the definition

$$\int_{\mathbf{c}} \mathbf{F} \cdot d\mathbf{s} = \int_a^b \mathbf{F}(\mathbf{c}(t)) \cdot \mathbf{c}'(t) \, dt$$

where we used the relation $ds = \|\mathbf{c}'(t)\| \, dt$. This definition holds even if $\|\mathbf{c}'(t)\| = 0$, as it may be justified using Riemann sum methods (see p. 430).

Alternatively, writing $d\mathbf{s} = (dx, dy, dz)$ and $\mathbf{F} = (F_1, F_2, F_3)$, we obtain

$$\begin{aligned} \int_{\mathbf{c}} \mathbf{F} \cdot d\mathbf{s} &= \int_{\mathbf{c}} F_1 dx + F_2 dy + F_3 dz \\ &= \int_a^b \left(F_1 \frac{dx}{dt} + F_2 \frac{dy}{dt} + F_3 \frac{dz}{dt} \right) \, dt. \end{aligned}$$

7.2.4 Orientation

When integrating a function of a single variable, we know that

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

that is, by changing the direction we progress along the axis we change the sign of the integral. This is due to our choice of orientation.

There are exactly two different orientations. In our every day life as well, not only on the x-axis, there are two different orientations (and no more other orientations). There's the one we live in our entire lives – where there's North, South, East, West, and then there's the orientation **seen through a mirror**. Obviously that is not a world we live in, but it shows us there are two different ways to look at objects.

The moral is that while a path in space is independent of our viewpoint, and any scalar valued function defined on it is independent of our viewpoint as well, the result of integration of a vector field *does* depend on our viewpoint, since we must fix a coordinate system around the path in order to perform the integration. And, by the previous discussion, there are two different possibilities to impose a coordinate system, each being the mirror image of the other.

Alternatively, this translates into the fact that for each curve C in space, there are two different types of paths that trace it – each going in a different direction.

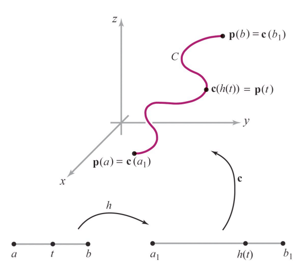
There are exactly two *essentially* different parameterizations. With all that said, we still did not address the question of the speed in which we advance along the path. Fortunately, it turns out that the speed in which we progress does not matter, as long as we don't backtrack along the curve. The only factor that matters is the direction of progression. This will be made clear in Theorem 7.6 below.

7.2.5 Dependence on Parameterization

Definition 7.5 (Reparametrization)**.** *Let* $\mathbf{c} : [a_1, b_1] \to \mathbb{R}^3$ *be a path, and let* h *be a differentiable function that maps another interval* $[a, b]$ *to the interval* $[a_1, b_1]$ *in a one-to-one manner. Then the composition*

$$\mathbf{p}(s) = \mathbf{c}(h(s))$$

is called a **reparametrization** *of* $\mathbf{c}$*, and is a path in its own right.*



We note that h maps endpoints to endpoints, but might reverse the order. If h keeps the same order, we say it is ***orientation-preserving***. If it reverses the order, we say it is ***orientation-reversing***.

Theorem 7.6 (Change of Parametrization for Line Integrals)**.** *Let* $\mathbf{c} : [a_1, b_1] \to \mathbb{R}^3$ *be a differentiable path, and let* $\mathbf{F}$ *be a vector field defined on its image. Let* $\mathbf{p} : [a, b] \to \mathbb{R}^3$ *be a reparametrization of* $\mathbf{c}$*.*

Then if $\mathbf{p}$ *is orientation-preserving, then*

$$\int_{\mathbf{p}} \mathbf{F} \cdot d\mathbf{s} = \int_{\mathbf{c}} \mathbf{F} \cdot d\mathbf{s},$$

and if $\mathbf{p}$ *is orientation-reversing, then*

$$\int_{\mathbf{p}} \mathbf{F} \cdot d\mathbf{s} = -\int_{\mathbf{c}} \mathbf{F} \cdot d\mathbf{s}.$$

Proof. For $\mathbf{p}$ to be a reparametrization of $\mathbf{c}$, means that there exists some one-to-one function $h : [a, b] \to [a_1, b_1]$ such that

$$\mathbf{p} = \mathbf{c} \circ h.$$

Applying the chain rule to this expression, we get

$$\mathbf{p}'(t) = \mathbf{c}'(h(t))h'(t),$$

so that

$$\int_{\mathbf{p}} \mathbf{F} \cdot d\mathbf{s} = \int_a^b \mathbf{F}(\mathbf{p}(t)) \cdot \mathbf{p}'(t)\, dt = \int_a^b [\mathbf{F}(\mathbf{c}(h(t))) \cdot \mathbf{c}'(h(t))]\; h'(t)\, dt.$$

We now change variables with $s = h(t)$, to get

$$\int_{h(a)}^{h(b)} \mathbf{F}(\mathbf{c}(s)) \cdot \mathbf{c}'(s)\, ds$$

$$= \begin{cases} \int_{a_1}^{b_1} \mathbf{F}(\mathbf{c}(s)) \cdot \mathbf{c}'(s)\, ds = \int_{\mathbf{c}} \mathbf{F} \cdot d\mathbf{s} & \mathbf{p} \text{ is orien.-preserving} \\ \int_{b_1}^{a_1} \mathbf{F}(\mathbf{c}(s)) \cdot \mathbf{c}'(s)\, ds = -\int_{\mathbf{c}} \mathbf{F} \cdot d\mathbf{s} & \mathbf{p} \text{ is orien.-reversing} \end{cases}$$

□

Having said that reversing the orientation changes the sign of the *line integral*, we now go back to the path integral and ask the same question. As opposed to line integrals, for which the sign is important (we want to know if the work put into moving an object is positive or negative), for *path integrals* the opposite is important – if we measure the weight of a wire by integrating its density, we want the result to be completely independent of our choice of path and its orientation. This is indeed guaranteed due to the fact that we only have the factor $\|\mathbf{c}(t)\|$ (the speed along the path) and not $\mathbf{c}(t)$ (the velocity along the path). Thus we have the following theorem:

Theorem 7.7 (Change of Parametrization for Path Integrals)**.** *Let* $f : \mathbb{R}^3 \to \mathbb{R}$ *be a continuous function on the image of a differentiable path* $\mathbf{c} : [a, b] \to \mathbb{R}^3$*, and let* $\mathbf{p}$ *be any reparametrization of* $\mathbf{c}$*. Then*

$$\int_{\mathbf{c}} f(x, y, z)\, ds = \int_{\mathbf{p}} f(x, y, z)\, ds.$$

7.2.6 Line Integral of a Gradient

We now wish to make a generalization of the fundamental theorem of calculus:

Theorem 7.8 (Line Integrals of Gradient Vector Fields). *Let $f : \mathbb{R}^3 \to \mathbb{R}$ be a differentiable function with continuous derivatives, and let $\mathbf{c} : [a, b] \to \mathbb{R}^3$ be a differentiable path. Then*

$$\int_{\mathbf{c}} \nabla f \cdot d\mathbf{s} = f(\mathbf{c}(b)) - f(\mathbf{c}(a)).$$

This theorem is ***crucial*** in physics! It means that if a force $\mathbf{F}$ is a gradient of a potential f, then the work done by $\mathbf{F}$ when moving a particle between two points in space is *independent* of the trajectory of the particle!

Proof. Consider the function

$$F : t \mapsto f(\mathbf{c}(t)).$$

The chain rule gives

$$F'(t) = \nabla f(\mathbf{c}(t)) \cdot \mathbf{c}'(t).$$

By the standard fundamental theorem of calculus,

$$\int_a^b F'(t)\, dt = F(b) - F(a) = f(\mathbf{c}(b)) - f(\mathbf{c}(a))$$

thus

$$\begin{aligned} \int_{\mathbf{c}} \nabla f \cdot d\mathbf{s} &= \int_a^b \nabla f(\mathbf{c}(t)) \cdot \mathbf{c}'(t)\, dt \\ &= \int_a^b F'(t)\, dt \\ &= F(b) - F(a) \\ &= f(\mathbf{c}(b)) - f(\mathbf{c}(a)). \end{aligned}$$

□

7.2.7 Line Integrals Over Geometric Curves

We are now poised to discuss integrals over the geometric curve C in $\mathbb{R}^3$, which is the image of the path $\mathbf{c}$ and any of its reparametriations. We may summarize as follows:

Path integrals. Since the curve C traced by the path $\mathbf{c}$ is a fixed object in space, and any function $f(x, y, z)$ defined on it is fixed too, in the sense that there's no dependence on our viewpoint, the integral

$$\int_C f\, ds$$

may be defined unambiguously.

Line integrals. Line integrals, however, depend on the orientation we choose for our curve C, and thus we can write the expression

$$\int_C \mathbf{F} \cdot d\mathbf{s}$$

only if we specify an orientation on C.

7.2.8 Examples

Example 7.9. *Find the work done by the force field $\mathbf{F}(x, y, z) = (x, y, 0)$ when a particle is moved along the straight-line segment from $(0, 0, 1)$ to $(3, 1, 1)$.*

Solution. *The straight line segment starting at $(0, 0, 1)$ and ending at $(3, 1, 1)$ may be parametrized by $\mathbf{c}(t) = (3t, t, 1)$, $0 \le t \le 1$. Thus the work done is the line integral*

$$\textit{Work} = \int_C \mathbf{F} \cdot d\mathbf{s} = \int_0^1 (3t, t, 0) \cdot (3, 1, 0)\, dt = 5.$$

Example 7.10. *Evaluate $\int_C \mathbf{F}(\mathbf{r}) \cdot d\mathbf{r}$ where $\mathbf{F}(x, y, z) = (\sin z, \cos \sqrt{y}, x^3)$ and C is the straight-line segment from $(1, 0, 0)$ to $(0, 0, 3)$.*

Solution. *The straight line segment starting at $(1, 0, 0)$ and ending at $(0, 0, 3)$ may be parametrized by $\mathbf{r}(t) = (1 - t, 0, 3t)$, $0 \le t \le 1$. Thus the work done is the line integral*

$$\begin{aligned} \int_C \mathbf{F} \cdot d\mathbf{r} &= \int_0^1 (\sin 3t, \cos \sqrt{0}, (1-t)^3) \cdot (-1, 0, 3)\, dt \\ &= \int_0^1 \left(-\sin 3t + 3(1-t)^3\right)\, dt \\ &= \frac{1}{3}(\cos 3 - 1) + \frac{3}{4}. \end{aligned}$$

Example 7.11. *Evaluate $\int_{\mathbf{c}} \mathbf{F} \cdot d\mathbf{s}$ where $\mathbf{F}(x, y, z) = (yz, xz, xy)$ and $\mathbf{c}(t) = \left(\cos^3 t, \sin^3 t, 0\right)$ where $0 \le t \le 2\pi$.*

Solution. *Notice that $\mathbf{F} = \nabla f$ where $f(x, y, z) = xyz$. In addition, $\mathbf{c}$ is a closed curve: $\mathbf{c}(0) = \mathbf{c}(2\pi) = (1, 0, 0)$. Thus*

$$\int_{\mathbf{c}} \mathbf{F} \cdot d\mathbf{s} = f(\textit{end}) - f(\textit{start}) = 0.$$

7.3 Parametrized Surfaces

Homework 7.12 (Reading due Friday 11/21; Problems due Monday 11/24).
Read *§7.4, §7.5*
Solve *§7.3: 1, 5, 10, 14; §7.4: 1, 2, 4, 18**

Example 7.13. *We have seen that a plane can be represented in two ways:*

1. *A point on it, and two directions.*
2. *A point on it, and the orthogonal direction to it.*

While the second method gives a nice formula of the form

$$ax + by + cz = d$$

we are interested in the first method: If P is a point on the plane, and $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are two distinct directions in the plane, then the plane may be described by the function

$$\mathbf{\Phi}(u, v) = \boldsymbol{\alpha} u + \boldsymbol{\beta} v + \mathbf{OP}.$$

This is called the **parametric equation of the plane**.

Usually it is this type of description of a surface in $\mathbb{R}^3$ that is useful for our purposes, and not a description as a function $z = f(x, y)$. This is so for two main reasons:

1. Most surfaces are not functions of x, y: the sphere and the torus are two obvious examples of such surfaces. In fact, any surface that has more than one point over a given pair (x_0, y_0) in the plane is an example.
2. As we saw in our discussion of path and line integrals, we prefer to parametrize lines to be able to integrate along them. The same holds for surfaces: To integrate functions over surface we need parameters which "control our progression" along the surface relative to some starting point.

Extending the idea of the parametric equation of the plane, we define a ***parametrization of a surface*** as a function

$$\mathbf{\Phi}(u, v) : D \subseteq \mathbb{R}^2 \to \mathbb{R}^3$$

where D is some domain in $\mathbb{R}^2$. The ***surface*** S is the image of $\mathbf{\Phi}$. We can write:

$$\mathbf{\Phi}(u, v) = (x(u, v), y(u, v), z(u, v))$$

Tangent Vectors to Parametrized Surfaces. It is quite clear that the parameters u and v induce two different directions of progression along the surface. Thus, taking partial derivatives with respect to either, will give two distinct tangential directions at each given point on the surface, which we call $\mathbf{T}_u$ and $\mathbf{T}_v$:

$$\mathbf{T}_u = \frac{\partial \mathbf{\Phi}}{\partial u} = \frac{\partial x}{\partial u}\mathbf{i} + \frac{\partial y}{\partial u}\mathbf{j} + \frac{\partial z}{\partial u}\mathbf{k}$$
$$\mathbf{T}_v = \frac{\partial \mathbf{\Phi}}{\partial v} = \frac{\partial x}{\partial v}\mathbf{i} + \frac{\partial y}{\partial v}\mathbf{j} + \frac{\partial z}{\partial v}\mathbf{k}$$

all evaluated at a point $(u_0, v_0) \in D$.

Regular Surfaces. Intuitively, a regular point on a surface is a point that is not a "corner". That is, at that point both partial derivatives exist and are nonzero, so that there exists a linear approximation (i.e. a tangent plane). At such a point, $\mathbf{T}_u \times \mathbf{T}_v$ will be a nonzero vector that is perpendicular to the surface at that point.

Thus, we say that a surface S is ***regular*** or ***smooth*** at $\mathbf{\Phi}(u_0, v_0)$ if $\mathbf{T}_u \times \mathbf{T}_v \neq \mathbf{0}$ at (u_0, v_0). The surface S is called ***regular*** if it is regular at each point.

Tangent Plane to a Parametrized Surface. Clearly, at each regular point on S, both tangent vectors are nonzero, and thus define a normal vector $\mathbf{n} = \mathbf{T}_u \times \mathbf{T}_v$. We use this normal vector to define the ***tangent plane*** to the surface S at the (regular) point $\mathbf{\Phi}(u_0, v_0) = (x_0, y_0, z_0) \in S$ by the equation

$$(x - x_0, y - y_0, z - z_0) \cdot \mathbf{n} = 0.$$

Example 7.14. *The most obvious surface S to investigate is one that is also a graph of a function $z = g(x, y)$. We may write S in parametric form*

$$\begin{aligned} x &= u \\ y &= v \\ z &= g(u, v). \end{aligned}$$

So that we get

$$\begin{aligned} \mathbf{T}_u &= \mathbf{i} + \frac{\partial g}{\partial u}\mathbf{k} \\ \mathbf{T}_v &= \mathbf{j} + \frac{\partial g}{\partial v}\mathbf{k} \end{aligned}$$

all evaluated at some point $(u_0, v_0) \in \mathbb{R}^2$. Thus we get

$$\mathbf{n} = \mathbf{T}_u \times \mathbf{T}_v = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 0 & g_u \\ 0 & 1 & g_v \end{vmatrix} = -\frac{\partial g}{\partial u}\mathbf{i} - \frac{\partial g}{\partial v}\mathbf{j} + \mathbf{k} \neq \mathbf{0}$$

which implies that the surface is always regular. From this expression we can get the equation of the tangent plane

$$z - z_0 = \left(\frac{\partial g}{\partial x}\right)(x - x_0) - \left(\frac{\partial g}{\partial y}\right)(y - y_0)$$

with the partial derivatives evaluated at (x_0, y_0).

7.4 Area of a Surface

In this and subsequent chapters, we are only interested in surfaces that are regular (except for possibly a finite number of points), and are graphs of parametrizations which are differentiable with continuous derivatives, and one-to-one.

Definition. Recalling that if $\mathbf{a}$ and $\mathbf{b}$ are vectors, then $\|\mathbf{a} \times \mathbf{b}\|$ gives the area of the parallelogram they form, it seems reasonable that the definition of the ***surface area*** $A(S)$ of a parametrized surface S would be given by

$$A(S) = \iint_D \|\mathbf{T}_u \times \mathbf{T}_v\| \, du \, dv.$$

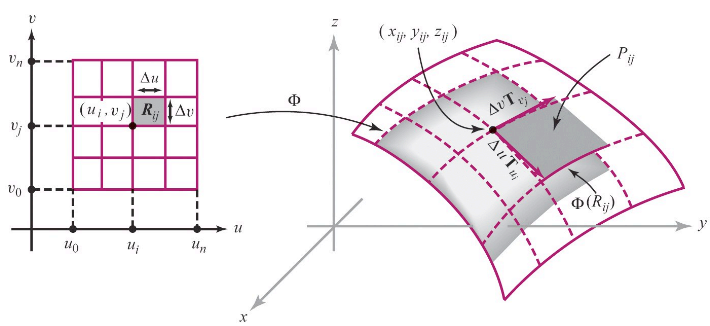
Noting that

$$\begin{aligned} \mathbf{T}_u \times \mathbf{T}_v &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_u & y_u & z_u \\ x_v & y_v & z_v \end{vmatrix} \\ &= (y_u z_v - z_u y_v)\mathbf{i} + (z_u x_v - x_u z_v)\mathbf{j} + (x_u y_v - y_u x_v)\mathbf{k}, \end{aligned}$$

we obtain

$$A(S) = \iint_D \sqrt{(y_u z_v - z_u y_v)^2 + (z_u x_v - x_u z_v)^2 + (x_u y_v - y_u x_v)^2} \, du \, dv.$$

Justification. Let us attempt to give some sort of rough justification to this formula. Assuming that D is a rectangle, and splitting it into n^2 small rectangles of size $\triangle u \times \triangle v$



each, we can see (up to a first approximation), that the ij^{th} rectangle is mapped to a patch that can be approximated by a parallelogram with sides $\triangle u \mathbf{T}_{u_i} \times \triangle v \mathbf{T}_{v_j}$. Thus the area of this small parallelogram is $\|\mathbf{T}_{u_i} \times \mathbf{T}_{v_j}\| \triangle u \triangle v$. Summing over all these parallelograms, and letting $n \to \infty$, we get our formula for $A(S)$.

Surface Area of a Graph. In the special case that S is actually the graph of a function $z = g(x, y)$, we again use the parametrization $x = u, y = v, z = g(u, v)$ for u, v in the domain D, to get

$$\mathbf{T}_u \times \mathbf{T}_v = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 0 & g_u \\ 0 & 1 & g_v \end{vmatrix} = -\frac{\partial g}{\partial u}\mathbf{i} - \frac{\partial g}{\partial v}\mathbf{j} + \mathbf{k} = -\frac{\partial g}{\partial x}\mathbf{i} - \frac{\partial g}{\partial y}\mathbf{j} + \mathbf{k}$$

so that

$$A(S) = \iint_D \left(\sqrt{\left(\frac{\partial g}{\partial x}\right)^2 + \left(\frac{\partial g}{\partial y}\right)^2 + 1} \right) dA$$

Surfaces of Revolution. Please read in the book (p. 466) how we justify the formulas for the area of a surface of revolution:

$$A = 2\pi \int_a^b |f(x)| \sqrt{1 + [f'(x)]^2}\, dx$$

when the graph of the function $y = f(x)$ is revolved about the x-axis, and

$$A = 2\pi \int_a^b |x| \sqrt{1 + [f'(x)]^2}\, dx$$

when the graph of the function $y = f(x)$ is revolved about the y-axis.

7.5 Integrals of Scalar Functions Over Surfaces

Homework 7.15 (Reading due Monday 11/24; Problems due Monday 11/24).
Read *§7.6*
Solve *§7.5: 2, 3, 6, 10, 11, 18*, 20**

We can now naturally define the integral of a scalar-valued function over a surface S: This extension of the definition of the surface area, is similar to the extension of the notion of *arc length* to the *path integral*.

If $f : \mathbb{R}^3 \to \mathbb{R}$, and S is a surface parametrized by $\mathbf{\Phi}(u, v) : D \subseteq \mathbb{R}^2 \to \mathbb{R}^3$, then we define the ***integral of*** f ***over*** S to be

$$\iint_S f \, dS = \iint_D f(\mathbf{\Phi}(u, v))\|\mathbf{T}_u \times \mathbf{T}_v\| \, du \, dv.$$

This is the obvious extension of our definition of the surface area: Recall, that for the surface area we saw that $\|\mathbf{T}_u \times \mathbf{T}_v\|\triangle u\triangle v$ approximates the area of the image of the R_{ij}^{th} rectangle in the uv plane (if both tangent vectors are evaluated at (u_i, v_j)). It is therefore reasonable to believe that

$$f(\mathbf{\Phi}(u_i, v_j))\|\mathbf{T}_u \times \mathbf{T}_v\|\triangle u\triangle v$$

is one summand in the corresponding Riemann sum. For a slightly more detailed explanation, see p. 475 of the book.

Surface Integrals Over Graphs. When S is in fact a differentiable function $z = g(x, y)$, we saw that

$$\|\mathbf{T}_u \times \mathbf{T}_v\| = \sqrt{\left(\frac{\partial g}{\partial x}\right)^2 + \left(\frac{\partial g}{\partial y}\right)^2 + 1}$$

so that we get

$$\iint_S f \, dS = \iint_D f(x, y, g(x, y))\sqrt{\left(\frac{\partial g}{\partial x}\right)^2 + \left(\frac{\partial g}{\partial y}\right)^2 + 1} \, dx \, dy.$$

Finally, if we recall that in the case of a graph, $\mathbf{T}_u \times \mathbf{T}_v = -\frac{\partial g}{\partial x}\mathbf{i} - \frac{\partial g}{\partial y}\mathbf{j} + \mathbf{k}$ we get that the unit normal vector to the surface is given by

$$\mathbf{n} = \frac{-\frac{\partial g}{\partial x}\mathbf{i} - \frac{\partial g}{\partial y}\mathbf{j} + \mathbf{k}}{\sqrt{\left(\frac{\partial g}{\partial x}\right)^2 + \left(\frac{\partial g}{\partial y}\right)^2 + 1}}$$

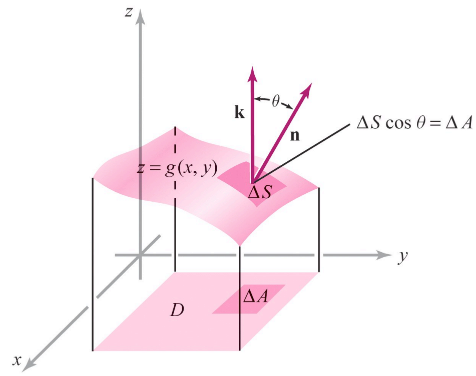


Figure 48: The area element on a graph is $dS = \frac{dxdy}{\cos\theta} = \frac{dxdy}{\mathbf{n}\cdot\mathbf{k}}$

so that

$$\frac{1}{\mathbf{n}\cdot\mathbf{k}} = \sqrt{\left(\frac{\partial g}{\partial x}\right)^2 + \left(\frac{\partial g}{\partial y}\right)^2 + 1} = \|\mathbf{T}_u \times \mathbf{T}_v\|$$

and we can thus write

$$\iint_S f\ dS = \iint_D f(x, y, g(x, y)) \frac{dx\ dy}{\mathbf{n}\cdot\mathbf{k}}.$$

$\frac{dx\ dy}{\mathbf{n}\cdot\mathbf{k}}$ thus gives us the area element on the graph, dS. We also observe that $\mathbf{n}\cdot\mathbf{k} = \cos\theta$ where θ is the angle of the normal to the graph from the z axis (see Figure 48), so that we can also write

$$\iint_S f\ dS = \iint_D \frac{f(x, y, g(x, y))}{\cos\theta} dx\ dy.$$

7.6 Surface Integrals of Vector Fields

Homework 7.16 (Problems and Reading due Monday 12/1)**.**
Read *§8.1*
Solve *§7.6: 3, 5, 9, 16, 18*

As we started with *arc length*, then moved on to *path integrals* (integrals of scalar-valued functions over curves) and finally discussed *line integrals* (integrals of vector-fields over curves), we do the same with surfaces: We saw how to calculate the *surface area*, in the previous section we saw how to integrate scalar-valued functions over surfaces, and now we want to define the concept of **integrating a vector-field over a surface**.

7.6.1 Physical Intuition: Flux

Consider a straight river with constant depth of h meters and constant water flow of $\mathbf{F}$ meters/second. Suppose a fisherman places a net of length L meters across the river at an angle θ. **How much water passes through the net each second?** Well,

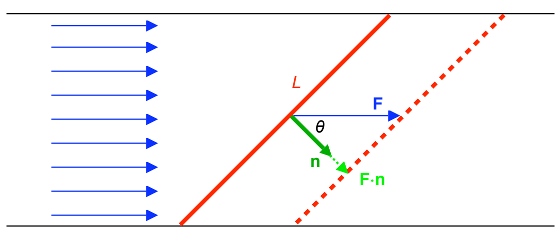


Figure 49: The amount of water passing in one second is $(Lh)(\mathbf{F}\cdot\mathbf{n})$

considering Figure 49 the answer is pretty clear: It is very basic geometry that we need.

The amount of water passing through the net each second, or the ***flux*** of the water through the net, is exactly the volume between the solid red line and the dotted red line, if their horizontal distance is $\|\mathbf{F}\|$. The volume of this parallelogram, if $\mathbf{n}$ is a unit normal vector to the surface *pointing in the direction of the flow*, is given by

$$\text{width across} \times \text{depth} \times \text{thikness} = L \times h \times (\mathbf{F}\cdot\mathbf{n})$$

(none of the $\times$ symbols here are a cross product)

If we denote the area of the net $L \times h = S$ and define the vector $\mathbf{S}$ to be $\mathbf{S} = S\mathbf{n}$, i.e. a normal vector to the surface of the net of magnitude that is equal to the area of the net, then we get that

$$Flux = \mathbf{F}\cdot\mathbf{S}$$

7.6.2 Orientation

The answer to the previous question – *"how much water passes through the net each second?"* – is obviously dependent on our choice of direction, or ***orientation***: Had we chosen the unit normal vector $\mathbf{n}$ to point the other way, the result of the dot product would have been negative.

That is to say, with one point of view the flux is positive, and with another point of view the flux is negative. Our intuition is clear about this, as well as about the fact that the *absolute value* of the flux is the same, it is only the sign that changes.

In short, if we decide that *right* is positive and *left* is negative, then two different people standing on the two banks of the river see the same *magnitude* of flux, but with opposite signs.

Extending the notion of orientation to a surface $S \subseteq \mathbb{R}^3$. Our intuition, again, is very clear about this issue: Most surfaces we encounter are ***oriented***, i.e. they have two sides – "above" and "below", "right" and "left", "inside" and "outside", etc. Our convention is that the **outside is positive** and **inside is negative**. This implies that when we need to specify a unit normal on the surface, we choose it so that it points *outwards*.

Given a parametrization $\mathbf{\Phi}$ of a surface S, we say that it is ***orientation-preserving*** if the normal it produces at each regular point points *outwards*, and we say it is ***orientation-reversing*** if the normal points *inwards*.

Remark 7.17 (Non-orientable surface). *As we mentioned before, there are examples of surfaces that do* not *have two proper sides. The typical example is that of the* Möbius strip. *See p. 486 of the book.*

Orientations of common surfaces. We discuss two common surfaces and their chosen orientations.

1. **The (unit) sphere.** For the sphere, as discussed above, we choose the orientation so that the unit normal vector faces *outward*. What does this imply? Let us consider the usual parametrization of the sphere

$$\mathbf{\Phi}(\theta, \phi) = (\sin\phi\cos\theta, \sin\phi\sin\theta, \cos\phi)$$

   where $\theta \in [0, 2\pi]$ and $\phi \in [0, \pi]$. Our strategy is to calculate the two tangent vectors that this parametrization defines at each point. We then take their cross product, which gives a normal vector. We make sure that it indeed faces *out*. If not, we have to change the order of θ and ϕ. We normalize this vector to get a *unit normal vector*.

(a) The tangent vectors are:

$$\begin{aligned}\mathbf{T}_\theta &= (-\sin\phi\sin\theta, \sin\phi\cos\theta, 0)\\ \mathbf{T}_\phi &= (\cos\phi\cos\theta, \cos\phi\sin\theta, -\sin\phi)\,.\end{aligned}$$

(b) Taking the cross product, we get

$$\mathbf{T}_\theta \times \mathbf{T}_\phi = (x\mathbf{i} + y\mathbf{j} + z\mathbf{k})(-\sin\phi) = -\sin\phi\mathbf{r}$$

(verify that you are able to get this) which points *in*, since for the ϕ values that we consider, $\sin\phi \geq 0$, and $\mathbf{r}$ points out.

(c) Thus $\mathbf{\Phi}$ is orientation-reversing, and we have to change the order of θ and ϕ to make sure we have the desired orientation.

(d) The unit normal (with the right order of θ and ϕ) is given by

$$\mathbf{n} = \frac{\mathbf{T}_\phi \times \mathbf{T}_\theta}{\|\mathbf{T}_\phi \times \mathbf{T}_\theta\|} = \frac{\mathbf{r}}{\|\mathbf{r}\|}.$$

(Why is this answer obvious?)

2. **A graph.** The desired orientation for a surface that is a graph $z = g(x, y)$ is for the normal vector to point *up*. Check p. 488 to see what the expression for $\mathbf{n}$ is in this case. (We saw the expression for $\mathbf{n}$ before in §7.5 – see the notes).

7.6.3 Definition

We know that a small area element on a surface S which is the image of a parametrization $\mathbf{\Phi}$ is given by

$$\triangle S = \|\mathbf{T}_u \times \mathbf{T}_v\| \triangle u \triangle v.$$

As in the physical motivation, we can associate a *direction* to this area element: this can be done simply by dropping the norm from $\mathbf{T}_u \times \mathbf{T}_v$. This produces a vector

$$\triangle\mathbf{S} = \mathbf{T}_u \times \mathbf{T}_v \triangle u \triangle v$$

that has magnitude $\triangle S$, and is perpendicular to the surface. Thus, the following definition is reasonable:

$$\iint_{\mathbf{\Phi}} \mathbf{F} \cdot d\mathbf{S} = \iint_D \mathbf{F} \cdot (\mathbf{T}_u \times \mathbf{T}_v)\, du\, dv$$

Also, one may write

$$\iint_{\mathbf{\Phi}} \mathbf{F} \cdot d\mathbf{S} = \iint_{\mathbf{\Phi}} \mathbf{F} \cdot \mathbf{n}\, dS$$

where $\mathbf{n}$ is a unit normal pointing in the direction of the appropriate orientation.

8 The Integral Theorems

Homework 8.1 (Reading due Wednesday 12/3; Problems due Monday 12/8)**.**
Read *§8.2*
Solve *§8.1: 1, 3d, 7, 13, 16, 20*

We are finally ready to study the three important theorems that relate the differential calculus and the integral calculus we have studied during this course. In all proceeding sections, we wish to prove results that are analogous to the well known ***fundamental theorem of calculus***, stating

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

and, so, giving a relation between the behavior of some quantity *inside* the region $[a, b]$ with the behavior of a related quantity on its *boundary* – the points a and b.

8.1 Green's Theorem

8.1.1 Statement of the theorem

Setup.

1. Recall that we say that a region $D \subseteq \mathbb{R}^2$ is ***simple*** if it is both x-simple and y-simple. See page 341 of the book.

2. Recall that the ***line intergral*** on the plane may be written as

$$\int_C \mathbf{F} \cdot d\mathbf{s} \quad = \quad \int_C F_1 dx + F_2 dy$$

3. **Let us make a choice of orientation on the plane:** Let the positive orientation be *counterclockwise* and the negative orientation *clockwise*. A simple closed curve C (see page 441) is denoted C^+ if the positive orientation is associated with it. Otherwise, it is denoted C^-.

Theorem 8.2 (Green's Theorem)**.** *Let $D \subseteq \mathbb{R}^2$ be a simple region in the plane, and let $C = \partial D$ be its boundary. Suppose that $P : D \to \mathbb{R}$ and $Q : D \to \mathbb{R}$ are differentiable and have continuous partial derivatives. Then*

$$\int_{C^+} P\,dx + Q\,dy = \iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right)\,dx\,dy$$

8.1.2 Proof

We split the proof into two lemmas that together prove the theorem.

Lemma 8.3. *Let D be a y-simple region with boundary $\partial D = C$. Let $P : D \to \mathbb{R}$ be differentiable with continuous partial derivatives. Then*

$$\int_{C^+} P\,dx = -\iint_D \frac{\partial P}{\partial y}\,dx\,dy$$

Lemma 8.4. *Let D be an x-simple region with boundary $\partial D = C$. Let $Q : D \to \mathbb{R}$ be differentiable with continuous partial derivatives. Then*

$$\int_{C^+} Q\,dy = \iint_D \frac{\partial Q}{\partial x}\,dx\,dy$$

Proof. See the book, page 520. □

8.1.3 Applications

Green's theorem on more general domains. Green's theorem applies to more general domains. We can see this by breaking up a "complicated" domain into smaller, simple regions, and apply the theorem for them. We should always remember, that in case the boundary is more complicated, our convention is that ***the positive orientation along a curve is such that the domain is to the left***.

Calculating area.

Theorem 8.5. *Let D be a domain to which Green's theorem applies. Then*

$$A(D) = \frac{1}{2}\int_{\partial D} x\,dy - y\,dx$$

Expression in terms of curl.

Theorem 8.6. *Let D and ∂D be as above, ∂D positively oriented, and let $\mathbf{F} = P\mathbf{i} + Q\mathbf{j}$. Then*

$$\int_{\partial D} \mathbf{F}\cdot d\mathbf{s} = \iint_D (\text{curl }\mathbf{F})\cdot\mathbf{k}\,dA$$

where one should think of $\mathbf{F}$ as a vector with three components (the third being arbitrary. Can you think why?) to calculate the curl.

Expression in terms of divergence.

Theorem 8.7. *Let D and ∂D be as above, ∂D positively oriented, and let $\mathbf{n}$ denote the outward normal to ∂D. Then*

$$\int_{\partial D} \mathbf{F} \cdot \mathbf{n}\, ds = \iint_D \operatorname{div} \mathbf{F}\, dA$$

We can express $\mathbf{n}$ in terms of a positively oriented parametrization $\mathbf{c}(t) = (x(t), y(t))$ of ∂D:

$$\mathbf{n} = \frac{(y'(t), -x'(t))}{\sqrt{[x'(t)]^2 + [y'(t)]^2}}.$$

8.1.4 Examples

Example 8.8. *Let C be the perimeter of the rectangle D with sides $x = 1, y = 2, x = 3,$ and $y = 3$. Evaluate the integral*

$$\int_C (3x^4 + 5)dx + (y^5 + 3y^2 - 1)dy$$

Solution. *Using Green's theorem, we get*

$$\begin{aligned} \int_C (3x^4 + 5)dx + (y^5 + 3y^2 - 1)dy &= \iint_D \left(\frac{\partial (y^5 + 3y^2 - 1)}{\partial x} - \frac{\partial (3x^4 + 5)}{\partial y} \right) dA \\ &= \int_2^3 \int_1^3 0\, dx\, dy = 0. \end{aligned}$$

Example 8.9. *Let*

$$\mathbf{F}(x, y) = (2y + e^x)\mathbf{i} + (x + \sin y^2)\mathbf{j}$$

and C be the circle $x^2 + y^2 = 1$. Evaluate

$$\int_C \mathbf{F} \cdot d\mathbf{s}.$$

Solution. *Denote the unit disk D. Then $\partial D = C$. By Green's theorem*

$$\begin{aligned} \int_C \mathbf{F} \cdot d\mathbf{s} &= \iint_D \left[\frac{\partial}{\partial x}(x + \sin y^2) - \frac{\partial}{\partial y}(2y + e^x) \right] dA \\ &= \iint_D (1 - 2) dA = -\pi. \end{aligned}$$

8.2 Stokes' Theorem

Homework 8.10 (Reading due Friday 12/5; Problems due Monday 12/8)**.**
Read *§8.3, 8.4*
Solve *§8.2: 1, 3, 5, 11, 26*

We are now ready to state *Stokes' theorem*, which is an extension of Green's theorem onto surfaces in space (and not just domains in the plane). Recall our expression of Green's theorem in terms of curl

$$\int_{\partial D} \mathbf{F} \cdot d\mathbf{s} = \iint_D (\text{curl } \mathbf{F}) \cdot \mathbf{k} \, dA$$

and compare to this theorem:

Theorem 8.11 (Stokes' Theorem)**.** *Let S be an oriented surface defined by a one-to-one parametrization $\mathbf{\Phi} : D \subseteq \mathbb{R}^2 \to S$, where D is a region to which Green's theorem applies. Let ∂S denote the oriented boundary of S and let $\mathbf{F}$ be a differentiable (with continuous partial derivatives) vector field on S. Then*

$$\int_{\partial S} \mathbf{F} \cdot d\mathbf{s} = \iint_S (\text{curl } \mathbf{F}) \cdot d\mathbf{S}.$$

Proof. Read page 533 of the book for a proof in the case that S is a graph of a function $z = f(x, y)$. The main idea of the proof in that case is reducing the problem to the plane (by eliminating z which depends on x, y) and then applying Green's theorem.

In the case of a parametrized surface the idea of the proof is similar – reduce the problem from the three variables x, y, z to the two variables of the parametrization u, v and then apply Green's theorem. $\square$

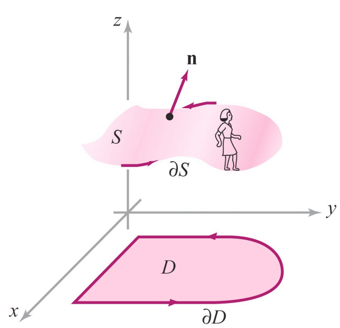


Figure 50: The correct orientation is so that the surface is to the left

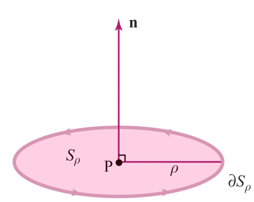
8.2.1 The correct orientation

1. In the case that S is a graph $z = f(x, y)$, our chosen orientation for the surface is so that the normal points in the positive $\mathbf{k}$ direction. The orientation on the boundary ∂S is induced from the orientation on the corresponding boundary ∂D in the plane, as discussed in the case of Green's theorem.

2. In the case that S is a parametrized surface, after choosing the orientation of the normal vector to it, the progression along ∂S should be so that the surface is to the left.

8.2.2 Results

Understanding what the *curl* really is. Consider the point P in space, and let the surface S be a very small disk centered at P, as in the figure. Denote it by S_ρ, where ρ is its radius. Let V be a fluid velocity field on S_ρ, let $\mathbf{n}$ be the unit normal as in the figure (this normal induces the orientation as discussed above). By Stokes' theorem, we



have

$$\int_{\partial S_\rho} \mathbf{V} \cdot d\mathbf{s} = \iint_{S_\rho} [(\text{curl } \mathbf{V}) \cdot \mathbf{n}]\, dS = \iint_{S_\rho} (\text{curl } \mathbf{V}) \cdot\ d\mathbf{S}$$

Now, since S_ρ is a very small disk, curl $\mathbf{V}$ is almost constant on it, so that we may approximate

$$\iint_{S_\rho} [(\text{curl } \mathbf{V}) \cdot \mathbf{n}]\, dS \approx (\text{curl } \mathbf{V})(P) \cdot \mathbf{n}\ A(S_\rho) = (\text{curl } \mathbf{V})(P) \cdot \mathbf{n}\ \pi\rho^2$$

(this approximation is made rigorous using the mean-value theorem for integrals). As ρ tends to 0, this approximation actually becomes an equality, and we get

$$(\text{curl } \mathbf{V})(P) \cdot \mathbf{n} = \lim_{\rho \to 0} \frac{1}{A(S_\rho)} \int_{\partial S_\rho} \mathbf{V} \cdot d\mathbf{s}.$$

This is actually true for any small domain containing P, i.e. S_ρ need not be a disk necessarily, as long as $\mathbf{n}$ remains normal to it.

The right hand side of this equation is a measurement of the flow around P in the plane perpendicular to $\mathbf{n}$. Please refer to page 540 of the book for a detailed explanation of this issue.

Different surfaces with the same boundary. One interesting result is that for two surfaces S_1 and S_2 that have the same boundary and orientation, we have

$$\iint_{S_1} (\text{curl } \mathbf{V}) \cdot \, d\mathbf{S} = \iint_{S_2} (\text{curl } \mathbf{V}) \cdot \, d\mathbf{S}.$$

Surfaces with no boundary. As a result, for a surface S with no boundary (like the sphere), we have

$$\iint_{S} (\text{curl } \mathbf{V}) \cdot \, d\mathbf{S} = 0.$$

8.2.3 Examples

Example 8.12. *Let*

$$\mathbf{F} = (yze^x + xyze^x)\mathbf{i} + xze^x\mathbf{j} + xye^x\mathbf{k}.$$

Show that the integral of $\mathbf{F}$ *around any oriented simple curve* C *that is the boundary of a surface* S *is zero.*

Solution. *By Stokes' theorem,*

$$\int_C \mathbf{F} \cdot d\mathbf{s} = \iint_S (\text{curl } \mathbf{F}) \cdot d\mathbf{S}.$$

Calculating $\nabla \times \mathbf{F}$*, we find that it is always* 0*, and thus the right hand side of Stokes' theorem is 0, so that* $\int_C \mathbf{F} \cdot d\mathbf{s} = 0$*.*

Alternatively, if we are lucky, we might notice that

$$\mathbf{F} = \nabla(xyze^x)$$

so that the integral around any closed curve is 0 *(recall, that we saw that the line integral of a gradient depends only on the endpoints).*

Homework 8.13 (Problems due Monday 12/8).
Solve *§8.3: 4, 12, 23, 24, 25* (read comment after Theorem 8 in the book); §8.4: 1, 3, 14*, 19*, 23*

8.3 Conservative Fields

We saw that if a vector field $\mathbf{F}$ is given by $\mathbf{F} = \nabla f$, then a line integral of $\mathbf{F}$ depends only on the endpoints:

$$\int_a^b \mathbf{F}(\mathbf{c}(t)) \cdot \mathbf{c}'(t)\, dt = f(\mathbf{c}(b)) - f(\mathbf{c}(a)).$$

As we mentioned, intuitively, this corresponds to the fact that a vector field given as a gradient of some function cannot have closed flow lines (water cannot flow down a hill and return to the same point).

Let us now make a rigorous statement that confirms just that.

Theorem 8.14 (Conservative Fields). *Let* $\mathbf{F}$ *be a differentiable (with continuous partial derivatives) vector field on* $\mathbb{R}^3$ *(it may fail to be differentiable at finitely many points, and the theorem will still hold). The following are equivalent (TFAE):*

1. *For any oriented simple closed curve* C, $\int_C \mathbf{F} \cdot d\mathbf{s} = 0$.
2. *For any two oriented simple curves* C_1 *and* C_2 *that have the same endpoints,*
$$\int_{C_1} \mathbf{F} \cdot d\mathbf{s} = \int_{C_2} \mathbf{F} \cdot d\mathbf{s}.$$
3. $\mathbf{F}$ *is the gradient of some function* f.
4. $\nabla \times \mathbf{F} = \mathbf{0}$.

A vector field that satisfies one (and thus all) of these conditions is called a **conservative vector field**.

Proof. To prove a theorem that states "the following are equivalent...", one must show that assuming any statement to be true, implies all other statements. It is enough to show that (1) implies (2), (2) implies (3), (3) implies (4) and finally, that (4) implies (1). □

8.3.1 Applications

As we said above, this theorem relates many ideas we mentioned intuitively in the past. As we saw, Earth's gravitational field is the gradient of the gravitational potential. This implies that the work done by the gravitational field on two objects moving from a common initial point to a common end point is the same, regardless of path.

8.3.2 Examples

Example 8.15 (§8.3 #15a). *Show that* $\mathbf{F} = (xy^2 + 3x^2y)\mathbf{i} + (x + y)x^2\mathbf{j}$ *is conservative. Calculate* $\int_C \mathbf{F} \cdot d\mathbf{s}$ *for* C *consisting of the line segments from* $(1, 1)$ *to* $(0, 2)$ *to* $(3, 0)$*.*

Solution. *If* $\mathbf{F}$ *were a vector field on* $\mathbb{R}^3$*, we could verify it is conservative by checking that* $\nabla \times \mathbf{F} = \mathbf{0}$ *(see the statement of the theorem). In our case,* $\mathbf{F}$ *is a two-dimensional vector field, but let us think of it as being three-dimensional. Taking the curl in the* z *direction will involve only the first two components of* $\mathbf{F}$ *anyway. Thus we want to check that* $(\nabla \times \mathbf{F}) \cdot \mathbf{k}$ *is* 0*:*

$$\frac{\partial}{\partial x}[(x + y)x^2] - \frac{\partial}{\partial y}[xy^2 + 3x^2y] = 3x^2 + 2xy - 2xy - 3x^2 = 0$$

as required.

This implies that $\mathbf{F}$ *is conservative, and thus there exists* $f : \mathbb{R}^2 \to \mathbb{R}$ *such that* $\mathbf{F} = \nabla f$ *by the theorem. Let us try to find this* f*, by solving the two equations that follow from* $\mathbf{F} = \nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right)$*:*

$$\begin{aligned} xy^2 + 3x^2y = \quad F_1 \quad &= \frac{\partial f}{\partial x} \\ (x + y)x^2 = \quad F_2 \quad &= \frac{\partial f}{\partial y}. \end{aligned}$$

Let us integrate the first equation with respect to x*:*

$$\frac{x^2y^2}{2} + x^3y + g(y) = f(x, y)$$

where $g(y)$ *is some unknown function of* y *(this is similar to the constant we get when taking the indefinite integral of a function of one variable). Integrating the second equation with respect to* y *we get*

$$\frac{x^2y^2}{2} + x^3y + h(x) = f(x, y).$$

Thus we get that $h(x) = g(y)$ *so that the both must equal some constant* c*.*

Now, since $\mathbf{F}$ *is conservative, that integral* $\int_C \mathbf{F} \cdot d\mathbf{s}$ *depends only on the endpoints, and we get:*

$$\int_C \mathbf{F} \cdot d\mathbf{s} = f(3, 0) - f(1, 1) = 0 + c - \frac{3}{2} - c = -\frac{3}{2}.$$

8.4 Gauss' Theorem

Recall the divergence form of Green's theorem

$$\int_{\partial D} \mathbf{F} \cdot \mathbf{n} \, ds = \iint_D (\nabla \cdot \mathbf{F}) \, dA.$$

The following theorem generalizes this result from two to three dimensions:

Theorem 8.16 (Gauss' Divergence Theorem)**.** *Let W be a 'nice' region in $\mathbb{R}^3$. Denote by ∂W the oriented closed surface that bounds W. Let $\mathbf{F}$ be a smooth vector field defined on W, and let $\mathbf{n}$ be a unit normal vector to the boundary of W pointing outwards. Then*

$$\iint_{\partial W} \mathbf{F} \cdot \mathbf{n} \, dS = \iiint_W (\nabla \cdot \mathbf{F}) \, dV.$$

By a 'nice' region we basically mean any region we will deal with in this course. More accurately, it means any region that can be broken up into finitely many elementary regions.

Recalling that $d\mathbf{S} = \mathbf{n} \, dS$, we may write the Gauss' divergence theorem as

$$\iint_{\partial W} \mathbf{F} \cdot d\mathbf{S} = \iiint_W (\nabla \cdot \mathbf{F}) \, dV.$$

8.4.1 Results

Understanding what the *divergence* really is. We proceed with an argument similar to the one we used to show what the curl was, to show that the divergence of a vector field $\mathbf{F}$ at some point $P \in \mathbb{R}^3$ is the rate of net outward flux at P per unit volume:

Let W_ρ be a small ball of radius ρ around P, so that

$$\iint_{\partial W_\rho} \mathbf{F} \cdot \mathbf{n} \, dS = \iiint_{W_\rho} \nabla \cdot \mathbf{F} \, dV \approx (\nabla \cdot \mathbf{F})(P) \cdot \text{vol } W_\rho.$$

Thus, exactly as we did in the case of the curl, we divide by vol W_ρ and let $\rho \to 0$, to get

$$(\nabla \cdot \mathbf{F})(P) = \lim_{\rho \to 0} \frac{1}{\text{vol } W_\rho} \iint_{\partial W_\rho} \mathbf{F} \cdot \mathbf{n} \, dS.$$

The integral on the right hand side is exactly the net flux coming out of the ball W_ρ, and by dividing by the ball's volume and taking the limit, we get that the divergence of $\mathbf{F}$ at P is the rate of net outward flux at P per unit volume.

If $(\nabla \cdot \mathbf{F})(P) > 0$, we say that P is a ***source*** for $\mathbf{F}$.

If $(\nabla \cdot \mathbf{F})(P) < 0$, we say that P is a ***sink*** for $\mathbf{F}$.

8.4.2 Examples

Example 8.17 (Gauss' Law)**.** *Let M be a domain in $\mathbb{R}^3$ to which Gauss' divergence theorem applies, and assume the boundary of M does not pass through the origin $(0,0,0)$. Consider the integral*

$$I = \iint_{\partial M} \frac{\mathbf{r}\cdot\mathbf{n}}{r^3}\, dS.$$

Then, if M contains the origin, $I = 4\pi$. Otherwise, $I = 0$.

(recall that $\mathbf{r}(x,y,z) = (x,y,z)$, and $r = \|\mathbf{r}\|$)

Proof.

1. *First, let us show that this is the case when $M = B_R$ is the ball of radius R centered at the origin. In this case, $\mathbf{n} = \mathbf{r}/r$ and $r = R$, so that we get*

$$I = \iint_{\partial B_R} \frac{\mathbf{r}\cdot(\mathbf{r}/r)}{r^3}\, dS = \iint_{\partial B_R} \frac{R^2}{R^4}\, dS = \frac{1}{R^2}\iint_{\partial B_R} dS = \frac{1}{R^2}\cdot 4\pi R^2 = 4\pi.$$

2. *Now, let us assume that M does not contain the origin. It is a somewhat lengthy calculation that yields that*

$$\nabla\cdot\left(\frac{\mathbf{r}}{r^3}\right) = 0$$

in M. Note, that we cannot perform this calculation at *the origin, because the vector field has a singularity there.*

Thus, applying Gauss' divergence theorem, we get

$$I = \iint_{\partial M} \frac{\mathbf{r}\cdot\mathbf{n}}{r^3}\, dS = \iiint_M \nabla\cdot\left(\frac{\mathbf{r}}{r^3}\right)\, dV = 0.$$

3. *Thus, we have proven the claim for any domain that does not contain the origin, and for any ball centered at the origin. The only remaining case is that of a general domain W containing the origin. But any such domain may be represented as the union of a small ball N centered at the origin, together with a domain M that is all of W, with N removed. Over N the integral is 4π, while over M is it 0. It follows that the integral is 4π over W as well.*

Example 8.18 (§8.4 #10)**.** *Evaluate the surface integral* $\iint_S \mathbf{F}\cdot\mathbf{n}\,dS$, *where*

$$\mathbf{F}(x,y,z) - \mathbf{i} + \mathbf{j} + z(x^2+y^2)^2\mathbf{k}$$

and S *is the surface of the cylinder* $x^2+y^2 \le 1, 0 \le z \le 1$, *including the sides and both lids.*

Solution. *By Gauss' divergence theorem*

$$\iint_S \mathbf{F}\cdot\mathbf{n}\,dS = \iiint_W (\nabla\cdot\mathbf{F})\,dV$$

where W *is the interior of the cylinder. Calculating the divergence, we get*

$$\nabla\cdot\mathbf{F} = \frac{\partial}{\partial x}(1) + \frac{\partial}{\partial y}(1) + \frac{\partial}{\partial z}z(x^2+y^2)^2 = (x^2+y^2)^2.$$

Thus, integrating in cylindrical coordinates, we get

$$\begin{aligned}\iint_S \mathbf{F}\cdot\mathbf{n}\,dS &= \iiint_W (\nabla\cdot\mathbf{F})\,dV \\ &= \iiint_W (x^2+y^2)^2\,dx\,dy\,dz \\ &= \iiint_W (r^2)^2\,r\,dr\,d\theta\,dz \\ &= \int_0^1\int_0^{2\pi}\int_0^1 r^5\,dr\,d\theta\,dz \\ &= \frac{2\pi}{6} = \frac{\pi}{3}.\end{aligned}$$

8.5 Examples and Applications

8.5.1 Conservative Fields

The Gravitational Force Field. Recall that the gravitational force that a mass m located at $\mathbf{r} = (x, y, z)$ feels is

$$\mathbf{F}(x, y, z) = -\frac{GmM}{r^3}\mathbf{r}$$

generated by the mass M located at the origin, where G is a constant and $r = \|\mathbf{r}\|$.

To calculate $\nabla \times \mathbf{F}$ we use the identity

$$\nabla \times (g\mathbf{G}) = g\,(\nabla \times \mathbf{G}) + (\nabla g) \times \mathbf{G}$$

to get

$$\nabla \times \mathbf{F} = -GmM\left[\frac{1}{r^3}\,(\nabla \times \mathbf{r}) + \nabla\left(\frac{1}{r^3}\right) \times \mathbf{r}\right].$$

The second term is $\mathbf{0}$ since $\nabla\left(\frac{1}{r^3}\right)$ is radial, and the first term is $\mathbf{0}$ by simple inspection. Since this calculation is valid whenever $r \neq 0$, we find that $\mathbf{F}$ is a conservative field. This fits well with our knowledge from §4.3 that the gravitational field is induced by the ***gravitational potential***

$$V = -\frac{GmM}{r}.$$

We have the relation $\mathbf{F} = -\nabla V$.

The Electric Field. The exact same inspection applies for the electric field, given by ***Coulomb's law***

$$\mathbf{E} = \frac{Q}{4\pi r^3}\mathbf{r}$$

where Q is a charge in the origin. The ***electric potential*** is given by

$$\phi = \frac{Q}{4\pi r}$$

and $\mathbf{E} = -\nabla\phi$.

By Gauss' law

$$\iint_{\partial W} \mathbf{E} \cdot d\mathbf{S} = \begin{cases} Q & \textit{origin} \in W \\ 0 & \textit{origin} \notin W \end{cases}$$

8.5.2 Non-Conservative Fields

The Magnetic Field. The magnetic field is a good example of a nonconservative field. It is given by ***Ampère's law***

$$\int_{\partial S} \mathbf{B} \cdot d\mathbf{s} = \iint_S \mathbf{J} \cdot d\mathbf{S} = I$$

where $\mathbf{J}$ is an electric current density and I is the flux of $\mathbf{J}$ through S.

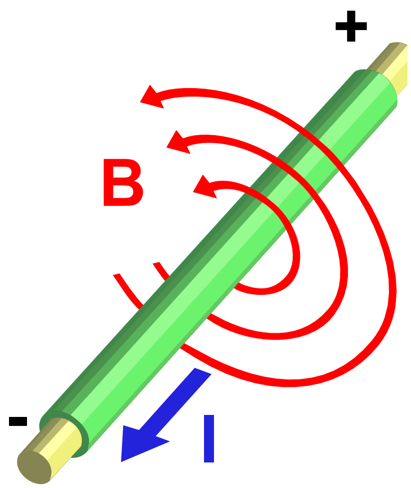


Figure 51: The magnetic field is induced by an electric current

8.5.3 Conservation Laws

We want to use the tools we have developed, in order to express mathematically some basic laws of nature, primarily the conservation of mass and the conservation of charge.

Mass conservation. We let $\mathbf{V}(t; x, y, z)$ be a vector field on $\mathbb{R}^3$ that changes with time, and we let $\rho(t; x, y, z)$ be a real-valued function on $\mathbb{R}^3$ that changes with time as well. Let $\mathbf{J} = \rho\mathbf{V}$.

Think of ρ as fluid density

Think of V as fluid velocity

Think of J as fluid density flow

Let $W \subseteq \mathbb{R}^3$ be any volume in space. Then

$$\iiint_W \rho(t_0; x, y, z)\, dV = \text{total mass of fluid in } W \text{ at time } t_0$$

and

$$\begin{aligned} -\iint_{\partial W} \mathbf{J}(t_0; x, y, z) \cdot d\mathbf{S} &= \text{flux of } \mathbf{J} \text{ into } W \text{ at time } t_0 \\ &= \text{fluid mass increase in } W \text{ at time } t_0 \end{aligned}$$

Thus, we get the ***law of conservation of mass***

$$\frac{d}{dt}\iiint_W \rho\, dV = -\iint_{\partial W} \mathbf{J} \cdot d\mathbf{S}.$$

Now, applying the divergence theorem to the RHS, and changing the order of integration and differentiation on the LHS, one arrives at

$$\iiint_W \frac{\partial \rho}{\partial t}\, dV = -\iiint_W (\nabla \cdot \mathbf{J})\, dV.$$

Moving both expression to the LHS, we get

$$\iiint_W \left(\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J}\right) dV = 0.$$

Since W was arbitrary, we arrive at the ***equation of continuity***

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = 0.$$

Recalling that $\mathbf{J} = \rho\mathbf{V}$, we get

$$\frac{\partial \rho}{\partial t} + \rho(\nabla \cdot \mathbf{V}) + (\nabla\rho) \cdot \mathbf{V} = 0.$$

Momentum "balance". Now we take $D \subseteq \mathbb{R}^3$ to be a fixed (**moving**) portion of the fluid. By Newton's Second Law $m\dot{\mathbf{v}} = m\mathbf{a} = \mathbf{f}$:

$$\frac{d}{dt}\iiint_D \rho\mathbf{V}\, dV = -\iint_{\partial D} p\, d\mathbf{S} \tag{8.1}$$

where p is the pressure inside the fluid.

We now proceed to explore the rule for differentiation of the left hand side of (8.1): For any "nice" function g defined on D, we have:

$$\frac{d}{dt}\iiint_D g\, dV = \iint_{\partial D} g\mathbf{V}\cdot\, d\mathbf{S} + \iiint_D \frac{\partial g}{\partial t}\, dV.$$

Considering the i^{TH} coordinate of the LHS of (8.1), we get:

$$\begin{aligned}\frac{d}{dt}\iiint_D \rho V_i\, dV &= \iint_{\partial D} \rho V_i\, \mathbf{V}\cdot d\mathbf{S} + \iiint_D \frac{\partial}{\partial t}(\rho V_i)\, dV \\ &= \iiint_D \left[\nabla\cdot(\rho V_i\mathbf{V}) + \frac{\partial}{\partial t}(\rho V_i)\right] dV, \qquad i = 1,2,3\end{aligned}$$

On the other hand, the RHS of (8.1) is:

$$-\iint_{\partial D} p\, d\mathbf{S} = -\iint_{\partial D}(pn_1, pn_2, pn_3)\, dS,$$

where $\mathbf{n} = (n_1, n_2, n_3)$ is a unit vector perpendicular to ∂D, and $\mathbf{dS}$=$\mathbf{n}$ dS. Considering the i^{TH} coordinate of this equation we have:

$$\begin{aligned}\left[-\iint_{\partial D} p\, d\mathbf{S}\right]_i &= -\iint_{\partial D} pn_i\, dS \\ &= -\iint_{\partial D}\left[(0,\ldots,p,\ldots,0)\cdot\mathbf{n}\right] dS \\ &= -\iiint_D \frac{\partial p}{\partial x_i}\, dV \qquad i = 1,2,3.\end{aligned}$$

So we have that the RHS of (8.1) is:

$$-\iint_{\partial D} p\, d\mathbf{S} = -\iiint_D \nabla p\, dV \tag{8.2}$$

Plugging these results into (8.1), we get:

$$\iiint_D \left[\nabla \cdot (\rho V_i \mathbf{V}) + \frac{\partial}{\partial t}(\rho V_i) + \frac{\partial p}{\partial x_i} \right] dV = 0$$

But since this is true for any "fixed in fluid" (*"Lagrangian"*) domain, we have that:

$$\frac{\partial}{\partial t}(\rho \mathbf{V}) + \nabla \cdot (\rho \mathbf{V} \otimes \mathbf{V}) = -\nabla p \tag{8.3}$$

Now if we return to our previous consideration of components, we have the equation:

$$\frac{\partial}{\partial t}(\rho V_i) + \nabla \cdot (\rho V_i \mathbf{V}) = -\frac{\partial p}{\partial x_i}, \qquad i = 1, 2, 3.$$

Which gives us:

$$\rho \frac{\partial V_i}{\partial t} + V_i \frac{\partial \rho}{\partial t} + V_i \nabla \cdot (\rho \mathbf{V}) + \rho \mathbf{V} \cdot \nabla V_i = -\frac{\partial p}{\partial x_i}, \qquad i = 1, 2, 3.$$

But by the equation of continuity we have $\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{V}) = 0$, so that the last equation reduces to:

$$\rho \frac{\partial V_i}{\partial t} + \rho(\mathbf{V} \cdot \nabla) V_i = -\frac{\partial p}{\partial x_i}, \qquad i = 1, 2, 3.$$

8.5.4 Euler's Equation of a Perfect Fluid

We can finally conclude:

$$\rho \frac{\partial}{\partial t} \mathbf{V} + \rho(\mathbf{V} \cdot \nabla) \mathbf{V} = -\nabla p, \tag{8.4}$$

which, together with

$$\nabla \cdot \mathbf{V} = 0, \tag{8.5}$$

make up ***Euler's equations of incompressible, inviscid fluid motion in free space***.

8.5.5 Heat Conservation and the Heat Equation

8.5.6 Maxwell's Equations

Maxwell's equations are the following four equations:

Gauss' Law

$$\nabla \cdot \mathbf{E} = \rho$$

No magnetic sources

$$\nabla \cdot \mathbf{B} = 0$$

Faraday's Law

$$\nabla \times \mathbf{E} + \frac{\partial \mathbf{H}}{\partial t} = \mathbf{0}$$

Ampère's Law

$$\nabla \times \mathbf{B} - \frac{\partial E}{\partial t} = \mathbf{J}$$