Multivariable Calculus Lectures

Richard J. Brown

Contents

LECTURE 1

Preliminaries

SYNOPSIS. This first lecture is just a bit of Linear Algebra backstory: As an introduction to the course, I thought to play with the structure of Euclidean space and linear algebra just to establish notation and begin the conversation. I also used a bit of Mathematica for visualization.

HELPFUL DOCUMENTS. Mathematica: `IntersectingPlanes`.

1.1. Real Euclidean Space $\mathbb{R}^n$.

1.1.1. The plane. The real plane is often described as the set of all ordered pairs of real numbers. We can write this as

$$\mathbb{R}^2 = \mathbb{R} \times \mathbb{R} = \left\{ (x, y) \;\middle|\; x, y \in \mathbb{R} \right\}.$$

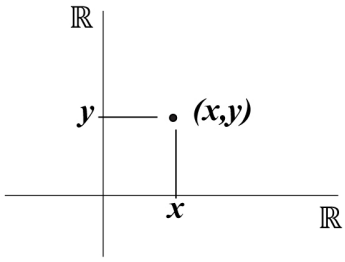


FIGURE 1. The plane $\mathbb{R}^2$.

The way the plane $\mathbb{R}^2$ is built out of two copies of the real line $\mathbb{R}$ is an example of a *Cartesian product*, a way of building a new set (called a product set) out of two sets, whose elements are pairs of elements of the two component sets, called factors, both $\mathbb{R}$ in this case. The set $\mathbb{R}^2$ is useful when studying functional relationships between sets because we can study the pairing given by the function as a subset "living inside" $\mathbb{R}^2$; We assigning the values of the input variable x to the function $f(x)$ to the first slot of the ordered pair, and then we assign the values of the output variable $y = f(x)$ to the other slot of the ordered pair (See Figure 1). This gives us a visual depiction of the functional relationship between x and y as the set of solutions of the equation $y = f(x)$ in the plane. Having this visual (read: geometric) depiction of the function facilitates the study of its properties, which is a central focus of what we call the calculus of functions of a single variable, in this case.

We can construct the operation of addition in the product set $\mathbb{R}^2$ by using the notion of addition in each factor $\mathbb{R}$ of $\mathbb{R}^2$ and forming an addition in $\mathbb{R}^2$ component-wise:

$$(a, b) + (c, d) = (a + c, b + d).$$

With this addition (and the identity element $(0, 0)$ and an inverse $(-a, -b)$ for every set element (a, b)), we can turn $\mathbb{R}^2$ into a group. Here we would

call $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ the *direct product* of the two groups $\mathbb{R}$. (A direct product is a Cartesian product on the underlying sets with whatever added structure the individual sets have and give to the product.) We can also multiply elements of $\mathbb{R}^2$ by real numbers (scalars multiplication), where

$$c \cdot (a, b) = (ca, cb), \quad \text{for} \quad a, b, c \in \mathbb{R},$$

and these two notions behave well together (meaning that they satisfy certain conditions that facilitate additional structure and study on $\mathbb{R}^2$).

1.1.2. Linear algebra. Now $\mathbb{R}$ is also a field, but $\mathbb{R}^2$ is not: One cannot construct a good notion of multiplication in $\mathbb{R}^2$ that satisfies all of the field axioms. However, with the notion of addition of ordered pairs, along with scalar multiplication, we can give $\mathbb{R}^2$ the structure of a *vector space* over $\mathbb{R}$.

DEFINITION 1.1 (Intuitive). A *linear* or *vector space* over a field is a set V of objects together with two operations which can be added together and multiplied by field elements in a "compatible" way.

It is common, in a linear space, to call the individual set elements "vectors". We also say that $\mathbb{R}^2$ is a vector space over $\mathbb{R}$. But it will be a good idea to make a very important distinction:

Using Figure 2 as a guide, we will distinguish between points in $\mathbb{R}^2$, given by all 2-tuples of numbers written as

$$\mathbb{R}^2 = \left\{ p = (x, y) \;\middle|\; x, y \in \mathbb{R} \right\},$$

and vectors in $\mathbb{R}^2$, denoted as the set of all possible 2×1-matrices, or 2-vectors

$$\mathbb{R}^2 = \left\{ \mathbf{p} = \begin{bmatrix} x \\ y \end{bmatrix} \;\middle|\; x, y \in \mathbb{R} \right\}.$$

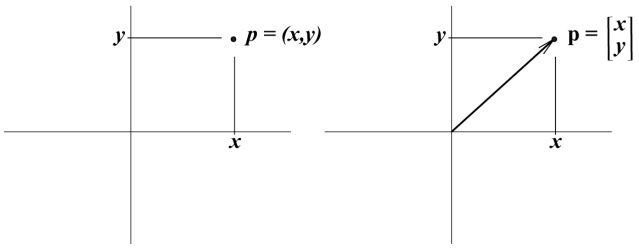


FIGURE 2. Points versus vectors, as elements of $\mathbb{R}^2$.

Some notes:

- Technically speaking, these two descriptions of the plane are quite different, even as there are "equivalent". Note that I am using quotes here because we have not yet defined this (mathematical) term. But intuitively we do see these two descriptions of the plane as the same. For now we will leave it as is.

- In time, we will need to be able to define vectors based at arbitrary points in $\mathbb{R}^2$. Noticing a difference between points and vectors (with the same entries) as descriptions of the elements of the plane will help greatly in this course when, for example, we define and understand vector fields.
- We can add still more structure to the vector space $\mathbb{R}^2$. There is a multiplication of vectors in $\mathbb{R}^2$ where the product is not a vector, but a real number (a scalar): a *scalar product*, sometimes called a *dot product* or an *inner product* on vectors (equivalently points):

$$\begin{bmatrix} a \\ b \end{bmatrix} \bullet \begin{bmatrix} c \\ d \end{bmatrix} = ac + bd \in \mathbb{R}.$$

  With this new structure, the plane becomes an example of an *inner product space*. This is very useful for vector spaces, since with this new structure, we can define notions of a distance between vectors, a vector's size, the angle between vectors, etc. And with these notions of measurement, the plane $\mathbb{R}^2$, as an inner product space, becomes a place where we can do Euclidean geometry. Hence, with this additional structure, we call the plane an example of a *Euclidean Space*.

1.1.3. The vector space $\mathbb{R}^n$. All of this still works if we generalize properly to ordered n-tuples of numbers: Define, for $n \in \mathbb{N}$,

$$\mathbb{R}^n = \underbrace{\mathbb{R} \times \mathbb{R} \times \ldots \times \mathbb{R}}_{n\text{-terms}} = \left\{ x = (x_1, x_2, \ldots, x_n) \;\middle|\; x_i \in \mathbb{R}, \text{ for } i = 1, \ldots, n \right\}$$

$$= \left\{ \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \;\middle|\; x_i \in \mathbb{R}, \text{ for } i = 1, \ldots, n \right\}.$$

Now, a set of k n-vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k \in \mathbb{R}^n$ are called *linearly independent* if for real scalars c_i, $i = 1, \ldots, k$,

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_k\mathbf{v}_k = \mathbf{0} \tag{1.1.1}$$

is only solved by $c_1 = c_2 = \ldots = c_k = 0$. If this is true, then none of the vectors can be written as a linear combination of the others.

EXAMPLE 1.1. $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$, and $\mathbf{v}_3 = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$ are *linearly dependent* since $3\mathbf{v}_1 - \mathbf{v}_2 - \mathbf{v}_3 = \mathbf{0}$. Thus, for instance, one can write $\mathbf{v}_3$ as a linear combination of the others;

$$3\mathbf{v}_1 - \mathbf{v}_2 = \mathbf{v}_3.$$

If one can find n vectors that are linearly independent in $\mathbb{R}^n$, then this set of n vectors can act as a *basis*, in that any vector in $\mathbb{R}^n$ can then be written uniquely as a linear combination of these basis vectors. So if $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in \mathbb{R}^n$ are linearly independent (that is, if they form a basis), then any vector in $\mathbb{R}^n$ can be uniquely determined by the coefficient c_i's:

$$\begin{aligned}\mathbb{R}^n &= \mathbf{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\} \\ &= \left\{ \mathbf{x} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} \middle| \; \mathbf{x} = c_1\mathbf{v}_1 + \ldots + c_n\mathbf{v}_n, \; c_i \in \mathbb{R} \right\}.\end{aligned}$$

Here, the term $\mathbf{span}\{\cdot\}$ is just the set of all linear combinations of the elements given. Note that we typically call the c_i's the *coordinates*, or *components* of the vector $\mathbf{x}$. And when we specify a vector in $\mathbb{R}^n$ without choosing explicit values for the components, we usually declare variables for the components as placeholders for calculation. Using the variables $x_1, \ldots, x_n$ for the components of $\mathbf{x} \in \mathbb{R}^n$, or, say, x, y, z, for $\mathbf{x} \in \mathbb{R}^3$, we can then study the structure of $\mathbb{R}^n$ and functions involving $\mathbf{x}$ via these variables. We will return to this idea very late in this course.

Here is an interesting side note from linear algebra: Using $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in \mathbb{R}^n$ as a basis for $\mathbb{R}^n$, the lines through the origin formed by taking the set of all multiples of each vector $\mathbf{v}_i$ can serve as axes for a coordinate system on $\mathbb{R}^n$. Indeed, if each $\mathbf{v}_i$ serves as a unit of measurement (a measuring stick) on the line that it determines, then the c_i's in any linear combination of basis vectors are the coordinates in that coordinate system, and different from what would be considered the *standard basis* of $\mathbb{R}^n$:

EXAMPLE 1.2. Construct the vectors

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \ldots, \quad \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

These vectors form a basis of $\mathbb{R}^n$, since $\mathbb{R}^n = \mathbf{span}\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$. This is called the *standard basis* for $\mathbb{R}^n$. See Figure 3 below.

Note that this standard basis can be used to define the equivalence between the notion of $\mathbb{R}^n$ defined as points and the notion of $\mathbb{R}^n$ defined as n-vectors.

1.2. Linear spaces inside $\mathbb{R}^n$

Given any set of vectors in $\mathbb{R}^n$, their span may or may not be all of $\mathbb{R}^n$ (if the number of vectors is less than n, they will definitely not). But they will generate a vector space. And in the case where that vector space is not

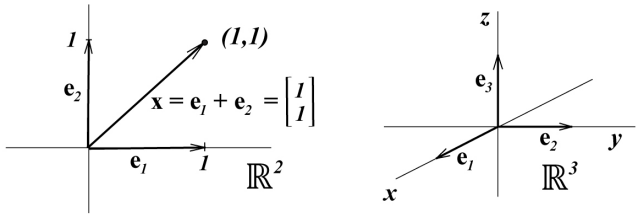


FIGURE 3. The standard bases in $\mathbb{R}^2$ and $\mathbb{R}^3$.

all of $\mathbb{R}^n$, how it sits inside $\mathbb{R}^n$ will be important. We call a vector space generated by a subset of vectors in a vector space a *vector subspace*:

DEFINITION 1.2. A *linear* or *vector subspace* W of a vector space V is a subset of the elements of V that satisfy

(1) $\mathbf{0} \in W \subset V$,
(2) If $\mathbf{w}_1, \mathbf{w}_2 \in W$, then $\mathbf{w}_1 + \mathbf{w}_2 \in W$, and
(3) if $\mathbf{w} \in W$, then for all $c \in \mathbb{R}$, $c\mathbf{w} \in W$.

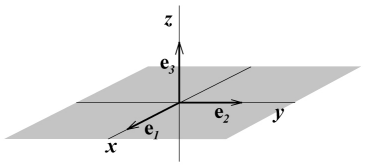


FIGURE 4. The xy-plane in R^3.

It is good to note here that ALL vector subspaces pass through the origin (contain the zero-vector).

And going back to Equation 1.1.1, note that for any $k \in \mathbb{N}$, the set of n-vectors $\mathbf{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is ALWAYS a linear subspace of $\mathbb{R}^n$. How big it is as a subspace depends on the number of $\mathbf{v}_i$s that are linearly independent.

EXAMPLE 1.3. The set $\mathbf{span}\left\{\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\2\\0\end{bmatrix}, \begin{bmatrix}2\\3\\0\end{bmatrix}\right\}$ is commonly referred to as the xy-plane in $\mathbb{R}^3$, thinking of the standard coordinates in $\mathbb{R}^3$. The span of these three vectors only makes a plane in three space since the third vector is simply twice the first plus 3/2 times the second. A basis for the span of these three 3-vectors can readily be the first two vectors in the standard basis of $\mathbb{R}^3$. Note that one can also call this linear subspace the $(z = 0)$-plane. In this way, the xy-plane is a version of $\mathbb{R}^2$ sitting inside $\mathbb{R}^3$ as a subspace of all 3-vectors with 0 in the last component. See Figure 4.

EXAMPLE 1.4. $\mathbf{span}\left\{\begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}2\\4\\6\end{bmatrix}, \begin{bmatrix}3\\6\\9\end{bmatrix}\right\}$ is a line passing through the origin in $\mathbb{R}^3$.

EXAMPLE 1.5. Let $V = \mathbf{span}\left\{\mathbf{a} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 2 \\ -2 \\ 3 \end{bmatrix}\right\}$. Then $V \subset \mathbb{R}^3$ is a 2-dimensional subspace, since $\mathbf{a}$ and $\mathbf{b}$ are linearly independent (recall that the dimension of a (finite-dimensional) vector space is the number of elements in any basis), and $V \subset \mathbb{R}^3$ will look like a plane passing through the origin (See Figure 5, with $\mathbf{a}$ and $\mathbf{b}$ in red). The two 3-vectors

$$\mathbf{c} = \begin{bmatrix} 1 \\ -4 \\ 0 \end{bmatrix}, \quad \mathbf{d} = \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}$$

differ in that $\mathbf{c} \in V$ (shown in blue in Figure 5, while $\mathbf{d} \notin V$ (shown in green in the figure). Indeed, $\mathbf{c} = -\mathbf{a} + \mathbf{b}$, but there are not constants $c_a, c_b \in \mathbb{R}$, where $c_a\mathbf{a} + c_b\mathbf{b} = \mathbf{d}$. We would say that $\mathbf{d}$ is linearly independent from V.

Further, by Example 1.5, we can view the lines passing through $\mathbf{a}$ and $\mathbf{b}$ as coordinate axes for V. And on each axis, we can use the length of the contained vector as the unit length along that axis, marking, for example, 0 at the origin and 1 at the head of $\mathbf{a}$. This provides a coordinate system directly on V, using the ordered pair (c_a, c_b) as the coordinates in V. Thus the vector $\mathbf{c} \in V \subset \mathbb{R}^3$ corresponds to the vector $\begin{bmatrix} 1 \\ -4 \\ 0 \end{bmatrix} \in \mathbb{R}^3$ (or the point $(1, -4, 0) \in \mathbb{R}^3$), but in the coordinates defined directly on V by the basis $\{\mathbf{a}, \mathbf{b}\}$, $\mathbf{c} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \in V$ (or the point $(-1, 1) \in V$) in the *parameterization* of V given by the basis. The idea of placing coordinates directly on a subspace instead of using the ambient coordinates of the larger space is an important one. We will spend much time on this.

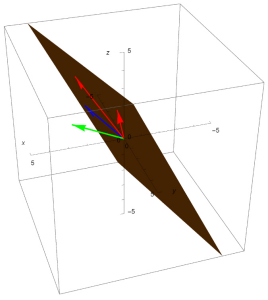


FIGURE 5. $V = \mathbf{span}\{\mathbf{a}, \mathbf{b}\}$.

1.2.1. Planes and Lines in $\mathbb{R}^3$. One way to describe a subspace like $V \in \mathbb{R}^3$ is through another form of multiplication of vectors, this one where the product of two 3-vectors is again a 3-vector. (Note that this is extremely rare and for now is limited to $\mathbb{R}^3$.) The cross product of two vectors $\mathbf{a} \times \mathbf{b} = \mathbf{n}$ is a vector normal (as in zero dot product) to both $\mathbf{a}$ and $\mathbf{b}$. Hence, for any vector $\mathbf{n}$, the set of all vectors normal to $\mathbf{n}$ is a two dimensional subspace $V \in \mathbb{R}^3$. And, if $\mathbf{n}$ is given as the cross product of two linearly independent vectors $\mathbf{a}$ and

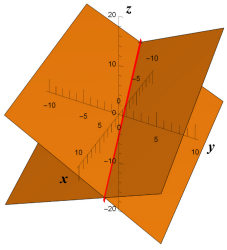


FIGURE 6. A line in $\mathbb{R}^3$ as the intersection of 2 planes.

b, then **a** and **b** serve as a basis for V. Indeed, endow $\mathbb{R}^3$ with the coordinates x, y, and z. Then the equation

$$\mathbf{n} \bullet \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix} \bullet \begin{bmatrix} x \\ y \\ z \end{bmatrix} = 0 = n_1 x + n_2 y + n_3 z,$$

defines a plane passing through the origin in $\mathbb{R}^3$. In Example 1.5, we have

$$\mathbf{n} = \begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \times \begin{bmatrix} 2 \\ -2 \\ 3 \end{bmatrix} = \begin{bmatrix} 2(3) - 3(2) \\ -(1)3 + 3(2) \\ 1(-2) - 2(2) \end{bmatrix} = \begin{bmatrix} 12 \\ 3 \\ -6 \end{bmatrix}.$$

Thus the vector (sub)space V is defined

$$V = \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} \in \mathbb{R}^3 \;\middle|\; 12x + 3y - 6z = 0 \right\}.$$

Check for yourself that, for the vectors Example 1.5, $\mathbf{a}, \mathbf{b}, \mathbf{c} \in V$, but $\mathbf{d} \notin V$.

But this automatically suggests a good generalization to the equation for planes that do not pass through the origin. Recall from linear algebra that one can describe a vector that is based not at the origin via the difference between two vectors based at the origin; a vector $\mathbf{v}_a$ based at $a = (a_1, a_2, a_3) \in \mathbb{R}^3$, with head at $b = (x, y, z)$ can be written as $\mathbf{v}_a = \mathbf{b} - \mathbf{a}$, with $\mathbf{a}, \mathbf{b}$ the vector equivalents of the points a, b, respectively. This follows from the geometric interpretation that one can envision the sum of two vectors by translating the base of one summand to the head of the other, as in Figure 7.

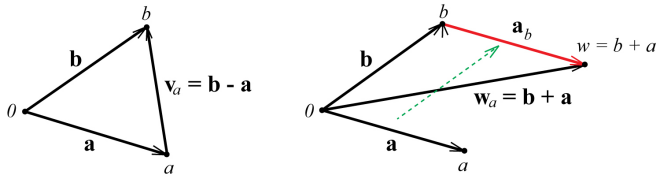


FIGURE 7. The geometric interpretation of vector addition and subtraction.

In a certain sense, however, the vector $\mathbf{v}_a$ is not an element of $\mathbb{R}^3$, since it is not based at the origin. But using the coordinates of $\mathbb{R}^3$, we can still describe vectors like $\mathbf{v}_a$ via vector addition (subtraction) and using actual origin-based vectors (elements of $\mathbb{R}^3$). And this is where we can talk about planes and lines that do not pass through the origin. Indeed, any non-trivial vector based at a with head at b will generate a line in $\mathbb{R}^3$ via its span (as the set of all vectors, based at a, that are multiples of $\mathbf{v}_a$). It can then be written as a vector solution to

$$\mathbf{b} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \quad \text{where} \quad \mathbf{v}_a = \mathbf{b} - \mathbf{a} = \begin{bmatrix} x - a_1 \\ y - a_2 \\ z - a_3 \end{bmatrix}.$$

In this fashion, we can now describe any plane in $\mathbb{R}^3$: Given any vector $\mathbf{n}_a$, based at a, one can describe a plane in $\mathbb{R}^3$ passing through a and normal to $\mathbf{n}_a$ via the equation

$$\mathbf{n}_a \bullet \mathbf{v}_a = \begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix} \bullet \begin{bmatrix} x - a_1 \\ y - a_2 \\ z - a_3 \end{bmatrix} = 0 = n_x(x - a_1) + n_y(y - a_2) + n_z(z - a_3).$$

Note that this plane is not a linear subspace of $\mathbb{R}^3$ as a vector space (the vector $x = y = z = 0$ is not in the plane, for example, when any of the three a_i's are nonzero). But it is a vector space, has a basis, and will play a very important role in understanding how functions involving more than one independent variable behave, as we will see.

One conclusion that can be drawn from this is that one can define a plane in $\mathbb{R}^3$ via a single equation. But then, what is the equation of a line in $\mathbb{R}^3$? Here is an example:

EXAMPLE 1.6. Consider the solution set for the set of equations:

$$\left.\begin{array}{rcll} x + 2y + 3z & = & 4 & (\text{eq1}) \\ 2x - 2y + 3z & = & 1 & (\text{eq2}) \end{array}\right\} \quad 2 \text{ equations in 3 unknowns.}$$

So what does this solution set in $\mathbb{R}^3$ look like? To see, solve as best as one can:

$$\begin{array}{rrcl} (\text{eq1}) + (\text{eq2}): & 3x + 6z & = & 5 \\ 2(\text{eq1}) - (\text{eq2}): & 6y + 3z & = & 7 \end{array}.$$

Then

$$x = \frac{5 - 6z}{3}, \quad y = \frac{7 - 3z}{6}, \quad z \text{ is free.}$$

Better yet, we can place a single parameter t directly on this set by setting $z = t$, so that $x = \frac{5-6t}{3}$ and $y = \frac{7-3t}{6}$, along with $z = t$ makes a parameterized curve (a line) in $\mathbb{R}^3$. One could also write this as a function (using vector notation):

$$\mathbf{c} : \mathbb{R} \to \mathbb{R}^3, \quad \mathbf{c}(t) = \begin{bmatrix} \frac{5-6t}{3} \\ \frac{7-3t}{6} \\ t \end{bmatrix}.$$

Note that, in this parameterization, we still have 3 equations in 4 unknowns. Do you notice a pattern between the number of equations, the number of unknowns and the "size" of the space of solutions?

1.3. Linear functions.

So, roughly speaking, a space V is called linear if any linear combination of two elements in V is still in V. So what, then, is a *linear function*?

DEFINITION 1.3. A function $f : \mathbb{R} \to \mathbb{R}$ is called *linear* if

$$f(c_1x_1 + c_2x_2) = c_1f(x_1) + c_2f(x_2), \quad \forall x_1, x_2 \in \mathbb{R}, \quad c_1, c_2 \in \mathbb{R}$$

Notes:

(1) With appropriate changes, this works equally well for $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$.

(2) Using this definition, then, the function $f(x) = 3x$ is linear, but the function $g(x) = 3x + 1$ is NOT! To see this,

$$\begin{aligned} g(2+3) = g(5) &= 3(5) + 1 = 16 \\ &\neq g(2) + g(3) = (3(2)+1) + (3(3)+1) = 17. \end{aligned}$$

The issue here is that for a function to be linear, the origin of the domain (the input space) must be mapped to the origin of the output space, so that $f(0) = 0$. But here $g(0) = 1$. And thus, $g(x)$ is not linear. It is an example of an *affine* function, one that can be seen as a composition of a linear function and a translation.

(3) Let $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ be linear. Then, given a basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ for the domain $\mathbb{R}^n$, we can write any $\mathbf{x} \in \mathbb{R}^n$ as

$$\mathbf{x} = c_1\mathbf{v}_1 + \ldots + c_n\mathbf{v}_n.$$

Then, since $\mathbf{f}$ is linear, we have

$$\begin{aligned} \mathbf{f}(\mathbf{x}) &= \mathbf{f}(c_1\mathbf{v}_1 + \ldots + c_n\mathbf{v}_n) = c_1\mathbf{f}(\mathbf{v}_1) + \ldots + c_n\mathbf{f}(\mathbf{v}_n) \\ &= m\Bigg\{ \underbrace{\begin{bmatrix} | & & | \\ \mathbf{f}(\mathbf{v}_1) & \ldots & \mathbf{f}(\mathbf{v}_n) \\ | & & | \end{bmatrix}}_{n} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = A_{m\times n}\mathbf{x}. \end{aligned}$$

Hence, any linear map between vector spaces can always be represented by a matrix.

LECTURE 2

Functions of Several Variables.

SYNOPSIS. Today we begin the course in earnest in Chapter 2, although, again like in Lecture 1, we will be covering the material mostly for notation and viewpoint. Pay close attention to why and how we visualize functions, through parameterizations, graphs, slices and sections. These will expose the visual clues to how we analyze functions.

HELPFUL DOCUMENTS.

- Mathematica: `CurvesInSpace`,
- Mathematica: `ParameterizedSurfaces`,
- Mathematica: `VisualizingFunctions`, and
- PDF: `LevelSets`.

2.1. Properties of Functions.

A function $f : X \to Y$ from a set X to another set Y is defined in a manner equal to what you have already studied in single variable calculus (and pre-calculus):

- f assigns to each $x \in X$ a single element $y \in Y$, and every element of X has an element of Y assigned to it.
- The set X is called the *domain* of the function, and Y is called the *codomain*.
- $f(X) \subset Y$ (as a set) is called the *range* of f, and more precisely called the *image* of X in Y under f. It is defined explicitly as
$$f(X) = \left\{ y \in Y \;\middle|\; y = f(x) \text{ for some } x \in X \right\}.$$
- For a subset $Z \subset Y$, the set
$$f^{-1}(Z) = \left\{ x \in X \;\middle|\; f(x) \in Z \right\}$$
is called the *inverse image* of Z in X under f, or the *preimage* of Z in X (under f). Note that if $y \notin f(X)$, then $f^{-1}(y) = \varnothing$ is still well-defined. Note also that the notation does not imply that the function f has an inverse function. The set $f^{-1}(Z) \subset X$ is only a set.
- f is called *one-to-one*, or *injective*, if
$$\# \left\{ x \in X \;\middle|\; f(x) = y \right\} \leq 1, \quad \forall y \in Y.$$
- f is called *onto* or *surjective* if $\forall y \in Y$, $y = f(x)$ for at least one $x \in X$.

- f is called *bijective* if f is both injective and surjective.

Note that, for this class, X and Y will be subsets of Euclidean space, although often not the same space nor the same dimension.

Here is some additional nomenclature and notation:

- Let $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^m$. If $m = 1$, we call $f : X \to Y$ a *real-valued* or *scalar-valued* function on X, or on n-variables (restricted to X). If $m > 1$, we say f is *vector-valued*. As we will see, vector-valued functions consist of expressions that are real-valued on each coordinate of $\mathbb{R}^m$.
- Where important to the discussion, we will denote scalars as $x \in \mathbb{R}$, and vectors as $\mathbf{x} \in \mathbb{R}^n$, $n > 1$. We will also denote a real-valued function as f, and a vector-valued function as $\mathbf{f}$. In lecture, we will employ the vector notation $\vec{x}$ and $\vec{f}$, since boldface is difficult in chalk. Note that when it is not important to the discussion, or for general situations, it is the case that we will use boldface for variables, and possibly write $f : X \to Y$, and $f(x) = y$, even if $X \in \mathbb{R}^n$, $n > 1$, and $y \in Y \subset \mathbb{R}^m$, $m > 1$. This is common in analysis and should be clear in context.
- a function $f : X \to Y$ is often called a *map* (or a *mapping*) from X to Y. In some contexts, a function and a map are not the same thing, but often they are used interchangeably.

DEFINITION 2.1. A map $p : X \to X$ is called a *projection* if $p(p(x)) = p(x)$, $\forall x \in X$.

  - Here, the set comprising the image $p(X) \subset X$ is called the *projection* of X onto $p(X)$. When X is a linear space and p a linear projection, then $p(X)$ is a linear subspace. See Example 2.1 below.
  - A projection p, restricted to its image, is the *identity map*. We can write this as $p\big|_{p(X)} = Id_{p(X)}$.
  - For $X = \mathbb{R}^n$, the map $p_i : \mathbb{R}^n \to \mathbb{R}^n$, defined by

$$p_i\left((x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n)\right) = (0, \ldots, 0, x_i, 0, \ldots, 0)$$

    is called the *ith projection*. Sometimes, one may write $p_i(\mathbf{x}) = x_i$, but this is not quite correct.
  - There are many extensions and generalizations of the idea of projection in various areas of mathematics, including some that do not seem to fit the definition above. (See, for instance, the separate document `StereographicProjection`.) For now, here are a couple of examples.

EXAMPLE 2.1. A common projection onto a linear subspace of $\mathbb{R}^n$ is to zero out one or more coordinates: In $\mathbb{R}^3$, the map

$$(x, y, z) \longmapsto (x, y, 0)$$

is a projection of 3-space onto the xy-plane (See the left side of Figure 8.

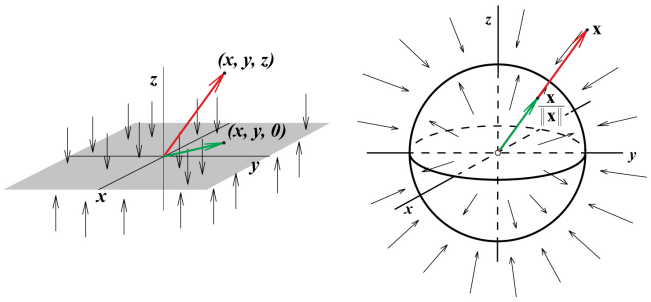


FIGURE 8. Projections in $\mathbb{R}^3$ onto the xy-plane (at left), and the unit sphere S^2 (at right).

EXAMPLE 2.2. The map $r : \mathbb{R}^3 - \{\mathbf{0}\} \to S^2$, $r(\mathbf{x}) = \frac{\mathbf{x}}{\|\mathbf{x}\|}$ is the map which normalizes every non-zero vector in $\mathbb{R}^3$. Here

$$S^2 = \left\{\mathbf{x} \in \mathbb{R}^3 \;\middle|\; \|\mathbf{x}\| = 1\right\}$$

is called the *unit sphere* in $\mathbb{R}^3$, seen on the right of Figure 8. Do you see why $\mathbf{0} \in \mathbb{R}^3$ cannot be in the domain of r?

2.2. Visualization of functions.

Visualizing functions either defined on subsets of $\mathbb{R}^n$ and/or to $\mathbb{R}^n$, when $n > 1$ can be tricky. Some tools that are useful include:

2.2.1. Graphs. In its most basic form, a *relation* is defined as any subset of the Cartesian product of two (or more) sets. And then a *graph* of a relation is just any visual depiction of that relation. When the two sets are subsets of real space $X \in \mathbb{R}^n$ and $Y \subset \mathbb{R}^m$, then the relation is a subset of $\mathbb{R}^n \times \mathbb{R}^m = \mathbb{R}^{n+m}$. Often, relations among real variables are given by equations, and in this case, the graph is the set of solutions to the equations "living" inside the direct product of copies of $\mathbb{R}$, one for each of the variables. And sometimes these relations are functional in one or more of the variables. In this case, solving the equation for one of the variables creates a function whose output is that solved-for variable and whose input(s) are the other

variables. In this case, the graph of that function takes on a particular look; that of a"height over a floor" schematic:

DEFINITION 2.2. For $f : X \subset \mathbb{R}^n \to \mathbb{R}$, the graph of f is the set

$$\mathbf{graph}(f) = \left\{(\mathbf{x}, f(\mathbf{x})) \in \mathbb{R}^n \times \mathbb{R} = \mathbb{R}^{n+1} \;\middle|\; x_{n+1} = f(\mathbf{x})\right\}.$$

Note that this is quite useful for $n = 2$ (so that the graph "lives" in $\mathbb{R}^3$, but not so useful for $n > 2$. Also, this is the proper generalization for the way graphs of functions were constructed in pre-calculus and single variable calculus. And, generally speaking, the "size" of $f(X) \in \mathbb{R}^3$ will be the same as that of X. It should be easy to see that it is always the case that $\mathbf{graph}(f) \subset \mathbb{R}^{n+1}$ always projects to (a copy of) $X \subset \mathbb{R}^n \times \mathbb{R}$ as

$$(x_1, x_2, \ldots, x_n, f(\mathbf{x})) \longmapsto (x_1, \ldots, x_n, 0).$$

See Figure 9. More generally, we have:

DEFINITION 2.3. For $\mathbf{f} : X \subset \mathbb{R}^n \to \mathbb{R}^m$, $m \geq 1$, where $\mathbf{f}(\mathbf{x}) = \mathbf{y}$, the graph of $\mathbf{f}$ is the set

$$\mathbf{graph}(\mathbf{f}) = \left\{(\mathbf{x}, \mathbf{f}(\mathbf{x})) \in \mathbb{R}^n \times \mathbb{R}^m = \mathbb{R}^{n+m} \;\middle|\; \mathbf{y} = \mathbf{f}(\mathbf{x})\right\}.$$

Consider the vector-valued function $\mathbf{g} : X \subset \mathbb{R}^2 \to \mathbb{R}^2$, defined by $\mathbf{g}(\mathbf{x}) = \mathbf{g}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} g_1(x,y) \\ g_2(x,y) \end{bmatrix}$. Here, for $i = 1, 2$, each $g_i : X \to \mathbb{R}$ is a real-valued function, called a *component function* or a *coordinate function*. But the graph of $\mathbf{g} \subset \mathbb{R}^4$ is the set

$$\mathbf{graph}(\mathbf{g}) = \left\{(x, y, z, u) \in \mathbb{R}^4 \;\middle|\; \begin{array}{l} z = g_1(x,y) \\ u = g_2(x,y) \end{array}\right\}.$$

It is already hard to visualize!

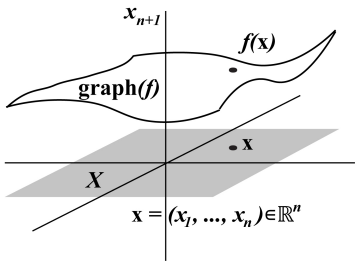


FIGURE 9. For $f : X \subset \mathbb{R}^n \to \mathbb{R}$, $\mathbf{graph}(f) \subset \mathbb{R}^{n+1}$.

An easier example to visualize is the function $\mathbf{h} : \mathbb{R} \to \mathbb{R}^2$, $\mathbf{h}(t) = (\cos t, \sin t)$. Its graph lives in $\mathbb{R}^3$ as a curve

$$\mathbf{graph}(\mathbf{h}) = \left\{(t, x, y) \in \mathbb{R}^3 \;\middle|\; x = \cos t, y = \sin t\right\}.$$

As one can see in Figure 10, this curve can be visualized and studied, but is still a bit tricky to analyze.

2.2.2. Parameterizations. Generalized coordinates can be placed directly on a subset of $\mathbb{R}^n$ through continuous functions so that points on the subset are distinguishable via parameter values instead of ambient coordinates. (One does this on a sphere when one speaks of the latitude and longitude of a point on our Earth.) A *parameterization* allows one to describe a subset of $\mathbb{R}^n$ by a smaller number of variables; one can generally talk of a subset having a dimension equal to the number of variables it takes to distinguish points on the subset, although the notion of dimension for a space is not always very well defined.

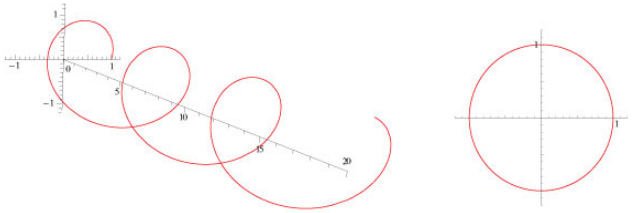


FIGURE 10. Projections in $\mathbb{R}^3$ onto the xy-plane (at left), and the unit sphere S^2 (at right).

Return to the function $\mathbf{h} : \mathbb{R} \to \mathbb{R}^2$, $\mathbf{h}(t) = (\cos t, \sin t) \in \mathbb{R}^2$, and consider only the image of $\mathbf{h} \subset \mathbb{R}^2$. Here, we say that h parameterizes the unit circle in the plane. In this case, t is a coordinate, defined directly (and only) on the circle of radius 1 in $\mathbb{R}^2$, and is a 1-dimensional parameterization. Note here that, broadly speaking, parameterizations should be one-to-one as functions, so that points are distinguished adequately. However, this is not true in general, and this example is telling. Here, we would say that this parameterization is locally-injective. We caution, though, that even this is not true in general.

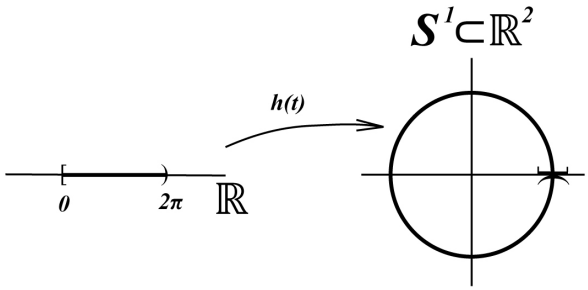


FIGURE 11. Parameterization of $S^1 \subset \mathbb{R}^2$ via $h : \mathbb{R} \to \mathbb{R}^2$, $h(t) = (\cos t, \sin t)$.

EXAMPLE 2.3. Let $\mathcal{D} \subset \mathbb{R}^2$ be the rectangle

$$\mathcal{D} = \left\{ (\theta, \psi) \in \mathbb{R}^2 \,\middle|\, \theta \in [0, 2\pi],\ \psi \in [0, \pi] \right\}$$

as a subset of $\mathbb{R}^2$. Then the function $\Phi : \mathcal{D} \to \mathbb{R}^3$, $\Phi(\theta, \psi) = (\sin\theta \sin\psi, \cos\theta \sin\psi, \cos\psi)$ provides coordinates directly on the unit sphere in three space that correspond to the azimuth angle θ and polar angle ψ of the standard spherical coordinate system in $\mathbb{R}^3$.

Note here two things:

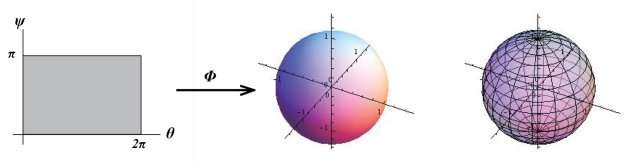


FIGURE 12. Parameterization of $S^2 \subset \mathbb{R}^3$ via $\Phi : \mathcal{D} \to \mathbb{R}^3$.

(1) The function Φ in Example 2.3 is injective, but only on the interior of $\mathcal{D}$, and maps the bottom and top edges of $\mathcal{D}$ to the north and south poles, respectively, and maps both the left and right edges of $\mathcal{D}$ on top of each other and to one of the half-great circles stretching from the north pole to the south. Can you draw this seam on the sphere in Figure 12.

(2) The parameterized objects in these examples are not graphs of functions. They are visual depictions, yes, but they do not satisfy Definition 2.3. They are actually the image of the function defining the parameterization. Hence we would call the unit circle $S^1 = \mathbf{image}(\mathbf{h})$ in Figure 11, and the two-sphere $S^2 = \mathbf{image}(\Phi)$ in Example 2.3. However, it is more common to write $S^1 = \mathbf{h}(\mathbb{R})$, and $S^2 = \Phi(\mathcal{D})$ to denote a functions's full image. We will use this notation going forward.

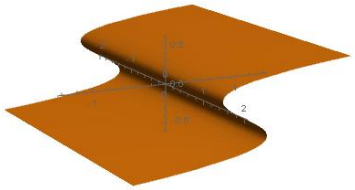


FIGURE 13. A surface which is not the graph of a function defined on the xy-plane in $\mathbb{R}^3$.

Now the domain of the graph of a function $f : X \subset \mathbb{R}^n \to \mathbb{R}$ always parameterizes $\mathbf{graph}(f) \subset \mathbb{R}^{n+1}$. See Figure 9. Can you see why? However, as seen with the sphere in Figure 12, it can also parameterize subsets of $\mathbb{R}^n$ that are not the graphs of functions. For example, in Figure 13, the planar rectangle $\mathcal{R} = [-2, 2] \times [-1.5, 1.5]$, has image $F(\mathcal{R}) \subset \mathbb{R}^3$ via the function $F : \mathcal{D} \to \mathbb{R}^3$, defined by $F(u, v) = \left(u, \frac{3(v^3 - v)}{4}, \frac{2v}{5}\right)$.

2.2.3. Slices and sections of graphs of functions. Understanding the features of graphs of functions of more than one variable can sometimes be facilitated by slicing through the graph, thus fixing the value of one or more variables, either parallel to the domain (a section), or perpendicular to the domain (a slice). First, some definitions:

DEFINITION 2.4. Let $f : X \subset \mathbb{R}^n \to \mathbb{R}$ be a real-valued function on X.

(a) A *c-level set* of f is $\left\{\mathbf{x} \in X \;\middle|\; f(\mathbf{x}) = c\right\}$.

(b) A *(horizontal) section* of f at c is the set

$$\left\{(\mathbf{x}, c) \in \mathbf{graph}(f) \subset \mathbb{R}^{n+1} \;\middle|\; c = f(\mathbf{x})\right\}.$$

Note that this is just the graph of a c-level set, and is sometimes called a *c-contour set*.

It is important to note here that a c-level set of a function $f : X \subset \mathbb{R}^n \to \mathbb{R}$ is a subset of the domain X, while a horizontal section is a subset of the graph of X under f.

EXAMPLE 2.4. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by $f(x, y) = x^2 + y^2$. This function is called a *parabolic bowl* due to the shape of the graph, as in Figure 14, at left.

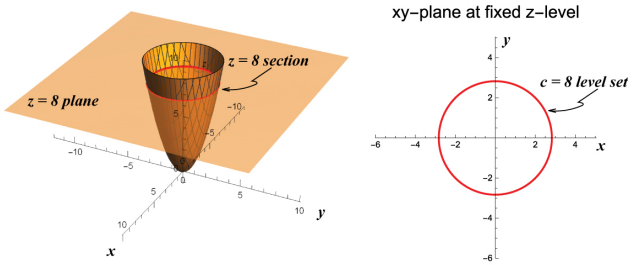


FIGURE 14. The $z = 8$-section and $c = 8$-level set of $z = f(x, y) = x^2 + y^2$.

Some notes:

- All c-level sets in Example!2.4 are circles in the plane, centered at the origin and of radius $\sqrt{c}$. They satisfy the equation $c = x^2 + y^2$.
- One can view c-level sets as the projections of a horizontal section back down into the domain, viewed as part of $\mathbb{R}^{n+1}$ representing the floor $z = 0$.
- One can also write a c-level set as the *inverse image* (as a set) of an output value $c \in \mathbb{R}$. Here $f^{-1}(c) \subset X$. Note that in Example 14, $f^{-1}(c) = \varnothing$, for $c < 0$. But still, $f^{-1}(c)$ is well-defined in these cases and $f^{-1}(-3) \subset X$, even as it is empty.

In contrast, a vertical section (or slice) of $\mathbf{graph}(f)$ is the intersection of $\mathbf{graph}(f)$ with a vertical subspace of $\mathbb{R}^{n+1}$ formed by setting one of the domain variables to a constant. So for

$$\mathbf{graph}(f) = \left\{(x_1, x_2, \ldots, x_n, z) \in \mathbb{R}^{n+1} \;\middle|\; z = f(\mathbf{x})\right\},$$

the x_i-slice at $x_i = c$ is the set

$$\left\{(\mathbf{x}, z) \in \mathbb{R}^{n+1} \;\middle|\; z = f(x_1, \ldots, x_{i-1}, c, x_{i+1}, \ldots, x_n)\right\}.$$

Back to $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x, y) = x^2 + y^2$ in Example 2.4, the $y = 2$-slice is shown in Figure 16, and given by the equation $z = x^2 + 4$ in the xz-plane defined at $y = 2$. The $y = 2$-slice is a curve in $\mathbb{R}^3$ parameterized by x, and is the set

$$\left\{(x, 2, x^2 + 4) \in \mathbb{R}^3\right\}.$$

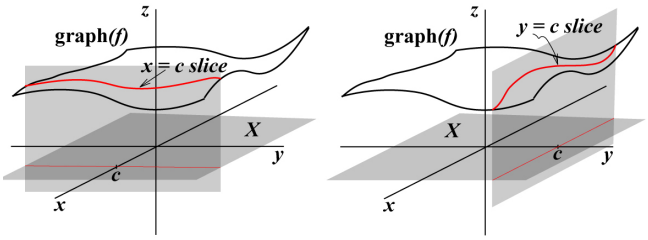


FIGURE 15. The 2 coordinate slices through $\mathbf{graph}(f)$, for $f : X \subset \mathbb{R}^2 \to \mathbb{R}$.

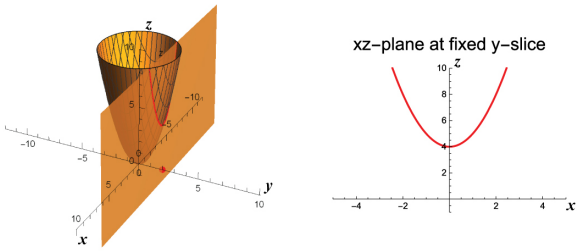


FIGURE 16. The y-slice at $y = 2$ of the function $z = f(x, y) = x^2 + y^2$.

LECTURE 3

Limits.

SYNOPSIS. Today, we define and investigate the notion of a limit in more than one dimension. This is much more subtle than in the Calculus I case, and much harder to fully investigate using the definition alone. Fortunately, all of the "nice" functions from Calculus I are still "nice" in their multivariable generalization. Also, all of the properties of limits developed in single variable calculus are still valid. We will not go deep in this section, but just survey some ideas which we will explore in more detail in the context of more advanced material. The accompanying Mathematica document details some of the more basic pathological functions, where limits do not exist even as intuition indicates they should.

HELPFUL DOCUMENTS.

- Mathematica: `PlottingSurfaces`, and
- PDF: `ProductRule`.

3.1. Definition.

Recall from Calculus I the definition of a limit of a function at a point:

DEFINITION 3.1. Let $I \subset \mathbb{R}$ be open and $f : I \to \mathbb{R}$ a real-valued function on I. Then f has a *limit* L at $x = c \in I$, denoted

$$\lim_{x \to c} f(x) = L,$$

if for every $\epsilon > 0$, there is a $\delta > 0$ such that if $0 < |x - c| < \delta$, then $|f(x) - L| < \epsilon$.

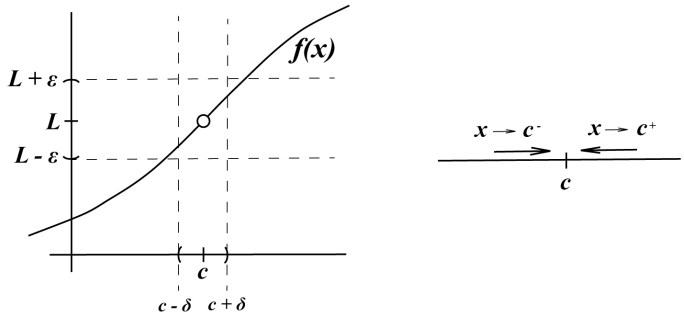


FIGURE 17. On the left, $f(x)$ has the limit L at $x = c$. On the right, x can only approach c in $\mathbb{R}$ form two directions.

Some notes:

- Defining a limit *at* c gives us a notion of what happens to the values of the function *near* the input point c. There is no stipulation or requirement that c be in the domain of the function. But if we are to study what happens to the values of a function near a point, then the domain either includes that point, or abuts to it (so that one can get arbitrarily close to it). In a certain sense, This is the essence of what calculus is really all about!
- If, anytime one can define a small (ϵ-)interval around L, one can find a small (δ-)interval of inputs (around c, but not necessarily at c) all of whose function values stay in the function-value interval (see the left side of Figure 17), then near c, all function values stay near L, and the limit will exist.
- Of course, limits can exist even when function values at c are different or nonexistent. In Figure 18, the limit is the same at $x = c$ for all three graphs.
- In $\mathbb{R}$, the idea of x approaching c involves only 2 possible directions, as shown on the right of Figure 17. These correspond to the one-sided limits

$$\lim_{x\to c^-} f(x), \quad \text{and} \quad \lim_{x\to c^+} f(x).$$

And only when these two "side" limits both exist and are equal, does the actual limit exist.

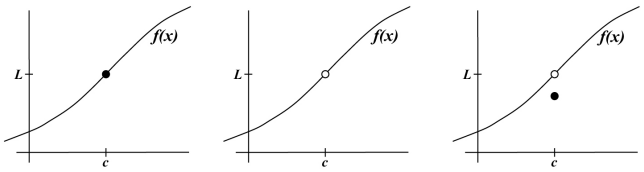


FIGURE 18. In all three cases here, $\lim_{x\to c} f(x) = L$.

In Figure 19, $\lim_{x\to c^-} f(x) \neq \lim_{x\to c^+} f(x)$. Hence $\lim_{x\to c} f(x)$ does not exist. To see this, make a choice for what the limit L could possibly be. Then choose an $\epsilon > 0$ which is small enough to not include both ends of $f(x)$ near $x = c$. Then there will always be points x arbitrarily close to c where $f(x) \notin (L-\epsilon, L+\epsilon)$.

In the case of a vector-valued function of more than one variable, $\mathbf{f} : X \in \mathbb{R}^n \to \mathbb{R}^m$ with $\mathbf{x} \in \mathbb{R}^n$, and $\mathbf{f}(\mathbf{x}) \in \mathbb{R}^m$, we need generalize Definition 3.1 only slightly. However, the ramifications of this generalization are quite complex.

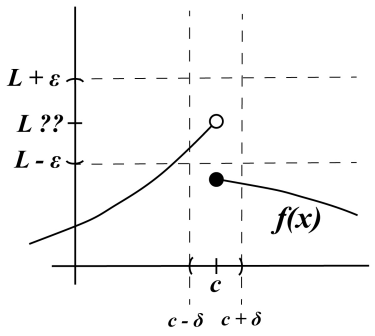


FIGURE 19. What possible choice for L could work as a limit for $f(x)$ at $x = c$?

DEFINITION 3.2. A function $\mathbf{f} : X \subset \mathbb{R}^n \to \mathbb{R}^m$ has a limit $\mathbf{L}$ at $\mathbf{x} = \mathbf{c}$, denoted

$$\lim_{\mathbf{x}\to\mathbf{c}} \mathbf{f}(\mathbf{x}) = \mathbf{L}$$

if, for every $\epsilon > 0$, there is a $\delta > 0$ such that if $0 < \|\mathbf{x} - \mathbf{c}\| < \delta$, then $\|\mathbf{f}(\mathbf{x}) - \mathbf{L}\| < \epsilon$.

Here, $\|\cdot\|$ is the Euclidean norm in real space, defined by

$$\|\mathbf{x} - \mathbf{y}\| = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}\ .$$

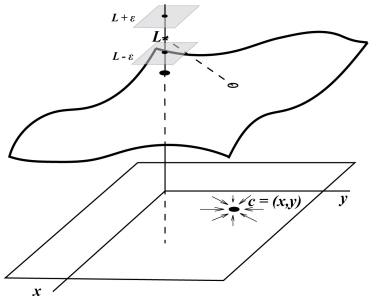


FIGURE 20. Approaching a domain point c in two dimensions.

Notice all of the similarities, and one big difference: The number of ways to approach $\mathbf{c}$ in the domain makes things a lot more complicated! See Figure 20

To understand this, we need to introduce some topology: The Euclidean metric on $\mathbb{R}^n$ allows for a nice definition of an "open" set, much like an open interval in $\mathbb{R}$.

3.2. Topology in $\mathbb{R}^n$.

DEFINITION 3.3. An *open ball* of radius $\epsilon > 0$, centered at $\mathbf{c} \in \mathbb{R}^n$ is

$$B_\epsilon(\mathbf{c}) = \left\{ \mathbf{x} \in \mathbb{R}^n \;\middle|\; \|\mathbf{x} - \mathbf{c}\| < \epsilon \right\}.$$

Some notes:

- In $\mathbb{R}^3$, this is the usual ball you played with as a kid (see Figure 21), but without the skin!
- In $\mathbb{R}^2$, it is the disk of radius ϵ without the circle edge. And in $\mathbb{R}$? How about $\mathbb{R}^1$7?
- One can think of this ball as the set of all vectors of length less than ϵ based at $\mathbf{c}$ (and not at $\mathbf{0}$.

DEFINITION 3.4. A *closed ball* of radius $\epsilon > 0$, centered at $\mathbf{c} \in \mathbb{R}^n$ is

$$\overline{B}_\epsilon(\mathbf{c}) = \left\{ \mathbf{x} \in \mathbb{R}^n \;\middle|\; \|\mathbf{x} - \mathbf{c}\| \le \epsilon \right\}.$$

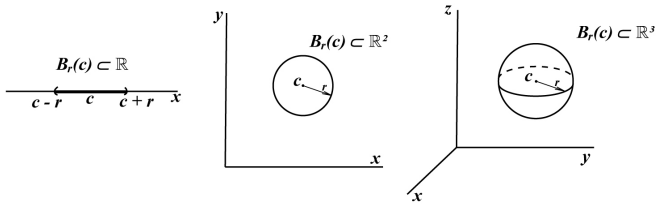


FIGURE 21. An r-ball about $c \in \mathbb{R}$, $c \in \mathbb{R}^2$ and $c \in \mathbb{R}^3$.

Here, the "skin" of the ball (technically, its *boundary*), is the set

$$\left\{\mathbf{x} \in \mathbb{R}^n \;\middle|\; \|\mathbf{x} - \mathbf{c}\| = \epsilon\right\}.$$

What does this skin look like in $\mathbb{R}^5$, for example? In $\mathbb{R}$? Does this thing have a name? It winds up being the boundary of both $B_\epsilon(\mathbf{c})$ and $\overline{B}_\epsilon(\mathbf{c})$.

DEFINITION 3.5. A set $X \in \mathbb{R}^n$ is called *open* if $\forall \mathbf{x} \in X$, $\exists \epsilon > 0$ such that $B_\epsilon(\mathbf{x}) \subset X$.

DEFINITION 3.6. A point $\mathbf{x} \in \mathbb{R}^n$ is a *boundary point* of $X \subset \mathbb{R}^n$ if $\forall \epsilon > 0$, $B_\epsilon(\mathbf{x})$ contains points in X and points not in X. See Figure 22.

DEFINITION 3.7. A set $X \in \mathbb{R}^n$ is called *closed* if it contains all of its boundary points.

EXAMPLE 3.1. Given $\epsilon > 0$, the set

$$D = \left\{\mathbf{x} \in \mathbb{R}^3 \;\middle|\; \|\mathbf{x}\| < \epsilon \text{ and } z \geq 0\right\}$$

is neither open nor closed in $\mathbb{R}^3$. It contains some but not all of its boundary points.

DEFINITION 3.8. Given $X \in \mathbb{R}^n$, A point $\mathbf{x} \in X$ is an *interior point* of X if $\exists \epsilon > 0$ such that $B_\epsilon(\mathbf{x}) \subset X$.

We note here that, given $X \subset \mathbb{R}^n$ and an interior point $x \in X$, we call X a *neighborhood* of x. A neighborhood X is open when, of course X is open in $\mathbb{R}^3$. We will often refer to open neighborhoods of a point x without regard to which one we choose.

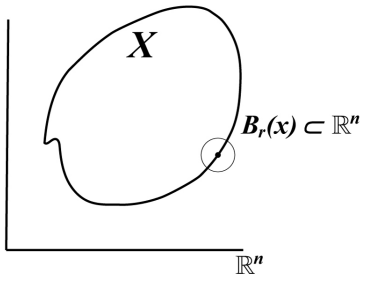


FIGURE 22. A boundary point $x \in \mathbb{R}^n$ of X.

Here is a better way to "see" a limit without a graph? Separate the domain and codomain spaces. Given a function $\mathbf{f} : X \subset \mathbb{R}^n \to \mathbb{R}^m$, $\mathbf{f}$ has a limit $\mathbf{L}$ at $\mathbf{x} = \mathbf{c}$ if, given any ϵ-ball $B_\epsilon(\mathbf{L})$, one can find a δ-ball $B_\delta(\mathbf{c})$ so the image $\mathbf{f}(B_\delta(\mathbf{c}))$ lies entirely inside $B_\epsilon(\mathbf{L})$ (except possibly at $\mathbf{c}$).

In practice,

(1) Limits are hard to calculate using the definition, as pathological functions create a diverse array of issues.
(2) Limits follow all of the typical rules found in Calculus I (See page 106.
(3) Most functions in vector calculus are "nice": They behave well on their full domain:
    - vector-valued functions are scalar-valued on each component.

- Scalar-valued functions involving trig, exponential, logarithmic, rational and polynomial functions are nice even if their arguments involve many variables.

EXAMPLE 3.2. The function $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x, y) = \cos(x + y)$ will still have limits everywhere for the same reasons the cosine function did in single variable calculus.

3.3. Techniques for study.

3.3.1. Directional approach. One technique to study whether a limit exists or not is to reduce the approach of $\mathbf{x}$ to $\mathbf{c}$ to one direction, and use all of the techniques one learns from Calculus I. While this can often be useful for establishing a limit may not exist (coming in from two different direction yields to different values), it is dangerous to use to establish a limit (See accompanying Mathematica files for examples).

EXAMPLE 3.3. Let $f : \mathbb{R}^2 \to R$ be defined by $f(x, y) = \frac{xy}{x^2+y^2}$. This function is not defined at the origin in $\mathbb{R}^2$. But does the limit exist there? Let's explore by looking only at certain directions. To start, suppose we approach the origin $(0, 0)$, along the linen $y = 0$ in the plane. Then

$$\lim_{(x,y)\to(0,0)} f(x, y) = \lim_{(x,0)\to(0,0)} \frac{x \cdot 0}{x^2 + 0^2} = \lim_{x\to 0} 0 = 0.$$

This result will be the same if we approached the origin along the line $x - 0$ (check this!). But nowm, let's approach the origin in $\mathbb{R}^2$ along the line $y = x$. Here

$$\lim_{(x,y)\to(0,0)} f(x, y) = \lim_{(x,x)\to(0,0)} \frac{x \cdot x}{x^2 + x^2} = \lim_{x\to 0} \frac{x^2}{2x^2} = \frac{1}{2}.$$

If approaching from different directions yields different values for a limit, then can a limit possibly exist?

EXAMPLE 3.4. Let $f : \mathbb{R}^2 \to R$ be defined by $g(x, y) = \frac{x^4y^4}{(x^2+y^4)^3}$. This function is again not defined only at the origin in $\mathbb{R}^2$. Does the limit exist at the origin? Let's explore by again looking at certain directions. Suppose we approach the origin $(0, 0)$, along the linen $y = cx$ in the plane. This should help to determine almost every direction of approach depending on

the value of $c \in \mathbb{R}$. (which directions are missed?) Then

$$\lim_{(x,y)\to(0,0)} g(x,y) = \lim_{(x,cx)\to(0,0)} \frac{x^4(cx)^4}{(x^2+(cx)^4)^3} = \lim_{x\to 0} \frac{c^4x^8}{(x^6+3c^4x^8+3c^8x^{10}+c^{12}x^{12})}$$
$$= \lim_{x\to 0} \frac{c^4x^2}{(1+3c^4x^2+3c^8x^4+c^{12}x^6)} = 0.$$

So it would seem here that the limit odes actually exist and is equal to 0 in this case. However, let's approach the origin along the parabola $x = y^2$. Then we have

$$\lim_{(x,y)\to(0,0)} g(x,y) = \lim_{(y^2,y)\to(0,0)} \frac{\left(y^2\right)^4 y^4}{\left((y^2)^2+y^4\right)^3} = \lim_{y\to 0} \frac{y^{12}}{8y^{12}} = \frac{1}{8}.$$

It turns out that approaching from all directions is more complicated than simply coming in linearly from each direction.

3.3.2. Polar coordinates. Switch to polar coordinates and use the fact that $B_\epsilon(\mathbf{x}) = B_\rho(\mathbf{x})$ where ρ is the "distance" variable in the spherical coordinate system on $\mathbb{R}^n$.

EXAMPLE 3.5. Back to $f : \mathbb{R}^2 \to R$, $f(x,y) = \frac{xy}{x^2+y^2}$, we convert the coordinate system in the plane to polar coordinates through the equations $x = \rho\cos\theta$, and $y = \rho\sin\theta$. Then

$$f(x,y) = f(\rho\cos\theta, \rho\sin\theta) = \frac{(\rho\cos\theta)(\rho\sin\theta)}{(\rho\cos\theta)^2+(\rho\sin\theta)^2} = \cos\theta\sin\theta = f(\rho,\theta).$$

But approaching from different linear directions to the origin means approaching along lines of fixed θ. As f will take different values for different fixed values of θ, the limit at the origin does not exist.

3.4. Continuity.

Continuity of functions in vector calculus is pretty much the same as for Calculus I, with a bit of extra structure:

DEFINITION 3.9. A function $\mathbf{f} : X \subset \mathbb{R}^n \to \mathbb{R}^m$ is said to be *continuous at* $\mathbf{a}$ if either $\mathbf{a}$ is an isolated point of X or if

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

And we say that $\mathbf{f}$ is a continuous function on X if it is continuous at $\mathbf{a}$ for every $\mathbf{a} \in X$.

Some notes:

- In the case of continuity, graphs do not have tears, holes, cliffs, or break in it (this is not a mathematical description).
- Like in Calculus I, sums and scalar multiples of continuous functions are continuous.
- Also, products of continuous functions are continuous, and also quotients where they make sense.
- Compositions of continuous functions are also continuous where they make sense. In this case, we have, if $\mathbf{f} : X \subset \mathbb{R}^n \to \mathbb{R}^m$ and $\mathbf{g} : Y \subset \mathbb{R}^m \to \mathbb{R}^p$ are continuous, and $\mathbf{f}(X) \subset Y$, the
$$(\mathbf{g} \circ \mathbf{f}) : X \subset \mathbb{R}^n \to \mathbb{R}^p$$
is continuous.
- the vector-valued function $\mathbf{f} : X \subset \mathbb{R}^n \to \mathbb{R}^m$ is continuous at $\mathbf{a}$ iff each component function $f_i : X \subset \mathbb{R}^n \to \mathbb{R}$, $i = 1, 2, \ldots, m$ is continuous at $\mathbf{a}$.

LECTURE 4

The Derivative.

SYNOPSIS. Today, we finish our discussion on limits and pass through the concept of continuity. Really, there is little to add to the mix since the only new idea is that the limit of a function not only exists but equals the function value at a point of continuity. But there are a few rules and extensions that we talk about here. Then on to differentiability, where things start to diverge from single variable calculus. Here we define what differentiability is for a vector-valued function on more than one variable, both from an analytical as well as geometric perspective, and start the discussion on its properties. The accompanying Mathematica notebook gives some geometric meaning to the derivative of a real-valued function on two variables and how the tangent plane to its graph in three space is defined and constructed.

HELPFUL DOCUMENTS. Mathematica: `PartialDerivatives`.

The Derivative. A *partial derivative* of a real-valued function $f : X \subset \mathbb{R}^n \to \mathbb{R}$ taken at a point is really a single variable calculus concept, where one studies how a function is changing in a particular direction:

DEFINITION 4.1. Let $\mathbf{a} \in X$ be an interior point, and $f : X \subset \mathbb{R}^n \to \mathbb{R}$ a real-valued function on X. Then the *partial derivative* of f, with respect to the coordinate x_i at the point $\mathbf{x} = \mathbf{a}$ is the real number

$$\frac{\partial f}{\partial x_i}(\mathbf{a}) = \lim_{h \to 0} \frac{f(a_1, \ldots, a_{i-1}, a_i + h, a_{i+1}, \ldots, a_n) - f(\mathbf{a})}{h}.$$

The *partial derivative* of f with respect to x_i is the real-valued function

$$\frac{\partial f}{\partial x_i}(\mathbf{x}) = \lim_{h \to 0} \frac{f(x_1, \ldots, x_{i-1}, x_i + h, x_{i+1}, \ldots, x_n) - f(\mathbf{x})}{h}.$$

.

Notes:

- It is simply the ordinary (read: Calculus I) derivative of f with respect to x_i, found by fixing all coordinates x_j, for $j \neq i$, and varying only x_i.
- Alternate notation: $D_{x_i} f(\mathbf{x})$, or $f_{x_i}(\mathbf{x})$.
- Geometrically, given $f(x, y)$ and $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$, the y-slice through $\mathbf{graph}(f)$ at $y = b$ is a 1-dimensional curve inside the xz-plane

at $y = b$. Then $\frac{\partial f}{\partial x}(a,b)$ is the slope of this curve inside the slice, evaluated at (a,b):

$$\frac{\partial f}{\partial x}(a,b) = \lim_{h\to 0} \frac{f(a+h,b) - f(a,b)}{h}.$$

In this case, a varies, but b is held constant. In this case,, we say that the partial derivative of f with respect to x, evaluated at $(x,y) = (a,b)$ is the slope of the line tangent to that portion of the $\mathbf{graph}(f)$ that intersects the xz-plane at $y = b$. item In turn,

$$\frac{\partial f}{\partial y}(a,b) = \lim_{h\to 0} \frac{f(a,b+h) - f(a,b)}{h}$$

is the slope of the line tangent to that portion of the $\mathbf{graph}(f)$ that intersects the xy-plane at $x = a$. These partial derivatives, as regular single variable calculus derivatives in a single direction, satisfy all of the rules that one developed in Calculus I.

Now, for a point (a,b) in the domain where these two quantities exist, the two tangent lines sitting in $\mathbb{R}^3$, cross at the point $(a,b,f(a,b)) \in \mathbb{R}^3$ and are perpendicular (form a right angle). They will determine a plane in $\mathbb{R}^3$: choose a non-zero vector inside each line, based at the corssing point. The plane determined by these two crossing lines is then the set of all linear combinations (in $\mathbb{R}^3$) of these two vectors.

EXAMPLE 4.1. In higher, dimensions, this setup generalizes well: For $f : X \subset \mathbb{R}^n \to \mathbb{R}$, with

$$\mathbf{graph}(f) = \left\{ \left[\begin{array}{c} x_1 \\ \vdots \\ x_n \\ z \end{array} \right] \in \mathbb{R}^{n+1} \;\middle|\; z = f(x_1,\ldots,x_n) \right\},$$

fix a point $\mathbf{a} = \left[\begin{array}{c} a_1 \\ \vdots \\ a_n \end{array} \right] \in X$. Now allow the ith coordinte to vary. Then the slice formed by fixing

$$x_1 = a_1, \ldots, x_{i-1} = a_{i-1}, x_{i+1} = a_{i+1}, \ldots, x_n = a_n$$

(note that this set of equations comprise $n-1$ equations in $\mathbb{R}^{n+1}$, where the graph of f lives), forms a two-dimensional space in $\mathbb{R}^{n+1}$ parameterized by the variable x_1 and z, which we will call the $x_i z$-plane at $\mathbf{a}$. Here, the intersection

$$\mathbf{graph}(f) \cap \{x_i z\text{-plane at } \mathbf{x} = \mathbf{a}\}$$

is a 1-dimensional curve. If $\frac{\partial f}{\partial x_i}(\mathbf{a})$ exists, then its value represents the slope of the line tangent to this curve in the $x_i z$-plane at $\mathbf{a}$. As a line in $\mathbb{R}^{n+1}$, it passes through $(\mathbf{a}, f(\mathbf{a}) \in \mathbb{R}^{n+1}$. Now if tangent lines exist for each of the

variables x_i, for $i = 1, \ldots, n$, they will form an n-dimensional space inside $\mathbb{R}^{n+1}$ passing through the point $(\mathbf{a}, f(\mathbf{a}) \in \mathbb{R}^{n+1}$. This space will be of vital importance to us.

Back to our two dimensional case, the plane formed by the two tangent lines to the slices of $f(x, y)$ at the point (a, b) is called the *tangent plane* to the graph of f at (a, b). So what is the equation defining this 2-dimensional plane in $\mathbb{R}^3$, what is its equation?

- It will consist of one linear equation in the three variables x, y, and z.
- One choice of vector in the tangent line to the curve in the xz-plane corresponding to $y = b$ will be based at (a, b) and have components $\begin{bmatrix} 1 \\ f_x(a,b) \end{bmatrix}$. (Why is this?) So, as a vector in $\mathbb{R}^3$, this vector will have components $\begin{bmatrix} 1 \\ 0 \\ f_x(a,b) \end{bmatrix}$, and be based at the point $(x, y, z) = (a, b, f(a, b))$.
- The other vector in the tangent line to the intersection of the yz-plane at $x = a$ with the graph of f, will have components $\begin{bmatrix} 0 \\ 1 \\ f_y(a,b) \end{bmatrix}$, again based at $(a, b, f(a, b))$.
- The tangent plane is then the set of all linear combinations of vectors, based at $(a, b, f(a, b0)$ that have components

$$c_1 \begin{bmatrix} 1 \\ 0 \\ f_x(a,b) \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 1 \\ f_y(a,b) \end{bmatrix}.$$

  A little cumbersome, but well-defined.

There is a better way to describe the tangent plane: The vector

$$\mathbf{n} = \begin{bmatrix} 1 \\ 0 \\ f_x(a,b) \end{bmatrix} \times \begin{bmatrix} 0 \\ 1 \\ f_y(a,b) \end{bmatrix} = \begin{bmatrix} -f_x(a,b) \\ -f_y(a,b) \\ 1 \end{bmatrix}$$

is normal to both of the tangent vectors. Thus it is also normal to every linear combination of these tangent vectors. In fact, then, one can *define* the tangent space defined by these two tangent vectors as the space of vectors

normal to $\mathbf{n}$, so with the dot product, we have

$$\underbrace{\begin{bmatrix} x-a \\ y-b \\ z-f(a,b) \end{bmatrix}}_{\text{generic vector at } (a,b,f(a,b))} \bullet \underbrace{\begin{bmatrix} -f_x(a,b) \\ -f_y(a,b) \\ 1 \end{bmatrix}}_{\text{normal vector to all}} = 0.$$

This works out to

$$-f_x(a,b)(x-a) - f_y(a,b)(y-b) + z - f(a,b) = 0,$$

or, with a bit of rearranging of terms

$$z = f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b).$$

Now, call the right hand side of this last equation $h(x,y)$, so that

$$z = h(x,y) = f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b)$$

is a linear function $h : \mathbb{R}^2 \to \mathbb{R}$. It's graph $z = h(x,y)$ is then a linear equation representing a plane in $\mathbb{R}^3$ that passes through $(a,b,f(a,b))$, and has precisely the partials $h_x(a,b) = f_x(a,b)$ and $h_y(a,b) = f_y(a,b)$, when it is defined, that is. When it is defined, the graph of this function becomes the *best* linear approximation to $\mathbf{graph}(f)$ at the point $(a,b) \in X$. So what does "best" actually mean in this context? It means:

(1) $z = h(x,y)$ is a linear function in the variables x, y, and z, and
(2) at (a,b), all of the following are true:
   - The functions are equal, so $h(a,b) = f(a,b)$;
   - the derivatives are equal, so $\frac{\partial h}{\partial x}(a,b) = h_x(a,b) = f_x(a,b) = \frac{\partial f}{\partial x}(a,b)$, and $\frac{\partial h}{\partial y}(a,b) = h_y(a,b) = f_y(a,b) = \frac{\partial f}{\partial y}(a,b)$.

EXAMPLE 4.2. Let $f(x,y) = x^2 + y^2$, and choose $(a,b) = (1,2)$. We can go directly to Definition 4.1 here and compute

$$\begin{aligned} \frac{\partial f}{\partial x}(1,2) &= \lim_{h\to 0} \frac{f(1+h,2) - f(1,2)}{h} \\ &= \lim_{h\to 0} \frac{\left((1+h)^2 + 2^2\right) - \left(1^2 + 2^2\right)}{h} \\ &= \lim_{h\to 0} \frac{(1 + 2h + h^2 + 4 - (1+4))}{h} = \lim_{h\to 0} \frac{2h + h^2}{h} = \lim_{h\to 0} 2 + h = 2. \end{aligned}$$

Similarly, $\frac{\partial f}{\partial y}(1,2) = 4$. Then

$$\begin{aligned} z = h(x,y) &= f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b) \\ &= 5 + 2(x-1) + 4(y-2) \end{aligned}$$

is the equation in x, y, and z, whose solutions comprise the tangent plane to the graph of $f(x,y)$ in $\mathbb{R}^3$. Of course, as mentioned in the first note after Definition 4.1, one can think directly that partial derivative are really single

variable derivatives, as far as calculation goes. What this means is that we can sidestep the definition, and simply write

$$\frac{\partial f}{\partial x}(1,2) = \frac{\partial f}{\partial x}(x,y)\bigg|_{(x,y)=(1,2)} = \frac{\partial}{\partial x}\left[x^2+y^2\right]\bigg|_{(x,y)=(1,2)} = (2x+0)\bigg|_{(x,y)=(1,2)} = 2.$$

There is a major caveat that we need to mention here: Just because the individual limits $\frac{\partial f}{\partial x}(a,b)$ and $\frac{\partial f}{\partial y}(a,b)$ may both exist, it does not automatically mean that f is differentiable at (a,b)! Example 4, on page 121 of the text is a great example of why the existence of these limits is not enough. I call this the rooftop function:

EXAMPLE 4.3. Let $g(x,y) = \Big||x|-|y|\Big| - |x| - |y|$. Here,

$$\frac{\partial g}{\partial x}(0,0) = \lim_{h\to 0}\frac{g(0+h,0)-g(0,0)}{h} = \lim_{h\to 0}\frac{\Big||h|-|0|\Big| - |h| - |0| - 0}{h} = 0.$$

Similarly, $\frac{\partial g}{\partial y}(0,0) = 0$. However, step off of the axes, and one can see the sharp edges of the graph. In fact, if one sliced the graph of g along the $x = y$ line (diagonally, with respect to the two axes), then the limits would not exist! Indeed, slice $\mathbf{graph}(g)$ along the line $y = x$. Call the plane forming this slice

$$P_x = \left\{(x,y,z) \in \mathbb{R}^3 \;\middle|\; y = x\right\}.$$

Then the piece of the graph of g inside P_x can be written as

$$z = g(x,x) = g(x) = \Big||x|-|x|\Big| - |x| - |x| = -2|x|.$$

But, as already known from Calculus I, This function has no derivative at $x = 0$, since

$$g'(0) = \lim_{x\to 0}\frac{g(x)-g(0)}{x-0} = \frac{-2|x|}{x}.$$

This limit does not exist, and one can see the corner at the origin of the graph of $g(x)$. Note: This idea of slicing a graph of a function along a line that is different from an axis in the domain will be an important tool in studying the properties of functions of more than one variable. This is the idea of a *directional derivative*, which we will explore soon.

The existence of a proper tangent space to the graph of a function relies on its ability to well-approximate the function from ALL directions. The best way to construct this is, again, to use the limit!

Let $f : \mathbb{R} \to \mathbb{R}$, and notice how we can rewrite

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}, \quad \text{as} \quad \lim_{x \to a} \frac{f(x) - (f(a) + f'(a)(x - a))}{x - a} = 0$$

when (and only when) the limit actually exists.

EXERCISE 1. Show that this is true.

But this means that, for $h(x) = f(a) + f'(a)(x - a)$, we can say that f is differentiable at $x = a$ precisely when the tangent line $y = h(x)$ to $y = f(x)$ at $x = a$ exists, so precisely when

$$\lim_{x \to a} \frac{f(x) - h(x)}{x - a}.$$

This is important, and establishes an alternate way to define differentiability for a function; A function $f(x)$ is differentiable

EXAMPLE 4.4. In 2-dimensions, this setup generalizes well: For $X \subset \mathbb{R}^2$ open, with $f : X \to \mathbb{R}$, f is *differentiable* at $(a, b) \in X$ if

- Both $\frac{\partial f}{\partial x}(a, b)$ and $\frac{\partial f}{\partial y}(a, b)$ exist, and
- if $h(x, y) = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b)$ satisfies

$$\lim_{(x,y) \to (a,b)} \frac{f(x, y) - h(x, y)}{||(x, y) - (a, b)||} = 0.$$

Note that, again, when $h(x, y)$ exists and satisfies the limit above, then $z = h(x, y)$ is the tangent plane to $\mathbf{graph}(f)$ at $(a, b, f(a, b)) \in \mathbf{graph}(f) \subset \mathbb{R}^3$.

More notes:

- An alternate, but equivalent, notion of differentiability: For $X \subset \mathbb{R}^2$ open, and $f : X \to \mathbb{R}$, f is differentiable at (a, b) if $f_x(x, y)$ and $f_y(x, y)$ are continuous in a neighborhood of (a, b) in X.
- Like in Calculus I, differentiability always implies continuity.
- Also true in n-dimensions: Given $X \subset \mathbb{R}^n$ open, and $f : X \to \mathbb{R}$, f is *differentiable* at $\mathbf{a} \in X$ if
  - Each of $\frac{\partial f}{\partial x_i}(\mathbf{a})$ exist for $i = 1, \ldots, n$, and
  - if

$$h(\mathbf{x}) = f(\mathbf{a}) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{a})(x_i - a_i) \quad \text{satisfies} \quad \lim_{\mathbf{x} \to \mathbf{a}} \frac{f(\mathbf{x}) - h(\mathbf{x})}{||\mathbf{x} - \mathbf{a}||} = 0.$$

There is an easier way to write this:

DEFINITION 4.2. For $f : X \subset \mathbb{R}^n \to \mathbb{R}$ differentiable at $\mathbf{a} \in X$, the *derivative* of f at $\mathbf{a}$ is the $1 \times n$ matrix

$$Df(\mathbf{a}) = \begin{bmatrix} f_{x_1}(\mathbf{a}) & f_{x_2}(\mathbf{a}) & \ldots & f_{x_n}(\mathbf{a}) \end{bmatrix}.$$

And the *derivative function* is the $1 \times n$ matrix of functions

$$Df(\mathbf{x}) = \begin{bmatrix} f_{x_1}(\mathbf{x}) & f_{x_2}(\mathbf{x}) & \dots & f_{x_n}(\mathbf{x}) \end{bmatrix}.$$

Knowing this, the tangent linear function, using the above notation and definition, can be written

$$\begin{aligned} h(\mathbf{x}) &= f(\mathbf{a}) + \sum_{i=1}^{n} f_{x_i}(\mathbf{a})(x_i - a_i) \\ &= f(\mathbf{a}) + \begin{bmatrix} f_{x_1}(\mathbf{a}) & \dots & f_{x_n}(\mathbf{a}) \end{bmatrix} \begin{bmatrix} x_1 - a_1 \\ x_2 - a_2 \\ \vdots \\ x_n - a_n \end{bmatrix} \\ &= f(\mathbf{a}) + Df(\mathbf{a})\,(\mathbf{x} - \mathbf{a}) \end{aligned}$$

where $Df(\mathbf{a})$ is a $n \times 1$ (row) matrix, and $(\mathbf{x} - \mathbf{a})$ is an $1 \times n$ matrix (an n-vector). The result is a number, as it should. Hence the limit, in the definition becomes

$$\lim_{\mathbf{x}\to\mathbf{a}} \frac{f(\mathbf{x}) - f(\mathbf{a})}{\|\mathbf{x} - \mathbf{a}\|} = \lim_{\mathbf{x}\to\mathbf{a}} \frac{f(\mathbf{x}) - (f(\mathbf{a}) - Df(\mathbf{a})(\mathbf{x} - \mathbf{a}))}{\|\mathbf{x} - \mathbf{a}\|} = \lim_{\mathbf{x}\to\mathbf{a}} \frac{f(\mathbf{x}) - h(\mathbf{x})}{\|\mathbf{x} - \mathbf{a}\|}.$$

Now what about $\mathbf{f} : X \subset \mathbb{R}^n \to \mathbb{R}^m$? Here, for $\mathbf{x} \in X$, we have $\mathbf{f}(\mathbf{x}) - \begin{bmatrix} f_1(\mathbf{x}) \\ \vdots \\ f_n(\mathbf{x}) \end{bmatrix}$. with n input variables and m output variables. If the derivative is to exist, then each component real-valued function $f_i : X \to \mathbb{R}$ must have a derivative (including all of the partials). We have

$$D\mathbf{f}(\mathbf{a}) = \begin{bmatrix} Df_1(\mathbf{a}) \\ \vdots \\ Df_n(\mathbf{a}) \end{bmatrix},$$

where each element in this matrix is, itself, a $1 \times n$ matrix. Hence

$$D\mathbf{f}(\mathbf{a}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{a}) & \frac{\partial f_1}{\partial x_2}(\mathbf{a}) & \cdots & \frac{\partial f_1}{\partial x_n}(\mathbf{a}) \\ \frac{\partial f_2}{\partial x_1}(\mathbf{a}) & \frac{\partial f_2}{\partial x_2}(\mathbf{a}) & \cdots & \frac{\partial f_2}{\partial x_n}(\mathbf{a}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1}(\mathbf{a}) & \frac{\partial f_n}{\partial x_2}(\mathbf{a}) & \cdots & \frac{\partial f_n}{\partial x_n}(\mathbf{a}) \end{bmatrix},$$

an $m \times n$ matrix with $\frac{\partial f_i}{\partial x_j}(\mathbf{a})$ as the ijth entry.

So our most general definition is:

DEFINITION 4.3. Let $\mathbf{f} : X \subset \mathbb{R}^n \to \mathbb{R}^m$ be a vector-valued function on an open X, and let $\mathbf{a} \in X$. $\mathbf{f}$ is differentiable at $\mathbf{x} = \mathbf{a}$ if

(1) $\frac{\partial f_i}{\partial x_j}(\mathbf{a})$ all exist, for $i = 1, \dots, n$ and $j = 1, \dots, n$, and
(2) the linear map $\mathbf{h}(\mathbf{x}) = \mathbf{f}(\mathbf{a}) + D\mathbf{f}(\mathbf{a})(\mathbf{x} - \mathbf{a})$ satisfies

$$\lim_{\mathbf{x}\to\mathbf{a}} \frac{\|f(\mathbf{x}) - h(\mathbf{x})\|}{\|\mathbf{x} - \mathbf{a}\|}.$$

Some final notes:

- In Definition 4.3, the term $\|f(\mathbf{x}) - h(\mathbf{x})\|$ measures the distance between $\mathbf{f}(\mathbf{x})$ and $\mathbf{h}(\mathbf{x})$ near $\mathbf{a}$, as vector-valued functions.
- $D\mathbf{f}(\mathbf{a})$, as a matrix of numbers, represents a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$. It's entries vary as $\mathbf{a}$ varies, but it represents the *best* linear map approximating $\mathbf{f}(\mathbf{x})$ near $\mathbf{x} = \mathbf{a}$.
- $D\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a}) \in \mathbb{R}^m$ is an m-vector for each value of $\mathbf{x}$ and represents a catalog of ways that moving around near $\mathbf{a}$ affects functions values in the codomain.
- $\mathbf{h}(\mathbf{x}) = \mathbf{f}(\mathbf{a}) + D\mathbf{f}(\mathbf{a})(\mathbf{x} - \mathbf{a})$ defines an affine map $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^m$ (a linear map with a translation). Recall that for $m = 1$, $z = h(\mathbf{x})$ has a graph in $\mathbb{R}^{n+1}$ which is tangent to $\mathbf{graph}(f)$ at the point $(\mathbf{a}, f(\mathbf{a}))$. It is the same for $m > 1$, once one understands the nature of a graph with more than one output, but geometrically, it is far less easy to "see".

LECTURE 5

The Rules of Differentiation

SYNOPSIS. Here, we bring back the rules for differentiation (used to derive new functions constructed using various combinations of other functions) from Calculus I and use them in our new context. The basic frame for this discussion is, "the rules are the same, but only precisely when they actually make sense." What this means is the focus of this class. Also, we will look at higher derivatives and the notion of a function being differentiable more than once. This involves defining the kth partial of a real-valued (and vector-valued) function and what it means for mixed partials to be equal. The differentiable class of a function is discussed, along with just what kind of object the kth derivative of a real-valued function on n variables is and how it encompasses its n^k partials.

5.0.1. The Rules of Differentiation. The nice thing about calculating derivatives in multivariable calculus is that, in many ways, they follow the same rules as they did in single variable calculus, suitably generalized, of course.

5.0.1.1. *The Constant Multiple Rule.* Multiplying a function $\mathbf{f} : X \subset \mathbb{R}^n \to \mathbb{R}^m$ by a real constant $c \in \mathbb{R}$ affects only the functions values, as in Calculus I functions. The new function $(c\mathbf{f}) : X \subset \mathbb{R}^n \to \mathbb{R}^m$ has a vector output, and a constant times a vector means simply multiplying each component by that constant. Indeed,

$$(c\mathbf{f})(\mathbf{x}) = c \cdot \begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_m(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} (cf_1)(\mathbf{x}) \\ (cf_2)(\mathbf{x}) \\ \vdots \\ (cf_m)(\mathbf{x}) \end{bmatrix}.$$

The partial derivatives of each f_i are single variable derivatives, where the Constant Multiple Rule held, so

$$\begin{aligned} \frac{\partial(cf_i)}{\partial x_j}(\mathbf{x}) &= \lim_{h\to 0} \frac{(cf_i)(x_1,\ldots,x_{j-1},x_j+h,x_{j+1},\ldots,x_n) - (cf)(\mathbf{x})}{h} \\ &= \lim_{h\to 0} \frac{c\,(f_i(x_1,\ldots,x_{j-1},x_j+h,x_{j+1},\ldots,x_n) - f(\mathbf{x}))}{h} \\ &= c \cdot \left(\lim_{h\to 0} \frac{f_i(x_1,\ldots,x_{j-1},x_j+h,x_{j+1},\ldots,x_n) - f(\mathbf{x})}{h} \right) = c \cdot \left(\frac{\partial f_i}{\partial x_j}(\mathbf{x}) \right). \end{aligned}$$

The effect is that the entire derivative matrix is multiplied by c, just as in matrix multiplication by a scalar, so that

$$D(c\mathbf{f})(\mathbf{x}) = c \cdot D\mathbf{f}(\mathbf{x}).$$

5.0.1.2. *The Sum/Difference Rule.* The Sum Rule (and hence the Difference Rule when one of the functions is multiplied by -1), is also precisely the same as that for single variable calculus, except that the two functions in the sum must have the same domain and codomain. Only then will the two derivative matrices have the same dimensions, allowing us to actually sum the derivative matrices. So consider $\mathbf{f}, \mathbf{g} : X \subset \mathbb{R}^n \to \mathbb{R}^m$ to m-vector-valued functions on $X \subset \mathbb{R}^n$, and let $\mathbf{h} = \mathbf{f} + \mathbf{g}$. Then

$$D\mathbf{h}(\mathbf{x}) = D\left(\mathbf{f} + \mathbf{g}\right)(\mathbf{x}) = D\mathbf{f}(\mathbf{x}) + D\mathbf{g}(\mathbf{x}).$$

5.0.1.3. *The Product Rule.* The Product Rule, in multivariable calculus, can be a bit trickier, given that the product of two vectors may or may not be a vector of the same size: It is for the cross product in $\mathbb{R}^3$). But the dot product of two vectors in $\mathbb{R}^n$ is a scalar. And sometimes, the output of a product of vectors may not even be a vector of any size (look up *outer product*, for example). The tricky part is to ensure that, if two vector-valued functions are differentiable, then so should be the product of those two functions, however, that is defined. And further, to find a rule to write the derivative of the product using the derivatives of the factors. The nice part of all this is that the Product Rule will always hold, as long as the parts and the product make sense.

Indeed, let $\mathbf{f} : X \subset \mathbb{R}^n \to \mathbb{R}^m$ and $\mathbf{g} : X \subset \mathbb{R}^n \to \mathbb{R}^p$ be two vector-valued functions, possibly of different sizes, but definitely defined on the same domain (why is this necessary?) The define $\mathbf{h}(\mathbf{x}) = \mathbf{f}(\mathbf{x}) \cdot \mathbf{g}(\mathbf{x})$. However that product is defined, it does need to make sense. But when it does, it means that the $m \times 1$-matrix of outputs $\mathbf{f}$ and the $p \times 1$-matrix of outputs of $\mathbf{g}$ are multiplied together in a well-defined way. But then the Product Rule is

$$D\mathbf{h}(\mathbf{x}) = D\left(\mathbf{f} \cdot \mathbf{g}\right)(\mathbf{x}) = D\mathbf{f}(\mathbf{x}) \cdot \mathbf{g}(\mathbf{x}) + \mathbf{f}(\mathbf{x}) \cdot D\mathbf{g}(\mathbf{x}).$$

Following the dimensions, at least, we would get that, if one could multiply an m-vector and a p-vector, then one can also multiple a $m \times n$-matrix to a p-vector, and add to that the product of an m-vector with a $p \times n$-matrix. Write this out to verify, but the genral idea is that an $m \times n$ matrix is just a collection of n m-vectors. Here are some examples:

EXAMPLE 5.1. Let $p = 1$. Then $g(x)$ is a scalar function, and $Dg(\mathbf{x})$ is a $1 \times n$-matrix. Then $\mathbf{h} = \mathbf{f} \cdot g$ makes sense, as the output is the product of an m-vector $\mathbf{f}(\mathbf{x})$ with a scalar $g(\mathbf{x})$ at every input. We get $\mathbf{h} : X \subset \mathbb{R}^n \to \mathbb{R}^m$, $\mathbf{h}(\mathbf{x}) = \mathbf{f}(\mathbf{x})g(\mathbf{x})$, and

$$\underbrace{D\mathbf{h}(\mathbf{x})}_{m\times n} = \underbrace{D\mathbf{f}(\mathbf{x})}_{m\times n}\,\underbrace{g(\mathbf{x})}_{\text{scalar}} + \underbrace{\mathbf{f}(\mathbf{x})}_{m\times 1}\,\underbrace{Dg(\mathbf{x})}_{1\times n}.$$

Here is a concrete example: Suppose that we wanted to create a function that was the product of $\mathbf{f}(x,y,z) = \begin{bmatrix} xy+y^2z \\ x^4z \end{bmatrix}$, and $g(x,y,z) = \ln(yz)$. Then we can certainly define

$$\mathbf{h}(x,y,z) = \mathbf{f}(x,y,z)g(x,y,z) = \begin{bmatrix} (xy+y^2z)\ln(yz) \\ (x^4z)\ln(yz) \end{bmatrix},$$

but we have to carefully choose our domain so that $\mathbf{h}$ makes sense. $\mathbf{f}$ is defined on all of $\mathbb{R}^3$, but the largest domain of g is the set

$$X = \left\{(x,y,z) \in \mathbb{R}^3 \;\middle|\; yz > 0\right\}.$$

Then $\mathbf{h} : X \subset \mathbb{R}^3 \to \mathbb{R}^2$ is defined as above. And on this open domain X, $\mathbf{h}$ will be differentiable (in fact, it is the product of differentiable functions). So we calculate the derivative in two ways: Directly, and via the Product Rule. Directly,

$$\begin{aligned} D\mathbf{h}(\mathbf{x}) &= \begin{bmatrix} \frac{\partial \mathbf{h_1}}{\partial x}(\mathbf{x}) & \frac{\partial \mathbf{h_1}}{\partial y}(\mathbf{x}) & \frac{\partial \mathbf{h_1}}{\partial z}(\mathbf{x}) \\ \frac{\partial \mathbf{h_2}}{\partial x}(\mathbf{x}) & \frac{\partial \mathbf{h_2}}{\partial y}(\mathbf{x}) & \frac{\partial \mathbf{h_2}}{\partial z}(\mathbf{x}) \end{bmatrix} \\ &= \begin{bmatrix} y\ln(yz) & (x+2yz)\ln(yz)+x+yz & y^2\ln(yz)+\frac{xy}{z}+y^2 \\ 4x^3z & \frac{x^4z}{y} & x^4\ln(yz)+x^4 \end{bmatrix}. \end{aligned}$$

Via the Product Rule, we have

$$\begin{aligned} D\mathbf{h}(\mathbf{x}) &= D\mathbf{f}(\mathbf{x}) \cdot g(\mathbf{x}) + \mathbf{f}(\mathbf{x}) \cdot Dg(\mathbf{x}) \\ &= \begin{bmatrix} \frac{\partial \mathbf{f_1}}{\partial x}(\mathbf{x}) & \frac{\partial \mathbf{f_1}}{\partial y}(\mathbf{x}) & \frac{\partial \mathbf{f_1}}{\partial z}(\mathbf{x}) \\ \frac{\partial \mathbf{f_2}}{\partial x}(\mathbf{x}) & \frac{\partial \mathbf{f_2}}{\partial y}(\mathbf{x}) & \frac{\partial \mathbf{f_2}}{\partial z}(\mathbf{x}) \end{bmatrix} \cdot g(\mathbf{x}) + \begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \end{bmatrix} \cdot \begin{bmatrix} \frac{\partial g}{\partial x}(\mathbf{x}) \\ \frac{\partial g}{\partial y}(\mathbf{x}) \\ \frac{\partial g}{\partial z}(\mathbf{x}) \end{bmatrix} \\ &= \begin{bmatrix} y & x+2yz & y^2 \\ 4x^3z & 0 & x^4 \end{bmatrix} \ln(yz) + \begin{bmatrix} xy+y^2z \\ x^4z \end{bmatrix} \cdot \begin{bmatrix} 0 & \frac{1}{y} & \frac{1}{z} \end{bmatrix}. \end{aligned}$$

I will leave it to the reader to see that these two matrices of functions are the same.

EXAMPLE 5.2. Now let $p = m > 1$, with the Dot Product on vectors. Note that, for ease of calculation here, we let $n = 1$. Then, for $\mathbf{f}, \mathbf{g} : X \subset \mathbb{R} \to \mathbb{R}^m$, the dot-product function is $h : X \subset \mathbb{R} \to \mathbb{R}$, where $h(x) = \mathbf{f}(\mathbf{x}) \cdot \mathbf{g}(\mathbf{x})$. Note that the product function h here is scalar-valued, but still has n inputs. Then

$$Dh(\mathbf{x}) = \begin{bmatrix} D_{x_1}h(\mathbf{x}) & \cdots & D_{x_n}h(\mathbf{x}) \end{bmatrix},$$

with

$$\begin{aligned} D_{x_i}h(\mathbf{x}) &= \frac{\partial}{\partial x_i}h(\mathbf{x}) = \frac{\partial}{\partial x_i}\left[\sum_{j=1}^{m} f_j(\mathbf{x})\cdot g_j(\mathbf{x})\right] \\ &= \sum_{j=1}^{n}\frac{\partial}{\partial x_i}\left[f_j(\mathbf{x})g_j(\mathbf{x})\right] \qquad \text{(Sum Rule)} \\ &= \sum_{j=1}^{m}\frac{\partial f_j}{\partial x_i}(\mathbf{x})g_j(\mathbf{x}) + f_j(\mathbf{x})\frac{\partial g_j}{\partial x_i}(\mathbf{x}) \qquad \text{(Calc I Product Rule)} \\ &= \sum_{j=1}^{m}\frac{\partial f_j}{\partial x_i}(\mathbf{x})g_j(\mathbf{x}) + \sum_{j=1}^{m} f_j(\mathbf{x})\frac{\partial g_j}{\partial x_i}(\mathbf{x}) \qquad \text{(Sum Rule)} \\ &= Df(\mathbf{x})\cdot g(\mathbf{x}) + f(\mathbf{x})\cdot Dg(\mathbf{x}), \end{aligned}$$

where each of the four pieces in this last sum of products is an m-vector. Notice that in the middle of this last calculation, we were simply multiplying together scalar-valued functions, so there was no $\cdot$ present.

Now, as a special note of caution: Be careful with vector products. The two examples above are symmetric products, named because

$$f(x)\cdot g(x) = g(x)\cdot f(x)$$

If the product is not symmetric, then the order of the factors matters. But the Product Rule will still work correctly. One jsut needs to pay attention to the order of the factors inside the product rule. For example, for $\mathbf{a},\mathbf{b}\in\mathbb{R}^4$, the cross product is called *antisymmetric*, since

$$\mathbf{a}\times\mathbf{b} = -\mathbf{b}\times\mathbf{a}.$$

Perhaps you already know this via a detailed calculation. However, we will see why this is true geometrically in a while. Hence for $\mathbf{f},\mathbf{g}: X\subset\mathbb{R}\to\mathbb{R}^3$, and $\mathbf{h}(x) = \mathbf{f}(x)\times\mathbf{g}(x)$, we have

$$D\mathbf{h}(x) = D\left(\mathbf{f}\times\mathbf{g}\right)(x) = D\mathbf{f}(x)\times\mathbf{g}(x) + \mathbf{f}(x)\times D\mathbf{g}(x).$$

And lastly, on this note, the Quotient Rule also hold, but only where it makes sense. One thing to keep in mind for the Quotient Rule is that the denonimator function myust be scalar-valued for even a quotient of functions to make sense. (Why?) At that point, the square of the denominator function in the Quotient Rule will also make sense.

A Note on Partial Derivatives. Given a differentiable real-valued function $f: X\subset\mathbb{R}^3\to\mathbb{R}$, say, we know that all of

$$\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} : X\to\mathbb{R}$$

are all continuous in a neighborhood of $\mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$. In this case, where all partials of a function are continuous on a domain, we say that the function is a C^1-*function*, or write $f \in C^1$.

DEFINITION 5.1. A *second partial derivative* of f with respect to a variable (x, say) is any one of

$$\begin{aligned} \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) &: X \to \mathbb{R}, \\ \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) &: X \to \mathbb{R}, \\ \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial z}\right) &: X \to \mathbb{R}. \end{aligned}$$

We also write

$$\begin{aligned} \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) &= \frac{\partial^2 f}{\partial x^2} = f_{xx} \\ \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) &= \frac{\partial^2 f}{\partial x \partial y} = f_{yx}, \\ \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial z}\right) &= \frac{\partial^2 f}{\partial x \partial z} = f_{zx}. \end{aligned}$$

Pay attention to the order of the variables, and hence the derivatives here. Indeed, the order of differentiation is written differently in the two notations, fractional and subscript-wise. Be careful here. Now If all 9 of these second partial derivative of f exist and are continuous on the domain X, then we say that $f \in C^2$.

Generalizing, let $f : X \subset \mathbb{R}^n \to \mathbb{R}$. For $i_1, i_2, \ldots, i_k \in \{1, 2, \ldots, n\}$, the kth partial derivative of f with respect to $x_{i_1}, \ldots, x_{i_k}$, is

$$\frac{\partial^k f}{\partial x_{i_k} \cdots \partial x_{i_1}}(\mathbf{x}) = \frac{\partial}{\partial x_{i_k}}\left(\cdots\left(\frac{\partial f}{\partial x_{i_1}}\right)\cdots\right) = f_{x_{i_1}\cdots x_{i_k}}.$$

EXAMPLE 5.3. Let $f : \mathbb{R}^3 \to \mathbb{R}$ be defined by $f(x, y, z) = z\cos(2xy)$. It should be readily apparent, and can be rigorously shown, that polynomials in many variables are continuous everywhere, and differentiable everywhere. The same it true for the cosine function. And since f is a product of a polynomial and a composition of the cosine function and a polynomial, $f \in C^1$, and we can calculate

$$\begin{aligned} f_x(x, y, z) &= -z\sin(2xy)2y = -2yz\sin(2xy), \\ f_y(x, y, z) &= -z\sin(2xy)2x = -2xz\sin(2xy), \quad \text{and} \\ f_z(x, y, z) &= \cos(2xy). \end{aligned}$$

But all three of these partial derivatives of f are also sums, products and compositions of differentiable functions, so that $f \in C^2$ also. Thus, we

find

$$f_{xy}(x,y,z) = \frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = -2z\sin(2xy) - 4xyz\cos(2xy), \quad \text{and}$$

$$f_{yx}(x,y,z) = \frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = -2z\sin(2xy) - 4xyz\cos(2xy).$$

Notice immediately that these two functions are the same. It turns out that this is true always when $f \in C^2$:

THEOREM 5.2. *Suppose $f : X \subset \mathbb{R}^n \to \mathbb{R}$ is C^2 on an open X. Then, for any choice of $i_1, i_2 \in \{1, 2, \ldots, n\}$,*

$$\frac{\partial^2 f}{\partial x_{i_1} \partial x_{i_2}} = \frac{\partial^2 f}{\partial x_{i_2} \partial x_{i_1}}.$$

The proof is constructive and in the book. We will not do it in class.

DEFINITION 5.3. A function $f : X \subset \mathbb{R}^n \to \mathbb{R}$ is of class C^k, $k \in \mathbb{N}$ if it has continuous partial derivative up to and including order k. A function $\mathbf{g} : X \subset \mathbb{R}^n \to \mathbb{R}^m$ is of class C^k if each component function $g_i : X \to \mathbb{R}$ is of class C^k.

And finally, a function like the above is of class C^∞ if it is *smooth*. This means that it has continuous partial derivatives of all orders. Some notes:

- This should be an obvious fact, but worth stating explicitly: If $f \in C^k$, then $f \in C^\ell$ for all $\ell < k$.
- A continuous function is said to be of class C^0.

So here is a thought experiment: Suppose $f : X \subset \mathbb{R}^n \to \mathbb{R}$ is of class C^∞, so it is a smooth function. Then we know the following:

(1) f has n first partial derivatives, and
(2) f has n^2 second partials, since each of the n first partials has n second partials.
(3) Thus f has n^k kth partials, for each $k \in \mathbb{N}$.

Now $Df(\mathbf{x})$ is a row matrix with n entries, each entry a real-valued function on n variables. Each of these entries is also differentiable. Plug in a value to evaluate the derivative of f at a point, and one gets a matrix of numbers. But without evaluation, $Df(\mathbf{x})$ is a (row)-matrix of functions. Just for a moment, view this row matrix as a column matrix. Then, in a way, $Df : X \subset \mathbb{R}^n \to \mathbb{R}^n$. And then $D(Df) = D^2 f(\mathbf{x})$ will be an $n \times n$ matrix of functions, with each entry, $\frac{\partial^2 f}{\partial x_i \partial x_j}$ a real-valued function on n variables. If we, for the moment think of $D^2 f$ as a function on X, then what is its codomain?

And, since f is smooth, the object $D\left(D^2 f\right) = D^3 f$ exists! What kind of object is this?

And in general, what kind of object is $D^k f(\mathbf{x})$, for $k \in \mathbb{N}$?

These objects will play a role in the multivariable Taylor expansion of a function like f, since Taylor series' of functions exist in multivariable calculus

and will (must) account for all of the derivatives of a function. Think about this....

LECTURE 6

The Chain Rule

SYNOPSIS. Here, we define and discuss the Chain Rule in the differential calculus of vector-valued functions of more than one independent variable. One can use the Calculus I version to define the multivariable calculus version, which works in the same fashion. However, care must be taken for two reasons: (1) the derivatives of functions here are not the same kinds of functions as the original functions, and (2) composition is tricky when the domains and codomains can be of different sizes. We discuss this at length here.

The Chain Rule.

6.0.0.1. *The Chain Rule in single variable calculus.* Recall from Calculus I: For $f, g : \mathbb{R} \to \mathbb{R}$, where $f, g \in C^1$,

$$\frac{d}{dx}\left(f \circ g\right)(x) = f'\left(g(x)\right) \cdot g'(x).$$

In essence, the derivative of a composition of functions is the product of the derivatives..., (but with a definite *twist*! - The derivative of the "outside" function is evaluated at the image of x under the "inside" function. This leads to the immediate question of just how the domain of a composition depends on the domains of the constituent pieces in the composition. To see this, let $f : J \subset \mathbb{R} \to \mathbb{R}$ and $g : I \subset \mathbb{R} \to \mathbb{R}$ be defined, but only on the subsets of the real line. Then

$$\textbf{domain}\left(f \circ g\right) = \left\{x \in I \;\middle|\; g(x) \in J\right\} = g^{-1}(J) \subset I.$$

Be careful here, though, as $g^{-1}(J)$ is the set inverse of g, which makes sense even if g does not have an inverse as a function.

EXAMPLE 6.1. Let $f(x) = \sqrt{x}$ and $g(x) = 2 - x^2$. Of course, without specifying a domain, the domain of each of these is automatically the largest set on which the function makes sense. In these cases, adn using the notation of the above discussion, $f : J \to \mathbb{R}$, with $J = [0, \infty)$, and $g : I \to \mathbb{R}$, where $I = \mathbb{R}$. So what is the domain of $(f \circ g)$? One way to see this is to simply construct the function:

$$\left(f \circ g\right)(x) = f(g(x)) = f(2 - x^2) = \sqrt{2 - x^2}.$$

With this, the domain can only include points that satisfy $2 - x^2 \geq 0$, so $x \in \left[-\sqrt{2}, \sqrt{2}\right]$. And thinking of this in terms of sets alone, we can calculate

$$\begin{aligned}\textbf{domain}\,(f \circ g)(x) = g^{-1}(J) = g^{-1}([0,\infty)) &= \left\{x \in I \;\middle|\; g(x) \in J\right\} \\ &= \left\{x \in \mathbb{R} \;\middle|\; 2 - x^2 \in [0,\infty)\right\} = \left[-\sqrt{2}, \sqrt{2}\right].\end{aligned}$$

But here is an issue: What is the domain of $(g \circ f)$? Calculating the function, we get

$$(g \circ f)(x) = g(f(x)) = g(\sqrt{x}) = 2 - \left(\sqrt{x}\right)^2 = 2 - x.$$

Without paying attention, one may wrongly assume that the domain is all of $\mathbb{R}$, since $(g \circ f)(x) = 2 - x$ is a degree-1 polynomial. However, the "inside" function has as its domain only the non-negative reals $[0, \infty)$. Hence so does $(g \circ f)$!

To use the set notation, note that J and I are switched here, and

$$\begin{aligned}\textbf{domain}\,(g \circ f)(x) = f^{-1}(I) = f^{-1}(\mathbb{R}) &= \left\{x \in J \;\middle|\; f(x) \in I\right\} \\ &= \left\{x \in [0,\infty) \;\middle|\; \sqrt{x} \in \mathbb{R}\right\} = [0,\infty).\end{aligned}$$

Note: In Leibniz notation, let $z = g(y)$, and $y = f(x)$, so that $z = (g \circ f)(x) = g(f(x))$. Then z is considered a function of x, and the Chain Rule looks like

$$\frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx}.$$

One can directly and easily again see this notion that the derivative of a product of functions is, in fact, the product of the derivatives. However, when evaluated the derivative of a composition at a point, the "twist" in the product again becomes clear, and

$$\left.\frac{dz}{dx}\right|_{x=a} = \left.\frac{dz}{dy}\right|_{y=g(a)} \cdot \left.\frac{dy}{dx}\right|_{x=a}.$$

EXAMPLE 6.2. Back to the previous example and translating into Leibniz notation, we have $y = f(x) = \sqrt{x}$, and $z = g(y) = 2 - y^2$. Then

$$z = (g \circ f)(x) = g(f(x)) = g(\sqrt{x}) = 2 - \left(\sqrt{x}\right)^2 = 2 - x, \quad \text{on} \quad [0,\infty).$$

Its derivative, defined on $(0,\infty)$, should be $\frac{dz}{dx} = -1$ everywhere. Here

$$\frac{dz}{dx} = \left.\frac{dz}{dy}\right|_{y=f(x)} \cdot \frac{dy}{dx} = -2y\Big|_{y=\sqrt{x}} \cdot \left(\frac{1}{2\sqrt{x}}\right) = \left(-2\sqrt{x}\right)\left(\frac{1}{2\sqrt{x}}\right) = -1.$$

6.0.1. The Chain Rule in multivariable calculus. In vector calculus, the Chain Rule still holds:

THEOREM (Theorem 2.5.3 in text). *Suppose $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^m$ are open, and $\mathbf{f} : Y \to \mathbb{R}^p$ and $\mathbf{g} : X \to \mathbb{R}^m$ are defined so that $\mathbf{g}(X) \subset Y$. Then, if $\mathbf{g}$ is differentiable at $\mathbf{x}_0 \in X$, and $\mathbf{f}$ is differentiable at $\mathbf{y}_0 = \mathbf{g}(\mathbf{x}_0) \in Y$, then $(\mathbf{f} \circ \mathbf{g})$ is differentiable at $\mathbf{x}_0$, with*

$$D(\mathbf{f} \circ \mathbf{g})(\mathbf{x}_0) = D\mathbf{f}(\mathbf{g}(\mathbf{x}_0))\, D\mathbf{g}(\mathbf{x}_0).$$

EXAMPLE 6.3. Let $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^3$, $\mathbf{f}(x, y) = (x^2 y, 1, e^{xy})$, and $g : \mathbb{R}^3 \to \mathbb{R}$, with $g(x, y, z) = xyz$. We calculate $D(g \circ \mathbf{f})(x, y)$ in two ways:

(1) **Composition before derivative.** Here, $(g \circ \mathbf{f}) : \mathbb{R}^2 \to \mathbb{R}$, and

$$(g \circ \mathbf{f})(x, y) = g(\mathbf{f}(x, y)) = g(x^2, 1, e^{xy}) = x^2 y e^{xy}.$$

Then

$$\begin{aligned} D(g \circ \mathbf{f})(x, y) &= \begin{bmatrix} \frac{\partial (g \circ \mathbf{f})}{\partial x}(x, y) & \frac{\partial (g \circ \mathbf{f})}{\partial y}(x, y) \end{bmatrix} \\ &= \begin{bmatrix} 2xye^{xy} + x^2 y^2 e^{xy} & x^2 e^{xy} + x^3 y e^{xy} \end{bmatrix}. \end{aligned}$$

(2) **Via the Chain Rule.** The derivatives of the constituent functions are

$$D\mathbf{f}(x, y) = \begin{bmatrix} 2xy & x^2 \\ 0 & 0 \\ ye^{xy} & xe^{xy} \end{bmatrix} \quad \text{and} \quad Dg(x, y, z) = \begin{bmatrix} yz & xz & xy \end{bmatrix}.$$

So $Dg(\mathbf{f}(x, y)) = Dg(x^2 y, 1, e^{xy}) = \begin{bmatrix} e^{xy} & x^2 y e^{xy} & x^2 y \end{bmatrix}$. With the Chain Rule, we get

$$\begin{aligned} D(g \circ \mathbf{f})(x, y) &= Dg(\mathbf{f}(x, y)) \cdot D\mathbf{f}(x, y) \\ &= \begin{bmatrix} e^{xy} & x^2 y e^{xy} & x^2 y \end{bmatrix} \begin{bmatrix} 2xy & x^2 \\ 0 & 0 \\ ye^{xy} & xe^{xy} \end{bmatrix} \\ &= \begin{bmatrix} 2xye^{xy} + x^2 y^2 e^{xy} & x^2 e^{xy} + x^3 y e^{xy} \end{bmatrix} \end{aligned}$$

as before.

Now you may be thinking that the variables can be confusing here, with x and y included in the two domains, $\mathbb{R}^2$ for $\mathbf{f}$, and $\mathbb{R}^3$ for g. In a very important way, they are not the same, and should not be considered so! One way to correct this error of notation, and also to make things much more clear, is to switch the names of the variables, using different variables for each domain. Indeed, Let us denote the function $\mathbf{f}$ as before, but noticing explicitly that it has three component functions

$$\mathbf{f}(x, y) = (f_1(x, y), f_2(x, y), f_3(x, y)) = (x^2 y, 1, e^{xy}),$$

and now define $g(u,v,w) = uvw$, the same function as before, but with new variable names. Then the two derivatives are, as before, but look like

$$Dg(u,v,w) = \begin{bmatrix} \frac{\partial g}{\partial u} & \frac{\partial g}{\partial v} & \frac{\partial g}{\partial w} \end{bmatrix}, \quad \text{and} \quad D\mathbf{f}(x,y) = \begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \\ \frac{\partial f_3}{\partial x} & \frac{\partial f_3}{\partial y} \end{bmatrix}.$$

Now, within the composition, we know that

$$\begin{aligned} u &= f_1(x,y) = x^2 y \\ v &= f_2(x,y) = 1 \\ w &= f_3(x,y) = e^{xy}. \end{aligned}$$

Hence the derivative of the composition, which is

$$Dg\left(\mathbf{f}(x,y)\right) \cdot D\mathbf{f}(x,y) = \begin{bmatrix} \frac{\partial (g\circ\mathbf{f})}{\partial x}(x,y) & \frac{\partial (g\circ\mathbf{f})}{\partial y}(x,y) \end{bmatrix},$$

where by direct matrix multiplication

$$\begin{aligned} \frac{\partial\left(g\circ\mathbf{f}\right)}{\partial x}(x,y) &= \frac{\partial g}{\partial u}\cdot\frac{\partial f_1}{\partial x} + \frac{\partial g}{\partial v}\cdot\frac{\partial f_2}{\partial x} + \frac{\partial g}{\partial w}\cdot\frac{\partial f_3}{\partial x} \\ &= \frac{\partial g}{\partial u}\cdot\frac{\partial u}{\partial x} + \frac{\partial g}{\partial v}\cdot\frac{\partial v}{\partial x} + \frac{\partial g}{\partial w}\cdot\frac{\partial w}{\partial x} \\ &= vw\Big|_{\substack{v=1\\ w=e^{xy}}} \cdot (2xy) + uw\Big|_{\substack{u=x^2y\\ w=e^{xy}}} \cdot (0) + uv\Big|_{\substack{u=x^2y\\ v=1}} \cdot (ye^{xy}) \\ &= 2xye^{xy} + x^2y^2e^{xy}. \end{aligned}$$

Here, the products of the partials in these derivative of compositions are always understood to have the "twist", as mentioned earlier, so that

$$\frac{\partial g}{\partial u}\cdot\frac{\partial u}{\partial x} = \frac{\partial g}{\partial u}\Big|_{\mathbf{u}=\mathbf{f}(\mathbf{x})} \cdot \frac{\partial u}{\partial x}\Big|_{\mathbf{x}}, \quad \text{where} \quad \mathbf{u} = \begin{bmatrix} u \\ v \\ w \end{bmatrix}, \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}.$$

Here is one more example:

EXAMPLE 6.4. Let $\mathbf{c}:\mathbb{R}\to\mathbb{R}^3$ be a C^1-curve in three-space, and $f:\mathbb{R}^3\to\mathbb{R}$ be a C^1-scalar-valued function on R^3. Then the composition $g = f\circ\mathbf{c}:\mathbb{R}\to\mathbb{R}$ is just f evaluated along the curve, and looks like a function from $\mathbb{R}$ to $\mathbb{R}$. One often writes

$$g = f\big|_{\mathbf{c}}.$$

In this sense, $g'(t) = \frac{df}{dt}(t)$ along $\mathbf{c}$. We calculate this quantity via the Chain Rule:

Here $\mathbf{c}(t) = \begin{bmatrix} x(t) \\ y(t) \\ z(t) \end{bmatrix}$ is C^1, and

$$\frac{d\mathbf{c}}{dt}(t) = \mathbf{c}'(t) = \begin{bmatrix} x'(t) \\ y'(t) \\ z'(t) \end{bmatrix} = \begin{bmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \\ \frac{dz}{dt} \end{bmatrix},$$

while $Df(x,y,z) = \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} & \frac{\partial f}{\partial z} \end{bmatrix}$.

Hence

$$\begin{aligned} g'(t) = D\left(f \circ \mathbf{c}\right)(t) = \frac{df\big|_{\mathbf{c}}}{dt} &= Df\left(\mathbf{c}(t)\right) \cdot D\mathbf{c}(t) \\ &= \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} & \frac{\partial f}{\partial z} \end{bmatrix} \begin{bmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \\ \frac{dz}{dt} \end{bmatrix} \\ &= \frac{\partial f}{\partial x} \cdot \frac{dx}{dt} + \frac{\partial f}{\partial y} \cdot \frac{dy}{dt} + \frac{\partial f}{\partial z} \cdot \frac{dz}{dt} \\ &= \left.\frac{\partial f}{\partial x}\right|_{\mathbf{c}(t)} \cdot x'(t) + \left.\frac{\partial f}{\partial y}\right|_{\mathbf{c}(t)} \cdot y'(t) + \left.\frac{\partial f}{\partial z}\right|_{\mathbf{c}(t)} \cdot z'(t). \end{aligned}$$

LECTURE 7

Directional Derivatives

SYNOPSIS. Today, we move into directional derivatives, a generalization of a partial derivative where we look for how a function is changing at a point in any single direction in the domain. This gives a powerful tool, both conceptually as well as technically, to discuss the role the derivative of a function plays in exposing the properties of both functions on and sets within Euclidean space. We define the gradient of a real-valued function (finally) and its interpretations and usefulness, and move toward one of the most powerful theorems of multivariable calculus, the Implicit Function Theorem.

The Directional Derivative.

7.0.0.1. *Vector form of a partial derivative.* Recall the definition of a partial derivative evaluated at a point: Let $f : X \subset \mathbb{R}^2 \to \mathbb{R}$, x open, and $(a, b) \in X$. Then the partial derivative of f with respect to the first coordinate x, evaluated at (a, b) is

$$\frac{\partial f}{\partial x}(a, b) = \lim_{h \to 0} \frac{f(a + h, b) - f(a, b)}{h}.$$

Here, we vary only the first coordinate, leaving the y coordinate value b fixed, and write $(a + h, b) = (a, b) + (h, 0)$. In vector notation, this is like taking the vector $\mathbf{a} = \begin{bmatrix} a \\ b \end{bmatrix}$, and adding to it a small amount h, but only in the x-direction. Indeed, this means adding to $\mathbf{a}$ the vector $h \begin{bmatrix} 1 \\ 0 \end{bmatrix} = h\mathbf{i}$, where here we use the standard convention for unit vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$, namely $\mathbf{i} = \mathbf{e_1}$, $\mathbf{j} = \mathbf{e_2}$, etc. We get

$$\mathbf{a} + h\mathbf{i} = \begin{bmatrix} a \\ b \end{bmatrix} + h \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix} + \begin{bmatrix} h \\ 0 \end{bmatrix} = \begin{bmatrix} a + h \\ b \end{bmatrix}.$$

Then the definition of a partial derivative becomes

$$\frac{\partial f}{\partial x}(\mathbf{a}) = \lim_{h \to 0} \frac{f(\mathbf{a} + h\mathbf{i}) - f(\mathbf{a})}{h}.$$

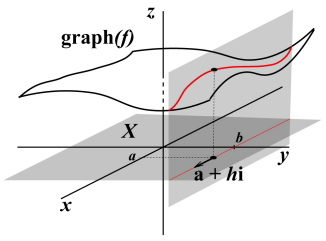


FIGURE 23. A directional derivative in the x-direction is the partial.

However, one can take a derivative of f at a point (a,b), or the point $\mathbf{a} = \begin{bmatrix} a \\ b \end{bmatrix}$ in *any* direction in the domain: Let $\mathbf{v} \in X$. Then

$$\lim_{h\to 0} \frac{f(\mathbf{a}+h\mathbf{v}) - f(\mathbf{a})}{h}$$

is perfectly well defined as long as the quantity $\mathbf{a}+h\mathbf{v}$ remains in X, which, since X is open, will be the case for small enough h. This is the derivative of f at (a,b) *in the direction of* $\mathbf{v}$, also known as the *directional derivative* of f at (a,b) with respect to $\mathbf{v}$:

$$D_{\mathbf{v}} f(\mathbf{a}) = \lim_{h\to 0} \frac{f(\mathbf{a}+h\mathbf{v}) - f(\mathbf{a})}{h}.$$

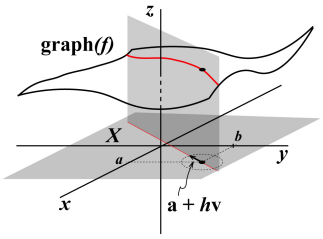


FIGURE 24. A directional derivative in the direction of $\mathbf{v} \in X$.

How does this work? For f differentiable at $\mathbf{a}$, compose f with the affine function $\mathbf{g} : \mathbb{R} \to \mathbb{R}^2$, where

$$\mathbf{g}(t) = \mathbf{a} + t\mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix} + t \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}.$$

Here, $\mathbf{g}$ parameterizes a line in $\mathbb{R}^2$ where at $t = 0$, $\mathbf{g}(0) = \mathbf{a}$, and at $t = 1$, $\mathbf{g}(0) = \mathbf{a}+\mathbf{v}$. $\mathbf{g}$ is also C^1, and $\mathbf{g}'(t) = \mathbf{v}$. In particular, $\mathbf{g}'(0) = \mathbf{v}$.

Now let $F(t) = f(\mathbf{g}(t)) = (f \circ \mathbf{g})(t) = f(\mathbf{a}+t\mathbf{v})$, like in our definition of directional derivative. Here, F, as the composition of two differentiable function, will also be differentiable, and

$$F'(0) = \lim_{t\to 0} \frac{F(t) - F(0)}{t - 0} = \lim_{t\to 0} \frac{f(\mathbf{a}+t\mathbf{v}) - f(\mathbf{a})}{t}.$$

But, using the Chain Rule, we can write

$$F'(0) = D_{\mathbf{v}} f(\mathbf{a}) = \frac{d}{dt}\bigg|_{t=0} f(\mathbf{a}+t\mathbf{v}) = Df(\mathbf{g}(0))\,\mathbf{g}'(0) = Df(\mathbf{a})\mathbf{v}.$$

DEFINITION 7.1. Let $f : X \subset \mathbb{R}^n \to \mathbb{R}$ be C^1. Then the *gradient function* of f is the function

$$\nabla f : X \subset \mathbb{R}^n \to \mathbb{R}^n, \quad \nabla f(\mathbf{x}) = \begin{bmatrix} f_{x_1}(\mathbf{x}) \\ f_{x_2}(\mathbf{x}) \\ \vdots \\ f_{x_n}(\mathbf{x}) \end{bmatrix}.$$

The *gradient vector* of f at $\mathbf{a} \in X$ is a vector in $\mathbb{R}^n$ based at $\mathbf{a}$:

$$\nabla f(\mathbf{a}) = \begin{bmatrix} f_{x_1}(\mathbf{a}) \\ f_{x_2}(\mathbf{a}) \\ \vdots \\ f_{x_n}(\mathbf{a}) \end{bmatrix}.$$

Notes:

- The gradient function carries the same information as the derivative matrix of f, but is a *vector* of functions so that

$$Df(\mathbf{x}) = (\nabla f)^T, \quad \text{where} \quad T = \text{ transpose.}$$

- The gradient is only defined for scalar-valued functions.

Using this gradient function, we can write

$$D_{\mathbf{v}} f(\mathbf{a}) = \underbrace{Df(\mathbf{a})\mathbf{v}}_{\text{matrix mult.}} = \underbrace{\nabla f(\mathbf{a}) \cdot \mathbf{v}}_{\text{dot product}}.$$

Warning! The choice of $\mathbf{v}$ is really a choice of direction only! Thus, it is vitally important that $\|\mathbf{v}\| = 1$ for this choice.

EXERCISE 2. Show that for $k \in \mathbb{R}$ and $\mathbf{w} = k\mathbf{v}$, that $D_{\mathbf{w}} f(\mathbf{a}) = kD_{\mathbf{v}} f(\mathbf{a})$.

The directional derivative specifies how f is changing in the direction of $\mathbf{v} \in X$. But what does this mean? Imagine standing in $X \in \mathbb{R}^n$ at a point $\mathbf{a}$ where a real-valued f is defined and differentiable. How is f changing in the particular direction that you are facing at the moment? For any $\mathbf{v} \in \mathbb{R}^n$, where $\|\mathbf{v}\| = 1$, $D_{\mathbf{v}} f(\mathbf{a}) = \nabla f(\mathbf{a}) \cdot \mathbf{v}$. So recall that

$$\mathbf{x} \cdot \mathbf{y} = \|\mathbf{x}\| \, \|\mathbf{y}\| \cos\theta,$$

where θ is the angle between $\mathbf{x}$ and $\mathbf{y}$. Remember that, for any $n > 1$, any two non-collinear (what does this mean?) vectors in $\mathbb{R}^n$ span a plane. Within that plane, there is a well-defined angle between them. So

$$D_{\mathbf{v}} f(\mathbf{a}) = \nabla f(\mathbf{a}) \cdot \mathbf{v} = \|\nabla f(\mathbf{a})\| \cos\theta,$$

since $\|\mathbf{v}\| = 1$.

But notice then that

$$-\|\nabla f(\mathbf{a})\| \leq D_{\mathbf{v}}(\mathbf{a}) \leq \|\nabla f(\mathbf{a})\|.$$

Thus the directional derivative of f at $\mathbf{a}$ will achieve its maximum when $\theta = 0$, and its minimum when $\theta = \pi$. And, of course, the directional derivative will be 0 precisely when $\theta = \pm\frac{\pi}{2}$. All of this comes from the Dot Product of the gradient vector and the chosen unit-length directional vector $\mathbf{v}$. Geometrically, what does this mean? Here is a beautiful and important interpretation:

THEOREM 7.2. *Let $X \in \mathbb{R}^n$ be open and $f : X \to \mathbb{R}$ a C^1-function. For $\mathbf{x}_0 \in X$, let*

$$S_{\mathbf{x}_0} = \left\{ \mathbf{x} \in X \;\middle|\; f(\mathbf{x}) = f(\mathbf{x}_0) = c \right\}.$$

The $\nabla f(\mathbf{x}_0) \perp S_{\mathbf{x}_0}$.

Another way to say this is that any vector $\mathbf{v}$ tangent to $S_{\mathbf{x}_0}$ will be perpendicular to $\nabla f(\mathbf{x}_0)$ (See Figure 25. The proof of this is constructive and very informative.

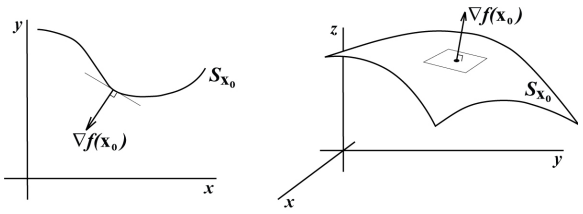


FIGURE 25. Geometrically, the gradient vector is always perpendicular to the level sets of a function.

PROOF. Let $I = (a, b)$, an open interval in $\mathbb{R}$, and $\mathbf{c} : I \to \mathbb{R}^n$ be a C^1-parameterized curve, with $\mathbf{c}(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix}$ such that

(1) $\mathbf{c}(t_0) = \mathbf{x}_0$, for some $t_0 \in I$, and
(2) $\mathbf{c}(I) \subset S_{\mathbf{x}_0}$.

Then the composition $(f \circ \mathbf{c}) : I \to \mathbb{R}$ is C^1 on I since both f and the curve are, and

$$f(\mathbf{c}(t)) = f(x_1(t), \ldots, x_n(t)) = c.$$

Differentiate this last equation inplicitly with respect to t. we get

$$\frac{d}{dt}\left[f(x_1(t), \ldots, x_n(t))\right] = \frac{d}{dt}\left[c\right] = 0$$
$$Df(\mathbf{c}(t))\,\mathbf{c}'(t) = 0.$$

Now, at $t = t_0$, $\mathbf{c}(t_0) = \mathbf{x}_0$, and $Df(\mathbf{x}_0)\,\mathbf{v} = 0$, where $\mathbf{v} = \mathbf{c}'(t_0)$, a vector tangent to the curve and hence tangent to $S_{\mathbf{x}_0}$. Hence the result follows. □

DEFINITION 7.3. For any $(n-1)$-dimensional hypersurface in $\mathbb{R}^n$ defined as the c-level set of a C^1 function $f : X \subset \mathbb{R}^n \to \mathbb{R}$,

$$S = \left\{\mathbf{x} \in X \;\middle|\; f(\mathbf{x}) = c\right\},$$

the tangent space to S at $\mathbf{a} \in S$ is the space of all vectors perpendicular to $\nabla f(\mathbf{a})$; it is defined by $h(\mathbf{x}) = \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}) = 0$, or

$$h(\mathbf{x}) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{a})(x_i - a_i) = 0.$$

Note: Compare this to the tangent space of $\mathbf{graph}(f) \subset \mathbb{R}^3$, where $f : \mathbb{R}^2 \to \mathbb{R}$ and $\mathbf{graph}(f)$ is defined by the equation in $\mathbb{R}^3$ given by $z = f(x, y)$.

LECTURE 8

Implicit and Inverse Function Theorems

SYNOPSIS. Here, give a treatment of both the Implicit Function Theorem (for real-valued functions), and the Inverse Function Theorem. These are very powerful theorems that expose some of the hidden structure of real-valued and vector-valued functions of more than one variable. We will study the ideas in class, and here is a proof of the Implicit Function Theorem for a function on (a subset of) three space. And here is a Mathematica Notebook for this class.

HELPFUL DOCUMENTS.

- Mathematica: `ImplicitFunctionTheoremExample`.
- PDF: `IFTproof`

8.1. The Implicit Function Theorem.

8.1.1. In three variables. Recall the definition of a c-level set of a function $F : X \subset \mathbb{R}^n \to \mathbb{R}$:

$$S_c = \left\{ \mathbf{x} \in \mathbb{R}^n \;\middle|\; f(\mathbf{x}) = c \right\}.$$

For this discussion, let $F \in C^1$, and $n = 3$. We are using an upper case F here for a reason, which should be clear in the following discussion. Here is an example for motivation:

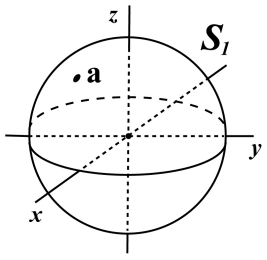


FIGURE 26. The unit 2-sphere $S_1 \in \mathbb{R}^3$.

EXAMPLE 8.1. Define $F : \mathbb{R}^3 \to \mathbb{R}$ by $F(x, y, z) = x^2 + y^2 + z^2$, and let $\mathbf{a} \in S_1$. Geometrically, here, $S_1 \in \mathbb{R}^3$ is the *unit sphere*, the sphere of radius-1 in three space. It is depicted in Figure 26.

Some questions:

Question 1. Is it possible to view S_1 as the graph of a function where we think of one variable as a dependent variable and all of the others still independent. Thus, in this case, can we write S_1 as the graph of $z = f(x, y)$ (this would be a different function than F)? The answer here is no! But, specifically, why not?

Question 2. Is it possible to write S_1 as $z = f(x, y)$ "locally", near $\mathbf{a} \in S_1$? The answer here is "depends...". But specifically, depends on what? Where $\mathbf{a}$ is located. Specifically whether the point in question is along the equator or not.

Question 3. So what information about F can be used to determine whether we can locally think of a level set of a function as the graph of (a different) function, with one variable a dependent variable and the other independent?

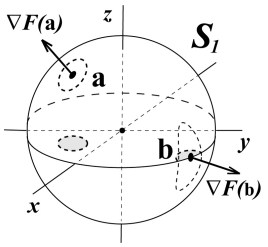


FIGURE 27. ∇F helps determine where level sets locally look like graphs of functions.

A central tool for this study will be the gradient of F: $\nabla F(\mathbf{a}) = \begin{bmatrix} F_x(\mathbf{a}) \\ F_y(\mathbf{a}) \\ F_z(\mathbf{a}) \end{bmatrix} = \begin{bmatrix} 2x \\ 2y \\ 2z \end{bmatrix}$. For $F(x, y, z) = x^2 + y^2 + z^2$, then, what property does ∇F have for points along the equator that is different from other points on S_1? Namely, for points like $\mathbf{b}$ along the equator in Figure 27, $F_z(\mathbf{b}) = 0$. Off of the equator, like point $\mathbf{a}$ in Figure 27, $F(\mathbf{a}) \neq 0$. So for $\mathbf{b} \in S_1$ along the equator, take any small open set *inside* S_1 containing $\mathbf{b}$ (This is the dotted oval sitting on the surface, where it bends around the sphere a bit). If we try to write these points in a form $z = f(x, y)$, we would wind up with some values for x and y with two points for z, following the function $z = \pm\sqrt{1 - x^2 - y^2}$. Here, $F_z(\mathbf{b}) = 0$ means that the gradient vector has no component in the z direction. It means that the gradient vector is "horizontal" (read: perpendicular to the z-direction). This means that the tangent plane to S_1 at the point $\mathbf{b}$ would look "vertical" here (all vector with only a z-component would be inside the tangent plane).

Hence the condition that $F_z(\mathbf{a}) \neq 0$ is a sufficient condition for being able to locally write $F(x, y, z) = c$ near $\mathbf{a}$ as $z = f(x, y)$ for some function f. This works equally well in n-dimensions:

THEOREM 8.1 (Theorem 2.6.5). *Let $F : X \subset \mathbb{R}^n \to \mathbb{R}$ be C^1 and $\mathbf{a} = (a_1, \ldots, a_{n-1}) \in S_c$, where*

$$S_c = \left\{ \mathbf{x} \in X \;\middle|\; F(\mathbf{x}) = c \right\}.$$

If $F_{x_n}(\mathbf{a}) \neq 0$, then there exists a neighborhood U of $(a_1, \ldots, a_{n-1}) \in \mathbb{R}^{n-1}$, a neighborhood V of $a_n \in \mathbb{R}$, and a C_1-function $f : U \subset \mathbb{R}^{n-1} \to V$, such that when $(x_1, \ldots, x_{n-1}) \in U$, and $x_n \in V$, then $x_n = f(x_1, \ldots, x_n)$.

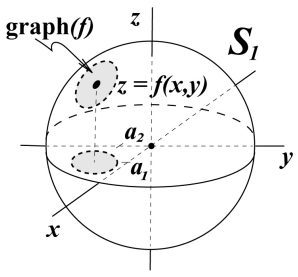


FIGURE 28. Near $\mathbf{a}$, S_1 looks like the graph of $z = f(x, y) = \sqrt{1 - x^2 - y^2}$.

Go back to our example of the 1-level set S_1 of the function $F(x,y,z) = x^2+y^2+z^2$. If we choose $\mathbf{a}$ strictly inside the northern hemisphere of S_1, as in Figure 28, then for these points, we can "solve" for z as a function of x and y:

$$z = f(x,y) = \sqrt{1-x^2-y^2}\,.$$

But to do this on a neighborhood of $U(\mathbf{a})$, we need to make sure that U includes no points from the equator. So choose $\mathbf{a} \subset S_1$ from the northern hemisphere. Now since $\mathbf{a} = (a_1,a_2,a_3)$ satisfies $a_1^2+a_2^2+a_3^2 = 1$, and $a_3 > 0$, it follows that $a_1^2+a_2^2 < 1$, so that (a_1,a_2), in the xy-plane, is inside the unit circle there. The distance between (a_1,a_2) and the unit circle in the xy-plane is $1-(a_1^2+a_2^2) > 0$, so choose $\delta = \frac{1}{2}\left(1-(a_1^2+a_2^2)\right)$. Then the neighborhood $U(a_1,a_2) = B_\delta(a_1,a_2)$ lies completely inside the unit circle in the xy-plane (See Figure 28). Take $V \in S_1$, where $V = f(U)$, and the theorem holds.

EXAMPLE 8.2. Let $G(x,y,z) = 2xy^2+xyz-2z^2$, and $\mathbf{a} = (2,-3,3)$. Can we write $z = f(x,y)$ near $\mathbf{a}$? In essense, this is a question of when it is possible to "solve" for z in terms of x and y. In practice, this theorem and idea provides the ability to solve for one variable in terms for the others even in the case where algebraically, it is extremely difficult or not possible.

For G, we can answer this question quickly: Since

$$G(\mathbf{a}) = G_z(2,-3,3) = (xy-4z)\Big|_{(2,-3,3)} = (2)(-3)-4(3) = -18 \neq 0,$$

the answer is yes!. Basically, since the z-component of the gradient is not 0 at $\mathbf{a}$, it will remain not 0 at all points *near* $\mathbf{a}$. Thus the gradient vector will not be horizontal near $\mathbf{a}$ and the tangent planes to the level-sets of G containing the nearby points will still not be vertical.

To continue with this example, at the point $\mathbf{b} = (0,4,0)$, we have

$$\nabla G(\mathbf{b}) = \begin{bmatrix} G_x(\mathbf{b}) \\ G_y(\mathbf{b}) \\ G_z(\mathbf{b}) \end{bmatrix} = \begin{bmatrix} 2y^2+yz \\ 4xy+xz \\ xy-4z \end{bmatrix}\Bigg|_{(0,4,0)} = \begin{bmatrix} 32 \\ 0 \\ 0 \end{bmatrix}.$$

Due to this, we cannot write z as a function of x and y, near $\mathbf{b}$. We also cannot write y as a function of x and z there. However, we can find a function (at least in theory) so that $x = g(y,z)$, near $\mathbf{b}$.

We can directly calculate the tangent plane to the level set of G near the points $\mathbf{a}$ and $\mathbf{a}$, again using the gradient, in any case that the gradient has at least one component that is not 0. Here,

$$\nabla G(\mathbf{a}) = \begin{bmatrix} G_x(\mathbf{a}) \\ G_y(\mathbf{a}) \\ G_z(\mathbf{a}) \end{bmatrix} = \begin{bmatrix} 2y^2+yz \\ 4xy+xz \\ xy-4z \end{bmatrix}\Bigg|_{(2,-3,3)} = \begin{bmatrix} 9 \\ -18 \\ -18 \end{bmatrix}.$$

Then the equation of the tangent plane is

$$\nabla G(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}) = 0$$
$$9(x-2) - 18(y+3) - 18(z-3) = 0$$
$$z = -1 + \frac{1}{2}x - y.$$

Note that $\mathbf{a}$ is actually inside this tangent plane.

For $\mathbf{b} = (0, 4, 0)$, we have $\nabla G(\mathbf{b}) = \begin{bmatrix} 32 \\ 0 \\ 0 \end{bmatrix}$, so

$$\nabla G(\mathbf{b}) \cdot (\mathbf{x} - \mathbf{b}) = 0 = 32(x-0) + 0(y-4) + 0(z-0) = 32x.$$

But this is simply the plane defined by the equation $x = 0$, or the yz-plane in $\mathbb{R}^3$. Again, note that $\mathbf{b}$ is inside the yz-plane.

Note: There is a general version of the Implicit Function Theorem for vector-valued functions, but for now, we will move on to a related idea:

8.2. The Inverse Function Theorem.

Here is a question: Let $y = f(x) = e^x$. Does $f(x)$ have an inverse? This question is really an existence question. One could answer it by actually constructing an inverse function. One can also answer it by appealing to the fact that e^x is injective, and noting that injective functions do have inverses. Specifically, the function here is $f : \mathbb{R} \to \mathbb{R}$, $y = f(x) = e^x$, but

$$\mathbf{image}(f) = \mathbb{R}_+ = \left\{ x \in \mathbb{R} \;\middle|\; x > 0 \right\}.$$

Hence only if we restrict the codomain of f to $\mathbb{R}_+$, can we actually construct the inverse: For $f : \mathbb{R} \to \mathbb{R}_+$, $f(x) = e^x$, construct $g : \mathbb{R}_+ \to \mathbb{R}$, $g(x) = \ln x$. Then one can show that $(f \circ g)(x) = x$ on $\mathbb{R}_+$ and $(g \circ f)(x) = x$ on $\mathbb{R}$.

In practice, at times, one would show that an inverse exists by simply taking $y = f(x)$, and attempting to "solve for x". Or one could graph the function and look to see that it satisfies the "horizontal line test", a graphical tool for establishing injectivity, since if a function satisfies the horizontal line test, then its inverse will satisfy the vertical line test, thus verifying that the inverse is actually a function. Without these tools, sometimes it is necessary to know if a function has an inverse even if the expression is not necessary. For example, does $y = x^2 + 5\cos x - e^x$ have an inverse? Does it have one on $[0, 1]$? Without other aids, graph this function to see.

In vector calculus, these questions become much more complicated (try graphing a non-linear function from three space to three space), even as the ideas behind them are precisely the same. Suppose $\mathbf{f} : X \subset \mathbb{R}^n \to \mathbb{R}^n$, $\mathbf{a} \in X \subset \mathbb{R}^n$ open, and $\mathbf{f} \in C^1$. If $\det D\mathbf{f}(\mathbf{a}) \neq 0$, then $\exists U \subset X$, and open neighborhood, where (1) $\mathbf{f}\big|_U$ is 1-1, (2) $\mathbf{f}(U) = V$ is open in $\mathbb{R}^n$, and (3) a

uniquely defined inverse function $\mathbf{g} : V \to U$, $g \in C^1$, where

$$(\mathbf{g} \circ \mathbf{f})(\mathbf{x}) = \mathbf{x} \quad \text{and} \quad (\mathbf{f} \circ \mathbf{g})(\mathbf{x}) = \mathbf{x}.$$

We say $\mathbf{f}$ and $\mathbf{g}$ are inverses of each other, and write $\mathbf{f}^{-1} = \mathbf{g}$ and $\mathbf{g}^{-1} = \mathbf{f}$.

Notes:

- Given $\mathbf{f} : X \subset \mathbb{R}^n \to \mathbb{R}^n$, then $\mathbf{f}(\mathbf{x}) = \begin{bmatrix} f_1(x_1, \ldots, x_n) \\ \vdots \\ f_n(x_1, \ldots, x_n) \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} =$ $\mathbf{y}$ is a system of n (nonlinear) equations, each writing a dependent variable y_i as a function of the n independent variables $x_1, \ldots, x_n$. Can we solve this system for the x-variables, writing each of them as a function of the variable $y_1, \ldots, y_n$? In essence, can we rewrite the system as
$$\mathbf{x} = \mathbf{g}(\mathbf{y}),$$
thereby finding the inverse function, where $\mathbf{g} = \mathbf{f}^{-1}$, at least locally to a point $\mathbf{a}$? The answer to this question is yes, but only if $\det D\mathbf{f}(\mathbf{a}) \neq 0$.
- If $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$ is linear, then $\mathbf{f}(\mathbf{x}) = A_{n\times n}\mathbf{x} = \mathbf{y}$, for some square matrix A. The question is: Is it possible to find a new matrix A^{-1} so that $\mathbf{x} = A^{-1}\mathbf{y}$? Again, the answer is yes, but only if $\det A \neq 0$.

The Inverse Function Theorem is simply the nonlinear (local) version of this!

EXAMPLE 8.3. Is it possible to solve $u = xy$, $v = x - y$ for x and y as functions of u and v near the point $\mathbf{a} = (1, 1)$ in the plane? How about near the point $\mathbf{b} = (-1, 1)$? Answer these questions, and where one can invert the system, do so.

To set up this problem, let $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$, where $\mathbf{f}(x, y) = (xy, x - y)$. The strategy here will be to calculate the derivative of $\mathbf{f}$, evaluated at $\mathbf{a}$ and $\mathbf{b}$, and see if its determinant is non-zero. Where it is non-zero, invert the system.

Here,

$$D\mathbf{f}(\mathbf{x}) = \begin{bmatrix} y & x \\ 1 & -1 \end{bmatrix}, \quad \text{so} \quad D\mathbf{f}(\mathbf{a}) = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad \text{and} \quad D\mathbf{f}(\mathbf{b}) = \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}.$$

It is easy to see that $\det D\mathbf{f}(\mathbf{a}) = -1 \neq 0$, while $\det D\mathbf{f}(\mathbf{b}) = 0$. Hence the system is invertible near $\mathbf{a}$ but not near $\mathbf{b}$.

To invert the system, write $v = x - y$ as $x = v + y$, and then

$$u = xy = (v + y)y = vy + y^2.$$

Solving this for y, we get $y = \frac{-v \pm \sqrt{v^2+4u}}{2}$. But without knowing which sign to choose, this is not yet a function. We then note here that when $(x, y) = (1, 1)$, then $(u, v) = (1, 0)$. Hence the plus sign in the y expression is

the one that is compatible to this, since when $u = 1$ and $v = 0$, y must equal 1. Hence we get the system

$$x = v + \frac{-v + \sqrt{v^2 + 4u}}{2} = \frac{v + \sqrt{v^2 + 4u}}{2}$$
$$y = \frac{-v + \sqrt{v^2 + 4u}}{2}.$$

Finally, note that at **b**, $x = -1$ and $y = 1$. This makes $u = -1$ and $v = -2$. Now, can you see why x and y cannot be functions of u and v near **b**?

LECTURE 9

Curves in Euclidean Space

SYNOPSIS. Today we begin the study of Chapter 3 on vector-valued functions. For the most part, there are only two topics of discussion here: paths or curves and vector fields, respectively defined as functions from the real line into n-space, or functions from n-space into itself. The reason for an entire chapter on these two items is that they play a huge role in a solid general understanding of all of the calculus of vector-valued functions of more than one variable. They also introduce the idea of a geometric object begin completely defined by a function, allowing us to fold geometry into the analysis of functions in a fundamental way. This is one of the core principles of higher mathematics. Today, curves in n-space and some of their properties. One defining characteristic of a curve in n-space is that its length should be independent of its parameterization, even though we calculate the length using the parameterization. This extra document details why this is so:

HELPFUL DOCUMENTS. PDF: `ParameterizationIndependence`.

Curves in $\mathbb{R}^n$. We start with a definition:

DEFINITION 9.1. A *curve* or *path* in $\mathbb{R}^n$ is a continuous function $\mathbf{x} : \mathcal{I} \subset \mathbb{R} \to \mathbb{R}^n$, where $\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix}$, defined on an interval $\mathcal{I}$.

Note that the image of $\mathbf{x}(t)$ is an n-vector for each value of $t \in \mathcal{I}$. If $\mathbf{x}$ is differentiable as a function, then its derivative is also an n-vector, and

$$\frac{d}{dt}\mathbf{x}(t) = \mathbf{x}'(t) = \begin{bmatrix} \frac{dx_1}{dt}(t) \\ \vdots \\ \frac{dx_n}{dt}(t) \end{bmatrix} = \begin{bmatrix} x_1'(t) \\ \vdots \\ x_n'(t) \end{bmatrix}.$$

We sometimes call the derivative vector the *velocity*, and denote it $\mathbf{v}(t) = \mathbf{x}'(t)$.

At a point $t_0 \in \mathcal{I}$, $\mathbf{v}(t_0)$ is represented by a vector in $\mathbb{R}^n$, although not one based at the origin, as usual. Rather, the vector $\mathbf{v}(t_0)$ is based at $\mathbf{x}(t_0)$ and tangent to the curve $\mathbf{image}(\mathbf{x})$. See Figure 29 below left. Here, we simply denote the entire curve as $\mathbf{x}$. Now, as long as $\mathbf{v}(t_0) \neq \mathbf{0}$, this vector

defines a unique tangent line to $\mathbf{x}$ at $t = t_0$, parameterized as

$$\ell(s) = \mathbf{x}(t_0) + s\mathbf{v}(t_0), \quad \text{or}$$
$$\ell(s) = \mathbf{x}(t_0) + (t - t_0)\mathbf{v}(t_0), \quad \text{for} \quad s = t - t_0.$$

Note that the line, shown in Figure 29 at right, $\ell = \mathbf{span}\{\mathbf{v}(t_0)\}$.

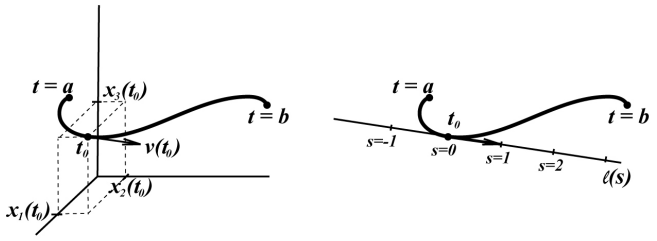


FIGURE 29. A parameterized curve in $\mathbb{R}^3$ and its tangent line.

Here, the *speed* of $\mathbf{x}(t)$ at $t = t_0$ is simply the size of the velocity vector at t_0, so $\|\mathbf{v}(t_0)\|$. The interpretation is of a bead moving along a piece of wire that is the curve. The bead is at $\mathbf{x}(t_0)$ at time $t = t_0$ and moving with (instantaneous) speed $\|\mathbf{v}(t_0)\|$ then. All of this is a topic of a standard single variable calculus course, since all of the derivatives here are calculated according to the component functions $x_i : I \to \mathbb{R}$, each of which is a real-valued on $\mathcal{I} \subset \mathbb{R}$.

Indeed, let $x = f(t)$ and $y = g(t)$, for $t \in \mathcal{I} \subset \mathbb{R}$, define a parametric curve in $\mathbb{R}^2$. If $f, g \in C^1$, then $\frac{dx}{dt} = f'(t)$ and $\frac{dy}{dt} = g'(t)$, and when defined,

$$\frac{dy}{dx} = \frac{\frac{dy}{dt}}{\frac{dx}{dt}}$$

defines the tangent line in $\mathbb{R}^2$ to the curve at (x_0, y_0):

$$y = \left(\left. \frac{dy}{dx} \right|_{(x_0,y_0)} \right)(x - x_0) + y_0, \quad x_0 = f(t_0), \text{ and } y_0 = g(t_0).$$

This construction was useful for studying curves that are defined only implicitly and not representable as functions $y(x)$ or $x(y)$: If a curve is defined as $F(x, y) = 0$, then we can calculate $\frac{dy}{dx}$ in two ways: (1) implicitly, or (2) via a parameterization like above. But we can use the language of vector calculus, now, to revisit these methods:

Implicit differentiation. Assume that $y = y(x)$ is an implicit function of x. The the equation $F(x, y) = 0$ looks like $F(x, y(x)) = 0$, and is only a function of x. Thus we can differentiate with respect to x and get

$$\frac{d}{dx}F(x, y(x)) = \frac{\partial}{\partial x}F(x, y) + \frac{\partial}{\partial y}F(x, y)\frac{dy}{dx} = 0.$$

Thus, we get

$$\frac{dy}{dx} = -\frac{F_x(x,y)}{F_y(x,y)}.$$

Via parameterization. Both $x = x(t)$ and $y = y(t)$ are functions of t, so $F(x,y) = F(x(t),y(t)) = 0$, adn F is only a function of t. Thus

$$\frac{d}{dt}F(x(t),y(t)) = F_x(x,y)\frac{dx}{dt} + F_y(x,y)\frac{dy}{dt} = 0.$$

This implies again the SAME result:

$$\frac{dy}{dx} = \frac{\frac{dy}{dt}}{\frac{dx}{dt}} = -\frac{F_x(x,y)}{F_y(x,y)}.$$

Thinking of a curve as a function affords us all of the tools of calculus to study the geometry of curves:

(1) We can attribute higher derivatives to geometric features like acceleration,
$$\mathbf{a}(t) = \frac{d}{dt}\mathbf{v}(t) = \frac{d^2}{dt^2}\mathbf{x}(t),$$
and jerk, etc.
(2) We can recover quantities like distance via integrating velocity, so that
$$\mathbf{x}(t) = \int_{t_0}^{t} \mathbf{v}(s)\, ds.$$
Do keep in mind, though, that integrating a vector means integrating each component, noting that the constant of integration is, again, a vector.
(3) Derivative rules, again, behave well with respect to curves. For example, the Product Rule and the Dot Product:
$$\frac{d}{dt}\left[\mathbf{x}\cdot\mathbf{y}(t)\right] = \frac{d\mathbf{x}}{dt}\cdot\mathbf{y}(t) + \mathbf{x}(t)\cdot\frac{d\mathbf{y}}{dt}.$$
(4) Facilitates geometric study:

EXAMPLE 9.1. If $\mathbf{x}(t) \subset \mathbb{R}^n$ is a C^1-curve, with $||\mathbf{x}(t)|| = c > 0$, for all $t \in I$, then $\mathbf{x}'(t)\cdot\mathbf{x}(t) = 0$, for every $t \in I$.

EXERCISE 3. Prove this result.

Recall from Calculua II, for $f : [\alpha,\beta] \to \mathbb{R}$, the length of $\mathbf{graph(f)} \subset \mathbb{R}^2$ on $[\alpha,\beta]$ is

$$L = \int_{\alpha}^{\beta} \sqrt{1 + (f(x))^2}\, dx,$$

or if the curve is a parametric curve,

$$L = \int_\alpha^\beta \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt.$$

EXERCISE 4. Show that these two quantities are the same.

This last formula is tailor-made for us: Let $\mathbf{x} : [a,b] \to \mathbb{R}^2$, $\mathbf{x}(t) = \begin{bmatrix} x(t) \\ y(t) \end{bmatrix}$. Here, given a partition on the interval $[a,b]$,

$$a = t_0 < t_1 < \ldots < t_{n-1} < t_n = b,$$

one looks for the approximate length of the curve on the subinterval $[t_{i-1}, t_i]$, and then adds up the approximations on each subinterval to get an approximation of the length of the curve. For each subinterval $[t_{i-1}, t_i]$, calculate Δt_i. Now approximate the length of the curve on a subinterval by using Euclidean distance between $\mathbf{x}(t_{i-1}$ and $\mathbf{x}(t_i)$. The approximate length of the curve in the ith subinterval is

$$\|\mathbf{x}(t_i) - \mathbf{x}(t_{i-1})\| = \sqrt{(x(t_i) - x(t_{i-1}))^2 + (y(t_i) - y(t_{i-1}))^2}.$$

But we can write

$$x(t_i) - x(t_{i-1}) = \Delta x_i = x'(t_i^*)\Delta t_i$$
$$y(t_i) - y(t_{i-1}) = \Delta y_i = y'(t_i^{**})\Delta t_i$$

by the Mean Value Theorem for some t_i^* and t_i^{**} in $[t_{i-1}, t_i]$. So the approximate length of the curve, given the partition, is

$$\text{approx } L = \sum_{i=1}^n \sqrt{\left(x'(t_i^*)\right)^2 + \left(y'(t_i^{**})\right)^2}\ \Delta t_i.$$

And the actual length is found by taking the limit as the largest $\Delta t_i \to 0$:

$$\begin{aligned} L &= \lim_{\max_i \Delta t_i \to 0} \sum_{i=1}^n \sqrt{\left(x'(t_i^*)\right)^2 + \left(y'(t_i^{**})\right)^2}\ \Delta t_i \\ &= \int_a^b \sqrt{(x'(t))^2 + (y'(t))^2}\ dt \\ &= \int_a^b \left\|\mathbf{x}'(t)\right\|\ dt, \end{aligned}$$

where the quantity $\|\mathbf{x}'(t)\|$ is the size of the velocity vector at time t, otherwise known as the *speed* of the curve at t. All of this works in $\mathbb{R}^n$, $n \in \mathbb{N}$.

DEFINITION 9.2. The length of $\mathbf{x} : [a,b] \to \mathbb{R}^n$, a C^1-parameterized curve in $\mathbb{R}^n$ is

$$L(\mathbf{x}) = \int_a^b \left\|\mathbf{x}'(t)\right\|\ dt.$$

Some notes:

- One integrates speed to recover distance (length traveled)!

- Even if the curve is only piecewise C^1 (so maybe it has corners), this still works, as integrals are additive.
- This formula seems to critically depend on the parameterization. But it does not! (See the Helpful document for a proof.) TO verify, reparameterize and reintegrate. or better yet, parameterize *intrinsically*, using length itself as the parameter on the curve.

Let $\mathbf{x} : [a, b] \to \mathbb{R}^n$ be a path with non-zero velocity everywhere (so that $\mathbf{v}(t) \neq \mathbf{0}$, $\forall t \in [a, b]$). Denote by $\mathbf{p}_0 = \mathbf{x}(a)$, and $\mathbf{p} = \mathbf{x}(s)$, where

$$s(t) = \int_a^t \left\|\mathbf{x}'(\tau)\right\| \, d\tau.$$

(Note that the use of τ is simply a dummy variable in a definite integral, and never actually appears on the curve.) Some things to think about:

- Since $\mathbf{x}'(t) \neq \mathbf{0}$, the length is always positive, and $s(t)$ is a strictly increasing function. As such, it is invertible, and we can reparameterize $\mathbf{x}(t)$ to
$$\mathbf{x}(s) = \mathbf{x}\left(t(s)\right)$$
as a function of s.
- In practice, $t(s)$ may be difficult of near impossible to find, but the total length of the curve is
$$s(b) = \int_a^t \left\|\mathbf{x}'(\tau)\right\| \, d\tau = \int_a^t \left\|\mathbf{x}'(t)\right\| \, dt,$$
which is just the length of the curve in the t parameter. Hence reparameterization does not change length.
- $s(t)$ is c^1 when $\mathbf{x}$ is, and
$$s'(t) = \frac{ds}{dt} = \frac{d}{dt}\left[\int_a^t \left\|\mathbf{x}'(\tau)\right\| \, d\tau\right] = \left\|\mathbf{x}'(t)\right\|.$$
So under this reparameterization, the derivative is just the spread of the curve at the old value of t.

So we can use this to calculate the tangent vector in the new parameter: Write $\mathbf{x}(t) = \mathbf{x}(s(t))$. The differentiate, using the Chain Rule:

$$\mathbf{x}'(t) = \frac{d}{dt}\mathbf{x}(s(t)) = \mathbf{x}'(s) \cdot s'(t) = \mathbf{x}(s)\left\|\mathbf{x}'(t)\right\|, \quad \text{so} \quad \mathbf{x}'(s) = \frac{\mathbf{x}'(t)}{\|\mathbf{x}'(t)\|}.$$

Conclusions?

(1) In the new parameter, the arclength traverses the curve at unit speed always.
(2) $\mathbf{x}'(s)$ is just the *normalization* of the tangent vector at the same point as $\mathbf{x}(t)$.

DEFINITION 9.3. For a C^1-path $\mathbf{x} : [a, b] \to \mathbb{R}^n$, the *unit tangent vector* to $\mathbf{x}$ at $t = t_0$ is

$$\mathbf{T}(t_0) = \frac{\mathbf{x}'(t_0)}{\|\mathbf{x}'(t_0)\|},$$

and is just the normalized velocity.

This concept of a normalized velocity vector will be very important later on in the course.

LECTURE 10

Vector Fields

SYNOPSIS. Vector fields, as geometric objects and/or functions, provide a backbone in which all of physics and engineering, really mathematical modeling is structured on. From force fields in physics to slope fields in differential equations and modeling, the notion of a vector field allows us to recover measureable quantities from models defined only by equations of motion. Here, we begin the study of their basic structure and properties.

Vector Fields. We start with a definition:

DEFINITION 10.1. A *vector field* on $\mathbb{R}^n$ is a map $\mathbf{F} : X \subset \mathbb{R}^n \to \mathbb{R}^n$, as assignment of a vector $\mathbf{F}(\mathbf{x})$ to every point $\mathbf{x} \in X$.

Examples of vector fields include

- Force fields in physics,
- slope fields in differential equations, and
- fluid (air) flow in climate models.

A vector field is of class C^n precisely when $\mathbf{F}$ is C^n. This means that vectors vary in both size and direction in a continuous (C^0), or differentiable (C^n, $n \geq 1$), etc.

DEFINITION 10.2. A vector field is called a *gradient field* on $\mathbb{R}^n$ if $\mathbf{F}$ is the gradient of a real-valued function $f : X \subset \mathbb{R}^n \to \mathbb{R}$.

Let $f : X \subset \mathbb{R}^n \to \mathbb{R}$ be C^1. Then $\nabla f : X \subset \mathbb{R}^n \to \mathbb{R}^n$. Here, we interpret this as a vector field on X, a gradient field on X.

- Here, f is called a *potential function* for the gradient field $\mathbf{F}(\mathbf{x}) = \nabla f(\mathbf{x})$.
- Recall for $f : \mathbb{R}^2 \to \mathbb{R}$, the level sets of f are generically curves in the domain of f, which is the plane. For f a potential function of a gradient field,
  (1) the level curves are *equipotential sets*, sets of equal potential, and
  (2) the gradient field along these sets always is orthogonal to (the tangent lines of) these sets.
  (3) The gradient field always points in the direction of the most rapid increase of f at each point.
  (4) In contrast to a vector field, a real-valued function is sometimes called a *scalar field*. The gradient takes a potential (scalar) field to a vector field.

EXAMPLE 10.1. Given a scalar field $f : X \subset \mathbb{R}^n \to \mathbb{R}$, finding its gradient field is straightforward: Take derivatives adn form the vector. But, given a gradient field, can one find a potential for it? In Example 5 on page, 231, the author tells you that the gradient field $\mathbf{F}(x,y,z) = \begin{bmatrix} 3x^2+y^2 \\ 2xy \\ x^3-2z \end{bmatrix}$ has potential $f(x,y,z) = x^3z + xy^2 - z^2$, and "leave[s] it to you to verify...". But without a candidate for a potential, how does one calculate one? The idea is to un-differentiate!

Pretend that $f : \mathbb{R}^n \to \mathbb{R}$ is unspecified. But since we know $\mathbf{F}(x,y,z) = \nabla f(x,y,z)$, then we know the following:

(1) $\dfrac{\partial f}{\partial y} = 2xy$. Hence $f(x,y,z) = xy^2 + h(x,z)$. (Why is this? Because, the partial derivative of f with respect to why would see every function of only x and z as a constant. Hence, when un-differentiating, one has to account for this fact by specifying the constant lost to differentiating (with respect to y) as something which is a function of possibly everything except for y. Got it? So we already know something about f. Namely, f is of the form $f(x,y,z) = xy^2 + h(x,z)$.

(2) Then $\dfrac{\partial f}{\partial x} = \dfrac{\partial}{\partial x}\left[xy^2 + h(x,z)\right] = y^2 + \dfrac{\partial h}{\partial x}(x,z) = 3x^2z + y^2$. This is bacause the last expression is the x-component of the gradient field $\mathbf{F}(x,y,z)$. Hence $h(x,z) = x^3z + g(z)$, where g is some unknown function of only z. Now we know even more about f. We know $f(x,y,z) = xy^2 + x^3z + g(z)$.

(3) And lastly, $\dfrac{\partial f}{\partial z} = \dfrac{\partial}{\partial z}\left[xy^2 + x^3z + g(z)\right] = x^3 + g'(z) = x^3 - 2z$. But this means that $g'(z) = -2z$, so that $g(z) = -z^2$.

Hence we have $f(x,y,z) = xy^2 + x^3z - z^2$.

DEFINITION 10.3. A *flow line*, or a *trajectory* of a vector field $\mathbf{F} : X \subset \mathbb{R}^n \to \mathbb{R}^n$ is a differentiable curve $\mathbf{x} : I \subset \mathbb{R} \to \mathbb{R}^n$ that satisfies

$$\mathbf{x}'(t) = \mathbf{F}\left(\mathbf{x}(t)\right), \quad \forall t \in I. \tag{10.0.1}$$

Here, the velocity vector of the curve at every point in the domain of the curve equals to vector field at that point. Finding such a path, given a vector field, is precisely the subject of a field of mathematics called differential equations! But simply verifying that a given path is a flow line of a vector field is a matter of just verifying Equation 10.0.1.

EXAMPLE 10.2. Is the path $\mathbf{x}(t) = \begin{bmatrix} e^{-t} + 2e^{2t} \\ -e^{-t} + e^{2t} \end{bmatrix}$ in the plane a flow line for the vector field $\mathbf{F}(\mathbf{x}) = \begin{bmatrix} x + 2y \\ x \end{bmatrix}$ on $\mathbb{R}^2$?

The answer is yes, since

$$\mathbf{x}'(t) = \begin{bmatrix} -e^{-t} + 4e^{2t} \\ e^{-t} + 2e^{2t} \end{bmatrix} = \begin{bmatrix} \left(e^{-t} + 2e^{2t}\right) + 2\left(-e^{-t} + e^{2t}\right) \\ \left(e^{-t} + e^{2t}\right) \end{bmatrix} = \begin{bmatrix} x(t) + 2y(t) \\ x(t) \end{bmatrix} = \mathbf{F}\left(\mathbf{x}(t)\right).$$

DEFINITION 10.4. An *linear operator* is a mapping from one linear (vector) space to another.

From linear algebra, this means that any matrix determines a linear operator from the domain to the codomain. This also means that any linear transformation is also called a linear operator. However, one can define linear spaces whose elements are functions, in the following way: Two real-valued functions, defined on the same space, can be added together to create another function from the same domain to the same codomain. Once can also multiple any function by a real number to create a new function. Hence any linear combination of real-valued functions from a domain to $\mathbb{R}$ is again a function from the domain to $\mathbb{R}$. And since there exists an additive identity function (the 0-function), and an additive inverse function for every function, the set of functions from a domain to $\mathbb{R}$ form a linear space (like a vector space). However, these "function" spaces are not finite dimensional, and hence there is not a finite basis, like for the standard vector spaces one sees in linear algebra. But one can define linear maps between these function spaces, and they behave much like the linear transformations you have seen in linear algebra. So think of linear operators as maps taking functions to functions. Note that the notion of an operator being *linear* is just the idea that the image of a linear combination of inputs is just a linear combination of the images of the inputs, or

$$f(c_1 x + c_2 y) = c_1 f(x) + c_2 f(y).$$

Keep this in mind.

DEFINITION 10.5. The *del operator* ∇ is the linear operator that takes a real-valued C^1-function $f : X \subset \mathbb{R}^n \to \mathbb{R}$ to its gradient vector field $\nabla f : X \to \mathbb{R}^n$.

Some Notation: $\nabla = \mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{k}\frac{\partial}{\partial z}$ in $\mathbb{R}^3$, or

$$\nabla = \begin{bmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{bmatrix} = \mathbf{e}_1 \frac{\partial}{\partial x_1} + \mathbf{e}_2 \frac{\partial}{\partial x_2} + \ldots + \mathbf{e}_n \frac{\partial}{\partial x_n}, \quad \text{in } \mathbb{R}^n.$$

This notation may seem a bit odd, but it is common, and implies

$$\nabla(\ ) = \sum_{i=1}^{n} \mathbf{e}_i \frac{\partial}{\partial x_i}(\ ) = \begin{bmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{bmatrix}(\ )$$

is to be interpreted as

$$\nabla f = \nabla(f) = \begin{bmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{bmatrix}(f) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}.$$

DEFINITION 10.6. For a C^1-vector field $\mathbf{F} : X \subset \mathbb{R}^n \to \mathbb{R}^n$, the *divergence* of $\mathbf{F}$, denoted $\mathbf{div}\,\mathbf{F}$, or $\nabla \cdot \mathbf{F}$, is the scalar function

$$\mathbf{div}\,\mathbf{F} = \nabla \cdot \mathbf{F} = \sum_{i=1}^{n} \frac{\partial F_i}{\partial x_i} = \frac{\partial F_1}{\partial x_1} + \frac{\partial F_2}{\partial x_2} + \ldots + \frac{\partial F_n}{\partial x_n},$$

for $\mathbf{x} = (x_1, \ldots, x_n) \in X$, and $\mathbf{F}(\mathbf{x}) = \begin{bmatrix} F_1(\mathbf{x}) \\ \vdots \\ F_n(\mathbf{x}) \end{bmatrix}$.

Some notes:

- Here $\mathbf{div}\,\mathbf{F}(\mathbf{x}) = \nabla \cdot \mathbf{F}(\mathbf{x}) = \begin{bmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{bmatrix} \cdot \begin{bmatrix} F_1(\mathbf{x}) \\ F_2(\mathbf{x}) \\ \vdots \\ F_n(\mathbf{x}) \end{bmatrix}$ uses the Dot Product, although the product on each component is meant to indicate "apply the partial operator to the component function".
- We will prove this later on in the course, but the divergence of a vector field measures the infinitesimal volume change caused by the vector field.
- A vector field $\mathbf{F}$, where $\nabla \cdot \mathbf{F} = 0$ is called *incompressible*.
- Viewed as an operator, ∇ can *operate* on functions in different ways:
  (1) As the gradient of a scalar field ∇f, for $f : X \subset \mathbb{R}^n \to \mathbb{R}$;
  (2) As the divergence of a vector field $\nabla \cdot \mathbf{F}$, for $\mathbf{F} : X \subset \mathbb{R}^n \to \mathbb{R}^n$; And
  (3) as the *curl* of a vector field $\nabla \times \mathbf{F}$, but only in $\mathbb{R}^3$.

DEFINITION 10.7. For a C^1-vector field $\mathbf{F} : X \subset \mathbb{R}^3 \to \mathbb{R}^3$, the *curl* of $\mathbf{F}$, denoted $\mathbf{curl}\,\mathbf{F}$, or $\nabla \times \mathbf{F}$, is

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

More Notes:

- It is worth noticing that (1) the gradient of a scalar field is a vector field, (2) the divergence of a vector field is a scalar field, and (3) the curl of a vector field (in R^3) is a vector field.
- We will again prove this later, but the curl of a vector field measures the infinitesimal twist in the vector field along the vector field at each point.
- If, for $\mathbf{F} : X \subset \mathbb{R}^3 \to \mathbb{R}^3$, we have $\nabla \times \mathbf{F} = \mathbf{0}$ everywhere, we say $\mathbf{F}$ is *irrotational*.

EXAMPLE 10.3. The vector field $\mathbf{F}(x, y, z) = \begin{bmatrix} y \\ -x \\ 0 \end{bmatrix}$ rotates each xy-plane at $z = c$. Here

$$\nabla \times \mathbf{F} = \begin{bmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{bmatrix} \times \begin{bmatrix} y \\ -x \\ 0 \end{bmatrix} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ y & -x & 0 \end{vmatrix}$$
$$= \left(\frac{\partial}{\partial y}(0) - \frac{\partial}{\partial z}(-x)\right)\mathbf{i} - \left(\frac{\partial}{\partial x}(0) - \frac{\partial}{\partial z}(y)\right)\mathbf{j} + \left(\frac{\partial}{\partial x}(-x) - \frac{\partial}{\partial y}(y)\right)\mathbf{k} = -2\mathbf{k}.$$

Notice, by the definition and properties of the cross product, that, as a vector field $\nabla \times \mathbf{F}$ must be orthogonal to $\mathbf{F}$ at every point.

EXAMPLE 10.4. Explosions and Implosions in $\mathbb{R}^3$: For $\mathbf{G}(x, y, z) = \pm c \begin{bmatrix} x \\ y \\ z \end{bmatrix}$, for $c \in \mathbb{R}$, we have $\nabla \times \mathbf{G} = \mathbf{0}$. These vector fields are irrotational.

EXERCISE 5. Show all constant vector fields $\mathbf{H}(x, y, z) = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}$ are irrotational.

And here are some beautiful facts, whose calculations provide excellent practice:

EXERCISE 6. Show that a gradient vector field in $\mathbb{R}^3$ is irrotational. That is, for $f : X \subset \mathbb{R}^3 \to \mathbb{R}$ a C^2-function, show that $\nabla \times (\nabla f) = \mathbf{0}$.

EXERCISE 7. Show that the curl of a vector field in $\mathbb{R}^3$ is incompressible. That is, for $F : X \subset \mathbb{R}^3 \to \mathbb{R}^3$ a C^2-vector field, show that $\nabla \cdot (\nabla \times \mathbf{F}) = 0$.

LECTURE 11

Differentials and Taylor Series

SYNOPSIS. Herein, we give a brief interpretation of the differential of a function. There are many interpretations of a function's differential, but we only deal with one currently. Then we delve into an even briefer description of the Taylor Series of a real-valued function on $\mathbb{R}^n$. The details, for now, are not important. But the relationship to the counterparts of both of these concepts to single variable calculus is quite important.

HELPFUL DOCUMENTS. Mathematica: `TaylorPolynomials`.

The differential of a function. Recall that for a variable x, a small change in x is denoted $\Delta x = (x+h) - x = h$, where h is a number near 0. As the value of h tends to 0, Δx also vanishes. But we can mark the vanishing of Δx via what is called an *infinitesimal change* in x, and denote it dx, so that

$$\Delta x \xrightarrow{h\to 0} dx.$$

Really, this has meaning almost exclusively in the context of how other quantities change that depend on x or when compared to x. The quantity dx is called the *differential* of x.

Now let $f : X \subset \mathbb{R} \to \mathbb{R}$ be a differentiable function, and $a \in X$. For the graph $y = f(x)$, the quantity

$$\Delta y = \Delta f = f(x + \Delta x) - f(x)$$

represents a small change in y, as it depends on Δx, the small change in x. As $h \to 0$, of course, Δf also goes to 0. But measuring *how* Δf goes to zero is important in calculus. Hence we mark teh infinitesimal change in y or f by its differential: $df = dy$. Studying just how the dependent variable y is changing as one varies x is vitally important in the study of functional relationships between entities, and is the motivation behind the Liebniz notation in calculus $\frac{dy}{dx} = \frac{df}{dx} = \frac{d}{dx}f(x)$ representing the derivative of $f(x)$ with respect to x:

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

since $\Delta x = (x + h) - x = h$. To even be able to discuss ideas that involve passing to a limit, one needs to be able to discuss quantities that are *infinitesimally* close to 0 or close to each other. One can say that an infinitesimally small positive number represents a positive number closer to 0 than any real positive number.

We note here that, as an alternate definition, one can call the quantities dx, and dy actual new variables, whose relationship is tied to the relationship between y and x, namely $y = f(x)$. This alternate definition provides a much more concrete foundation for which to use these quantities, but structurally does not change their meaning. We will visit this more concrete notion of a differential later when we discuss differential forms.

The quantity df (the differential of f), represents an infinitesimal change in f given an infinitesimal change in its independent variable x: at $x = a$, we have

$$df(a) = f'(a)dx, \quad \text{or} \quad \frac{df}{dx}(a) = f'(a)$$

to reflect the idea that this differential will change as we vary the point $x = a$. More generally, $df = f'(x)\,dx$.

Some notes:

- This will make more sense later, when we discuss differential forms, but df, the differential of f, is an example of a *differential* 1*-form*.
- This concept embodies the Substitution Rule (the Anti-Chain Rule) in single variable calculus:

$$\int_a^b f\left(g(x)\right) g'(x)\,dx \xrightarrow[du=g'(x)\,dx]{u=g(x)} \int_{g(a)}^{g(b)} f(u)\,du.$$

Indeed, let f be a function of u, with $u = \alpha$ a point in its domain$f$, and F an antiderivative of f, so that $F'(u) = f(u)$. Then

$$dF(\alpha) = F'(\alpha)\,du = f(\alpha)\,du = \left(f(u)\Big|_{u=\alpha}\right) du.$$

If $u = g(x)$ is a function of x, then du, the differential of u, is related to dx, and $du = g'(x)\,dx$. But also, f and hence F are functions of x, via composition: $f(u) = f\left(g(x)\right)$ and $F(u) = F\left(g(x)\right)$. Thus, their differentials also vary with respect to x,

$$\begin{aligned} f(\alpha)\,du = dF(\alpha) = F'(\alpha)\,du &= \left(F'(u)\Big|_{u=\alpha}\right) du \\ &= \left(F'(u)\Big|_{u=g(a)=\alpha}\right)\left(g'(x)\Big|_{x=a}\right) dx \\ &= \left(f(u)\Big|_{u=g(a)}\right)\left(g'(x)\Big|_{x=a}\right) dx \\ &= f'\left(g(a)\right) g'(a)\,dx. \end{aligned}$$

Hence, we are left with, on their appropriate domains,

$$f(u)\,du = f'\left(g(x)\right) g'(x)\,dx.$$

Finally, recall that integration is just a form of infinitesimal addition. Now does the form of the Substitution Rule make more sense?

In many variables, let $f : X \subset \mathbb{R}^n \to \mathbb{R}$ be a differentiable function, and $\mathbf{a} \in X$. df is the sum of the *partial differentials* (differentials in the coordinate directions), $\frac{\partial f}{\partial x_i}\, dx_i$, and

$$df = \frac{\partial f}{\partial x_1}\, dx_1 + \ldots + \frac{\partial f}{\partial x_n}\, dx_n = \sum_{n=1}^{n} \frac{\partial f}{\partial x_i}\, dx_i = Df(\mathbf{a})d\mathbf{x} \tag{11.0.1}$$

represents the *total differential* of f. This quantity represents an infinitesimal change in f in terms of the infinitesimal changes in its coordinate directions dx_i. The use of the vector term $d\mathbf{x} = \begin{bmatrix} dx_1 \\ \vdots \\ dx_n \end{bmatrix}$ will make more sense later on in the course.

As a function, $\Delta f = f(\mathbf{a} + \Delta\mathbf{x}) - f(\mathbf{a}) = f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a})$, where $\Delta\mathbf{x} = \mathbf{h}$ is a vector of small changes in each of the coordinate directions. Written out, Δf will contain many terms which are not linear in $\Delta\mathbf{x}$. As $\Delta\mathbf{x}$ tends to 0, any terms which contain products of the various small changes in the coordinate directions will tend toward 0 much faster, so that only the linear parts of these terms will contribute to the limit (the higher-degree terms will die off quickly, leaving only the linear terms). One can then see directly how the differential of a function operates:

EXAMPLE 11.1. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by $f(x, y) = x^2 + xy - x - y + \sin x$. Here $\Delta\mathbf{x} = (\Delta x, \Delta y))^T$, and

$$\begin{aligned}
\Delta f(\pi, 0) &= f\left((\pi, 0)^T + (\Delta x, \Delta y)^T\right) - f(\pi, 0) \\
&= (\pi + \Delta x)^2 + (\pi + \Delta x)(\Delta y) - (\pi + \Delta x) - \Delta y + \sin(\pi + \Delta x) - \pi^2 + \pi \\
&= \pi^2 + 2\pi\Delta x + (\Delta x)^2 + \pi\Delta y + \Delta x\Delta y - \pi - \Delta x - \Delta y - \sin(\Delta x) - \pi^2 + \pi.
\end{aligned}$$

Notice here that all of the terms not containing a Δx or a Δy cancel out. Recall also that for very small values of Δx, the function $\sin(\Delta x) \approx \Delta x$. This is called a first-order approximation of the sine function near $x = 0$, and reflects the idea that the sine function has a Taylor expansion at $x = 0$ containing a linear term with coefficient 1 (its first Taylor polynomial is $T_1(x) = x$). Likewise, for very small values of Δx and Δy, all of the other higher-order terms vanish double fast, leaving only the linear terms:

$$\Delta f(\pi, 0) = (2\pi - 1)\Delta x - \Delta x + (\pi - 1)\Delta y = (2\pi - 2)\Delta x + (\pi - 1)\Delta y.$$

Passing to the infinitesimals, we get $\Delta f \longrightarrow df$, and $\Delta\mathbf{x} \longrightarrow d\mathbf{x} = \begin{bmatrix} dx \\ dy \end{bmatrix}$, and we get

$$df(\pi, 0) = (2\pi - 2)\, dx + (\pi - 1)\, dy.$$

Now compare this to the direct computation, using Equation 11.0.1 above. Here

$$\frac{\partial f}{\partial x}(\pi,0) = (2x - y - 1 + \cos x)\Big|_{\substack{x=\pi \\ y=0}} = (2\pi - 2)$$

and

$$\frac{\partial f}{\partial y}(\pi,0) = (x-1)\Big|_{\substack{x=\pi \\ y=0}} = (\pi - 1)$$

so that

$$df(\pi,0) = Df(\pi,0)\, d\mathbf{x} = \begin{bmatrix} 2\pi - 2 & \pi - 1 \end{bmatrix} \begin{bmatrix} dx \\ dy \end{bmatrix} = (2\pi - 2)\, dx + (\pi - 1)\, dy.$$

The result is the same.

And finally, going back to the notion of dx and dy being actual coordinates, tied together via $y = f(x)$ so that $dy = f'(x)\, dx$, we can extend this notion to the multidimensional case. Here, we can view each of the dx_i as an actual coordinate on the local linearization of f at the point $\mathbf{x} = \mathbf{a}$. Thus, the set $\{dx_1, \ldots, dx_n\}$ become a set of coordinates on each tangent space to the domain $X \subset \mathbb{R}^n$, and $df(\mathbf{a})$ becomes a vector measuring just how f is changing infinitesimally. Again,we will explore this notion in detail at the end of the course.

The Taylor series. Recall that the Taylor series of a C^∞-function $f : I \subset \mathbb{R} \to \mathbb{R}$ at a point $a \in I$ is

$$f(x) = \sum_{i=0}^{\infty} \frac{f^{(i)}(a)}{i\,!}(x-a)^i,$$

and is defined on the largest interval where the series converges. Here, one may also truncate this series to obtain the mth Taylor polynomial

$$T_m(x) = \sum_{i=0}^{m} \frac{f^{(i)}(a)}{i\,!}(x-a)^i.$$

The mth Taylor polynomial is considered the "best" mth-degree polynomial that approximates $f(x)$ near $x = a$, and we define the term "best" to mean that all of the derivatives of f and T_m are the same up to and including the mth derivative. So, for $i = 0, 1, \ldots, m$,

$$\frac{d^i}{dx^i} T_m(a) = f^{(i)}(a).$$

Now let $g : X \subset \mathbb{R}^n \to \mathbb{R}$ also be C^∞. We may ask very similar questions, like: What is the best mth degree polynomial (in the variables defining X) that approximates g near $\mathbf{x} = \mathbf{a}$. Again, the criteria for "best" will be the one that matches g at $\mathbf{a}$ for all (partial) derivatives up to and including the order-m ones.

Obviously, the best 0-degree polynomial to approximate $g(\mathbf{x})$ at $\mathbf{x} = \mathbf{a}$ is the one whose function value is $g(\mathbf{a})$, so

$$T_0(\mathbf{x}) = g(\mathbf{a}).$$

And we have already calculated the best first-degree polynomial, where the derivative matrix of g played a vital role:

$$T_1(\mathbf{x}) = g(\mathbf{a}) + Dg(\mathbf{a})(\mathbf{x} - \mathbf{a}).$$

Perhaps a better way to write this is to appeal to the individual variables explicitly, so

$$\begin{aligned} T_1(\mathbf{x}) &= g(\mathbf{a}) + Dg(\mathbf{a})(\mathbf{x} - \mathbf{a}) \\ &= g(\mathbf{a}) + \sum_{i=1}^{n} \frac{\partial g}{\partial x_i}(\mathbf{a})(x_i - a_i) \\ &= g(\mathbf{a}) + \underbrace{g_{x_1}(\mathbf{a})(x_1 - a_1) + g_{x_2}(\mathbf{a})(x_2 - a_2) + \ldots + g_{x_n}(\mathbf{a})(x_n - a_n)}_{\text{all first partials}}. \end{aligned}$$

Now it should be straightforward to see that not only does $T_1(\mathbf{a}) = g(\mathbf{a})$, but

$$\frac{\partial T_1}{\partial x_i}(\mathbf{a}) = \frac{\partial g}{\partial x_i}(\mathbf{a}),$$

for all $i = 1, \ldots, n$.

So follow the pattern. What would be the "best" degree-2 polynomial to approximate $g(\mathbf{x})$ at $\mathbf{x} = \mathbf{a}$? Of course, one such that $T_2(\mathbf{a}) = g(\mathbf{a})$, and for all $i, j = 1, \ldots, n$, we have

$$\frac{\partial T_2}{\partial x_i}(\mathbf{a}) = \frac{\partial g}{\partial x_i}(\mathbf{a}), \quad \text{and} \quad \frac{\partial^2 T_2}{\partial x_j \partial x_i}(\mathbf{a}) = \frac{\partial^2 g}{\partial x_j \partial x_i}(\mathbf{a}).$$

It is apparent by reverse engineering that the only degree-2 polynomial that would work is

$$T_2(\mathbf{x}) = g(\mathbf{a}) + \sum_{i=1}^{n} \frac{\partial g}{\partial x_i}(\mathbf{a})(x_i - a_i) + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 g}{\partial x_j \partial x_i}(\mathbf{a})(x_i - a_i)(x_j - a_j).$$

A big question to answer here is: Why the $\frac{1}{2}$ coefficient? Think about this.

There is a much better way to view the form of T_2:

DEFINITION 11.1. Given $f : X \subset \mathbb{R}^n \to \mathbb{R}$ a C^2 function, the $n \times n$ matrix whose ijth entry is $\frac{\partial^2 f}{\partial x_j \partial x_i}$,

$$Hf = \begin{bmatrix} f_{x_1 x_1} & \cdots & f_{x_1 x_n} \\ \vdots & \ddots & \vdots \\ f_{x_n x_1} & \cdots & f_{x_n x_n} \end{bmatrix}$$

is called the *Hessian* of f.

Now, denote by $\mathbf{h} = \mathbf{x} - \mathbf{a}$, so that $h_i = x_i - a_i$. Then we can write

$$\begin{aligned} T_2(\mathbf{x}) &= g(\mathbf{a}) + \sum_{i=1}^{n} \frac{\partial g}{\partial x_i}(\mathbf{a})(x_i - a_i) + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 g}{\partial x_j \partial x_i}(\mathbf{a})(x_i - a_i)(x_j - a_j) \\ &= g(\mathbf{a}) + \begin{bmatrix} g_{x_1}(\mathbf{a}) & \cdots & g_{x_n}(\mathbf{a}) \end{bmatrix} \begin{bmatrix} h_1 \\ \vdots \\ h_n \end{bmatrix} + \frac{1}{2} \begin{bmatrix} h_1 & \cdots & h_n \end{bmatrix} \begin{bmatrix} g_{x_1 x_1}(\mathbf{a}) & \cdots & g_{x_1 x_n}(\mathbf{a}) \\ \vdots & \ddots & \vdots \\ g_{x_n x_1}(\mathbf{a}) & \cdots & g_{x_n x_n}(\mathbf{a}) \end{bmatrix} \begin{bmatrix} h_1 \\ \vdots \\ h_n \end{bmatrix} \\ &= g(\mathbf{a}) + Dg(\mathbf{a})\mathbf{h} + \frac{1}{2}\mathbf{h}^T H f \mathbf{h}. \end{aligned}$$

So what would the third Taylor polynomial look like? Generalize in the obvious fashion, and get

$$\begin{aligned} T_3(\mathbf{x}) = g(\mathbf{a}) &+ \sum_{i=1}^{n} \frac{\partial g}{\partial x_i}(\mathbf{a})(x_i - a_i) + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 g}{\partial x_j \partial x_i}(\mathbf{a})(x_i - a_i)(x_j - a_j) \\ &+ \frac{1}{6} \sum_{i,j,k=1}^{n} \frac{\partial^3 g}{\partial x_k \partial x_j \partial x_i}(\mathbf{a})(x_i - a_i)(x_j - a_j)(x_k - a_k), \end{aligned}$$

of course. And the polynomial $T_\ell(\mathbf{x})$, for natural number $\ell > 3$?

You may notice that I did not write $T_3(\mathbf{x})$ in a more elegant fashion, using some three dimensional verion of the derivative matrix or the Hessian matrix. It get difficult now since we would be creating and using objects that are higher dimensional arrays (All $8 = 2^3$ of the third-order partials of $f : \mathbb{R}^2 \to \mathbb{R}$ would be arranged into a three dimensional array. These objects do exist and are manifestations of what are called tensors. Getting a handle on the notation and working with these objects would involve a bit more time than we can devote to it at the moment. So we rely simply on the summation notation, and basically stop here.

EXERCISE 8. Devise a mathematical notation that would provide an array-based version of the third-order terms in $T_3(\mathbf{x})$.

LECTURE 12

Extrema

SYNOPSIS. Local and global extrema are much like their counterparts in single variable calculus. They are just points in the domain of a real-valued function where the function value is locally the lowest or highest. And they occur, if at all, at critical points of the function. If the function is differentiable everywhere, then extrema only occur at places where the derivative (matrix) has zeros in all of its elements. Thus all of the directional derivatives are 0 here also. But since directional derivatives are just derivatives along slices through the function, we can also check the concavity of these slice functions along vector directions in the domain. This leads to a notion of a second directional derivative, and also to one major application of the Hessian matrix of second partials. Relating this to a quadratic form, we construct the Second Derivative Test for a C^2-real-valued function of more than one variable. We then end with the multidimensional counterpart of the Extreme Value Theorem, once we understand what closed and bounded mean for a domain in real n-space.

HELPFUL DOCUMENTS. Mathematica: `Extrema`, `MoreExtrema`.

Local extrema. Like in single variable calculus, local extrema are important properties of functions:

DEFINITION 12.1. The function $f : X \subset \mathbb{R}^n \to \mathbb{R}$, for X open, has a *local minimum* at $\mathbf{a} \in X$ if there exists a neighborhood $U(\mathbf{a}) \subset X$ such that $f(\mathbf{x}) \geq f(\mathbf{a})$, for every $\mathbf{x} \in U$. And f has a *local maximum* at $\mathbf{a} \in X$ if there exists a neighborhood $U(\mathbf{a}) \subset X$ such that $f(\mathbf{x}) \leq f(\mathbf{a})$, for every $\mathbf{x} \in U$.

Some Notes:

- A local minimum (maximum) is *global* if $U = X$.
- If $f \in C^1$, then local extrema have a special quality:

  THEOREM 12.2. *Given $X \subset \mathbb{R}^n$ open and $f : X \to \mathbb{R}$ a C^1-function, if f has a local extremum at $\mathbf{a} \in X$, then $Df(\mathbf{a})$ is the zero matrix (every entry in the matrix is 0).*

- The proof of this theorem shows that a directional derivative, evaluated at $\mathbf{a}$ in this case, would see a local extremum here in every direction. In particular, in the coordinate directions. And the only matrix $A_{1\times n}$ that takes every possible n-vector to 0 is the 0-matrix.

DEFINITION 12.3. Given $f : X \subset \mathbb{R}^n \to \mathbb{R}$, with X open, a point $\mathbf{a} \in X$ is a *critical point* of f if either

(1) $Df(\mathbf{a}) = 0_{1\times n}$, or
(2) $Df(\mathbf{a})$ is undefined.

Just like in single variable calculus, extrema happen at critical points, but not all critical point need be extrema. Some examples of critical points include:

- sharp mountain ridges or roof peaks, or mountain top points, where the derivative matrix is not defined,
- smoothed over mountain tops where the derivative matrix is the zero matrix,
- saddle points,
- mesas and flood plains.

Recall that the graph of a function $f : X \subset \mathbb{R}^n \to \mathbb{R}$ is a subset of $\mathbb{R}^{n+1}$ where $X \subset \mathbb{R}^n$ is pictured as the "floor", and the height above the floor is hte value of the last variable x_{n+1}. At a place $(\mathbf{a}, f(\mathbf{a}))$, where the derivative matrix is the zero matrix, the equation of the tangent space would be

$$x_{n+1} = f(\mathbf{a}) + Df(\mathbf{a})\,(\mathbf{x} - \mathbf{a}) = f(\mathbf{a})$$

and the tangent space is parallel to the floor, or "horizontal".

However, like in Calculua I, extrema do not need to exist at all for particular functions:

EXAMPLE 12.1. Let $f(x, y) = x^2 + y^2$, on the domain $X = \mathbb{R}^2 - (0, 0)$, the plane without the origin. What would you consider the point in X where f achieves its maximum? How about it's minumum?

So how does one detect an extremum, given a critical point? Really, it is all about the structure. Some ideas:

(1) Look for extreme bahavior by simply testing functions values for points "near" the critical point. In single variable calculus, we sometimes call this the 0th Derivative Test. This method is sometimes employed when the derivative matrix is not defined at a critical point.
(2) If f is differentiable at a critical point $\mathbf{a}$, then the derivative matrix is the zero matrix there. Again, this means that every directional derivative will also be 0 at $\mathbf{a}$. Recall that directional derivatives are defined via vertical slices through the graph of the function f along directions through $\mathbf{a}$ in the domain. If one follows the curve where $\mathbf{graph}(f)$ intersects the slice, adn sees the derivative go from negative (before $\mathbf{a}$), to 0 (at $\mathbf{a}$), to positive (after $\mathbf{a}$), and this happens in every direction, then you have a local minimum at $\mathbf{a}$. Of course, one can generalize this stipulation and say that

in some or all directions, the derivative can stay 0 near $\mathbf{a}$, and we would still have a local min there. One can characterize this as a form of First Derivative Test for a critical point.

(3) Or if, within each slice, f, restricted to the slice is concave up or down and stays that way for all of the slices, then f is locally extreme at $\mathbf{a}$. (Again, all that is really necessary is that the concavity is not mixed along the slices.) If locally extreme is every direction, then locally extreme.

EXAMPLE 12.2. For the parabolic bowl, $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x,y) = x^2 + y^2$, we know that $f \in C^1$, since it is a polynomial, and

$$Df(\mathbf{x}) = \begin{bmatrix} 2x & 2y \end{bmatrix} = \begin{bmatrix} 0 & 0 \end{bmatrix}$$

only at the origin $x = y = 0$. So we have

$$D_{\mathbf{v}}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}\right) = Df(\mathbf{0})\mathbf{v} = \begin{bmatrix} 0 & 0 \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0, \quad \forall \mathbf{v} \in \mathbb{R}^2.$$

So choose a direction $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \in \mathbb{R}^2 \subset \mathbb{R}^3$ in the floor of $\mathbb{R}^3$, considered the domain of f. Then the vertical plane containing $\mathbf{v}$ is defined by the vector orthogonal to both $\mathbf{v}$ and the vector $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ (why is this true?), and is

$$\begin{bmatrix} v_1 \\ v_2 \\ 0 \end{bmatrix} \times \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ v_1 & v_2 & 0 \\ 0 & 0 & 1 \end{vmatrix} = \begin{bmatrix} v_2 \\ -v_1 \\ 0 \end{bmatrix}, \quad \text{or} \quad v_2 x - v_1 y = 0.$$

Notice that, in the domain, this is just the line $y = \frac{v_2}{v_1}x$, at least when $v_1 \neq 0$.

Now the graph of this line is the set of points

$$\left(x, \frac{v_2}{v_1}, f\left(x, \frac{v_2}{v_1}\right)\right) = \left(x, \frac{v_2}{v_1}, x^2\left(1 + \frac{v_2^2}{v_2^2}\right)\right) \in \mathbb{R}^3.$$

So let $f_{\mathbf{v}} = f\Big|_{y=\frac{v_2}{v_1}x} : \mathbb{R} \to \mathbb{R}$, where $f_{\mathbf{v}}(x) = x^2\left(1 + \frac{v_2^2}{v_1^2}\right)$. Here, of course, we have the data

$$f_{\mathbf{v}}'(0) = 2x\left(1 + \frac{v_2^2}{v_1^2}\right)\Big|_{x=0} = 0, \quad \text{and} \quad f_{\mathbf{v}}''(0) = 2\left(1 + \frac{v_2^2}{v_1^2}\right)\Big|_{x=0} > 0.$$

Hence, within the slice formed by $\mathbf{v}$, $f_{\mathbf{v}}$ is concave up, and, at least in this direction, according to the Second Derivative Test for an extremum from Calculus I, the point $\mathbf{0}$ corresponds to a local min of $f\Big|_{\mathbf{v}}$. And since this will be true of all choices of $\mathbf{v}$ (we did ntoot specify values of the entries

v_1 and v_2), we can safely conclude that $\mathbf{a} = \mathbf{0}$ is a local minimum for f. Yes, I know, it is a global minimum on any domain that contains the origin

Now for $f : X \subset \mathbb{R}^n \to \mathbb{R}$ a C^2-function, we already have access to all of its second derivative information in the form of the Hessian of f,

$$HF(\mathbf{a}) = \begin{bmatrix} f_{x_1x_1}(\mathbf{a}) & \cdots & f_{x_1x_n}(\mathbf{a}) \\ \vdots & \ddots & \vdots \\ f_{x_nx_1}(\mathbf{a}) & \cdots & f_{x_nx_n}(\mathbf{a}) \end{bmatrix}.$$

Recall that $D_{\mathbf{v}}f(\mathbf{a}) = Df(\mathbf{a})\mathbf{v}$. One can also show that $D^2_{\mathbf{v}}f(\mathbf{a}) = \mathbf{v}^T Hf(\mathbf{a})\mathbf{v}$ is the second directional derivative of f in the direction of $\mathbf{v}$. It directly measures the concavity of the curve which is the intersection of $\mathbf{graph}(f)$ with the slice determined by $\mathbf{v}$ at $\mathbf{a}$.

EXERCISE 9. Show that the second directional derivative of f in the direction of $\mathbf{v}$ is given by $D^2_{\mathbf{v}}f(\mathbf{a}) = \mathbf{v}^T Hf(\mathbf{a})\mathbf{v}$.

Hence if $D^2_{\mathbf{v}}f(\mathbf{a}) = \mathbf{v}^T Hf(\mathbf{a})\mathbf{v} > 0$, for every direction $\mathbf{v}$ at $\mathbf{a} \in X$, then we can be assured that there is a local minimum of f at $\mathbf{a}$.

Some notes:

- For every $n \times n$ matrix A, one can construct a *quadratic form*: A real-valued function $Q : \mathbb{R}^n \to \mathbb{R}$ defined by
$$Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} = \sum_{i=1}^{n} a_{ij}x_ix_j.$$
In dimension-1, any quadratic form will look like $Q(x) = ax^2$, and in dimension-2, with $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, we have $Q(x, y) = ax^2 + (b + c)xy + dx^2$. In general, a quadratic form will be a polynomial in the variables given, with each monomial corresponding either to the product of two variables or the square of one of them, with total exponent 2 (sum the exponents of each of the factors).
- Quadratic forms are invariant under conjugations of the matrix defining the form. Hence we can always take A to be symmetric (which means $a_{ij} = a_{ji}$ throughout A, or $A^T = A$.
- Hessians are always symmetric! (why?)
- A quadratic form $Q(\mathbf{x})$ is called *positive definite* if $Q(\mathbf{x}) > 0$, for every $\mathbf{x} \neq \mathbf{0}$. (And *negative definite* if $Q(\mathbf{x}) < 0$, for every $\mathbf{x} \neq \mathbf{0}$.)

THEOREM 12.4. *For $X \subset \mathbb{R}^n$ open, let $f : X \to \mathbb{R}$ be C^2 with a critical point $\mathbf{a} \in X$.*

(1) if $Hf(\mathbf{a})$ is positive definite, then f has a local minimum at $\mathbf{a}$.
(2) if $Hf(\mathbf{a})$ is negative definite, then f has a local maximum at $\mathbf{a}$.
(3) If $\det Hf(\mathbf{a}) \neq 0$, and neither positive nor negative definite, then $\mathbf{a}$ is a saddle.

There is a mechanical process for determining when a matrix if positive or negative definite and it is all linear algebra. In essence, it involves testing the *leading principal minors* of $Hf(\mathbf{a})$ to see if $Hf(\mathbf{a})$ is positive definite, negative definite, or neither.

Indeed, let $Q(\mathbf{a}) = \mathbf{x}^T a_{n\times n}\mathbf{x}$ be a quadratic form. Define the *kth leading principal minor* of A to be the determinant of

$$A_k = \begin{bmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{k1} & \cdots & a_{kk} \end{bmatrix}.$$

This A_k is a $k \times k$ submatrix of A consisting of entries that are both in the first k rows and the first k columns of A. Of course, A has n of these:

$$A_1 = a_{11}. \quad A_2 = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, A_3 = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{31} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \quad \ldots, \quad A_n = A.$$

So what can we say?

- If all of these leading principal minors are positive, so if $\mathbf{det}\, A_k > 0$, for $k = 1, \ldots, n$, then A, and hence $Q(\mathbf{x})$, is positive definite.
- A, and hence $Q(\mathbf{x})$, is negative definite if $\mathbf{det}\, A < 0$ for k-odd and $\mathbf{det}\, A > 0$ for k-even.
- A is called *indefinite* if neither of the two cases above holds but all of the leading principal minors are non-zero.
- $Q(\mathbf{x})$ is called *degenerate*, as is A, if $\mathbf{det}\, A = 0$, and nondegenerate otherwise. Note here that it is certainly possible that $Q(\mathbf{x})$ is nondegenerate but at least one of the leading principal minors is 0. Just take any nonsingular matrix with $a_{11} = 0$.

Lastly, the Extreme Value Theorem from single variable calculus has a counterpart in vector calculus. Recall that a set $X \subset \mathbb{R}^n$ is *closed* if it contains all of its boundary points. A set $X \subset \mathbb{R}^n$ is called *bounded* if there exists a real number $M > 0$ such that

$$||\mathbf{x}|| < M, \quad \forall \mathbf{x} \in X.$$

And a set $X \subset \mathbb{R}^n$ is called *compact* if it is both closed and bounded in $\mathbb{R}^n$.

THEOREM 12.5 (The Extreme Value Theorem). *If $X \subset \mathbb{R}^n$ is compact and $f : X \to \mathbb{R}$ is continuous, then f has a global maximum and a global minimum on X.*

Just like in single variable calculus, it is certainly possible for a function f on a possibly nonclosed or unbounded (or both)X to have global extrema. But it is only guaranteed to have each when X is compact and f is continuous.

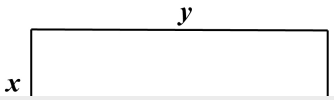
LECTURE 13

Optimization

SYNOPSIS. Just as in single variable calculus, optimizing a function of one variable is a matter for the Extreme Value Theorem and local extrema. But often, situations arise where the objective function involves more than one variable. In this case, there are usually relationships between the variables that allow for rewriting the function as a function of one variable. This is a form of constrained optimization that generalizes well to multivariable calculus. Today we explore this idea, using geometry to "see" our way through to a technique. This leads to the technique of Lagrange multipliers, which we develop here.

HELPFUL DOCUMENTS. Mathematica: `LagrangeMult`.

One variable optimization. Recall optimization in single variable calculus:



EXAMPLE 13.1. Using 1800 linear feet of fencing, construct a rectangular yard along a straight river with the largest area possible. The idea here is to maximize area of a rectangular region. Given the two unknowns of length and width, say, x and y, maximize area $A = xy$. Of course, there is a constraint in that you can only use up to 1800 feet of fencing. Mathematically speaking, this means that $1800 = 2x + y$, given the arrangement of the rectangle in Figure 30. We call the area equation here the *objective function*, and the perimater fencing equation the *constraint*.

The constraint facilitates calculation by

- allowing us to change the objective function, via substitution, into a function of only one var and
- allows us to use single variable calculus techniques to help locate the extrema of the obj function within the constraints.

Now, since $1800 = 2x + y$, we know $y = 1800 - 2x$, so that

$$A = xy = x(1800 - 2x) = 1800x - 2x^2.$$

This is a clue that we are on the right track here, as $A(x)$ has a graph which is a parabola opening down (the leading coefficient is negative). Hence

it will have a max at the vertex. We also know that the variables must be nonnegative numbers, as they denote lengths. Hence $0 \le x \le 900$ and $0 \le y \le 1800$. Hence $A(x)$ has a domain $[0, 900]$ and by the Extreme Value Theorem, must achieve its maximum either at an endpoint or at a critical point. And as $A(x)$ is differentiable, all critical points will occur at places where $a'(x) = 0$. Here

$$A'(x) = 1800 - 4x = 0 \quad \text{is solved only by} \quad x = 450.$$

Immediately, using the Second Derivative Test, we also see that $A''(450) = -4 > 0$. Hence $x = 450$ corresponds to a local maximum, and since $A(0) = A(900) = 0$, $x = 450$ corresponds to a global maximum.

The solution, then, is to construct the pen in Figure 30 with $x = 450$ feet, and $y = 900$ feet.

Here is a different viewpoint of the same problem: Leave the function $A = A(x, y) = xy$ as a function on two variables, and consider the level sets of $A(x, y)$ on the domain

$$R = \left\{(x, y) \in \mathbb{R}^2 \;\middle|\; 0 \le x \le 900, 0 \le y \le 1800\right\},$$

a few of which are graphed in Figure 31. Also on R an in the figure, we can graph the constraint curve (as the red line), thinking of it as the 0-level set of the function $P(x, y) = 2x + y - 1800$. Now, if we are forced to stay on the constraint line, is there a place on this red line where we touch or cross the level set of $A(x, y)$ corresponding to the largest area? One can possibly see it in the figure. But can one "calculate" it?

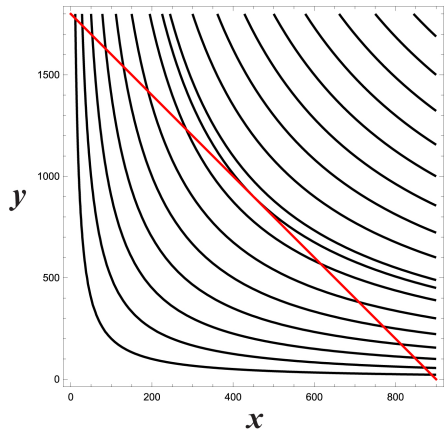


FIGURE 31. Level set of $A(x, y) = xy$ in black, and 0-level set of $P(x, y) = 1800 - 2x - y$ in red.

In this new approach, both the objective function and the constraint are left as functions of the two variables. And we search for a geometric solution to locating an extremum of one function constrained by a second one. One can see in the figure that, as we move along the red line, we are cutting through the level sets of A for a while. At some point, we go tangent to a particular level set and then we start cutting through level sets of A again, although in the other direction (first form lower to higher, then from higher to lower values of A). SO what to we see as the values of A along the red line? We see A rising for a while (cutting through blue lines of increasing A, topping out at some point (the red line becomes tangent to a blue line), then declining in value (again

cutting through level set of decreasing A). We have found our maximum of A along the red line via the point of tangency with a blue line!

This new, geometric, idea of optimization can be generalized: Optimize $f : X \subset \mathbb{R}^n \to \mathbb{R}$, subject to $g : X \to \mathbb{R}$, where $g(\mathbf{x}) = c$. We look for extrema of f while constrained to the c-level set of g. We note here that the idea we started with, that of using the constraint to remove one of the variables in the objective function is less helpful in this multivariable setting. And possibly impossible:

EXAMPLE 13.2. Maximize $f(x, y, z) = x^2 + 3y^2 + y^2z^4$, subject to $g(x, y, z) = e^{xy} - xyz + \cos\left(\frac{xy}{z}\right) = 2$. Try to solve for one of the variables in g as a function of the other two, and substitute that into f to remove a variable!

THEOREM 13.1. *For $X \subset \mathbb{R}^n$ open, $f, g : X \to \mathbb{R}$ both C^1-functions, let*

$$S = \left\{\mathbf{x} \in X \;\middle|\; g(\mathbf{x}) = c\right\}$$

be the c-level set of g. Then, if $f\big|_S$ has an extremum at $\mathbf{x}_0 \in S$, where $\nabla g(\mathbf{x}_0) \neq \mathbf{0}$, then $\exists \lambda \in \mathbb{R}$ such that

$$\nabla f(\mathbf{x}_0) = \lambda \nabla g(\mathbf{x}_0).$$

Some notes:

- The extrema of f will happen at places where ∇f is a multiple of ∇g, as vectors. These wind up being places of tangency between level sets, and places where often the level set of g stops cutting through the level sets of f (for a moment).
- The equation $\nabla f(\mathbf{x}_0) = \lambda \nabla g(\mathbf{x}_0)$ is actually set of n equaitons (non-linear) in $n + 1$ unknowns (each component of the vector $\mathbf{x}$, along with the real number λ). So there are lots of solutions!
- But if we add in the constraint itself, we arrive at $n + 1$ equations in $n + 1$ unknowns:

$$\begin{aligned} f_{x_1}(\mathbf{x}) &= \lambda g_{x_1}(\mathbf{x}) \\ &\vdots \\ f_{x_n}(\mathbf{x}) &= \lambda g_{x_n}(\mathbf{x}) \\ g(\mathbf{x}) &= c. \end{aligned}$$

- The variable λ is called a *Lagrange multiplier*. It's actual value is not nearly as important as its existence!

EXAMPLE 13.3. Identify all critical points of $f(x, y) = 5x + 2y$, subject to $g(x, y) = 5x^2 + 2y^2 = 14$.

Here, $\nabla f(\mathbf{x}) = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$ and $\nabla g(\mathbf{x}) = \begin{bmatrix} 10x \\ 4y \end{bmatrix}$. The system is then

$$\begin{aligned} f_x(\mathbf{x}) &= \lambda g_x(\mathbf{x}) & 5 &= \lambda 10x \\ f_y(\mathbf{x}) &= \lambda g_y(\mathbf{x}) & 2 &= \lambda 4y \\ g(\mathbf{x}) &= c & 5x^2 + 2y^2 &= 14. \end{aligned}$$

Solving, we find by the first and second equations, that $x = \frac{1}{2\lambda} = y$. So the last equations becomes

$$\frac{5}{4\lambda^2} + \frac{2}{4\lambda^2} = 14, \quad \Longrightarrow \lambda = \pm\frac{1}{2\sqrt{2}}.$$

Hence the critical points are $(x, y) = \left(\sqrt{2}, \sqrt{2}\right)$ and $(x, y) = \left(-\sqrt{2}, -\sqrt{2}\right)$

Geometrically, one can see whether these points are extrema or not, and why the gradient condition is quite telling. And analytically?

Handling multiple constraints is done in the same general fashion:

- Each constraint tends to reduce the number of independent variables by one.
- Each constraint tends to reduce the dimension of the space that we evaluate the objective function along by one.
- in $\mathbb{R}^3$, one objective function has level sets which are generically surfaces. Each constraint will also have level sets which are mostly surfaces. two surfaces typically meet in a curve. We then evaluate the objective function along this curve, looking for extrema in a very single variable calculus fashion.

EXAMPLE 13.4. Find the extrema of $f(x, y, z) = 2x + y^2 - z^2$, subject to $g_1(x, y, z) = x - 2y = 0$ and $g_2(x, y, z) = x + z = 0$.

Here, one could simply replace z with $-x$ and y with $\frac{x}{2}$, and look for extrema of $f(x) = 2x + \frac{x^2}{4} - x^2 = 2x - \frac{3}{4}x^2$. One would find that $x = \frac{4}{3}$ is the only extremum and that it is a maximum. So the point $\mathbf{x}_0 = \left(\frac{4}{3}, \frac{2}{3}, -\frac{4}{3}\right)$ is the only critical point of f.

Geometrically, How do we construct a system that we can solve for?

THEOREM 13.2. *For $X \subset \mathbb{R}^n$ open, $f, g_1, \ldots, g_k : X \to \mathbb{R}$ be C^1-functions, with $k < n$. Let*

$$S = \left\{ \mathbf{x} \in X \;\middle|\; g_1(\mathbf{x}) = c_1, \ldots, g_k(\mathbf{x}) = c_k \right\}$$

be the intersection of the level sets of the g_i, $i = 1, \ldots, k$. Then, if $f\big|_S$ has an extremum at $\mathbf{x}_0 \in S$, where $\nabla g_1(\mathbf{x}_0), \ldots, g_k(\mathbf{x}_0)$ are all linearly independent

as vectors, then there exist scalars $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ *such that*

$$\nabla f(\mathbf{x}_0) = \lambda_1 \nabla g_1(\mathbf{x}_0) + \ldots + \lambda_k \nabla g_k(\mathbf{x}_0). \tag{13.0.1}$$

Notes:

- Recall that linear independence as vectors as means that every vector must be nonzero!
- Basically, as a vector and at an extremum, $\nabla f(\mathbf{x}_0)$ must be in the space spanned by the $\nabla g_i(\mathbf{x}_0)$, for $i = 1, \ldots, k$.

Example 13.5. In Example 13.4 above, we sought the extrema of $f(x, y, z) = 2x + y^2 - z^2$, subject to the two constraints $g_1(x, y, z) = x - 2y = 0$ and $g_2(x, y, z) = x + z = 0$. To use Theorem 13.2, we form Equation 13.0.1 directly via the vectors

$$\nabla f(\mathbf{x}) = \begin{bmatrix} 2 \\ 2y \\ -2z \end{bmatrix}, \quad \nabla g_1(\mathbf{x}) = \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix}, \quad \text{and} \quad \nabla g_2(\mathbf{x}) = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

Here, the constraint vectors are linearly independent everywhere (why is this?), so the system is

$$\begin{aligned}
f_x(\mathbf{x}) &= \lambda_1 \frac{\partial g_1}{\partial x}(\mathbf{x}) + \lambda_2 \frac{\partial g_2}{\partial x}(\mathbf{x}) & 2 &= \lambda_1 + \lambda_2 \\
f_y(\mathbf{x}) &= \lambda_1 \frac{\partial g_1}{\partial y}(\mathbf{x}) + \lambda_2 \frac{\partial g_2}{\partial y}(\mathbf{x}) & 2y &= -2\lambda_1 \\
f_z(\mathbf{x}) &= \lambda_1 \frac{\partial g_1}{\partial z}(\mathbf{x}) + \lambda_2 \frac{\partial g_2}{\partial z}(\mathbf{x}) & -2z &= \lambda_2 \\
g_1(\mathbf{x}) &= c_1 & x - 2y &= 0 \\
g_2(\mathbf{x}) &= c_2 & x + z &= 0.
\end{aligned}$$

There are many ways to solve these 5 equations in 5 unknowns. One way it so eliminate the lambdas in the first equation via substitution using the second and third. One obtains $\lambda_1 = -y$ and $\lambda_2 = -2z$, so that the first equation is $2 = -y - 2z$. And eliminating x in the last two equations yields the single equation $0 = 2y + z$. Together, the system

$$\begin{aligned}
2 &= -y - 2z \\
0 &= 2y + z
\end{aligned}$$

is solved by $z = -\frac{4}{3}$ and $y = \frac{2}{3}$. One then calculates $x = \frac{4}{3}$, so that the only critical point of f is again $\mathbf{x}_0 = \left(\frac{4}{3}, \frac{2}{3}, -\frac{4}{3}\right)$.

LECTURE 14

The Definite Integral

SYNOPSIS. The integral calculus of functions of more than one variable also follows closely the structure and patterns of single variable calculus. However, noting that graphs of functions, even of two independent variables are no longer curves, but hypersurfaces in R^{n+1}, the idea of "area under a curve" must be suitably generalized. In this lecture, we lay the groundwork to understand volumes in many dimensions and what is means to calculate. Then we alter the idea of single variable integration to fit this new multi-dimensional arena and build the tools and structures we need to create the integral calculus.

Volumes of regions. The area of a two dimensional region $\mathcal{R}$ (say the difference between the graphs of two function over the same domain in single variable calculus, is really just a "sum" of the lengths of all of the vertical lines formed by slicing (along lines of constant values of the independent variable x) on some interval of x comprising the region. Using Figure **??** below, we have

$$\text{Area}(\mathcal{R}) = \int_a^b \ell(x)\,dx,$$

where for each value of $x \in [a,b]$, the value of $\ell(x)$ is $f(x) - g(x)$ on the left and $f(x) - 0 = f(x)$ on the right.

This remains true in higher dimensions, at least once we understand the notions of lengths and areas in higher dimensions: Given $f(x,y)$, a nonnegative function defined and continuous on the rectangle

$$\mathcal{R} = \left\{(x,y) \in \mathbb{R}^2 \;\middle|\; a \le x \le b,\ c \le y \le d\right\}, \tag{14.0.1}$$

its graph lies "over" the region in the xy-plane in $\mathbb{R}^3$ as the set of points $(x,y,0) \in \mathbb{R}^3$, where $(x,y) \in \mathcal{R}$. Add in the vertical walls connecting the four edges of $\mathcal{R}$ in the floor to the corresponding graphs of the edges in $\mathbf{graph}(f)$, and one obtains a solid region in $\mathbb{R}^3$, as in Figure **??**, which we define as

$$\mathcal{S} = \left\{(x,y,z) \in \mathbb{R}^3 \;\middle|\; a \le x \le b,\ c \le y \le d,\ 0 \le z \le f(x,y)\right\}.$$

We can calculate the volume of $\mathcal{S}$ as the sum of all of the areas of the "vertical" slices through $\mathcal{S}$ by, say, slicing along lines of constant x, which

we will call $\mathcal{R}_x$. In this way, we could write

$$\text{Volume}(\mathcal{S}) = \int_a^b \text{Area}(\mathcal{R}_x)\, dx.$$

Note that we could also slice vertically along lines of constant y, creating regions $\mathcal{R}_y$, so that

$$\text{Volume}(\mathcal{S}) = \int_c^d \text{Area}(\mathcal{R}_y)\, dy.$$

We will stick with the former for now.

So, for each value of x, what is the area of each $\mathcal{R}_x$? At a point $x_0 \in [a, b]$, the area of $\mathcal{R}_{x_0}$ is

$$\text{Area}(\mathcal{R}_{x_0}) = \int_c^d \ell(y)\, dy = \int_c^d f(x_0, y)\, dy.$$

Nesting these two concepts together, we arrive at

$$\text{Volume}(\mathcal{S}) = \int_a^b \text{Area}(\mathcal{R}_x)\, dx = \int_a^b \left(\int_c^d f(x, y)\, dy \right) dx = \int_a^b \int_c^d f(x, y)\, dy\, dx.$$

Some notes:

- The parentheses distinguishing the "inside" integral from the "outside" (in the penultimate expression) are not strictly needed (and hence removed) if one understands that the integrals are always taken to be nested.
- The use of the choice of $x = x_0$ subscript is also not needed, and hence removed. It is understood here that as one integrates with respect to one variable, the other is considered fixed, like a parameter. Do you recall this idea from the notion of partial differentiation?
- If the limits of the variables do not depend on each other, then the region one is integrating over is rectangular. In this case, one can reverse the process and form a nested pair of integrals with the order of integration reversed, but with the same result. So

$$\text{Volume}(\mathcal{S}) = \int_a^b \int_c^d f(x, y)\, dy\, dx = \int_c^d \int_a^b f(x, y)\, dx\, dy.$$

The general notion of parameterizing the parallel slices (in any chosen direction) through a solid to find its volume is known as *Cavalieri's Principle*. Let $\mathcal{S} \subset \mathbb{R}^n$ be an n-dimensional solid in n-space, bounded in the x_1-direction by $[a, b]$. Then

$$\text{Volume}(\mathcal{S}) = \int_a^b \text{Volume}(\mathcal{R}_{x_1})\, dx_1,$$

where $\text{Volume}(\mathcal{R}_{x_1})$ is the volume of the $(n-1)$-dimensional x_1-slice through $\mathcal{S}$ at $x_1 \in [a, b]$. Recursively speaking, calculating the volume of $\mathcal{S}$ will involve a nested set of n integrals, or an *n-tuple integral*. Note that, while we would easily use terms like *quadruple integral* or *quintuple integral* for

volumes in, respectively, $\mathbb{R}^4$ and $\mathbb{R}^5$, we commonly refer to a nested set of three integrals as a *triple integral*, and in two dimensions, a *double integral*.

Now one can define a double integral on a rectangular region $\mathcal{R}$ via a 2-dimensional Riemann Sum:

Define a nonnegative $f(x, y)$ on the region $\mathcal{R}$ defined above in Equation 14.0.1, and partition $\mathcal{R}$ into boxes by partitioning the two intervals $[a, b]$ and $[c, d]$:

$$a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b, \quad \text{and}$$
$$c = y_0 < y_1 < \cdots < y_{m-1} < y_m = d,$$

so that $\Delta x_i = x_i - x_{i-1}$ and $\Delta y_j = y_j - y_{j-1}$. Then the area of the ijth box is then $\Delta A_{ij} = \Delta x_i \Delta y_j$.

Now choose a point (p_i, q_j) within each box, where $p_i \in [x_{i-1}, x_i]$ and $q_j \in [y_{j-1}, y_j]$, $i = 1, \ldots, n$ and $j = 1, \ldots, m$. Then we can approximate the volume of the solid $\mathcal{S}$ between the floor (the domain $\mathcal{R}$ in the xy-plane in $\mathbb{R}^3$, the ceiling (the $\mathbf{graph}(f)$ over $\mathcal{R}$, by the sum of all of the volumes of the small cuboids whose base in $[x_{i-1}, x_i] \times [y_{j-1}, y_j]$, height is $f(p_i, q_j)$. Hence

$$\text{Volume}(\mathcal{S}) \approx \sum_{i=1}^{n} \sum_{j=1}^{m} f(p_i, q_j) \Delta A_{ij}.$$

This is a 2-dimensional Riemann Sum.

Definition 14.1. The double integral of f on $\mathcal{R}$ is

$$\iint_{\mathcal{R}} f \, dA = \lim_{n,m \to \infty} \sum_{i=1}^{n} \sum_{j=1}^{m} f(p_i, q_j) \Delta A_{ij}$$

when the limits exists.

Notes:

(1) Actually, as stated, the definition has a serious flaw in it. I will leave it unspecified to see if you can see it. It is a flaw in the nature of the limit. Find it!
(2) If the limit exists, then we say f is *integrable* on $\mathcal{R}$.
(3) The notation used, without specific upper and lower limits but the more general $\mathcal{R}$ under the double integral sign, is common and accentuates the region $\mathcal{R}$ instead of the coordinates used. But, using the standard cartesian coordinates x, and y, we automatically know then, that, in this case,
$$\iint_{\mathcal{R}} f \, dA = \int_a^b \int_c^d f(x, y) \, dy \, dx = \int_c^d \int_a^b f(x, y) \, dx \, dy.$$
(4) Over any base box in a Riemann Sum, if $f(x, y) < 0$, then we interpret the volume of that box as negative, just like in single variable calculus.
(5) Also like in single variable calculus, the same problems and caveats that occur with the limit can occur here also:

- In single variable calculus, piecewise continuous functions on an interval $[a, b]$ are integrable. Recall that piecewise continuous functions are those that are continuous everywhere, except on a finite set of points where "jump" discontinuities can occur.
- In two dimensions, if f is bounded on $\mathcal{R}$ with the set of all discontinuities having zero area, then f is integrable. One way to see this is to think of $\mathbf{graph}(f)$ as smooth but possibly cut up into a finite number of pieces.
- Continuous functions on closed, bounded domains are always integrable.

THEOREM 14.2 (Fubini). *Let f be bounded on $\mathcal{R} = [a, b] \times [c, d]$ and assume that the set S of discontinuities of f on $\mathcal{R}$ has zero area. If every line parallel to the coordinate axes meets S in, at most, a finite number of places, then*

$$\iint_{\mathcal{R}} f \, dA = \int_a^b \int_c^d f(x, y) \, dy \, dx = \int_c^d \int_a^b f(x, y) \, dx \, dy.$$

Notes:

- The fact that $\mathcal{R}$ is a rectangle is vital here.
- The stipulation that all lines parallel to coordinate axes meet S in at most a finite number of places is sufficient but not strictly necessary. It forces the function intersecting the slice to be piecewise continuous, and thus integrable. But this is not the only way to have an integrable funciton on each slice.
- For all intents and purposes, aero area means that the set of discontinuities has smaller dimension as a set than $\mathcal{R}$.

The properties of double integrals reflect those of their 1-dimensional cousins. See Proposition 5.2.7 on page 320 of the text.

Keep in mind that Cavalieri's Principle will still hold for solids in 3-space defined for regions of the plane (as domains for functions) more general than rectangles. However, the order of integration, when defining and calculating a double integral, may matter. Hence, we need to understand why and how.

DEFINITION 14.3. A region $\mathcal{D} \subset \mathbb{R}^2$ is called *elementary* if it can be described via an interval in one variable and as the difference between two functions of that variable in the other. There are three types:

(1) **Type I:** $\mathcal{D} = \left\{(x, y) \in \mathbb{R}^2 \,\middle|\, a \le x \le b, \gamma(x) \le y \le \delta(x)\right\}$, where $\gamma(x)$ and $\delta(x)$ are continuous functions on $[a, b]$.
(2) **Type II:** $\mathcal{D} = \left\{(x, y) \in \mathbb{R}^2 \,\middle|\, c \le y \le d, \alpha(x) \le x \le \beta(x)\right\}$, where $\alpha(y)$ and $\beta(y)$ are continuous functions on $[c, d]$.
(3) **Type III:** $\mathcal{D}$ is of both Type I and Type II.

A region $\mathcal{D}$ is called *non-elementary* if it is neither Type I nor Type II. We immediately have:

THEOREM 14.4. *If $\mathcal{D} \subset \mathbb{R}^2$ is elementary and f is C^0 on $\mathcal{D}$, then*

(1) **Type I:** $\displaystyle\iint_{\mathcal{D}} f\, dA = \int_a^b \int_{\gamma(x)}^{\delta(x)} f(x,y)\, dy\, dx,$

(2) **Type II:** $\displaystyle\iint_{\mathcal{D}} f\, dA = \int_c^d \int_{\alpha(y)}^{\beta(y)} f(x,y)\, dx\, dy.$

Note here that the proof is noneventful, and relies on a notion of extending $\mathcal{D}$ to some rectangular region $\mathcal{R} \supset \mathcal{D}$ by creating a new function f^{ext} on $\mathcal{R}$ which equals f on $\mathcal{D}$ and is 0 outside of $\mathcal{D}$ in $\mathcal{R}$. This creates a discontinuous function on $\mathcal{R}$, but one that is integrable according to Fubini. Do not worry about this technique. It works for the theorem, but is not necessary to know for a good understanding of integration.

EXAMPLE 14.1. Let $f(x) = -\frac{3}{2}x + 4$, and $g(x) = \frac{1}{2}x$. The region $\mathcal{D}$ in the (closed) first quadrant of the plane between these two functions is a triangle of height 2, with base along the y-axis from 0 to 4. Integrate the function $h(x,y) = 2x + 2y$ on $\mathcal{D}$.

Strategy: View $\mathcal{D}$ as elementary of either type and construct the double integral according to Theorem 14.4. Then use the Fundamental Theorem of Calculus (from single variable calculus) on the "inside" integral, then again on the "outside" integral.

Solution: Viewing $\mathcal{D}$ as a Type I elementary region, we set $\gamma(x) = g(x)$ and $\delta(x) = f(x)$, and use the formula of Theorem 14.4 to set up the integral. We get

$$\iint_{\mathcal{D}} f\, dA = \int_0^2 \int_{\frac{1}{2}x}^{-\frac{3}{2}x+4} (2x + 2y)\; dy\, dx.$$

Then we calculate:

$$\begin{aligned}
\iint_{\mathcal{D}} f\, dA &= \int_0^2 \int_{\frac{1}{2}x}^{-\frac{3}{2}x+4} (2x + 2y)\; dy\, dx = \int_0^2 \left[\left(2xy + y^2\right) \Big|_{\frac{1}{2}x}^{-\frac{3}{2}x+4} \right] dx \\
&= \int_0^2 \left[2x\left(-\frac{3}{2}x + 4\right) + \left(-\frac{3}{2}x + 4\right)^2 - \left(2x\left(\frac{1}{2}x\right) + \left(\frac{1}{2}x\right)^2\right) \right] dx \\
&= \int_0^2 \left[-3x^2 + 8x + \frac{9}{4}x^2 - 12x + 16 - x^2 - \frac{1}{4}x^2 \right] dx \\
&= \int_0^2 \left(-2x^2 - 4x + 16\right) dx \\
&= \left(-\frac{2}{3}x^3 - 2x^2 + 16x \right) \Big|_0^2 = -\frac{16}{3} - 8 + 32 = \frac{56}{3}.
\end{aligned}$$

Notice that we can also deem $\mathcal{D}$ as elementary of Type II, using

$$\alpha(y) = 0, \quad \text{and} \quad \beta(y) = \begin{cases} -\frac{2}{3}(y-4) & y \in [1,4] \\ 2y & y \in [0,1]. \end{cases}$$

Then the construction and calculation become

$$\begin{aligned}
\iint_{\mathcal{D}} f\, dA &= \int_0^4 \int_{\alpha(y)}^{\beta(y)} (2x+2y)\, dx\, dy \\
&= \int_0^1 \int_0^{2y} (2x+2y)\, dx\, dy + \int_1^4 \int_0^{-\frac{2}{3}(y-4)} (2x+2y)\, dx\, dy \\
&= \int_0^1 \left[\left(x^2+2xy\right)\Big|_0^{2y} \right] dy + \int_1^4 \left[\left(x^2+2xy\right)\Big|_0^{-\frac{2}{3}(y-4)} \right] dy \\
&= \int_0^1 \left((2y)^2 + 2(2y)y\right) dy + \int_1^4 \left(\left(-\frac{2}{3}(y-4)\right)^2 + 2\left(-\frac{2}{3}(y-4)\right) y \right) dy \\
&= \int_0^1 8y^2\, dy - \frac{2}{3}\int_1^4 \left(-\frac{2}{3}(y^2-8y+16) + 2y^2 - 8y \right) dy \\
&= \left[\frac{8}{3}y^3 \Big|_0^1 \right] - \frac{2}{3}\int_1^4 \left(\frac{4}{3}y^2 - \frac{8}{3}y - \frac{32}{3} \right) dy \\
&= \frac{8}{3} - \frac{8}{9}\left[\left(\frac{y^3}{3} - y^2 - 8y \right)\Big|_1^4 \right] \\
&= \frac{8}{3} - \frac{8}{9}\left(\frac{64}{3} - 16 - 32 - \frac{1}{3} + 1 + 8 \right) = \frac{8}{3} - \frac{8}{9}(-18) = \frac{8}{3} + 16 = \frac{56}{3}.
\end{aligned}$$

And lastly, more complicated regions (those that are not elementary), can usually be broken up into a set of elementary regions that meet along boundaries. Then the integrals over these adjacent regions can be added together, noting that the contributions along the boundaries will be zero.

EXAMPLE 14.2. Consider the *annular* region $\mathcal{D}$ between the two planar equations $x^2+y^2=1$ and $x^2+y^2=4$. This region is not elementary! But in the plane, slice up $\mathcal{D}$ into four regions using the two vertical lines $x = \pm 1$, as in Figure **??**. Then we have

- $\mathcal{D}_1 = \left\{ (x,y) \in \mathbb{R}^2 \;\middle|\; -4 \le x \le -1,\ -\sqrt{4-x^2} \le y \le \sqrt{4-x^2} \right\}$,
- $\mathcal{D}_2 = \left\{ (x,y) \in \mathbb{R}^2 \;\middle|\; -1 \le x \le -1,\ -\sqrt{4-x^2} \le y \le -\sqrt{1-x^2} \right\}$,
- $\mathcal{D}_3 = \left\{ (x,y) \in \mathbb{R}^2 \;\middle|\; -1 \le x \le -1,\ \sqrt{1-x^2} \le y \le \sqrt{4-x^2} \right\}$,

- $\mathcal{D}_4 = \left\{(x,y) \in \mathbb{R}^2 \;\middle|\; 1 \le x \le 4,\ -\sqrt{4-x^2} \le y \le \sqrt{4-x^2}\right\}$. As written, all of these are Type I.

EXAMPLE 14.3. ***Find the area of a circle of radius*** $r \ge 0$. Here, the circle of radius r centered at the origin is the set of points that satisfy the equation $x^2 + y^2 = r^2$. The region is then the *closed disk* $\mathcal{D}_r$ consisting of the interior of this circle and the circle itself. It is elementary of Type III, and can be written as elementary of Type I as

$$\mathcal{D}_r = \left\{(x,y) \in \mathbb{R}^2 \;\middle|\; -r \le x \le r,\ -\sqrt{r^2-x^2} \le y \le \sqrt{r^2-x^2}\right\}.$$

Using this, then, we have

$$\text{Area}(\mathcal{D}_r) = \int_{-r}^{r} \int_{-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}} dy\, dx.$$

Two things here: First, what is the integrand here? And why does this work? And secondly, finish this calculation. Note that you will have to use a inverse trig substitution to solve this. Perhaps THAT is why you will wind up with the answe: $\text{Area}(\mathcal{D}_r) = \pi r^2$.

LECTURE 15

The Definite Triple Integral

SYNOPSIS. Today we continue the general idea of integration of a real-valued function on more than one variable by generalizing the 2-dimensional version to three dimensions. There is little that is new here except for the pattern of the generalization that leads to the n-dimensional version. Fubini's Theorem still holds, and switching the order of integration outside of a cuboid region still involves checking that the region is elementary in different permutations of the variables of integration and that, if so, one can rewrite the limits as functions of some of the variable properly.

Volumes of higher dimensional regions. Let

$$\mathcal{B} = \left\{ (x,y,z) \in \mathbb{R}^3 \;\middle|\; \begin{array}{l} a \le x \le b \\ c \le y \le d \\ p \le z \le q \end{array} \right\}$$

be a *cuboid*. We can approximate the volume of the four-dimensional solid $\mathcal{S}$ with $\mathcal{B}$ as its base (can you envision this?) and $\mathbf{graph}(f)$ as its *roof*, where $f : \mathcal{B} \to \mathbb{R}$ is a nonnegative C^0-function, by:

- Partitioning all three dimensions of $\mathcal{B}$ so that

$$\mathcal{B}ijk = \left\{ (x,y,z) \in \mathbb{R}^3 \;\middle|\; \begin{array}{l} x \in [x_{i-1}, x_i] \\ y \in [y_{j-1}, y_j] \\ z \in [z_{k-1}, z_k] \end{array} \right\},$$

  with lengths $\Delta x_i = x_i - x_{i-1}$, $\Delta y_j = y_j - y_{j-1}$, and $\Delta z_k = z_k - z_{k-1}$. Note that, in this construction, we will choose a partition size of n for all three dimensions. This will greatly simplify the construction;
- Find $\mathbf{volume}(\mathcal{B}_{ijk}) = \Delta V_{ijk} = \Delta x_i \Delta y_j \Delta z_k$;
- Choose a point $\mathbf{c}_{ijk} \in \mathcal{B}_{ijk}$, $\forall i,j,k \in \{1,\dots,n\}$;
- Sum over indices,

$$\mathbf{volume}(\mathcal{S}) \approx \sum_{i,j,k=1}^{n} f(\mathbf{c}_{ijk}) \Delta V_{ijk}.$$

Then the triple integral of f over $\mathcal{B}$ is

$$\mathbf{volume}(\mathcal{S}) = \iiint_{\mathcal{B}} f \, dV = \lim_{n\to\infty} \sum_{i,j,k=1}^{n} f(\mathbf{c}_{ijk}) \Delta V_{ijk} = \lim_{n\to\infty} \sum_{i,j,k=1}^{n} f(x_i, y_j, z_k) \Delta x_i \Delta y_j \Delta z_k.$$

Here are some facts about triple integrals:

(1) Fubini's Theorem still holds in higher dimensions.

(2) A region $\mathcal{W} \in \mathbb{R}^3$ is called *elementary* if it can be written as

$$\mathcal{W} = \left\{ (x,y,z) \in \mathbb{R}^3 \;\middle|\; \begin{array}{c} a \le x \le b \\ \alpha(x) \le y \le \beta(x) \\ \varphi(x,y) \le z \le \psi(x,y) \end{array} \right\},$$

or some permutation of these variables. Here, one direction should look like an interval, a second direction should look like the difference between two functions of that first variable, and the third direction should look like the difference between two functions of the other two.

(3) Given an elementary $\mathcal{W}$, then, if f is continuous on $\mathcal{W}$, we have

$$\iiint_{\mathcal{W}} f\,dV = \int_a^b \int_{\alpha(x)}^{\beta(x)} \int_{\varphi(x,y)}^{\psi(x,y)} f(x,y,z)\,dz\,dy\,dx,$$

or, again, some permutation of the three variables. Note again that the integrals are nested here.

(4) The volume of the solid $\mathcal{W}$ can be found by integrating the *unit function* $f(x,y,z) = 1$, so

$$\textbf{volume}(\mathcal{W}) = \iiint_{\mathcal{W}} 1\,dV = \iiint_{\mathcal{W}} dV.$$

(5) Sometimes, it is advantageous to understand that

$$\iiint_{\mathcal{W}} f\,dV = \iint_{\mathcal{D}} \int_{\varphi(x,y)}^{\psi(x,y)} f(x,y,z)\,dz\,dA,$$

for $\mathcal{D}$ elementary in x and y, and $\mathcal{W}$ elementary in all three variables.

EXAMPLE 15.1. ***What is the volume of the unit sphere in*** $\mathbb{R}^3$***?*** Here we define the unit 2-sphere as

$$S^2 = \left\{ (x,y,z) \in \mathbb{R}^3 \;\middle|\; x^2+y^2+z^2 = 1 \right\},$$

noting that the notation is common in mathematics and generalizes to S^n, the unit n-sphere, $n \in \mathbb{N}$, as a subset of $\mathbb{R}^{n+1}$. One ways to think of this is the set of all unit-length vectors in $(n+1)$-space. So what does S^1 look like? How about S^0??

Here, then, the space consisting of S^2 and its interior is sometimes called the (unit) 3-ball, or B^3. So we are looking for the volume of B^3. Note that B^3 is elementary, and can be written as the difference between two functions $z = -\sqrt{1-x^2-y^2}$ and $z = \sqrt{1-x^2-y^2}$, on the domain

$$\mathcal{D} = \left\{ (x,y) \in \mathbb{R}^2 \;\middle|\; x^2+y^2 = 1 \right\} = \left\{ (x,y,z) \in \mathbb{R}^2 \;\middle|\; x^2+y^2 = 1,\ z = 0 \right\}.$$

These graphs of these two functions are the southern and northern hemispheres of S^2, respectively, and meet at the equator.

So here,

$$\textbf{volume}(B^3) = \iiint_B i\,dV = \iint_{\mathcal{D}} \int_{-\sqrt{1-x^2-y^2}}^{\sqrt{1-x^2-y^2}} dz\,dA$$
$$= \int_{-1}^{1} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \int_{-\sqrt{1-x^2-y^2}}^{\sqrt{1-x^2-y^2}} dz\,dy\,dx.$$

EXAMPLE 15.2. ***Let $\mathcal{W}$ be the region bounded by $y = x^2$ and $y+z = 9$ and the xy-plane. Integrate $f(x,y,z) = 8xyz$ over $\mathcal{W}$.*** Here, the roof of this solid is the inclined plane $z = 9-y$, the floor is the xy-plane, and the wall is given by the parabolic $y = x^2$, projected vertically out of the floor. We get Figure 32 below. $\mathcal{W}$ is an elementary region, and one way to see this is the following: As z goes from the floor at 0 to the roof at $9-y$, the variable y goes from 9 to x^2, and x ranges from -3 to 3. Hence the integration is

$$\begin{aligned}
\int_{-3}^{3}\int_{x^2}^{9}\int_{0}^{9-y} 8xyz\,dz\,dy\,dx &= \int_{-3}^{3}\int_{x^2}^{9}\left[8xy\left(\frac{z^2}{2}\right)\Big|_0^{9-y}\right] dy\,dz \\
&= \int_{-3}^{3}\int_{x^2}^{9} 4xy(9-y)^2\,dy\,dx \\
&= \int_{-3}^{3}\int_{x^2}^{9} 4x\left(81y - 18y^2 + y^3\right)\,dy\,dx \\
&= \int_{-3}^{3} 4x\left[\left(\frac{81}{2}y^2 - \frac{18}{3}y^3 + \frac{y^4}{4}\right)\Big|_{x^2}^{9}\right] dx \\
&= \int_{-3}^{3} 4x\left[\frac{81^2}{2} - 6(9)^3 + \frac{9^4}{4} - \frac{81}{2}x^4 + 6x^6 - \frac{x^8}{4}\right] dx = 0.
\end{aligned}$$

A good question to ask is: Why is this quantity 0? One can "see" that this is true at this point due to the properties of the integral one learned in single variable calculus. Indeed, notice that the integrand is actually an *odd function*, symmetric with respect to the origin. In this case, the integrans is a polynomial with all of the monomials of odd degree, and that the interval one is integrating over is of the form $[-a,a]$, for some $a \geq 0$. Hence one can cease calculating here and conclude.

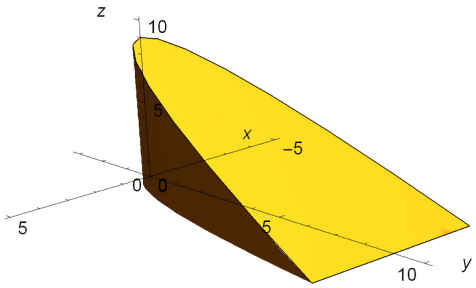


FIGURE 32. The solid $\mathcal{W}$ in Example 15.2.

LECTURE 16

Changing Variables in Integration

SYNOPSIS. Here, we focus on the idea of changing the coordinates in an integral. In Calculus I, the Substitution Method was an actual change of coordinates used usually to make the integrand easier to play with. Here, and in more generality, changing the coordinate system on a region is used more to make the region easier to integrate over. Of course, it must be true that the value of the definite integral should be the same no matter the coordinates used. Hence one must be careful to properly account for the change, precisely as in the Substitution Method, where a change of variable creates a new variable corresponding to the "inside function" of the composition of functions in the integrand (this is a function of the old variable). The extra piece was the derivative of the inside function. This generalizes as the 1-dimensional version of a similar phenomena in higher dimensions. We detail this today.

HELPFUL DOCUMENTS. PDF: `SphereVolume`.

Parameterization. Placing new coordinates on a space involves again a function: Let $T : X \subset \mathbb{R}^2 \to \mathbb{R}^2$, $T(u, v) = (x(u, v), y(u, v))$ be a C^1-map. Then, any subset $D \subset X$, in the uv-coordinates, is mapped to its image $T(D)$, another region (in the xy-coordinates). See Figure 33. To do this, one needs to write the old coordinates as functions of the new coordinates. In functional form, this involves a function whose domain is in the new coordinates and whose codomain uses the old coordinates.

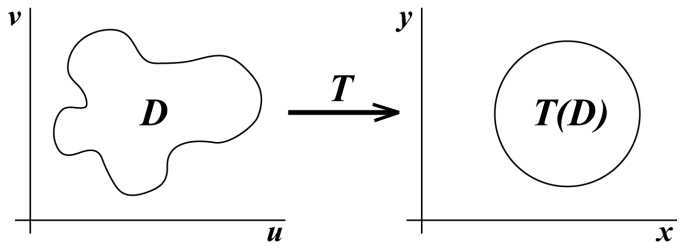


FIGURE 33. A coordinate change in the plane.

EXAMPLE 16.1. ***Polar Coordinates.*** Working in polar coordinates in the plane involves switching the variables from x and y, to r and θ, representing, respectively, range (Euclidean distance from the origin) and angle (from a reference line in the xy-plane, often the positive x-axis. The resulting coordinate equations are $x = r\cos\theta$ and $y = r\sin\theta$, which takes the domain $D = [0,1] \times [0, 2\pi]$ in the $r\theta$-plane to the unit disk $T(D)$ in the xy-plane. See Figure 34 below.

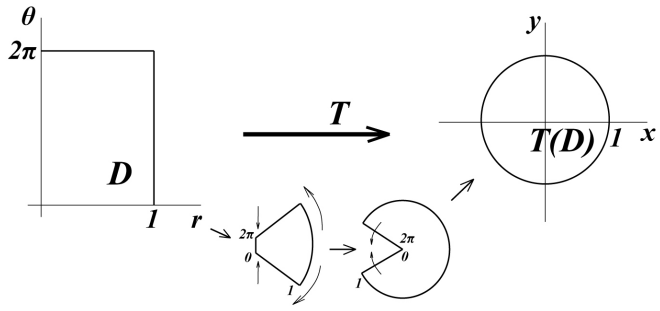


FIGURE 34. Here, $T(r,\theta) = (x(r,\theta), y(r,\theta)) = (r\cos\theta, r\sin\theta) = (x, y)$.

Notice here that, in this coordinate change, the edges of the box D under T do not behave so well. For instance, the entire left edge is mapped to the origin (all points $(0,\theta)$ in the $r\theta$-plane represent the origin no matter the value of θ), and the top and bottom edges of D under T are *identified* (mapped point by point to the same image). This represents the fact that $(r, \theta + 2\pi) = (r,\theta)$. However, on the inside of the box D, the map T is injective. This will be vitally important.

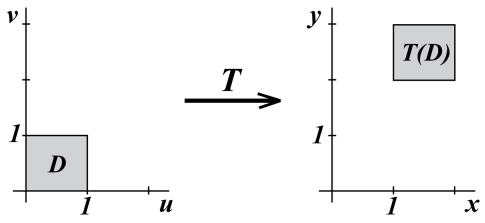


FIGURE 35. Here, $T(u,v) = (x(u,v), y(u,v))) = (u+1, v+2)$.

EXAMPLE 16.2. ***Translations.*** In the process of studying the properties of an equilibrium solution in a system of ordinary differential equations, one often changes coordinates by a translation, moving the equilibrium solution (really a point in Euclidean space) to the origin, before analyzing. For

example, the coordinate change $u = x - 1$ and $v = y - 2$ is a transformation of the plane that moves the origin down 2 and to the left 1. Then, writing x and y in terms of the new variables u and v, we see that the unit square $= [0,1] \times [0,1]$ in the uv-plane is moved by T to $T(D) = [1,2] \times [2,3]$. Under this transformation, if, for example, one were to integrate a function over $[1,2]$ in x and $[2,3]$ in y, then by changing coordinates, one could instead integrate over the unit square in the uv-plane.

EXAMPLE 16.3. ***A Linear Map of*** $\mathbb{R}^2$**.** Domains which are parallelograms have the quality that a linear transformation can take them to squares (or at least rectabgles). Consider T the linear planar transformation $x = u+v$ and $y = u - v$. As a transformation, T takes the unit square D to a bigger square $T(D)$, though here it is one that is not oriented so that its sides are parallel to the coordinate axes. See Figure 36

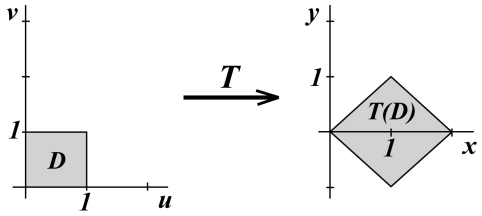


FIGURE 36. Here, $T(u,v) = (u(x,y), v(x,y))) = (x+1, y+2)$.

Notes:

- In general, it takes practive to "see" a transformation, and to construct one.
- One good practice to well-understand how to transform the boundary first, or at least any vertices.
- Example 16.3 is an example of a transformation that is often easily constructible.

PROPOSITION 16.1. *Let* $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$*, where* $\boldsymbol{det}\, A = ad - bc \neq 0$*. Then the planar transformation* $T(u,v) = A \begin{bmatrix} u \\ v \end{bmatrix}$ *is one-to-one, onto, takes parallelograms to parallelograms (vertices to vertices), and*

$$\boldsymbol{area}(T(D)) = |\boldsymbol{det}\, A| \boldsymbol{area}(D).$$

Some notes:

- As a linear transformation, note that it is always the case that the origin goes to the origin, so no translations here.
- Notice the use of the absolute value for the determinant. One needs this to relate the areas, which are positive measures of size in the plane.
- Of course, this proposition generalizes to n-dimensions, for $n \in \mathbb{N}$.
- The proof is really just linear algebra with a bit of geometry thrown in.

So here is the new question: Why would someone interested in integration over a domain want to change the shape and/or size of the domain? The answer is that rectangles are easier to integrate over than more general domains, and often a nonelementary domain can be made elementary via a coordinate change.

Special Note: Notice in the above examples that the more complicated regions are in the codomain of the transformation $T(u,v) = (x,y)$. This is because one writes the given variables x and y in terms of the new variables u and v, so $x = x(u,v)$ and $y = y(u,v)$. Then, by composition, any function $f(x,y)$ can be composed with T to generate a version of f written in terms of u and v:

$$(f \circ T)(u,v) = f(T(u,v)) = f(x(u,v), y(u,v)) = f(u,v).$$

You have seen this before in single variable calculus in the form of the *Substitution Method* for the evaluation of an integral. If $g'(x)$ is continuous on $[a,b]$, and f is continuous on the range of $u = g(x)$, then

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

Here, one changed the variable x to u in order to make the integrans easier to integrate. Part of the simplifying effect of the change in variables was the required term given by the relationship between differentials $du = g'(x)\, dx$. We will also need this in our new multi-dimensional case.

Let $T(u,v) = (x(u,v), y(u,v))$ be a C^1-transformation of $\mathbb{R}^2$. Then

$$DT(u,v) = \begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{bmatrix}, \quad \text{and}$$

$$\mathbf{Jac}(T) = \mathbf{det}(DT) = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}$$

$$= \frac{\partial(x,y)}{\partial(u,v)} = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}.$$

We have

THEOREM 16.2. *Let D and D^* be elementary regions in the xy-plane and uv-plane, respectively, and suppose $T : \mathbb{R}^2 \to \mathbb{R}^2$ is C^1, $D = T(D^*)$, and*

T is injective on te interior of D^*. Then for any integrable $f : D \to \mathbb{R}$,

$$\iint_D f(x,y)\,dx\,dy = \iint_{D^*} f\left(x(u,v), y(u,v)\right) \left|\frac{\partial(x,y)}{\partial(u,v)}\right| du\,dv.$$

Again, the absolute value of the Jacobian determinant is an important part of this formula. We will see why later on in teh course. Here is a fun fact: By this theorem, you actually learned the Substitution Method backwards!

In 1-dimension, the theorem says

$$\int_I f(x)\,dx = \int_{I^*} f(x(u))x'(u)\,du,$$

where $I = [a,b]$ and $I^* = [u(a), u(b)]$, and the 1-dimensional Jacobian was $x'(u)$ for the coordinate change $T(u) = x(u)$. Some notes:

- You learned it backwards because you prupose back then was to simplify the integrand.
- The purpose in more generality (in multivariable calculus) is to simplfy the integrating region.
- This was never stressed in Calculus I, but even in 1-dimension, the coordinate change to switch to the new variable does need to be 1-1 for the transformation to work correctly. Funny how the examples you worked on always did work out that way...!

New question: So why does the Jacobian arise is the way that it does when changing variables? Here is the answer:

Under the transformation T, a small rectangular region R centered at (u_0, v_0) in D^* given by $\Delta u = u - u_0$ and $\Delta v = v - v_0$ is taken to a new region $R = T(R^*) \subset D$: Here, $\mathbf{aera}(R^*) = \Delta u \Delta v > 0$, and, since T is C^1 and bijective, $\mathbf{area}(R) > 0$ also. We cannot calculate $\mathbf{area}(R)$ directly, but we can approximate it: A linear approximation to $T(u,v)$, near (u_0, v_0) is the function

$$h(u,v) = T(u_0,v_0) + DT(u_0,v_0)\begin{bmatrix} u-u_0 \\ v-v_0 \end{bmatrix}.$$

This linear approximation maps $(u_0, v_0) \mapsto T(u_0, v_0) = (x_0, y_0)$, and takes R^* to a parallelogram $h(R^*)$ which is close to $R = T(R^*)$, so that

$$\mathbf{area}(R) \approx \mathbf{area}\left(h(R^*)\right).$$

We will focus on this later in the course, but it is true that the area of a parallelogram in the plane can be computed by using a form of the cross product adapted to vectors in $\mathbb{R}^2$. Indeed, given two vectors $\mathbf{a}$ and $\mathbf{b}$ in $\mathbb{R}^2$, embed them as vectors in $\mathbb{R}^3$ simply by giving each a 0 in the last coordinate. Then, the area of the parallelogram that has these two vectors as sides is given by the quantity $\|\mathbf{a} \times \mathbf{b}\|$. This is a general feature of the

cross product in R^3. So here, then,

$$\begin{aligned}
\mathbf{area}(R^*) &= \left\| \begin{bmatrix} \Delta u \\ 0 \end{bmatrix} \times \begin{bmatrix} 0 \\ \Delta v \end{bmatrix} \right\| = \Delta u \Delta v, \quad \text{and} \\
\mathbf{area}(h(R^*)) &\approx \left\| h\left(\begin{bmatrix} \Delta u \\ 0 \end{bmatrix} \right) \times h\left(\begin{bmatrix} 0 \\ \Delta v \end{bmatrix} \right) \right\| \\
&= |\mathbf{det}\, DT(u_0, v_0)| \cdot \mathbf{area}(R^*) \\
&= \left| \frac{\partial(x,y)}{\partial(u,v)} \right| \Delta u \Delta v.
\end{aligned}$$

Now consider a cahnge of coordinates from polar to Cartesian: $x = r\cos\theta$ and $y = r\sin\theta$. Then

$$dA = dx\,dy = \left| \frac{\partial(x,y)}{\partial(r,\theta)} \right| dr\,d\theta = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} dr\,d\theta = r\,dr\,d\theta.$$

Hence

$$\iint_D f(x,y)\,dx\,dy = \iint_{D^*} f(r\cos\theta, r\sin\theta) r\,dr\,d\theta,$$

where the region D^* is rectangular in the $r\theta$-plane. This is precisely where the extra r in the integrand comes from when converting to polar coordinates.

Note again, here, though, that the transformation T from polar to Cartesian is not 1-1 on D^*. But it is on the interior, and hence is fine. So why is this the case when integrating??

Recall the famous formula from single variable calculus: Given a function $r = f(\theta)$, one can find the area of the region D formed from the origin to $f(\theta)$ between two values θ_1 and θ_2 as

$$\mathbf{area}(D) = \int_{\theta_1}^{\theta_2} \frac{1}{2} [f(\theta)]^2\, d\theta.$$

So why is this true?

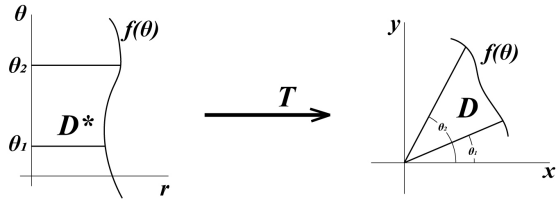


FIGURE 37. The classic formula from Calculus I on the area of a sector capped by $r = f(\theta)$.

The region D is actually written in the Cartesian plane (the xy-plane), as in the right side of Figure 37. Then, under the polar coordinate transformation, D^* looks like the left side. Note that D^* is elementary in θ. Then

the area of D can be calculated:

$$\begin{aligned}\mathbf{area}(D) &= \iint_D 1\,dx\,dy = \iint_{D^*} r\,dr\,d\theta \\ &= \int_{\theta_1}^{\theta_2}\int_0^{f(\theta)} r\,dr\,d\theta = \int_{\theta_1}^{\theta_2}\left[\frac{r^2}{2}\Big|_0^{f(\theta)}\right]d\theta \\ &= \int_{\theta_1}^{\theta_2}\frac{1}{2}\left[f(\theta)\right]^2\,d\theta.\end{aligned}$$

Ans lastly, a quick comment: ALl of this discussion was in the context of functions defined in a region in the plane. But all of this is readily generalizable to $\mathbb{R}^n$: For example, a C^1-transformation in three space would look like

$$T(u,v,w) = \left(x(u,v,w), y(u,v,w), z(u,v,w)\right),$$

and the resulting Jacobian would look like

$$\mathbf{Jac}(T) = \left|\frac{\partial(x,y,z)}{\partial(u,v,w)}\right| = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} & \frac{\partial z}{\partial w} \end{vmatrix}.$$

EXAMPLE 16.4. ***Cartesian-to-Spherical Coordinate Jacobian.*** The *spherical coordinate system* is one of the natural generalizations of polar coordinates in the plane. One set of equations constructing the transformation is $x = \rho\sin\varphi\cos\theta$, $y = \rho\sin\varphi\sin\theta$, and $z = \rho\cos\varphi$. Given a point $\mathbf{p} \in \mathbb{R}^3$ in this coordinate system, ρ is the distance from $\mathbf{p}$ to the origin in the Euclidean metric, φ is angle the ray from $\mathbf{0}$ to $mathbfp$ makes with respect to the positive z-axis, and θ is the angle between the positive x-axis and the ray formed inside the xy-plane between the origin and the projection of $\mathbf{p}$ into the xy-plane inside $\mathbb{R}^3$. If we were to switch from Cartesian coordinates to spherical coordinates in order to integrate, we would need the Jacobian.

We have

$$\left|\frac{\partial(x,y,z)}{\partial(\rho,\varphi,\theta)}\right| d\rho\, d\varphi\, d\theta = \begin{vmatrix} \frac{\partial x}{\partial \rho} & \frac{\partial x}{\partial \varphi} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial \rho} & \frac{\partial y}{\partial \varphi} & \frac{\partial y}{\partial \theta} \\ \frac{\partial z}{\partial \rho} & \frac{\partial z}{\partial \varphi} & \frac{\partial z}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \sin\varphi\cos\theta & \rho\cos\varphi\cos\theta & -\rho\sin\varphi\sin\theta \\ \sin\varphi\sin\theta & \rho\cos\varphi\sin\theta & \rho\sin\varphi\cos\theta \\ \cos\varphi & -\rho\sin\varphi & 0 \end{vmatrix}$$

$$\begin{aligned} &= \cos\varphi \begin{vmatrix} \rho\cos\varphi\cos\theta & -\rho\sin\varphi\sin\theta \\ \rho\cos\varphi\sin\theta & \rho\sin\varphi\cos\theta \end{vmatrix} \\ &\qquad - \left(-\rho\sin\varphi \begin{vmatrix} \sin\varphi\cos\theta & -\sin\varphi\sin\theta \\ \rho\sin\varphi\sin\theta & \rho\sin\varphi\cos\theta \end{vmatrix} \right) \\ &= \rho^2\cos\varphi\left(\sin\varphi\cos\varphi\cos^2\theta + \sin\varphi\cos\varphi\sin^2\theta\right) \\ &\qquad + \rho^2\sin\varphi\left(\sin^2\varphi\cos^2\theta + \sin^2\varphi\sin^2\theta\right) \\ &= \rho^2\sin\varphi\cos^2\varphi + \rho^2\sin\varphi\sin^2\varphi \\ &= \rho^2\sin\varphi. \end{aligned}$$

Hence the integration becomes

$$\iiint_{W^*} f(x,y,z)\, dx\, dy\, dz = \iiint_{W} f\left(\rho\sin\varphi\cos\theta, \rho\sin\varphi\sin\theta, \rho\cos\varphi\right) \rho^2\sin\varphi\, d\rho\, d\varphi\, d\theta.$$

LECTURE 17

The Line Integral

SYNOPSIS. Here, we dive deeper into integration under the idea that in multiple directions, there are more ways to study the properties of functions. The first type is called the line integral, where one integrates over a curve. The two varieties, the *scalar* and *vector line integrals*, have interesting geometric interpretations as well as simple meaning on their own. Today, we define and study these two types.

HELPFUL DOCUMENTS. GIF: `LineIntegralOfScalarField`, `LineIntegralOfVectorField`.

17.0.1. Return to notation. Recall the notation we used when discussing integration:

- In single variable calculus,

$$\int_a^b f(x)\,dx = \int_{\mathcal{I}} f\,dx,$$

  where $x \in \mathcal{I} \subset \mathbb{R}$ and f is assumed to be a function of x.
- In vector calculus to date: For a double integral,

$$\iint_{\mathcal{D}} f\,dA, \quad \mathcal{D} \subset \mathbb{R}^2.$$

  And for triple integrals,

$$\iiint_{\mathcal{W}} f\,dV, \quad \mathcal{W} \subset \mathbb{R}^3,$$

  where f is real-valued, so $f : \mathcal{W} \to \mathbb{R}$.

These are definite integrals whose value would remain unchanged upon a change of variables. Hence the notation denotes a coordinate-free attempt to write the quantities: For $\mathcal{D}$ a region in the xy-plane, for example, then

$$\int_{\mathcal{D}} f\,dA = \iint_{\mathcal{D}} f(x,y)\,dx\,dy.$$

In this lecture, we will define some new ways to study properties of functions over relevant domains in $\mathbb{R}^n$.

17.0.2. Line Integrals.

17.0.2.1. *Real-valued, scalar functions.* Recall that differentiating a C^1-function $f : X \subset \mathbb{R}^n \to \mathbb{R}$ along a curve $\mathbf{x} : I \subset \mathbb{R} \to \mathbb{R}^n$, where $\mathbf{x}(I) \subset x \subset \mathbb{R}^n$, looks a lot like a single variable calculus endeavor:

$$\begin{aligned}\frac{df}{dt}(t) &= \frac{d}{dt} f(\mathbf{x}(t)) = Df(\mathbf{x}(t)) \cdot \frac{d\mathbf{x}}{dt}(t) \\ &= \frac{\partial f}{\partial x_1}\frac{dx_1}{dt} + \cdots + \frac{\partial f}{\partial x_n}\frac{dx_n}{dt} \in \mathbb{R}.\end{aligned}$$

One would expect, then, that integrating a funciton $f : X \subset \mathbb{R}^n \to \mathbb{R}$ over a curve (in essence, adding up its values along the curve) should also seem like a 1-dimenaional calculation. In fact, it is quite similar, once we can interpret things correctly:

DEFINITION 17.1. Given an integrable $f : X \subset \mathbb{R}^n \to \mathbb{R}$, and a C^1-curve $\mathbf{x} : \mathcal{I} \to \mathbb{R}^n$, where $\mathcal{I} = [a, b] \subset \mathbb{R}$ and $\mathbf{x}(\mathcal{I}) \subset X$, the *scalar line integral* of f over $\mathbf{x}$ is

$$\int_{\mathbf{x}} f\, ds = \int_a^b f(\mathbf{x}(t)) \left\|\mathbf{x}'(t)\right\| dt.$$

Some notes:

- The symbolic s denotes arclength: Recall that, for any curve parameterization $\mathbf{x} : [a, b] \to \mathbb{R}$, the arclength parameter is

  $$s(t) = \int_a^t \left\|\mathbf{x}'(\tau)\right\| d\tau.$$

  Seen as a change in variables, the new differential is then, by the Fundamental Theorem of Calculus

  $$ds = s'(t)\, dt = \left(\frac{d}{dt}\left[\int_a^t \left\|\mathbf{x}'(\tau)\right\| d\tau\right]\right) dt = \left\|\mathbf{x}'(t)\right\| dt.$$

  This suggests that the (scalar) line integral is *parameterization independent*. This suggestion is correct.
- The scalar line integral is also sometimes called the *path integral*, or the *line integral of a scalar field*.
- Here is a good geometric interpretation of the scalar line integral, using a function on two variables so that we can see the graph, $f : X \subset \mathbb{R}^2 \to \mathbb{R}$: The $\mathbf{graph}(f)$, as a subset of $\mathbb{R}^3$ defined as the set of solutions to $z = f(x, y)$, sits vertically over X sitting within the xy-plane in $\mathbb{R}^3$. The curve $\mathbf{x}$ sits inside X. Form a vertical wall by drawing a vertical line from each point in $\mathbf{x} \subset X$ to $\mathbf{graph}(\mathbf{x}) \subset \mathbf{graph}(f)$. Then the scalar line integral of f along $\mathbf{x}$ is the total area of this wall. This is simply the curved version of the standard geometric interpretation of the integral of a function $f(x)$ over an interval in Calculus I. See Figure **??**.
- So, again, the scalar line integral of f along $\mathbf{x}$ is essentially a 1-dimensional calculation, but now in $\mathbb{R}^n$, $n > 1$.

EXAMPLE 17.1. ***A constant function along a straight line.*** Let $f(x,y) = 4$, and $\mathbf{x}(t) = \begin{bmatrix} 3 \\ 4 \end{bmatrix} t$, defined on $\mathcal{I} = [0,1]$. Then

$$\begin{aligned}\int_{\mathbf{x}} f\,ds &= \int_0^1 f\left(\mathbf{x}(t)\right)\left\|\mathbf{x}'(t)\right\|\,dt \\ &= \int_0^1 f\left(3t,4t\right)\sqrt{3^2+4^2}\,dt \\ &= \int_0^1 4\sqrt{25}\,dt = 20t\Big|_0^1 = 20.\end{aligned}$$

Perhaps this was rather obvious, since the rectangle (in $\mathbb{R}^3$), whose base if $\mathbf{x}([0,1])$, and whose height is 4 has area which is the length of the curve times the height: The length of the curve is 5, so the area is $4 \cdot 5 = 20$.

EXAMPLE 17.2. ***Surface area of a cylinder.*** The surface area SA of a cylinder of radius r and height h is $SA = 2\pi rh$ (there is no top nor bottom to this cylinder). We can functionally calculate this: Let $g(x,y) = h$, and define $\mathbf{c}(t) = \begin{bmatrix} r\cos t \\ r\sin t \end{bmatrix} t$ be a curve, defined on $\mathcal{I} = [0,2\pi]$. Then

$$\int_{\mathbf{c}} g\,ds = \int_0^{2\pi} h\sqrt{r^2\cos^2 t + r^2\sin^2 t}\,dt = \int_0^{2\pi} hr\,dt = 2\pi rh.$$

EXAMPLE 17.3. ***A curve in*** $\mathbb{R}^3$***.*** Integrate $f(x,y,z) = \frac{xz}{y}$ over $\mathbf{x}(t) = \begin{bmatrix} 4 \\ 2t \\ \frac{t^2}{2} \end{bmatrix}$, defined on $\mathcal{I} = [1,2]$. First, we calculate

$$\left\|\mathbf{x}'(t)\right\| = \sqrt{0^2+2^2+t^2} = \sqrt{4+t^2}.$$

Then we can calculate the integral:

$$\int_{\mathbf{x}} f\,ds = \int_1^2 \frac{(4)\left(\frac{t^2}{2}\right)}{2t}\sqrt{4+t^2}\,dt = \int_1^2 2t\sqrt{4+t^2}\,dt.$$

With the substitution $u = 4+t^2$, along with $du = 2t\,dt$, we get

$$\int_1^2 2t\sqrt{4+t^2}\,dt = \int_5^8 \sqrt{u}\,du = \frac{2}{3}u^{\frac{3}{2}}\Big|_5^8 = \frac{2}{3}\left(16\sqrt{2} - 5\sqrt{5}\right).$$

17.0.2.2. *Real-valued, vector functions (vector fields).* For a curve $\mathbf{x} : \mathcal{I} \to \mathbb{R}^n$, $\mathcal{I} = [a, b] \subset \mathbb{R}$, where $\mathbf{x}(I) \subset x \subset \mathbb{R}^n$, inside a vector field $\mathbf{F} : X \subset \mathbb{R}^n \to \mathbb{R}^n$ (it is again understood that the entire curve is in X), one can ask how much of the vector field can be "seen" by (a point on) the curve. Recall for a curve where $\mathbf{x}'(t) \neq 0$, $\forall t \in \mathcal{I}$, that $\mathbf{T}(t) = \frac{\mathbf{x}'(t)}{\|\mathbf{x}'(t)\|}$ is the unit tangent vector at t. Then the component of $\mathbf{F}$ along $\mathbf{x}$ is

$$\mathbf{F}(\mathbf{x}(t)) \cdot \mathbf{T}(t).$$

We can, again, "add up" these values all along the curve to get

$$\int_{\mathbf{x}} (\mathbf{F} \cdot \mathbf{T})\, ds,$$

the scalar line integral of $\mathbf{F} \cdot \mathbf{T}$, as a scalar-valued function, along $\mathbf{x}$.

Note, that, in a way, this represents the aggregate boost or hindrance a particle moving along the curve would feel by the vector field.

But

$$\begin{aligned}\int_{\mathbf{x}} (\mathbf{F} \cdot \mathbf{T})\, ds &= \int_a^b \left(\mathbf{F}(\mathbf{x}(t)) \cdot \frac{\mathbf{x}'(t)}{\|\mathbf{x}'(t)\|}\right) \left\|\mathbf{x}'(t)\right\| dt \\ &= \int_a^b \mathbf{F}(\mathbf{x}(t)) \cdot \mathbf{x}'(t)\, dt \\ &= \int_a^b \mathbf{F}(\mathbf{x}(t)) \cdot d\mathbf{s} = \int_{\mathbf{x}} \mathbf{F} \cdot d\mathbf{s},\end{aligned}$$

where

$$d\mathbf{s} = \begin{bmatrix} dx_1 \\ \vdots \\ dx_n \end{bmatrix} = \begin{bmatrix} x_1'(t)\, dt \\ \vdots \\ x_n'(t)\, dt \end{bmatrix} = \begin{bmatrix} x_1'(t) \\ \vdots \\ x_n'(t) \end{bmatrix} dt = \mathbf{x}'(t)\, dt$$

is the vector infinitesimal displacement and represents an infinitesimal change in displacement along each coordinate direction (instead of just along the curve).

Definition 17.2. For a C^1-curve $\mathbf{x} : [a, b] \to X \subset \mathbb{R}^n$ in a vector field $\mathbf{F} : X \subset \mathbb{R}^n \to \mathbb{R}^n$, the *vector line integral* of $\mathbf{F}$ along $\mathbf{x}$ is

$$\int_{\mathbf{x}} \mathbf{F} \cdot d\mathbf{s} = \int_a^b \mathbf{F}(\mathbf{x}(t)) \cdot \mathbf{x}'(t)\, dt.$$

Some notes:

(1) Recall that the work W done by a vector field $\mathbf{F}$ (force field) on a particle is just the force times the displacement $\mathbf{d}$. As vector quantities, when the particle's motion is
   - linear:

$$W = \mathbf{F} \cdot \mathbf{d} = \|\mathbf{F}\| \|\mathbf{d}\| \cos\theta = \|\mathbf{F}\| \cdot (\text{displacement in direction of } \mathbf{F}).$$

- curved: Here we must measure infinitesimally, and

$$W = \int_a^b \mathbf{F}(\mathbf{x}(t)) \bullet \mathbf{x}'(t)\, dt,$$

a scalar integral.

(2) Some facts:

THEOREM 17.3. *A scalar line integral is independent of the curve's parameterization.*

THEOREM 17.4. *A vector line integral depends on a curve's parameterization only in the direction of travel.*

(3) Some curve facts:
- For $\mathcal{I} = [a, b]$, and $\mathbf{x} : \mathcal{I} \to \mathbb{R}^n$, a piecewise C^1-continuous curve, a C^1-function $h : \mathcal{J} = [c, d] \to \mathcal{I}$, which is 1-1, onto, and has a C^1-inverse, induces $\mathbf{p} : \mathcal{J} \to \mathbb{R}^n$, $\mathbf{p} = \mathbf{x} \circ h$ a *reparameterization.*
- For $\mathbf{x}$ 1-1 on $\mathcal{I}$, then $\mathbf{x}(\mathcal{I})$ has only two directions of travel. A choice of direction, given by the notion of increasing values of t along $\mathbf{x}$, is called an *orientation* of the curve.
- A reparameterization is called *oriantation preserving* if the direction of travel along $\mathbf{p}(\mathcal{J})$ is the same as that along $\mathbf{x}(\mathcal{I})$. Otherwise, the reparameterization is called *orientation reversing.*
- Again, for $\mathcal{I} = [a, b]$, a curve $\mathbf{x} : \mathcal{I} \to \mathbb{R}^n$ is called *simple* if $\mathbf{x}$ is 1-1 on (a, b), and *closed* is $\mathbf{x}(a) = \mathbf{x}(b)$.
- In general, scalar line (path) integrals are defined on curves, and vector line integrals are defined on oriented curves.

(4) If $\mathbf{x}$ is a *simple, closed curve*, then the notation for a vector line integral is

$$\oint_{\mathbf{x}} \mathbf{F} \bullet d\mathbf{s}$$

and is called the *circulation* of $\mathbf{F}$ along $\mathbf{x}$. We will see the geometric interpretation of this in time.

(5) For a vector line integral, the term

$$\mathbf{F} \bullet d\mathbf{s} = F_1\, dx_1 + \ldots + F_n\, dx_n = \sum_{i=1}^{n} F_i\, dx_i$$

is called a *differential 1-form.* We will devote more time to this later.

EXAMPLE 17.4. ***Evaluate the integral***

$$\oint_{\mathbf{c}} (x^2 - y^2)\, dx + (x^2 + y^2)\, dy,$$

where* $\mathbf{c}$ *is the boundary of the unit square, oriented clockwise. Recognizing that this integral is simply a vector line integral of the vector field $\mathbf{F} = (x^2 - y^2)\mathbf{i} + (x^2 + y^2)\mathbf{j}$ over the closed, simple curve $\mathbf{c}$ given by the edge of the unit square, one sees that

$$(x^2 - y^2)\,dx + (x^2 + y^2)\,dy = \mathbf{F} \bullet d\mathbf{s}$$

is just a differentiable 1-form. The process here would be, then, the parameterize the unit square perimeter by time, and integrate under the parameterization: We get

$$\mathbf{c}(t) = \begin{cases} (0,t) & 0 \le t \le 1 \\ (t-1,1) & 1 \le t \le 2 \\ (1,3-t) & 2 \le t \le 3 \\ (4-t,0) & 3 \le t \le 4 \end{cases}$$

as our clockwise parameterization, beginning and ending at the origin. To understand the switch to the parameterization, we highlight the first "piece": Along the left-side edge of the unit square only, the parameterization is the path $\mathbf{c}_1$, going from $(0,0)$ to $(0,1)$ and parameterized by t in the y-direction only. We get

$$\begin{aligned} \int_{\mathbf{c}_1} \mathbf{F} \bullet d\mathbf{s} &= \int_{\mathbf{c}_1} (x^2 - y^2)\,dx + (x^2 + y^2)\,dy \\ &= \int_0^1 F_1(x(t), y(t))x'(t)\,dt + F_2(x(t), y(t))y'(t)\,dt \\ &= \int_0^1 \left((0)^2 - (t)^2\right)(0\,dt) + \left((0)^2 + (t)^2\right)(1\,dt) \\ &= \int_0^1 t^2\,dt = \frac{t^3}{3}\bigg|_0^1 = \frac{1}{3}. \end{aligned}$$

Hence on the four pieces (so once around the square), we get

$$\begin{aligned} \oint_{\mathbf{c}} \mathbf{F} \bullet d\mathbf{s} &= \oint_{\mathbf{c}} (x^2 - y^2)\,dx + (x^2 + y^2)\,dy \\ &= \int_0^1 t^2\,dt + \int_1^2 \left((t-1)^2 - 1^2\right)dt + \int_2^3 \left(1^2 + (3-t)^2\right)dt + \int_3^4 (4-t)^2\,dt \\ &= \int_0^1 t^2\,dt + \int_1^2 \left(t^2 - 2t\right)dt + \int_2^3 \left(10 - 6t + t^2\right)dt + \int_3^4 \left(16 - 8t + t^2\right)dt \\ &= \frac{1}{3} + \left(\frac{t^3}{3} - t^2\right)\bigg|_1^2 + \left(10t - 3t^2 + \frac{t^3}{3}\right)\bigg|_2^3 + \left(16t - 4t^2 + \frac{t^3}{3}\right)\bigg|_3^4 \\ &= \frac{1}{3} + \left(\frac{8}{3} - 4 - \frac{1}{3} + 1\right) + \left(30 - 27 + 9 - 20 + 12 - \frac{8}{3}\right) + \left(64 - 64 + \frac{64}{3} - 48 + 36 - 9\right) \\ &= \frac{1}{3} - \frac{2}{3} + \frac{4}{3} + \frac{1}{3} = \frac{4}{3}. \end{aligned}$$

EXAMPLE 17.5. ***Calculate the circulation of*** $\mathbf{F} = x\mathbf{i} + y\mathbf{j}$ ***over the unit circle in the plane.*** Here, we can parameterize the unit circle as the simple closed curve $\mathbf{x}(t) = \begin{bmatrix} \cos t \\ \sin t \end{bmatrix}$, for t on the interval $t \in [0, 2\pi]$. So

$$\begin{aligned} \oint_{\mathbf{x}} \mathbf{F} \cdot d\mathbf{s} &= \int_0^{2\pi} \mathbf{F}(\mathbf{x}(t)) \cdot \mathbf{x}'(t)\, dt \\ &= \int_0^{2\pi} \left(\begin{bmatrix} \cos t \\ \sin t \end{bmatrix} \cdot \begin{bmatrix} -\sin t \\ \cos t \end{bmatrix} \right) dt = \int_0^{2\pi} 0\, dt = 0. \end{aligned}$$

Note that, here, the vector field is perpendicular to the curve everywhere. Hence there is no circulation of $\mathbf{F}$ along $\mathbf{x}$ in this case.

EXAMPLE 17.6. ***Circulation of a vector field along one if its integral curves.*** By definition, if $\mathbf{x}$ is an integral curve of a vector field $\mathbf{F}$ on some interval $[a, b]$, then for all $t \in [a, b]$, we have $\mathbf{x}'(t) = \mathbf{F}(\mathbf{x}(t))$. So

$$\begin{aligned} \int_{\mathbf{x}} \mathbf{F} \cdot d\mathbf{s} &= \int_a^b \mathbf{F}(\mathbf{x}(t)) \cdot \cdot \mathbf{x}'(t)\, dt \\ &= \int_a^b \mathbf{x}'(t) \cdot \mathbf{x}'(t)\, dt = \int_a^b \left\| \mathbf{x}'(t) \right\|^2 dt. \end{aligned}$$

In this case, we are again simply adding up the vector field components along the curve, but here they equal the velocity vectors all along the curve.

LECTURE 18

The Theorem of Green

SYNOPSIS. Today, we go directly into on of the three big theorem's of vector calculus, Green's Theorem. This theorem exposes a deep relationship between the aggregate behavior of a vector field along the boundary of a relatively nice region in the plane (the vector line integral), to the integral of a related function on the interior of the region. Since Green's Theorem is restricted to regions in the plane, there are a number of ways to craft the related integrals, giving different geometric meaning to the quantities. One interesting geometric interpretation is that the theorem relates the total twisting effect of a vector field in the region (measured by integrating the curl of the vector field as it sits in three space with no vertical component), to the total tangent component of the vector field along the closed boundary. Proving this theorem is neither deep nor long, and we go over the idea here in lecture. We finish with a general definition and discussion of the properties of a special kind of vector field that shows up in an lot of physical applications.

18.0.1. Green's Theorem. We begin directly with the theorem:

THEOREM 18.1. *Let $\mathcal{D}$ be a closed, bounded region in $\mathbb{R}^2$, whose boundary $\mathbf{c} = \partial\mathcal{D}$ is a finite union of simple, closed, curves, oriented so that $\mathcal{D}$ is always on the left. For a C^1-vector field on $\mathcal{D}$ $\mathbf{F}(x,y) = M(x,y)\mathbf{i} + N(x,y)\mathbf{j}$, we have*

(18.0.1)
$$\left(\int_{\mathbf{c}} \mathbf{F}\bullet\, d\mathbf{s} =\right) \oint_{\mathbf{c}} M\,dx + N\,dy = \iint_{\mathcal{D}} \left(\frac{\partial N}{\partial x} - \frac{\partial M}{\partial y}\right) dx\,dy \left(= \iint_{\mathcal{D}} (\nabla \times \mathbf{F})\bullet \mathbf{k}\, dA\right).$$

Some notes:

- The first equal sign is obvious since $\mathbf{F} = \begin{bmatrix} M \\ N \end{bmatrix}$ and $d\mathbf{s} = \begin{bmatrix} dx \\ dy \end{bmatrix}$.
- The last equal sign is also obvious since $(\nabla \times \mathbf{F}) = \left(\frac{\partial N}{\partial x} - \frac{\partial M}{\partial y}\right)\mathbf{k}$. To see this, think of a vector field in $\mathbb{R}^2$ as a vec tor field in $\mathbb{R}^3$ with z-component 0. Then calculate $\nabla \times \mathbf{F}$.
- The theorem basicallt says that the vector line integral (the circulation) of $\mathbf{F}$ along $\partial\mathcal{D}$ equals the curl of $\mathbf{F}$ on $\mathcal{D}$.
  (1) $\displaystyle\int_{\mathbf{c}} \mathbf{F}\bullet\, d\mathbf{s}$ measures the aggregate component of $\mathbf{f}$ tangent to $\mathbf{c}$ and in the direction of travel along $\mathbf{c}$.

(2) Curl in two dimensions measures the rotation or twisting one would experience if flowing along $\mathbf{F}$.

So the sum total of the push or pull of a particle by $\mathbf{F}$ along $\mathcal{D}$ equals the total rotational effect of $\mathbf{F}$ on $\mathcal{D}$.

EXAMPLE 18.1. ***A constant vector field.*** For a constant vector field $\mathbf{F}$, we know $\nabla \times \mathbf{F} = \mathbf{0}$. So what can we say about what is happening on the left-hand side of Equation 18.0.1? How about in the middle?

- The proof is elementary, and relies on three facts:
  (1)

  LEMMA 18.2. *If $\mathcal{D}$ is elementary of Type I, then* $\oint_{\partial\mathcal{D}} M\,dx = -\iint_{\mathcal{D}} \frac{\partial M}{\partial y}\,dA$.

  (2)

  LEMMA 18.3. *If $\mathcal{D}$ is elementary of Type II, then* $\oint_{\partial\mathcal{D}} N\,dy = \iint_{\mathcal{D}} \frac{\partial N}{\partial x}\,dA$.

  (3) Any region $\mathcal{D}$, valid for Green's Theorem, can be cut up into a finite number of elementary regions, so that (i) the ends of each cut line are in $\partial\mathcal{D}$, (ii) the cuts do not intersect, (iii) $\mathcal{D} = \bigcup \mathcal{D}_i$, with each $\mathcal{D}_i$ elementary of Type III, and (iv) each cut intersects exactly 2 $\mathcal{D}_i$'s with each cut oriented in each $\mathcal{D}_i$ oppositely. Note there, that then, vector line integral along the cut lines will cancel out. So there will be no contribution of the cut lines in the calculation. And within the double integral, the contributions of all cut lines will also be 0.

Here is the idea of the proof of Lemma 18.2:

PROOF OF LEMMA 18.2. With $\mathcal{D}$ elementary of Type I, write

$$\mathcal{D} = \left\{(x,y) \in \mathbb{R}^2 \;\middle|\; a \le x \le b,\ \alpha(x) \le y\beta(x)\right\}.$$

Orient $\partial\mathcal{D} = \mathbf{c}_1^+ \cup \mathbf{d}_1^+ \cup \mathbf{c}_2^- \cup \mathbf{d}_2^-$, as needed, where the plus sign means compatible with variable values, and the minus sign means contrary to parameter values.

Then, on the right-hand side, we have

$$\begin{aligned} -\iint_{\mathcal{D}} \frac{\partial M}{\partial y}\, dA &= \int_a^b \int_{\alpha(x)}^{\beta(x)} -\frac{\partial M}{|partialy}\, dy\, dx \\ &= -\int_a^b \left(M(x,\beta(x)) - M(x,\alpha(x))\right)\, dx \\ &= \int_a^b \left(M(x,\alpha(x)) - M(x,\beta(x))\right)\, dx \end{aligned}$$

by the Fundamental Theorem of Calculus, suitably adapted. But here

$$-\int_a^b M(x,\beta(x))\, dx = \int_{\mathbf{c}_2^-} M\, dx, \quad \text{and} \quad \int_a^b M(x,\alpha(x))\, dx = \int_{\mathbf{c}_1^+} M\, dx.$$

Now add to this the fact that

$$\int_{\mathbf{d}_1^+} M\, dx = \int_{\mathbf{d}_2^-} M\, dx = 0$$

since x is constant along these curves, and we have

$$\begin{aligned} -\iint_{\mathcal{D}} \frac{\partial M}{\partial y}\, dA &= \int_{\mathbf{c}_1^+} M\, dx + \int_{\mathbf{d}_1^+} M\, dx + \int_{\mathbf{c}_2^-} M\, dx + \int_{\mathbf{d}_2^-} M\, dx \\ &= \oint_{\partial \mathcal{D}} M\, dx. \end{aligned}$$

□

Exercise 10. Prove Lemma 18.3.

18.0.2. Conservative Vector Fields. Here, we have a definition:

Definition 18.4. A C^0-vector field $\mathbf{F}$ has *path-independent line integrals* if for any two piecewise C^1-curves with the same endpoints $\mathbf{x}_1$ and $\mathbf{x}_2$, we have

$$\int_{\mathbf{x}_1} \mathbf{F} \bullet d\mathbf{s} = \int_{\mathbf{x}_2} \mathbf{F} \bullet d\mathbf{s}.$$

Now notice that, since the two curves have the same endpoints, they together form a closed curve (traversing one of them backwards, that is). Sometimes this closed curve is simple.

Theorem 18.5. *A C^0-vector field $\mathbf{F}$ has path-independent line integrals iff $\int_{\mathbf{c}} \mathbf{F} \bullet d\mathbf{s} = 0$ for every piecewise C^1-simple, closed curve $\mathbf{c}$ in the domain of $\mathbf{F}$.*

Note: If $\mathbf{x}_1^+ \cup \mathbf{x}_2^-$ is not simple, yet consists of a finite number of isolated intersections and/or intervals that coincide, then the result is still true. Why is this the case? Can you reason it through?

DEFINITION 18.6. A C^0-vector field $\mathbf{F}$ is called *conservative* or a *gradient field* if there is a C^1-real valued function f, where $\mathbf{F} = \nabla f$. Such an f is called a *potential* for $\mathbf{F}$, adn is said to *generate* the vector field $\mathbf{F}$.

Notes:

- Conservative vector fields always and only (read: iff) have path-independent line integrals. Indeed,
$$\begin{aligned}\int_{\mathbf{x}} \mathbf{F} \bullet d\mathbf{s} &= \int_{\mathbf{x}} \nabla f \bullet d\mathbf{s} = \int_a^b \nabla f\left(\mathbf{x}(t)\right) \bullet \mathbf{x}'(t)\, dt \\ &= \int_a^b \frac{d}{dt}\left[f\left(\mathbf{x}(t)\right)\right] dt \\ &= f\left(\mathbf{x}(t)\right)\Big|_a^b = f\left(\mathbf{x}(b)\right) - f\left(\mathbf{x}(a)\right),\end{aligned}$$
depends only on the endpoints of the path. This is Theorem 6.3.3 in the text. And when the curve is closed, the endpoints are the same, so....
- In $\mathbb{R}^2$ and $\mathbb{R}^3$, conservative vector fields are irrotational: If $\mathbf{F}$ is conservative, then $\nabla \times \mathbf{F} = \nabla \times \nabla f = \mathbf{0}$. The converse of this statement, suitably modified, is: If $\mathbf{F}$ is irrotational, AND the domain of $\mathbf{F}$ is *simply connected*, then $\mathbf{F}$ is conservative. This is Theorem 6.3.5 in the text.

DEFINITION 18.7. A region in either $\mathbb{R}^2$ or $\mathbb{R}^3$ is *simply connected* if it is connected (comes in one piece) and every simply closed curve in the region has its entire interior inside the region.

So a disk in the plane in simply connected, but an annulus in the plane is not.

EXAMPLE 18.2. ***A nonconservative vector field on a non-simply connected domain.*** Let
$$\mathbf{F} = \frac{-y}{x^2+y^2}\mathbf{i} + \frac{x}{x^2+y^2}\mathbf{j} + 0\mathbf{k}$$
on $\mathcal{W} = \mathbb{R}^3 - \left\{(0,0,z) \in \mathbb{R}^3 \,\middle|\, z \in \mathbb{R}\right\}$. Here, $\mathbf{F}$ is irrotational ($\nabla \times \mathbf{F} = \mathbf{0}$), but $\mathbf{F}$ is not conservative. It is also the case, though, that $\mathcal{W}$ is not simply connected, as the unit circle in the xy-plane inside $\mathbb{R}^3$ cannot be continuously shrunk to a point in $\mathcal{W}$. I will offer this as an exercise to show that $\mathbf{F}$ is not conservative. As a hint, find a simple, closed curve where the vector line integral of $\mathbf{F}$ along the curve is not 0.

So a good question is: how does one tell if a vector field is conservative? The answers are varied, but include tools like (1) checking the mixed partials of what the potential function would be, (2) integrating to attempt to find

the potential, and (3) searching for a closed simple curve on which the vector line integral of the vector field does not vanish.

EXAMPLE 18.3. ***Find a potential for*** $\mathbf{F} = (y+z)\mathbf{i} + 2z\mathbf{j} + (x+y)\mathbf{k}$***, if conservative.*** The solution here: If a potential f exists, then $\frac{\partial f}{\partial x} = y + z$. But this means that $f(x, y, z) = xy + xz + g(y, z)$. But then

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

Do you see why this cannot happen? Also,

$$\nabla \times \mathbf{F} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ y+z & 2z & x+y \end{vmatrix} = (1-2)\mathbf{i} + \ldots$$

and at this point, we can halt the calculation, since the vector field is not irrotational, and hence cannot be conservative.

EXAMPLE 18.4. ***Find a potential for*** $\mathbf{F} = (2x+y)\mathbf{i} + (z\cos yz + x)\mathbf{j} + (y\cos yz)\mathbf{k}$**.** Here, one first step is to identify that if a function $f : \mathbb{R}^3 \to \mathbb{R}$ exists so that $\nabla f = \mathbf{F}$, then $\frac{\partial f}{\partial x} = 2x + y$. Thus, we can see that

$$f(x, y, z) = x^2 + xy + g(y, z)$$

by anti-differentiation. Thus,

$$\frac{\partial f}{\partial y} = x + \frac{\partial g}{\partial y}(y, z) = x + z\cos yz,$$

so that $\frac{\partial g}{\partial y}(y, z) = z\cos yz$ and $g(y, z) = \sin yz + h(z)$. We now know that

$$f(x, y, z) = x^2 + xy + \sin yz + h(z).$$

Lastly, we have

$$\frac{\partial f}{\partial z} = y\cos yz + h'(z) = y\cos yz$$

so that $h'(z) = 0$, and thus $h(z) = \text{const}$. Thus we can say that

$$f(x, y, z) = x^2 + xy + \sin yz.$$

LECTURE 19

Surface Parameterizations

SYNOPSIS. The two dimensional counterpart to a curve in n-space is a surface in n-space, and today we define and discuss the properties of parameterized surfaces in (mostly) three-space (and sometimes n-space.) The parallels to curves will be obvious, and discussing these parallels will bring up very interesting contrasts, which we will highlight. Then we will begin the discussion of how a parameterization of a surface in space allows us to discuss the properties of the surface, including how functions behave when defined along the surface.

HELPFUL DOCUMENTS. Mathematica: `ParameterizedSurfaces`.

19.0.1. Coordinates on a surface. We typically parameterize a curve in $\mathbb{R}^n$ via a map $\mathbf{c} : \mathcal{I} \to \mathbb{R}$, where $\mathcal{I} = [a, b] \subset \mathbb{R}$, and

$$\mathbf{c}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix} \in \mathbb{R}^n.$$

For $n = 3$, we would usually write the three coordinate functions as $x(t)$, $y(t)$, and $z(t)$, all real-valued functions on $\mathcal{I}$.

Now let $\mathcal{D} \subset \mathbb{R}^2$ be a connected, open set, along with some or all of its boundary points. A parameterized surface in $\mathbb{R}^n$ is then a C^0-function $\mathbf{X} : \mathcal{D} \to \mathbb{R}^n$ $(n > 2)$, that is one-to-one on the interior of $\mathcal{D}$. The corresponding image of $\mathbf{X}$ is then

$$\mathbf{X}(s, t) = \begin{bmatrix} x_1(s, t) \\ x_2(s, t) \\ \vdots \\ x_n(s, t) \end{bmatrix} \in \mathbb{R}^n.$$

EXAMPLE 19.1. Let $\mathcal{D} = \left\{(s, t) \in \mathbb{R}^2 \,\middle|\, 0 \le s \le 2\pi,\ 0 \le t \le \pi\right\} = [0, 2\pi] \times [0, \pi]$ in the st-plane. Then the map $\mathbf{X} : \mathcal{D} \to \mathbb{R}^3$, defined by

$$\mathbf{X}(s, t) = \begin{bmatrix} x(s, t) \\ y(s, t) \\ z(s, t) \end{bmatrix} = \begin{bmatrix} a \cos s \sin t \\ a \sin s \sin t \\ a \cos t \end{bmatrix}$$

has image S^2_a the sphere of radius a centered at the origin in $\mathbb{R}^3$. In this parameterization, every point on the interior of $\mathcal{D}$ is mapped uniquely to a

point on $S_a^2 = \mathbf{X}(\mathcal{D})$. However, the top and bottom edges of $\mathcal{D}$ are each all mapped to a point (the north and south poles, respectively, while the two walls are together mapped to a line from the north pole to the south pole. Perhaps we can call that line the "seam" of the ball? See Figure **??**.

Note: Just as a curve sits inside $\mathbb{R}^n$, $n > 2$, a surface can have $\mathbb{R}^n$, $n > 3$ as a codomain. For now, we will restrict our examples to three space $\mathbb{R}^3$ for the convenience and clarity of visualization.

EXAMPLE 19.2. ***Graphs of functions are parameterizations.*** Any $f : \mathcal{D} \subset \mathbb{R}^2 \to \mathbb{R}$ has $\mathbf{graph}(f) \subset \mathbb{R}^3$ parameterized directly by the coordinates of $\mathcal{D}$, where $(x, y) \mapsto (x, y, f(x, y))$. The points in the domain parameterize $\mathbf{graph}(f) \subset \mathbb{R}^3$ because the third coordinate is uniquely specified once the first two are known.

EXAMPLE 19.3. ***The square torus.*** The *2-torus* T^2 has a nice description as the product of two copies of the circle $T^2 = S^1 \times S^1$, where each coordinate is an angular one. This is fundamentally different form the two angular coordinates that comprise the 2-sphere, though. Here, let $\mathcal{D} = [0, 2\pi] \times [0, 2\pi] = [0, 2\pi]^2$. Then, for $a > b > 0$ positive constants, the map $\mathbf{X} : \mathcal{D} \to \mathbb{R}^3$, defined by

$$\mathbf{X}(s,t) = \begin{bmatrix} x(s,t) \\ y(s,t) \\ z(s,t) \end{bmatrix} = \begin{bmatrix} (a + b\cos t)\cos s \\ (a + b\cos t)\sin s \\ b\sin t \end{bmatrix}$$

has the shape of the surface of a doughnut. Here b is the cross-sectional radius and a is the radial distance from the z-axis to the center of any corss-sectional circle. The torus T^2 is often described as simply the unit square in the plane with the "opposite sides identified".

DEFINITION 19.1. A surface $\mathcal{S} = \mathbf{X}(\mathcal{D})$ is *differentiable* if its coordinate functions are. When this is the case, then

$$\mathbf{X}_s(s_0, t_0) = \frac{\partial \mathbf{X}}{\partial s}(s_0, t_0) = \begin{bmatrix} \frac{\partial x}{\partial s}(s_0, t_0) \\ \frac{\partial y}{\partial s}(s_0, t_0) \\ \frac{\partial z}{\partial s}(s_0, t_0) \end{bmatrix}.$$

The same is true for $\mathbf{X}_t(s_0, t_0)$.

If these two first partials are continuous functions, then the derivative matrix $D\mathbf{X} = [\mathbf{X}_s, \mathbf{X}_t]$ exists and is a 3×2 matrix. Each of $\mathbf{X}_s$ and $\mathbf{X}_t$ is

a vector of functions, and when evaluated at the point (s_0, t_0), each represents a vector tangent to the embedded surface $\mathcal{S} = \mathbf{X}(\mathcal{D})$, at the point $\big(x(s_0,t_0), y(s_), t_0), z(x_0,t_0)\big)$. Hence $\mathbf{X}_s(s_0,t_0)$ and $\mathbf{X}_t(s_0,t_0)$ are member of the tangent space to $\mathcal{S}$ at (s_0,t_0), where s, t are parameter coordinates on the surface, and $(x(s_0,t_0), y(s_0,t_0, z(s_0,t_0))$ is the corresponding point in the ambient coordinates in $\mathbb{R}^3$.

Now as $\mathbf{X}_s$ and $\mathbf{X}_t$ are always members of the tangent space to $\mathbf{X}(\mathcal{D}) \subset \mathbb{R}^3$, then $\mathbf{X}_s \times \mathbf{X}_t$ is normal to the surface (when it is nonzero and the surface is C^1, that is). Call this normal vector $\mathbf{N}(s_0,t_0) = (\mathbf{X}_s \times \mathbf{X}_t)(s_0,t_0)$.

DEFINITION 19.2. The surface $\mathcal{S} = \mathbf{X}(\mathcal{D})$ is called *smooth at the point* $\mathbf{X}(s_0,t_0)$ if $\mathbf{X}$ is C^1 in an open neighborhood of (s_0,t_0) and if $\mathbf{N}(s_0,t_0) \neq \mathbf{0}$. The surface $\mathcal{S}$ is called *smooth* if it is smooth everywhere.

Note that C^1 and smoothness ensure that the embedded surface has no sharp edges only if the normal vector is nonzero. This is similar to the image of a parameterized curve, where if the parameter function is differentiable, one may still have a corner in the image of the curve if the tangent vector is $\mathbf{0}$.

EXAMPLE 19.4. For $S_a^2 = \mathbf{X}(\mathcal{D})$, given in Example 19.1 above, we have

$$\mathbf{X}_s = \begin{bmatrix} -a\sin s \sin t \\ a\cos s \sin t \\ 0 \end{bmatrix}, \quad \text{and} \quad \mathbf{X}_t = \begin{bmatrix} a\cos s\cos t \\ a\sin s \cos t \\ -a\sin t \end{bmatrix}.$$

Hence $\mathbf{N} = \mathbf{X}_s \times \mathbf{X}_t = -a\sin t \begin{bmatrix} x \\ y \\ z \end{bmatrix}$. Hence, under this parameterization, $\mathbf{N}$ is nonzero everywhere except for when $t = 0, \pi$, so everywhere except for at the poles.

EXERCISE 11. Do the calculation that shows that $\mathbf{N} = \mathbf{X}_s \times \mathbf{X}_t = -a\sin t \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ in Example 19.4 above.

However, S_a^2 is smooth everywhere, even at the poles. But not according to this parameterization. To see the poles as smooth, one would have to reparameterize so that the points corresponsding to the poles lie somewhere inside the corresponding region domain $\mathcal{D}$. In a sense, the parameterization we specified does give meaning to the phrase "One cannot walk east or west when standing at the north pole. One can only walk south!"

DEFINITION 19.3. A *piecewise smooth* parameterized surface is the union of images of finitely many parameterized surfaces $\mathbf{X}_i : \mathcal{D}_i \to \mathbb{R}^3$, where

each $\mathcal{D}_i$ is (1) elementary, (2) C^1 except possibly along $\partial\mathcal{D}_i$, and (3) each $\mathcal{S}_i = \mathbf{X}_i(\mathcal{D}_i)$ is smooth except at possibly a finite set of points.

19.0.2. Surface area. Recall that the length of a parameterized curve $\mathbf{c} : [a, b] \to \mathbb{R}^n$ can be calculated using the parameterization

$$\int_a^b \|\mathbf{c}'(t)\| \, dt,$$

even as the actual length of the curve in $\mathbb{R}^n$ is independent of the parameterization. In 2-dimensions, we can develop a similar construction: Given a surface parameterization $\mathbf{X} : \mathcal{D} \to \mathbb{R}^n$, for a region $\mathcal{D}$, a small rectangle $\mathcal{R} \subset \mathcal{D}$, based at a point (s_0, t_0) and of size Δs and Δt, will have image $\mathbf{X}(\mathcal{R})$ a small region inside the surface $\mathcal{S} = \mathbf{X}(\mathcal{D})$. This image will most likely not be a parallelogram. But it can be approximated by one with sides $\mathbf{X}_s(s_0, t_0)\Delta s$ and $\mathbf{X}_t(s_0, t_0)\Delta t$. Then, the area of this image region is

$$\begin{aligned}\textbf{area}\,(\mathbf{X}(\mathcal{R})) &\approx \|\mathbf{X}_s(s_0, t_0)\Delta s \times \mathbf{X}_t(s_0, t_0)\Delta t\| \\ &= \|\mathbf{X}_s(s_0, t_0) \times \mathbf{X}_t(s_0, t_0)\| \, \Delta s \Delta t = \|\mathbf{N}(s_0, t_0)\| \, \Delta s \Delta t.\end{aligned}$$

Note that this quantity is the area of the unit square inside the tangent space to $\mathcal{S}$ at (s_0, t_0), suitably scaled by Δs and Δt.

In the limit, as $\Delta s, \Delta t \to 0$, we get $\|\mathbf{X}_s \times \mathbf{X}_t\| \, ds \, dt$. With the idea that $\textbf{area}(\mathcal{D}) = \iint_{\mathcal{D}} dA$, we get

$$\textbf{area}(\mathcal{S}) = \iint_{\mathcal{S}} dS = \iint_{\mathcal{D}} \|\mathbf{X}_s \times \mathbf{X}_t\| \, ds \, dt = \iint_{\mathcal{D}} \|\mathbf{N}(s, t)\| \, ds \, dt.$$

Here $dS = \|\mathbf{N}(s, t)\| \, dA$ is the differential of area, or an *area form* on $\mathcal{S}$.

Some notes:

- The expression $dS = \|\mathbf{N}(s, t)\| \, dA$ is the 2-dimensional analog to $dc = \|\mathbf{c}'(t)\| \, dA$ in the scalar-line integral.
- For $\mathbf{X}(s, t) = (x(s, t), y(s, t), z(s, t)) \in \mathbb{R}^3$,

$$\mathbf{X}_s \times \mathbf{X}_t = \begin{bmatrix} \frac{\partial(y,z)}{\partial(s,t)} \\ -\frac{\partial(x,z)}{\partial(s,t)} \\ \frac{\partial(x,y)}{\partial(s,t)} \end{bmatrix},$$

so

$$\textbf{area}(\mathcal{S}) = \iint_{\mathcal{D}} \sqrt{\left(\frac{\partial(y,z)}{\partial(s,t)}\right)^2 + \left(\frac{\partial(x,z)}{\partial(s,t)}\right)^2 + \left(\frac{\partial(x,y)}{\partial(s,t)}\right)^2} \, ds \, dt,$$

where each of the summands under the radical is the square of a Jacobian determinant. Compare this to the calculation of arclength for a parameterized curve in the plane in single variable calculus: Given $\mathbf{c}(t) = (x(t), y(t))$,

$$\textbf{arclength}(\mathbf{c}) = \int_{\mathbf{c}} \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2} \, dt.$$

EXAMPLE 19.5. ***The surface area of S_a^2 a 2-sphere of radius a.*** Recall that $\mathbf{X}(s,t) = (a\cos s\sin t, a\sin s\sin t, a\cos t)$. So, as detailed in Example 19.4 above.

$$\mathbf{X}_s = \begin{bmatrix} -a\sin s\sin t \\ a\cos s\sin t \\ 0 \end{bmatrix}, \quad \text{and} \quad \mathbf{X}_t = \begin{bmatrix} a\cos s\cos t \\ a\sin s\cos t \\ -a\sin t \end{bmatrix},$$

leading to

$$\mathbf{N} = \mathbf{X}_s \times \mathbf{X}_t = -a\sin t \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} a^2\cos s\sin^2 t \\ -a^2\sin s\sin^2 t \\ -a^2\sin t\cos t \end{bmatrix}.$$

This leads to

$$\begin{aligned} \|\mathbf{X}_s \times \mathbf{X}_t\| &= \sqrt{\left(\frac{\partial(y,z)}{\partial(s,t)}\right)^2 + \left(\frac{\partial(x,z)}{\partial(s,t)}\right)^2 + \left(\frac{\partial(x,y)}{\partial(s,t)}\right)^2} \\ &= \sqrt{a^4\cos^2 s\sin^4 t + a^4\sin^2 s\sin^4 t + a^4\sin^2 t\cos^2 t} \\ &= \sqrt{a^4\sin^2 t} = a^2\sin t. \end{aligned}$$

where we do not need to shroud this last term in absolute values since $\sin t$ is nonnegative for $t \in [0,\pi]$.

Thus we have

$$\begin{aligned} \mathbf{area}(S_a^2) &= \iint_{S_a^2} dS = \int_0^\pi \int_0^{2\pi} \|\mathbf{X}_s \times \mathbf{X}_t\|\, ds\, dt \\ &= \int_0^\pi \int_0^{2\pi} a^2\sin t\, ds\, dt = \int_0^\pi 2\pi a^2\sin t\, dt \\ &= -2\pi a^2\cos t\Big|_0^\pi = 2\pi a^2 + 2\pi a^2 = 4\pi a^2. \end{aligned}$$

EXAMPLE 19.6. ***The surface area of a graph.*** For $f : \mathcal{D} \subset \mathbb{R}^2 \to \mathbb{R}$,

$$\mathcal{S} = \mathbf{X}(\mathcal{D}) = \left\{(x,y,z) \in \mathbb{R}^3 \;\middle|\; z = f(x,y)\right\} = \mathbf{graph}(f).$$

Here

$$\mathbf{X}_x(x,y) = \begin{bmatrix} 1 \\ 0 \\ f_x \end{bmatrix}, \quad \text{and} \quad \mathbf{X}_y(x,y) = \begin{bmatrix} 0 \\ 1 \\ f_y \end{bmatrix},$$

so that

$$\mathbf{X}_x \times \mathbf{X}_y = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 0 & f_x \\ 0 & 1 & f_y \end{vmatrix} = -f_x\mathbf{i} - f_y\mathbf{j} + \mathbf{k} = \begin{bmatrix} -f_x \\ -f_y \\ 1 \end{bmatrix}.$$

Then the surface area of $\mathcal{S} = \mathbf{X}(\mathcal{D}) = \mathbf{graph}(f)$ is

$$\mathbf{area}\,(\mathcal{S}) = \iint_{\mathcal{S}} dS = \iint_{\mathcal{D}} \|\mathbf{X}_s \times \mathbf{X}_t\|\, dt = \iint_{\mathcal{D}} \sqrt{(f_x)^2 + (f_y)^2 + 1}\, dA.$$

Now compare this to the single variable calculation of the length of a curve which is the graph of a function $f : [a, b] \to \mathbb{R}$:

$$\mathbf{length} = \int_a^b \sqrt{1 + (f'(x))^2}\, dx.$$

LECTURE 20

Surface Integration

SYNOPSIS. We continue with the idea of understanding how the calculus of functions behaves along parameterized surfaces (instead of along parameterized curves.) Today, we define and study both scalar and vector surface integrals, of real-valued functions and vector fields along surfaces embedded in three space, respectively. We sill stick to surfaces in three space for the expediency of understanding these concepts without too much intricate machinery. But we will allude regularly to the idea that we can embed and parameterize a surface in n-space, ($n > 2$), and play the same game. We also discuss the idea of reparameterizations, orientation of a surface, and geometric interpretations, all as a lead up to another of the three big theorems, Stoke's Theorem.

20.0.1. Functions defined on surfaces. In a similar fashion that we integrate functions (real-valued) and vector fields (vector-valued) over curves, we can do so over surfaces:

- Like for scalar line integrals, if the surface in $\mathbb{R}^3$ lies inside the domain of a real-valued function on (a part of) $\mathbb{R}^3$, we can restrict the domain of the function to only the surface. Then adding up the values of the function, essentially integrating the function, over the surface is straightforward.
- If the surface is parameterized, so has coordinates defined directly on it, then the integration process can utilize the parameterization and we integrate using a double integral. However, it should also be obvious that the value of such an integration must be independent of any particular choice of parameterization.
- Basically, for a surface $\mathcal{S}$, we look to define the quantity

$$\iint_{\mathcal{S}} f \, dS, \quad \text{where} \quad dS = ||\mathbf{N}|| \, dA,$$

  where, by the previous lecture, the vector $\mathbf{N}$ is the normal to the surface at a point, defined via the cross product of the tangent partial derivative vectors of the parameterization, and dA is the area-form given also by the parameterization.
- Given a parameterization $\mathbf{X} : \mathcal{D} \subset \mathbb{R}^2 \to \mathbb{R}^n$, we will refer to the surface as simply $\mathbf{X}$ or $\mathbf{X}(s,t)$. This is entirely similar to our reference to a curve parameterization $\mathbf{x} : \mathcal{I} \subset \mathbb{R} \to \mathbb{R}^n$ as simply $\mathbf{x}$, or $\mathbf{x}(t)$, when appropriate.

DEFINITION 20.1. Let $\mathbf{X} : \mathcal{D} \subset \mathbb{R}^2 \to \mathbb{R}^3$ be a smooth parameterized surface, with $\mathcal{D}$ bounded. Let f be a C^0 function defined on a domain that includes $\mathbf{X}$. Then the *scalar surface integral* of f along $\mathbf{X}$ is

$$\begin{aligned}\iint_{\mathbf{X}} f\, dS &= \iint_{\mathcal{D}} f\left(\mathbf{X}(s,t)\right) \|\mathbf{X}_s \times \mathbf{X}_t\|\, ds\, dt \\ &= \iint_{\mathbf{X}} f\left(x(s,t), y(s,t), z(s,t)\right) \sqrt{\left(\frac{\partial(y,z)}{\partial(s,t)}\right)^2 + \left(\frac{\partial(x,z)}{\partial(s,t)}\right)^2 + \left(\frac{\partial(x,y)}{\partial(s,t)}\right)^2}\, ds\, dt.\end{aligned}$$

Some notes:

- Like for line integrals, dS is a scalar 2-form (as ds is a scalar 1-form), and represents an infinitesimal change in surface area along the surface.
- For $f(x,y,z) = 1$, the scalar surface integral of f gives the surface area of $\mathbf{X}$.
- In the parameter coordinates (s,t), this looks like a standard double integral.
- If $\mathbf{X}$ is not smooth, but has edges, so is piecewise smooth, then each smooth piece must be integrated separately, and the results added together.

DEFINITION 20.2. Let $\mathbf{X} : \mathcal{D} \subset \mathbb{R}^2 \to \mathbb{R}^3$ be a smooth parameterized surface, with $\mathcal{D}$ bounded. Let $\mathbf{F}$ be a C^1-vector field defined on a domain that includes $\mathbf{X}(\mathcal{D})$. Then the *vector surface integral* of $\mathbf{F}$ along $\mathbf{X}$ is

$$\begin{aligned}\iint_{\mathbf{X}} \mathbf{F} \cdot d\mathbf{S} &= \iint_{\mathcal{D}} \mathbf{F}\left(\mathbf{X}(s,t)\right) \cdot \mathbf{N}(s,t)\, ds\, dt \\ &= \iint_{\mathbf{X}} \mathbf{F}\left(x(s,t), y(s,t), z(s,t)\right) \cdot \begin{bmatrix} \frac{\partial(y,z)}{\partial(s,t)} \\ -\frac{\partial(x,z)}{\partial(s,t)} \\ \frac{\partial(x,y)}{\partial(s,t)} \end{bmatrix} ds\, dt.\end{aligned}$$

Some notes:

- Here $d\mathbf{S} = \mathbf{N}(s,t)\, ds\, dt$ is a vector 2-form. It is the differential of surface area written in terms of the normal to the surface at (s,t).
- If we normalize the normal vector

$$\mathbf{n}(s,t) = \frac{\mathbf{N}(s,t)}{\|\mathbf{N}(s,t)\|},$$

then

$$\begin{aligned}\iint_{\mathbf{X}} \mathbf{F} \cdot d\mathbf{S} &= \iint_{\mathcal{D}} \mathbf{F}\left(\mathbf{X}(s,t)\right) \cdot \mathbf{N}(s,t)\, ds\, dt \\ &= \iint_{\mathcal{D}} \mathbf{F}\left(\mathbf{X}(s,t)\right) \cdot \mathbf{n}(s,t)\, \|\mathbf{N}(s,t)\|\, ds\, dt \\ &= \iint_{\mathbf{X}} \left(\mathbf{F} \cdot \mathbf{n}\right) dS.\end{aligned}$$

Thus, the vector surface integral of a vector field along a surface is just the scalar surface integral of the component of the vector field normal to the surface, along the surface. This concept wil be very important in the near futur

So what is the geometric interpretation of a vector surface integral? The normal component of the vector field at a point on the surface measures the flow at that point through the surface. Hence the quantity $\iint_{\mathbf{X}} \mathbf{F} \bullet d\mathbf{S}$ measures the total vector field flow through the surface. This is also called the *flux* of $\mathbf{F}$ through $\mathbf{X}$. Compare this to the interpretation of the vector line integral $\int_{\mathbf{c}} \mathbf{F} \bullet d\mathbf{s}$, the circulation, which measures the vector field flow in the direction of $\mathbf{c}$ along $\mathbf{c}$.

DEFINITION 20.3. Let $\mathbf{X} : \mathcal{D}_1 \subset \mathbb{R}^2 \to \mathbb{R}^3$ and $\mathbf{Y} : \mathcal{D}_2 \subset \mathbb{R}^2 \to \mathbb{R}^3$ be two parameterizations, such that $\mathbf{X}(\mathcal{D}_1) = \mathbf{Y}(\mathcal{D}_2)$. The $\mathbf{Y}$ is called a reparameterization of $\mathbf{X}$ if there exists a one-to-one and onto $\mathbf{H} : \mathcal{D}_2 \to \mathcal{D}_1$, with inverse $\mathbf{H}^{-1} : \mathcal{D}_1 \to \mathcal{D}_2$, such that $\mathbf{Y} = \mathbf{X} \cdot \mathbf{H}$

Note that a reparameterization is called smooth if both $\mathbf{X}$ and $\mathbf{Y}$ are smooth and if $\mathbf{H}$ is C^1. Here are some facts:

(1)

THEOREM 20.4. *For f a C^0-function on a domain including a smooth $\mathbf{X} : \mathcal{D} \to \mathbb{R}^3$, then for any smooth reparameterization $\mathbf{Y}$ of $\mathbf{X}$,*

$$\iint_{\mathbf{Y}} f \, dS = \iint_{\mathbf{X}} f \, dS.$$

(2) For a smooth curve $\mathbf{c}$, an *orientation* is a choice of a continuously varying *unit tangent vector* along $\mathbf{c}$. This is entirely consistent with the idea of simply choosing a direction of travel along the curve. Note that an orientation is given automatically when one parameterizes the curve (the direction of increase in values of the parameter), but one may also "choose" to move in the opposite direction, as we did when proving Green's Theorem.

For a smooth surface $\mathbf{X}$, an *orientation* is a choice of continuously varying *unit normal vector* along $\mathbf{X}$. In effect, you are choosing a "side", as in above vs. below, or inside vs. outside. Note that there are surfaces that are not orientable. But note that any parameterization $\mathbf{X}$ automatically orients the surface, since

$$\mathbf{n}(s,t) = \frac{\mathbf{N}(s,t)}{\|\mathbf{N}(s,t)\|} = \frac{\mathbf{X}_s(s,t) \times \mathbf{X}_t(s,t)}{\|\mathbf{X}_s(s,t) \times \mathbf{X}_t(s,t)\|}.$$

(3) One may ask if two smooth parameterizations $\mathbf{X}$ and $\mathbf{Y}$ defining the same surface are oriented compatibly. Essentially, we are asking if a smooth reparameterization *preserves orientation*. The answer is yes, if the two calculated unit normal vectors at each point are "on the same side". But this can be determined by $\mathbf{H} : \mathcal{D}_2 \to \mathcal{D}_1$, the function defining the reparameterization. Indeed, let $\mathbf{Y}(s,t)$,

defined on $\mathcal{D}_2$ be a reparameterization of $\mathbf{X}(u,v)$, defined on the domain $\mathcal{D}_1$, with $\mathbf{H}$, defined as above. Then $\mathbf{Y}(s,t) = \mathbf{X}(u,v) = \mathbf{X}(\mathbf{H}(s,t))$. Then one can use the Chain Rule to show

$$\mathbf{N}_{\mathbf{Y}}(s,t) = \frac{\partial(u,v)}{\partial(s,t)}\mathbf{N}_{\mathbf{X}}(u,v),$$

where the subscripts on the normal vectors here only denote to which parameterization the normal vector belongs. Hence if the Jacobian determinant of $\mathbf{H}$ is positive, the reparameterization is *orientation preserving*. If negative, then the reparameterization is *orientation reversing*: We have

THEOREM 20.5. *For $\mathbf{F}$ a C^0-vector field on a domain including a smooth $\mathbf{X} : \mathcal{D} \to \mathbb{R}^3$, then for any orientation preserving smooth reparameterization $\mathbf{Y}$,*

$$\iint_{\mathbf{Y}} \mathbf{F} \cdot d\mathbf{S} = \iint_{\mathbf{X}} \mathbf{F} \cdot d\mathbf{S}.$$

If $\mathbf{Y}$ is orientation reversing, then

$$\iint_{\mathbf{Y}} \mathbf{F} \cdot d\mathbf{S} = -\iint_{\mathbf{X}} \mathbf{F} \cdot d\mathbf{S}.$$

(4) Orienting a surface automatically orients the boundary of the surface. Indeed, let $\mathcal{S}$ be an oriented surface with boundary in $\mathbb{R}^3$ such that $\partial\mathcal{S}$ is a piecewise-C^1 closed curve. Let $\mathbf{p} = (s_0, t_0) \in \partial\mathcal{S}$, where

$$\mathbf{p} = (s_0, t_0) = (x(s_0, t_0), y(s_0, t_0), z(s_0, t_0)),$$

and choose $\mathbf{c} : [a,b] \to \mathcal{S} \subset \mathbb{R}^3$ a smooth curve in the surface so that $\mathbf{c}(a) = \mathbf{p}$ and $\mathbf{c} \cap \partial\mathcal{S} = \{\mathbf{p}\}$. Now define

$$\mathbf{n}(\mathbf{p}) = \lim_{t\to a} \mathbf{n}(\mathbf{c}(t)), \quad \text{and} \quad \mathbf{v}(a) = \lim_{t\to a} \mathbf{c}'(t).$$

Here, $\mathbf{n}$ and $\mathbf{v}$ are vectors based at $\mathbf{p}$ and are orthogonal to each other. Hence they determine a two dimensional plane in $\mathbb{R}^3$ as the set of all linear combinations. Then $\mathbf{n} \times \mathbf{v}$ is perpendicular to both, and using the right-hand rule, determines a unique direction on $\partial\mathcal{S}$.

This is the direction specified in Green's Theorem!

LECTURE 21

The Theorem of Stokes

SYNOPSIS. In this lecture, we begin to finish the foundational material of what makes a vector calculus course with a full discussion of one of the two other Big Theorems, those of Stokes and Gauss. Here, we present and discuss Stokes' Theorem, developing the intuition of what the theorem actually says, and establishing some main situations where the theorem is relevant. Then we use Stokes' Theorem in a few examples and situations.

THEOREM 21.1 (Stokes' Theorem). *Let $\mathcal{S}$ be a bounded, piecewise smooth, oriented surface in $\mathbb{R}^3$, where $\partial\mathcal{S}$ consists of finitely many piecewise smooth closed curves oriented compatibly. FOr $\mathbf{F}$ a C^1-vector field on a domain containing $\mathcal{S}$,*

$$\iint_{\mathcal{S}} \nabla \times \mathbf{F} \bullet d\mathbf{S} = \oint_{\partial\mathcal{S}} \mathbf{F} \bullet d\mathbf{s}.$$

Some notes:

(1) Here, the surface integral of the curl of a vector field along a surface is equal to the circulation of the vector field along the boundary of the surface.
(2) This is a lot like Green's Theorem:
   - The left-hand side measure the normal component of the curl of $\mathbf{F}$ along $\mathcal{S}$, so measures the amount of twisting in the direction through $\mathcal{S}$).
   - The right-hand side measures the tangent component of $\mathbf{F}$ along $\partial\mathcal{S}$.
(3) In a way, the shape of the surface doesn;t matter as much as what is happening on the boundary. According to Stokes' Theorem, in each of the surfaces in Figure **??**, the value of $\iint_{\mathcal{S}} \nabla \times \mathbf{F} \bullet d\mathbf{S}$ is the same.
(4) Typical use: Sometimes the flux of the curl of $\mathbf{F}$ is hard to calculate across a bounded surface. But the circulation along its boundary is not!

EXAMPLE 21.1. ***Compute the flux of the curl of*** $\mathbf{F} = xz\,\mathbf{i} + yz\,\mathbf{j} + xy\,\mathbf{k}$ ***through the surface of the sphere*** $x^2 + y^2 + z^2 = 4$ ***inside the cylinder*** $x^2 + y^2 = 1$ ***and above the*** xy***-plane in*** $\mathbb{R}^3$**.** The strategy for this calculation is that, since both the vector field and the surface satisfy

Stokes' Theorem ($\mathbf{F}$ is C^1 and $\mathcal{S}$, using the outward normal is orientable and bounded with a closed, smooth boundary curve, which we can orient compatibly as counterclockwise, or with $\mathcal{S}$ on the left, walking upright on the curve), we look to calulate the surface integral by instead calculating the circulation of $\mathbf{F}$ along $\partial\mathcal{S}$.

First, let's parameterize $\partial\mathcal{S}$ so that the orientation given by the parameterization is conpatible with the surface orientation. Here, $\partial\mathcal{S}$ is on both the sphere $x^2+y^2+z^2=4$, as well as the cylinder $x^2+y^2=1$. Hence

$$(x^2+y^2)+z^2=1+z^2=4,$$

so that $z^2=3$ and $z=\sqrt{3}$ (recall we are only using the positive hemisphere here). So parameterize $\partial\mathcal{S}$ as $\mathbf{c}:[0,2\pi]\to\mathbb{R}^3$ via

$$\mathbf{c}(t)=\begin{bmatrix}\cos t\\ \sin t\\ \sqrt{3}\end{bmatrix}\in\mathbb{R}^3.$$

The next step is to calculate the circulation of $\mathbf{F}$ over $\mathbf{c}$. Here we have

$$\begin{aligned}\iint_{\mathcal{S}}\nabla\times\mathbf{F}\bullet d\mathbf{S}\overset{\text{Stokes'}}{=\!=\!=}\int_{\partial\mathcal{S}}\mathbf{F}\bullet d\mathbf{s}&=\int_{\mathbf{c}}\mathbf{F}\left(\mathbf{c}(t)\right)\bullet\mathbf{c}'(t)\,dt\\ &=\int_0^{2\pi}\begin{bmatrix}\sqrt{3}\cos t\\ \sqrt{3}\sin t\\ \cos t\sin t\end{bmatrix}\bullet\begin{bmatrix}-\sin t\\ \cos t\\ 0\end{bmatrix}dt\\ &=\int_0^{2\pi}0\,dt=0.\end{aligned}$$

Now suppose that we did this calculation directly, without using Stoke's. In this case, the strategy is to parameterize the surface, which we will do using spherical coordinates (note that the "curved disk" here, which comprises $\mathcal{S}$, lies on the $\rho=4$ sphere). In spherical coordinates, we find that the region is, in fact, a rectangle in the two angles. Then we calculate the resulting double integral.

Here, we start with calculating $\mathbf{curl}(\mathbf{F})$:

$$\mathbf{curl}(\mathbf{F})=\nabla\times\mathbf{F}=\begin{vmatrix}\mathbf{i}&\mathbf{j}&\mathbf{k}\\ \frac{\partial}{\partial x}&\frac{\partial}{\partial y}&\frac{\partial}{\partial z}\\ xz&yz&xy\end{vmatrix}=(x-y)\,\mathbf{i}-(y-x)\,\mathbf{j}+0\,\mathbf{k}=\begin{bmatrix}x-y\\ x-y\\ 0\end{bmatrix}.$$

Next, we parameterize $\mathcal{S}$. First, we again note that the entirety of $\mathcal{S}$ lies on the sphere at $\rho=2$. This leaves us already with the parameterization $\mathbf{X}:\mathcal{D}\to\mathbb{R}^3$ based solely on the angles

$$\mathbf{X}(\theta\varphi)=\begin{bmatrix}2\cos\theta\sin\varphi\\ 2\sin\theta\sin\varphi\\ 2\cos\varphi\end{bmatrix}.$$

Second, we seek to define the region $\mathcal{D}$ in the $\theta\varphi$-plane that leads to $\mathcal{S} = \mathbf{X}(\mathcal{D})$. To this end, we note that the azimuth angle goes all the way to describe $\mathcal{S}$, so $\theta \in [0, 2\pi]$. However, the polar angle φ will only go from $\varphi = 0$, at the north pole, to the edge of $\mathcal{S}$, so we need to find the value of φ that corresponds to the edge. To see this, look directly into the xz-plane, and note that the $\rho = 2$ sphere forms a semicircle of radius 2, and the intersection of $\partial\mathcal{S}$ in this plane occurs at a point on this semicircle with x-coordinate 1. One can calculate that the y-coordinate here is $\sqrt{3}$, and that the radial line from the origin to $\partial\mathcal{S}$ has angle $\alpha = \frac{\pi}{3}$. This means that the polar angle $\varphi = \frac{\pi}{2} - \alpha = \frac{\pi}{6}$.

Hence the region $\mathcal{D}$ in the $\theta\varphi$-plane corresponds to $\mathcal{D} = [0, 2\pi] \times \left[0, \frac{\pi}{6}\right]$.

Next, before we integrate, we need to check to ensure that our idea of orienting $\mathcal{S}$ with the normal pointing outward is correct, using this parameterization. We need this since we have already oriented our curve $\mathbf{c}$ in the previous calculation to be compatible with the outward pointing normal. Here, we have

$$\mathbf{X}_\theta = \begin{bmatrix} -2\sin\theta\sin\varphi \\ 2\cos\theta\sin\varphi \\ 0 \end{bmatrix}, \quad \text{and} \quad \mathbf{X}_\varphi = \begin{bmatrix} 2\cos\theta\cos\varphi \\ 2\sin\theta\cos\varphi \\ -2\sin\varphi \end{bmatrix},$$

and $\mathbf{N} = \mathbf{X}_\theta \times \mathbf{X}_\varphi = \begin{bmatrix} \cos\theta\sin\varphi \\ \sin\theta\sin\varphi \\ \cos\varphi \end{bmatrix}(-4\sin\varphi)$. But this means

$$\mathbf{n} = \frac{\mathbf{N}}{||\mathbf{X}_\theta \times \mathbf{X}_\varphi||} = -\begin{bmatrix} \cos\theta\sin\varphi \\ \sin\theta\sin\varphi \\ \cos\varphi \end{bmatrix}.$$

Unfortunately, this unit normal points inward (toward the origin). This is fine, but it is incompatible with our orientation of $\partial\mathcal{S}$. To fix this, we simply reparameterize to generate the other orientation. The simplest way to do this is to rewrite the region $\mathcal{D}$ as lying inside the $\varphi\theta$-place instead of the $\theta\varphi$-plane. Then the region $\mathcal{D} = \left[0, \frac{\pi}{6}\right] \times [0, 2\pi]$ in the $\varphi\theta$-plane and the normal vector to $\mathcal{S}$ using this new orientation reversing reparameterization will be

$$\mathbf{N} = \mathbf{X}_\varphi \times \mathbf{X}_\theta = -\mathbf{X}_\theta \times \mathbf{X}_\varphi$$

by the properties of the cross product.

Lastly, we calculate the flux of the curl:

$$\begin{aligned}
\iint_{\mathcal{S}} \nabla \times \mathbf{F} \bullet d\mathbf{S} &= \iint_{\mathcal{D}} \nabla \times \mathbf{F}\left(\mathbf{X}\left(\varphi,\theta\right)\right) \bullet \left(\mathbf{X}_{\varphi} \times \mathbf{X}_{\theta}\right)\, dA \\
&= \iint_{\mathcal{D}} \left(\nabla \times \mathbf{F}\left(\mathbf{X}\left(\varphi,\theta\right)\right) \bullet \mathbf{n}\right)\, dA \\
&= \int_0^{\frac{\pi}{6}} \int_0^{2\pi} \begin{bmatrix} 2\sin\varphi(\cos\theta - \sin\theta) \\ 2\sin\varphi(\cos\theta - \sin\theta) \\ 0 \end{bmatrix} \bullet \begin{bmatrix} \cos\theta\sin\varphi \\ \sin\theta\sin\varphi \\ \cos\varphi \end{bmatrix} (4\sin\varphi)\, d\theta\, d\varphi \\
&= \int_0^{\frac{\pi}{6}} \int_0^{2\pi} 8\sin^3\varphi\left(\cos^2\theta - \sin^2\theta\right)\, d\theta\, d\varphi \\
&= \int_0^{\frac{\pi}{6}} 8\sin^3\varphi \left(\int_0^{2\pi} \sin 2\theta\, d\theta\right) d\varphi.
\end{aligned}$$

But the inside integral is 0, since

$$\int_0^{2\pi} \sin 2\theta\, d\theta = \left[-\frac{1}{2}\cos 2\theta \Big|_0^{2\pi}\right] = 0.$$

Hence the entire double integral is 0.

So which was easier??

Here is a less typical example of the use of Stokes' Theorem: Sometimes, one can use Stokes' to change the surface in a way that leaves the boundary fixed. So if a calculation of the flux of the curl of a vector field across $\mathcal{S}$ is difficult, and the circulation of the vector field along $\partial\mathcal{S}$ is also difficult, if Stokes applies, one can just find a different surface, with the same boundary, where the flux of the curl is easier to integrate.

EXAMPLE 21.2. ***Example 7.3.2 of the text.*** Calculate the flux of the curl of

$$\mathbf{F} = \begin{bmatrix} e^{y+z} - 2y \\ xe^{y+z} + y \\ e^{x+y} \end{bmatrix} \quad \text{across} \quad \mathcal{S} = \left\{ (x,y,z) \in \mathbb{R}^3 \;\middle|\; z \geq \frac{1}{e},\ z = e^{-(x^2+y^2)} \right\}.$$

By inspection and a few quick calculations (check the book), one can use Stokes' Theorem here, but both sides of the equal sign in the theorem are quite difficult calculations! However, by Stokes', any surface with the same boundary as $\mathcal{S}$ will do, when calculating the flux of the curl of $\mathbf{F}$ across it.

So here, choose

$$\widehat{\mathcal{S}} = \left\{ \left(x, y, \frac{1}{e}\right) \in \mathbb{R}^3 \;\middle|\; x^2 + y^2 \leq 1 \right\}.$$

Then, we have

$$\partial\widehat{\mathcal{S}} = \partial\mathcal{S} = \left\{ \left(x, y, \frac{1}{e} \right) \in \mathbb{R}^3 \;\middle|\; x^2 + y^2 = 1 \right\}.$$

So by Stokes' Theorem,

$$\iint_{\mathcal{S}} \nabla \times \mathbf{F} \bullet d\mathbf{S} = \iint_{\widehat{\mathcal{S}}} \nabla \times \mathbf{F} \bullet d\mathbf{S} = \iint_{\widehat{\mathcal{S}}} (\nabla \times \mathbf{F} \bullet \mathbf{n})\, dS.$$

So for this calculation, we find

$$\nabla \times \mathbf{F} = \begin{bmatrix} e^{x+y} - xe^{y+z} \\ e^{y+z} - e^{x+y} \\ 2 \end{bmatrix}, \quad \text{and} \quad \mathbf{n} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

So then $\nabla \times \mathbf{F} \bullet \mathbf{n} = 2$. Now the calculation is simply

$$\begin{aligned} \iint_{\mathcal{S}} \nabla \times \mathbf{F} \bullet d\mathbf{S} &= \iint_{\widehat{\mathcal{S}}} (\nabla \times \mathbf{F} \bullet \mathbf{n})\, dS \\ &= \iint_{\widehat{\mathcal{S}}} 2\, dS = 2\left(\mathbf{area}(\widehat{\mathcal{S}})\right) = 2\left(\pi(1)^2\right) = 2\pi. \end{aligned}$$

Here ae some great uses for Stokes' Theorem:

(1) A surface is called *compact* if it is closed as a set, and bounded. A surface is called *closed* if it is compact and has no boundary.
   - Surfaces like the 2-sphere S^2, and the 2-torus T^2 are closed, while the disk, or a surface which is the continuous injective graph of a closed rectangle in the plane (we tend to call this a *(flying) carpet*.
   - Recall that a curve in $\mathbb{R}^n$ is *simple* if it does not intersect itself. Hence a bounded simple curve with its endpoints is *compact*. Its boundary are the two endpoints. A *closed* curve forms a loop, and hence has no boundary.
   - Also, the surface of teh unit cube in $\mathbb{R}^3$ is closed. Creases and corners are not considered boundaries of a surface. So, in a very mathematical sense, there is no real difference between the surface of a cube, and the surface of a ball. One does have edges and corners and the other does not, but each does enclose space. Now, there is a difference, though. The surface of the unit cube is not smooth, while the surface of the ball is. But we can still integrate over each. The difference is that, to integrate over the cube (a piece-wise smooth surface), we would have to break up the integral calculation into each face, and then add the results at the end.

(2) In the case of a closed surface, the curl of any vector field in $\mathbb{R}^3$ over a closed surface will be 0, by order of Stokes' Theorem: The total flux of the curl of a vector field over a surface is equal to the

circulation of that vector field over the boundary of the surface. If the surface has no boundary, then there is no circulation. Then by Stokes', the curl of $\mathbf{F}$ has no flux across the surface.

(3) In contrast, let $\mathbf{F}$ be a conservative vector field, so $\mathbf{F} = \nabla f$ for a real-valued, C^1-function. Then, for any surface $\mathcal{S}$ that satisfies Stokes', the circulation of $\mathbf{F}$ along $\partial\mathcal{S}$ is 0 (we say it vanishes), or

$$\int_{\partial\mathcal{S}} \mathbf{F} \bullet d\mathbf{s} = 0.$$

Why is this? For any conservative vector field, $\nabla \times \mathbf{F} = \nabla \times \nabla f = \mathbf{0}$. Hence by Stokes'

$$\iint_{\mathcal{S}} \nabla \times \mathbf{F} \bullet d\mathbf{S} = \iint_{\mathcal{S}} \mathbf{0} \bullet d\mathbf{S} = 0 = \oint_{\partial\mathcal{S}} \mathbf{F} \bullet d\mathbf{s}.$$

LECTURE 22

Thh Theorem of Gauss

SYNOPSIS. In this lecture, we finish the foundational material of what makes a vector calculus course with a full discussion of the last of the Big Theorems, the theorem of Gauss. Really, the BIg Three theorems we are discussing are all similar in nature yet vary in dimension. Again, for Gauss' Theorem, we state the theorem and discuss its constituent pieces, develop the intuition needed to see what the theorem states, and establish some main situations where the theorem is relevant. Then we use Gauss' and Stokes' Theorems to give a much more precise idea of just what the divergence and the curl of a vector field actually is and how to understand these concepts geometrically.

22.0.1. Gauss' Theorem.

THEOREM 22.1 (Gauss' Theorem). *Let $\mathcal{W} \in \mathbb{R}^3$ be a solid region, whith $\partial\mathcal{W}$ a finite set of piecewise smooth, closed, orientable surfaces, oriented outwardly from $\mathcal{W}$. For $\mathbf{F}$ a C^1-vector field on a domain containing $\mathcal{W}$,*

$$\left(\oiint_{\partial\mathcal{W}} (\mathbf{F}\cdot\mathbf{n})\, dS \;=\right) \oiint_{\partial\mathcal{W}} \mathbf{F}\cdot d\mathbf{S} = \iiint_{\mathcal{W}} \nabla\cdot\mathbf{F}\, dV \left(= \iiint_{\mathcal{W}} (\mathbf{div}\,\mathbf{F})\, dV\right).$$

Special Notes:

(1) Recall that any compact domain in $\mathbb{R}^2$ with nonempty interior has a set of closed curves as boundary. In $\mathbb{R}^3$, any compact domain with nonempty interior has a set of closed surfaces as boundary. And this generalizes to higher dimensions readily. But this leads to the conclusion: *The boundary of a compact region with nonempty interior has no boundary!* Think about this.

(2) The proof of Gauss' Theorem is elementary and quite straightforward: Denoting points in $\mathbb{R}^3$ in the obvious way using the variable x, y, and z, and with $\mathbf{F} = F_1(x,y,z)\,\mathbf{i} + F_2(x,y,z)\,\mathbf{j} + F_3(x,y,z)\,\mathbf{k}$, we first see that

$$\iiint_{\mathcal{W}} (\mathbf{div}\,\mathbf{F})\, dV = \iiint_{\mathcal{W}} \frac{\partial F_1}{\partial x}\, dV + \iiint_{\mathcal{W}} \frac{\partial F_2}{\partial y}\, dV + \iiint_{\mathcal{W}} \frac{\partial F_3}{\partial z}\, dV.$$

Then, it should also be clear that

$$\iint_{\partial\mathcal{W}} (\mathbf{F}\cdot\mathbf{n})\, dV = \iint_{\partial\mathcal{W}} F_1\,\mathbf{i}\cdot\mathbf{n}\, dV + \iint_{\partial\mathcal{W}} F_2\,\mathbf{j}\cdot\mathbf{n}\, dV + \iint_{\partial\mathcal{W}} F_3\,\mathbf{k}\cdot\mathbf{n}\, dV.$$

And finally, one simply shows that each is equal to each, respectively.

EXERCISE 12. Show $\iiint_{\mathcal{W}} \frac{\partial F_1}{\partial x}\, dV = \iint_{\partial \mathcal{W}} F_1\, \mathbf{i} \bullet \mathbf{n}\, dV$, when $\mathcal{W}$ is elementary in all directions.

So we are now fully in a position to understand some concepts that we have previously only vaguely discussed:

- Divergence of a vector field.
- Curl of a vector field.

22.0.2. Divergence.

- Intuitive definition: Measures the infinitesimal expansion of volume under the flow of a vector field.
- Actual definition: Measures the aggregate flux of a vector field across the boundary of an infinitesimal ball centered at a point.

THEOREM 22.2. *Let* $\mathbf{F}$ *be a* C^1*-vector field defined in some (open) neighborhood of a point* $\mathbf{p} \in \mathbb{R}^3$*. For*

$$\mathcal{S}_a = \left\{ \mathbf{x} \in \mathbb{R}^3 \;\middle|\; \|\mathbf{x} - \mathbf{p}\| = a \right\},$$

a 2-sphere of radius $a > 0$ *centered at* $\mathbf{p}$ *and oriented outwardly,*

$$\mathbf{div}\, \mathbf{F}(\mathbf{p}) = \lim_{a \to 0^+} \frac{3}{4\pi a^3} \oiint_{\mathcal{S}_a} \mathbf{F} \bullet d\mathbf{S}.$$

PROOF. For any $f \in C^0[\mathcal{W} \subset \mathbb{R}^3, \mathbb{R}]$, $\mathcal{W}$ a bounded solid region, there exists $\mathbf{q} \in \mathbb{R}^3$ where

$$\iiint_{\mathcal{W}} f(x,y,z)\, dV = f(\mathbf{q}) \cdot \mathbf{volume}(\mathcal{W}).$$

This is the Mean Value Theorem for triple integrals.

Now since $\mathbf{F}$ is C^1 in an open neighborhood of $\mathbf{p}$, there exists $\epsilon > 0$ such that $\mathbf{F}$ is C^1 on

$$\overline{\mathcal{B}}_\epsilon = \left\{ \mathbf{x} \in \mathbb{R}^3 \;\middle|\; \|\mathbf{x} - \mathbf{p}\| \le \epsilon \right\}.$$

Then $\epsilon = a$ in the theorem and $\mathcal{S}_a = \partial \overline{\mathcal{B}}_\epsilon$. Then, there exists a $\mathbf{q} \in \overline{\mathcal{B}}_\epsilon$ such that

$$\begin{aligned} \iiint_{\overline{\mathcal{B}_\epsilon}} \mathbf{div}(\mathbf{F})\, dV &= \mathbf{div}\left(\mathbf{F}(\mathbf{q})\right) \cdot \mathbf{volume}\left(\overline{\mathcal{B}}_\epsilon\right) \\ &= \frac{4\pi\epsilon^3}{3} \mathbf{div}\left(\mathbf{F}(\mathbf{q})\right), \end{aligned}$$

since divergence is simply a scalar field on $\overline{\mathcal{B}}_\epsilon$. Now, we can use Gauss' Theorem:

$$\begin{aligned} \lim_{\epsilon \to 0^+} \frac{3}{4\pi\epsilon^3} \oiint_{\mathcal{S}_\epsilon} \mathbf{F} \bullet d\mathbf{S} &\overset{\textbf{Gauss}}{=\!=\!=} \lim_{\epsilon \to 0^+} \frac{3}{4\pi\epsilon^3} \iiint_{\overline{\mathcal{B}}_\epsilon} \mathbf{div}\left(\mathbf{F}\right)\, dV \\ &= \lim_{\epsilon \to 0^+} \frac{3}{4\pi\epsilon^3} \left(\frac{4\pi\epsilon^3}{3} \cdot \mathbf{div}\left(\mathbf{F}(\mathbf{q})\right) \right) \\ &= \lim_{\epsilon \to 0^+} \mathbf{div}\left(\mathbf{F}(\mathbf{q})\right) = \mathbf{div}\left(\mathbf{F}(\mathbf{p})\right). \end{aligned}$$

□

So Gauss' Theorem says that the amount of volume created or lost upon flowing along a vector field in a compact solid $\mathcal{W}$ is equal to the total amount flowing through the boundary $\partial\mathcal{W}$.

22.0.3. Curl.

- Intuitive definition: Measures the twisting effect of a vector field in $\mathbb{R}^3$ felt by flowing along it.
- Actual definition: Measures the total circulation of a vector field along an edge of an infinitesimal disk normal to the vector field at a point.

THEOREM 22.3. *Let* $\mathbf{F}$ *be a* C^1*-vector field defined in some (open) neighborhood of a point* $\mathbf{p} \in \mathbb{R}^3$*. Let* $\mathbf{n}$ *be a unit vector based at* $\mathbf{p}$*, with*

$$D_a = \left\{ \mathbf{x} \in \mathbb{R}^3 \;\middle|\; \|\mathbf{x} - \mathbf{p}\| \leq a,\ (\mathbf{x} - \mathbf{p}) \bullet \mathbf{n} = 0 \right\},$$

a 2-disk of radius $a > 0$ *centered at* $\mathbf{p}$ *and normal to* $\mathbf{n}$*. Orient* D_a *compatibly with* $\mathbf{n}$ *and also* $C_a = \partial D_a$*. Then the component of curl in the direction of* $\mathbf{n}$ *is*

$$\mathbf{curl}\,\mathbf{F}(\mathbf{p}) \bullet \mathbf{n} = \lim_{a \to 0^+} \frac{1}{\pi a^2} \oint_{C_a} \mathbf{F} \bullet d\mathbf{s}.$$

Hence, curl at a point is the infinitesimal circulation of $\mathbf{F}$ along a loop perpendicular to the direction of flow. In essence, choose a unit vector $\mathbf{n}$ based at $\mathbf{p}$, and form a small disk normal to $\mathbf{n}$ and containing $\mathbf{p}$. If the vector field generally points in a direction along the boundary of the disk compatibly with its orientation with the disk on its left (counterclockwise), then the circulation will be positive. If, in the aggregate, it points in the opposite direction to the orientation on the boundary of the disk, then the circulation will be negative. And if, in the aggregate, the vector field is orthogonal to the boundary of the disk, then the circulation will be 0.

Now allow $\mathbf{n}$ to vary. The magnitude of $\mathbf{curl}\,\mathbf{F}(\mathbf{p}) \bullet \mathbf{n}$ will be maximized at $\mathbf{p}$ precisely when

$$\mathbf{n} = \frac{\mathbf{curl}\,\mathbf{F}(\mathbf{p})}{\|\mathbf{curl}\,\mathbf{F}(\mathbf{p})\|}.$$

Therefore the twisting or rotating effect of the vector field $\mathbf{F}$ at $\mathbf{p}$ is greatest about the axis parallel to $\frac{\mathbf{curl}\,\mathbf{F}(\mathbf{p})}{\|\mathbf{curl}\,\mathbf{F}(\mathbf{p})\|}$. One can use this as a definition of the curl of a vector field.

PROOF. Exactly like the previous theorem but using Stokes instead of Gauss. □

Notes:

- The quantity $\mathbf{div}\,(\mathbf{F}(\mathbf{p}))$ is also called the *flux density* of $\mathbf{F}$ at $\mathbf{p}$: It is the limit of the flux per unit volume.
- The quantity $\mathbf{curl}\,(\mathbf{F}(\mathbf{p}))$ is also called the *circulation density* of $\mathbf{F}$ at $\mathbf{p}$: It is the limit of the flux per unit volume.

Stokes' Theorem says that the total rotational effect of a vector field on a surface in $\mathbb{R}^3$ is equal to the aggregate boost or hindrance of a particle on the edge.

Green's Theorem is simply Stokes' Theorem limited to domains in the plane.

EXAMPLE 22.1. ***For*** $\mathbf{F} = 2x\,\mathbf{i} + y^2\,\mathbf{j} + z^2\,\mathbf{k}$ ***and***

$$\mathcal{S} = \left\{(x,y,z) \in \mathbb{R}^3 \;\middle|\; x^2 + y^2 + z^2 = 1\right\},$$

the unit 2-sphere in $\mathbb{R}^3$***, find the flux of*** $\mathbf{F}$ ***through*** $\mathcal{S}$***.***

The solution here is a calculation of

$$\oiint_{\mathcal{S}} (\mathbf{F} \bullet \mathbf{n})\, dS,$$

where $\mathbf{n}$ is the unit normal vector to $\mathcal{S}$. We use Gauss' Theorem to instead integrate the divergence of $\mathbf{F}$ on the unit ball

$$\overline{\mathcal{B}} = \left\{(x,y,z) \in \mathbb{R}^3 \;\middle|\; x^2 + y^2 + z^2 \leq 1\right\}.$$

By Gauss,

$$\begin{aligned}
\oiint_{\mathcal{S}} (\mathbf{F} \bullet \mathbf{n})\, dS &= \iiint_{\overline{\mathcal{B}}} \mathbf{div}\,\mathbf{F}\, dV \\
&= \iiint_{\overline{\mathcal{B}}} 2(1 + y + z)\, dV \\
&= 2\iiint_{\overline{\mathcal{B}}} dV + 2\iiint_{\overline{\mathcal{B}}} y\, dV + 2\iiint_{\overline{\mathcal{B}}} z\, dV.
\end{aligned}$$

We do the middle integral first: We have

$$\begin{aligned}
2\iiint_{\overline{\mathcal{B}}} y\, dV &= 2\int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}\int_{-\sqrt{1-x^2-z^2}}^{\sqrt{1-x^2-z^2}} y\, dy\, dz\, dx \\
&= 2\int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \underbrace{\left(\frac{y^2}{2}\bigg|_{-\sqrt{1-x^2-z^2}}^{\sqrt{1-x^2-z^2}}\right)}_{0} dz\, dx \\
&= 0.
\end{aligned}$$

The same will be true for the third integral. And so we have ,

$$\begin{aligned}
\oiint_{\mathcal{S}} (\mathbf{F} \bullet \mathbf{n})\, dS &\overset{\textbf{Gauss}}{=\!=\!=} \iiint_{\overline{\mathcal{B}}} \mathbf{div}\,\mathbf{F}\, dV \\
&= 2\iiint_{\overline{\mathcal{B}}} dV = 2 \cdot \mathbf{vol}\left(\overline{\mathcal{B}}\right) = 2\left(\frac{4\pi(1)^3}{3}\right) = \frac{8\pi}{3}.
\end{aligned}$$

LECTURE 23

Differential Forms

SYNOPSIS. During these last three lectures, I will discuss the structure of differential forms form the perspective of multi-linear algebra and n-forms on vector spaces. This is basically not done in the book. This allows me to give a much more foundational treatment of just what forms are and not just how they work. We learn their structure, how to integrate them and how to differentiate them, all with an eye toward what works regardless of the dimension. We show how many of the things we learned in the past, from the product rule and the Substitution Method in Calculus I to the Change of Variables Theorem and Fubini's Theorem in Calculus III, are all just examples of more general structure. We then finish with the Generalized Stoke's Theorem, and show how the various big theorems of Gauss, Stokes and Green are also simply particular examples. We end with the same result of the Fundamental Theorem of Calculus. In fact, one can easily say that the Generalized Stokes Theorem is just the Fundamental Theorem of Multivariable Calculus.

23.0.1. Multilinear algebra. Let V be an n-dimensional vector space on $\mathbb{R}$. Then, relative to some basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$, (note that we will use the notation of the *standard basis* in $\mathbb{R}^n$ for convenience, but one can adapt this argument to any basis) any element $\mathbf{v} \in V$ can be written as

$$\mathbf{v} = v_1\mathbf{e}_1 + \ldots + v_n\mathbf{e}_n.$$

Here, v_i is the ith coordinate of $\mathbf{v}$ (in the given basis) and, by convention, one often denotes elements of V by their set of coordinates in the form of a *(column) vector*

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \in V, \quad v_i \in \mathbb{R}, \quad \text{for } i = 1, \ldots, n,$$

so that we say $\mathbf{v} \in V$ is a *vector* in V.

A *linear functional*, or *(linear) 1-form*, or *covector*, is a linear map $f : V \to \mathbb{R}$, so satisfies

$$f(c_1\mathbf{v} + c_1\mathbf{w}) = c_1 f(\mathbf{v}) + c_1 f(\mathbf{w}), \quad \forall \mathbf{v}, \mathbf{w} \in V, \quad \forall c_1, c_2 \in \mathbb{R}.$$

The set of all covectors of V is again a n-dimensional vector space V^* called the *dual space* to V.

EXERCISE 13. Show that V^* is a vector space.

So what is a basis for V^*? For each $i = 1, \ldots, n$, let $\mathbf{e}_i : V \to \mathbb{R}$ be defined so that

$$\mathbf{e}_i^*(\mathbf{e}_j) = \begin{cases} 1 & i = j \\ 0 & i \neq j. \end{cases}$$

Then, by linearity, $\mathbf{e}_i^*(\mathbf{w}) = w_i$, and the ith basis covector strips off the ith entry of $\mathbf{w}$. One can readily show that $\{\mathbf{e}_1^*, \ldots, \mathbf{e}_n^*\}$ forms a basis for V^*, and that for any $\mathbf{v}^* \in V^*$,

$$\mathbf{v}^* = v_1\mathbf{e}_1^* + \ldots + v_n\mathbf{e}_n^*, \quad v_i \in \mathbb{R}.$$

Here, $\mathbf{v}^* : V \to \mathbb{R}$ satisfies

$$\begin{aligned} \mathbf{v}^*(\mathbf{w}) &= v_1\mathbf{e}_1^*(\mathbf{w}) + \ldots + v_n\mathbf{e}_n^*(\mathbf{w}) \\ &= v_1w_1 + \ldots + v_nw_n = \mathbf{v} \cdot \mathbf{w} \\ &= \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}. \end{aligned}$$

Some notes:

- In this way, we often write covectors as *row vectors*, since written this way, they can readily "act" on vectors as functionals.
- The dot product $\mathbf{dot} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is not a linear function. It is linear, however, on each of its factors separately, and is an example of a *mulitlinear function*. Indeed, for $v \in V$, the dot product $\mathbf{dot}(\mathbf{v}, \cdot) = \mathbf{dot}_\mathbf{v}(\cdot)$ with one slot filled is a linear functional, so that one can identify, for $\mathbf{v}^* \in V^*$,

  $$\mathbf{v}^*(\mathbf{w}) = \mathbf{dot}_\mathbf{v}(\mathbf{w}) = \mathbf{v} \cdot \mathbf{w},$$

  as before.
- In $\mathbb{R}^3$, each $\mathbf{p} \in \mathbb{R}^3$ has a *tangent space* $T_\mathbf{p}\mathbb{R}^3$, which is another copy of $\mathbb{R}^3$, but with its origin based at $\mathbf{p}$. It is a different space than the one where $\mathbf{p}$ "lives".

On this last bullet point, for coordinates $(x_1, \ldots, x_n)$ on $\mathbb{R}^n$, define a coordinate system on $T_\mathbf{p}\mathbb{R}^n$ as $(dx_1, \ldots, dx_n)$, where each dx_i is the infinitesimal change in the coordinate x_i at $\mathbf{p}$ in $\mathbb{R}^n$, but ranges over all real numbers in a particular direction in $T_\mathbf{p}\mathbb{R}^n$. Here, each dx_i is a linear functional on $T_\mathbf{p}\mathbb{R}^n$ since, for a choice of $\mathbf{v} \in T_\mathbf{p}\mathbb{R}^n$, $dx_i(\mathbf{v}) = v_i$.

Some notes:

- Think of a parameterized hypersurface $\mathcal{S} \in \mathbb{R}^n$, and it is easier to see how a tangent vector $\mathbf{v} \in T_\mathbf{p}\mathcal{S}$, but $\mathbf{v} \notin \mathcal{S}$.
- This definition of dx_i works because coordinates themselves are actually linear functionals on a space (at least the Cartesian one are). They are projections onto the factors of the space, which are linear functions. INdeed, let $\mathbf{p} = (p_1, p_2) \in \mathbb{R}^2$. Then the functions $x : \mathbb{R}^2 \to \mathbb{R}$ and $y : \mathbb{R}^2 \to \mathbb{R}$ can be defined as $x(\mathbf{p}) = p_1$ and

$y(\mathbf{p}) = p_2$. These coordinate functions are linear and hence are not only continuous but differentiable, and the derivative functions are

$$\begin{aligned} Dx_{\mathbf{p}} &: T_{\mathbf{p}}\mathbb{R}^2 \to \mathbb{R}, \quad Dx_{\mathbf{p}} = \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad \text{and} \\ Dy_{\mathbf{p}} &: T_{\mathbf{p}}\mathbb{R}^2 \to \mathbb{R}, \quad Dy_{\mathbf{p}} = \begin{bmatrix} 0 & 1 \end{bmatrix}, \end{aligned}$$

where each is a 1×2-matrix. Then, given $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \in T_{\mathbf{p}}\mathbb{R}^2$, we have

$$\begin{aligned} Dx_{\mathbf{p}}(\mathbf{v}) &= \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = v_1, \\ Dx_{\mathbf{p}}(\mathbf{v}) &= \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = v_1. \end{aligned}$$

So we use this to <u>define</u> coordinates directly inside $T_{\mathbf{p}}\mathbb{R}^2$, (dx, dy), where

$$dx = Dx_{\mathbf{p}} = \begin{bmatrix} 1 & 0 \end{bmatrix}, \qquad dy = Dy_{\mathbf{p}} = \begin{bmatrix} 0 & 1 \end{bmatrix}.$$

EXAMPLE 23.1. Let $\mathbf{v} \in \mathbb{R}^3$, so $\mathbf{v} = v_1\,\mathbf{i} + v_2\,\mathbf{j} + v_3\,\mathbf{k}$. Then we can write

$$\begin{aligned} x &: \mathbb{R}^3 \to \mathbb{R} \quad x(\mathbf{v}) = \mathbf{v} \bullet \mathbf{i} = \mathbf{e}_1^*(\mathbf{v}) = v_1, \\ y &: \mathbb{R}^3 \to \mathbb{R} \quad y(\mathbf{v}) = \mathbf{v} \bullet \mathbf{j} = \mathbf{e}_2^*(\mathbf{v}) = v_2, \\ z &: \mathbb{R}^3 \to \mathbb{R} \quad z(\mathbf{v}) = \mathbf{v} \bullet \mathbf{k} = \mathbf{e}_3^*(\mathbf{v}) = v_3. \end{aligned}$$

Defined, implicitly at least, this way, we often "abuse notation" for convenience and clarity of concept and simply write

$$\mathbf{v} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \in \mathbb{R}^3,$$

as one would normally see in a calculus text.

Geometrically, a linear functional on $\mathbb{R}^n$ looks like

$$\omega = a_1\,dx_1 + \ldots + a_n\,dx_n = \mathbf{a}\,d\mathbf{x},$$

where the (row matrix) covector $\mathbf{a}$ is called the *coefficient vector* of the functional, and $d\mathbf{x} = \begin{bmatrix} dx_1 \\ \vdots \\ dx_n \end{bmatrix}$ is a basis of covectors (linear functionals) in $\mathbb{R}^n$. However, we could write $\mathbf{a}$ as a column vector. If we did, then, we would be forced to write $\omega = \mathbf{a} \bullet d\mathbf{x}$. We do see this at times, and context should make it clear. See Equation 23.0.1 below.

EXAMPLE 23.2. Let $\mathbf{a} = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} -4 \\ -5 \\ -6 \end{bmatrix} \in \mathbb{R}^3$. Then

$$\omega = \mathbf{a}\, d\mathbf{x} = a_1\, dx + a_2\, dy + a_3\, dz = dx + 2\, dy + 3\, dz,$$

and

$$\begin{aligned} \omega(\mathbf{v}) &= a_1\, dx(\mathbf{v}) + a_2\, dy(\mathbf{v}) + a_3\, dz(\mathbf{v}) \\ &= a_1 v_1 + a_2 v_2 + a_3 v_3 = \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} -4 \\ -5 \\ -6 \end{bmatrix} \\ &= 1(-4) + 2(-5) = 3(-6) = -32. \end{aligned}$$

EXAMPLE 23.3. It is also a good idea to keep in mind where different mathematical objects "live": Let $\mathbf{v} = \begin{bmatrix} -1 \\ -2 \end{bmatrix} \in T_{\mathbf{p}}\mathbb{R}^2$, for $\mathbf{p} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Then, while we envision $\mathbf{v}$ as a vector in $\mathbb{R}^2$ based at $\mathbf{p}$, it is really a vector based at the origin (like vectors should) of $T_{\mathbf{p}}\mathbb{R}^2$, considered a different space.

Let $\mathbf{c} : [a, b] \to \mathbb{R}^2$ be a C^1-curve. For $t_0 \in (a, b)$, $\mathbf{p} = \mathbf{c}(t_0) \in \mathbb{R}^2$, the space $T_{\mathbf{p}}\mathbb{R}^2$ is not the same plane as $\mathbb{R}^2$. For one, it has different coordinates. We can write the tangent line $\ell_{\mathbf{p}}$ to the curve at $\mathbf{p}$ via the coordinates (dx, dy) of $T_{\mathbf{p}}\mathbb{R}^2$; This is because $\ell_{\mathbf{p}}$ is defined as the set of all tangent vectors to $\mathbf{c}$ at $\mathbf{p}$, so $\ell_{\mathbf{p}} \subset T_{\mathbf{p}}\mathbb{R}^2$, and not really in the plane where $\mathbf{c}$ is defined. In fact, the line $\ell_{\mathbf{p}}$ is a *vector subspace* of $T_{\mathbf{p}}\mathbb{R}^2$: The equation for $\ell_{\mathbf{p}}$ is

$$dy = (\text{constant})\, dx.$$

Can you guess what the constant is?

EXAMPLE 23.4. Let $\mathbf{c} : [0, 4] \to \mathbb{R}^2$ be defined by $\mathbf{c}(t) = (t, t^2)$. In the xy-plane, the equation for the tangent line to $\mathbf{c}$ at $\mathbf{p} = \mathbf{c}(1) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is

$$(y - 1) = 2(x - 1), \quad \text{or} \quad y = 2x - 1.$$

However, in $T_{\mathbf{p}}\mathbb{R}^2$, a copy of $\mathbb{R}^2$, but with the origin at $\mathbf{p}$ and coordinates (dx, dy), the equation for $\ell_{\mathbf{p}}$ is

$$dy = 2\, dx, \quad \text{or} \quad \frac{dy}{dx} = 2.$$

Just for contrast, the equation for $\ell_{\mathbf{q}} \in T_{\mathbf{q}}\mathbb{R}^2$, when $\mathbf{q} = \mathbf{c}(3) = \begin{bmatrix} 3 \\ 9 \end{bmatrix}$ is $dy = 6\,dx$.

Now compare this to the de-parameterized curve: Let $x = t$, so that $y = f(x) = x^2$. Now the curve $\mathbf{c}$ is the graph of the function $f : [0, 4] \to \mathbb{R}$ (and parameterized by x). Using dy as the infinitesimal change in y, its relationship to dx, an infinitesimal change in x, is then $dy = f'(x)\,dx = 2x\,dx$. And we are back in Calculus I.

And now, we can generalize:

DEFINITION 23.1. A *one form* on a smooth region $\mathcal{D} \subset \mathbb{R}^n$ is a choice of a linear one form on each tangent space to $\mathcal{D}$ which varies continuously with $\mathbf{p} \in \mathcal{D}$.

Some notes:

- This definition sounds a lot like that of a vector field, a choice of a vector in each tangent space to $\mathcal{D}$ which varies continuously with $\mathbf{p} \in \mathcal{D}$. It is actually quite close!
- Instead of a vector choice in our vector field, a one-form is a choice of a covector, or linear functional, in each tangent space. In this sense, a 1-form on $\mathcal{D}$ is a *covector field* on $\mathcal{D}$.

On $\mathbb{R}$, a generic 1-form looks like $\omega = f(x)\,dx$, for f a C^0-function on $\mathbb{R}$. At a point $x_0 \in \mathbb{R}$, $f(x_0) = a$, and the linear functional (the covector) at $T_{x_0}\mathbb{R}$, which is a copy of $\mathbb{R}$ but with the origin at x_0, is $\omega_{x_0} = a\,dx$. Then, for $v \in T_{x_0}\mathbb{R}$, we have

$$\omega_{x_0}(v) = a\,dx(v).$$

On $\mathbb{R}^n$, a generic 1-form looks like

$$\omega = f_1(\mathbf{x})\,dx_1 + \ldots + f_n(\mathbf{x})\,dx_n = \sum_{i=1}^{n} f_i(\mathbf{x})\,dx_i = \mathbf{F} \bullet d\mathbf{x}, \tag{23.0.1}$$

where $\mathbf{F}(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) \\ \vdots \\ f_n(\mathbf{x}) \end{bmatrix}$, and $d\mathbf{x} = \begin{bmatrix} dx_1 \\ \vdots \\ dx_n \end{bmatrix}$. Do you recognize this formula? A common way to construct 1-forms on a domain is to use vector fields as the *coefficient functions* of the form. But really, a 1-form is a covector field. We are simply writing the coefficients a column vectors instead of their more properly written row vectors. But this is what we alluded to in the discussion just after Example 23.1

Some final notes:

- This $d\mathbf{x}$ is precisely the $d\mathbf{s}$ in the definition of the vector line integral $\int_{\mathbf{c}} \mathbf{F} \bullet d\mathbf{s}$. In a sense, integrating a vector field along a curve IS just the adding up of the values of a 1-form along the curve.

- A 1-form is called a *differential 1-form* if the coefficient functions $f_i(\mathbf{x})$ are C^1-functions for all $i = 1, \ldots, n$.
- For any real-valued C^2-function $f : \mathcal{D} \subset \mathbb{R}^n \to \mathbb{R}$, its *differential*

$$df(\mathbf{x}) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{x})\, dx_i = \frac{\partial f}{\partial x_1}(\mathbf{x})\, dx_1 + \ldots + \frac{\partial f}{\partial x_n}(\mathbf{x})\, dx_n$$

  is a differential 1-form since each of the functions $\frac{\partial f}{\partial x_i}(\mathbf{x})$ is a C^1-function. But 1-forms do not have to arise in this fashion (that is, not all 1-forms arise as the differentials of functions).

LECTURE 24

More about Forms

SYNOPSIS. A continuation of the last three lectures on differential forms and their structure.

24.0.1. A covector product. Since dx and dy are linear functionals on $\mathbb{R}^2$, viewed as coordinates of $T_{\mathbf{p}}\mathbb{R}^2$ for $\mathbf{p} \in \mathcal{D} \subset \mathbb{R}^2$, they are covectors. And like vectors, beyond summing and constant multiples, one can define products of covectors. But like vectors, products of covectors are not always vectors. Think of the inner, outer, and cross products as forms of multiplication where the output may have the same or a different structure from the inputs. Here, we define a new product on covectors:

DEFINITION 24.1. The *wedge product* of two linear 1-forms ω and ν on $\mathbb{R}^n$ is

$$\omega \wedge \nu(\mathbf{v}_1, \mathbf{v}_2) = \begin{vmatrix} \omega(\mathbf{v}_1) & \omega(\mathbf{v}_2) \\ \nu(\mathbf{v}_1) & \nu(\mathbf{v}_2) \end{vmatrix} = \omega(\mathbf{v}_1)\nu(\mathbf{v}_2) - \omega(\mathbf{v}_2)\nu(\mathbf{v}_1).$$

EXERCISE 14. Show $\omega \wedge \nu : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is linear on each factor, but is not a linear function (it is multi-linear).

Here is a good geometric interpretation: Think of a plane with two vectors

$$\widehat{\mathbf{v}}_1 = \begin{bmatrix} \omega(\mathbf{v}_1) \\ \nu(\mathbf{v}_1) \end{bmatrix}, \quad \text{and} \quad \widehat{\mathbf{v}}_2 = \begin{bmatrix} \omega(\mathbf{v}_2) \\ \nu(\mathbf{v}_2) \end{bmatrix}.$$

Then $\omega \wedge \nu(\mathbf{v}_1, \mathbf{v}_2) = (\widehat{\mathbf{v}}_1 \times \widehat{\mathbf{v}}_2) \cdot \mathbf{k}$ is the *signed area* of the parallelogram in the plane whose sides are $\widehat{\mathbf{v}}_1$ and $\widehat{\mathbf{v}}_2$.

Some notes:

- The wedge product is *skew-symmetric*, or *anti-symmetric*:
$$\nu \wedge \omega(\mathbf{v}_1, \mathbf{v}_2) = -\omega \wedge \nu(\mathbf{v}_1, \mathbf{v}_2).$$
- The wedge product of two 1-forms is also anti-symmetric in its arguments:

$$\begin{aligned} \omega \wedge \nu(\mathbf{v}_2, \mathbf{v}_1) &= \omega(\mathbf{v}_2)\nu(\mathbf{v}_1) - \omega(\mathbf{v}_1)\nu(\mathbf{v}_2) \\ &= -\left(\omega(\mathbf{v}_1)\nu(\mathbf{v}_2) - \omega(\mathbf{v}_2)\nu(\mathbf{v}_1)\right) = -\omega \wedge \nu(\mathbf{v}_1, \mathbf{v}_2). \end{aligned}$$

- $\omega \wedge \nu(\mathbf{v}_1, \mathbf{v}_1) = 0$, always. (Exercise)
- $\omega \wedge \omega(\mathbf{v}_1, \mathbf{v}_2) = 0$, always. (Exercise)
- $(\omega + \mu) \wedge \nu(\mathbf{v}_1, \mathbf{v}_2) = \omega \wedge \nu(\mathbf{v}_1, \mathbf{v}_2) + \mu \wedge \nu(\mathbf{v}_1, \mathbf{v}_2)$. (Exercise)

- The wedge product of two 1-forms is multilinear:

$$\omega \wedge \nu(c_1\mathbf{w}_1 + c_2\mathbf{w}_2, \mathbf{v}) = c_1\omega \wedge \nu(\mathbf{w}_1, \mathbf{v}) + c_2\omega \wedge \nu(\mathbf{w}_2, \mathbf{v}), \quad \text{and}$$
$$\omega \wedge \nu(\mathbf{v}, c_1\mathbf{w}_1 + c_2\mathbf{w}_2) = c_1\omega \wedge \nu(\mathbf{v}, \mathbf{w}_1) + c_2\omega \wedge \nu(\mathbf{v}, \mathbf{w}_2).$$

For 1-forms ω and ν on $\mathbb{R}^n$, $\omega\wedge\nu$ is called a *2-form*, where $\omega\wedge\nu : \mathbb{R}^n\times\mathbb{R}^n \to \mathbb{R}$ acts on paris of vectors. Indeed, let $\omega = \sum a_i\, dx_i$ and $\nu = \sum b_j\, dx_j$ be 2 linear 1-forms. Then

$$\omega \wedge \nu = \sum_{i,j=1}^{n} a_i\, dx_i \wedge b_j\, dx_j = \sum_{i,j=1}^{n} a_i b_j\, dx_i \wedge dx_j,$$

since forms are linear on each factor. However, when $i = j$, Definition 24.1 implies

$$dx_i\wedge dx_i(\mathbf{v}_1, \mathbf{v}_2) = \begin{vmatrix} dx_i(\mathbf{v}_1) & dx_i(\mathbf{v}_2) \\ dx_i(\mathbf{v}_1) & dx_i(\mathbf{v}_2) \end{vmatrix} = dx_i(\mathbf{v}_1)\, dx_i(\mathbf{v}_2) - dx_i(\mathbf{v}_2)\, dx_i(\mathbf{v}_1) = 0.$$

Couple this with the skew-symmetry of the wedge product, so that $dx_i \wedge dx_j = -dx_j \wedge dx_i$, and it becomes obvious that many terms in a generic 2-form can be neglected or combined, simplifying the form.

EXERCISE 15. Show that, for $\omega = \sum_{i=1}^{n} f_i(x)\, dx_i$ and $\nu = \sum_{j=1}^{n} g_j(x)\, dx_j$, the 2-form $\omega \wedge \nu$ is a sum of, at most, $\binom{n}{2}$ distinct, non-zero terms, which are some function times $dx_i \wedge dx_j$, after all simplifications.

EXAMPLE 24.1. In $\mathbb{R}^3$, with coordinates x, y, and z, let

$$\omega = a_1\, dx + a_2\, dy + a_3\, dz$$
$$\nu = b_1\, dx + b_2\, dy + b_3\, dz.$$

Then

$$\begin{aligned} \omega \wedge \nu &= a_1b_1\, dx \wedge dx + a_1b_2\, dx \wedge dy + a_1b_3\, dx \wedge dz \\ &\quad + a_2b_1\, dy \wedge dx + a_2b_2\, dy \wedge dy + a_2b_3\, dy \wedge dz \\ &\quad + a_3b_1\, dz \wedge dx + a_3b_2\, dz \wedge dy + a_3b_3\, dz \wedge dz \\ &= (a_2b_3 - a_3b_2)\, dy \wedge dz + (a_3b_1 - a_1b_3)\, dz \wedge dx + (a_1b_2 - a_2b_1)\, dx \wedge dy. \end{aligned}$$

Do you recognize the structure of the coefficient (row)-vector here?

DEFINITION 24.2. A *differential 2-form* on a region in $\mathbb{R}^n$ is just a choice of a linear 2-form on each tangent space to the region that varies differentiably with respect to the region points.

For $\omega = \sum_{i=1}^{n} f_i(\mathbf{x})\,dx_i$ and $\nu = \sum_{j=1}^{n} g_j(\mathbf{x})\,dx_j$, we have

$$\omega \wedge \nu = \sum_{i,j=1}^{n} h_{ij}(\mathbf{x})\,dx_i \wedge dx_j,$$

with all appropriate cancellations and skew symmetries.

EXAMPLE 24.2. Let $\omega = x^2 y\,dx \wedge dy - xz\,dy \wedge dz$ be a 2-form on $\mathbb{R}^3$, and $\mathbf{p} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$. Then, at $\mathbf{p}$, we have

$$\begin{aligned} \omega_{\mathbf{p}} &= \left(x^2 y\Big|_{\mathbf{p}}\right) dx \wedge dy + \left(xz\Big|_{\mathbf{p}}\right) dy \wedge dz \\ &= 2\,dx \wedge dy - 3\,dy \wedge dz, \end{aligned}$$

a linear 2-form on $T_{\mathbf{p}}\mathbb{R}^3$. Now, if we choose two vectors $\mathbf{v}_1 = \begin{bmatrix} 4 \\ 0 \\ -1 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}$ in $T_{\mathbf{p}}\mathbb{R}^3$, then

$$\begin{aligned} \omega_{\mathbf{p}}(\mathbf{v}_1, \mathbf{v}_2) &= 2\,dx \wedge dy\left(\begin{bmatrix} 4 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}\right) - 3\,dy \wedge dz\left(\begin{bmatrix} 4 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}\right) \\ &= 2\begin{vmatrix} 4 & 3 \\ 0 & 1 \end{vmatrix} - 3\begin{vmatrix} 0 & 1 \\ -1 & 2 \end{vmatrix} = 2(4) - 3(1) = 5. \end{aligned}$$

Note that

$$\begin{vmatrix} 4 & 3 \\ 0 & 1 \end{vmatrix} = \begin{vmatrix} dx\left(\begin{bmatrix} 4 \\ 0 \\ -1 \end{bmatrix}\right) & dx\left(\begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}\right) \\ dy\left(\begin{bmatrix} 4 \\ 0 \\ -1 \end{bmatrix}\right) & dy\left(\begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}\right) \end{vmatrix}.$$

Written out, then, a 2-form on $\mathbb{R}^3$, using the coordinates x, y, and z, looks like

$$\mu = f_1(x, y, z)\,dx \wedge dy + f_2(x, y, x)\,dx \wedge dz + f_3(x, y, z)\,dy \wedge dz.$$

And in the case that μ is itself the wedge product of two 1-forms,

$$\omega = g_1\,dx + g_2\,dy + g_3\,dz, \quad \text{and} \quad \nu = h_1\,dx + h_2\,dy + h_3\,dz,$$

then

$$f_1(x,y,z) = g_1(x,y,z)h_2(x,y,z) - g_2(x,y,z)h_1(x,y,z),$$

with the other two coefficients defined similarly.

Notes:

(1) Completely generalizes to $\mathbb{R}^n$, $n > 3$ with a very similar structure.
(2) A generic way to write a differential 2-form on $\mathbb{R}^n$ with coordinates $x_1, \ldots, x_n$ is
$$\mu = \sum_{i,j=1}^{n} f_{ij}\, dx_i \wedge dx_j$$
and leave all cancellations and skew-symmetries up to the reader.
(3) We can continue to construct higher-order forms via the wedge product:
   - Let ω_i be a set of m differentiable 1-forms on $\mathbb{R}^n$, for $i = 1, \ldots, m$. Then
   $$\mu = \omega_1 \wedge \cdots \wedge \omega_m$$
   is a differential m-form on $\mathbb{R}^n$ which will ultimately look like
   $$\mu = \sum F_{i_1 \cdots i_k}\, dx_{i_1} \wedge \cdots \wedge dx_{i_k}$$
   with a lot of terms vanishing and other simplifications. At a point $\mathbf{p} \in \mathbb{R}^n$,
   $$\mu_{\mathbf{p}} : \underbrace{T_{\mathbf{p}}\mathbb{R}^n \times \cdots \times T_{\mathbf{p}}\mathbb{R}^n}_{m\text{-terms}} \to \mathbb{R},$$
   is a linear m-form, and
   $$\mu_{\mathbf{p}}(\mathbf{v}_1, \ldots, \mathbf{v}_m) = \begin{vmatrix} \omega_1(\mathbf{v}_1) & \cdots & \omega_1(\mathbf{v}_m) \\ \vdots & \ddots & \vdots \\ \omega_m(\mathbf{v}_1) & \cdots & \omega_m(\mathbf{v}_m) \end{vmatrix}.$$
   The computation is a very mechanical process.
   - Note also that $\mu_{\mathbf{p}}$, a linear m-form, is multilinear, so linear on each argument:

$$\begin{aligned}\mu_{\mathbf{p}}(\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}, c_1\mathbf{u}_1 + c_2\mathbf{u_2}, \mathbf{v}_{i+1}, \ldots, \mathbf{v}_m) \\ &= c_1\mu_{\mathbf{p}}(\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}, \mathbf{u}_1, \mathbf{v}_{i+1}, \ldots, \mathbf{v}_m) \\ &\quad + c_2\mu_{\mathbf{p}}(\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}, \mathbf{u}_2, \mathbf{v}_{i+1}, \ldots, \mathbf{v}_m).\end{aligned}$$

   - And for ω a differentiable k-form and ν a differentiable ℓ-form, we have $\omega \wedge \nu$ is a differentiable $(k + \ell)$-form
(4) Call a C^1-function f on $\mathbb{R}^n$ a differentiable 0-form. Then, for ω a differentiable k-form,
$$\begin{aligned} f \wedge \omega &= f \cdot \left(\sum F_{i_1 \cdots i_k}\, dx_{i_1} \wedge \cdots \wedge dx_{i_k}\right) \\ &= \sum f \cdot F_{i_1 \cdots i_k}\, dx_{i_1} \wedge \cdots \wedge dx_{i_k}\end{aligned}$$
is still a k-form.

(5) There exist 2-forms that do not arise as the wedge products of 1-forms.
(6) There are no m-forms on $\mathbb{R}^n$, where $m > n$. Why not?
(7) An n-form on $\mathbb{R}^n$ is also called a *volume form.*
(8) The wedge product is also called the *exterior product* on forms.

DEFINITION 24.3. A differential m-form on $\mathbb{R}^n$, $n \geq m$

$$\mu = \sum_{i_1,\ldots,i_m=1}^{n} F_{i_1\cdots i_m}\, dx_{i_1} \wedge \cdots \wedge dx_{i_m}$$

is a continuous family of linear m-forms $\mu_{\mathbf{p}}$, parameterized by $\mathbf{p} \in \mathbb{R}^n$, such that, at each $\mathbf{p}$,

$$\mu_{\mathbf{p}} : \underbrace{T_{\mathbf{p}}\mathbb{R}^n \times \cdots \times T_{\mathbf{p}}\mathbb{R}^n}_{m\text{-terms}} \to \mathbb{R}$$

is multilinear, so while not a linear function, it is linear on each argument separately.

Here is an alternate view: For each $\mathbf{p} \subset \mathcal{D}\mathbb{R}^n$, and each factor $T_{\mathbf{p}}\mathbb{R}^n$ of $\mu_{\mathbf{p}}$, a choice of $\mathbf{v}_{\mathbf{p}} \in T_{\mathbf{p}}\mathbb{R}^n$ is a vector field on $\mathcal{D} \subset \mathbb{R}^n$. Hence a differential m-form on $\mathcal{D}$ "acts" on a set of m vector fields on $\mathcal{D}$ simultaneously, and returns a function on $\mathcal{D}$.

EXAMPLE 24.3. Let $\mathbf{F} = 2y\,\mathbf{i} - x\,\mathbf{k}$ be a vector field on $\mathbb{R}^3$, and

$$\omega = x^2 y\, dx - x\, dy + y^2 z\, dz$$

be a differentiable 1-form. Then, at any given $\mathbf{p} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ and $\mathbf{v} \in \mathbf{F}$, we have

$$\begin{aligned} \omega(\mathbf{v}) &= (x^2 y\, dx - x\, dy + y^2 z\, dz)(\mathbf{v}) \\ &= x^2 y\, dx(\mathbf{v}) - x\, dy(\mathbf{v}) + y^2 z\, dz(\mathbf{v}) \\ &= x^2 y(2y) - x(0) + y^2 z(-x) = 2x^2 y^2 - x y^2 z. \end{aligned}$$

Hence we can write $\omega(\mathbf{F}) : \mathbb{R}^n \to \mathbb{R}$, as

$$\omega(\mathbf{F})(x, y, x) = 2x^2 y^2 - x y^2 z.$$

24.0.2. Integrating forms. Yet another alternate view: Forms are "generalized" integrands; One can add them up on each tangent space over a domain. Interpreting them as such and also geometrically will depend on the form and the space we integrate over, though. Here are some examples:

I. Integrating n-forms in $\mathbb{R}^n$:

(a) $n-1$: Let $f:\mathbb{R}\to\mathbb{R}$ be C^0. Then $\omega = f(x)\,dx$ is a continuous 1-form. For $\mathcal{I}=[a,b]\subset\mathbb{R}$ an interval (already parameterized),

$$\int_{\mathcal{I}}\omega = \int_a^b f(x)\,dx.$$

Any continuous function $f(x)$ on an interval $\mathcal{I}\subset\mathbb{R}$ can be associated to a 1-form in this way, and integrating this form on the interval is performed in the fashion one would employ in first semester, single variable calculus.

(b) If $\omega = f(x,y)\,dx\wedge dy$, on a region $\mathcal{D}\subset\mathbb{R}^2$, then

$$\int_{\mathcal{D}}\omega = \iint_{\mathcal{D}} f(x,y)\,dx\,dy. \tag{24.0.1}$$

(c) This generalizes quite naturally, and if $\omega = f(\mathbf{x})\,dx_1\wedge\cdots\wedge dx_n$ is a continuous $n-form$ in $\mathbb{R}^n$, then on some compact, n-dimensional region $\mathcal{R}\subset\mathbb{R}^n$, we can write

$$\int_{\mathcal{R}}\omega = \underbrace{\iint\cdots\int_{\mathcal{R}}}_{n\text{-integrals}} f(mathbfx)\,dx_1\cdots dx_n. \tag{24.0.2}$$

This helps to understand the use of the term *volume form* for an n-form in $\mathbb{R}^n$; If $f(\mathbf{x})\equiv 1$, then integrating ω over $\mathcal{R}$ yields the volume of $\mathcal{R}$:

- For a form defined on a region, we always integrate the form over a subset of that region of the same "size" as that of the order of the form.
- Notice that in both Eqns 24.0.1 and 24.0.2, we use only a single integral sign, even for multiple integrals. The form and the region identify the type of integration, so we do not need to use multiple integrals. This will be very useful later.

I. Integrating m-forms in $\mathbb{R}^n$, $m<n$: Here, we highlight, via a few examples, the integration of forms on spaces and see how one can interpret these quantities in ways that we have already developed and discussed, but using this new language of forms:

EXAMPLE 24.4. ***Integrating 1-forms and the circulation of a vector field.*** Let $m=1$. Then, on a curve $\mathbf{c}:[a,b]\to\mathbb{R}^n$, with

$$\omega = f_1(\mathbf{x})\,dx_1 + \ldots + f_n(\mathbf{x})\,dx_n = \sum_{i=1}^{n} f_i(\mathbf{x})\,dx_i,$$

we can write $\omega = \mathbf{F}\bullet d\mathbf{s}$, where $\mathbf{F}(\mathbf{x}) = \begin{bmatrix} f_1 \\ \vdots \\ f_n \end{bmatrix}$ can be interpreted as a vector field, and $d\mathbf{s} = \begin{bmatrix} dx_1 \\ \vdots \\ dx_n \end{bmatrix}$ is a vector form of infinitesimal displacement. Then we have

$$\begin{aligned}\int_{\mathbf{c}} \omega = \int_{\mathbf{c}} \mathbf{F} \bullet d\mathbf{s} &= \int_{\mathbf{c}} \omega_{\mathbf{c}}\left(\mathbf{c}(t)\right)\, dt \\ &= \int_a^b \mathbf{F}\left(\mathbf{c}(t)\right) \bullet d\mathbf{s}\left(\mathbf{c}'(t)\right) \\ &= \int_a^b \mathbf{F}\left(\mathbf{c}(t)\right) \bullet \begin{bmatrix} dx_1\left(\mathbf{c}'(t)\right) \\ \vdots \\ dx_n\left(\mathbf{c}'(t)\right) \end{bmatrix} dt \\ &= \int_a^b \mathbf{F}\left(\mathbf{c}(t)\right) \bullet \mathbf{c}'(t)\, dt.\end{aligned}$$

Hence our interpretation is that integrating a 1-form over a curve in $\mathbb{R}^n$ is the same as calculating the circulation of a vector field over the curve (a vector line integral of the field), when the vector field is the coefficient vector of the form. In this interpretation, we have

$$d\mathbf{s} = \begin{bmatrix} dx_1 \\ \vdots \\ dx_n \end{bmatrix} = \begin{bmatrix} x_1'(t)\, dt \\ \vdots \\ x_n'(t)\, dt \end{bmatrix} = \begin{bmatrix} x_1'(t) \\ \vdots \\ x_n'(t) \end{bmatrix} dt = \mathbf{c}'(t)\, dt.$$

EXAMPLE 24.5. ***Integrating 2-forms and surface integrals.*** For $m = 2$. Then, on a surface $\mathbf{X} : \mathcal{D} \subset \mathbb{R}^2 \to \mathbb{R}^n$, with

$$\omega = \sum_{i,j=1}^{n} F_{ij}(\mathbf{x})\, dx_i \wedge dx_j, \text{(neglecting simplifications)}$$

we have, with $\mathbf{X}(\mathcal{D}) = \mathcal{R} \subset \mathbb{R}^n$,

$$\int_{\mathcal{R}} \omega = \int_{\mathbf{X}} \omega_{\mathbf{X}}\left(\mathbf{X}_s, \mathbf{X}_t\right)\, ds \wedge dt,$$

where

$$\mathbf{X}_s = \begin{bmatrix} \frac{\partial x_1}{\partial s}(s,t) \\ \vdots \\ \frac{\partial x_1}{\partial s}(s,t) \end{bmatrix}, \quad \text{and} \quad \mathbf{X}_t = \begin{bmatrix} \frac{\partial x_1}{\partial t}(s,t) \\ \vdots \\ \frac{\partial x_1}{\partial t}(s,t) \end{bmatrix}$$

are the partial derivative vectors determined by the parameterization $\mathbf{X}(s,t) = (x_1(s,t), \ldots, x_n(s,t))$. Thus, we have

$$\int_{\mathcal{R}} \omega = \int_{\mathbf{X}} \omega_{\mathbf{X}}(\mathbf{X}_s, \mathbf{X}_t)\, ds \wedge dt = \iint_{\mathcal{R}} \mathbf{F} \bullet d\mathbf{S},$$

which looks something like a vector surface integral, but where *we need to understand and interpret these pieces appropriately.* Here $\mathbf{F}$ is a vector of all of the $\mathbf{F}_{ij}$s, but is not a vector field in the sense that we have already defined it. For evidence of this, there will be $\binom{n}{2}$ elements in $\mathbf{F}$ after simplifications. This is typically too many to actually serve as a vector field. Also, the vector $d\mathbf{S}$ is a vector of the associated term $dx_i \wedge dx_j$, and is also a vector of size $\binom{n}{2}$, too many to serve geometrically as infinitesimal displacement like it did in $\mathbb{R}^3$. However, at least formally, we can write

$$\int_{\mathbf{X}} \omega_{\mathbf{X}}(\mathbf{X}_s, \mathbf{X}_t)\, ds \wedge dt = \int_{\mathcal{D}} \mathbf{F}(\mathbf{X}(s,t)) \bullet d\mathbf{S}(\mathbf{X}_s, \mathbf{X}_t),$$

where, for each term in $d\mathbf{S}$, we have

$$dx_i \wedge dx_j(\mathbf{X}_s, \mathbf{X}_t) = \begin{vmatrix} dx_i(\mathbf{X}_s) & dx_i(\mathbf{X}_t) \\ dx_j(\mathbf{X}_s) & dx_j(\mathbf{X}_t) \end{vmatrix} ds \wedge dt = \begin{vmatrix} \frac{\partial x_i}{\partial s} & \frac{\partial x_i}{\partial t} \\ \frac{\partial x_j}{\partial s} & \frac{\partial x_j}{\partial t} \end{vmatrix} ds \wedge dt$$
$$= \frac{\partial(x_i, x_j)}{\partial(s,t)}\, ds \wedge dt.$$

For instance, if the surface $\mathbf{X} : \mathcal{D} \to \mathbb{R}^4$ with coordinates (x, y, z, u), then the 2-form ω can look like

$$\omega = F_{12}\, dx \wedge dy + F_{13}\, dx \wedge dz + F_{14}\, dx \wedge du + F_{23}\, dy \wedge dz + F_{24}\, dy \wedge du + F_{34}\, dz \wedge du.$$

Then we can write

$$\mathbf{F} = \begin{bmatrix} F_{12}(x,y,z,u) \\ F_{13}(x,y,z,u) \\ \vdots \\ F_{34}(x,y,z,u) \end{bmatrix}, \quad \text{and} \quad \mathbf{dS} = \begin{bmatrix} dx \wedge dy \\ dx \wedge dz \\ \vdots \\ dz \wedge du \end{bmatrix} = \begin{bmatrix} \frac{\partial(x,y)}{\partial(s,t)} \\ \frac{\partial(x,z)}{\partial(s,t)} \\ \vdots \\ \frac{\partial(z,u)}{\partial(s,t)} \end{bmatrix} ds \wedge dt,$$

even as these $6 = \binom{4}{2}$-vectors do not correspond to geometric objects on $\mathbb{R}^4$.

EXAMPLE 24.6. ***Integrating 2-forms in $\mathbb{R}^3$ and the flux of a vector field.*** In the special case where the surface in in $\mathbb{R}^3$, so $n = 3$ and $m = 2$, we do have a good geometric interpretation of the integral of a 2-form on the surface. For a parameterized surface $\mathcal{S} \subset \mathbb{R}^3$, with $\mathbf{X} : \mathcal{D} \subset \mathbb{R}^2 \to \mathbb{R}^3$, and $\mathcal{S} = \mathbf{X}(\mathcal{D})$, we have ω with only $3 = \binom{3}{2}$ terms, so

$$\omega = F_{12}(\mathbf{x})\, dx \wedge dy + F_{13}(\mathbf{x})\, dz \wedge dx + F_{23}(\mathbf{x})\, dx \wedge dy.$$

Then we have

$$\int_{\mathcal{S}} \omega = \int_{\mathcal{D}} \omega_{\mathbf{X}}\left(\mathbf{X}_s, \mathbf{X}_t\right)\, ds \wedge dt = \int_{\mathcal{S}} \mathbf{F} \bullet d\mathbf{S}.$$

where $\mathbf{F}(\mathbf{x}) = F_{12}(\mathbf{x})\,\mathbf{i} + F_{13}(\mathbf{x})\,\mathbf{j} + F_{23}(\mathbf{x})\,\mathbf{k}$ is an actual vector field on $\mathbb{R}^3$ (it has the correct dimension), and

$$d\mathbf{S} = \begin{bmatrix} dy \wedge dz \\ dz \wedge dx \\ dx \wedge dy \end{bmatrix} = \begin{bmatrix} \frac{\partial(y,z)}{\partial(s,t)} \\ \frac{\partial(z,x)}{\partial(s,t)} \\ \frac{\partial(x,y)}{\partial(s,t)} \end{bmatrix} ds \wedge dt.$$

Hence we can interpret the integration of a 2-form on a surface in $\mathbb{R}^3$ as the calculation of the flux of a vector field through a surface, where the vector field is the coefficient vector of the 2-form. This is a *vector surface integral* of the field.

Special note here: Notice how we defined the "middle" term in ω here using $dz \wedge dx$ instead of $dx \wedge dz$. Geometrically speaking, there is a reason for this, and we will get to that with a little more structure in the next (and final) lecture. But for now, to associate the three functions F_{12}, F_{13}, and F_{23} with the actual components of a vector field $\mathbf{F}$, we need to address an issue of a minus sign introduced in this middle term. We choose to respect the minus sign by reversing the terms in the middle wedge. Keep track of this and hold your thoughts for now.

More generally, let $\mathcal{R} \subset \mathbb{R}^n$ be an m-dimensional region parameterized by $\mathbf{X} : \mathcal{D} \subset \mathbb{R}^m \to \mathbb{R}^n$, where $\mathbf{X}(\mathcal{D}) = \mathcal{R}$. Then, for ω a differential m-form on $\mathbb{R}^n$,

$$\omega\Big|_{\mathbf{X}(\mathcal{D})} = \omega_{\mathbf{X}(\mathcal{D})} = \omega_{\mathbf{X}}$$

is an m-form on $\mathcal{R}$ which can be expressed and integrated via the parameterization. Indeed, with the coordinates $(s, t, \ldots, u)$ for $\mathcal{D} \in \mathbb{R}^m$, we have

$$\begin{aligned} \int_{\mathcal{R}} \omega &= \int_{\mathbf{X}} \omega_{\mathbf{X}}\left(\mathbf{X}_s, \mathbf{X}_t, \ldots, \mathbf{X}_u\right)\, ds \wedge dt \wedge \cdots \wedge du \\ &= \int_{\mathcal{R}} \mathbf{F} \bullet d\mathbf{S}, \end{aligned}$$

where the vector $\mathbf{F}$ of all of the form coefficient functions $F_{i_1 i_2 \cdots i_m}$, and the vector $d\mathbf{S}$, containing all of the respective form pieces $dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_m}$, will each have only $\binom{n}{m}$-distinct elements. They will not generally correspond to vector fields in the way we define them (there are often too many elements). We will go any deeper than this here, but do note the following: Things tend to line up well, combinatorially, when one

integrates a 1-form over a curve (the number of elements if $\mathbf{F}$ and $d\mathbf{s}$ is $\binom{n}{1} = n$), and when integrating an $(n-1)$-form over a hypersurface (of dimension again $(n-1)$) since then, again, the constituents vectors $\mathbf{F}$ and $d\mathbf{S}$ have size $\binom{n}{n-1} = n$ each. At that point, we can again think of $\mathbf{F}$ as a vector field.

EXERCISE 16. Let $\mathbf{F} = \begin{bmatrix} y \\ x \end{bmatrix}$ a vector field on $\mathbb{R}^2$ and $ds = \begin{bmatrix} dx \\ dy \end{bmatrix}$, the quantity $\omega = \mathbf{F} \bullet d\mathbf{s}$ is the 1-form $\omega = y\,dx + x\,dy$. Calculate $\int_{\mathbf{c}} \omega$, where $\mathbf{c} : [0,2] \to \mathbb{R}^2$ is defined by $\mathbf{c}(t) = \begin{bmatrix} t^2 \\ t^3 \end{bmatrix}$. (Hint: The answer is 32.)

And now we end this lecture with some examples of techniques and quantities you are already familiar with, but now revisited in the new language of forms.

EXAMPLE 24.7. ***Integrating a 1-form in $\mathbb{R}$ and the Substitution Method in Calculus I.*** Let $\omega = f(u)\,du$, a differential 1-form on $\mathcal{I} = [c,d]$. Then $\displaystyle\int_{\mathcal{I}} \omega = \int_a^b f(u)\,du$ like in Calculus I. But let's reparameterize $\mathcal{I}$ via the function

$$g : \mathcal{J} \to \mathcal{I}, \quad g : [a,b] \to [c,d]$$

so that $u = g(x)$, and $c = u(a)$ and $d = u(b)$.

Now using the reparameterization, we get

$$\int_{\mathcal{I}} \omega = \int_{\mathcal{J}} \omega_g\left(g'(x)\right)\,dx = \int_a^b f\left(g(x)\right) \cdot g'(x)\,dx,$$

so that

$$\int_{c=g(a)}^{d=g(b)} f(u)\,du = \int_a^b f\left(g(x)\right) \cdot g'(x)\,dx.$$

Do you remember the structure of the *Substitution Method* in first semester calculus?

EXAMPLE 24.8. ***Fubini and the Change of Variables Theorem in the language of forms.*** Here is a curious and beautiful fact: Recall that forms are skew-symmetric, so $dx \wedge dy = -dy \wedge dx$. Hence for $\omega = f(x,y)\,dx \wedge dy$, we have $-\omega = f(x,y)\,dy \wedge dx$. Let $\mathcal{R} \subset \mathbb{R}^2$ be rectangular. Then, by Fubini's Theorem,

$$\iint_{\mathcal{R}} f(x,y)\,dx\,dy = \iint_{\mathcal{R}} f(x,y)\,dy\,dx.$$

However,

$$\int_{\mathcal{R}} \omega = \iint_{\mathcal{R}} f(x,y)\,dx \wedge dy \stackrel{?}{=} \iint_{\mathcal{R}} f(x,y)\,dy \wedge dx = -\int_{\mathcal{R}} \omega.$$

What is going on here? Actually, nothing out of the ordinary. Switching the order of integration is like a reparameteriztion of the plane, from the xy-plane to the yx-plane, and the switching function is $T : (y, x) = (x, y)$. This, in fact, is an orientation-reversing reparameterization, since

$$\mathbf{Jac}(T) = \begin{vmatrix} \frac{\partial}{\partial y}[x] & \frac{\partial}{\partial y}[y] \\ \frac{\partial}{\partial x}[x] & \frac{\partial}{\partial x}[y] \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = -1.$$

Now the standard Change of Variables formula is

$$\iint_{\mathcal{R}} f(x, y)\, dx\, dy = \iint_{\mathcal{R}} f(x, y) \left| \frac{\partial(x, y)}{\partial(y, x)} \right| dy\, dx$$

with the *absolute value* of the Jacobian determinant. But the absolute value is artificial, and is a convenient shortcut to mask a much deeper structure.

Indeed, this curious fact relies on the idea that for Fubini's Theorem, we avoid orientation and forms, and simply state that the order of integration (over a rectangle) does not matter. But with forms, it does matter, as orientation is critical. In the above case, the change in orientation due to the reparameterization introduces a minus sign. But the switch from the form $dx \wedge dy$ to $dy \wedge dx$ introduces another, which conveniently cancels out the former. Simply changing the order of integration and taking the absolute value of the Jacobian works. Something to think about. Without forms (and orientation), we can conveniently simply change the order of integration in Fubini's Theorem without violating rules regarding orientation of variable changes.

EXAMPLE 24.9. ***Integrating a 2-form on a surface in three space.*** Let

$$\mathcal{M} = \left\{ (x, y, z) \in \mathbb{R}^3 \;\middle|\; z = \sqrt{1 - x^2 - y^2} \right\}$$

be the unit sphere above the xy-plane, and $\omega = z^2\, dx \wedge dy$ be a differentiable 2-form on $\mathbb{R}^3$. Evaluate $\int_{\mathcal{M}} \omega$.

STRATEGY. Parameterize the hemisphere $\mathcal{M}$ and calculate the integral via the parameterization.

SOLUTION. Use the function

$$\mathbf{X}(r, \theta) = \left(r\cos\theta, r\sin\theta, \sqrt{1 - r^2} \right)$$

so that the paramter region is the rectangle $\mathcal{D} = [0, 1] \times [0, 2\pi]$ in the $r\theta$-plane, as shown in Figure 38 below:

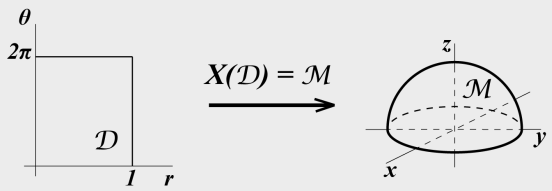


FIGURE 38. A parameterization of the unit northern hemisphere in $\mathbb{R}^3$.

Then we can calculate using the parameterization:

$$\begin{aligned}
\int_{\mathcal{M}} \omega &= \int_{\mathcal{D}} \omega_{\mathbf{X}(r,\theta)} \left(\frac{\partial X}{\partial r}(r,\theta), \frac{\partial X}{\partial \theta}(r,\theta) \right) dr \wedge d\theta \\
&= \int_{\mathcal{D}} \omega_{\mathbf{X}(r,\theta)} \left(\begin{bmatrix} \cos\theta \\ \sin\theta \\ \frac{-r}{\sqrt{1-r^2}} \end{bmatrix}, \begin{bmatrix} -r\sin\theta \\ r\cos\theta \\ 0 \end{bmatrix} \right) dr \wedge d\theta.
\end{aligned}$$

Here, think of these as part of a cylindrical coordinate system on $\mathbb{R}^3$, with the last coordinate $z = 1 - r^2$. Then continuing:

$$\int_{\mathcal{M}} \omega = \int_{\mathcal{D}} (1 - r^2) \begin{vmatrix} dx\left(\begin{bmatrix} \cos\theta \\ \sin\theta \\ \frac{-r}{\sqrt{1-r^2}} \end{bmatrix} \right) & dx\left(\begin{bmatrix} -r\sin\theta \\ r\cos\theta \\ 0 \end{bmatrix} \right) \\ dy\left(\begin{bmatrix} \cos\theta \\ \sin\theta \\ \frac{-r}{\sqrt{1-r^2}} \end{bmatrix} \right) & dy\left(\begin{bmatrix} -r\sin\theta \\ r\cos\theta \\ 0 \end{bmatrix} \right) \end{vmatrix} dr \wedge d\theta$$

$$= \int_{\mathcal{D}} (1 - r^2) \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{vmatrix} dr \wedge d\theta = \int_{\mathcal{D}} (1 - r^2) \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} dr \wedge d\theta$$

$$= \int_{\mathcal{D}} (1 - r^2) r\, dr \wedge d\theta = \int_0^{2\pi} \int_0^1 (r - r^3)\, dr\, d\theta$$

$$= \int_0^{2\pi} \left[\left(\frac{r^2}{2} - \frac{r^4}{4} \right) \Big|_0^1 \right] d\theta = \frac{\theta}{4} \Big|_0^{2\pi} = \frac{\pi}{2}.$$

LECTURE 25

Generalized Stokes' Theorem

SYNOPSIS. A continuation of the last three lectures on differential forms and their structure.

25.0.1. More notation. For $\omega = \sum F_{i_1 i_2 \cdots i_m} dx_{i_1} \wedge \cdots \wedge dx_{i_m}$ a differential m-form on $\mathcal{M} \subset \mathbb{R}^n$, $n \geq m$,

$$\int_{\mathcal{M}} \omega = \underbrace{\int \cdots \int_{\mathcal{M}}}_{n\text{-integrals}} \sum F_{i_1 i_2 \cdots i_m} dx_{i_1} \wedge \cdots \wedge dx_{i_m},$$

where $\mathcal{M}$ is an m-dimensional region in $\mathbb{R}^n$. Note that the order of the form and the dimension of the region integrated over will agree.

DEFINITION 25.1. Let $f : \mathcal{D} \subset \mathbb{R}^n \to \mathbb{R}$ be a C^1-function. Then the *exterior derivative* of f, denoted df, is the 1-form

$$df = \frac{\partial f}{\partial x_1} dx_1 + \ldots + \frac{\partial f}{\partial x_n} dx_n = Df(\mathbf{x})\, d\mathbf{x} = \nabla f \bullet d\mathbf{x}.$$

For $\omega = \sum F_{i_1 i_2 \cdots i_m} dx_{i_1} \wedge \cdots \wedge dx_{i_m}$ a differential m-form, the differential $(m+1)$-form

$$d\omega = \sum d(F_{i_1 i_2 \cdots i_m}) \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_m}$$

is called the *exterior derivative* of ω.

Some notes:

- We call a C^1-function $f : \mathcal{D} \subset \mathbb{R}^n \to \mathbb{R}$ a (differential) 0-form. Thus the exterior derivative of a function is simply its differential, a 1-form. Thus theexterior derivative of any differential m-form is always an $(m+1)$-form.
- For each set of indices, the term $d(F_{i_1 i_2 \cdots i_m})$ is the standard differential of a function, and is a 1-form. Upon writing it out, one must then address any and all simplifications and cancellations, which can be many.

EXAMPLE 25.1. Let $\omega = x^2y\,dx - x\,dy$ be a C^∞ 1-form on $\mathbb{R}^2$. Then

$$\begin{aligned} d\omega &= d(x^2y) \wedge dx - d(x) \wedge dy \\ &= (2xy\,dx + x^2\,dy) \wedge dx - (1\,dx - 0\,dy) \wedge dy \\ &= 2xy\,dx \wedge dx + x^2\,dy \wedge dx - dx \wedge dy \\ &= -(1 + x^2)\,dx \wedge dy. \end{aligned}$$

So what is $d(d\omega) = d^2\omega$? (Hint: Is it possible to have a 3-form on the plane?) Here

$$d(d\omega) = d(-(1 + x^2))\,dx \wedge dy = -2x\,dx \wedge dx \wedge dy = 0. \quad \text{(Why?)}$$

EXAMPLE 25.2. Let $f : \mathbb{R}^3 \to \mathbb{R}$ be defined as $f(x, y, z) = x^2ye^{2z}$. Calculate df and $d(df) = d^2f$.

First, we have $df = 2xye^{2z}\,dx + x^2e^{2z}\,dy + 2x^2ye^{2z}\,dz$. Then

$$\begin{aligned} d(df) &= d(2xye^{2z}) \wedge dx + d(x^2e^{2z}) \wedge dy + d(2x^2ye^{2z}) \wedge dz \\ &= 2ye^{2z}\,dx \wedge dx + 2xe^{2z}\,dy \wedge dx + 4xye^{2z}\,dz \wedge dx \\ &\quad + 2xe^{2z}\,dx \wedge dy + 0\,dy \wedge dy + 2x^2e^{2z}\,dz \wedge dy \\ &\quad + 4xye^{2z}\,dx \wedge dz + 2x^2e^{2z}\,dy \wedge dz + 4x^2ye^{2z}\,dz \wedge dz \\ &= 0 \end{aligned}$$

due to skew-symmetry and term-by-term cancelations.

But this is a general feature of exterior differentiation, and has broad implications. We say that the exterior derivative is *nilpotent*; it has a positive power, in this case its square, that is 0:

PROPOSITION 25.2. *For ω a differential k-form, $d(d\omega) = d^2\omega = 0$.*

We will not prove this here, but in coordinates, the proof relies on the fact that mixed partials are equal for a sufficiently differentiable function.

EXERCISE 17. $\omega = F(x, y, z)\,dx + G(x, y, z)\,dy + H(x, y, z)\,dz + J(x, y, z)\,du$, a C^2 1-form on $\mathbb{R}^4$, show that $d^2\omega = 0$.

Here are some other properties of the exterior derivative:

(1) If ω is a k-form, and ν is an ℓ-form, then

$$d(\omega \wedge \nu) = d\omega \wedge \nu + (-1)^k\omega \wedge d\nu. \tag{25.0.1}$$

Note that we call this equation the *Wedge Product Rule* for exterior differentiation.

EXERCISE 18. Verify, using the Wedge Product Rule, that $d^2(\omega \wedge \nu) = 0$.

(2) As a special case of the Wedge Product Rule, let $k = \ell = 0$. Then $f \wedge g = f \cdot g$, since both f and g are just functions. But then the Wedge Product Rule is simply the Product Rule for the (regular) derivative of functions you learned in Calculus I. Indeed,

$$\begin{aligned} d\big(f(x)g(x)\big) &= d(f \wedge g) = df \wedge g + (-1)^0 f \wedge dg \\ &= f'(x)\,dx \cdot g(x) + f(x) \cdot g'(x)\,dx = \big(f'(x) \cdot g(x) + f(x) \cdot g'(x)\big)\,dx \\ &= df \cdot g + f \cdot dg. \end{aligned}$$

(3) Now look at forms in $\mathbb{R}^3$ only: What one sees is the following:
- d(0-form) = gradient of the coefficient function.
- d(1-form) = curl of the coefficient vector field.
- d(2-form) = divergence of the coefficient vector field.

Perhaps this is another way to think of the ideas that the curl of the gradient is always the zero vector field, and the divergence of the curl of a vector field is always 0. In the language of differential forms on $\mathbb{R}^3$, they both are just $d^2\omega = 0$.

(4) Again, only in $\mathbb{R}^3$, there is a one-to-one correspondence between 0-forms and 3-forms:

$$f(x,y,z) \longleftrightarrow f(x,y,z)\,dx \wedge dy \wedge dz.$$

And, there is a one-to-one correspondence between 1 forms and 2-forms:

$$F_1\,dx + F_2\,dy + F_3\,dz \longleftrightarrow F_1\,dx \wedge dy + F_2\,dx \wedge dz + F_3\,dy \wedge dz.$$

THEOREM 25.3 (Generalized Stokes' Theorem). *Let $\mathcal{D} \subset \mathbb{R}^k$ be a compact region with nonempty interior, and $\mathcal{M} = \mathbf{X}(\mathcal{D})$ be an oriented, parameterized k-dimensional hypersurface in $\mathbb{R}^n$, with $k \le n$ and $\partial\mathcal{M}$ oriented compatibly. Then, for a $(k-1)$-form defined on an open set in $\mathbb{R}^n$ containing $\mathcal{M}$, we have*

$$\int_{\mathcal{M}} d\omega = \int_{\partial\mathcal{M}} \omega.$$

Note: if $\mathcal{M}$ is closed, so that $\partial\mathcal{M} = \varnothing$, then $\int_{\partial\mathcal{M}} \omega = 0$, since integrating over nothing is nothing.

Now we are in a position to understand the big theorems that we have already studied individually.

25.0.2. In the language of forms, The Theorem of Gauss. In Theorem 25.3, let $n = k = 3$, and $\mathcal{M}$ be a (3-dimensional) compact solid in $\mathbb{R}^3$, with

$$\omega = F_1(\mathbf{x})\,dy \wedge dz + F_2(\mathbf{x})\,dz \wedge dx + F_3(\mathbf{x})\,dx \wedge dy$$

a differential 2-form on a superset of $\mathcal{M}$ in $\mathbb{R}^3$. Then, we know that ω is a 2-form on the closed surface $\partial\mathcal{M}$, and that one interpretation of the

integral of ω over $\partial\mathcal{M}$ is just the vector surface integral of the vector field $\mathbf{F}(\mathbf{x}) = F_1(\mathbf{x})\,\mathbf{i} + F_2(\mathbf{x})\,\mathbf{j} + F_3(\mathbf{x})\,\mathbf{k}$ over the surface, so

$$\int_{\partial\mathcal{M}} \omega = \oiint_{\partial\mathcal{M}} \mathbf{F} \bullet d\mathbf{S}, \quad \text{where} \quad d\mathbf{S} = \begin{bmatrix} dy \wedge dz \\ dz \wedge dx \\ dx \wedge dy \end{bmatrix}.$$

This is the left-hand-side of Gauss' Theorem.

For the right-hand-side of Gauss' Theorem, note that $d\omega$ will be a 3-form. We have

$$\begin{aligned}
\int_{\mathcal{M}} d\omega &= \int_{\mathcal{M}} d\left(F_1(\mathbf{x})\,dy \wedge dz + F_2(\mathbf{x})\,dz \wedge dx + F_3(\mathbf{x})\,dx \wedge dy\right) \\
&= \int_{\mathcal{M}} \left(\frac{\partial F_1}{\partial x}\,dx + \frac{\partial F_1}{\partial y}\,dy + \frac{\partial F_1}{\partial z}\,dz\right) \wedge dy \wedge dz \\
&\qquad + \left(\frac{\partial F_2}{\partial x}\,dx + \frac{\partial F_2}{\partial y}\,dy + \frac{\partial F_2}{\partial z}\,dz\right) \wedge dz \wedge dx \\
&\qquad + \left(\frac{\partial F_3}{\partial x}\,dx + \frac{\partial F_3}{\partial y}\,dy + \frac{\partial F_3}{\partial z}\,dz\right) \wedge dx \wedge dy \\
&= \int_{\mathcal{M}} \left(\frac{\partial F_1}{\partial x} + \frac{\partial F_2}{\partial y} + \frac{\partial F_3}{\partial z}\right) dx \wedge dy \wedge dz = \int_{\mathcal{M}} (\mathbf{div}\,\mathbf{F})\,dV.
\end{aligned}$$

Note that all summands with like terms in their wedge products are 0, and every permutation needed to make the only surviving term $dx \wedge dy \wedge dz$ introduces a minus sign, but there are an even number of permutations to generate the coefficient sum. The end result is precisely the right-hand-side of Gauss' Theorem. So, when the dimensions match, *The Generalized Stokes' Theorem is Gauss' theorem.*

25.0.3. In the language of forms, The Theorem of Stokes. In Theorem 25.3, let $k = 2$ and $n = 3$. In this case, let $\mathcal{D} \subset \mathbb{R}^2$ be a compact region (with boundary), and $\mathcal{S} = \mathbf{X}(\mathcal{D}) \subset \mathbb{R}^3$ be an oriented parameterized surface, with the closed curve $\partial\mathcal{S}$ oriented compatibly. And let

$$\omega = F_1(\mathbf{x})\,dx + F_2(\mathbf{x})\,dy + F(\mathbf{x})\,dz$$

be a differential 1-form, defined on a superset of $\mathcal{S}$ in $\mathbb{R}^3$. Then, we know that ω is a 1-form on the closed curve $\partial\mathcal{S}$, and that one interpretation of the integral of ω over $\partial\mathcal{S}$ is just the vector line integral (the circulation) of the vector field $\mathbf{F}(\mathbf{x}) = F_1(\mathbf{x})\,\mathbf{i} + F_2(\mathbf{x})\,\mathbf{j} + F_3(\mathbf{x})\,\mathbf{k}$ over the curve. So

$$\int_{\partial\mathcal{S}} \omega = \oint_{\partial\mathcal{S}} \mathbf{F} \bullet d\mathbf{s}, \quad \text{where} \quad d\mathbf{s} = \begin{bmatrix} dx \\ dy \\ dz \end{bmatrix}.$$

This is the left-hand-side of Stokes' Theorem.

For the right-hand-side of Stokes' Theorem, note that $d\omega$ will be a 2-form. We have

$$\begin{aligned}
\int_{\mathcal{S}} d\omega &= \int_{\mathcal{S}} d\left(F_1(\mathbf{x})\,dx + F_2(\mathbf{x})\,dy + F_3(\mathbf{x})\,dz\right) \\
&= \int_{\mathcal{S}} \left(\frac{\partial F_1}{\partial x}\,dx + \frac{\partial F_1}{\partial y}\,dy + \frac{\partial F_1}{\partial z}\,dz\right) \wedge dx \\
&\qquad + \left(\frac{\partial F_2}{\partial x}\,dx + \frac{\partial F_2}{\partial y}\,dy + \frac{\partial F_2}{\partial z}\,dz\right) \wedge dy \\
&\qquad + \left(\frac{\partial F_3}{\partial x}\,dx + \frac{\partial F_3}{\partial y}\,dy + \frac{\partial F_3}{\partial z}\,dz\right) \wedge dz \\
&= \int_{\mathcal{M}} \left(\frac{\partial F_3}{\partial y} - \frac{\partial F_2}{\partial z}\right) dy \wedge dz + \left(\frac{\partial F_1}{\partial z} - \frac{\partial F_3}{\partial x}\right) dz \wedge dx + \left(\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial z}\right) dx \wedge dy \\
&= \int_{\mathcal{S}} \nabla \times \mathbf{F} \bullet d\mathbf{S}.
\end{aligned}$$

As before, all summands with like terms in their wedge products are 0, and the minus signs come from the permutations needed to combine the remaining terms, if possible. The end result is precisely the right-hand-side of Stokes' Theorem. So, again, when the dimensions are right, *The Generalized Stokes' Theorem is Stokes' Theorem.*

25.0.4. In the language of forms, The Theorem of Green. Once more in Theorem 25.3, let $k = n = 2$. In this case, let $\mathcal{D} \subset \mathbb{R}^2$ be a compact region (with boundary), and

$$\omega = F_1\,dx + F_2(\mathbf{x})\,dy$$

be a differential 1-form, defined on a superset of $\mathcal{D}$ in $\mathbb{R}^2$. As in the discussion above involving Stokes' Theorem, integrating ω over $\partial\mathcal{D}$ is akin to calculating the circulation of $\mathbf{F}$ over $\partial\mathcal{D}$, so

$$\int_{\partial\mathcal{D}} \omega = \oint_{\partial\mathcal{D}} \mathbf{F} \bullet d\mathbf{s} = \oint_{\partial\mathcal{D}} F_1\,dx + F_2\,dy = \oint_{\partial\mathcal{D}} M(x,y)\,dx + N(x,y)\,dy,$$

where here we expose the notation used in Green's Theorem by setting $M(x,y) = F_1(x,y)$ and $N(x,y) = F_2(x,y)$. This is the left-hand-side of Green's Theorem.

For the right-hand-side of Green's Theorem, note that $d\omega$ will be a 2-form. We have

$$\begin{aligned}\int_{\mathcal{D}} d\omega &= \int_{\mathcal{D}} d\left(M(x,y)\,dx + N(x,y)\,dy\right) = \int_{\mathcal{D}} dM \wedge dx + dN \wedge dy \\ &= \int_{\mathcal{D}} \left(\frac{\partial M}{\partial x}\,dx + \frac{\partial M}{\partial y}\,dy\right) \wedge dx + \left(\frac{\partial N}{\partial x}\,dx + \frac{\partial N}{\partial y}\,dy\right) \wedge dy \\ &= \int_{\mathcal{D}} \frac{\partial M}{\partial x}\,dx \wedge dx + \frac{\partial M}{\partial y}\,dy \wedge dx + \frac{\partial N}{\partial x}\,dx \wedge dy + \frac{\partial N}{\partial y}\,dy \wedge dy \\ &= \int_{\mathcal{D}} \left(\frac{\partial N}{\partial x} - \frac{\partial M}{\partial y}\right) dx \wedge dy = \iint_{\mathcal{D}} \left(\frac{\partial N}{\partial x} - \frac{\partial M}{\partial y}\right) dx\,dy.\end{aligned}$$

Again, the end result is precisely the right-hand-side of Green's Theorem. And, once again, when the dimensions are right, *The Generalized Stokes' Theorem is Green's Theorem.*

25.0.5. In the language of forms, The Fundamental Theorem of Calculus. One last time, suppose that in Theorem 25.3, we let $k = n = 1$. Then, for $\mathcal{I} = [a,b] \subset \mathbb{R}$ a closed, bounded, (compact) interval in $\mathbb{R}$, with $\partial\mathcal{I} = \{a,b\}$ 2 points, and for ω a 0-form (just a function) on a superset of $\mathcal{I}$ in $\mathbb{R}$. Then, upon orienting $\mathcal{I}$, we automatically orient $\partial\mathcal{I}$. Going from a to b renders the orientation on $\partial\mathcal{I}$ in such a fashion that the upper endpoint is considered positive and the lower endpoint is considered negative. Using this,

$$\begin{aligned}\int_{\partial\mathcal{I}} \omega &= \text{adding up all values of } f(x) \text{ on the set of} \\ &\qquad \text{points } \{a,b\}, \text{ oriented compatibly with } \mathcal{I} \\ &= f(b) - f(a).\end{aligned}$$

This is just the right-hand-side of the Fundamental Theorem of Calculus.

Note: We have yet had no reason to understand the orientation of a discrete set of points, or a 0-dimensional set. One does this simply by assigning a plus or minus to each point separately. By convention, then, the orientation induced on the boundary of an interval upon orienting the interval assigns a minus sign to the lower point, and a plus to the higher point. So here, $f(a)$ is considered negative, and $f(b)$ is considered positive.

And for ω a 0-form (a function), We know that $d\omega$ will be a 1-form (its differential). We have

$$\int_{\mathcal{I}} d\omega = \int_{\mathcal{I}} df = \int_a^b f'(x)\,dx.$$

But putting these together, we get

$$\int_{\partial\mathcal{I}} \omega = \underbrace{f(b) - f(a) = \int_a^b f'(x)\,dx}_{\text{Fund. Thm of Calc.}} = \int_{\mathcal{I}} d\omega.$$

Hence, *The Generalized Stokes' Theorem is also just the Fundamental Theorem of Calculus.*

Put all of this together and one can easily see that Theorem 25.3, being a dimensionless, and coordinate-less statement on a relationship between quantities defined on a region and related quantities restricted to its boundary, is just *The Fundamental Theorem of Vector Calculus.*