MATH 326 Course Notes

August 13, 2013

1 Day 1

1.1 Point Sets

In this class a *point set in n dimensions* will mean a collection of points of real n-dimensional space. That might sound unfamiliar, but we have all seen point sets before! For example the set of *natural numbers*

$$\mathbb{N} = \{0, 1, 2, 3, \dots\},$$

the set of *integers*

$$\mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\},$$

as well as the set of rational numbers $\mathbb{Q}$ and the set of real numbers $\mathbb{R}$ are all examples of point sets in one dimension. Given any real numbers $a, b \in \mathbb{R}$ with $a < b$, we get two more very important example of one-dimensional point sets: the *open interval* (a, b), defined by

$$(a, b) = \{x \in \mathbb{R} : a < x < b\},$$

and the *closed interval* $[a, b]$, defined by

$$[a, b] = \{x \in \mathbb{R} : a \leq x \leq b\}.$$

Two dimensional point sets can be even more interesting! Here we consider subsets of real two-dimensional space $\mathbb{R}^2$. For example, given a real number $r > 0$ and a point $P = (a, b)$ in $\mathbb{R}^2$, the *open ball centered at P of radius r* is the set

$$\begin{aligned} B_r(P) &= \{\text{points in } \mathbb{R}^2 \text{ with distance from } P \text{ less than } r\} \\ &= \{(x, y) \in \mathbb{R}^2 : \sqrt{(x-a)^2 + (y-b)^2} < r\} \\ &= \{(x, y) \in \mathbb{R}^2 : (x-a)^2 + (y-b)^2 < r^2\}. \end{aligned}$$

The open ball will be a very important tool later on when we talk about open and closed sets.

A very striking and dramatic example of a two-dimensional point set is a Julia set. These sets arise naturally in the study of complex dynamics, and can be described as

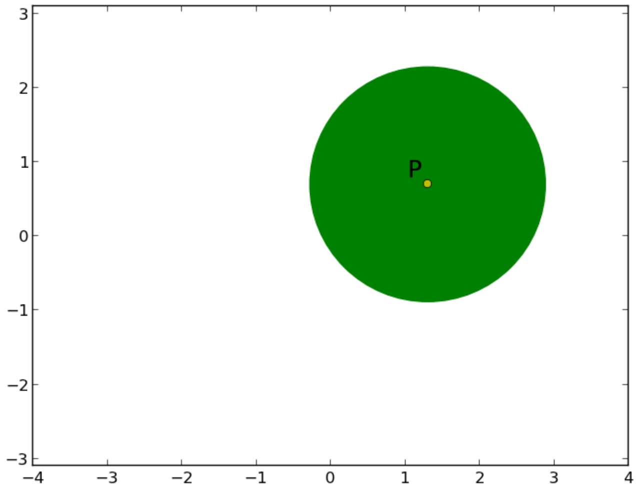


Figure 1: The green region represents an open ball centered about P of radius r, for some point P in $\mathbb{R}^2$ and $r > 0$.

the inverse orbit of a point under a complex rational map. An example of such a set is shown in Figure (2).

Lastly, given two one-dimensional point sets $A, B \subseteq \mathbb{R}$, we can form an example of a two dimensional point set by taking the *cartesian product*

$$A \times B = \{(a, b) : a \in A, b \in B\}.$$

For example, the cartesian product of two open intervals is the open rectangle.

Most interesting of all, and sometimes very hard to visualize, are point sets in higher dimensions. A point P in n-dimensional space $\mathbb{R}^n$ is denoted as an n-tuple of real numbers $P = (a_1, a_2, \ldots a_n)$. The distance between any two points P and $Q = (b_1, b_2, \ldots, b_n)$ in $\mathbb{R}^n$, denoted by $d(P, Q)$ is given by the equation

$$d(P, Q) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \cdots + (a_n - b_n)^2}.$$

Just like in two dimensions, we can define the open ball centered about P of radius r as

$$B_r(P) = \{\text{points } Q \text{ in } \mathbb{R}^n \text{ with } d(P, Q) < r\}.$$

Notice that in three dimensions this is the in region on the inside of a sphere of radius r centered at P, and in one dimension, this is an open interval.

Given two n-dimensional point sets S and T, there are some basic, but very handy ways of constructing new n-dimensional point sets. We can form their *union*

$$S \cup T = \{P : P \in S \text{ or } P \in T\},$$

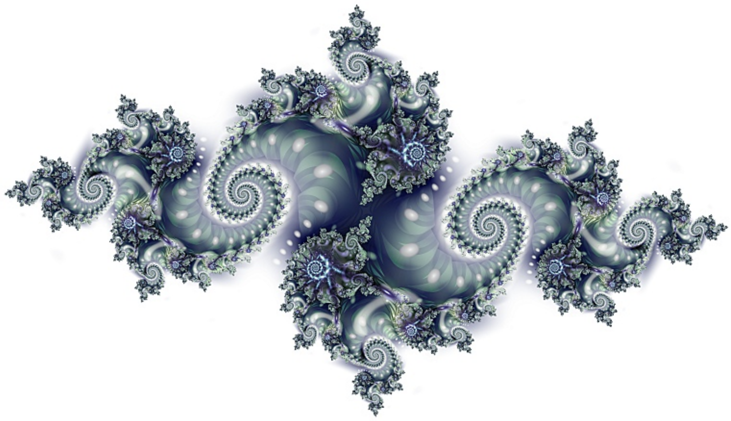


Figure 2: An example of a Julia set.

their *intersection*

$$S \cap T = \{P : P \in S \text{ and } P \in T\},$$

We can also form the complement of S by

$$S' = \{P : P \notin S\}.$$

Exercise 1.1. Draw a sketch of the the two dimensional point set S defined by

$$S = \{(x, y) \in \mathbb{R}^2 : 1 < x^2 + y^2 < 4,\ y > 0\}.$$

Exercise 1.2. Let A and B be the one-dimensional point sets defined by

$$A = (-2, -1) \cup (1, 2)$$

and

$$B = (-2, 2).$$

Draw a sketch of the the two dimensional point set S defined by

$$S = (A \times B) \cup (B \times A).$$

1.2 Interior Points and Open Sets

Next, we define what it means for a point set to be open or to be closed. We start by defining what it means for a point to be interior to a set.

Definition 1.3. Let S be a point set. A point P in S is called an *interior point of* S if there exists a real number $r > 0$ such that the open ball of radius r centered at P is contained in S, ie. $B_r(P) \subseteq S$. The *interior of* S, denoted $\mathrm{int}(S)$, is the subset of S of all interior points of S.

For example, consider the set $S = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 4\}$. The point $(0, 2)$ is not an interior point of S, since for any $r > 0$ the open ball of radius r about $(0, 2)$ will contain some points not in S. However, S has lots of interior points! In fact, we claim that if $P = (x, y)$ is any point in S whose distance from the origin $\mathcal{O} = (0, 0)$ is less than 2, then P is interior to S. In other words, $\text{int}(S) = B_2(\mathcal{O})$. To see this, let $r = 2 - d(\mathcal{O}, P)$. Then $r > 0$, and if we can show that $B_r(P) \subseteq S$, then we will have shown that P is an interior point of S, and justified our claim. To show our desired inclusion, we will use the following very common tactic: we will prove that if Q is a point in $B_r(P)$, then Q is also a point in S. With this in mind, suppose that $Q \in B_r(P)$. Then by definition, we know $d(Q, P) < r$. We now invoke the *triangle inequality*, which says the following: the length of one side of a triangle is no greater than the sum of the lengths of the other two sides. The points $\mathcal{O}, P$, and Q form a triangle, as shown in Figure (3). Applying this to the triangle $\mathcal{O}PQ$, we find

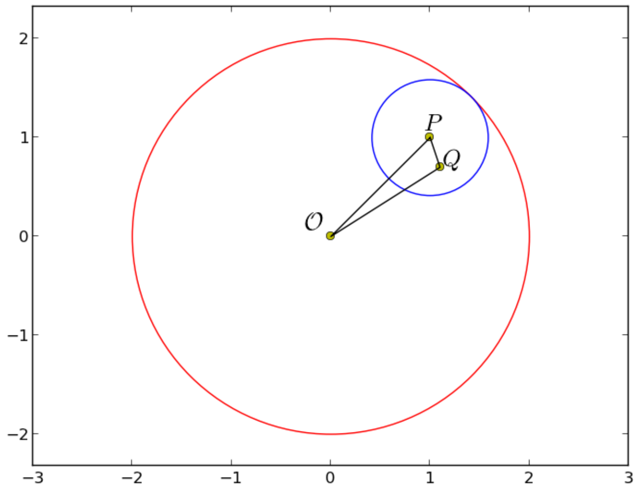


Figure 3: The region within the red circle is $B_2(\mathcal{O})$, and the region within the blue circle is $B_r(P)$. The area within these circles is not colored, in order to aid legibility. Notice that the distances between $\mathcal{O}$, P and Q are the lengths of the three sides of the black triangle in the figure.

$$d(\mathcal{O}, Q) \leq d(\mathcal{O}, P) + d(P, Q).$$

Since $d(P, Q) < r$, this means that

$$d(\mathcal{O}, Q) < d(\mathcal{O}, P) + r = d(\mathcal{O}, P) + 2 - d(\mathcal{O}, P) = 2.$$

Thus $d(\mathcal{O}, Q) < 2$, and this implies that $Q \in S$. We conclude that $B_r(P) \subseteq S$. This verifies our claim.

The argument of the previous paragraph serves two purposes

- it demonstrates the level we should strive for in mathematically justifying a statement in this class
- it illustrates the use of one of the most basic and powerful tools in mathematics; the triangle inequality

Take time to understand the argument of the previous paragraph. Take time to write your own argument along the same lines. Try to identify points where you are tempted to "wave your hands", as opposed to construct a clear and obvious line of reasoning. In this class, one of our goals is to avoid this sort of ambiguous reasoning, so that we can all become math studs. Also take time to think about the triangle inequality: its importance in higher-level math classes, such as real analysis, is hard to overemphasize. As a mathematician I knew once said

> I really only know how to do three things; add, subtract, and use the triangle inequality.

While this quote should be taken with a grain of salt, it stresses the importance and power of the concept of the triangle inequality.

The notion of interior points allows us to define what it means for a set to be open.

Definition 1.4. Let S be a point set. Then the set S is *open* if every point of S is an interior point of S and is *closed* if the complement of S is open. Equivalently, S is open if $\text{int}(S) = S$.

The most basic example of an open set is an open interval. More generally if P is a point in $\mathbb{R}^n$, then the open ball $B_r(P)$ is open. The empty set $\varnothing$ is also technically open, since it doesn't have any points, and therefore trivially satisfies the requirement that all of the points that it has are interior points.

The study of open and closed sets in $\mathbb{R}^n$ is a good sneak peak at the wonderful world of *point set topology.* Moreover the brave, hardy, mathematically inclined soul can study abstract topological spaces wherein the concept of an open set is generalized.

Exercise 1.5. For each of the following two-dimensional point sets S, decide whether or not S is open. If S is open, provide a *mathematical argument.* If S is closed, find a point of S that is not in the interior of S, and provide a *mathematical argument* that it is not in the interior.

(a) $S = \mathbb{R}^2$

(b) $S = \mathbb{Z} \times \mathbb{Z}$

(c) $S = (a, b) \times (c, d)$, where $a, b, c, d \in \mathbb{R}$, with $a < b$ and $c < d$

(d) $S = \{(x, y) \in \mathbb{R}^2 : y = x^2\}$

Exercise 1.6. Show that if A and B are open sets, then $A \cup B$ is open.

Exercise 1.7. Show that if A and B are open sets, then $A \cap B$ is open.

Exercise 1.8. Show that if A is any set, then $\text{int}(A)$ is an open set.

2 Day 2

2.1 Closed Sets and Boundary Points

As a sort of opposite of open sets, we have closed sets, which are defined in the following way

Definition 2.1. A point set S is *closed* if its complement S' is an open set.

Notice that $S'' = S$, and therefore if S is open, then S' must also be closed. For example if $S = B_1(\mathcal{O}) \subseteq \mathbb{R}^2$. Then S is open, and therefore

$$S' = \{(x, y) \in \mathbb{R}^2 : \sqrt{x^2 + y^2} \geq 1\}$$

is a closed set.

Caution 2.2. *Not all sets are open or closed. For example, the* half-open *interval* $[a, b) = \{x \in \mathbb{R} : a \leq x < b\}$ *is a one-dimensional point set which is neither open nor closed.*

Another concept that it is convenient to have is that of a *boundary point*

Definition 2.3. Let S be a point set. A point P (not necessarily in S) is called a *boundary point* of S if for every $r > 0$, the ball $B_r(P)$ contains at least one point of S and at least one point of S'. The collection of all boundary points of S is called the *boundary of* S, and is denoted by $\text{bdry}(S)$.

For example, if $S = B_r(\mathcal{O}) \subseteq \mathbb{R}^2$, then $\text{bdry}(S) = \{(x, y) \in \mathbb{R}^2 : \sqrt{x^2 + y^2} = r^2\}$. If S is closed, then $\text{bdry}(S) \subseteq S$, and if S is open then $S \cap \text{bdry}(S) = \varnothing$. In this class, we will often be interested in n-dimensional regions, which are defined in the following way

Definition 2.4. A *region* is a point set S which is either

- a nonempty open set

or else is

- a nonempty open set together with some or all of its boundary points.

For example, the set

$$S = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 4, y > 0\}$$

is a region, while the set

$$S = \{(x, y) \in \mathbb{R}^2 : y = x^2\}$$

is not a region.

Exercise 2.5. Define a point set S by

$$S = \{(x, y) \in \mathbb{R}^2 : y \geq x^2,\ y \leq 1\}.$$

Make a graph of S and determine if it is open, closed, or neither. What is $\text{bdry}(S)$?

Exercise 2.6. Define a point set S by

$$S = \{(x, y) \in \mathbb{R}^2 : y = \sin(1/x),\ x > 0.\}.$$

See Fig (4). Does S have any interior points? Is S closed? What is bdry(S)?

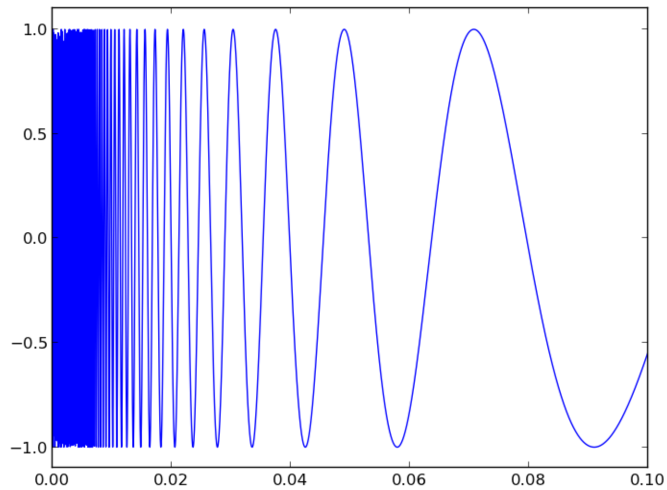


Figure 4: Graph of $y = \sin(1/x)$ for $x > 0$.

2.2 Limits

Now that we have talked about open and closed sets, we have the necessary tools to discuss the concept of a *limit*. We will define limits for functions of two variables, but the reader should note that there is no difference between this and the more general definition of the limit of a function of n variables for any integer $n \geq 1$.

In this section, $f(x, y)$ will always refer to a function of two real variables x, y defined on a region R (see previous section) of $\mathbb{R}^2$, and (x_0, y_0) will always refer to a point either in the interior of R or on the boundary of R. The limit of $f(x, y)$ as the point (x, y) approaches (x_0, y_0) is a real value L satisfying the following property

- As (x, y) gets closer and closer to (x_0, y_0) in distance, the value of $f(x, y)$ gets closer and closer to L in absolute value

Note first of all that the limit as (x, y) appraches (x_0, y_0) does not have to exist! When no such value of L satisfying this property exists, we say that the limit as (x, y) approaches (x_0, y_0) does not exist. When a value L satisfying the above property does exist, we say that the limit as (x, y) approaches (x_0, y_0) exists and is equal to L. Mathematically this is denoted by

$$\lim_{(x,y)\to(x_0,y_0)} f(x, y) = L.$$

Unfortunately the afore mentioned property of limits is not mathematically rigorous. What do we mean by "closer and closer"? The next definition makes our meaning completely clear.

Definition 2.7. We say that *the limit as* (x, y) *approaches* (x_0, y_0) *exists and is equal to* L and write

$$\lim_{(x,y)\to(x_0,y_0)} f(x, y) = L$$

if for any real number $\epsilon > 0$ there exists a real number $\delta > 0$ (which may depend on ϵ) such that $|f(x, y) - L| < \epsilon$ for all points $(x, y) \in R$ with $d((x, y), (x_0, y_0)) < \delta$. We say that *the limit as* (x, y) *approaches* (x_0, y_0) *does not exist* if there is no L satisfying $\lim_{(x,y)\to(x_0,y_0)} f(P) = L$.

In other words, $\lim_{(x,y)\to(x_0,y_0)} f(x, y) = L$ means that for all $\epsilon > 0$ there exists $\delta > 0$ so that f maps the elements of of the open ball $B_\delta((x_0, y_0))$ to within epsilon of L.

To help us understand the above definition, we give the following example

Example 2.8. *Show that*

$$\lim_{(x,y)\to(0,0)} \frac{2x^3 - y^3}{x^2 + y^2} = 0.$$

To show this, we need to come up with a solid *mathematical argument.* An argument like this should always start out the same way

Let $\epsilon > 0$.

What this tells the reader of our argument is that ϵ is some positive quantity. We aren't sure what it is exactly, or how small it might be. But that's okay! If it makes the reader feel more comfortable, the author assures them that there is a wise old man at the top of Mount Rainier that knows exactly what the value of ϵ is in our argument. Unfortunately, there's no phone line up there, so we'll have to ask him about the value later.

Now we need to show that there is a $\delta > 0$ (which can depend on ϵ) so that

$$\left|\frac{2x^3 - y^3}{x^2 + y^2}\right| < \epsilon \quad \text{for all } (x, y) \text{ with } \sqrt{x^2 + y^2} < \delta.$$

Notice that $|x| \leq \sqrt{x^2 + y^2}$ and $|y| < \sqrt{x^2 + y^2}$ so that

$$|2x^3 - y^3| \leq 2|x|^3 + |y|^3 \leq 3(x^2 + y^2)^{3/2},$$

and therefore

$$\left|\frac{2x^3 - y^3}{x^2 + y^2}\right| = \frac{|2x^3 - y^3|}{x^2 + y^2} \leq \frac{3(x^2 + y^2)^{3/2}}{x^2 + y^2} = 3\sqrt{x^2 + y^2}.$$

Thus if we choose $\delta = \epsilon/3$, then for $\sqrt{x^2 + y^2} < \delta$ we have

$$\left|\frac{2x^3 - y^3}{x^2 + y^2}\right| \leq 3\sqrt{x^2 + y^2} < 3\delta = \epsilon.$$

This completes our argument, proving the limit exists and is equal to 0.

Exercise 2.9. Prove that

$$\lim_{(x,y)\to(0,0)} \frac{x^4+y^4}{x^2+y^2} = 0$$

3 Day 3

3.1 Proving Limits Do Not Exist

A not uncommon occurence is that a limit of a function will not exist. To prove that limits do not exist, it is usually useful to use another fact, which allows us to compare limits in two dimensions to their simpler one-dimensional counterparts. Before discussing this fact, we require a technical definition

Definition 3.1. We define a *path* in $\mathbb{R}^n$ to be a function γ whose domain is some open subset U of $\mathbb{R}$ and whose range takes values in $\mathbb{R}^n$. In other words, a path is a function $\gamma : U \to \mathbb{R}^n$ satisfying

$$\gamma(t) = (\gamma_1(t), \gamma_2(t), \ldots, \gamma_n(t)),$$

for some real-valued functions $\gamma_1, \gamma_2, \ldots, \gamma_n$. A path γ is called *continuous* if each of the functions γ_i are continuous for $i = 1, 2, \ldots, n$.

For example $\gamma(t) = (t, t^2)$ is an example of a path in $\mathbb{R}^2$. It is a continuous path, since t and t^2 are continuous functions. With this notion under our belt, we have the following theorem

Theorem 3.2. *If the limit as (x,y) approaches (x_0,y_0) exists and is equal to L, and if γ is a continuous path in $\mathbb{R}^2$ satisfying $\gamma(t_0) = (x_0,y_0)$ for some value t_0, then*

$$\lim_{t\to t_0} f(\gamma(t)) = L.$$

The power of this theorem is that in order to show that a limit does not exist, it suffices to find two different paths approacing the limit point giving us different limit values. This concept is demonstrated in the next example.

Example 3.3. *Show that*

$$\lim_{(x,y)\to(0,0)} \frac{x^2-y^2}{x^2+y^2}$$

does not exist.

Let $f(x,y) = \frac{x^2-y^2}{x^2+y^2}$. To show this we consider two paths:

Path 1: For our first path, we take the straight path along the x-axis

$$\gamma(t) = (t, 0),$$

which is equal to the limit point $(0,0)$ when $t = 0$. Composition of functions tells us that (for $t \neq 0$)

$$f(\gamma(t)) = \frac{t^2-0^2}{t^2+0^2} = \frac{t^2}{t^2} = 1.$$

Therefore

$$\lim_{t \to 0} f(\gamma(t)) = \lim_{t \to 0} 1 = 1.$$

Path 2: For our second path, we consider the straight path along the y-axis

$$\gamma(t) = (0, t).$$

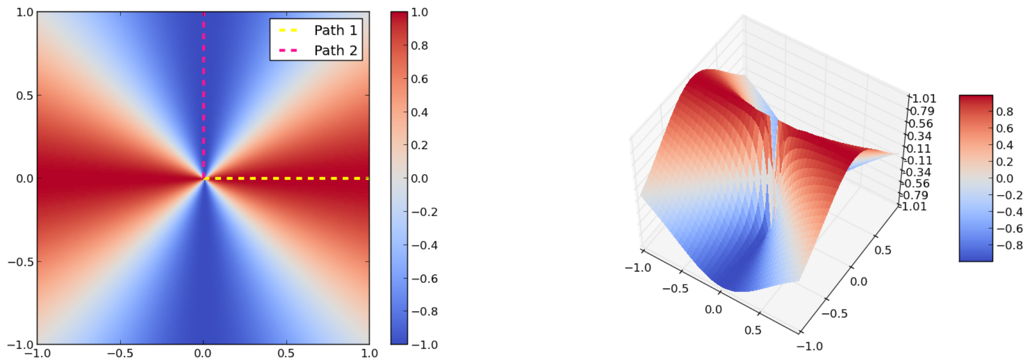
Again $\gamma(t)$ is equal to the limit point $(0, 0)$ when $t = 0$. Composition of functions tells us that (for $t \neq 0$)

$$f(\gamma(t)) = \frac{0^2 - t^2}{0^2 + t^2} = \frac{-t^2}{t^2} = -1.$$

Therefore

$$\lim_{t \to 0} f(\gamma(t)) = \lim_{t \to 0} -1 = -1.$$

Thus as we approach the limit point along the first path, we get a limit of 1, but along the second path, we get a limit of -1. As a consequence of Theorem (3.2), we conclude that the limit does not exist. For a graphical view of this situation, see Figure (5).



(a) A contour plot of the function in Example (3.3). Taking the limit along the two different paths gives two different values

(b) A surface plot of the function in question, showing the pinched-looking point at the origin where the limit of the function does not exist

Figure 5: Behavior near the origin of $f(x, y) = (x^2 - y^2)/(x^2 + y^2)$

A common question that may occur to the reader at this point is the following:

Question 3.4. *What if the limit of a function exists and is equal along every straight path tending to a certain point. Does this mean that the limit exists?*

The answer, as it turns out is not necessarily. Consider the following example which describes one of the coolest rational functions we have seen thus far.

Example 3.5. *Let $f(x, y)$ be the function*

$$f(x, y) = \frac{x^4 y^4}{(x^2 + y^4)^3}.$$

Then for every straight line path γ satisfying $\gamma(0) = (0, 0)$ we have

$$\lim_{t \to 0} f(\gamma(t)) = 0,$$

but the limit

$$\lim_{(x,y) \to (0,0)} f(x, y)$$

does not exist.

To show this, let γ be a straight line path satisfying $\gamma(0) = (0, 0)$. Then $\gamma(t) = (at, bt)$ for some constants a, b not both zero. It follows that

$$f(\gamma(t)) = \frac{(at)^4 (bt)^4}{((at)^2 + (bt)^4)^3} = \frac{a^4 b^4 t^2}{(a^2 + b^4 t^2)^3}.$$

If $a = 0$, then $f(\gamma(t)) = 0$, and $\lim_{t \to 0} f(\gamma(t)) = 0$ trivially. Otherwise $f(\gamma(t))$ is continuous at 0 and therefore

$$\lim_{t \to 0} f(\gamma(t)) = f(\gamma(0)) = \frac{0}{a^2} = 0.$$

Hence in any case, $\lim_{t \to 0} f(\gamma(t)) = 0$.

Now to show that $\lim_{t \to 0} f(\gamma(t))$ does not exist, consider the path

$$\gamma(t) = (t^2, t).$$

Then $\gamma(0) = (0, 0)$ and

$$f(\gamma(t)) = \frac{(t^2)^4 (t)^4}{((t^2)^2 + (t)^4)^3} = \frac{t^{12}}{(2t^4)^3} = \frac{1}{8}.$$

Hence

$$\lim_{t \to 0} f(\gamma(t)) = \lim_{t \to 0} \frac{1}{8} = \frac{1}{8}.$$

This proves that the limit does not exist. To summarize, we have the following word of caution

Caution 3.6. *Sometimes one has to rely on more than straight-line paths to prove that the limit of a function does not exist.*

The actual surface plot of $f(x, y)$ in this case has kind of an interesting shape to it, and is available for the interested reader in Figure (6).

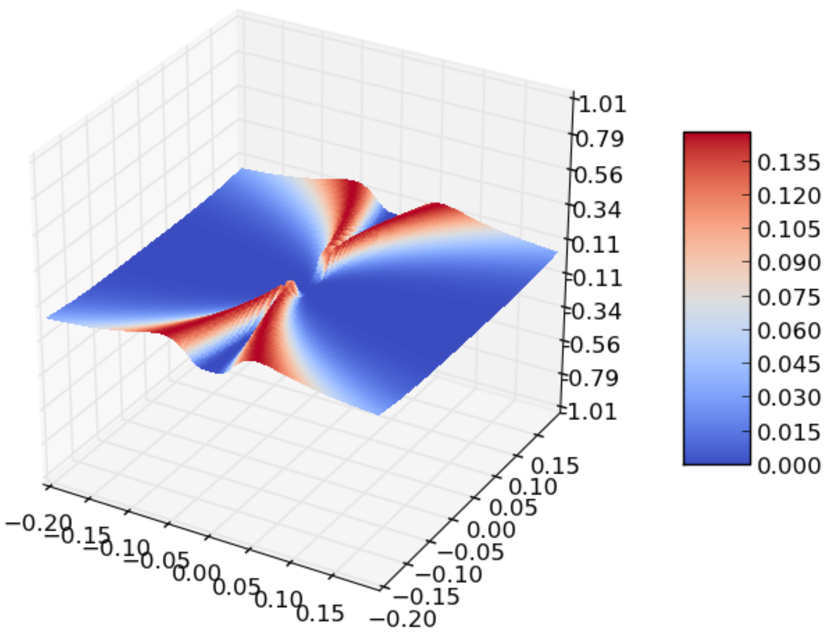


Figure 6: The limit along any straight line to the origin of the function $f(x, y) = \frac{x^4 y^4}{(x^2+y^4)^3}$ is the same. However, the limit along the curved path $x = y^2$ curved path gives a different value.

Exercise 3.7. Prove that

$$\lim_{(x,y)\to(0,0)} \frac{xy^2}{x^2+y^4}$$

does not exist.

Exercise 3.8. Prove that

$$\lim_{(x,y)\to(0,0)} \frac{x^4+3x^2y^2+2xy^3}{(x^2+y^2)^2}$$

does not exist.

4 Day 4

4.1 Real Vector-Valued Functions

In the next few sections, it will be handy to extend the idea of a real-valued function to the more general notion of a *vector-valued function*, ie. a function whose values are actually vectors in $\mathbb{R}^m$. Consider the following example

Example 4.1. *The function*

$$F(x,y) = (x^2, xy, y^2)$$

is a vector-valued function from $\mathbb{R}^2$ to $\mathbb{R}^3$. As input, it takes a point (x,y), and as output it gives us back a point (or vector) in $\mathbb{R}^3$.

The reader may justifiably be annoyed that we are not making as much a distinction as we should about the difference between m-dimensional real vectors and points in $\mathbb{R}^m$. However, we may canonically associate a point P in $\mathbb{R}^m$ with the usual vector corresponding to the line segment $\mathcal{O}P$, so this breach in formality will not cause us any problems.

With the previous example of a vector-valued function in mind, we give the following definition

Definition 4.2. A *vector-valued function* defined on a region R of $\mathbb{R}^n$ is a function F of the form

$$F(x_1,x_2,\ldots,x_n) = (f_1(x_1,x_2,\ldots,x_n), f_2(x_1,x_2,\ldots,x_n), \ldots f_m(x_1,x_2,\ldots,x_n))$$

for some real-valued functions $f_1(x_1,x_2,\ldots,x_n), f_2(x_1,x_2,\ldots,x_n), \ldots f_m(x_1,x_2,\ldots,x_n)$ defined on a region R of $\mathbb{R}^n$. The functions $f_1,\ldots,f_j$ are called the *component functions of* F, and f_j is called the j'th component of F.

Like a plumber with his pipes, we should always be thinking about how various functions can fit together to make a new function. In other words, how can we take the composition of two functions to make a new one. We have the following definition.

Definition 4.3. Suppose that F and G are vector-valued functions with F defined on some region R of $\mathbb{R}^n$ and taking values in $\mathbb{R}^m$; G defined on some region S of $\mathbb{R}^n$ with $f(R) \subseteq S$ and taking values in $\mathbb{R}^\ell$. Furthermore, suppose

$$F(x_1, x_2, \dots, x_n) = (f_1(x_1, x_2, \dots, x_n), f_2(x_1, x_2, \dots, x_n), \dots, f_m(x_1, x_2, \dots, x_n)),$$

and

$$G(y_1, y_2, \dots, y_m) = (g_1(y_1, y_2, \dots, y_m), g_2(y_1, y_2, \dots, x_m), \dots, g_\ell(y_1, y_2, \dots, y_m)),$$

for some real-valued functions f_i and g_j. Then we define the composition of F and G to be the new vector-valued function H defined on the region R of $\mathbb{R}^n$ and taking values in $\mathbb{R}^\ell$ with

$$H = (h_1(x_1, x_2, \dots, x_n), h_2(x_1, x_2, \dots, x_n), \dots, h_\ell(x_1, x_2, \dots, x_n))$$

where for each $j = 1, 2, \dots, \ell$

$$h_j = g_j(f_1(x_1, x_2, \dots, x_n), f_2(x_1, x_2, \dots, x_n), \dots, f_m(x_1, x_2, \dots, x_n)).$$

The composition H is denoted as $G \circ F$.

Composition, in the level of generality presented in the previous definition, are somewhat of a notational nightmare. We hope to dispell some of the incipient confusion with the next example

Example 4.4. *Consider the vector valued functions*

$$F(t) = (\sin(t), \cos(t))$$

and

$$G(x, y) = (x^2, xy, y^2).$$

Find an expression for the composition $G \circ F$.

Since F takes $\mathbb{R}$ to $\mathbb{R}^2$ and G takes $\mathbb{R}^2$ to $\mathbb{R}^3$, we can form their composition $G \circ F$, which will be a function which takes $\mathbb{R}$ to $\mathbb{R}^3$. The way that we accomplish this is by following a generalization of the rule used for calculating the composition of two functions of a single variable: everywhere we see an x in $G(x, y)$ we will replace it with $\sin(t)$, and everywhere we see a y we will replace it with $\cos(t)$. Doing so, we get

$$(G \circ F)(t) = (\sin^2(t), \sin(t)\cos(t), \cos^2(t)).$$

We can also define the limit of a vector-valued function in the following way

Definition 4.5. Suppose that

$$F(x_1, x_2, \dots, x_n) = (f_1(x_1, x_2, \dots, x_n), f_2(x_1, x_2, \dots, x_n), \dots, f_m(x_1, x_2, \dots, x_n)),$$

is a vector-valued function defined on a region R of $\mathbb{R}^n$, and suppose that $(a_1, a_2, \ldots, a_n)$ is a point either in the interior or on the boundary of R. We say *the limit as* $(x_1, \ldots, x_n)$ *approaches* $(a_1, \ldots, a_n)$ *of* $F(x_1, \ldots, x_n)$ *exists and is equal to* $(L_1, L_2, \ldots, L_m)$ and write

$$\lim_{(x_1,\ldots,x_n)\to(a_1,\ldots,a_n)} F(x_1, \ldots, x_n) = (L_1, \ldots, L_m)$$

if and only if for each $j = 1, 2, \ldots, m$ we have

$$\lim_{(x_1,\ldots,x_n)\to(a_1,\ldots,a_n)} f_j(x_1, \ldots, x_n) = L_j.$$

In other words, the limit as $(x_1, \ldots, x_n)$ approaches $(a_1, \ldots, a_n)$ of $F(x_1, \ldots, x_n)$ exists if and only if it exists for each of the component functions f_j. Thus to show a limit does not exist, we need only show that it does not exist for one of the component functions.

Exercise 4.6. Consider the multivariate functions

$$F(x, y) = (x^2, xy)$$

and

$$G(x, y) = (y, -x)$$

Calculate the compositions $F \circ G$ and $G \circ F$.

Exercise 4.7. Consider the multivariate functions

$$F(x, y, z) = (x^2 + y, xyz)$$

and

$$G(x, y) = 3xy$$

Calculate the compositions $G \circ F$.

Exercise 4.8. Let

$$F(x_1, x_2, \ldots, x_n) = (f_1(x_1, x_2, \ldots, x_n), f_2(x_1, x_2, \ldots, x_n), \ldots, f_m(x_1, x_2, \ldots, x_n)),$$

be any vector-valued function, and consider

$$G(y_1, y_2, \ldots, y_m) = y_j.$$

Show that

$$G \circ F(x_1, \ldots, x_n) = f_j(x_1, \ldots, x_n).$$

4.2 Continuity of Vector-Valued Functions

With the concept of a limit firmly established, we are now able to define what it means for a function to be continuous.

Definition 4.9. A vector-valued function $F(x_1, \dots, x_n)$ is said to be continuous at a point $(a_1, \dots, a_n)$ in its domain if

$$\lim_{(x_1,\dots,x_n)\to(a_1,\dots,a_n)} F(x_1, \dots, x_n) = F(a_1, \dots, a_n).$$

In other words, a vector-valued function F is continuous at a point (x_0, y_0) if the limit as $(x_1, \dots, x_n)$ approaches $(a_1, \dots, a_n)$ of $F(x_1, \dots, x_n)$ exists and is equal to the value of the function at that point. A function is said to be *continuous on the set* S if it is continuous at every point of S. When S consists of the entire domain of the function, we will say that the function is continuous.

In actuality, there are few equivalent ways of defining continuity of a function at a point which may be more or less convenient than the one given in the definition under various circumstances.

Theorem 4.10. *Let $F(x_1, \dots, x_n)$ be a vector-valued function, and let $(a_1, \dots, a_n)$ be a point in its domain. Then the following are equivalent*

(a) $F(x_1, \dots, x_n)$ is continuous at $(a_1, \dots, a_n)$

(b) for all $\epsilon > 0$, there exists a $\delta > 0$ such that whenever $(x_1, \dots, x_n)$ is a point in the domain of F with $d((x_1, \dots, x_n), (a_1, \dots, a_n)) < \delta$, we have that

$$d(F(x_1, \dots, x_n), F(a_1, \dots, a_n)) < \epsilon$$

(c) for all $\epsilon > 0$, there exists a $\delta > 0$ such that

$$F(B_\delta((a_1, \dots, a_n))) \subseteq B_\epsilon(F(a_1, \dots, a_n))$$

(d) for all $j = 1, 2, \dots, m$, the component functions $f_j(x_1, \dots, x_n)$ of F are continuous

Proof. Exercise. □

One of the most important properties of continuity is that it is preserved under composition, and therefore composition gives us a powerful way of constructing new continuous functions from old ones. This is explained formally by the following theorem.

Theorem 4.11. *Suppose that F and G are vector-valued functions with F defined on some region R of $\mathbb{R}^n$ and taking values in $\mathbb{R}^m$; G defined on some region S of $\mathbb{R}^n$ with $f(R) \subseteq S$ and taking values in $\mathbb{R}^\ell$. Furthermore, suppose that F is continuous at a point $(a_1, \dots, a_n) \in R$ and G is continuous at $F(a_1, \dots, a_n)$. Then the composition $G \circ F$ is continuous at $(a_1, \dots, a_n)$.*

Proof. Let $\epsilon > 0$ and set $H = F \circ G$. By Theorem (4.10.c), it suffices to show that there exists $\delta > 0$ such that $H(B_\delta((a_1, \ldots, a_n))) \subseteq B_\epsilon(H(a_1, \ldots, a_n))$. Since G is continuous at $F(a_1, \ldots, a_n)$, Theorem (4.10.c) tells us that there exists $\eta > 0$ such that

$$G(B_\eta(F(a_1, \ldots, a_n))) \subseteq B_\epsilon(G(F(a_1, \ldots, a_n))) = B_\epsilon(H(a_1, \ldots, a_n)).$$

Moreover, since F is continuous, there exists $\delta > 0$ such that $F(B_\delta((a_1, \ldots, a_n))) \subseteq B_\eta(F(a_1, \ldots, a_n))$. Therefore

$$H(B_\delta(a_1, \ldots, a_n)) = G(F(B_\delta((a_1, \ldots, a_n)))) \subseteq G(B_\eta(F(a_1, \ldots, a_n))) \subseteq B_\epsilon(H(a_1 \ldots, a_n))$$

we conclude that H is continuous at $(a_1, \ldots, a_n)$. □

As an example of applying this theorem, consider the following.

Example 4.12. *Show that the function $h(x, y) = \sin(1 + xy)^2$ is continuous.*

The function $g(t) = \sin(t)$ is continuous on all of $\mathbb{R}$, and $f(x, y) = (1 + xy)^2$ is continuous on all of $\mathbb{R}^2$ (since it's a polynomial), and therefore by Theorem (4.11) $h(x, y) = g(f(x, y))$ is continuous on all of $\mathbb{R}^2$.

The following simple corollary of Theorem (4.11)

Theorem 4.13. *Suppose that $f(x, y)$ is a function of two variables that is continuous at the point (x_0, y_0), and that $G(x, y, z)$ is a function of three variables that is continuous at the point $(x_0, y_0, f(x_0, y_0))$. Then the function $G(x, y, f(x, y))$ is a function of two variables that is continuous at the point (x_0, y_0).*

Proof. Define a vector valued function $F(x, y) = (x, y, f(x, y))$. Then $F(x, y)$ is continuous at (x_0, y_0) by Theorem (4.10.d) Therefore by Theorem (4.11), $G(x, y, f(x, y)) = G \circ F(x, y)$ is continuous at (x_0, y_0). □

An example application of the previous theorem is the following

Example 4.14. *The function $F(x, y, z) = x^2 + y^2 + z^2$ is continuous on all of $\mathbb{R}^3$ (since it is a polynomial), and $f(x, y) = x/(1 + x^2 + y^2)^{3/2}$ is continuous on all of $\mathbb{R}^2$ since it is a quotient of continuous functions and the denominator is never zero. Therefore by the previous theorem $x^2 + y^2 + \frac{x^2}{(1+x^2+y^2)^3}$ is continuous on all of $\mathbb{R}^2$.*

Exercise 4.15. Recall the continuous function

$$\operatorname{sinc}(t) = \begin{cases} \sin(t)/t & \text{if } t \neq 0 \\ 1 & \text{if } t = 0 \end{cases}$$

Prove that

$$\lim_{(x,y)\to(0,0)} \frac{\sin(xy)}{xy} = 1$$

by relating this limit to the limit of the composition of continuous functions.

Exercise 4.16. Suppose that $f(x, y)$ defined on an open set containing a point (x_0, y_0). Furthermore, suppose that $f(x, y)$ is continuous at (x_0, y_0) and that $f(x_0, y_0) \neq 0$. Prove that there is an open ball containing (x_0, y_0) on which $f(x, y)$ is defined and has the same sign as $f(x_0, y_0)$.

Exercise 4.17. Prove Theorem (4.10).

4.3 Continuous Real-Valued Functions and the Extreme Value Theorem

We now turn our attention to properties of continuous, real-valued functions (as opposed to the more general vector-valued ones). The property of being continuous behaves rather nicely with respect to addition and multiplication. In particular, if $f(x, y)$ and $g(x, y)$ are both real-valued functions which are continuous at a point (x_0, y_0), then we have the following facts

- $f(x, y) + g(x, y)$ is continuous at (x_0, y_0)
- $f(x, y)g(x, y)$ is continuous at (x_0, y_0)
- if $g(x_0, y_0) \neq 0$, then $f(x, y)/g(x, y)$ is continuous at (x_0, y_0)

From these simple observations, it follows that if $f(x, y)$ and $g(x, y)$ are any polynomials, then $f(x, y)/g(x, y)$ is continuous at every point that is not a root of $g(x, y)$. In particular, any polynomial $f(x, y)$ is continuous (see exercises below).

A subject near and dear to any person studying calculus is that of optimization. In particular, when can we find a point in the domain of a function where it achieves its maximum or minimum value. A powerful theorem which provides a partial answer to this question is the so-called Extreme Value Theorem. To state it, we require the definition of a bounded point set.

Definition 4.18. Let S be a point set. Then S is called *bounded* if there exists a real number $R > 0$ such that $S \subseteq B_R(\mathcal{O})$. If S is both closed and bounded, then S is called *compact.*

Using this definition, we have the following theorem.

Theorem 4.19 (Extreme Value Theorem)**.** *Suppose $f(x, y)$ is a function which is continuous on a compact set C. Then there exist points (x_{min}, y_{min}) and (x_{max}, y_{max}) in C such that for any point $(x, y) \in C$ we have that*

$$f(x_{min}, y_{min}) \leq f(x, y) \leq f(x_{max}, y_{max}).$$

To rephrase the statement of the theorem, a continuous function on a compact set attains it largest and smallest values. We will not try to prove this here, since the actual proof is firmly entrenched in the subject of real analysis, and is therefore slightly beyond the scope of this course. Instead, we will demonstrate it's meaning with a couple examples.

Example 4.20. *The set $C = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$ is a compact set. The function $f(x, y) = x^2 - y^2$ is a polynomial, and is therefore continuous. The previous theorem then tells us that $f(x, y)$ attains it's maximal values on C. As a matter of fact, one could take $(x_{min}, y_{min}) = (0, 1)$ and $(x_{max}, y_{max}) = (1, 0)$. (Notice that the points where the minima and maxima are achieved are not unique).*

5 Day 5

5.1 Total Differentials

Just as in the case of real-valued functions of a single variable, differentiability in higher terms is defined in terms of the limit of a difference quotient. However, the higher-dimensional situation is a good deal more fun, due to the existence of various different *types* of derivatives. To give us a taste of what we have in mind, suppose that $f(x, y)$ is a real-valued function of two variables. The main kinds of derivatives we will deal with are the following:

- the *partial derivatives* $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$
- the *total differential* $df(x, y)$
- the *total derivatives* $\frac{df(x,y)}{dx}$ and $\frac{df(x,y)}{dy}$

Partial derivatives are defined in exactly the same way as in earlier calculus classes, and when you differentiate with respect to one variable, you treat the second variable as a constant. In essence, partial derivatives record the rate of change of f with respect to a given variable, assuming that there are no dependencies between the variables. Total derivatives are a somewhat smarter version of this – they take into account the possibility of dependencies between the variables defining a function f. Most exotic of all (compared to the simpler case of functions of a single variable) is the total differential, $df(x, y)$ which is a vector of values, often called the gradient. More generally, we can define the total differential dF of a vector-valued function F, whose values take the form of a matrix, often called the Jacobian.

We begin by defining the total differential of a real, vector-valued function. To simplify the notation of this section, we will always suppose that

$$F(x_1, \ldots, x_n) = (f_1(x_1, \ldots, x_n), f_2(x_1, \ldots, x_n), \ldots, f_m(x_1, \ldots, x_n))$$

is a real vector-valued function defined on a region R of $\mathbb{R}^n$ and that

$$G(y_1, \ldots, y_m) = (g_1(y_1, \ldots, y_m), g_2(y_1, \ldots, y_m), \ldots, g_\ell(y_1, \ldots, y_m))$$

is a real vector-valued function defined on a region S of $\mathbb{R}^m$ with $F(R) \subseteq S$. With this in mind, we establish the following definition

Definition 5.1. The vector-valued function $F(x_1, \ldots, x_n)$ is said to be *differentiable* at a point $(a_1, \ldots, a_m)$ in the region R where it is defined if there exists an $m \times n$ matrix M satisfying

$$\lim_{(h_1,\ldots,h_n)\to(0,\ldots,0)} \frac{F(a_1 + h_1, \ldots, a_n + h_n) - F(a_1, \ldots, a_n) - M \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix}}{\|(h_1, \ldots, h_n)\|} = (0, 0, \ldots, 0).$$

When $F(x_1, \ldots, x_n)$ is differentiable at a point $(a_1, \ldots, a_n)$, we call M the *differential* or *total differential* of F at $(a_1, \ldots, a_n)$ and denote it as $dF(a_1, \ldots, a_n)$.

To help clarify what we mean, we offer the following example.

Example 5.2. *Consider the vector-valued function $F(x, y) = (x^2, xy, y^2)$. Show that F is differentiable at the point $(1, 1)$ and that its total differential is*

$$dF(1,1) = M = \begin{pmatrix} 2 & 0 \\ 1 & 1 \\ 0 & 2 \end{pmatrix}.$$

To show this, we must verify that

$$\lim_{(u,v)\to(0,0)} \frac{F(1+u, 1+v) - F(1,1) - M\binom{u}{v}}{\|(u,v)\|} = (0,0,0).$$

We first calculate

$$M\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 1 & 1 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = (2u, u+v, 2v)$$

and therefore

$$\begin{aligned}
&\lim_{(u,v)\to(0,0)} \frac{F(1+u, 1+v) - F(1,1) - M\binom{u}{v}}{\|(u,v)\|} \\
&= \lim_{(u,v)\to(0,0)} \frac{((1+u)^2, (1+u)(1+v), (1+v)^2) - (1,1,1) - (2u, u+v, 2v)}{\sqrt{u^2+v^2}} \\
&= \lim_{(u,v)\to(0,0)} \frac{(u^2, uv, v^2)}{\sqrt{u^2+v^2}} = \lim_{(u,v)\to(0,0)} \left(\frac{u^2}{\sqrt{u^2+v^2}}, \frac{uv}{\sqrt{u^2+v^2}}, \frac{v^2}{\sqrt{u^2+v^2}} \right).
\end{aligned}$$

Notice that

$$0 \leq \frac{u^2}{\sqrt{u^2+v^2}} \leq \frac{u^2+v^2}{\sqrt{u^2+v^2}} = \sqrt{u^2+v^2} \to 0 \text{ as } (u,v) \to (0,0),$$

so that $\lim_{(u,v)\to(0,0)} \frac{u^2}{\sqrt{u^2+v^2}} = 0$, and similarly $\lim_{(u,v)\to(0,0)} \frac{v^2}{\sqrt{u^2+v^2}} = 0$. Moreover, by the Cauchy-Schwartz inequality $|uv| \leq \frac{1}{2}(u^2+v^2)$ so

$$0 \leq \frac{uv}{\sqrt{u^2+v^2}} \leq \frac{u^2+v^2}{2\sqrt{u^2+v^2}} = \frac{1}{2}\sqrt{u^2+v^2} \to 0 \text{ as } (u,v) \to (0,0),$$

and therefore $\lim_{(u,v)\to(0,0)} \frac{u^2}{\sqrt{u^2+v^2}} = 0$. Thus we have that

$$\lim_{(u,v)\to(0,0)} \left(\frac{u^2}{\sqrt{u^2+v^2}}, \frac{uv}{\sqrt{u^2+v^2}}, \frac{v^2}{\sqrt{u^2+v^2}} \right) = (0,0,0).$$

This verifies our example.

One powerful fact about the total differential is that it satisfies a version of the chain rule.

Theorem 5.3 (Chain Rule)**.** *Suppose F is differentiable at a point $(a_1, \ldots, a_n)$ and G is differentiable at the a point $F(a_1, \ldots, a_n)$. Then the composition $G \circ F$ is differentiable at $(a_1, \ldots, a_n)$ and*

$$d(G \circ F)(a_1, \ldots, a_n) = dG(F(a_1, \ldots, a_n))dF(a_1, \ldots, a_n).$$

We omit the proof for the time being, and instead show its application in the next example.

Example 5.4. *Let $F(x, y) = (x^2, xy, y^2)$ and $G(x, y, z) = (x + 2y + z, y)$. Show that at any point (a, b),*

$$d(G \circ F)(a, b) = \begin{pmatrix} 2(a+b) & 2(a+b) \\ b & a \end{pmatrix}.$$

Proceeding exactly like in the previous example, it is easy enough to show that

$$dF(a, b) = \begin{pmatrix} 2a & 0 \\ b & a \\ 0 & 2b \end{pmatrix},$$

and that

$$dG(c, d) = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 0 \end{pmatrix},$$

and therefore by the chain rule

$$d(G \circ F) = dG(F(a, b))dF(a, b) = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 2a & 0 \\ b & a \\ 0 & 2b \end{pmatrix} = \begin{pmatrix} 2(a+b) & 2(a+b) \\ b & a \end{pmatrix}$$

Exercise 5.5. Suppose that the vector-valued function F is defined by the equation

$$F(x_1, \ldots, x_n) = M \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

for some $m \times n$ matrix M. Show that F is differentiable at every point $(a_1, \ldots, a_n)$ in $\mathbb{R}^n$ and that

$$dF(a_1, \ldots, a_n) = M.$$

5.2 Partial Derivatives

Partial derivatives are something defined for real-valued functions (we will not define them for vector-valued functions) and turn out to be intimately related to the total differential discussed in the previous section. For this section, we suppose that $f(x_1, \ldots, x_n)$ is a real-valued function defined on a region R of $\mathbb{R}^n$. We first recall the definition of the partial derivative of a real-valued function f:

Definition 5.6. Suppose $(a_1, \ldots, a_n)$ is a point in R, and that for some $L > 0$

$$\lim_{h\to 0} \frac{f(a_1, \ldots, a_i - h, \ldots, a_n) - f(a_1, \ldots, a_n)}{h} = L$$

Then we say that the partial derivative of f with respect to x_i exists and is equal to L, and we write

$$\frac{\partial f}{\partial x_i}(a_1, \ldots, a_n) = L.$$

Talk about the geometric significance of the partial derivatives.

5.3 Relationship between Partial Derivatives and the Total Differential

The partial derivatives of the component functions of a vector-valued function, and the total differential of the vector valued function are intimately related. In this section, we will see that differentiability of F implies the existence of all of the first partial derivatives of the component functions of F. A partial converse (no pun intended) of this statement is also true: if the partial derivatives of the component functions exist and are continuous, then F is differentiable. However, we will see examples showing that if we drop the assumption that the partial derivatives are continuous, then the converse does not hold.

To begin, we start with a simple theorem showing us that if F is differentiable, then it is continuous. The proof duplicates the similar proof in the case of functions of a single variable, and follows the intuition that if $F(x_1, \ldots, x_n)$ differentiable at $(a_1, \ldots, a_n)$, then F is linear close to $(a_1, \ldots, a_n)$ with

$$F(a_1 + h_1, \ldots, a_n + h_n) \approx dF(a_1, \ldots, a_n) \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} + F(a_1, \ldots, a_n).$$

Theorem 5.7. *If F is differentiable at $(a_1, \ldots, a_n)$, then F is continuous at $(a_1, \ldots, a_n)$.*

Proof. We wish to show,

$$\lim_{(x_1,\ldots,x_n)\to(a_1,\ldots,a_n)} F(x_1, \ldots, x_n) = F(a_1, \ldots, a_n).$$

Equivalently, we must show that,

$$\lim_{(h_1,\ldots,h_n)\to(0,\ldots,0)} F(a_1 + h_1, \ldots, a_n + h_n) = F(a_1, \ldots, a_n),$$

or in other words

$$\lim_{(h_1,\ldots,h_n)\to(0,\ldots,0)} F(a_1 + h_1, \ldots, a_n + h_n) - F(a_1, \ldots, a_n) = (0, \ldots, 0)$$

Let M be the $m \times n$ matrix satisfying $M = dF(a_1, \ldots, a_n)$ and define

$$E(h_1, \ldots, h_n) = F(h_1 + a_1, \ldots, h_n + a_n) - F(a_1, \ldots, a_n) - M\vec{h}.$$

Then by the definition of the derivative,

$$\lim_{\vec{h}\to(0,\ldots,0)} \frac{E(\vec{h})}{\|\vec{h}\|} = (0,\ldots,0).$$

Consequently,

$$\lim_{\vec{h}\to(0,\ldots,0)} E(\vec{h}) = (0,\ldots,0),$$

and therefore since $M\vec{h}$ is component-wise a polynomial in $h_1,\ldots,h_n$ (and therefore continuous),

$$\begin{aligned}
&\lim_{\vec{h}\to(0,\ldots,0)} F(h_1+a_1,\ldots,h_n+a_n) - F(a_1,\ldots,a_n) \\
&= \lim_{\vec{x}\to(0,\ldots,0)} (E(h_1,\ldots,h_n) + M\vec{x}) \\
&= 0 + M\vec{0} = (0,\ldots,0).
\end{aligned}$$

It follows that

$$\lim_{(x_1,\ldots,x_n)\to(a_1,\ldots,a_n)} F(x_1,\ldots,x_n) = F(a_1,\ldots,a_n).$$

□

As a consequence, we have the following corollary, which is sometimes useful in the proofs of various theorems

Corollary 5.8. *Suppose that F is differentiable at $(a_1,\ldots,a_n)$. Then there exists a vector-valued function $E(h_1,\ldots,h_n)$ from $\mathbb{R}^n$ to $\mathbb{R}^m$ satisfying the following three properties*

- *$E(h_1,\ldots,h_n)$ is continuous at $(0,\ldots,0)$*
- $$\lim_{(h_1,\ldots,h_n)} \frac{E(h_1,\ldots,h_n)}{\sqrt{h_1^2+\cdots+h_n^2}} = (0,\ldots,0)$$
- $$F(x_1,\ldots x_n) = F(a_1,\ldots,a_n) + (dF(a_1,\ldots,a_n))\cdot(\vec{x}-\vec{a}) + E(x_1-a_1,\ldots,x_n-a_n)$$

  for all $(x_1,\ldots,x_n)$ for which $F(x_1,\ldots,x_n)$ is defined.

Proof. It is easy to see that the function

$$E(h_1,\ldots,h_n) = F(h_1+a_1,\ldots,h_n+a_n) - F(a_1,\ldots,a_n) - M\vec{h}$$

satisfies the required properties. □

We next take our first look at the relationship between F and the partial derivatives $\partial f_k/\partial x_j$.

Theorem 5.9. *If F is differentiable at $(a_1, \dots, a_n)$, then $\partial f_k / \partial x_j$ exists at $(a_1, \dots, a_n)$ for all $1 \le j \le n$ and $1 \le k \le m$, and*

$$dF(a_1, \dots, a_n) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix} \tag{1}$$

Proof. Let $M = dF(a_1, \dots, a_n)$ be the $m \times n$ matrix which is the total differential of F at $(a_1, \dots, a_n)$. Define $E : \mathbb{R} \to \mathbb{R}^n$ by

$$E(t) = (a_1, a_2, \dots, a_{j-1}, t, a_{j+1}, \dots, a_n)$$

and also define $G : \mathbb{R}^m \to \mathbb{R}$ by

$$G(y_1, \dots, y_m) = y_k.$$

Then the composition $h(t) = (G \circ F \circ F)(t)$ is the real-valued function of one variable given by

$$h(t) = f_k(a_1, a_2, \dots, a_{j-1}, t, a_{j+1}, \dots, a_n).$$

Notice that $h'(a_j)$, if it exists, is exactly equal to the partial derivative $\frac{\partial f_k}{\partial x_j}$ at the point $(a_1, \dots, a_n)$. We next will prove this exists, and is equal to the k, j'th entry of the matrix M (ie. M_{kj}).

Since the functions E and G are both linear, vector-valued functions, and therefore differentiable everywhere. Moreover

$$dE(t) = \left. \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \right\} \text{1 in } j\text{'th row.}$$

$$dG(y_1, \dots, y_m) = \overbrace{(0, 0, \dots, 0, 1, 0, \dots, 0)}^{k \text{ in } i\text{'th column}}.$$

Since E is differentiable at a_j, F is differentiable at $(a_1, \dots, a_n)$ and G is differentiable at $F(a_1, \dots, a_n)$, the chain rule tells us that $h(t)$ is differentiable at a_j and

$$\begin{aligned} h'(a_j) &= dG(F(a_1, \dots, a_n)) \cdot dF(a_1, \dots, a_n) \cdot dE(a_j) \\ &= (0, 0, \dots, 0, 1, 0, \dots, 0) \cdot M \cdot \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \\ &= M_{kj} \end{aligned}$$

Since this is true for any choice of integers $1 \leq j \leq n$ and $1 \leq k \leq m$, our proof is complete. $\square$

Theorem 5.10. *If the partial derivative $\frac{\partial f_k}{\partial x_j}$ exists and is continuous at $(a_1, \dots, a_n)$ for all $1 \leq j \leq n$ and $1 \leq k \leq m$, then F is differentiable at $(a_1, \dots, a_n)$ and Equation (1) holds.*

Proof. Omitted. $\square$

Exercise 5.11. Use Corollary (5.8) to prove the chain rule for total differentials.

6 Day 6

more on chain rule

7 Day 7

7.1 Implicit Functions

Many times in problems and applications, functions are not provided explicitly, but instead are given in terms of relations. Consider the following example

Example 7.1. *Consider the function $z(x, y)$ defined implicitly by the relation (ie. constraint)*

$$x^2 + y^2 + z^2 = 1.$$

Notice that in this example, $z(x, y)$ will not be defined for all values of x and y, since if $x^2 + y^2$ is too big, then $x^2 + y^2 + z^2$ cannot be equal to 1. Also notice that in this case there's actually more than one function, defined implicitly by this equation. In particular,

$$z = \pm\sqrt{1 - x^2 - y^2}$$

are two different functions satisfying the above relation.

More generally, we can consider a function $z(x, y)$ defined implicitly by the constraint that

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

for some function $g(x, y, z)$. Two questions immediately come to mind

- Given a specific relation, for what values of x and y is my function $z(x, y)$ defined?
- What can we say regarding the uniqueness of a function defined implicitly?
- How do changes to the values of x and y affect the value of z?

If $g(x, y, z)$ is even a little unfriendly, it might in practice be very difficult to actually solve for z as a function of x and y explicitly, so when trying to answer these questions, what methods we use should, from a practical point of view, not depend on our ability to obtain an explicit expression.

The constraint equation can also be considered from a geometric perspective, if we think of it as defining a surface in $\mathbb{R}^3$. In this case, the problem of finding a function $z(x, y)$ is related in an obvious way to finding a parametrization of the surface. In particular, $(x, y) \mapsto (x, y, z(x, y))$ describes a parametrization of the surface described by $g(x, y, z) = 0$. In the case of the example constraint $g(x, y, z) = x^2 + y^2 + z^2 - 1 = 0$, the surface is a sphere. For other choices of g, the surface could look much more fun.

If we carry this idea a bit further, we could also consider functions defined by a *system of constraints*, ie

$$\begin{aligned} g_1(x, y, z) &= 0 \\ g_2(x, y, z) &= 0. \end{aligned}$$

From a geometric point of view, variables x, y, z satisfying these two equations simultaneously lie in the intersection of the surfaces defined by $g_1(x, y, z) = 0$ and $g_2(x, y, z) = 0$. When two surfaces intersect (given some certain assumptions about the niceness of the intersection) we should expect to get a *curve*. Since curves are parametrizable in terms of a single variable, we should then expect that two constraints might allow us to write both y and z as functions of x, ie. $z = z(x)$ and $y = y(x)$. Consider the following example.

Example 7.2. *Let y and z be the functions of x defined implicitly by the system of equations*

$$\begin{aligned} x^2 + y^2 + z^2 &= 1 \\ x + y + z &= 0 \end{aligned}$$

In this case, $x \mapsto (x, y(x), z(x))$ is a parametrization of the circle formed by the intersection of the sphere $x^2 + y^2 + z^2 = 1$ with the plane $x + y + z = 0$.

More generally, we can consider n variables $x_1, \ldots, x_n$ and m constraints

$$\begin{aligned} g_1(x_1, \ldots, x_n) &= 0 \\ g_2(x_1, \ldots, x_n) &= 0 \\ &\vdots \\ g_m(x_1, \ldots, x_n) &= 0 \end{aligned}$$

and ask the question of when some m of the variables can be written in terms of the remaining $n-m$. As it will turn out, whether or not we can do so will depend somewhat sensitively on the region of interest and the m variables we choose to try to express in terms of the remaining $n-m$.

7.2 Partial Derivatives of Implicit Functions

Setting aside for now the problem of whether one may express some variables in terms of the remaining ones, we turn now to focus on how changing the values of some of the constrained variables affect the rest. For example, if z is defined implicitly in terms of x and y via an equation

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

then if we change x to $x + \delta x$ and y to $y + \delta y$, then the constraint forces z to change to a value of $z + \delta z$ so that the constraint remains satisfied

$$g(x + \delta x, y + \delta y, z + \delta z) = 0.$$

Using the language of calculus, the real question we're trying to address here is the following: what are $\partial z/\partial x$ and $\partial z/\partial y$? For implicit functions defined by a single constraint, the answer is given by implicit differentiation, as we found in earlier calculus classes. As a reminder of this process, consider the following example.

Example 7.3. *Suppose z is defined implicitly in terms of x and y by the constraint*

$$x^2 + y^2 + z^2 = 1.$$

Determine $\partial z/\partial x$ and $\partial z/\partial y$.

To do so, we first differentiate the constraint equation with respect to x, treating y as a constant and z as a function of x and y:

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

Solving for $\frac{\partial z}{\partial x}$, we find $\frac{\partial z}{\partial x} = -x/z$. Similar work shows that $\frac{\partial z}{\partial y} = -y/z$.

What about the case when there is more than one constraint equation? For example, what if y and z are defined in terms of x implicitly by some constraints

$$\begin{aligned} g_1(x, y, z) &= 0 \\ g_2(x, y, z) &= 0. \end{aligned}$$

How can we determine dz/dx and dy/dx? The answer, as usual, comes down to the chain rule! Consider the functions $f(x) = (x, y(x), z(x))$ and $G(u, v, w) = (g_1(u, v, w), g_2(u, v, w))$. By the constraint equations,

$$G \circ f(x) = (g_1(x, y, z), g_2(x, y, z)) = (0, 0)$$

for all x on which f is defined, and therefore

$$d(G \circ f)(x) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

However, by the chain rule

$$\begin{aligned} d(G \circ f)(x) &= dG(x, y, z) \cdot df(x) \\ &= \begin{pmatrix} \frac{\partial g_1}{\partial x} & \frac{\partial g_1}{\partial y} & \frac{\partial g_1}{\partial z} \\ \frac{\partial g_2}{\partial x} & \frac{\partial g_2}{\partial y} & \frac{\partial g_2}{\partial z} \end{pmatrix} \begin{pmatrix} 1 \\ \frac{\partial y}{\partial x} \\ \frac{\partial z}{\partial x} \end{pmatrix} \\ &= \begin{pmatrix} \frac{\partial g_1}{\partial x} + \frac{\partial g_1}{\partial y}\frac{\partial y}{\partial x} + \frac{\partial g_1}{\partial z}\frac{\partial z}{\partial x} \\ \frac{\partial g_2}{\partial x} + \frac{\partial g_2}{\partial y}\frac{\partial y}{\partial x} + \frac{\partial g_2}{\partial z}\frac{\partial z}{\partial x} \end{pmatrix}. \end{aligned}$$

Therefore

$$\begin{aligned} \frac{\partial g_1}{\partial x} + \frac{\partial g_1}{\partial y}\frac{\partial y}{\partial x} + \frac{\partial g_1}{\partial z}\frac{\partial z}{\partial x} &= 0 \\ \frac{\partial g_2}{\partial x} + \frac{\partial g_2}{\partial y}\frac{\partial y}{\partial x} + \frac{\partial g_2}{\partial z}\frac{\partial z}{\partial x} &= 0 \end{aligned}$$

This is a system of linear equations, and we can solve it for $\frac{\partial y}{\partial x}$ and $\frac{\partial z}{\partial x}$. To make this more concrete, consider the following example.

Example 7.4. *Suppose that y and z are defined implicitly in terms of x by the following constraints*

$$\begin{aligned} x^2 + y^2 + z^2 &= 1 \\ x + y + z &= 0 \end{aligned}$$

Determine $\frac{dy}{dx}$ and $\frac{dz}{dx}$ at the point $(1/\sqrt{6}, 1/\sqrt{6}, -2/\sqrt{6})$.

To do so, let $g_1(x, y, z) = x^2 + y^2 + z^2 - 1$ and $g_2(x, y, z) = x + y + z$. Then from the work done above, we know that

$$\begin{aligned} 2x + 2y\frac{\partial y}{\partial x} + 2z\frac{\partial z}{\partial x} &= 0 \\ 1 + \frac{\partial y}{\partial x} + \frac{\partial z}{\partial x} &= 0 \end{aligned}$$

Solving for $\frac{\partial y}{\partial x}$ and $\frac{\partial z}{\partial x}$, we get

$$\frac{\partial y}{\partial x} = -\frac{x-z}{y-z}, \qquad \frac{\partial z}{\partial x} = -\frac{x-y}{z-y}$$

Putting in the actual values of x, y, and z we find $\partial y/\partial x = -1$ and $\partial z/\partial x = 0$.

Next time, we will derive a general equation for expressing the partial derivatives of the dependent variables with respect to the independent variables in terms of the partial derivatives of the constraint equations.

8 Day 8

Yesterday, we took a glance at how different variables may have dependencies on one-another, expressed in the form of constraints. Our main observation was that if there are n variables and m constraints, then we should expect to be able to write m of the variables as functions of the remaining $n-m$ variables. Today, we will formalize this notion by proving the implicit function theorem.

The main question that we address today is the following: suppose that we have n variables $y_1, \dots, y_\ell, z_1, \dots, z_m$ (with $\ell + m = n$), and that they satisfying m constraints

$$\begin{aligned} g_1(y_1, \dots, y_\ell, z_1, \dots, z_m) &= 0 \\ g_2(y_1, \dots, y_\ell, z_1, \dots, z_m) &= 0 \\ &\vdots \\ g_m(y_1, \dots, y_\ell, z_1, \dots, z_m) &= 0 \end{aligned} \tag{2}$$

What are sufficient conditions on the functions $g_1, \dots, g_m$ so as to guarantee that we can express $z_1, \dots, z_m$ as functions of $y_1, \dots, y_\ell$? Better yet, suppose that for some fixed point $(a_1, \dots, a_\ell, b_1, \dots, b_m) \in \mathbb{R}^n$, we know that $(y_1, \dots, y_\ell, z_1, \dots, z_m) = (a_1, \dots, a_\ell, b_1, \dots, b_m)$ is a solution to Eq. (2). What are sufficient conditions on the functions $g_1, \dots, g_m$ that guarantee we may write

$$\begin{aligned} z_1 &= z_1(y_1, \dots, y_\ell) \\ z_2 &= z_2(y_1, \dots, y_\ell) \\ &\vdots \\ z_m &= z_m(y_1, \dots, y_\ell) \end{aligned}$$

for some functions $z_j(y_1, \dots, y_\ell)$ satisfying $z_j(a_1, \dots, a_\ell) = b_j$ for all $1 \le j \le m$?

One answer to this question is given by the implicit function theorem. Here we state the implicit function theorem without proof, since the slickest proof of this fact relies on the inverse function theorem, which we will state and prove later on.

Theorem 8.1 (Implicit Function Theorem)**.** *Suppose that* $(a_1, \dots, a_\ell, b_1, \dots, b_m)$ *is a fixed point in* $\mathbb{R}^n$ *and that* $(y_1, \dots, y_\ell, z_1, \dots, z_m) = (a_1, \dots, a_\ell, b_1, \dots, b_m)$ *is a solution* *to Eq. (2). Additionally, suppose each of the g_i are continuously differentiable in a neighborhood of $(a_1, \ldots, a_\ell, b_1, \ldots, b_m)$ and the $m \times m$ matrix*

$$\begin{pmatrix} \frac{\partial g_1}{\partial z_1} & \frac{\partial g_1}{\partial z_2} & \cdots & \frac{\partial g_1}{\partial z_m} \\ \frac{\partial g_2}{\partial z_1} & \frac{\partial g_2}{\partial z_2} & \cdots & \frac{\partial g_2}{\partial z_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial z_1} & \frac{\partial g_m}{\partial z_2} & \cdots & \frac{\partial g_m}{\partial z_m} \end{pmatrix}$$

is invertible at $(a_1, \ldots, a_\ell, b_1, \ldots, b_m)$. Then in a small enough open ball centered at $(a_1, \ldots, a_\ell)$, for all $1 \le j \le m$ there exist continuously differentiable functions $z_j = z_j(y_1, \ldots, y_\ell)$ satisfying $z_j(a_1, \ldots, a_\ell) = b_j$.

Proof. Postponed. □

One may notice immediately that the implicit function theorem does not tell us what the functions $z_j(y_1, \ldots, y_\ell)$ are. However, the chain rule may be used to determine what the partial derivatives must be in terms of the partial derivatives of the g_i's.

Corollary 8.2. *If $(a_1, \ldots, a_\ell, b_1, \ldots, b_m)$ and $g_1, \ldots, g_m$ satisfy the conditions of the implicit function theorem, then:*

$$\begin{pmatrix} \frac{\partial z_1}{\partial y_1} & \frac{\partial z_1}{\partial y_2} & \cdots & \frac{\partial z_1}{\partial y_\ell} \\ \frac{\partial z_2}{\partial y_1} & \frac{\partial z_2}{\partial y_2} & \cdots & \frac{\partial z_2}{\partial y_\ell} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial z_m}{\partial y_1} & \frac{\partial z_m}{\partial y_2} & \cdots & \frac{\partial z_m}{\partial y_\ell} \end{pmatrix} = - \begin{pmatrix} \frac{\partial g_1}{\partial z_1} & \frac{\partial g_1}{\partial z_2} & \cdots & \frac{\partial g_1}{\partial z_m} \\ \frac{\partial g_2}{\partial z_1} & \frac{\partial g_2}{\partial z_2} & \cdots & \frac{\partial g_2}{\partial z_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial z_1} & \frac{\partial g_m}{\partial z_2} & \cdots & \frac{\partial g_m}{\partial z_m} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial g_1}{\partial y_1} & \frac{\partial g_1}{\partial y_2} & \cdots & \frac{\partial g_1}{\partial y_\ell} \\ \frac{\partial g_2}{\partial y_1} & \frac{\partial g_2}{\partial y_2} & \cdots & \frac{\partial g_2}{\partial y_\ell} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial y_1} & \frac{\partial g_m}{\partial y_2} & \cdots & \frac{\partial g_m}{\partial y_\ell} \end{pmatrix} \tag{3}$$

Proof. Define functions

$$F(y_1, \ldots, y_\ell) = (y_1, \ldots, y_\ell, z_1(y_1, \ldots, y_\ell), \ldots, z_m(y_1, \ldots, y_\ell))$$

and

$$G(x_1, \ldots, x_n) = (g_1(x_1, \ldots, x_n), \ldots, g_m(x_1, \ldots, x_n)).$$

Then

$$(G \circ F)(y_1, \ldots, y_\ell) = (g_1(y_1, \ldots, y_\ell, z_1, \ldots, z_m), \ldots, g_m(y_1, \ldots, y_\ell, z_1, \ldots, z_m)) = (0, \ldots, 0),$$

for all $y_1, \ldots, y_\ell$ for which F is defined. Therefore

$$d(G \circ F)(y_1, \ldots, y_\ell) = \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix}.$$

However, by the chain rule

$$d(G \circ F)(y_1, \ldots, y_\ell) = dG(y_1, \ldots, y_\ell, z_1 \ldots, z_\ell) \cdot dF(y_1, \ldots, y_\ell)$$

$$= \begin{pmatrix} \frac{\partial g_1}{\partial y_1} & \cdots & \frac{\partial g_1}{\partial y_\ell} & \frac{\partial g_1}{\partial z_1} & \cdots & \frac{\partial g_1}{\partial z_\ell} \\ \frac{\partial g_2}{\partial y_1} & \cdots & \frac{\partial g_2}{\partial y_\ell} & \frac{\partial g_2}{\partial z_1} & \cdots & \frac{\partial g_2}{\partial z_\ell} \\ \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial y_1} & \cdots & \frac{\partial g_m}{\partial y_\ell} & \frac{\partial g_m}{\partial z_1} & \cdots & \frac{\partial g_m}{\partial z_\ell} \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \\ \frac{\partial z_1}{\partial y_1} & \frac{\partial z_1}{\partial y_2} & \cdots & \frac{\partial z_1}{\partial y_\ell} \\ \frac{\partial z_2}{\partial y_1} & \frac{\partial z_1}{\partial y_2} & \cdots & \frac{\partial z_1}{\partial y_\ell} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial z_m}{\partial y_1} & \frac{\partial z_m}{\partial y_2} & \cdots & \frac{\partial z_m}{\partial y_\ell} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{\partial g_1}{\partial y_1} & \cdots & \frac{\partial g_1}{\partial y_\ell} \\ \frac{\partial g_2}{\partial y_1} & \cdots & \frac{\partial g_2}{\partial y_\ell} \\ \vdots & \cdots & \vdots \\ \frac{\partial g_m}{\partial y_1} & \cdots & \frac{\partial g_m}{\partial y_\ell} \end{pmatrix} + \begin{pmatrix} \frac{\partial g_1}{\partial z_1} & \cdots & \frac{\partial g_1}{\partial z_\ell} \\ \frac{\partial g_2}{\partial z_1} & \cdots & \frac{\partial g_2}{\partial z_\ell} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial z_1} & \cdots & \frac{\partial g_m}{\partial z_\ell} \end{pmatrix} \cdot \begin{pmatrix} \frac{\partial z_1}{\partial y_1} & \frac{\partial z_1}{\partial y_2} & \cdots & \frac{\partial z_1}{\partial y_\ell} \\ \frac{\partial z_2}{\partial y_1} & \frac{\partial z_2}{\partial y_2} & \cdots & \frac{\partial z_2}{\partial y_\ell} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial z_m}{\partial y_1} & \frac{\partial z_m}{\partial y_2} & \cdots & \frac{\partial z_m}{\partial y_\ell} \end{pmatrix}$$

Thus we have that

$$\begin{pmatrix} \frac{\partial g_1}{\partial y_1} & \cdots & \frac{\partial g_1}{\partial y_\ell} \\ \frac{\partial g_2}{\partial y_1} & \cdots & \frac{\partial g_2}{\partial y_\ell} \\ \vdots & \cdots & \vdots \\ \frac{\partial g_m}{\partial y_1} & \cdots & \frac{\partial g_m}{\partial y_\ell} \end{pmatrix} + \begin{pmatrix} \frac{\partial g_1}{\partial z_1} & \cdots & \frac{\partial g_1}{\partial z_\ell} \\ \frac{\partial g_2}{\partial z_1} & \cdots & \frac{\partial g_2}{\partial z_\ell} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial z_1} & \cdots & \frac{\partial g_m}{\partial z_\ell} \end{pmatrix} \cdot \begin{pmatrix} \frac{\partial z_1}{\partial y_1} & \frac{\partial z_1}{\partial y_2} & \cdots & \frac{\partial z_1}{\partial y_\ell} \\ \frac{\partial z_2}{\partial y_1} & \frac{\partial z_2}{\partial y_2} & \cdots & \frac{\partial z_2}{\partial y_\ell} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial z_m}{\partial y_1} & \frac{\partial z_m}{\partial y_2} & \cdots & \frac{\partial z_m}{\partial y_\ell} \end{pmatrix} = \begin{pmatrix} 0 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & 0 \end{pmatrix}$$

from which the statement of the corollary follows immediately. □

It may take some time to appreciate what we have here, but the idea involved is very zen:

> One of the very best ways to make progress on a math problem is to reduce that problem to linear algebra.

Our method in the previous corollary is exactly that: we reduced the problem of finding an expression for the various antiderivatives to a problem of determining the inverse of a certain matrix.

9 Day 9

Last time, we derived an equation for finding the partial derivative of functions defined implicitly. Today, we'll look at some examples of how that equation may be applied. In this section, the following lemma will be handy

Lemma 9.1. *The inverse of a 2×2 matrix (if it exists) is given by*

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

Proof. Exercise. $\square$

Example 9.2. *Suppose u and v are defined implicitly as functions of x and y by*

$$\begin{aligned} u\cos(v) - x &= 0 \\ u\sin(v) - y &= 0 \end{aligned}$$

Determine $\partial u/\partial x, \partial u/\partial y, \partial v/\partial x,$ and $\partial v/\partial y$.

To complete this problem, take $y_1 = x$,$y_2 = y$, $z_1 = u$, $z_2 = v$, as well as

$$\begin{aligned} g_1(y_1, y_2, z_1, z_2) &= z_1 \cos(z_2) - x_1 \\ g_2(y_1, y_2, z_1, z_2) &= z_1 \sin(z_2) - x_2 \end{aligned}$$

in Equation (3). Then we get

$$\begin{aligned} \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix} &= - \begin{pmatrix} \cos(v) & -u\sin(v) \\ \sin(v) & u\cos(v) \end{pmatrix}^{-1} \cdot \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \\ &= \begin{pmatrix} \cos(v) & \sin(v) \\ -\sin(v)/u & \cos(v)/u \end{pmatrix} \end{aligned}$$

This expression immediately gives us the values of all the partial derivatives.

10 Day 10

Optimization of functions with constraints

10.1 Review of Optimizing a Function of Two Variables

The student is probably already acquainted with the problem of minimizing or maximizing a function of two variables $f(x, y)$ on a compact region R of $\mathbb{R}^2$. To do so, the general process follows three simple steps:

(i) Search for *critical points* of f in the interior of R.

(ii) Parametrize the boundary of R and find critical points of $f(x, y)$ on the boundary subject to this parametrization.

(iii) Compare the value of f on the boundary critical points to the values on the critical points in the interior; choose the biggest and the smallest.

There are a couple technical points that are worth mentioning

- For many regions, the boundary may have to be parametrized in several segments. The points of the boundary where these different parametrizations meet are sometimes called "corners". If the boundary has any corners, then these are automatically considered critical points of the parametrization of the boundary. This is a technical but extremely important observation! What we consider a corner depends on the parametrization we choose. For example, if our region is $R = \{(x, y) : x^2 + y^2 \leq 1\}$, then two possible parametrizations of bdry$R$ are

$$\begin{cases} x(t) & = t \\ y(t) & = \pm\sqrt{1 - t^2} \end{cases}, \quad 0 \leq t \leq 1$$

  and

$$\begin{cases} x(t) & = \cos(t) \\ y(t) & = \sin(t) \end{cases}, \quad 0 \leq t \leq 2\pi.$$

  With respect to the first parametrization, there are two corners, located at $(\pm 1, 0)$. With respect to the second parametrization, there are no corners.

- The reason this process is successful is due to the extreme value principal: a continuous function on a compact domain will achieve its extreme values at some point in the domain. The functions that we can apply this process to are at least differentiable on R (otherwise critical points make little sense), and therefore continuous on R.

Example 10.1. *Consider the function*

$$f(x, y) = x^2 + y^2.$$

Find the maximum and minimum value of $f(x, y)$ on the compact region R defined by

$$R = [-1, 1] \times [-1, 1].$$

The region R is compact because it is closed (being the cartesian product of two closed sets), and bounded (being contained in the ball $B_2(\mathcal{O})$. In order to optimize $f(x, y)$ on R, we follow the steps outlined above

(i) Since $df(x, y) = \begin{pmatrix} 2x & 2y \end{pmatrix}$, f has a single critical point at $(0, 0)$.

(ii) We parametrize the boundary of R as four line segments L_1, L_2, L_3, L_4 progessing counter-clockwise around the square formed by bdry$R$, starting at the point $(-1, -1)$. The corresponding parametrizations are

$$L_1 : \begin{cases} x(t) & = 2t - 1 \\ y(t) & = -1 \end{cases}, \quad 0 \leq t \leq 1$$

$$L_2 : \begin{cases} x(t) & = 1 \\ y(t) & = 2t - 1 \end{cases}, \quad 0 \leq t \leq 1$$

$$L_3 : \begin{cases} x(t) &= 1 - 2t \\ y(t) &= 1 \end{cases}$$

$$L_4 : \begin{cases} x(t) &= -1 \\ y(t) &= 1 - 2t \end{cases}$$

On L_1, $f(x, y)$ takes the value

$$f(t) = f(x(t), y(t)) = (2t - 1)^2 + (-1)^2 = 4t^2 - 4t + 2,$$

which has a single critical point where $f'(t) = 0$ at $t = 1/2$. The corresponding point on the boundary is $(0, -1)$. Similarly, on the remaining boundaries, we find critical points at $(1, 0), (0, 1)$, and $(-1, 0)$. Additionally, the chosen parametrization has "corners" at $(-1, -1), (-1, 1), (1, 1), (1, -1)$.

(iii) Finally, we compare the values of $f(x, y)$ at the various critical points we found

(x,y)	f(x,y)
(0,0)	0
(0,-1)	1
(1,0)	1
(0,1)	1
(-1,0)	1
(-1,-1)	2
(1,-1)	2
(1,1)	2
(-1,1)	2

We conclude that the minimum of $f(x, y)$ on R is 0 and occurs at $(0, 0)$, while the maximum is 2 and occurs at the four corners of bdry(R). (This calculation may also be verified from the fact that $f(x, y)$ is just the square of the distance from the origin).

10.2 Optimizing Functions of Several Variables

This idea extends quite naturally to optimizing a function of several variables $f(x_1, \ldots, x_n)$ on the interior of a a compact region R of $\mathbb{R}^n$. To do so, we can follow similar steps as in the two-variable case, searching first for critical points in the interior of R, then for critical points on the boundary of R, and concluding with a comparison of some finite collection of values. However, this immediately poses some challenges

- Boundaries of compact regions in higher dimensions can be computationally much more involved than their counterparts in $\mathbb{R}^2$. For example, think of the region $R = [-1, 1] \times [-1, 1] \times [-1, 1]$. In order to maximize a function on R, we'd have to search for critical points in the interior, then on the 6 sides, then on the 12 edges, and also throw in the 8 vertices. Whew!

- To find critical points on the boundary in our old method, we had to parametrize the boundary. However, parametrizing surfaces can be even harder than parametrizing curves in the (x, y)-plane.

In practice, finding critical points inside the regions is not too tough; just look for points where all the partial derivatives are simultaneously zero. It's dealing with the boundary that can make things so much more complicated, with trying to find appropriate parametrizations of the various segments of the boundary. Therefore we turn our attention to developing more advanced methods of optimizing functions on boundaries, ie. finding maximum and minimum values of a function where the input values are constrained to live on the boundary of some compact region in $\mathbb{R}^n$. We will call this problem *optimizing a function with constraints*.

10.3 Optimizing a Function with Constraints

In this section, we will outline various ways in which we can optimize a function $f(x_1, \ldots, x_n)$ subject to some constraints. We will restrict our attention to situations in which the constraints on the values of $x_1, \ldots, x_n$ may be expressed in terms of a single equation

$$g(x_1, \ldots, x_n) = 0.$$

where g is assumed to be differentiable at every point of $\mathbb{R}^n$ with continuous partial derivatives and no critical points occuring at points where $g = 0$. We view this geometrically as optimizing the function on the hypersurface described by the equation $g = 0$. Note that the methods presented have obvious generalizations to systems of constraint equations. We will present three main methods of dealing with situation

Method I: Parametrization

Method II: Implicit Differentiation

Method III: Lagrange Multipliers

The first method is what we have already discussed: one finds a parametrization of all the values satisfying the constraint equation, and then maximizes the function with respect the the new parameters. This method is somewhat impractical in general, since it is often hard to find an explicit expression for such a parametrization, given various constraint equations.

The second method is based on the observation that under certain conditions, the constraint equation allows us to treat one of the variables, say x_n as a function of the other $n - m$ variables, without actually writing down explicit equations for them. Then by the magic of implicit differentiation, one can find critical points and use them to optimize f subject to the specified constraints.

The third method is the method of Lagrange Multipliers, and is based on a certain geometric observation of the situation we are in. The rest of this section will be dedicated to a discussion of Method II; we will discuss the method of Lagrange multipliers in more detail next time.

By an argument that will be made rigorous later when we discuss the implicit function theorem, as long as the partial derivatives of the function g is sufficiently nice and $\partial g/\partial x_n \neq 0$, the constraint equations allow us to express x_n in terms of $x_1, \ldots, x_{n-1}$. For convenience, we label the relabel $x_1, \ldots, x_{n-1}$ as $y_1, \ldots, y_{n-1}$ and x_n as z, with z being a function of the y_j. Under this relabelling, we are trying to minimize $f(y_1, \ldots, y_{n-1}, z)$ subject to the the constraint equations

$$g(\vec{y}, z) = 0.$$

Since the z is really now just functions of $y_1, \ldots, y_{n-1}$, we see that f is really just a function of $y_1, \ldots, y_{n-1}$ also. The chain rule then gives

$$\frac{\partial f}{\partial y_j} = \frac{\partial f}{\partial x_j} + \frac{\partial f}{\partial x_n}\frac{\partial z}{\partial y_j}.$$

where by Equation (3)

$$\frac{\partial z}{\partial y_j} = -\frac{\partial g}{\partial x_j}/\frac{\partial g}{\partial x_n}.$$

Substituting this into the previous equation and multiplying by $\partial g/\partial x_n$, we see that the maxima and minima of f subject to the constraint $g = 0$ occur at points $(x_1, \ldots, x_n)$ on the surface $g(x_1, \ldots, x_n) = 0$ satisfying the equation

$$\frac{\partial g}{\partial x_n}\frac{\partial f}{\partial x_j} - \frac{\partial f}{\partial x_n}\frac{\partial g}{\partial x_j} = 0, \quad 1 \leq j \leq n \tag{4}$$

Notice, that the above equation only applies where $\partial g/\partial x_n \neq 0$. If $\partial g/\partial x_n$ vanishes at some point, then we must instead use one of the other variables in place of x_n. Thus in general we have the following theorem

Theorem 10.2. *If g is differentiable everywhere with continuous partial derivatives and the critical points of g occur away from the surface $g(x_1, \ldots, x_n) = 0$, then a local/global minimum/maximum of $f(x_1, \ldots, x_n)$ subject to the constraint $g(x_1, \ldots, x_n) = 0$ will occur at a point $(x_1, \ldots, x_n)$ satisfying*

$$\frac{\partial g}{\partial x_k}\frac{\partial f}{\partial x_j} - \frac{\partial f}{\partial x_k}\frac{\partial g}{\partial x_j} = 0, \quad 1 \leq j \leq n, \tag{5}$$

for all $1 \leq k \leq n$ for which $\partial g/\partial x_k \neq 0$.

Example 10.3. *Find the maximum and minimum value of $f(x, y, z) = x^3 + y^3 + z^3$ on the surface of the sphere $x^2 + y^2 + z^2 = 1$.*

Notice that this is maximizing $f(x, y, z)$ subject to the constraint $g(x, y, z) = 0$ for $g(x, y, z) = x^2 + y^2 + z^2 - 1$. Assuming that the maximum occurs at a point where x, y, z are all nonzero, Equation (5) with $x_k = z$, says it will satisfy

$$(2z)(3x^2) - (3z^2)(2x) = 0$$
$$(2z)(3y^2) - (3z^2)(2y) = 0.$$

Only two such points also satisfy the constraint $x^2+y^2+z^2=1$, namely $(-1/\sqrt{3},-1/\sqrt{3},-1/\sqrt{3})$ and $(1/\sqrt{3},1/\sqrt{3},1/\sqrt{3})$, on which f takes the values $-3^{-1/2}$ and $3^{-1/2}$, respectively. If $z=0$, then Equation (5) with $x_k=y$ says that

$$(2y)(3x^2)-(3y^2)(2x)=0$$

and therefore $y=x=\pm 1/\sqrt{2}$, at which the value of f is $\pm 1/\sqrt{2}$, which by comparison to the two values already found is not the local max/min. Similarly, if $y=0$ or $x=0$, we do not get the max/min of f on the sphere. Thus the absolute max/min of f on the sphere are $1/\sqrt{3}$ and $-1/\sqrt{3}$, respectively.

11 Day 11

Last time, we obtained an equation for finding critical points of a function subject to a constraint. This time, we develop a geometric picture of this, and use it to obtain a third method of optimization: the method of Lagrange multipliers. To get this geometric picture, we require the notion of a level surface of a function.

Definition 11.1. Given a function, a *level surface* of the function $f(x_1,\dots,x_n)$ is the collection of points where the value of f is constant. It is given by the equation

$$f(x_1,\dots,x_n)=k,$$

for some constant k.

We have the following theorem

Theorem 11.2. *If f and g are differentiable everywhere with continuous partial derivatives, and the critical points of g occur away from the surface $g=0$, then a local maximum/minimum value k of $f(x_1,\dots,x_n)$ subject to the constraint $g(x_1,\dots,x_n)=0$ occurs at a point where the level surface $f(x_1,\dots,x_n)=k$ is tangent to the surface $g(x_1,dots,x_n)=0$.*

Proof. Suppose that $(a_1,\dots,a_n)$ is a local minimum/maximum of $f(x_1,\dots,x_n)$ subject to the constraint $g(x_1,\dots,x_n)=0$, and let $k=f(a_1,\dots,a_n)$. Then the normal vector of the surface $f(x_1,\dots,x_n)=k$ at $(a_1,\dots,a_n)$ is

$$\left\langle \frac{\partial f}{\partial x_1},\dots,\frac{\partial f}{\partial x_n}\right\rangle$$

and the normal vector of the surface $g(x_1,\dots,x_n)=0$ at $(a_1,\dots,a_n)$ is

$$\left\langle \frac{\partial g}{\partial x_1},\dots,\frac{\partial g}{\partial x_n}\right\rangle,$$

with all derivatives evaluated at the point $(a_1, \ldots, a_n)$. Since the critical points of g occur away from where $g = 0$, there exists a k such that $\partial g/\partial x_k \neq 0$ at $(a_1, \ldots, a_n)$. Therefore by Equation (5), we have that

$$\begin{aligned}
\left(\frac{\partial g}{\partial x_k}\right)\left\langle \frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n} \right\rangle \\
&= \left\langle \frac{\partial g}{\partial x_k}\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial g}{\partial x_k}\frac{\partial f}{\partial x_n} \right\rangle \\
&= \left\langle \frac{\partial f}{\partial x_k}\frac{\partial g}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_k}\frac{\partial g}{\partial x_n} \right\rangle \\
&= \left(\frac{\partial f}{\partial x_k}\right)\left\langle \frac{\partial g}{\partial x_1}, \ldots, \frac{\partial g}{\partial x_n} \right\rangle
\end{aligned}$$

Since the two surfaces intersect at this point and their normal vectors are scalar multiples of one-another there, we conclude that the surfaces are tangent at $(a_1, \ldots, a_n) = 0$. □

The previous theorem often immediately gives us a simple geometric method of determining the location of an absolute max/min of a function, as the next example demonstrates.

Example 11.3. *Find the maximum value of the function $f(x, y, z) = x + y + z$ on the surface of the sphere $x^2 + y^2 + z^2 = 1$.*

By the previous theorem, it suffices to find a value k and point (x, y, z) such that the surface $f(x, y, z) = k$ is tangent to the sphere described by $g(x, y, z) = x^2 + y^2 + z^2 - 1 = 0$. The surface $f(x, y, z) = k$ is a plane with normal vector $\langle 1, 1, 1\rangle$. Since the normal vector of the sphere points in the radial direction at any point on the sphere, it follows that the max/min of f must occur where $x = y = z$; ie. at $(1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})$ or at $(-1/\sqrt{3}, -1/\sqrt{3}, -1/\sqrt{3})$. Hence the absolute max/min are $3/\sqrt{3}$ and $-3/\sqrt{3}$, respectively.

This idea is formalized in the method of Lagrange multipliers. The method is simple to describe. Suppose we are maximizing f subject to the constraint $g = 0$. For Lagranges method, we form the following auxilary function

$$u(x_1, \ldots, x_n, \lambda) = f(x_1, \ldots, x_n) + \lambda g(x_1, \ldots, x_n).$$

Theorem 11.4. *The critical points of f subject to $g = 0$ occur at the critical points of*

u, ie. points $(x_1, \dots, x_n, \lambda)$ satisfying the $n+1$ constraints.

$$\begin{aligned}
\frac{\partial u}{\partial x_1} &= \frac{\partial f}{\partial x_1} + \lambda \frac{\partial g}{\partial x_1} = 0 \\
\frac{\partial u}{\partial x_2} &= \frac{\partial f}{\partial x_2} + \lambda \frac{\partial g}{\partial x_2} = 0 \\
&\vdots \\
\frac{\partial u}{\partial x_n} &= \frac{\partial f}{\partial x_n} + \lambda \frac{\partial g}{\partial x_n} = 0 \\
\frac{\partial u}{\partial \lambda} &= g = 0
\end{aligned}$$

Proof. The critical points of u are exactly those points where the level surface $f = k$ and the surface $g = 0$ are tangent, and therefore this is a restatement of the previous theorem. □

The scalar λ in the above is called the *Lagrange multiplier*. The power of the method of Lagrange multipliers lies in its simplicity; after forming the function u, we can forget about any dependencies between variables that are caused by the constraints, and simply find critical points of a function of several variables. The method of solution also has the wonderful bonus of preserving the natural symmetry in the form of the solutions.

Example 11.5. *Find the dimensions of the box of largest volume which can be fitted inside the ellipsoid*

$$x^2/a^2 + y^2/b^2 + z^2/c^2 = 1$$

assuming each edge of the box is parallel to a coordinate axis.

To solve this we note that the dimensions of the box will be $2x, 2y, 2z$, so that the volume of the box will be

$$V = 8xyz.$$

Obviously to maximize V, x, y, z must all be nonzero. Thus we wish to find critical points of the function

$$u(x, y, z, \lambda) = 8xyz + \lambda\left(\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} - 1\right).$$

In this case, the equations for the critical points of u become

$$\begin{aligned}
8yz + 2\lambda x/a^2 &= 0 \\
8xz + 2\lambda y/b^2 &= 0 \\
8xy + 2\lambda z/c^2 &= 0 \\
x^2/a^2 + y^2/b^2 + z^2/c^2 - 1 &= 0
\end{aligned}$$

Multiplying the first equation by x, the second by y and the third by z, summing them and applying the fourth equation, we find

$$24xyz + 2\lambda = 0.$$

Therefore $2\lambda - 24xyz$. Putting this into the first equation, and dividing by yz, we find

$$8 = 24x^2/a^2.$$

Solving for x, we see $x = a/\sqrt{3}$. Similarly, we obtain $y = b/\sqrt{3}$ and $z = c/\sqrt{3}$. Thus the dimensions of the box are $2a/\sqrt{3} \times 2b/\sqrt{3} \times 2c/\sqrt{3}$.

The method of Lagrange multipliers is also easily generalized to maximizing and minimizing subject to several constraints. See the text for details.

12 Day 12

Exam 1

13 Day 13

Today, we will prove a theorem mentioned previously regarding sufficient conditions for differentiability of a function of several variables.

Theorem 13.1. *Suppose that $F : \mathbb{R}^n \to \mathbb{R}^m$ is defined on a region R of $\mathbb{R}^n$ and that $(a_1, \ldots, a_n) \in R$. Then if the various partial derivatives of the component functions of F, $\partial f_k/\partial x_j$, exist for all $1 \leq j \leq n, 1 \leq k \leq m$ in a ball $B_r(a_1, \ldots, a_n) \subseteq R$ and are continuous at $(a_1, \ldots, a_n)$, then F is differentiable at $(a_1, \ldots, a_n)$.*

Proof. Let $g_{jk} = \frac{\partial f_k}{\partial x_j}$. Then by assumption

$$\lim_{\vec{x} \to \vec{a}} g_{jk}(\vec{x}) = g_{jk}(\vec{a}).$$

Hence □

Sufficient Conditions for Differentiability

14 Day 14

Changing the Order of Differentiation

Theorem 14.1 (Clairaut)**.** *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is a real valued function defined in some region R of $\mathbb{R}^n$. If for some point $\vec{a} \in R$ the second partial derivatives $\frac{\partial^2 f}{\partial x \partial y}$ and $\frac{\partial^2 f}{\partial y \partial x}$ exist in some $B_r(\vec{a})$ and are continuous at $\vec{a}$, then*

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}.$$

Slight generalization by Schwarz:

Theorem 14.2 (Schwarz)**.** *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is a real valued function defined in some region R of $\mathbb{R}^n$. If for some point $\vec{a} \in R$ the second partial derivative $\frac{\partial^2 f}{\partial x \partial y}$ exists in some $B_r(\vec{a})$ and is continuous at $\vec{a}$, then the second partial derivative $\frac{\partial^2 f}{\partial y \partial x}$ must also exist at $\vec{a}$ and*

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}.$$

An alternative sufficient condition:

Theorem 14.3. *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is a real valued function defined in some region R of $\mathbb{R}^n$. If for some point $\vec{a} \in R$ the first partial derivatives $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ exists in some $B_r(\vec{a})$ and are differentiable at $\vec{a}$, then the second partial derivatives $\frac{\partial^2 f}{\partial x \partial y}$ and $\frac{\partial^2 f}{\partial y \partial x}$ must also exist at $\vec{a}$ and*

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}.$$

15 Day 15

The Mean Value Theorem

Theorem 15.1. *Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is a real-valued function defined in some region R of $\mathbb{R}^n$. Suppose further that $\vec{a} = (a_1, \dots, a_n), \vec{b} = (b_1, \dots, b_n)$ are points in R, with the line segment $L(\vec{a}, \vec{b})$ connecting them also in R. Further assume that f is continuous on the line segment, and differentiable on the line segment, except possibly at the endpoints. Then there exists a point $\vec{c}$ on $L(\vec{a}, \vec{b})$ such that*

$$f(\vec{b}) - f(\vec{a}) = \frac{\partial f}{\partial x_1}(\vec{c})(b_1 - a_1) + \frac{\partial f}{\partial x_2}(\vec{c})(b_2 - a_2) + \cdots + \frac{\partial f}{\partial x_n}(\vec{c})(b_n - a_n).$$

16 Day 16

Taylor's Formula

Theorem 16.1 (Taylor's Theorem)**.** *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is a k-times differentiable function defined in some region R of $\mathbb{R}^n$. If for some $\vec{a} \in R$, f is k-times differentiable at $\vec{a}$, then using multi-index notation*

$$f(\vec{x}) = f(\vec{a}) + \sum_{j=1}^{k} \sum_{|\alpha|=j} \frac{1}{\alpha_1! \alpha_2! \dots \alpha_n!} \left(\frac{\partial^j f}{\partial x_1^{\alpha_1} \dots \partial x_n^{\alpha_n}}(\vec{a}) \right) (x_1 - a_1)^{\alpha_1} \dots (x_n - a_n)^{\alpha_n} + R_k(\vec{x})$$

where $R_k : \mathbb{R}^n \to \mathbb{R}$ is a function defined on R satisfying

$$\lim_{\vec{x} \to \vec{a}} \frac{\|R(\vec{x})\|}{\|\vec{x} - \vec{a}\|^k} = 0.$$

The function R is sometimes called the remainder term.

As a special case, we have the following

Corollary 16.2 (Taylor's Theorem for Functions with Two variables)**.** *Suppose $f(x, y)$ is a k-times differentiable function defined in some region R of $\mathbb{R}^2$. If for some $(a, b) \in R$, f is k-times differentiable at (a, b), then*

$$f(x,y) = f(a,b) + \sum_{j=1}^{k}\sum_{i=0}^{j} \frac{1}{i!(j-i)!}\left(\frac{\partial^j f}{\partial x^i \partial y^{j-i}}(a,b)\right)(x-a)^i(y-b)^{j-i} + R_k(x,y).$$

for some function $R_k(x, y)$ defined on the region R, and satisfying

$$\lim_{(x,y)\to(0,0)} \frac{R(x,y)}{((x-a)^2+(y-b)^2)^{k/2}} = 0.$$

If we assume a little more information, we can also obtain an explicit expression for the remainder term.

17 Day 17

Similar Matrices, Diagonalizability, and the Spectral Theorem

18 Day 18

The Second Derivative Test

Definition 18.1. If f is a real valued function on $\mathbb{R}^n$ that is twice differentiable at a point $\vec{a}$, then the *Hessian* of f at $\vec{a}$ is defined to be the matrix of second partial derivatives

$$H_f(\vec{a}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(\vec{a}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\vec{a}) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\vec{a}) \\ \frac{\partial^2 f}{\partial x_2 x_1}(\vec{a}) & \frac{\partial^2 f}{\partial x_2^2}(\vec{a}) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(\vec{a}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n x_1}(\vec{a}) & \frac{\partial^2 f}{\partial x_n \partial x_2}(\vec{a}) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(\vec{a}) \end{pmatrix}$$

Using the Hessian, we have the following characterization of the critical points of f.

Theorem 18.2 (Second Derivative Test)**.** *Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is defined on some region R of $\mathbb{R}^n$. Assume further that f is twice differentiable at a point $\vec{a}$ and that $\vec{a}$ is a critical point of f. If the eigenvalues of the Hessian $H_f(\vec{a})$ are $\lambda_1, \lambda_2, \ldots, \lambda_n$, the have the following*

- *If $\lambda_1, \ldots, \lambda_n$ are all strictly greater than 0, then f has a local minimum at $\vec{a}$*
- *If $\lambda_1, \ldots, \lambda_n$ are all strictly less than 0, then f has a local maximum at $\vec{a}$*

- *If some $\lambda_i < 0$ and some $\lambda_j > 0$, then f has a saddle point at $\vec{a}$, regardless of the other eigenvalues (some of which may be zero)*
- *If $\lambda_1, \ldots, \lambda_n$ are all greater than or equal to zero, but at least one is equal to zero, then the second derivative test fails, and we are no sure about the type of critical point*
- *If $\lambda_1, \ldots, \lambda_n$ are all less than or equal to zero, but at least one is equal to zero, then the second derivative test fails, and we are no sure about the type of critical point*

19 Day 19-20

The Inverse Function Theorem

Theorem 19.1 (Inverse Function Theorem)**.** *Suppose that $F : \mathbb{R}^n \to \mathbb{R}^n$ is a vector-valued function defined on an open set U of $\mathbb{R}^n$, and let $\vec{a} \in U$. Further assume that*

(i) there is a ball $B_r(\vec{a})$ on which F is differentiable with continuous partial derivatives

(ii) the total differential $dF(\vec{a})$ of F at $\vec{a}$ is a nonsingular (ie. invertible) matrix

Then there exists $0 < s \leq r$ such that the following is true

(a) F is a 1-to-1 function on $B_s(\vec{a})$

(b) the set $V = f(B_s(\vec{a}))$ is open

(c) F has an inverse function F^{-1} defined on V which is differentiable on V with continuous partial derivatives

20 Day 21

The Implicit Function Theorem

Theorem 20.1 (Implicit Function Theorem)**.** *Suppose that we are given functions $g_1(x_1, \ldots, x_n), \ldots, g_m(x_1, \ldots, x_n)$ defined in some region R of $\mathbb{R}^n$, with $m < n$. Let $\ell = n - m$, and for convenience relabel $x_1, \ldots, x_n$ as $y_1, \ldots, y_\ell, z_1, \ldots, z_m$. Suppose that we have the following*

(i) a point $(a_1, \ldots, a_\ell, b_1, \ldots, b_m) = (\vec{a}, \vec{b}) \in \mathbb{R}^n$ satisfying $g_j(\vec{a}, \vec{b}) = 0$ for all j.

(ii) the functions $g_1, \ldots, g_n$ are differentiable with continuous partial derivatives in some ball $B_r((\vec{a}, \vec{b}))$.

(iii) the $m \times m$ square matrix

$$\begin{pmatrix} \frac{\partial g_1}{\partial z_1}(\vec{a},\vec{b}) & \frac{\partial g_1}{\partial z_2}(\vec{a},\vec{b}) & \dots & \frac{\partial g_1}{\partial z_m}(\vec{a},\vec{b}) \\ \frac{\partial g_2}{\partial z_1}(\vec{a},\vec{b}) & \frac{\partial g_2}{\partial z_2}(\vec{a},\vec{b}) & \dots & \frac{\partial g_2}{\partial z_m}(\vec{a},\vec{b}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial z_1}(\vec{a},\vec{b}) & \frac{\partial g_m}{\partial z_2}(\vec{a},\vec{b}) & \dots & \frac{\partial g_m}{\partial z_m}(\vec{a},\vec{b}) \end{pmatrix}$$

is nonsingular.

Then there exists a ball $B_s(\vec{a})$ and functions $u_1(y_1,\dots,y_\ell),\dots,u_m(y_1,\dots,y_\ell)$ defined on $B_s(\vec{a})$ such that

(a) $u_1,\dots,u_m$ are differentiable on $B_s(\vec{a})$

(b) for all $\vec{y}=(y_1,\dots,y_\ell)$ in $B_s(\vec{a})$, we have that

$$g_j(\vec{y},u_1(\vec{y}),u_2(\vec{y}),\dots,u_m(\vec{y}))=0, \text{ for all } j.$$

In other words, in the ball $B_s(\vec{a})$, the constraint equations

$$\begin{aligned} g_1(y_1,\dots,y_\ell,z_1,\dots,z_m) &= 0 \\ g_2(y_1,\dots,y_\ell,z_1,\dots,z_m) &= 0 \\ &\vdots \\ g_m(y_1,\dots,y_\ell,z_1,\dots,z_m) &= 0 \end{aligned}$$

define $z_1,\dots,z_m$ implicitly as functions of the $y_1,\dots,y_\ell$.