LIMITS IN MULTIVARIABLE CALCULUS

MATH 195, SECTION 59 (VIPUL NAIK)

Corresponding material in the book: Section 14.2

What students should definitely get: The rough $\varepsilon - \delta$ of limit (modulo knowledge from one variable). Computation techniques and rules for limits for polynomials, rational functions, and other kinds of functions.

What students should hopefully get: The distinction between multiple inputs and multiple outputs, the distinction between joint continuity and separate continuity, the extent to which concepts from functions of one variable generalize and don't generalize to functions of several variables.

Executive summary

Words ...

(1) Conceptual definition of limit $\lim_{x \to c} f(x) = L$: For any neighborhood of L, however small, there exists a neighborhood of c such that for all $x \neq c$ in that neighborhood of c, $f(x)$ is in the original neighborhood of L.
(2) Other conceptual definition of limit $\lim_{x \to c} f(x) = L$: For any open ball centered at L, however small, there exists an open ball centered at c such that for all $x \neq c$ in that open ball, $f(x)$ lies in the original open ball centered at L.
(3) $\varepsilon - \delta$ definition of limit $\lim_{x \to c} f(x) = L$: For any $\varepsilon > 0$, there exists $\delta > 0$ such that for all $x = \langle x_1, x_2, \dots, x_n \rangle$ satisfying $0 < |x - c| < \delta$, we have $|f(x) - L| < \varepsilon$. The definition is the same for vector inputs and vector outputs, but we interpret subtraction as vector subtraction and we interpret $|\cdot|$ as length/norm of a vector rather than absolute value if dealing with vectors instead of scalars.
(4) On the range/image side, it is possible to break down continuity into continuity of each component, i.e., a vector-valued function is continuous if each component scalar function is continuous. This cannot be done on the domain side.
(5) We can use the above definition of limit to define a notion of continuity. The usual limit theorems and continuity theorems apply.
(6) The above definition of continuity, when applied to functions of many variables, is termed *joint continuity*. For a jointly continuous function, the restriction to any continuous curve is continuous with respect to the parameterization.
(7) We can define a function of many variables to be a continuous in a particular variable if it is continuous in that variable when we fix the values of all other variables. A function continuous in each of its variables is termed *separately continuous*. Any jointly continuous function is separately continuous, but the converse is not necessarily true.
(8) Geometrically, separate continuity means continuity along directions parallel to the coordinate axes.
(9) For homogeneous functions, we can talk of the order of a zero at the origin by converting to radial/polar coordinates and then seeing the order of the zero in terms of r.

Actions ...

(1) Polynomials and sin and cos are continuous, and things obtained by composing/combining these are continuous. Rational functions are continuous wherever the denominator does not blow up. The usual *plug in to find the limit* rule, as well as the usual list of indeterminate forms, applies.
(2) Unlike the case of functions of one variable, the strategy of canceling common factors is not sufficient to calculate all limits for rational functions. When this fails, and we need to compute a limit at the origin, try doing a polar coordinates substitution, i.e., $x = r\cos\theta$, $y = r\sin\theta$, $r > 0$. Now try to find the limit as $r \to 0$. If you get an answer independent of θ in a strong sense, then that's the limit. This method works best for homogeneous functions.

(3) For limit computations, we can use the usual chaining and stripping techniques developed for functions of one variable.

1. Limits: basic definition

1.1. **Recall of the definition in one variable.** Let's first recall the definition of limit in the context of functions from subsets of $\mathbb{R}$ to $\mathbb{R}$.

Suppose f is a function from a subset of $\mathbb{R}$ to $\mathbb{R}$, and c is a point in the *interior* of the domain of f (i.e., f is defined on an open interval around c). For a real number L, we say that $\lim_{x \to c} f(x) = L$ if the following holds:

For every $\varepsilon > 0$, there exists $\delta > 0$ such that if $0 < |x - c| < \delta$, then $|f(x) - L| < \varepsilon$.

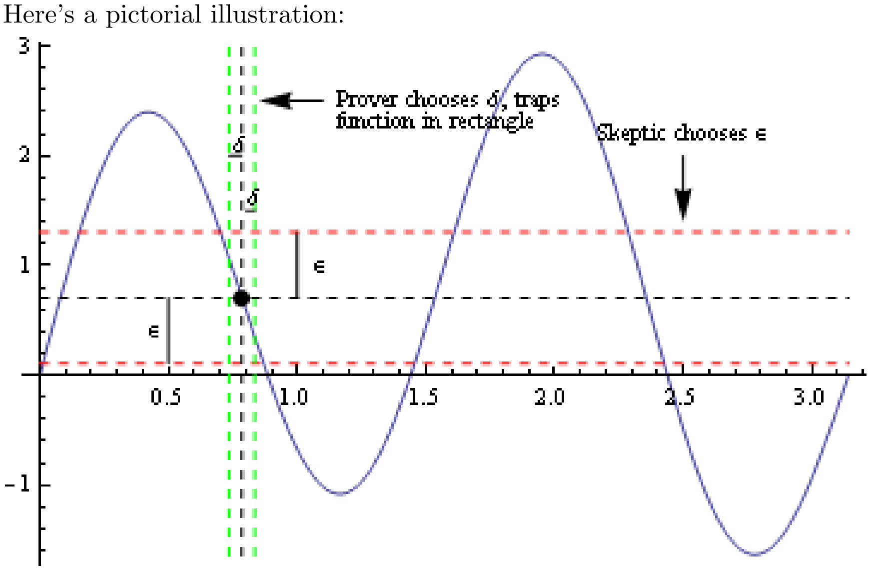
The way I think about the definition is in terms of a *cage* (or *trap*). And the reason why we need this notion of a cage or trap is precisely to avoid these kinds of oscillations that give rise to multiple limits. So, here is the formal definition:

We say that $\lim_{x \to c} f(x) = L$ (as a two-sided limit) if, for every $\varepsilon > 0$, there exists $\delta > 0$ such that, for every x such that $0 < |x - c| < \delta$, we have $|f(x) - L| < \varepsilon$.

That's quite a mouthful. Let's interpret it graphically. What it is saying is that: "for every ε" so we consider this region $(L - \varepsilon, L + \varepsilon)$, so there are these two horizontal bars at heights $L - \varepsilon$ and $L + \varepsilon$. Next it says, there exists a δ, so there exist these vertical bars at $c + \delta$ and $c - \delta$. So we have the same rectangle that we had in the earlier definition.

Here is another way of thinking of this definition that I find useful: as a *prover-skeptic game*. Suppose I claim that as x tends to c, $f(x)$ tends to L, and you are skeptical. So you (the skeptic) throw me a value $\varepsilon > 0$ as a challenge and say – can I (the prover) trap the function within ε? And I say, yeah, sure, because I can find a $\delta > 0$ such that, within the ball of radius δ about c, the value $f(x)$ is trapped in an interval of size ε about L. So basically you are challenging me: can I create an ε-cage? And for every ε that you hand me, I can find a δ that does the job of this cage.

Here's a pictorial illustration:



In other words, we have the following sequence of events:

- The skeptic chooses $\varepsilon > 0$ in order to challenge or refute the prover's claim that $\lim_{x \to c} f(x) = L$
- The prover chooses $\delta > 0$ aiming to trap the function within an ε-distance of L.
- The skeptic chooses a value of x in the interval $(c - \delta, c + \delta) \setminus \{c\}$.
- The judge now computes whether $|f(x) - L| < \varepsilon$. If yes, the prover wins. If not, the skeptic wins.

We say that $\lim_{x \to c} f(x) = L$ if the prover has a *winning strategy* for this game, i.e., a recipe that allows the prover to come up with an appropriate δ for any ε. And the statement is false if the skeptic has a winning strategy for this game.

Now, in single variable calculus (possibly in the 150s sequence if you took that sequence) you mastered a bunch of generic winning strategies for specific forms of the function f. In particular, you saw that for constant functions, *any* strategy is a winning strategy. For a linear function $f(x) := ax + b$, the strategy $\delta = \varepsilon/|a|$ is a winning strategy. The winning strategy for a quadratic function $f(x) := ax^2 + bx + c$ at a point $x = p$ is more complicated; one formula is $\min\{1, \varepsilon/(|a| + |2ap + b|)\}$.

Conversely, to show that a limit does *not* exist, we try to show that the skeptic has a winning strategy, i.e., we find a value of ε such that when the skeptic throws that at the prover, any δ that the prover throws back fails, in the sense that the skeptic can find a value of x satisfying $0 < |x - c| < \delta$ but $|f(x) - L| \geq \varepsilon$.

1.2. **Beyond δs and εs: thinking using neighborhoods.** The strict $\varepsilon - \delta$ definition does not make clear what the key element is that's ripe for generalization. To see how this can be generalized, we need to take a more abstract perspective. Here is the more abstract definition:

> For any neighborhood of L, however small, there exists a neighborhood of c such that for all $x \neq c$ in that neighborhood of c, $f(x)$ is in the original neighborhood of L.

This *is* the $\varepsilon - \delta$ definition, albeit without an explicit use of the letters ε and δ. Rather, I have used the term *neighborhood* which has a precise mathematical meaning. Making things more formal in the language we are familiar with, we can say:

> For any open ball centered at L, however small, there exists an open ball centered at c such that for all $x \neq c$ in that open ball, $f(x)$ lies in the original open ball centered at L.

An open ball is described by means of its radius, so if we use the letter ε for the radius of the first open ball and the letter δ for the radius of the second open ball, we obtain:

> For any $\varepsilon > 0$, there exists $\delta > 0$ such that for all $x \in (c - \delta, c + \delta) \setminus \{c\}$, we have $f(x) \in (L - \varepsilon, L + \varepsilon)$.

Or, equivalently:

> For any $\varepsilon > 0$, there exists $\delta > 0$ such that for all x satisfying $0 < |x - c| < \delta$, we have $|f(x) - L| < \varepsilon$.

Although it is the final formulation that we use, the first two formulations are conceptually better because they avoid unnecessary symbols and are also easier to generalize to other contexts.

The key advantage, from our perspective, of thinking about *neighborhoods* and *open balls* instead of *intervals* is that these ideas continue to work in higher dimensions. The main difference is that, in higher dimension, the open "balls" are now disks (the interiors of spheres) rather than merely intervals.

1.3. **The formal definition of limit.** Suppose f is a function from a subset D of $\mathbb{R}^n$ to a subset of $\mathbb{R}$. Suppose $c = (c_1, c_2, \ldots, c_n)$ is a point in the *interior* of D, i.e., D contains an open ball about c. Then, for a real number L, we say that $\lim_{x \to c} f(x) = L$ if we have the following:

> For any $\varepsilon > 0$, there exists $\delta > 0$ such that for all $x = (x_1, x_2, \ldots, x_n)$ such that the distance between x and c is greater than 0 and less than δ, we have $|f(x) - L| < \varepsilon$.

Formally, this definition is exactly the same as before, but now, the geometric distance between x and c plays the role that the absolute value $|x - c|$ played in the past.

If we thought of the inputs as vectors instead of points, so $c = \langle c_1, c_2, \ldots, c_n \rangle$ and $x = \langle x_1, x_2, \ldots, x_n \rangle$, then the distance between x and c is $|x - c|$, i.e., the *length* of the vector $x - c$. With this notation, the definition looks exactly like it does for functions of one variable:

> For any $\varepsilon > 0$, there exists $\delta > 0$ such that for all $x = \langle x_1, x_2, \ldots, x_n \rangle$ satisfying $0 < |x - c| < \delta$, we have $|f(x) - L| < \varepsilon$.

1.4. **Case of vector inputs, vector outputs.** if f is a function from a subset of $\mathbb{R}^n$ to a subset of $\mathbb{R}^m$, i.e., a vector-valued function with vector inputs, then we can use the same definition, but now, we use the "length" notion, instead of absolute value on both the domain and the range side. In particular, for $c = \langle c_1, c_2, \ldots, c_n \rangle$ in the interior of the domain of f, and $L = \langle L_1, L_2, \ldots, L_m \rangle$, we say that $\lim_{x \to c} f(x) = L$ if the following holds:

> For any $\varepsilon > 0$, there exists $\delta > 0$ such that for all $x = \langle x_1, x_2, \ldots, x_n \rangle$ satisfying $0 < |x - c| < \delta$, we have $|f(x) - L| < \varepsilon$.

However, there is a critical distinction to keep in mind. In the case of multiple outputs, there is an alternative definition of limit: namely, the limit of a vector-valued function is the vector of the limits of each of its component scalar functions. *There is no such shortcut on the domain side.* We'll talk more about isolating coordinates a little later.

1.5. **Everything you thought of as true is true.** The following results are true, with the usual conditional existence caveats:

- The limit, if it exists, is unique.
- The limit of the sum is the sum of the limits.
- Constant scalars can be pulled out of limits.
- The limit of the difference is the difference of the limits.
- The limit of the quotient is the quotient of the limits.
- Post-composition with a continuous function can be pulled in and out of limits.

2. Continuity: basic definition and theorems

2.1. **The corresponding definition of continuity.** This comes as no surprise: f is continuous at a point if the limit of f at the point equals the value of f at the point.

The concept of continuity on a subset is trickier, because of the existence of boundary points. Boundary points are points which do not lie in the interior, i.e., for which there is no open ball containing the point that lies completely inside the subset. For boundary points, we modify the definition somewhat – for the boundary point c in a subset D of $\mathbb{R}^n$, we say that $\lim_{x \to c} f(x) = L$ with respect to D if it satisfies the following:

> For any $\varepsilon > 0$, there exists $\delta > 0$ such that for all $x = \langle x_1, x_2, \dots, x_n \rangle$ *is in the subset* D and satisfies $0 < |x - c| < \delta$, we have $|f(x) - L| < \varepsilon$.

With this caveat, we can now define a function to be continuous on a subset of $\mathbb{R}^n$ if it is continuous with respect to the subset at all points in the subset.

2.2. **Continuity theorems.** The limit theorems give rise to corresponding continuity theorems:

- A sum of continuous functions is continuous.
- A scalar times a continuous function is continuous.
- A product of continuous functions is continuous.
- A quotient of continuous functions is continuous at all points where the denominator function is nonzero.

2.3. **Fixing all but one coordinate.** Suppose f is a function from a subset of $\mathbb{R}^n$ to $\mathbb{R}$. One way of thinking about the concept of continuity is that if we tweak all the coordinates just a little bit, the function value changes only a little bit. This suggests another notion of continuity:

> Suppose i is a natural number between 1 and n. We say that a function f of n variables $x_1, x_2, \dots, x_n$ is continuous in the variable x_i if, *once we fix the value of all the other variables*, the corresponding function is continuous in the (single) variable x_i which can still vary freely.

Further:

> We say that a function f of n variables $x_1, x_2, \dots, x_n$ is *separately continuous* in each of the variables x_i if it is continuous in each variable x_i once we fix all the other variable values.

The notion of continuity defined earlier is *joint continuity* and this is the default notion of continuity for a function of several variables. It turns out that a (jointly) continuous function is also separately continuous, i.e., it is continuous in each variable. However, the converse is not true, i.e., it is possible for a function to be separately continuous but not jointly continuous. The reason is roughly that separate continuity only guarantees continuity *if we change only one variable at a time* whereas joint continuity guarantees continuity *under simultaneous changes in the values of multiple variables.* This also has a geometric interpretation in terms of directions of approach.

2.4. **Directions of approach: left, right, up, down, sideways, spiral.** When we deal with functions of one variable, there are two directions of approach on the *domain* side: left and right. These two directions of approach give rise to the notions of *left hand limit* and *right hand limit* respectively (a limit from one side is generically termed a *one-sided limit*).

How many directions of approach are there for a function of 2 variables? In one sense, there are 4 directions of approach: the positive and negative directions of approach in each coordinate. Pictorially, if the two inputs are put on the xy-plane as the x-axis and y-axis, then the four directions of approach are *left*, *right*, *up*, and *down*. In cartographic terminology, up is north, right is east, down is south, and left is west.

Now, if we were only interested in *continuity in each variable in isolation*, then these would be the only four directions of approach that concern us. In other words, as far as *separate continuity* is concerned, there are only 4 directions of approach. However, we are concerned with *joint continuity*, which allows us to simultaneously change the values of multiple variables. Thus, we need to seriously consider (i) diagonal directions of approach, i.e., approach along arbitrary linear directions, and (ii) non-linear directions of approach, such as spiral approach or other curved approach.

This helps clarify the significant difference between joint and separate continuity. With separate continuity, we only care about the directions of approach along or parallel to the coordinate axes (left and right) so if there are n variables, there are only $2n$ directions of possible approach. With joint continuity, on the other hand, we care about infinitely many different directions of approach, and want the function to be continuous when restricted to any of these possible curves. Joint continuity is thus considerably stronger than separate continuity.

The key thing to remember is the following:

> If the limit of a function exists in the joint continuity/limit sense, then this is the same as the limit for any direction of approach, whether linear or curved. Thus, if we find multiple directions of approach with different limits, or any direction of approach with no limit, then the limit does not exist in a joint snese.

2.5. **Example of a separately continuous, not jointly continuous function.** Consider the following function defined on the plane $\mathbb{R}^2$. It is defined as follows:

$$f(x,y) := \left\{ \begin{array}{rl} \frac{xy}{x^2+y^2}, & (x,y) \neq (0,0) \\ 0, & (x,y) = (0,0) \end{array} \right.$$

For any input other than the origin, this is $(1/2)\sin(2\theta)$ where θ is the polar angle under the polar coordinate system where the pole is the origin and the polar axis is the x-axis.

We claim the following:

- *This function is (jointly) continuous, and hence separately continuous, at all points other than the origin*: The form $(xy)/(x^2+y^2)$ involves quotients of continuous functions, so it is continuous everywhere that the denominator is nonzero, which means everywhere other than the origin. Alternatively, we can see this from the polar description.
- *The function is separately continuous in each variable at the origin*: From either the algebraic or the polar description, we see that the function is zero everywhere on the x-axis and te y-axis, hence it is continuous at the origin for directions of approach along the axes.
- *The function is not jointly continuous at the origin*: To see this, note that for any linear direction of approach other than the axes,we do not get a limit of 0. For instance, for $m \neq 0$ on the line $y = mx$ (minus the origin) the function is a constant function $m/(1+m^2)$, and the limit of this at the origin is thus also $m/(1+m^2)$, which is nonzero. This can also be seen from the polar description as $(1/2)\sin(2\theta)$: from that description, it is clear that for every linear direction of approach, the function is a constant, but this constant differs as we change the linear direction of approach.

2.6. **Approach along straight lines: is it enough?** We just saw that continuity in each variable is not enough to guarantee joint continuity, and hence, to show that a limit exists, it is not enough simply to consider approach along the coordinate axes. The next question might be: *what about all straight line directions of approach*? If we compute the limit, or verify continuity, along all straight line directions of approach, is that enough? Or is it possible that the limit/continuity fails when we consider parabolic or spiral approach?

The answer is *no*. It is possible for there to exist a point and a function such that the limit of the function along any straight line approach to the point equals the value, *but* there exist non-straight line approaches where the limit is not equal to the value. The explanation for this, though, is usually some more obvious and glaring discontinuities around other points and in other lines.

2.7. **Separate and joint continuity: real world.** Joint continuity is a typical assumption made in modeling real world situations, particularly situations where the quantities being measured are large aggregates. Let's think about the example of quantity demanded by a household for a good as a function of the unit price of the good, the unit prices of substitute goods, the unit prices of complementary goods, and other variables. *Separate continuity* would mean that if one of these variables is changed *ceteris paribus* on the other variables, then the quantity demanded varies continuously with the variable being changed. Joint continuity would mean that simultaneous slight perturbations in multiple determinants of demand lead to only a slight perturbation in the quantity demanded.

3. Limits of functions of two variables: practical tricks

3.1. **Summary of ideas.** For a function of several variables, if we want to *compute the limit*, we try to use the various limit theorems to compute limits: limit of sums, differences, products, pull out scalar multiples, post-composition with continuous functions.

If we want to *show the limit does not exist*, we try one of these two methods: (i) find a direction of approach for which the limit does not exist, or (ii) find two directions of approach that give different limits.

The easiest way to implement a "direction of approach" is to simply fix one coordinate and make the other coordinate approach the point, i.e., the "continuous in each variable" thinking. However, as the example of $(xy)/(x^2 + y^2)$ illustrates, simply using these directions of approach may paint a misleading picture: the limit may exist/the function may appear continuous using only these directions of approach, but there may be others that give a different result.

From now on, as far as most actual examples are concerned, we restrict attention to functions of two variables. However, most of what we say applies to functions of n variables.

3.2. **Polynomial functions.** A polynomial function of two (or more) variables is *jointly continuous everywhere*. This means that in order to calculate the limit of such a function at a point, it suffices to plug in the value at the point.

For instance:

$$\lim_{(x,y)\to(2,5)} x^3 - xy^2 + y^4 = 2^3 - (2)(5)^2 + 5^4 = 8 - 50 + 625 = 583$$

Here's one way of seeing this. Any polynomial in x and y is a sum of monomials in x and y, and each monomial is the product of a power of x times a power of y.

First, the functions $(x, y) \mapsto x$ and $(x, y) \mapsto y$ are themselves continuous. Thus, each of the functions of the form $x^a y^b$ is continuous (because it's a product of continuous functions). Finally, the polynomial is continuous because it is a sum of these continuous monomial functions.

3.3. **General indeterminate form rules.** These rules are pretty much the same as for functions of one variable. For typical situations involving polynomial and trigonometric functions, the first thing we try is to plug in the point. If we get a numerical answer, then that is the limit. *Always plug in first.*

If we have a fraction, then it could happen that the denominator approaches 0. $(\to 0)/(\to 0)$ is an indeterminate form, and means that *more work* is needed to determine whether a limit exists and what its value is. If the numerator approaches a nonzero number, and the denominator approaches 0, then the limit does not exist.

3.4. **The case of rational functions.** A rational function in two variables is the quotient where both the numerator and the denominator are polynomials in the two variables. For instance, $(x^2+y^2-x^3y-1)/(x^3+x^2y^4+5)$ is a rational function. As mentioned above, for a rational function, the following basic rules apply:

- If, at the point where we need to calculate the limit, the denominator is nozero, we can compute the limit by evaluation.
- If, at the point where we need to calculate the limit, the denominator is zero and the numerator is nonzero, the limit does not exist.
- If, at the point where we need to calculate the limit, both the numerator and the denominator become zero, we have an indeterminate form and need to do more work.

However, unlike the case of functions of one variable, this strategy of finding and canceling factors proves grossly inadequate both in cases where the limit does exist and in cases where it does not. Roughly, this is because there is no precise analogue of the factor theorem for polynomials in more than one variable, and in particular, an expression of x and y being zero at a point does not guarantee the existence of a "factor" of a particular form for that expression.[1]

Thus, we need an alternate way of thinking about these limits. We tackle the problem by first restricting attention to the special case of *homogeneous polynomials* and the rational functions obtained as quotients of such polynomials.

3.5. **Homogeneous polynomials and rational functions.** A *homogeneous polynomial* of homogeneous degree d in the variables $x_1, x_2, \dots, x_n$ is a function $F(x_1, x_2, \dots, x_n)$ with the property that the *total degree* of $x_1, x_2, \dots, x_n$ in every monomial that constitute that polynomial is d. For instance, the polynomial $F(x,y) = x^2 - xy + 3y^2$ is homogeneous of degree 2 in x and y , but the polynomial $x^2 - xy^3$ is not homogeneous because its monomials have different degrees (2 and 4 respectively).

A *homogeneous function* of homogeneous degree d in the variables $x_1, x_2, \dots x_n$ is a function with the property that, for any $a \in \mathbb{R}$ (or perhaps restricted to some large subset of $\mathbb{R}$ if there are domain restrictions on the function):

$$F(ax_1, ax_2, \dots, ax_n) = a^d F(x_1, x_2, \dots, x_n)$$

Any homogeneous polynomial of degree d is also a homogeneous function of degree d.

Here are some rules for homogeneous functions:

- The zero function is homogeneous of any degree (sort of)
- The sum of homogeneous functions of the same homogeneous degree is also homogeneous of the same degree, unless it is identically the zero function.
- The product of homogeneous functions of degrees d_1 and d_2 is homogeneous of degree $d_1 + d_2$.
- The reciprocal of a homogeneous function of degree d is homogeneous of degree $-d$.
- The composite (in a painful sense, don't take this at face value) of homogeneous functions of degrees d_1 and d_2 is homogeneous of degree $d_1 d_2$.
- The k^{th} power of a homogeneous function of degree d is homogenous of degree kd.

3.6. **Rational functions, homogeneous and otherwise, and radial coordinates.** We discuss the radial/polar coordinate approach now. This approach is particularly useful for homogeneous functions, although it also has applications to some non-homogeneous functions.

The idea is as follows: Suppose we want to compute $\lim_{(x,y)\to(0,0)} F(x,y)$. This is equivalent to trying to compute $\lim_{r\to 0} F(r\cos\theta, r\sin\theta)$. More precisely, $\lim_{(x,y)\to(0,0)} F(x,y)$ exists if and only if the limit $\lim_{r\to 0} F(r\cos\theta, r\sin\theta)$ exists as an actual number, with no appearance of θ in the final expression. In other words, the answer is independent of θ in a strong sense (*joint* rather than *separate*).

Conceptually, any fixed value of θ describes an approach to the origin/pole from the ray making that angle with the x-axis. The limit for a fixed value of θ is the limit for approach along such a ray. By saying that we get a constant answer with no appearance of θ, we are basically saying that the limit does not depend on the direction of approach.[2]

In particular, we note that if F is a homogeneous function of degree d in x and y, we can write:

[1]More sophisticated versions of the result are true even in multiple variables, but this gets us into pretty deep mathematics.
[2]There is a "separate" versus "joint" subtlety here, but it's too tricky to explain, so we're glossing over it.

$$F(x, y) = r^d g(\theta)$$

where g is a new function of the "dimensionless" (in the sense of being free of length units) variable θ.

If g is continuous, then it is a continuous function on the closed interval $[0, 2\pi]$, hence it is bounded from both above and below. In particular, we see that under these conditions:

- If $d > 0$, then the limit is 0.
- If $d = 0$, then the limit is well defined *only if* g is a constant function, which means that F to begin with is a constant function.
- If $d < 0$, then the limit is not defined because magnitudes of function values are going to ∞.

Note that in case of homogeneous rational functions, the homogeneous degree is the difference of homogeneous degrees of numerator and denominator, so we obtain the following:

- If the (homogeneous) degree of the numerator is greater than the degree of the denominator, the limit is 0.
- If the degrees are equal, the limit is undefined (unless the numerator is a constant multiple of the denominator).
- If the degree of the denominator is greater, the limit is undefined, because magnitudes of function values are going to ∞.

4. Nostalgia time: limits in one variable

We now review some ideas from single variable calculus, and try to understand what they tell us about life with many variables.

4.1. Zeroeyness: order of zero. Consider a function f of one variable x. Suppose $\lim_{x\to c} f(x) = 0$. The *order* of this zero is defined as the least upper bound of the set of values β such that $\lim_{x\to c} |f(x)|/|x-c|^\beta = 0$. If we denote this order by r, the following are true:

- For $\beta < r$, $\lim_{x\to c} |f(x)|/|x - c|^\beta = 0$.
- For $\beta > r$, $\lim_{x\to c} |f(x)|/|x - c|^\beta$ is undefined, or $+\infty$.
- The limit $\lim_{x\to c} |f(x)|/|x - c|^r$ may be zero, infinity, or a finite nonzero number, or undefined for other reasons.

Roughly speaking, the order describes *how zeroey* the zero of f is around c.

For an infinitely differentiable function f , the order of any zero, if finite, must be a positive integer. Further, it can be computed as follows: the order of the zero is the smallest positive integer k such that the k^{th} derivative of f at c is nonzero.

For convenience, in the subsequent discussion, we restrict attention to the case that $c = 0$, i.e., the point in the domain at which we are taking the limit is 0. Thus, instead of $x - c$, we just write x.

We note the following:

- If f_1 and f_2 have zeros of orders r_1 and r_2 respectively at c, then f_1+f_2 has a zero of order $\min\{r_1, r_2\}$ at c if $r_1 \neq r_2$, and *at least* $\min\{r_1, r_2\}$ at c if $r_1 = r_2$.
- If f_1 and f_2 have zeros of orders r_1 and r_2 respectively at c, the pointwise product $f_1 f_2$ has a zero of order $r_1 + r_2$ at c.
- If f_1 has a zero of order r_1 at c and f_2 has a zero of order r_2 at 0, then $f_1 \circ f_2$ has a zero of order $r_1 r_2$ at c.

4.2. Strippable functions. I will call a function f *strippable* if f is differentiable at 0, $f(0) = 0$ and $f'(0) = 1$. In particular, this means that $\lim_{x\to 0} f(x)/x = 1$. Strippable functions have a zero of order 1 at zero.

Here are some strippable functions: sin, tan, $x \mapsto \ln(1+x)$, $x \mapsto e^x - 1$, arcsin, arctan. The significance of strippable functions is as follows: if the quantity inside of a strippable function is going to zero, and we are in a multiplicative situation, then the strippable function can be stripped off to compute the limit. Composing with strippable functions does not affect the order of a zero.

4.3. **Stripping: some examples.** To motivate stripping, let us look at a fancy example:

$$\lim_{x\to 0} \frac{\sin(\tan(\sin x))}{x}$$

This is a composite of three functions, so if we want to chain it, we will chain it as follows:

$$\lim_{x\to 0} \frac{\sin(\tan(\sin x))}{\tan(\sin x)} \frac{\tan(\sin x)}{\sin x} \frac{\sin x}{x}$$

We now split the limit as a product, and we get:

$$\lim_{x\to 0} \frac{\sin(\tan(\sin x))}{\tan(\sin x)} \lim_{x\to 0} \frac{\tan(\sin x)}{\sin x} \lim_{x\to 0} \frac{\sin x}{x}$$

Now, we argue that each of the inner limits is 1. The final limit is clearly 1. The middle limit is 1 because the inner function $\sin x$ goes to 0. The left most limit is 1 because the inner function $\tan(\sin x)$ goes to 0. Thus, the product is $1 \times 1 \times 1$ which is 1.

If you are convinced, you can further convince yourself that the same principle applies to a much more convoluted composite:

$$\lim_{x\to 0} \frac{\sin(\sin(\tan(\sin(\tan(\tan x)))))}{x}$$

However, *writing that thing out takes loads of time*. Wouldn't it be nice if we could just strip off those sins and tans? In fact, we can do that.

The key stripping rule is this: *in a multiplicative situation* (i.e. there is no addition or subtraction happening), if we see something like $\sin(f(x))$ or $\tan(f(x))$, and $f(x) \to 0$ in the relevant limit, then we can strip off the sin or tan. In this sense, both sin and tan are *strippable* functions. A function g is strippable if $\lim_{x\to 0} g(x)/x = 1$.

The reason we can strip off the sin from $\sin(f(x))$ is that we can multiply and divide by $f(x)$, just as we did in the above examples.

Stripping can be viewed as a special case of the l'Hopital rule as well, but it's a much quicker shortcut in the cases where it works.

Thus, in the above examples, we could just have stripped off the sins and tans all the way through.

Here's another example:

$$\lim_{x\to 0} \frac{\sin(2\tan(3x))}{x}$$

As $x \to 0$, $3x \to 0$, so $2\tan 3x \to 0$. Thus, we can strip off the outer sin. We can then strip off the inner tan as well, since its input $3x$ goes to 0. We are thus left with:

$$\lim_{x\to 0} \frac{2(3x)}{x}$$

Cancel the x and get a 6. We could also do this problem by chaining or the l'Hopital rule, but stripping is quicker and perhaps more intuitive.

Here's yet another example:

$$\lim_{x\to 0} \frac{\sin(x\sin(\sin x))}{x^2}$$

As $x \to 0$, $x\sin(\sin x) \to 0$, so we can strip off the outermost sin and get:

$$\lim_{x\to 0} \frac{x\sin(\sin x)}{x^2}$$

We cancel a factor of x and get:

$$\lim_{x\to 0} \frac{\sin(\sin x)}{x}$$

Two quick sin strips and we get x/x, which becomes 1.

Yet another example:

$$\lim_{x \to 0} \frac{\sin(ax)\tan(bx)}{x}$$

where a and b are constants. Since this is a multiplicative situation, and $ax \to 0$ and $bx \to 0$, we can strip the sin and tan, and get:

$$\lim_{x \to 0} \frac{(ax)(bx)}{x}$$

This limit becomes 0, because there is a x^2 in the numerator and a x in the denominator, and cancellation of one factor still leaves a x in the numerator.

Here is yet another example:

$$\lim_{x \to 0} \frac{\sin^2(ax)}{\sin^2(bx)}$$

where a, b are nonzero constants. We can pull the square out of the whole expression, strip the sins in both numerator and denominator, and end up with a^2/b^2.

Here's another example:

$$\lim_{x \to 0} \frac{\arcsin(2\sin^2 x)}{x \arctan x}$$

Repeated stripping reveals that the answer is 2. Note that arcsin and arctan are also strippable because $\lim_{x \to 0}(\arcsin x)/x = 1$ and $\lim_{x \to 0}(\arctan x)/x = 1$.

4.4. **Thinking of L'Hôpital's rule.** L'Hôpital's rule is a rule to compute limits of the indeterminate form $(\to 0)/(\to 0)$. The key idea is that for an indeterminate form of this sort, we differentiate both the numerator and the denominator and try to compute the limit again.

In terms of orders of zero, this can be viewed as follows: each application of the L'Hôpital's rule reduces the order of zero in the numerator by one *and* reduces the order of zero in the denominator by one. In particular, we see the following:

(1) When the numerator has higher order of zero than the denominator, then the quotient approaches zero. In the case where both orders are positive integers, repeated application of the LH rule will get us to a situation where the denominator becomes nonzero (because the order of the zero in the denominator becomes zero) while the numerator is still zero (because the order of the zero in the denominator is still positive) – yes, you read that correct.
(2) When the numerator and the denominator have the same order, the quotient *could* approach something finite and nonzero. In most cases, repeated application of the LH rule gets us down to a quotient of two nonzero quantities.
(3) When the denominator has the higher order, the quotient has an undefined limit (the one-sided limits are usually $\pm\infty$). In the case where both orders are positive integers, repeated application of the LH rule will get us to a situation where the numerator becomes nonzero while the denominator is still zero (because the order of the zero is still positive).

4.5. **Taylor polynomials and Taylor series.** To compute the order of a zero of f at a point c, we can consider Taylor polynomials/Taylor series of f at $x - c$, and look at the smallest r such that the coefficient of $(x - c)^r$ is nonzero. This is the order of the zero. Note that this definition of order is the same as the definition we gave earlier as the number of times we need to differentiate to get a nonzero value. Moreover, the *value* of the limit $\lim_{x \to c} f(x)/(x - c)^r$ is that nonzero coefficient.

For convenience, as before, we set $c = 0$, so $x - c$ can simply be written as x. The order is thus the smallest power with a nonzero coefficient of x in the Taylor series. The value $\lim_{x \to 0} f(x)/x^r$ is the nonzero coefficient.

4.6. **Quick order computations and application to limits.** We know that the functions sin, tan, arcsin, arctan, $x \mapsto e^x - 1$, $x \mapsto \ln(1+x)$ are all strippable and in particular have order 1. All these facts can also be seen in terms of the Taylor series for these functions.

Let's now consider some examples of zeros of order 2 at zero: $1 - \cos x$, $1 - \cosh x$, $\sin^2 x$, $\sin(x^2)$.

Here are some examples of zeros of order 3 at zero: $\sin^3 x$, $\sin(x^3)$, $\tan(x \sin x) \arctan x$, $x \sin(e^{\sin^2 x} - 1)$, $x - \sin x$, and $x - \tan x$. With the exception of the last two examples, all of these can be justified using the way order of zero interacts with multiplication and composition. For $x - \sin x$ and $x - \tan x$, we can use either the Taylor series/power series expansions *or* we can just see how many times we need to differentiate in order to hit a nonzero number.

Similarly, the function $x^2 \sin^3(x^3)$ has a zero of order $2 + (3)(3) = 2 + 9 = 11$ at zero.

The order of a zero can also be fractional. This does *not* happen for infinitely differentiable functions, but can happen in other cases. For instance, $\sin^3(x^{7/5})$ has order of zero at zero of 3 times 7/5 which is 21/5 or 4.2.

5. The multi-variable generalizations

Now that we've recalled how things worked with one variable, it is time to study the generalization to multiple variables. Specifically, *does it make sense to talk of the order of a zero for a function of two variables*, and can this be used to compute limits?

For simplicity, we restrict attention to limit computations at the point $(0, 0)$, just as in the single variable case, we largely restricted attention to the case $c = 0$. However, most of the ideas we present continue to work for limit computations at other points.

Also, our ideas generalize to functions of more than two variables.

5.1. **Stripping still works!** It continues to be true that in a *multiplicative* situation, we can strip off all the strippable functions as long as the input to these functions is approaching zero. For instance, consider the limit computation:

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

The sin in the numerator can be stripped, because we can multiply and divide by $x^4 y$, and, *crucially*, we know that as $(x, y) \to (0, 0)$, $x^4 y \to 0$. Thus, we get the limit:

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

Now, using the general rules on homogeneous degree, or performing the polar coordinate substitution, we see that this limit is zero.

5.2. **Concept of order: too many variables!** The concept of order of zero does not *quite* make sense, but there are some situations where it does.

The most interesting case is that of *homogeneous* functions, which we have already discussed. In the case of a homogeneous function of degree $d > 0$, the "order" of the zero is also d, when viewed as a function of the radial coordinate r. This provides a fresh perspective on some of the observations made earlier about homogeneous functions.

In general, the concept of order of zero does not make sense because it differs depending upon the direction of approach. For instance, consider the function $x^3 + xy + y^5$. If we consider approach along the x-axis, the order of zero (as a function of x) is 3. If we consider approach along the y-axis, the order of zero (as a function of y) is 5. If we consider approach along any other linear direction, say $y = mx$, the order of zero turns out to be 2, because we get:

$$x^3 + mx^2 + m^5 x^5$$

and the smallest power with nonzero coefficient is x^2.

Thus, the order of zero could depend on the direction of approach. Nonetheless, it often makes sense to talk of the generic order of zero, which is the order of the zero for most directions of approach. As we can see from the above, in case of a polynomial, this is the minimum of the total degrees in x and y of all the monomoials constituting that polynomial.