Functions of several variables

While a parametrized curve is given by a point-valued function, that is a function that takes a scalar (a number) as input and gives a point as output, the main object of study in this class is the reverse: a function that takes a point as input and gives a number as output. Since a point is given by a list of numbers (its coordinates), a function of this sort can also be viewed as taking a list of numbers as input; for this reason, we call it a **function of several variables** (the variables in question being those that stand for the input numbers).

The hierarchy of functions and relations

There are many different types of mathematical objects that we could study in this class. Some of them are relation-like objects:

- truth values,
- sets,
- relations,
- ternary relations,
- quaternary relations,
- etc;

some of them are function-like objects:

- constants,
- functions,
- binary functions,
- ternary functions,
- etc.

As you go along these lists, both the number of variables and the number of dimensions needed for graphing increase, as in the following diagram:

$$\begin{tikzcd}
& \text{truth values} \arrow[dl] \arrow[dr, dash] & & \\
\text{zero dimensions} & & \text{constants} \arrow[dl, dash] \arrow[dr] & \\
& \text{sets} \arrow[dl] \arrow[dr, dash] & & \text{no variables} \\
\text{one dimension} & & \text{functions} \arrow[dl, dash] \arrow[dr] & \\
& \text{relations} \arrow[dl] \arrow[dr, dash] & & \text{one variable} \\
\text{two dimensions} & & \text{binary functions} \arrow[dl, dash] \arrow[dr] & \\
& \text{ternary relations} \arrow[dl] \arrow[dr, dash] & & \text{two variables} \\
\text{three dimensions} & & \text{ternary functions} \arrow[dl, dash] \arrow[dr] & \\
& \text{quaternary relations} \arrow[dl, dotted] \arrow[dr, dotted, dash] & & \text{three variables} \\
 & & &
\end{tikzcd}$$

A **truth value** is either true or false; any statement with no variables in it, such as the statement that $0 < 2$, should evaluate to true or false (in this case, true). To indicate that you are talking about the truth value of this statement, rather than asserting the statement itself, you can put curly braces around it (although there are several other notations used for this); for example, $\{0 < 2\}$ is the truth value that 0 is less than 2, which is the true truth value rather than the false one. You can also use a variable to give a name to a truth value, so maybe p stands for $\{0 < 2\}$; we won't need to do this in this course, but you'll do this constantly if you take a course on Logic.

A **constant** is, in this class, a *real number*, such as -2. Any expression with no variables should evaluate to a constant, but we use one dimension to graph a constant on a number line. Again, you can use a variable to stand for a constant, so maybe a stands for -2; in other words, $a = -2$.

A **set** is, in the simplest case, a *set of real numbers*. A statement with one variable defines a set, such as $\{x \mid x < 2\}$, the set of real numbers that are less than 2. We again use one dimension to graph a set. If A stands for the set $\{x \mid x < 2\}$, then these two statements mean the same thing:

- $x \in A$, usually pronounced 'x in A';
- $x < 2$.

The first of these says that x *belongs* to the set A, while the second uses the definition of A to say precisely what that means about x.

A **function**, or *unary function* for emphasis, is a rule for taking a number (the *input*) and using it to calculate a number (the *output*). An example is $(x \mapsto x - 2)$, the rule which subtracts 2 from any number. To graph a function, we need two dimensions, one for the input and one for the output. If f stands for the function $(x \mapsto x - 2)$, then these two expressions mean the same thing:

- $f(x)$, usually pronounced 'f of x';
- $x - 2$.

The first of these is the *value* of the function f at the *argument* x, while the second uses the definition of f to say precisely what that means in terms of x.

A **relation**, or *binary relation* for emphasis, is a set of ordered pairs instead of a set of individual numbers. An example is $\{x, y \mid x + y < 2\}$. We again use two dimensions to graph a relation. If R stands for the relation $\{x, y \mid x + y < 2\}$, then these two statements mean the same thing:

- $(x, y) \in R$;
- $x + y < 2$.

The first of these says that x and y are *related* by the relation R, while the second uses the definition of R to say precisely what that means in terms of x and y.

A **binary function**, or *function of two variables*, is a rule for taking an ordered pair of two inputs and using it to calculate an output. An example is $(x, y \mapsto x + y - 2)$, the rule which subtracts 2 from the sum of the two inputs. To graph a binary function, we need three dimensions, two for the inputs and one for the output. If g stands for the function $(x, y \mapsto x + y - 2)$, then these two expressions mean the same thing:

- $g(x, y)$;
- $x + y - 2$.

A **ternary relation**, or *relation between three variables*, is a set of ordered triples instead of a set of ordered pairs. An example is $\{x, y, z \mid x + y + z < 2\}$. We again use three dimensions to graph a ternary relation.

A **ternary function**, or *function of three variables*, is a rule for taking an ordered triple of three inputs and using it to calculate an output. An example is $(x, y, z \mapsto x + y + z - 2)$, the rule which subtracts 2 from the sum of the three inputs. To graph a ternary function, we need four dimensions, three for the inputs and one for the output.

A **quaternary relation**, or *relation between four variables*, is a set of ordered quadruples. An example is $\{x_1, x_2, x_3, x_4 \mid x_1 + x_2 + x_3 + x_4 < 2\}$. We again use four dimensions to graph a quaternary relation.

We can continue with quaternary functions, quinary functions, etc, which are functions of four or more variables; and we can continue with quinary relations, senary relations, etc, which are relations between five or more variables. (But around this point, most people stop using the '-ary' terms, because few people can remember them.)

There are various relationships between these different kinds of objects:

- The domain of a function of n variables is a relation between n variables (the same n variables).
- The range of a function of any number of variables is a set (a relation with 1 variable, the output).
- The graph of a function of n variables is the graph of a relation between $n+1$ variables (the n input variables plus the 1 output variable), which contains all of the information in the function.

For example, a binary function (a function of 2 variables) has a relation (a binary relation, a relation between 2 variables) as its domain, a set (a unary relation, a relation with 1 variable) as its range, and a ternary relation (a relation between 3 variables) as its graph. In particular, if $f(a,b)=c$, then $(a,b) \in \operatorname{dom} f$ (where $\operatorname{dom} f$ is the domain of f), $c \in \operatorname{ran} f$ (where $\operatorname{ran} f$ is the range of f), and $(a,b,c) \in \operatorname{gr} f$ (where $\operatorname{gr} f$ is the graph of f).

Definitions for functions of several variables

In order to form precise definitions of various concepts related to functions of several variables, it's handy to piggyback on the definitions for functions of one variable. This is *not* the way that the book writes its definitions, but it's the way that I prefer. So here are my definitions.

General principles

Recall that a *parametrized curve*, or *point-valued function*, takes a number to a point (in however many dimensions we're dealing with, typically 2 or 3 dimensions). That is, if C is a parametrized curve and t is a real number, then $C(t)$ is a point $P=(x,y)$, $P=(x,y,z)$, etc. Meanwhile, a *function of several variables* (however many variables we're dealing with, typically 2 or 3 variables) takes a point to a number; that is, if f is a function of 2 or 3 variables and $P=(x,y)$ or $P=(x,y,z)$ is a point in 2 or 3 dimensions, then $f(P)=f(x,y)$ or $f(P)=f(x,y,z)$ is a real number c. If we combine these by composition of functions, then $f \circ C$ is an ordinary function; that is, if t is a real number, then so is $(f \circ C)(t)$:

$$(f \circ C)(t) = f\big(C(t)\big) = f(P) = c.$$

From one-variable Calculus, you should know how to define various concepts (continuity, limits, differentiability, derivatives, differentials) for ordinary functions. It's easy to extend these concepts to vector- and point-valued functions (parametrized curves), since these simply consist of several ordinary functions (the coordinates or components). So to define these concepts for functions of several variables, we typically use a formula like this:

> If $f \circ C$ has a certain property whenever C does, no matter what C might be (as long as it has the property), then that's what it means for f to have that property.

This formula doesn't always work perfectly; for one thing, we often want to say more than just a Yes/No property, and it may not be obvious what matters about C or how to extract the appropriate information from the composites. Besides that, even when this formula would make perfect sense, sometimes some of the nice theorems that we would expect aren't always true, which means that we should look for a modifed definition that makes the theorems work. (That's what mathematicians really want from a definition; they're not handed down from on high but developed for the purpose of getting correct results.) Nevertheless, all of the definitions here will be based on something like this formula.

Continuity

Continuity follows the general formula precisely. A function f of several variables is **continuous** if, whenever C is a continuous parametrized curve, the composite $f \circ C$ is a continuous function. (It wouldn't be fair to expect $f \circ C$ to be continuous unless C is continuous as well as f, but if both C and f are continuous, then their composite ought to be as well.)

Sometimes we want to look at continuity in more detail; in general, to say that a function is continuous really means that it's continuous at every number in its domain. So for a function of several variables, we want to talk about continuity at particular points in its domain. A function f is **continuous** at a point P_0 in the domain of f if, whenever C is a parametrized curve and t_0 is a number such that

$C(t_0) = P_0$ and C is continuous at t_0, then $f \circ C$ is also continuous at t_0. Again, it wouldn't be fair to demand more than this if we're only asking f to be continuous at P_0.

An equivalent definition is to say that f is continuous at P_0 if f is defined at P_0 and, for every positive number ϵ, there is some positive number δ such that, whenever $\|P - P_0\| < \delta$ and f is defined at P, then $|f(P) - f(P_0)| < \epsilon$. This is essentially how it is defined in the textbook. However, this ϵ-δ stuff is rather less fun to work with. Ultimately, you have to say something like this some time, but I prefer to say it once, when giving the first definition in one-variable Calculus, and then never again.

Any function with a formula that is built out of the coordinate variables using only the usual operations is continuous wherever it is defined. (To be definite, the usual operations are addition, subtraction, multiplication, division, taking opposites, taking reciprocals, taking absolute values, raising to powers with constant exponents and/or positive bases, extracting roots with constant indexes and/or positive radicands, logarithms, the six trigonometric operations, and the six inverse trigonometric operations. Some notable operations *not* on this list are piecewise definitions and powers where the exponent varies and the base may be zero or negative.)

To prove this, you use the continuity of each component of a continuous parameterized curve and the one-variable theorem that any function built out of continuous functions using these operations is continuous. For example, if f and g are continuous at P_0 and I want to prove that $f + g$ is continuous at P_0, consider a parametrized curve C and a number t_0 such that $C(t_0) = P_0$ and C is continuous at t_0; by definition, $f + g$ is continuous at P_0 if, for each such C and t_0, $(f + g) \circ C$ is continuous at t_0. Since f is continuous at P_0, this means (by definition) that $f \circ C$ is continuous at t_0; similarly, since g is continuous at P_0, this means that $f \circ C$ is continuous at t_0. By a theorem in one-variable Calculus, since $f \circ C$ and $g \circ C$ are both continuous at t_0, so is their sum $(f \circ C) + (g \circ C)$. But $(f \circ C) + (g \circ C)$ is the same function as $(f + g) \circ C$, since they do the same thing to any input t:

$$\Big((f \circ C) + (g \circ C)\Big)(t) = (f \circ C)(t) + (g \circ C)(t) = f\Big(C(t)\Big) + g\Big(C(t)\Big);$$
$$\Big((f + g) \circ C\Big)(t) = (f + g)\Big(C(t)\Big) = f\Big(C(t)\Big) + g\Big(C(t)\Big).$$

Therefore, $(f + g) \circ C$ is continuous at t_0. Since this argument works for any relevant C and t_0, this proves that $f + g$ is continuous at P_0, as desired. (Similar arguments work for all of the other operations.)

Limits

To keep things simple, we'll only look at finite limits approaching a finite value; none of our limits will involve infinity in any role. (Things will become more complicated in another way shortly!)

There is a technicality about limits that's often ignored in one-variable Calculus, which is that the expression whose limit you're taking must be defined at numbers arbitrarily close to the number that the variable is approaching. It's often treated as a big deal that the function doesn't have to be defined at that number precisely, which is certainly true and important, but it still has to be defined *near* that number. For example (and assuming that we're only working with real numbers), you can't talk about the limit of $\sqrt{t}$ as $t \to -1$, because t can't get very close to -1 while $\sqrt{t}$ is defined. On the other hand, it's fine to talk about the limit as $t \to 0$, because even though $\sqrt{t}$ is undefined when $t < 0$, still $\sqrt{t}$ is defined when $t > 0$, which allows t to get arbitrarily close to 0. (But on the other other hand, you can't talk about the limit as $t \to 0^-$, because now this requires $t < 0$, which leaves $\sqrt{t}$ undefined again.)

A number t_0 is a **limit point** of a set D if it makes sense to talk about a function defined on D as having a limit approaching t_0, in other words if there exists a function whose domain is D (a constant function will do) that has a limit approaching t_0. (The term 'limit point' is traditional even in one dimension, even though I would normally call t_0 a number rather than a point.) This is equivalent to saying that there are numbers in D (other than possibly t_0 itself) that are arbitrarily close to t_0, in other words if, given any positive distance $\delta > 0$, there is at least some number t in the set D such that $0 < |t - t_0| < \delta$. But I prefer to think of the definition that has no δ (or ϵ) in it.

Keeping this technicality in mind, the **limit** approaching a point P_0 of a function f of several variables (which in symbols we can write as

$$\lim_{P \to P_0} f(P),$$

that is

$$\lim_{(x,y) \to P_0} f(x, y)$$

in 2 dimensions or

$$\lim_{(x,y,z) \to P_0} f(x, y, z)$$

in 3 dimensions) is the unique number L (if this exists) such that, whenever C is a parametrized curve and t_0 is a number, if $C(t) = P_0$ when and only when $t = t_0$, and if C is continuous at t_0, and if t_0 is a limit point of the domain of $f \circ C$, then L is the limit of $f \circ C$ approaching t_0. In other words (ignoring the fine print),

$$\lim_{P \to P_0} f(P) = L$$

if

$$\lim_{t \to t_0} f\big(C(t)\big) = L$$

whenever

$$\lim_{t \to t_0} C(t) = P_0.$$

The point of all of that is this: the limit of one of these composites is basically the limit of the function along a particular curve. If the function is undefined along the curve, then we don't expect its limit to exist, and this is what the clause about limit points takes care of. We also don't want to worry about $f(P_0)$ itself, since f might not be continuous at P_0, which is why $C(t)$ is not allowed to be P_0 except when $t = t_0$. So we're only looking at certain curves that are *appropriate* to the problem. Then, in order for the limit to exist overall, the limit must exist along each appropriate curve and be the same along all of them.

If for any appropriate curve, there is no limit along that curve, then the limit overall does not exist. Besides that, if there are two appropriate curves such that the limits along them are different, then again the limit does not exist overall. It is in this way that one generally proves that a limit does not exist, when it doesn't. When the limit does exist, however, then you usually need to find a general argument to show that it does and what it is, because you can't actually check every individual curve. Fortunately, we have a theorem that

$$\lim_{P \to P_0} f(P) = f(P_0)$$

whenever f is continuous at P_0 (assuming that P_0 is a limit point of the domain of f), as in one-variable Calculus.

One often talks about limits with restrictions on the manner of approaching the point. For example, instead of saying (x, y) approaches $(2, 3)$, we might say that (x, y) approaches $(2, 3)$ *while* $x \neq y$. (An analogue in one-variable Calculus is, for example, $t \to 0^-$; that is, $t \to 0$ while $t < 0$.) Technically, this is handled by modifying the function so that it is defined only when the given restriction is met (so in this example, the function would be undefined when $x = y$). That is,

$$\lim_{\substack{(x,y) \to (2,3) \\ x \neq y}} f(x, y) = \lim_{(x,y) \to (2,3)} (f(x, y) \text{ for } x \neq y),$$

where by '$f(x, y)$ for $x \neq y$' I mean $f(x, y)$ if $x \neq y$ but something undefined if $x = y$.

Differentiability

The way that differentiability fits in with composition of functions is the chain rule

$$(f \circ g)'(t) = f'\big(g(t)\big)g'(t).$$

Following the general principle, we replace g with a parametrized curve C, and the values of the derivatives of this (replacing $g'(t)$) are vectors. However, the composite is an ordinary function, so the derivative of f should multiply by a vector to get a scalar. One way to do this is to multiply a vector by a vector with the dot product, so the derivative of a function of several variables should also be a vector. (Since we want this concept to make sense even when lengths and angles don't apply, this vector is going to have to be a *row* vector; see page 9 of the handout on vectors.) There are actually several sorts of derivatives in higher dimensions, and we'll come back to this subject later; but the one which is a vector will provide the definition of differentiability.

We say that the function f is **differentiable** at some point P_0 if there exists a (row) vector $\mathbf{v}$ such that, whenever C is a parametrized curve and t_0 is a number such that $C(t_0) = P_0$ and C is differentiable at t_0, then $f \circ C$ is also differentiable at t_0 and furthermore $(f \circ C)'(t_0) = \mathbf{v} \cdot C'(t_0)$. If f is differentiable at every point P_0 in its domain, then f is simply *differentiable*.

This vector $\mathbf{v}$ is called the **gradient** of f at P_0 and may be written as $\nabla f(P_0)$ (although $f'(P_0)$ would make a lot of sense), so the basic rule is

$$(f \circ C)'(t) = \nabla f\big(C(t)\big) \cdot C'(t).$$

Higher differentiability

Where a function f is differentiable, the components of its gradients define some more functions, called the **partial derivatives** of f. (We will do more with these partial derivatives later on.) Wherever the partial derivatives are themselves continuous, the original function is **continuously differentiable**. Where the partial derivatives are themselves differentiable, the original function is **twice differentiable**. Where the partial derivatives are continuously differentiable, the original function is **twice continuously differentiable**. Etc etc etc. (As in one-variable Calculus, there is a theorem that a differentiable function must be continuous, so a twice-differentiable function must be continuously differentiable, etc.)

Where this goes on forever, the function is **infinitely differentiable**: it is differentiable, its partial derivatives are differentiable, their partial derivatives are differentiable, etc. Any function built out of the usual operations is infinitely differentiable except at certain exceptional places where a derivative fails to exist, such as when taking the absolute value or square root of zero. But to prove this, it's best to look at how to calculate the derivatives, which I'll get to next.