PARAMETRIC STUFF

MATH 195, SECTION 59 (VIPUL NAIK)

Corresponding material in the book: Section 10.1, 10.2. *We are omitting the topic of surface area mentioned at the end of Section 10.2 of the book.*

What students should definitely get: Parametric descriptions of curves, relationship with functional and relational descriptions, domain of parameters, parametric differentiation, arc length.

What students should hopefully get: Distinction between curve and parameterization, self-intersections, going back and forth between parametric and relational descriptions, the subtletly behind higher derivatives.

Note on lecture notes and book

The lecture notes are intended to be a reasonably faithful version of what was planned for the lecture, though the actual contents of the lecture may deviate somewhat.

I generally avoid doing a lot of worked examples of the kind that are already present in the book. You are encouraged to read the worked examples and other discussions in the book to get a different perspective when you sit down to do the weekly homework.

The book often has some fancy examples (for instance, cycloids, strophoids, conchoids, cissoids, cochleoids, cardioids, etc.) that are fun to read, but you are not expected to have a thorough mastery of this stuff for the course. The level of difficulty of the tests is roughly going to be like the routine homeworks, with a few multiple choice questions of the level of difficulty of the quizzes and one question at the level of the advanced homeworks.

Executive summary

Words ...

(1) A parametric description of a curve is one where both coordinates are expressed as functions of of a parameter, typically denoted t. Parametric descriptions offer an alternative to functional and implicit (relational) descriptions of curves. Here, t varies over some subset of the real numbers. In symbols, we have something like $x = f(t)$, $y = g(t)$, where t varies over some subset D of the real numbers.

(2) Descriptions where x is a function of y or y is a function of x are special cases of parametric descriptions.

(3) The same curve may admit multiple parametrizations, and different parameterizations may correspond to different speeds and different orderings of traversal of the point. The curve itself only contains the information of *what* points were traversed, not the information of the *sequence* and *pace* in which they were traversed.

(4) The curve traced by a parameterization depends not only on the coordinate functions but also the domain for the parameter. The larger the domain, in general, the larger the curve traced. However, in some cases, expanding the domain may not make the curve strictly larger. This happens in cases where both coordinate functions are even or have commensurable periods.

(5) A parameterization of a curve may involve self-intersections, retracings (e.g., tracing back for even function pairs), or even wrapping around itself (for periodic function pairs).

(6) Function composition allows us to switch between multiple parameterizations.

(7) In some cases, it is possible to move back and forth between parametric and relational descriptions.

(8) Parametric differentiation: if $x = f(t)$ and $y = g(t)$, then $dy/dx = (dy/dt)/(dx/dt) = g'(t)/f'(t)$. This can also be used to differentiate repeatedly. Note that the derivative is a function of t rather than of (x, y), so to find the derivative given the point (x, y) we need to go back and determine t.

(9) Higher derivatives can be computed iteratively using parametric differentiation. But note that it is *not* true that $d^2y/dx^2 = (d^2y/dt^2)/(d^2x/dt^2)$. The actual formula/procedure is more complicated (see lecture notes).

(10) Arc length: The formula for arc length from $t = a$ to $t = b$ (with $a < b$) is $\int_a^b \sqrt{(dx/dt)^2 + (dy/dt)^2}\, dt$.

Actions ...

(1) *Parametric to relational: elimination of parameter*: In many cases, it is possible to eliminate a parameter from a parametric description. The idea is to use some well known identities or manipulation techniques to try to directly relate x and y by finding some equation between them that is true for all t. However, this is not the full story. We next need to see if there are additional restrictions on x and y deducible from the fact that they arose as function of t, also keeping in mind the domain restrictions on t.

For instance, the parameterization $x = t^2, y = t^4$ for $t \in \mathbb{R}$ can be rewritten as $y = x^2$, but we need the additional condition that $x \geq 0$.

See more examples in the lecture notes, quizzes, and homeworks.

(2) *Relational to parametric*: Here, we see a relation between x and y, and try to choose a parametric description that would give rise to the relation. Again, the domain of choice for the parameter needs to be chosen wisely.

See more examples in the lecture notes and quizzes.

(3) *Parametric differentiation and geometric consequences*: We use the formula $(dy/dt)/(dx/dt)$. If $x = f(t)$ and $y = g(t)$, then this becomes $g'(t)/f'(t)$. This is valid for all t in the interior of the domain of definition where both f' and g' are defined and $f' \neq 0$. If $f'(t) = 0$ but $g'(t) \neq 0$, we have a vertical tangent situation. If $g'(t) = 0$ but $f'(t) \neq 0$, we have a horizontal tangent situation.

1. An introduction to the parametric approach

1.1. **Describing curves: the old ways.** In single variable calculus, you saw two main kinds of ways that curves in the xy-coordinate plane can be described:

- Explicit functional descriptions, where one coordinate is written as a function of the other, i.e., $y = f(x)$ or $x = f(y)$ type form. The $y = f(x)$ type description must satisfy the vertical line test.
- Implicit descriptions or relational descriptions, where an expression in both variables is declared to be zero, i.e., $R(x, y) = 0$ where R is an expression that involves both x and y. Implicit descriptions need not define functions globally, though under suitable conditions, they may still define functions locally. You saw implicit descriptions when looking at implicit differentiation, and later you saw that the general solution to a differential equation often appears naturally first in the form of an implicit description, and depending on the circumstances, it may or may not be possible to write y *explicitly* as a function of x.

1.2. **A new approach: a parameter.** In addition to *functional* and *relational* descriptions of curves, there is a third kind of description called a *parametric* description. Here, we find functions f and g and a parameter t and write:

$$x = f(t), y = g(t), t \in D$$

where D is a subset of $\mathbb{R}$. The curve traced by this parametric description is precisely the set of points:

$$\{(f(t), g(t)) : t \in D\}$$

One way of thinking of this is that t is time, and there is a particle such that the x-coordinate of its location at time t is $f(t)$ and the y-coordinate is $g(t)$. In other words, the location of the particle at time t is functionally dependent on t, with each coordinate given by its own function. The set D is the set of times for which this is valid.

1.3. **The relation between functional and parametric descriptions.** Functional descriptions (where one of the coordinates x, y is expressed in terms of the other) are special cases of parametric descriptions as follows:

- If f is the identity function (i.e., $f(t) = t$ by definition) then the curve is the graph of the function $y = g(x)$ with $x \in D$.
- If g is the identity function (i.e., $g(t) = t$ by definition) then the curve is the graph of the function $x = f(y)$ with $y \in D$.

1.4. **Parameterization.** Given a curve, a particular parametric description of the curve is termed a *parameterization* of the curve. We say that the parameterization is *continuous* if the function f and g are continuous (and also, typically, we assume that D is connected) and we say that the parameterization is *smooth* if it is continuous and the functions f and g are continuously differentiable on the interior of D (in some definitions, "smooth" requires us to assume that f and g are both infinitely differentiable).

1.5. **Parameterization: a story reconstruction.** A curve, or a subset of the plane, is a static entity. It's just there, all at once.

A parameterization creates a *story* behind the curve. The story reveals, how, with the passage of time (the parameter t) the curve was gradually constructed.

However, the important point to note here is that we have a lot of leeway in how to construct this story. We can think of the curve as a collection of trails left behind by a wild animal – but it's up to us to connect the dots and decide just *how fast* the wild animal moved. We even have leeway in determining *which direction* the wild animal moved – there are two mutually opposite directions of motion that would leave the same trail. The trails don't give us conclusive information in this regard.

The upshot is that the same curve can have multiple parameterizations. These parameterizations differ in the speed with which they go through stuff. They could also differ in other subtler ways.

For instance, the curve $y = x^2$ can be represented using the parameterization $x = t$, $y = t^2$ or using the parameterization $x = t^3$, $y = t^6$, or using the parameterization $x = \sinh t$, $y = \sinh^2 t$.

There is an additional leeway we have in some cases: in case the curve is self-intersecting, we have further leeway in choosing the order of paths taken by the curves, i.e., the sequence in which various trails were left.

1.6. **Choice of domain.** The same pair of function f and g can define different curves depending on the domain of the variable t. In general, if $D_1 \subseteq D_2$, then the curve obtained by making t vary over D_1 is a subset of the curve obtained by varying t over D_2.

However, just because D_1 is a *proper* subset of D_2 does not necessarily mean that the corresponding curve is a *proper* subset. For instance, the circle $x^2 + y^2 = 1$ can be described by $x = \cos t$, $y = \sin t$, for $t \in [0, 2\pi]$, but allowing t to vary over the much bigger set of *all* real numbers gives the *same* circle. Similarly, the curve given by $x = t^2, y = t^4$ is the same whether we let t vary over all nonnegative reals or whether we let t vary over all reals. In both cases, it is the graph of the function $y = x^2$ restricted to $x \geq 0$.

1.7. **Different letter for the parameter.** In most cases, we use the parameter t and we're secretly thinking of time. However, there is nothing sacred about t, and we could use any other letter such as u or θ.

1.8. **Switching between multiple parametric descriptions.** Consider a parametric description $x = f(t)$, $y = g(t)$. Suppose we consider a function h. Then we can define a new parametric description using $t = h(u)$, i.e., we now define $x = f(h(u))$ and $y = g(h(u))$ with u the new parameter. We do need to worry about domain issues, but this is best seen in context.

2. The topology and geometry of parametric descriptions

2.1. **Various kinds of repetitions.** The same point (x, y) may occur in a given parametric description for multiple values of t. Some possible ways that this repetition could occur are given below – note that the list is not exhausitve:

- *Self-intersection at an isolated point*: Here, just a single point is repeated, and the behavior around the point is not repeated. The curve thus intersects itself. For instance, the curve $x = t(t-1)(t-2)$, $y = t(t-1)(t-3)$ gives the same point pair $(0, 0)$ for $t = 0$ and $t = 1$, but the curve intersects itself only at this isolated point.

- *Reverse trace*: Here, the curve traces back along itself. This occurs when both functions have a common axis of mirror symmetry. For instance, $x = t(1-t)$ and $y = \sin(\pi t)$ are both functions with mirror symmetry about $t = 1/2$. Thus, the path traced from $-\infty$ to $1/2$ is retraced in reverse from $1/2$ to ∞. As an easier example, if both coordinates are defined by even functions, then the path traced from $-\infty$ to 0 is traced back in reverse from 0 to ∞.
- *Wrap around itself*: Here, the curve wraps around itself. This occurs if both x and y are periodic functions with equal or commensurable periods. The canonical example is $x = \cos t$, $y = \sin t$.

2.2. **Dense fillings and such curves.** Some parameterizations can fill an area pretty densely. One example is the parameterization $x = \sin t$, $y = \sin(\alpha t)$ where α is an irrational number. Note first that both x and y remain in $[-1, 1]$, so this curve remains in a closed bounded region, namely the square with vertices with (x, y)-coordinates $(1, 1)$, $(-1, 1)$, $(-1, -1)$, and $(1, -1)$. It turns out that the curve covers the square densely as $t \in \mathbb{R}$.

2.3. **Closed intervals, open intervals, disconnections, etc.** We have an intuitive idea of what it means for a subset of the real numbers to be *connected*. For subsets of the real numbers to be connected is equivalent to its being an *interval*, which may be open, closed, or infinite at one or both ends.

Typically, when we perform a parameterization $x = f(t)$, $y = g(t)$, $t \in D$, then D is taken to be an interval. We make some notes:

- If D is a closed bounded interval, i.e., of the form $[a, b]$ for $a < b$ both finite, then the curve traced is a curve that includes its endpoints, and hence is also closed and bounded. Basically, the image of something closed and bounded is closed and bounded. This is a result of topology that is also responsible for the *extreme value theorem* which you have seen in single variable calculus.
- In other cases, the curve traced may or may not be closed and bounded. We cannot say anything for sure.

3. Converting back and forth between the various forms

Here we discuss the back and forth:

Parametric description $\leftrightarrow$ Implicit (relational) description

You will not be expected to actually execute this back and forth except in some simple cases. The purpose of this discussion is simply to illustrate this with simple examples.

3.1. **Proceeding from a parametric to a relational description.** The key idea here is to notice how x and y relate to the parameter t, and then try to deduce from this a relationship between x and y that does not involve t. For instance, consider $x = \cos t$ and $y = \sin^2 t$. In this case, we know that:

$$\cos^2 t + \sin^2 t = 1 \ \forall \ t \in \mathbb{R}$$

Thus, we get that x and y satisfy the relation:

$$x^2 + y = 1$$

Note that in this case, we see that y is in fact a function of x:

$$y = 1 - x^2$$

However, and this is the key point, there may well be a loss of information when we move to the relational description. Specifically, the parameterization using t imposes the constraints that $-1 \leq x \leq 1$ and $0 \leq y \leq 1$. On the other hand, the relation $x^2 + y = 1$ covers a larger swath of possibilities for x and y. Thus, we need to *explicitly* include the restrictions based on the range of possible values for x and y, in this case, the restriction that $x \in [-1, 1]$. Since y is determined based on x, the additional restriction on the value of y is not necessary.

Here are some other examples:

- $x = \cosh t$, $y = \sinh t$, $t \in \mathbb{R}$. We obtain that $x^2 = 1 + y^2$, with the additional restriction that $x > 0$ (and no restriction on y). Alternatively, we can express this as $x = \sqrt{1 + y^2}$. If we write it in this form, the additional restriction is unnecessary.

- $x = \cos t$, $y = \cos(2t)$, $t \in \mathbb{R}$. We obtain that $y = 2x^2 - 1$, with the additional restriction that $x \in [-1, 1]$.
- $x = t^3$, $y = t^7$. We can rewrite this as $x^7 = y^3$, with no additional restrictions on x and y. Note that the situation becomes a little more tricky if we have even exponents instead of the odd numbers 3 and 7.
- $x = t^4$, $y = t^2 + 1$, $t \in \mathbb{R}$. We get $x = (y-1)^2$, with the restriction that $y \geq 1$.
- $x = t^2$, $y = t^3$, $t \in [1, 2]$. We get $x^3 = y^2$, with $y \in [1, 8]$.

3.2. **Proceeding from a relational to a parametric description.** Converting a relational description to a *nice* parametric description is tricky, and occasionally requires the invention of new branches of mathematics. For instance, trigonometry is the branch of mathematics that was developed to provide a nice parametric description for the circle $x^2 + y^2 = 1$. There are some simple cases where parametric descriptions are easy to construct, but these are usually exceptions.

For instance, for $x^2 - y^2 = 1$, $x > 0$, we could take the parametric description $x = \cosh t$ and $y = \sinh t$ for $t \in \mathbb{R}$ or the parametric description $x = \sec t$ and $y = \tan t$ for $t \in (-\pi/2, \pi/2)$.

4. Calculus with parameters

4.1. **Implicit differentiation recall.** You're already aware of how to do differentiation and integration, and apply these to geometric situations (tangents, normals, angle of intersection, etc.) for curves given explicitly in functional form. And you've also seen how to do differentiation of curves given in implicit or relational form (it's called *implicit differentiation*).

We'll spend a few seconds recalling implicit differentiation, since there are important parallels and similarities between implicit and parametric differentiation.

Implicit differentiation goes something like this – we start with a mixed expression in x and y that describes the curve, e.g.:

$$\sin(x + y) = xy$$

We differentiate both sides:

$$\frac{d(\sin(x+y))}{dx} = \frac{d(xy)}{dx}$$

Now, how would we handle something like $\sin(x + y)$? It is something in terms of $x + y$, so we use the chain rule on the left side, thinking of $v = x + y$ as the intermediate function:

$$\frac{d(\sin(x+y))}{d(x+y)} \frac{d(x+y)}{dx} = x\frac{dy}{dx} + y\frac{dx}{dx}$$

This simplifies to:

$$\cos(x+y)\left[1 + \frac{dy}{dx}\right] = x\frac{dy}{dx} + y$$

Opening up the parentheses, we get:

$$\cos(x+y) + \cos(x+y)\frac{dy}{dx} = x\frac{dy}{dx} + y$$

Now, we move stuff together to one side, to get:

$$(\cos(x+y) - x)\frac{dy}{dx} = y - \cos(x+y)$$

And we now isolate dy/dx:

$$\frac{dy}{dx} = \frac{y - \cos(x+y)}{\cos(x+y) - x}$$

Notice first that the right side is a *mixed expression in both x and y*, rather than an expression purely in terms of x. This is to be expected since y is not explicitly a function of x.

But something subtler is going on. When we do differentiation with the English letter "d" then that notation makes sense *only* if y is (locally) a function of x. If it isn't, then the answer above notwithstanding, the derivative dy/dx does not exist.

Here's another example: the circle of radius 1 centered at the origin. This is given by the equation $x^2 + y^2 = 1$. Note that in this case, y is *not* a function of x, because for many values of x, there are two values of y. For instance, for $x = 0$, we have $y = 1$ and $y = -1$. For $x = 1/2$, we have $y = \sqrt{3}/2$ and $y = -\sqrt{3}/2$. So, y is not a function of x.

However, *locally* y is still a function of x, in the following sense. If you just restrict yourself to the part above the x-axis, then you do get y as a function of x. This is the function $y := \sqrt{1-x^2}$ for $-1 \leq x \leq 1$. If we restrict ourselves to the part below the x-axis, we consider the function $y := -\sqrt{1-x^2}$ for $-1 \leq x \leq 1$.

Now, how do we calculate dy/dx? Well, it depends on whether we are interested in the part above the x-axis or in the part below the x-axis. For the part above the x-axis, we have the function $\sqrt{1-x^2}$, and we get that the derivative is:

$$\frac{d(\sqrt{1-x^2})}{dx} = \frac{d(\sqrt{1-x^2})}{d(1-x^2)}\frac{d(1-x^2)}{dx} = \frac{1}{2\sqrt{1-x^2}} \cdot (2x) = \frac{-x}{\sqrt{1-x^2}}$$

If we are interested in the lower side, we get $x/\sqrt{1-x^2}$.

Now, in this case, we have to split into two cases, and do a painful calculation involving differentiating a square root via the chain rule.

Let's now consider how to use implicit differentiation to do the same example.

We start with the original expression:

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

This is an *identity*, which means that it's true for every point on the curve. When we have an equation that is identically true, it is legitimate to differentiate both sides and still get an identity. Differentiating both sides with respect to x, we get:

$$\frac{d(x^2)}{dx} + \frac{d(y^2)}{dx} = 0$$

Simplifying and using the chain rule, we get:

$$2x + 2y\frac{dy}{dx} = 0$$

We thus get:

$$\frac{dy}{dx} = \frac{-x}{y}$$

Notice that with this method, we get $-x/y$, which works in *both* cases. When $y = \sqrt{1-x^2}$, we get $-x/\sqrt{1-x^2}$, and when $y = -\sqrt{1-x^2}$, we get $x/\sqrt{1-x^2}$. The method that we used is called *implicit differentiation*.

So, with a functional description, we need to split into cases. With an implicit description, we can give the general answer in one case, but it involves *both* variables x and y. With these examples in mind, let's consider the parametric situation.

4.2. **Parametric differentiation.** For the curve given parametrically by $x = f(t), y = g(t), t \in D$, where f and g are both differentiable at some point t_0 in the interior of D, the derivative dy/dx is given by:

$$\frac{dy}{dx}|_{t=t_0} = \frac{dy/dt}{dx/dt}|_{t=t_0} = \frac{g'(t)}{f'(t)}|_{t=t_0}$$

In short:

$$\frac{dy}{dx} = \frac{g'(t)}{f'(t)}$$

This is an application of the chain rule, or thinking of the derivative as the relative rate of change. Basically, we are saying that the relative rate of change of y with respect to x at time t_0 is the rate of change of y with respect to time divided by the rate of change of x with respect to time. That makes sense.

Also note that the answer that we get is *purely* a function of t, and x and y do not appear in the description. This is important at many levels. The first is that it takes account of self-intersections. Remember that given a parametric description, the same point on the curve can be repeated at different times, with the curve moving in different directions at these different times. Since the differentiation is with respect to t, *we can thus get different derivative values at different times for the same physical point location.*

Second, if given a point (x, y), we first need to reverse engineer and find the time t that gave rise to the point (x, y) and *then* plug in the formula to find the derivative value.

Let's apply this to the circle $x^2 + y^2 = 1$ using the parameterization $x = \cos t$, $y = \sin t$. We get:

$$\frac{dy}{dx} = \frac{\sin' t}{\cos' t} = \frac{\cos t}{-\sin t} = -\cot t$$

Now, suppose we want to find the derivative value at the point $(1/2, \sqrt{3}/2)$. We first need to determine that this point arises from $t = 2n\pi + \pi/3$. Note that since the curve repeats periodically, the choice of n is irrelevant, so we can choose $t = \pi/3$. Then, the derivative value is $-\cot(\pi/3) = -1/\sqrt{3}$.

You can compare this to the derivative value obtained using implicit differentiation – $-x/y$ – and see that they are in fact the same.

4.3. **Vertical/horizontal tangents and cusps.** Here $x = f(t)$, $y = g(t)$, and $t \in D$. Even if f and g are both differentiable functions everywhere on the domain D, it is still possible for there to be vertical/horizontal tangents and cusps. Below are given some *sufficient* conditions for a horizontal/vertical tangent.

- *Horizontal tangent*: This ocurs when $g'(t) = 0$ and $f'(t) \neq 0$. Note that the horizontal tangent itself could be of many kinds – a local extremum type or a point of inflection type, or something different.
- *Vertical tangent*: This occurs when $f'(t) = 0$ and $g'(t) \neq 0$. Again, there are many possibilities for this.

If both $f'(t)$ and $g'(t)$ are zero, then we need to determine which is more zeroey, and also whether there is a sign change, etc. to determine whether we have a horizontal/vertical tangent/cusp. We refrain from the details here.

4.4. **Repeated differentiation.** Note that the second derivative d^2y/dx^2 is the derivative of the first derivative and the roles of x and y are not symmetric. In particular, the *freshman chain rule* does not work for second derivatives – we need something more sophisticated. Some of you have worked out chain rule equivalents for second derivatives, e.g., we have that $(f \circ g)'' = (f'' \circ g) \cdot (g')^2 + (f' \circ g) \cdot (g'')$ and also that $(f^{-1})'' = -(f'' \circ f^{-1})/(f' \circ f^{-1})^3$. You don't need to memorize these, since they can always be derived from the rules for first derivatives. The key point is that the *naive* chain rules are false. In particular, d^2y/dx^2 is *not* the same as $(d^2y/dt^2)/(d^2x/dt^2)$.

So how do we differentiate twice in practice? We first compute the first derivative in parametric form, and then differentiate that with respect to x, again using parametric differentiation. Specifically, if $x = f(t)$ and $y = g(t)$, we have:

$$\frac{dy}{dx} = \frac{g'(t)}{f'(t)}$$

Thus, we get:

$$\frac{d^2y}{dx^2} = \frac{\frac{d}{dt}\left(\frac{g'(t)}{f'(t)}\right)}{f'(t)}$$

where the denominator is just dx/dt. The same idea can be used to calculate higher derivatives. Note that we could simplify the above as a *general expression* using the quotient rule for differentiation, and thus get another version of the same formula:

$$\frac{d^2y}{dx^2} = \frac{f'(t)g''(t) - g'(t)f''(t)}{(f'(t))^3}$$

It's up to you which form you choose to remember. I recommend that for better conceptual understanding and generalizability, you definitely remember the first version, since the second version can easily be derived from it on the spot.

4.5. **Arc length, speed, and unit speed parameterization.** This is new stuff, in the sense that you have probably not seen this even in the context of functional descriptions. It turns out that the description is easier to give in parametric form anyway. Basically, the arc length is given by:

$$\int \sqrt{(dx)^2 + (dy)^2}$$

More concretely, if $x = f(t)$, $y = g(t)$, then the arc length from time $t = a$ to $t = b$ (with $a < b$) is given by:

$$\int_a^b \sqrt{(dx/dt)^2 + (dy/dt)^2}\, dt$$

or equivalently:

$$\int_a^b \sqrt{(f'(t))^2 + (g'(t))^2}\, dt$$

The *speed* of a curve is defined as the speed with which it is traveling with respect to t, and is hence simply:

$$\sqrt{(dx/dt)^2 + (dy/dt)^2}$$

In other words, if $x = f(t)$ and $y = g(t)$, the speed as a function of time t is:

$$\sqrt{(f'(t))^2 + (g'(t))^2}$$

We say that a particular parameterization of a curve is a *unit speed parameterization* if the speed at any time t is 1. What this means is that the arc length from $t = a$ to $t = b$ is simply $b - a$.

An example of a unit speed parameterization is the parameterization $x = \cos t$, $y = \sin t$ for a circle. Another example is $x = t/\sqrt{2}$, $y = t/\sqrt{2}$ for the line $y = x$.

Note that for the particular case where y is a function of x, the arc length from $x = a$ to $x = b$, with $a < b$, becomes:

$$\int_a^b \sqrt{1 + (dy/dx)^2}\, dx$$

Note on the difficulty of arc length computations. Actually computing the arc length as a number is not easy in most cases because the integration that we need to do is not possible within the world of elementary integrations. In fact, whole new branches of mathematics (such as the theory of elliptic integrals) were invented to be able to perform some of the integrations that arose in arc length computations.

Roughly, here are the kinds of curves for which arc lengths can be computed:

- Straight lines (yeah, and we didn't even need calculus for those)
- Circles (again, no need for calculus here – the radian measure directly defines arc length).
- Parabolas such as $x = y^2$ and $y = x^2$. To compute the arc length, we need to do a trigonometric or hyperbolic trigonometric substitution. The final answer we get involves something like integration $\sec^3\theta$: For instance, consider $y = x^2$. We get:

$$\int \sqrt{1 + (2x)^2}\, dx$$

Put $\theta = \arctan(2x)$, so $x = (1/2)\tan\theta$, and we get:

$$\frac{1}{2}\int \sqrt{1+\tan^2\theta}\sec^2\theta\,d\theta$$

This simplifies to $(1/2)\int \sec^3\theta\,d\theta$, which can be done using integration by parts.

- The graph of the hyperbolic cosine function (catenary): The derivative of cosh is sinh, so we get:

$$\int_a^b \sqrt{1+(\sinh^2 x)}\,dx$$

This becomes:

$$\int_a^b \cosh x\,dx$$

giving $[\sinh x]_a^b = \sinh b - \sinh a$.

While this list above is by no means exhaustive, most of the computable examples you will see will be variations on this theme.

POLAR COORDINATES

MATH 195, SECTION 59 (VIPUL NAIK)

Corresponding material in the book: Section 10.3.

What students should definitely get: The information needed to set up a polar coordinate system, how to go between a physical point and its polar coordinates, conversion back and forth between Cartesian and polar coordinates. Thinking of polar coordinate descriptions as parametric descriptions when viewed in Cartesian coordinates. Using this to compute the slope of the tangent line.

What students should hopefully get: The dimensionality of the plane is 2, so 2 parameters are needed to describe a point in any decent coordinate system. What happens when we fix one coordinate in a Cartesian or polar coordinate system. Why spirals are easy to describe in polar coordinates. How various symmetries in implicit and functional descriptions correspond to geometric symmetries.

Executive summary

Words ...

(1) *Specifying a polar coordinate system*: To specify a polar coordinate system, we need to choose a point (called the *origin* or *pole*), a half-line starting at the point (called the *polar axis* or *reference line*) and an orientation of the plane (chosen counter-clockwise in the usual depictions).

(2) *Finding the polar coordinates of a point and vice versa*: The radial coordinate r is the distance between the point and the pole. The angular coordinate θ is the angle (measured in the counter-clockwise direction) from the polar axis to the line segment from the pole to the point. Note that θ is uniquely defined up to addition of multiples of 2π, and it becomes truly unique if we restrict it to a half-open half-closed interval of length 2π. *The exception is the pole itself, for which θ is undefined* in the sense that any value of θ could be chosen.

(3) *Converting between Cartesian and polar coordinates*: If we take the polar axis as the positive x-axis and the axis at an angle of $+\pi/2$ from it as the positive y-axis, we get a Cartesian coordinate system. The point defined by polar coordinates (r, θ) has Cartesian coordinates $(r\cos\theta, r\sin\theta)$. Conversely, given a point with Cartesian coordinates (x, y) the corresponding polar coordinates are $r = \sqrt{x^2 + y^2}$ and θ is the unique angle (up to addition of multiples of 2π) such that $x = r\cos\theta$, $y = r\sin\theta$.

Actions ...

(1) A functional description of the form $r = F(\theta)$ gives rise to a parametric description in Cartesian coordinates: $x = F(\theta)\cos\theta$ and $y = F(\theta)\sin\theta$. We can do the usual things (like find slopes of tangent lines) using this parametric description. Note that here, θ is typically allowed to vary over all of $\mathbb{R}$ rather than simply being restricted to an interval of length 2π. The slope of the tangnet line in Cartesian terms is given by:

$$\frac{dy}{dx} = \frac{dy/d\theta}{dx/d\theta} = \frac{d(F(\theta)\sin\theta)/d\theta}{d(F(\theta)\cos\theta)/d\theta} = \frac{F'(\theta)\sin\theta + F(\theta)\cos\theta}{F'(\theta)\cos\theta - F(\theta)\sin\theta}$$

(2) An implicit (relational) description in Cartesian coordinates can be converted to a description in polar coordinates by replacing x by $r\cos\theta$ and y by $r\sin\theta$.

(3) An implicit (relational) description in polar coordinates can sometimes be converted to a description in Cartesian coordinates, but with some ambiguity. General idea: replace r by $\sqrt{x^2 + y^2}$, $\cos\theta$ by $x/\sqrt{x^2 + y^2}$, and $\sin\theta$ by $y/\sqrt{x^2 + y^2}$.

1. Review of Cartesian coordinates

1.1. **Descartes' achievement.** We're so used to Cartesian coordinates that we don't often give them much thought. But the idea of using Cartesian coordinates to describe a plane, or to describe three-dimensional

space, was a fundamental breakthrough when it did occur. The idea here being that something as geometric as a plane or space could be represented purely by a tuple of real numbers.

There are some aspects of the *rectangular Cartesian coordinate system* that are worth disaggregating:

(1) The *number* of parameters used is equal to the *dimensionality* of the system being studied. The concept of dimensionality as the *number of free parameters* or *number of degrees of freedom* is probably not something totally new to you.
(2) In order to actually specify a Cartesian coordinate system, we need to choose a pair of orthogonal lines, an ordering of these lines, and a direction to be labeled positive within each line.
(3) Once we choose an origin and an ordered pair of orthogonal directed lines through it, we can use Cartesian coordinates to proceed from an ordered pair of real numbers to a point in the plane, and back. The two procedures are reverses of each other.
(4) Different choices of origin and different choices of direction for the pair of perpendicular lines give different choices of Cartesian coordinate systems. Moving from one to the other *geometrically* corresponds to translations, rotations, and reflections. Moving from one to the other *algebraically* corresponds to some specific algebraic operations on the coordinates.

The most important of these ideas is (1). In some sense, the dimensionality of the plane – namely 2, is far more fundamental than the specific choice of coordinate system used. We will soon construct a new kind of coordinate system called a *polar coordinate system*. This looks very different from a Cartesian coordinate system, but the number of real parameters needed to describe a point remains 2.

1.2. **Fixing one coordinate in the Cartesian system.** Let's consider another aspect of the Cartesian coordinate system. A point in the Cartesian coordinate system is given by a pair of coordinates (x, y). What happens if we fix one coordinate and let the other vary over $\mathbb{R}$. Specifically:

- If we fix a value of x to x_0 and let y vary over $\mathbb{R}$, we get a vertical line given by $x = x_0$. For different choices of x_0, we get parallel lines. Overall, we get a family of parallel lines.
- If we fix a value of y to y_0 and let x vary over $\mathbb{R}$, we get a horizontal line given by $y = y_0$. Overall, we get a family of parallel lines.

2. Polar coordinates

2.1. **The key definitions.** The key idea behind polar coordinates is to specify the *distance* from a specified origin and the *direction* of the line joining the point to the origin.

To create a polar coordinate system, we need the following pieces of data:

- A point selected as the *pole* or *origin* for the coordinate system.
- A half-line (ray) with the point at its endpoint. We will call this the *polar axis* or *reference line*.
- An orientation (counter-clockwise) on the plane.

Every point has two coordinates:

- The *radial coordinate*, denoted r, which is the distance from the origin to that point.
- The *angular coordinate*, denoted θ, which is the angle made between the reference line and the line segment joining the origin to that point, measured counter-clockwise.

Some important notes:

- The radial coordinate is a nonnegative real number.
- The angular coordinate for the pole is not defined. In fact, *any* value of θ could be used for the pole and it would serve to describe the pole. This is a kind of degeneracy or singularity.
- For any other point, the angular coordinate is unique up to multiples of 2π. To make it truly unique, we usually adopt the convention that the angle θ must satisfy $0 \leq \theta < 2\pi$.

The role of coordinates in a polar coordinate system is asymmetric. The r-coordinate is a length coordinate, and the θ-coordinate is a dimensionless angle coordinate. This contrasts with the Cartesian coordinate system where both coordinates play a symmetric role as lengths.

2.2. **What happens if we fix one coordinate?** We note that:

- If we fix a given value of r but allow θ to vary freely, we get a circle centered at the origin. The exceptional case $r = 0$ gives us the single point namely the origin. Thus, we get a family of concentric circles centered at the origin.
- If we fix a given value of θ but allow r to vary freely, we get a ray (half-line) starting at the origin.

2.3. **Polar and Cartesian coordinates: conversion.** Every polar coordinate system has a corresponding Cartesian coordinate system, and vice versa. For a Cartesian coordinate system, we convert to a polar coordinate system by selecting the same origin, taking the reference line as the positive x-axis, and choosing the orientation as counter-clockwise, from the positive x-axis to the positive y-axis.

With this back-and-forth, here are the conversion rules:

- The Cartesian coordinates (x, y) gives $r = \sqrt{x^2 + y^2}$ and θ is the angle $0 \leq \theta < 2\pi$ such that $r\cos\theta = x$ and $r\sin\theta = y$.
- The polar coordinates (r, θ) gives the Cartesian coordinates $x = r\cos\theta$ and $y = r\sin\theta$.

With these conversion rules, we can derive formulas involving polar coordinates from the corresponding formulas involving Cartesian coordinates.

2.4. **Polar coordinates with negative radial coordinate.** A slight variation on the polar coordinate theme is the case where we consider polar coordinates with *negative* radial coordinate value. Here r is the negative of the distance from the pole to the point, and θ is the usual θ shifted by π, so it is the angle made from the polar axis to the half-line in the *opposite* direction to that joining the pole to the point.

3. Descriptions of curves in polar coordinates

3.1. r **as a function of** θ**.** When we give this kind of description, we *usually* allow θ to vary over all real numbers, rather than just restrict it to an interval of length 2π. Typically, the curves for which these descriptions work well are *spirals* starting out at the origin and spiraling outward. For instance, the curve with equation $r = e^\theta$ is a spiral.

One way of thinking of these curves in terms of the *usual Cartesian coordinate system* is in parametric terms – θ is a parameter. If we denote $r = F(\theta)$, then the two coordinate functions x and y are given by $x = F(\theta)\cos\theta$ and $y = F(\theta)\sin\theta$.

3.2. **Case of negative** r **values.** In some cases, the expression $r = F(\theta)$ gives rise to negative r-values for some values of θ. In this case, we interpret these the way we discussed the interpretation of negative r-values.

3.3. **Slope of tangent line in Cartesian terms.** We can determine the slope of the tangent line as follows:

$$\frac{dy}{dx} = \frac{dy/d\theta}{dx/d\theta} = \frac{d(F(\theta)\sin\theta)/d\theta}{d(F(\theta)\cos\theta)/d\theta} = \frac{F'(\theta)\sin\theta + F(\theta)\cos\theta}{F'(\theta)\cos\theta - F(\theta)\sin\theta}$$

Using this, we can determine the equation of the tangent line in a Cartesian coordinate system.

3.4. **Relational description.** In addition to functional descriptions of the form $r = F(\theta)$, we could more generally have implicit (relational) descriptions between r and θ, i.e., descriptions of the form $H(r, \theta) = 0$ where H is a function of two variables. These are more general than functional descriptions.

Note that it is generally possible to convert a relational description in a Cartesian coordinate system to a relational description in a polar coordinate system. Starting with a relational description $G(x, y) = 0$ in a Cartesian coordinate system we get the corresponding polar relation by setting $x = r\cos\theta$ and $y = r\sin\theta$. For instance, the parabola $y = x^2$, in polar coordinates, becomes:

$$r\sin\theta = r^2\cos^2\theta$$

which simplifies to:

$$r(\sin\theta - r\cos^2\theta) = 0$$

It turns out that the r solution is subsumed in the other, so we get:

$$\sin\theta = r\cos^2\theta$$

Similarly if we have the equation:

$$xy = \sin(x^2 + y^2)$$

Then, in polar coordinates, this becomes:

$$r^2\cos\theta\sin\theta = \sin(r^2\cos^2\theta + r^2\sin^2\theta)$$

This simplifies to:

$$r^2\cos\theta\sin\theta = \sin(r^2)$$

Going the other way around, the rule is to plug in $r = \sqrt{x^2+y^2}$, $\cos\theta = x/\sqrt{x^2+y^2}$ and $\sin\theta = y/\sqrt{x^2+y^2}$. Note that directly plugging θ in reverse is tricky, because there is no single shorthand expression for θ – the expression depends on the signs of x and y and gets a little messy.

3.5. **Symmetries for polar equations.** Here are two symmetries of note:

- *Mirror symmetry*: A functional or relational description is symmetric about the polar axis (the reference line) if it satisfies the condition that the condition is satisfied by replacing θ with $-\theta$. More generally, if replacing θ by $2\alpha - \theta$ preserves the condition, then there is symmetry about the line $\theta = \alpha$. Thus, for instance, if replacing θ by $\pi - \theta$ preserves the condition, the curve has mirror symmetry about the y-axis.
- *Half turn symmetry*: If replacing θ by $\theta + \pi$ preserves the condition, then the curve has half turn symmetry about the pole (origin). Similarly, if replacing r by $-r$ preserves the condition, then the curve has half turn symmetry about the pole (origin).

THREE-DIMENSIONAL GEOMETRY

MATH 195, SECTION 59 (VIPUL NAIK)

Corresponding material in the book: Section 12.1

What students should definitely get: The use of three coordinates to describe points in space. Right-hand rule and orientation issues. Octants and coordinate planes. Distance formula. Equation of sphere.

What students should hopefully get: Three-dimensionality. Relation between describing equations and dimensionality (dimensionality of surfaces, curves). The special case of the linear situation.

Executive summary

Words ...

(1) Three-dimensional space is coordinatized using a Cartesian coordinate system by selecting three mutually perpendicular axes passing through a point called the origin: the x-axis, y-axis, and z-axis. These satisfy the right-hand rule. The coordinates of a point are written as a 3-tuple (x, y, z).

(2) There are $2^3 = 8$ octants based on the signs of each of the coordinates. There are three coordinate planes, each corresponding to the remaining coordinate being zero (the xy-plane corresponds to $z = 0$, etc.). There are three axes, each corresponding to the other two coordinates being zero (e.g., the x-axis corresponds to $y = z = 0$).

(3) The distance formula between points with coordinates (x_1, y_1, z_1) and (x_2, y_2, z_2) is:

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

This is similar to the formula in two dimensions and the squares and square root arises from the Pythagorean theorem.

(4) The equation of a sphere with center having coordinates (h, k, l) and radius r is $(x-h)^2 + (y-k)^2 + (z-l)^2 = r^2$. Given an equation, we can try completing the square to see if it fits the model for the equation of a sphere.

1. Quick introduction

1.1. **What's up with the number three?** Why is three-dimensional space considered *three*-dimensional? Because, in order to describe a point in this space, it requires three coordinates, or three independent pieces of information. Remember, dimensionality is the *number of degrees of freedom*, or the *number of freely varying parameters*.

For now, the focus is on describing the behavior of a *Cartesian* three-dimensional coordinate system. However, it is important to note that *any* successful coordinate system to describe three-dimensional space must use three freely varying parameters. There does exist (though we will not be talking about it right now) a *polar* coordinate system in three dimensions.

1.2. **Three coordinates: x, y, and z.** The three-dimensional Cartesian coordinate system involves the following: choice of a point called the *origin*, and three mutually perpendicular directed lines called the x-axis, y-axis, and z-axis respectively. To specify a point with three coordinates, we move parallel to the x-axis by the value of the x-coordinate, then parallel to the y-axis by the value of the y-coordinate, then parallel to the z-axis by the value of the z-coordinate.

The way to denote/describe a point with x-coordate x_0, y-coordinate y_0, and z-coordinate z_0 is as (x_0, y_0, z_0). Or, more generally, we just write (x, y, z).

1.3. **Orientation issues.** Recall that, for the two-dimensional plane, we adopt an orientation convention, namely, the convention that the shorter angle from the positive x-axis to the positive y-axis (measured $\pi/2$) be counter-clockwise. In a three-dimesional system, we must adopt a similar orientation convention. This convention is somewhat artificial and in fact if you were doing mathematics properly then such a convention would be meaningless. But since we're trying to do mathematics the shortcut way, we will need to introduce this convention.

The convention is called the *right-hand rule*. This says that if you grip the z-axis with your right hand and curl the fingers of your right hand so as to move from the x-axis to the y-axis, and make your thumb point along the z-axis away from the fingers, then it points along the positive z-axis. The book has a nice picture of this, which I will not reproduce here.

We will thus deal with *right-handed coordinate systems*. The mirror reflection of a right-hand coordinate system is a left-handed coordinate system. There isn't really any difference between left-handed and right-handed coordinate systems as far as their geometry is concerned. However, they are different in the sense that no amount of rotation can turn a right-handed coordinate system into a left-handed coordinate system.

1.4. **Octants, axes, planes.** Since there are three coordinates, there are a total of $2^3 = 8$ possible sign combinations for the coordinates. These eight sign combination give rise to eight chambers of three-dimensional space, and these chambers are called *octants*. They are the three-dimensional analogues of the *quadrants* in two-dimensional space, of which there are $2^2 = 4$.

The eight octants are:

- Positive x, positive y, positive z
- Positive x, positive y, negative z
- Positive x, negative y, positive z
- Positive x, negative y, negative z
- Negative x, positive y, positive z
- Negative x, positive y, negative z
- Negative x, negative y, positive z
- Negative x, negative y, negative z

There are also cases where one or more of the coordinates takes the value 0. These are discussed below:

- The x-coordinate takes the value 0: The yz-plane.
- The y-coordinate takes the value 0: The xz-plane.
- The z-coordinate takes the value 0: The xy-plane.
- Both the y and z coordinates take the value 0: The x-axis.
- Both the x and z coordinates take the value 0: The y-axis.
- Both the y and z coordinates take the value 0: The z-axis.
- All three coordinates are zero: We get the origin.

1.5. **The distance formula.** The distance between the points (x_1, y_1, z_1) and (x_2, y_2, z_2) is given by the formula:

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

Note that this is very similar to the two-dimensional distance formula, and has a similar justification. Please note that the *appearance of squares and square roots is independent of dimensionality.* It has to do with the Pythagorean theorem. So when we are working in three dimensions, we still deal with squares and square roots, not with cubes and cube roots.

2. Curves and surfaces

2.1. **The arithmetic of dimensionality of curves and surfaces.** A surface is something two-dimensional, a curve is something one-dimensional, and a point is something zero-dimensional.

Every condition, relation, or equation creates a constraint and reduces the dimensionality of the set of possibilities by one. In particular:

- A single equation reduces the dimensionality of the solution space by one. Since $3 - 1 = 2$, the solution set to a single equation in three dimensions is something two-dimensional, such as a plane or surface, or a union of finitely many planes and surfaces.
- A pair of two equations reduces the dimensionality of the solution space by two. Since $3 - 2 = 1$, the solution set to a pair of such equations is something one-dimensional, such as a line or curve, or a union of finitely many lines and curves.
- A triple of equations reduces the dimensionality of the solution space by three. Since $3 - 3 = 0$, the solution set to a tripe of such equations is something zero-dimensional, such as a point or a finite collection of points.

2.2. **Linear equations and systems.** We will talk about these in considerably more detail later on, but here we simply note how linear equations fit into the model above.

- The solution set to a *linear equation* in x, y, and z is a plane, which is a flat example of a surface.

  In particular, the solution set to an equation of the form $x = x_0$ is a plane parallel to the yz-plane. Similarly, $y = y_0$ gives a plane parallel to the xz-plane, and $z = z_0$ gives a plane parallel to the xy-plane.
- The solution set to a *consistent but inequivalent pair of linear equations* in x, y, and z is a line which is a flat example of a curve.

  In particular, the solution set to a pair of equations of the form $x = x_0$, $y = y_0$ is a line parallel to the z-axis. Similarly with the coordinate roles interchanged.
- The solution set to a system of three consistent but independent linear equations in x, y and z is a single point.

2.3. **Ignoring one coordinate.** Suppose we have a relation $R(x, y) = 0$. The solution space to this in *three* dimensions is the surface obtained by taking the solution set in the xy-plane and then translating it along the z-direction to cover every possible z-coordinate. We can think of it as something cylinder-like.

For instance, the relation $x^2 + y^2 = 25$ gives a circle of radius 5 centered at the origin when viewed purely in the xy-plane. The solution set overall is a cylinder stretching infinitely in both directions, whose axis is the z-axis and whose cross sections are these circles.

Similar remarks hold for a relation purely in terms of y and z or a relation purely in terms of x and z.

2.4. **Equation of a sphere.** The equation of a sphere whose center has coordinates (h, k, l) and whose radius is r is given by:

$$(x - h)^2 + (y - k)^2 + (z - l)^2 = r^2$$

Given an equation, we can try doing things like completing the square and see if, after we do that, we get the equation of a sphere.

VECTOR STUFF

MATH 195, SECTION 59 (VIPUL NAIK)

Corresponding material in the book: Sections 12.2, 12.3, 12.4.

What students should definitely get: Vector as a n-tuple, special attention to case $n = 3$, geometric interpretation of vectors, free and localized vectors, computation and properties of dot product, computation and properties of length, computation and properties of cross product, computation and properties of scalar triple product. Geometric applications to computing angles, areas of triangles and parallelograms, finding a vector orthogonal to two vectors, testing coplanarity and orthogonality.

What students should hopefully get: The idea of correlation, cosine as measuring closeness, the significance of n-dimensional vectors, the notion of vector operations and meaningfulness depending on the kind of scale from which the coordinates are drawn, the notions of ordinal scale, difference scale, and ratio scale.

Executive summary

0.1. **n-dimensional generality.** Words ...

(1) A vector is an ordered n-tuple of real numbers (or quantities measured using real numbers). The space of such n-tuples is a n-dimensional vector space over the real numbers. Vectors can be used to store tuples of prices, probabilities, and other kinds of quantities.

(2) There is a zero vector. We can add vectors and we can multiply a vector by a scalar. Note that these operations may or may not have an actual meaning based on the thing we are storing using the vector.

(3) We can take the dot product $v \cdot w$ of two vectors v and w in n-dimensional space. if $v = \langle v_1, v_2, \ldots, v_n \rangle$ and $w = \langle w_1, w_2, \ldots, w_n \rangle$, then $v \cdot w = \sum_{i=1}^{n} v_i w_i$. The dot product is a real number (though if we put units to the coordinates of the vector, it gets corresponding squared units; if we use different units for the different vectors, the units are the product units).

(4) The length or norm of a vector v, denoted $|v|$, is defined as $\sqrt{v \cdot v}$. It is a nonnegative real number.

(5) The correlation between two vectors v and w is defined as $(v \cdot w)/(|v||w|)$. It is in $[-1, 1]$. (For geometric interpretation, see the three-dimensional case).

(6) *Properties of the dot product*: The dot product is symmetric, the dot product of any vector with the zero vector is 0, the dot product is additive (distributive) in each coordinate and scalars can be pulled out.

(7) *Properties of length*: The only vector with length zero is the zero vector, all other vectors have positive length. The length of λv is $|\lambda|$ times the length of v. We also have $|v + w| \leq |v| + |w|$ for any vectors v and w, with equality occurring if either is a positive scalar multiple of the other or one of them is the zero vector.

0.2. **Three-dimensional geometry.** Words ...

(1) We can identify points in three-dimensional space with three-dimensional vector as follows: the vector corresponding to a point (x, y, z) is the vector $\langle x, y, z \rangle$. Physically, this can be thought of as a directed line segment or arrow from the origin to the point (x, y, z).

(2) We can also consider vectors starting at any point in three-dimensional space and ending at any point. The corresponding vector can be obtained by subtracting the coordinates of the points. The vector from point P to point Q is denoted $\overrightarrow{PQ}$.

(3) There are unit vectors $\mathbf{i} = \langle 1, 0, 0 \rangle$, $\mathbf{j} = \langle 0, 1, 0 \rangle$, and $\mathbf{k} = \langle 0, 0, 1 \rangle$. These are thus the vectors of length 1 along the positive x, y, and z directions respectively. A vector $\langle x, y, z \rangle$ can be written as $x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$.

(4) Vectors can be added geometrically using the *parallelogram law*. This procedure gives the same answer as the usual coordinate-wise addition.
(5) Scalar multiplication also has a geometric interpretation – the length gets scaled by the scalar multiple, and the direction remains the same or is reversed depending on the scalar's sign.
(6) For vectors v and w, we have $v \cdot w = |v||w| \cos\theta$ where θ is the angle between v and w. We can use this procedure to find the angle between two vectors. The correlation between the vectors is thus $\cos\theta$. We can interpret this specifically for $\theta = 0$, θ an acute angle, $\theta = \pi/2$, θ an obtuse angle, and $\theta = \pi$ (see the table in the lecture notes).
(7) We can define the vector cross product $v \times w$ using a matrix determinant. Equivalently, if $v = \langle v_1, v_2, v_3\rangle$ and $w = \langle w_1, w_2, w_3\rangle$, then $v \times w = \langle v_2w_3 - v_3w_2, v_3w_1 - v_1w_3, v_1w_2 - v_2w_1\rangle$. *This is a specifically three-dimensional construct.*
(8) The cross product has the property that cross product of any two qcollinear vectors is zero, cross product of any vector with the zero vector is zero, the cross product is skew-symmetric, distributive in each variable, and allows scalars to be pulled out. It is not associative in general. There is an identity relating cross product and dot product: $a \times (b \times c) = (a \cdot c)b - (a \cdot b)c$. Also, the cross product satisfies the relation:

$$a \times (b \times c) + b \times (c \times a) + c \times (a \times b) = 0$$

(9) The cross product of a and b satisfies $|a \times b| = |a||b| \sin\theta$ where θ is the angle between a and b, and further, the cross product vector is perpendicular to both a and b, and its direction is given by the right hand rule.
(10) There is a scalar triple product. The scalar triple product of vectors a, b, and c is defined as the number $a \cdot (b \times c)$. It can also be viewed as the determinant of a matrix whose rows are the coordinates of a, b, and c respectively. The scalar triple product is preserved under cyclic permutations of the input vectors and gets negated under flipping two of the input vectors. It is linear in each input variable (i.e., distributive and pulls out scalars). The scalar triple product is zero if and only if the three input vectors can all be made to lie in the same plane.

Actions ...

(1) Vector and scalar projections: Given vectors a and b, the *vector projection* of b onto a, denoted $\text{proj}_a\, b$, is given by the vector $\frac{a \cdot b}{|a|^2} a$. The scalar projection or component of b along a, denoted $\text{comp}_a\, b$, is given by $\frac{a \cdot b}{|a|}$. The vector projection is what we obtain by taking the vector from the origin to the foot of the perpendicular from the head of b to the line of a. The scalar projection is the *directed* length of this vector, measured positive in the direction of a.
(2) Finding the angle between vectors: This is done using the dot product. The angle between vectors v and w is $\arccos((v \cdot w)/|v||w|)$.
(3) Finding the area of a triangle or a parallelogram: We first find two adjacent sides as vectors both with the same starting vertex (by taking the differences of coordinates of endpoints). For the parallelogram, we take the length of the cross product of these two vectors. For the triangle, we take *half* the length.
(4) Finding the volume of a parallelopiped: We find three sides as vectors, all with the same starting vertex. Then we take the scalar triple product of these sides.
(5) Finding a vector orthogonal to two given vectors: Simply take the cross product if they are linearly independent. Otherwise, just pick anything that dots with one of them to zero.
(6) Testing orthogonality: We check whether the dot product is zero.
(7) Testing coplanarity of points: We take one point as the basepoint, compute difference vectors to it from the other three points. We then take the scalar triple product of these three vectors. If we get zero, then the four points are coplanar, otherwise they are not.

1. Vectors – what's the big deal?

In this lecture (and the corresponding notes) I will take a somewhat different approach to vector spaces from what is to be found in the book. The reasons are multi-fold. First, the highly geometric approach to vectors that the book takes, although standard fare, is somewhat misleading for applications of vector space

theory to the social sciences. The approach taken in the book *is* right for applying vector spaces to the physical sciences, particularly physics. But the intended as well as actual audience for this course comprises economics majors, so I've determined the focus accordingly.

1.1. **Vector as a tuple of data, and vector-valued functions.** A *real number* can be used to store a single quantitative variable. What happens when we need to store more than one variable? In this case, we need to store a bunch of real numbers. One way of storing a bunch of data is as a tuple, or a system of comma-separated values. This kind of bunch of data can be thought of as a vector.

For instance, let's think of the share prices of five different companies as a function of time. At any given time, each share price is a number (measured in fixed units). The collection of five share prices is a tuple of 5 numbers, which can be represented as comma-separated values. If the prices are p_1, p_2, p_3, p_4, and p_5, then the tuple in question is $\langle p_1, p_2, p_3, p_4, p_5 \rangle$.

Since the prices vary *with time*, each of the prices is actually a function of time, so what we really have is a *vector-valued function* whose coordinates are p_1, p_2, p_3, p_4, and p_5, given by:

$$t \mapsto \langle p_1(t), p_2(t), p_3(t), p_4(t), p_5(t) \rangle$$

We say that this is a vector-valued function in a five-dimensional vector space.

Vector-valued functions are a new perspective on the parametric descriptions we saw a little while ago. In fact, these parametric descriptions of curves in the plane can be thought of as vector-valued functions in two dimensions.

Here are some other examples of vectors and vector-valued functions:

- To keep track of prices, a bundle of commodities is chosen and unit prices for these commodities are tracked as a function of time. If 15 commodities are chosen, each of the unit prices is a function of time. The collection of all unit prices at a given time is a vector in 15 dimensions and the function that sends a time to the collection of these 15 unit prices is a vector-valued function.
- In a given market, the quantity demanded and quantity supplied are both functions of price, i.e., at a given price, there is a certain quantity demanded and a certain quantity supplied for a commodity. The pair of these quantities thus gives a 2-dimensional vector-valued function of price.
- The scores of 20 students on a given test are stored as a 20-dimensional tuple, or a vector in 20-dimensional space.

1.2. **The theory of measurement: a little aside.** When we talk of scales for measuring stuff, there are three different kinds of scales we typically use:

- *Ordinal scales* or *comparison scales* where we can only compare things as greater or less – the quantification of *how* different two things are doesn't really make sense.
- *Difference scales* where we can measure how different two things are – but there need not be an absolute zero quantity.
- *Ratio scales* where there is an absolute zero and hence we can perform operations like scalar multiplication, addition, subtraction, ratios.

Ordinal scales or comparison scales are the weakest, in the sense that they reflect very little understanding of what is being measured. Difference scales reflect a better local understanding but a poor global understanding. Ratio scales are pretty good.

Consider, for instance, the example of temperature. In the olden days, the human understanding of temperature was simply as a means of comparison – some things were hotter than others, some things were colder than others. We didn't have an idea how to address the question "Is the hotness difference between A and B more than the hotness difference between C and D?" Then, temperature scales such as the Celsius and Fahrenheit scale were invented/discovered, and it was now possible to address the question of *difference in temperature* precisely. One could say that the temperature difference between 20 degrees Celsius and 30 degrees Celsius is more than the temperature difference between 40 degrees Celsius and 45 degrees Celsius. Note that the conclusions are the same whether we measure temperatures in degrees Celsius or degrees Fahrenheit.

With a difference scale, we can also talk of *average* temperatures. The average of the temperatures 20 degrees Celsius and 30 degrees Celsius is 25 degrees Celsius. Note that we arrive at the same average whether we measure temperatures in Celsius or Fahrenheit.

However, what does *not* make sense is the *addition* of temperatures. We cannot add 20 degrees Celsius and 30 degrees Celsius to get 50 degrees Celsius. We can, but it does not make sense. Also, this addition does not give the same answers in Celsius and Fahrenheit. Similarly, it does not make sense to multiple a temperature of 20 degrees Celsius by 1.7.

The reason these fail is that in neither Celsius nor Fahrenheit is the zero temperature an *absolute zero*.

Prior to the discovery of absolute zero by Kelvin and the creation of the Kelvin scale for temperature, a difference scale was the best one could do. Now that the Kelvin scale exists, and we have an absolute zero, we can measure temperatures in degrees Kelvin. Measured in these units, temperatures can be added and the ratios of temperatures can be taken.

1.3. **Return to vectors: vector operations coordinate-wise.** Recall that a vector is just a bunch of scalars stored side by side. Now, we can define some vector operations, but *those operations may not have any meaning for some interpretations of vectors*. The basic operations are:

- The *zero vector*, denoted 0, is defined as the vector all of whose coordinates are zero.
- We can multiply a vector by a scalar (i.e., a real number). Given a real number λ and a vector $v = \langle v_1, v_2, \ldots, v_n \rangle$, we achieve this multiplication by multiplying λ by each of the coordinates v_i, to get the new vector $\langle \lambda v_1, \lambda v_2, \ldots, \lambda v_n \rangle$.
- We can add two vectors in the same vector space $\langle v_1, v_2, \ldots, v_n \rangle$ added to $\langle w_1, w_2, \ldots, w_n \rangle$ is $\langle v_1 + w_1, v_2 + w_2, \ldots, v_n + w_n \rangle$.

Both addition and scalar multiplication are carried out *coordinate-wise*. In particular, we can use these to compute the *difference* of two vectors (by subtracting the values in each coordinate) and the *average* of two or more vectors (by averaging the values in each coordinate). Note that it is imperative that the vectors live in the same space in order for us to perform the operations.

Further, note that although we can perform these operations arithmetically, whether they make sense depends on whether the quantities being measured are on the right kind of scale. If the quantities are merely on an ordinal scale, and the quantification is purely for ease of understanding, then *none of the vector operations have any meaning*. If the quantities are on a difference scale, then taking differences and averages makes sense but taking sums and scalar multiples does not. If the quantities are on a ratio scale, then all the vector operations make sense.

1.4. **Dot product or scalar product.** Given two vectors v and w in the same vector space, say $v = \langle v_1, v_2, \ldots, v_n \rangle$ and $w = \langle w_1, w_2, \ldots, w_n \rangle$, their *dot product* is defined as:

$$\sum_{i=1}^{n} v_i w_i$$

or, in long form, as:

$$v_1 w_1 + v_2 w_2 + \cdots + v_n w_n$$

As such, this number is a *scalar*, not a vector, because it is a single number obtained as a sum of products of real numbers.

However, in the real world interpretation where the coordinates of the vector represent quantities that have units attached to them, we can interpret this dot product too as having a unit. Note that:

- Quantities of different kinds cannot be added. Thus, for a dot product to make sense, it is necessary that all coordinates of the vector be measuring the same kind of quantity in the same units. In other words, all coordinates of v should use the same units as each other, and all coordinates of w should use the same units as each other.
- It is not necessary that the vectors v and w be measured in the same units as each other. The units for the dor product are the units for v times the units for w. For instance, if the units for v are meters per second, and the units for w are seconds, the units for $v \cdot w$ are meters. In case both v and w are measured in the same units, the units for $v \cdot w$ are the corresponding squared units.

Here's one (somewhat contrived?) application to *total money value* calculations (as come up when calculating GDP). The *money value* of a commodity is the product of its unit price p and its quantity q. If we have a bunch of commodities and the price vector (i.e., the collection of unit prices) is given by $p = \langle p_1, p_2, \ldots, p_n \rangle$ while the quantity vector (i.e., the collection of quantities) is given by $q = \langle q_1, q_2, \ldots, q_n \rangle$, then the total money value is the dot product $p \cdot q$ of the two vectors. Explicitly, it is the sum $\sum_{i=1}^{n} p_i q_i$.

Here's another example. Suppose you make a gamble which has n possible outcomes. The pay-off to you for outcomes i, with $1 \leq i \leq n$, is m_i. The probability of outcome i is p_i. The *expected value of pay-off* for this gamble is $\sum_{i=1}^{n} p_i m_i$. We can think of this as a dot product of the *probabilities vector* $\langle p_1, p_2, \ldots, p_n \rangle$ and the *pay-offs vector* $\langle m_1, m_2, \ldots, m_n \rangle$.

Another example is calorie consumption: your overall calorie consumption in a day is the dot product of the vector giving the caloric contents of the foods you eat, with the vector describing how much of each food you eat.

1.5. **Length of a vector and correlation between vectors.** The *length* or *norm* of a vector $v = \langle v_1, v_2, \ldots, v_n \rangle$ is denoted $|v|$ and defined as:

$$|v| = \sqrt{v \cdot v} = \sqrt{\sum_{i=1}^{n} v_i^2}$$

Note that it makes sense to take the square root of the expression because a sum of squares is nonnegative. Also note that, by definition, $|v|$ is a nonnegative real number.

When we study the geometric interpretation of vectors, we will see that the length or norm is the same as the length of the corresponding geometric notion of vector when we are in the familiar three-dimensional setting.

The *correlation* between two vectors v and w is defined as:

$$\frac{v \cdot w}{|v||w|}$$

This correlation basically measures how similar the *directions* of the vectors v and w are to each other. In particular, this correlation is in the interval $[-1, 1]$, where -1 occurs if v and w are *negative* scalar multiples of each other, and 1 occurs if v and w are *positive* scalar multiples of each other. We will see the relation with some geometry and trigonometry a little later.

1.6. **Why square up to measure the length?** There is a real-world reason for defining the norm or length by taking squares, adding, and then taking the square root, rather than just adding up the absolute values. The reason is that we want each coordinate to be weighed *by itself*, so that larger magnitude coordinates get a higher weighting. An example is when considering population densities. If we average the population density between an area with high density and an area with zero density, the effective population density comes out to be a lot less than that of the area with high density. However, the population density *lived/experienced* by people is only the density in the high density area, not in the zero density area. Ideally, we want to weight the zero density area by zero, because nobody actually lives there. More generally, we want to weight low density areas by their low densities or low populations, reflecting the fact that since fewer people actually live there, they should count for less in the average. The use of squares in the norm concept captures this idea.

1.7. **Properties of the dot product and length.** Here are some key properties of the dot product:

- The dot product of any vector with the zero vector is the number zero: $a \cdot 0 = 0 \cdot a = 0$ for all a. Note that the right-most 0 is the *scalar* 0 and the other two 0s are vector 0s.
- The dot product is *symmetric*, i.e., $a \cdot b = b \cdot a$ for all vectors a and b.
- The dot product is *distributive* in both coordinates, i.e., $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$ and $(a + b) \cdot c = (a \cdot c) + (b \cdot c)$.
- Scalars can be pulled out of the dot product: $(\lambda a) \cdot b = \lambda(a \cdot b)$.

Note: The last two properties *together* constitute what is called *linearity* in each variable, and hence, *bilinearity*. This may make more sense to you if/when you become familiar with linear algebra.

Here are the key properties of length:

- The length of a vector is 0 if and only if the vector is the zero vector.
- For any vector a and scalar λ, the length of λa is $|\lambda|$ times the length of a.
- *Triangle inequality*: For any vectors a and b, we have $|a + b| \leq |a| + |b|$ with equality holding if and only if a and b are positive scalar multiples of each other (i.e., the angle between them is 0) or one of the vectors a and b is 0.

2. Vectors in the physical world: in three dimensions

2.1. **Vectors as arrows or directed line segments.** We now shift attention to the notion of vectors in the physical world. This concept is not important as an end in itself, but some of these ideas are necessary for later applications to vector integration and differentiation, as well as developing the theory of vector-valued functions.

Consider a three-dimensional space equipped with a coordinate system. Geometrically, a *vector based at the origin* in the space can be thought of as an arrow (a directed line segment) with tail at the origin and head at some point in the space. Thus, there is a correspondence:

Points in three-dimensional space $\leftrightarrow$ Vectors based at the origin in three-dimensional space

More generally, we can talk of a vector *based at any point*. This is just a directed line segment starting at one point and ending at another (possibly the same) point. Given a vector based at a point, there is a unique vector based at the origin that is parallel to it and has the same length and same direction. Basically, this is obtained by taking the original vector and translating it till its tail becomes the origin.

Consequently, there is a notion of a *free vector*. A free vector can be thought of as an equivalence class of vectors, with different starting points, all of which have the same length and direction. Given any free vector, there is a unique way of representing it using a vector based at the origin.

Some notation: the vector with tail at a point A and head at a point B is denoted $\overline{AB}$, and we sometimes put an arrow in the left-to-right direction.

2.2. **Vectors and coordinates.** In a coordinate system, the vector from the origin to the point with coordinates (x, y, z) is denoted as the vector $\langle x, y, z \rangle$. The operations of vector addition, scalar multiplication, and dot product, are all done in the context of this description.

Before describing these, I will briefly note the **i**,**j**,**k** notation. We define the three unit vectors **i**, **j**, and **k** as follows:

- **i** is the unit vector $\langle 1, 0, 0 \rangle$, i.e., it is a vector of length 1 along the positive x-axis.
- **j** is the unit vector $\langle 0, 1, 0 \rangle$, i.e., it is a vector of length 1 along the positive y-axis.
- **k** is the unit vector $\langle 0, 0, 1 \rangle$, i.e., it is a vector of length 1 along the positive z-axis.

We can thus write:

$$\langle x, y, z \rangle = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$$

In other words, the vector $\langle x, y, z \rangle$ is obtained by traveling x in the x-axis direction, y in the y-axis direction, z in the z-axis direction.

2.3. **Vector operations: scalar multiplication and addition.** We can do a bunch of vector operations, as discussed earlier, and relate it to geometric operations:

- *Scalar multiplication*: Given a vector $\langle x, y, z \rangle$ and a scalar λ, we get the vector $\langle \lambda x, \lambda y, \lambda z \rangle$. Geometrically, this new vector is parallel to the old vector, its length is $|\lambda|$ times the length of the old vector, and its direction (forward/backward) is the same as or reverse of the old vector depending on whether $\lambda > 0$ or $\lambda < 0$.
- *Addition*: Given a vector $v_1 = \langle x_1, y_1, z_1 \rangle$ and a vector $v_2 = \langle x_2, y_2, z_2 \rangle$, the sum of the vectors is $\langle x_1 + x_2, y_1 + y_2, z_1 + z_2 \rangle$. Geometrically, this can be obtained as follows: we take the vector v_2, treat it as a free vector, and translate it so that its tail is at the head of v_1. Now, we make a vector from the origin to the *head* of this translated v_2. This new vector is $v_1 + v_2$. Note that we could interchange the roles of v_1 and v_2 in this process, and get the same answer. When we make the whole picture, we get a parallelogram, so this method sometimes goes under the name of the *parallelogram law of addition*.

2.4. **Terminological note.** We say that two vectors are *collinear* or *parallel* or *linearly dependent* if either is a scalar multiple of the other.

The *zero vector* is the vector all of whose coordinates are zero. It corresponds to the arrow from the origin to itself.

2.5. **Bold face conventions.** Note that the book uses a bold face letter convention for vectors, e.g., the vector with name v will be written as $\mathbf{v}$ instead of just v.

I will not be bothered too much about this convention, and it's anyway hard to make bold face conventions clear when writing things with pen, pencil, or chalk. So you don't have to bother much about it.

2.6. **Vector operations: dot product.** Recall the general definition of dot product, $\langle v_1, v_2, \ldots, v_n \rangle \cdot \langle w_1, w_2, \ldots, w_n \rangle$ is given by $\sum_{i=1}^{n} v_i w_i$.

In particular, if $v = \langle v_1, v_2, v_3 \rangle$ and $w = \langle w_1, w_2, w_3 \rangle$, we get:

$$v \cdot w = v_1 w_1 + v_2 w_2 + v_3 w_3$$

We also define:

$$|v| = \sqrt{v \cdot v} = \sqrt{v_1^2 + v_2^2 + v_3^2}$$

This is the *length* of the vector v and is the same as the length of the actual line segment corresponding to the vector.

It turns out that if v and w are vectors, then:

$$v \cdot w = |v||w| \cos\theta$$

where θ is the angle between v and w.

2.7. **Thinking about dot product: what does it measure?** As noted above:

$$v \cdot w = |v||w| \cos\theta$$

This can be rearranged to give:

$$\cos\theta = \frac{v \cdot w}{|v||w|}$$

In other words, taking θ as the angle between 0 and π:

$$\theta = \arccos\left[\frac{v \cdot w}{|v||w|}\right]$$

The quantity $\cos\theta$, which is also equal to the right side, is the *correlation* between the vectors v and w, or, in other words, it is the degree of similarity in the direction of the vectors v and w. This is a number in the interval $[-1, 1]$. Let's think more clearly about what this means:

θ	$\cos\theta$	Interpretation on the relation between v and w
0	1	w is a *positive* scalar multiple of v, i.e., they are in the same direction
In $(0, \pi/2)$	In $(0, 1)$	v and w are positively correlated, form an acute angle
$\pi/2$	0	v and w are *uncorrelated*. They are perendicular or orthogonal.
In $(\pi/2, \pi)$	In $(-1, 0)$	v and w are negatively correlated, form an obtuse angle.
π	-1	w is a negative scalar multiple of v, i.e., they are in exactly opposite directions.

2.8. **Rethinking trigonometry: components and projections.** The preceding discussion can give you a deeper perspective on trigonometry. The cos function measures the *closeness* or *correlation* or *degree of agreement* between two things separated by the given angle. The sin function measures how big the *rest of the stuff is*. The relation:

$$\cos^2\theta + \sin^2\theta = 1$$

basically says that if you add up the parts that agree and the rest of the stuff you get 100%.

Let's discuss a related notion that helps make this clearer. Given vectors a and b, both based at the origin, the *vector projection* of b onto a is the vector starting at the origin and ending at the foot of the perpendicular from the head of b to the line of a. The *component* or *scalar projection* of b along a is the length of this vector projection measured in the direction of a (so it could be negative). Specifically:

- The vector projection of b onto a, denoted $\text{proj}_a\, b$, is $\frac{a \cdot b}{|a|^2} a$.
- The scalar projection of b onto a, denoted $\text{comp}_a\, b$, is $\frac{a \cdot b}{|a|}$.

2.9. **Sidenote: orthogonal as unrelated.** People who are mathematically and statistically literate have an interesting new adjective to add to their vocabulary for daily use: "orthogonal" in the sense of being unrelated or irrelevant. This usage stems from the above discussion of dot products and correlation. However, this usage is not common among people without a math/stats background, and it is a way for you to show off your mathematical/statistical erudition. An amusing exchange was recorded in Supreme Court proceedings in the United States in 2006:

`http://www.volokh.com/2010/01/11/orthogonal-ooh/`

Short version: Supreme Court justices were flummoxed when a law professor used the word "orthogonal" in the sense of unrelated.

3. Vagaries of three dimensions: cross product and scalar triple product

The constructions we will discuss now are specific to three dimensions. Although they do have some analogues or generalizations in higher dimensions, the analogues are not obvious or intuitive and require considerable mathematical machinery.

3.1. **The cross product.** The cross product of two vectors $a = \langle a_1, a_2, a_3 \rangle$ and $b = \langle b_1, b_2, b_3 \rangle$ is defined as follows:

$$a \times b = \langle a_2 b_3 - a_3 b_2, a_3 b_1 - a_1 b_3, a_1 b_2 - a_2 b_1 \rangle$$

The formula looks a bit cryptic. Here is the idea: we arrange the coordinates in the cyclic ordering $1 \to 2 \to 3 \to 1$. A particular coordinate in $a \times b$ is then the next coordinate of a times the next to next coordinate of b minus the next coordinate of b minus the next-to-next coordinate of a. We can also write this as a *determinant* as follows:

$$\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}$$

For those of you who have seen determinants, this should make sense. Basically, we evaluate the determinant by expansion using the first row. Recall that $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$, are the unit vectors in the x, y, and z directions, so their coefficients are the coordinates of the vector we finally want.

3.2. **Properties of the cross product.** Here are some key properties of the cross product:

- The cross product of any vector with the zero vector is the zero vector: $a \times 0 = 0 \times a = 0$ for any vector a.
- The cross product is *anti-symmetric* or *skew-symmetric*: For any vectors a and b, $a \times b = -(b \times a)$.
- The cross product of two vectors that are scalar multiples of each other is zero: $a \times b = 0$ if $b = \lambda a$ for some scalar λ.
- The cross product is *distributive* in both vectors: $a \times (b + c) = (a \times b) + (a \times c)$ and $(a + b) \times c = (a \times c) + (b \times c)$.
- Scalars can be pulled out of cross product: $(\lambda a) \times b = \lambda(a \times b)$ and $a \times (\lambda b) = \lambda(a \times b)$.
- The cross product is *not associative*: $a \times (b \times c) \neq (a \times b) \times c$.
- We have the following relation: $a \times (b \times c) = (a \cdot c)b - (a \cdot b)c$.
- The cross product $a \times b$ is orthogonal to a and to b. In particular, we get that $a \cdot (a \times b) = 0$ and $b \cdot (a \times b) = 0$.

- We have the following identity involving the cross product:

$$(a \times (b \times c)) + (b \times (c \times a)) + (c \times (a \times b)) = 0$$

Also:

$$((a \times b) \times c) + ((b \times c) \times a) + ((c \times a) \times b) = 0$$

3.3. **Geometric interpretation of the cross product.** The cross product of non-collinear vectors a and b is a vector perpendicular to both a and b (and hence, perpendicular to the plane spanned by a and b). Its direction is given by the right hand rule: curl the fingers of the right hand from a to b, make the thumb point away from the fingers, then the thumb points in the direction of $a \times b$. Its magnitude is given by the formula:

$$|a \times b| = |a||b| \sin \theta$$

where θ is the angle between a and b.

In particular, we note some important relations between the unit vectors $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$:

$$\begin{aligned} \mathbf{i} \times \mathbf{j} &= \mathbf{k} \\ \mathbf{j} \times \mathbf{i} &= -\mathbf{k} \\ \mathbf{j} \times \mathbf{k} &= \mathbf{i} \\ \mathbf{k} \times \mathbf{j} &= -\mathbf{i} \\ \mathbf{k} \times \mathbf{i} &= \mathbf{j} \\ \mathbf{i} \times \mathbf{k} &= -\mathbf{j} \end{aligned}$$

3.4. **Scalar triple product.** There is a special construction in three dimensions called the *scalar triple product*. Given vectors $a = \langle a_1, a_2, a_3 \rangle$, $b = \langle b_1, b_2, b_3 \rangle$, $c = \langle c_1, c_2, c_3 \rangle$. The scalar triple product, denoted $[a\ b\ c]$, is defined as the number:

$$a_1b_2c_3 - a_1b_3c_2 + a_2b_3c_1 - a_2b_1c_3 + a_3b_1c_2 - a_3b_2c_1$$

It can be defined also as the determinant of the matrix:

$$\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}$$

Finally, $[a\ b\ c]$ can be defined in terms of dot and cross product:

$$[a\ b\ c] = a \cdot (b \times c) = b \cdot (c \times a) = c \cdot (a \times b)$$

The scalar triple product has the following properties:

- If a, b, and c are in a common plane (i.e., there is a plane containing the heads of all the vectors, when their tails are at the origin), then the scalar triple product is 0. In particular, if any two of the vectors are collinear (i.e., one is a scalar multiple of the other), the scalar triple product is 0.
- The scalar triple product is invariant under cyclic permutations, i.e., $[a\ b\ c] = [b\ c\ a] = [c\ a\ b]$.
- The scalar triple product gets negated if we just switch two of the pieces: $[a\ b\ c] = -[b\ a\ c]$.
- The scalar triple product is distributive in each input. For instance, here's distributivity in the third input: $[a\ b\ (c+d)] = [a\ b\ c] + [a\ b\ d]$.
- Scalars can be pulled out from any of the three inputs to the outside.

4. Geometric applications of dot product and cross product

Examples are not included in the lecture notes, but you can see some in the book and you will also do more practice of examples when doing your homework.

4.1. **Area of triangle, parallelogram, volume of parallelopided.** To compute the area of a triangle or parallelogram, we need to first treat the sides as free vectors. To do this, we subtract the coordinates of the endpoints of the sides from each other.

For instance, consider a triangle with vertices $P(p_1, p_1, p_3)$, $Q(q_1, q_2, q_3)$, and $R(r_1, r_2, r_3)$. The side PQ, viewed as a vector, is $\langle q_1 - p_1, q_2 - p_2, q_3 - p_3 \rangle$. Similarly, the side PR, viewed as a vector, is $\langle r_1 - p_1, r_2 - p_2, r_3 - p_3 \rangle$. The area of the triangle is given by *half* the length of the cross product:

$$\text{Ar}(\triangle PQR) = \frac{1}{2}|\overline{PQ} \times \overline{PR}|$$

For a parallelogram, we compute two of its adjacent sides as vectors, and the area of the parallelogram is then the length of the cross product of these sides. Thus, for a parallelogram $PQRS$, the area is given by:

$$\text{Ar}(PQRS) = |\overline{PQ} \times \overline{PR}|$$

A parallelopiped (sometimes written parallelepided) is a three-dimensional analogue of a parallelogram. The volume of a parallelopiped is the absolute value of the scalar triple product of three adjacent vectors emanating from a single vertex.

4.2. **Determining the angle between vectors.** This can be done using the dot product (see earlier). Note that *we cannot use the cross product approach* – the cross product approach is sufficient to find $\sin\theta$, but there could be two different values of θ (an acute and an obtuse angle) giving rise to the same value of $\sin\theta$.

4.3. **Finding a vector orthogonal to two given vectors.** If the two given vectors are not collinear, we simply take the cross product.

If they are collinear, we can choose any vector orthogonal to either of them, and we basically do this by looking at the dot product.

4.4. **Testing coplanarity and orthogonality.** We can use the dot product, cross product, and scalar triple product to do various kinds of tests:

- To test whether three points are collinear, take the difference vectors of the first and second point, and then of the second and third point. Then, check if these difference vectors are scalar multiples of each other.
- To test whether four points are coplanar, pick one as basepoint, and calculate difference vectors from it of the other three points. The four points are coplanar if and only if this scalar triple product is zero.

VECTOR-VALUED FUNCTIONS

MATH 195, SECTION 59 (VIPUL NAIK)

Corresponding material in the book: Sections 13.1. 13.2.

What students should definitely get: Definition of vector-valued function, relation with parametric description of curves, basic operations on vector-valued functions, limit and continuity definitions and theorems, definition of derivative and integral, notion of tangent vector.

What students should hopefully/eventually get: Top-down and bottom-up descriptions of curves, finding intersections of curves with various kinds of descriptions, particularly in $\mathbb{R}^2$ and $\mathbb{R}^3$.

Executive summary

0.1. Vector-valued functions, limits, and continuity.

(1) *Not for review discussion*: A vector-valued function is a function from $\mathbb{R}$, or a subset of $\mathbb{R}$, to a vector space $\mathbb{R}^n$. It comprises n scalar functions, one for each of the coordinates. For instance, given scalar functions $f_1, f_2, \ldots, f_n$, we can construct a vector-valued function $f = \langle f_1, f_2, \ldots, f_n \rangle$ defined by $t \mapsto \langle f_1(t), f_2(t), \ldots, f_n(t) \rangle$.

(2) *Not for review discussion*: A vector-valued function in n dimensions corresponds to a parametric description of a curve in $\mathbb{R}^n$ whose points are just the heads of the corresponding vectors. The vector-valued function from the previous observation has corresponding curve $\{(f_1(t), f_2(t), \ldots, f_n(t) : t \in D\}$ where D is the appropriate domain.

(3) To add two vector-valued functions in n dimensions, we add them coordinate-wise, where the corresponding scalar functions are added pointwise as usual. This sum is also a vector-valued function in n dimensions.

(4) We can multiply a scalar function and a vector-valued function to get a new vector-valued function. At each point in the domain, this is just multiplication of the corresponding scalar number and the corresponding vector.

(5) We can take the dot product of two vector-valued functions in n dimensions. The dot product is a scalar-valued function. At each point in the domain, it is obtained by taking the dot product of the corresponding vector values.

(6) For $n = 3$, we can take the cross product of two vector-valued functions and get a vector-valued function. This cross product is taken pointwise.

(7) To calculate the limit of a vector-valued function at a point, we calculate the limit separately for each coordinate. We use this idea to define the *limit*, *left hand limit*, and *right hand limit* at any point in the domain.

(8) Limit theorems: Limit of sum is sum of limits, constant scalars pull out of limits, limit of scalar-vector product is product of scalar limit and vector limit, limit of dot product is dot product of limits, limit of cross product (case $n = 3$) is cross product of limits.

(9) A vector-valued function is *continuous* at a point in its domain if each coordinate function is continuous, or equivalently, if the limit equals the value. We say it is continuous on its interval if it is continuous at every point in the interior of the interval and has one-sided continuity at one of the endpoints.

(10) Continuity theorems: Sum of continuous vector-valued functions is continuous, product of continuous scalar function and continuous vector-valued function is continuous, dot product of continuous vector-valued functions is continuous, cross product (case $n = 3$) of continuous vector-valued functions is continuous.

(11) There is no n-dimensional analogue of the intermediate value theorem, multiple things fail.

Actions ...

(1) If no domain is specified, the domain of a vector-valued function is the intersection of the domains of all the constituent scalar functions.

0.2. **Top-down and bottom-up descriptions.** Words ...

(1) A top-down description of a subset of $\mathbb{R}^n$ is in terms of a system of equations and inequality constraints. Each equation (equality constraint) is expected to reduce the dimension by 1 (we start from n) whereas inequality constraints usually have no effect on the dimension. So if there are m independent equality constraints describing a subset of $\mathbb{R}^n$, we expect the subset to have dimension $n - m$.
(2) A bottom-up description is a parametric description with possibly more than one parameter. The number of parameters needed is the dimension of the subset. The parametric descriptions we have seen so far are 1-parameter descriptions and hence they describe curves – 1-dimensional subsets.
(3) The codimension of a m-dimensional subset is $n - m$.
(4) When intersecting, codimensions are expected to add. If the total codimension we get after adding is greater than the dimension of the space, the intersection is expected to be empty.
(5) In $\mathbb{R}^3$, curves are one-dimensional, surfaces are two-dimensional. Thus, curves are not expected to intersect each other, but curves and surfaces are expected to intersect at finite collections of points (in general).

Actions ...

(1) Strategy for finding intersection of subsets in $\mathbb{R}^n$ (specifically, curves and surfaces in $\mathbb{R}^3$) given with top-down descriptions: Take all the equations together and solve simultaneously.
(2) Strategy for finding intersection of curve given parametrically and curve or surface given by top-down description: Plug in the functions of the parameter for the coordinates in the top-down description.
(3) Strategy for finding intersection of curves given parametrically: Choose different letters for parameter values, and then equate coordinate by coordinate. We get a bunch of equations in two variables (the two parameter values).
(4) Strategy for finding collision of curves given parametrically: Just equate coordinates, using the same letter for parameter values. Get a bunch of equations all in one variable.

0.3. **Differentiation, tangent vectors, integration.**

(1) The derivative of a n-dimensional vector-valued function is again a n-dimensional vector-valued function. It can be defined by differentiating each coordinate with respect to the parameter, or by using a difference quotient expression. These definitions are equivalent.
(2) This derivative operation satisfies the sum rule, pulling out constant scalars, and product rules for scalar-vector multiplication, dot product, and cross product (case $n = 3$).
(3) As a free vector, the tangent vector at $t = t_0$ to a parametric description of a curve is just the derivative vector for the corresponding vector-valued function. As a localized vector, it starts off at the corresponding point in $\mathbb{R}^n$.
(4) The tangent vector for a curve with parametric description depends on the choice of parameterization. The *unit tangent vector* does not, apart from the issue of direction (forward or backward). The unit tangent vector is a unit vector (i.e., length 1 vector) in the direction of the tangent vector. It is unique for a given curve (independent of parameterization) up to forward-backward issues.
(5) To perform definite or indefinite integration of a vector-valued function, we perform the integration coordinate-wise.

1. Vector-valued functions, parametric descriptions and more

As is my wont, I will, wherever possible, state things in n-dimensional terms and then discuss any geometric significance of the case where $n = 2$ or $n = 3$. As mentioned in a previous lecture, restricting to $n = 3$ is a somewhat artificial thing to do from the perspective of the social sciences because the number of quantities that we are interested in simultaneously studying is often substantially more than 3.

1.1. **Vector-valued functions and parametric descriptions of curves.** We hinted at this last time, when motivating vectors, but let's make this formal.

A *vector-valued function* in n dimensions on a subset D of $\mathbb{R}$ is a collection of n functions $f_1, f_2, \ldots, f_n : D \to \mathbb{R}$, which are pieced together as coordinates of a vector as follows:

$$t \mapsto \langle f_1(t), f_2(t), \ldots, f_n(t) \rangle, \qquad t \in D$$

Thus, a vector-valued function is a *vector of functions in the usual sense*.

A vector-valued function corresponds to a parametric description of a curve in $\mathbb{R}^n$, and the curve is simply the set of corresponding points to the vectors:

$$\{(f_1(t), f_2(t), \ldots, f_n(t)) : t \in D\}$$

Note that there is the usual distinction between a curve and its parameterization. The curve is simply the subset of $\mathbb{R}^n$, whereas the parameterization is a particular story about how that curve was built. The same curve could admit multiple parameterizations that differ in timing, speed, direction, and choices made at self-intersection points.

1.2. **Domain convention for vector-valued functions.** If we are given a vector-valued function $f = \langle f_1, f_2, \ldots, f_n \rangle$ *without* a domain being specified, the domain is implicitly taken to be the *largest possible subset* of $\mathbb{R}$ on which f makes sense. This turns out to be the *intersection* of the domains of the functions $f_1, f_2, \ldots, f_n$.

1.3. **The two-dimensional and three-dimensional cases.** We previously examined the case $n = 2$, and this was what we called parametric descriptions of curves in the plane. The case $n = 3$ gives parametric descriptions of curves in space. These are sometimes called *space curves*. We will talk about these a little later, as a follow-up to a general discussion about top-down versus bottom-up approaches to finding subsets in $\mathbb{R}^n$.

1.4. **Multiple inputs and multiple outputs.** There are two ways in which multivariable calculus generalizes single variable calculus. The first is that we can now have *outputs* which are vectors, or tuples of real numbers, instead of single real numbers. The second is that we can have *inputs* which are tuples of real numbers, instead of single real numbers.

Of these, the challenge that we will currently deal with is the outputs challenge. This turns out to be not much of a challenge at all, and the key idea is to simply deal with things one output coordinate at a time.

The other challenge is the inputs challenge, namely, how do we deal with functions *of more than one variable*. This is fundamentally a deeper challenge. One of the ideas is to deal with the function one input at a time, but the other inputs cannot be completely ignored. The upshot of it all is that dealing with multiple inputs is something we will have to defer till a little later in the course.

Okay, now we move to the baby stuff.

2. Operations on vector-valued functions

2.1. **Four kinds of additions.** If f and g are vector-valued functions in n dimensions, given by $f = \langle f_1, f_2, \ldots, f_n \rangle$ and $g = \langle g_1, g_2, \ldots, g_n$, then $f + g$ is given by the vector-valued function $\langle f_1 + g_1, f_2 + g_2, \ldots, f_n + g_n \rangle$. Explicitly, it is given by the function:

$$t \mapsto \langle f_1(t) + g_1(t), f_2(t) + g_2(t), \ldots, f_n(t) + g_n(t) \rangle$$

Overall, we have seen four kinds of additions:

- Addition of scalar numbers.
- Addition of scalar-valued functions, which is done *pointwise*, i.e., to add two scalar-valued functions, we add their values at each point in the input domain.
- Addition of vectors, which is done *coordinate-wise*.
- Addition of vector-valued functions, which is done *pointwise and coordinate-wise*, i.e., to add two vector-valued functions, we add the vector values at each point in the input domain. To add these vectors, we in turn do coordinate-wise addition.

2.2. **Scalar-vector multiplication, dot product, and cross product.** Suppose f is a scalar-valued function and g is a n-dimensional vector-valued function. We can define the product fg. This is the function:

$$t \mapsto \langle f(t)g_1(t), f(t)g_2(t), \ldots, f(t)g_n(t)\rangle$$

Suppose f and g are n-dimensional vector-valued functions of one variable. The *dot product* $f \cdot g$ is the function that sends each t to the scalar obtained by taking the dot product $f(t) \cdot g(t)$. Thus, $f \cdot g$ is a scalar-valued function of one variable.

In the case $n = 3$, we can define, for 3-dimensional vector-valued function f and g, a vector-valued function $f \times g$. This sends each point t to the vector $f(t) \times g(t)$.

3. Limits and continuity

3.1. **Definition of limit.** Suppose we have a vector-valued function:

$$f(t) := \langle f_1(t), f_2(t), \ldots, f_n(t)\rangle$$

Then, for a real number a, we define:

$$\lim_{t\to a} f(t) = \langle \lim_{t\to a} f_1(t), \lim_{t\to a} f_2(t), \ldots, \lim_{t\to a} f_n(t)\rangle$$

In other words, we take limits coordinate-wise. We can similarly define various notions such as *left hand limit* and *right hand limit*.

3.2. **Limit theorems.** We have vector limit theorems that are analogous to, and follow from, the limit theorems for scalars. The usual caveats apply with the interpretation of these results.

- The limit of the sum is the sum of the limits.
- The limit of a scalar multiple is the same scalar multiple of the limit.
- The limit of a product of a scalar-valued function and a vector-valued function is the product of the limit of the scalar-valued function and the limit of the vector-valued function.
- The limit of the dot product is the dot product of the limits.
- In the case $n = 3$, the limit of the cross product is the cross product of the limits.

3.3. **Continuous vector-valued functions.** Continuing with the de ja vu, we define a vector-valued function to be continuous at a point if the limit at the point equals the value at the point. Similarly, we define *left continuous* and *right continuous*. We say that a vector-valued function is continuous on an interval if it is continuous at all points in the interior of the interval and has one-sided continuity from the appropriate side at boundary points.

For each of these notions of continuity, a vector-valued function satisfies that notion if and only if all the corresponding scalar functions satisfy that fnotion.

3.4. **Continuity theorems.** We have the following:

- The sum of continuous vector-valued functions is continuous.
- The product of a constant and a continuous vector-valued function is continuous.
- The product of a continuous scalar-valued function and a continuous vector-valued function is continuous.
- The dot product of two continuous vector-valued functions is a continuous scalar-valued function.
- In the case $n = 3$, the cross product of two continuous vector-valued functions is also a continuous vector-valued function.

3.5. **Intermediate value theorem fails to have an analogue.** Recall the intermediate value theorem, which states that any continuous function f that takes a value $f(a)$ at a and $f(b)$ at b must take all values between $f(a)$ and $f(b)$ on inputs between a and b.

This statement fails to have an analogue for vector-valued functions. First, there isn't a very precise notion of *between* for vector-valued functions. In other words, there is no natural total ordering on vectors that preserves the algebraic and geometric properties we are interested in. So it doesn't even make sense to formulate the statement. We *could* come up with some plausible formulations, but they are false.

It *is* true that each of the coordinates passes through all intermediate values. However, it is not necessary that every *vector* of intermediate values is achieved, because the different coordinates could be changing in different ways.

4. Top-down versus bottom-up: revisiting dimensionality arithmetic

4.1. **Top-down in n dimensions.** n-dimensional space, which we sometimes denote $\mathbb{R}^n$, is n-dimensional because there are n degrees of freedom in specifying a point, i.e., we need to provide n real numbers in order to specify a unique point in the space.

There are two ways of constructing subsets of $\mathbb{R}^n$, the *top-down* approach and the *bottom-up* approach.

In the top-down approach, we start with all of $\mathbb{R}^n$. Then, we decide to whittle down. To whittle the subset down, we provide constraints that a point must satisfy in order to lie in the subset. Here are two key things to note:

- As a *general heuristic*, every new constraint that is *a single scalar equality constraint* and is *mostly independent of previously introduced constraints* reduces the dimensionality of the space by 1. So, if we start with 24-dimensional space, and introduce 3 mutually independent scalar equality constraints, then the subset of the whole space that satisfies all of these constraints is expected to be 21-dimensional (whatever that means).
- As a *general heuristic*, a new constraint that is *a single scalar ineuqality constraint* has no effect on the dimensionality. For instance, in 2-dimensional space with coordinates x and y, introducing the constraint $y > 0$ restricts us to the upper half-plane, but this still has full dimensionality, i.e., 2.
- Combining these, we see that as a *general heuristic*, if we have more than n scalar equality constraints in a n-dimensional space, the solution space is expected to be empty – because its dimension is *less than zero*, which isn't possible. The exception arises when there is some hidden consistency or dependency between the equality constraints.

It's worth noting that there are many exceptions. For instance, the single scalar constraint $x^2+y^2+z^2 = 0$ in three-dimensional space with coordinates (x, y, z) has a solution space that is zero-dimensional (a single point). Similarly, the constraint $x^2 + y^2 + z^2 = -1$ has an empty solution space, while the constraint $(x + y - 1)^2 + (y + z - 1)^2 = 0$ has a one-dimensional solution space even though $3 - 1 = 2$.

However, these are the exceptions that prove the rule, in the following sense. Over the real numbers, setting a sum of squares of a bunch of quantities to be equal to zero is likely secrely compressing the condition that all of them are equal to zero into a single equation. So it's a bit of cheating compression of multiple constraints into what ostensibly looks like a single constraint. This is a sleight of hand.[1]

4.2. **Bottom-up in n dimensions.** In the previous subsection, we discussed an approach where we start with the whole space and then whittle down the set of points under consideration by adding constraints. I dubbed this the *top-down* approach. In constrast, there is a *bottom-up* approach. Here, we start with a clean slate and then draw in new points.

The parametric descriptions of curves, and vector-valued functions, are the *one-dimensional* case of bottom-up descriptions. A parametric description uses *one* real input (one degree of freedom) and traces a curve in $\mathbb{R}^n$ from that.

What analogoue of parametric description can we use to get higher-dimensional subsets of $\mathbb{R}^n$? For instance, how do we obtain parametric descriptions of *surfaces* in $\mathbb{R}^n$, i.e., two-dimensional subsets of $\mathbb{R}^n$?

[1] If you were working over the complex numbers instead of the real numbers, this kind of sleight of hand would not be possible. There are some deep results in mathematics, which you may never see in your life, which basically say that over the complex numbers, some rigorous version of the statements I made above *is* true if we restrict ourselves to things involving polynomials.

The general idea is as follows: a parametric description of a k-dimensional subset involves a k-parameter description. For instance, a parametric description of a surface in $\mathbb{R}^n$ corresponds to a n-dimensional vector-valued function with *two* real inputs, so that each of the coordinates itself is a function of two-variables. Concretely, it looks like:

$$(t, u) \mapsto \langle f_1(t, u), f_2(t, u), \ldots, f_n(t, u) \rangle$$

In other words, we now have *two degrees of freedom for the input*. The input pair is free to vary over some subset of $\mathbb{R}^2$, and the output traces a surface in $\mathbb{R}^n$.

We will return to this idea a little later in the course. It requires a new way of thinking about functions since we now have to deal with functions of many variables, a topic that has been taboo so far. Later in the course, we will develop a theory of continuity, differentiability, and derivative computations for such functions.

4.3. **Dimension, codimension, and intersection theory.** Suppose we are in a vector space of dimension n. We are given subsets M and P which are m-dimensional and p-dimensional respectively, where both m and p are less than or equal to n. What should we *expect* regarding the dimension of the intersection of the two subsets?

Intersection is a top-down approach, so we need to rethink of the subsets in terms of the constraints imposed to get them. The m-dimensional subset M arose because of $n - m$ independent constraints. The number $n - m$, i.e., the difference in dimension of the whole space and the subset, is termed the *codimension* of the subset. Similarly, the subset P is defined using $n - p$ independent constraints, so its codimension is $n - p$.

When we take intersections, then, *generically*, the codimensions add, i.e., the number of constraints needed to define the intersection is, generically, the sum of the number of constraints needed to define each subset. This need not always be true, but it is what we should *expect* in general. This means that the codimension of the intersection $M \cap P$ is $(n - m) + (n - p) = 2n - m - p$ and the dimension is therefore $m + p - n$. If $m + p < n$, then we *should expect, generically*, that the intersection is empty.

Here are some numerical examples:

- In 20-dimensional space, the intersection of an 18-dimensional subset and a 17-dimensional subset is expected to be 15-dimensional: The 18-dimensional subset has codimension 2, the 17-dimensional subset has codimension 3. When we intersect, the codimensions are expected to add, and we get codimension 5, which means dimension 15.
- In 5-dimensional space, the intersection of a 3-dimensional subset and a 2-dimensional subset is expected to be 0-dimensional, and thus is expected to be a point or finite collection of points: The 3-dimensional subspace has codimension 2, and the 2-dimensional subspace has codimension 3. When we intersect, the codimensions are expected to add, so we get $2 + 3 = 5$, so its dimension is 0.
- In 7-dimensional space, the intersection of a 3-dimensional subset and a 2-dimensional subset is expected to be empty, because its expected dimension comes out to be negative: The 3-dimensional subset has codimension 4, and the 2-dimensional subspace has codimension 5, so the codimensions add, and we get $4 + 5 = 9$. But codimension 9 in a 7-dimensional space yields a negative value for the dimension, which is not possible, so the intersection is likely empty.

Note that *in reality*, the intersection may have bigger dimension than expected, which happens if the constraints describing the two subsets are not independent of each other. Also, it is possible that the intersection is a lot smaller, or is even empty. So, while the above is a good thumb rule of what to expect, in practice you actually need to set up equations and solve.

4.4. **Curves in the plane.** A curve in the plane is something 1-dimensional in a 2-dimensional space, so both its dimension and codimension are 1. It could be described in either of these ways:

- The *top-down description* of the curve uses a single equational constraint in $\mathbb{R}^2$. This is a *relational* or *implicit* description, something of the form $F(x, y) = 0$ for some function F of two variables.
- The *bottom-up description* of the curve uses a *parametric* description of the form $x = f(t)$, $y = g(t)$.

We now note that the *intersection* of two curves in the plane is *expected* to be zero-dimensional, which means it is expected to be a finite collection of points. It may well happen that this intersection is zero-dimensional.

How do we *find* the point of intersection? We make various cases:

- If the two curves are given by relational descriptions $F_1(x, y) = 0$ and $F_2(x, y) = 0$, then we try to solve the system of two equations in two variables.
- If one curve is given by the relational description is $F(x, y) = 0$ and the other curve is given by the parametric description is $x = f(t)$, $y = g(t)$, then to solve this, we solve $F(f(t), g(t)) = 0$ as one equation in one variable. For each t-value we get, we compute the corresponding values of x and y.
- If both curves are parametric, say the first curve is given by $x = f_1(t)$, $y = g_1(t)$, and $x = f_2(t)$, $y = g_2(t)$, then to find the intersection, we first make sure that the dummy parameter letters are different. So we rewrite the parameter for the first curve as t_1 and the parameter for the second curve as t_2. We thus have $x = f_1(t_1)$, $y = g_1(t_1)$, and $x = f_2(t_2)$, $y = g_2(t_2)$. To find the intersection, solve the system of two equations: $f_1(t_1) = f_2(t_2)$ and $g_1(t_1) = g_2(t_2)$ in the two variables t_1 and t_2. After finding t_1 and t_2, plug in the function values to find the points x and y.

4.5. **Collision of curves.** The above described the *intersection* of curves. To intersect curves, we are simply interested in where their paths cross. There is a related notion of *collision* of curves with a time parameter, which occurs if two curves intersect with the *same value of time parameter for both*.

In symbols, if $x = f_1(t)$, $y = g_1(t)$ is one parametric curve and $x = f_2(t)$, $y = g_2(t)$ is another parametric curve, then to find whether they collide, we need to solve the system $f_1(t) = f_2(t)$ and $g_1(t) = g_2(t)$. This is a *system of two equations in one variable*.

What does this mean? If the number of equations is more than the number of variables, then we should in general expect *no solutions*. Specifically, in our case, we should not in general expect that the (usually finite) set of solutions to the equation $f_1(t) = f_2(t)$ has a non-empty intersection with the (usually finite) set of solutions to the equation $g_1(t) = g_2(t)$.

However, this *could* happen, even if unexpected.

Collision is much stronger and rarer than intersection, because collision requires the same value of time parameter where the curves intersect.

4.6. **Space curves.** A space curve, or a curve in space, is a 1-dimensional subset in 3-dimensional space. In particular, it has codimension 2. It could be described in two ways:

- A *top-down description*, which involves setting up *two* equations in the three coordinates, i.e., equations of the form $F(x, y, z) = 0$ and $G(x, y, z) = 0$.
- A *bottom-up description*, which involves writing all coordinates as functions of a parameter, i.e., $x = f(t)$, $y = g(t)$, and $z = h(t)$.

What can we say about the *intersections* of space curves? In general, space curves are not expected to intersect, because the codimension of the intersection turns out to be 4, which is bigger than the dimension of the space. However, they just might intersect. Here's how we compute the intersections:

- Suppose the two curves are both given in relational (top-down) form. Say one curve is given by $F_1(x, y, z) = 0$ and $G_1(x, y, z) = 0$ and the other curve is given by $F_2(x, y, z) = 0$ and $G_2(x, y, z) = 0$. To find the intersection, we need to solve the *four* equations $F_1(x, y, z) = 0$, $F_2(x, y, z) = 0$, $G_1(x, y, z) = 0$, $G_2(x, y, z) = 0$ in *three* variables. In general, we do not expect any solutions, because the dimension of the solution space appears to be -1. However, it may happen by sheer chance that there is a solution, i.e., that the curves intersect.
- Suppose one curve is given in relational form and the other curve is given in parametric form. Say, the top-down description for the first curve is $F_1(x, y, z) = 0$ and $G_1(x, y, z) = 0$, and the bottom-up description for the second curve is $x = f_2(t)$, $y = g_2(t)$, and $z = h_2(t)$. To solve these, we plug in the parametric description into the relational description, so we're trying to solve the two equations $F_1(f_2(t), g_2(t), h_2(t)) = 0$ and $G_1(f_2(t), g_2(t), h_2(t)) = 0$. This is s system of *two* equations in one variable, and we *expect* no solutions, because the solution space is expected to have dimension -1. But there may be a solution by chance.

- Suppose both curves are given in parametric form, say the first one is given as $x = f_1(t)$, $y = g_1(t)$, and $z = h_1(t)$. The second curve is given by $x = f_2(t)$, $y = g_2(t)$, and $z = h_2(t)$. In order to find the intersection, we change the dummy parameters to different letters, say t_1 for the first curve and t_2 for the second curve. We now have to solve the three equations $f_1(t_1) = f_2(t_2)$, $g_1(t_1) = g_2(t_2)$, and $h_1(t_1) = h_2(t_2)$ in the two variables.

4.7. **Collision of space curves.** We say that space curves collide if they intersect with the same value of the parameter on both. Note that collision of space curves is even less likely than intersection – the expected dimension of the space of intersection points is -1, the expected dimension of the space of collision points is -2.

Given the parametric description $x = f_1(t)$, $y = g_1(t)$, $z = h_1(t)$ and the parametric description $x = f_2(t)$, $y = g_2(t)$, $z = h_2(t)$. Then, to find the collision points, we need to solve the *three* equations $f_1(t) = f_2(t)$, $g_1(t) = g_2(t)$, $h_1(t) = h_2(t)$ in the *one* variable t.

5. Differentiation of vector-valued functions

5.1. **Definition of derivative.** Given a vector-valued function of the form:

$$f = t \mapsto \langle f_1(t), f_2(t), \ldots, f_n(t) \rangle$$

We define:

$$f'(t) = \langle f_1'(t), f_2'(t), \ldots, f_n'(t) \rangle$$

In other words, the derivative of a vector-valued function is the vector of the derivatives of each of its scalar functions.

Another equivalent definition is as a *limit of a difference quotient*:

$$f'(t) := \lim_{h \to 0} \frac{f(t+h) - f(t)}{h}$$

Here, h is a *real number*, the subtraction $f(t+h) - f(t)$ is *vector subtraction*, and the division by h is *scalar multiplication* of the scalar $1/h$ by the vector $f(t+h) - f(t)$.

Unpacking this definition gives the same as the definition in terms of coordinates given earlier.

5.2. **Definition of higher derivatives.** Differentiating once was the hard part, now we can just keep going on and on repeatedly. The k^{th} derivative of a vector-valued function is the vector of the k^{th} derivatives of each of its component scalar functions.

In other words, if we have $f(t) = \langle f_1(t), f_2(t), \ldots, f_n(t) \rangle$, then:

$$f^{(k)}(t) = \langle f_1^{(k)}(t), f_2^{(k)}(t), \ldots, f_n^{(k)}(t) \rangle$$

5.3. **Rule for addition and scalar multiplication.** If f and g are n-dimensional vector-valued functions, then $f + g$ is defined as the n-dimensional vector-valued function that, at each point t, is $f(t) + g(t)$. In turn, to add $f(t) + g(t)$, we add them coordinate-wise. Thus, we're being doubly wise: we're first adding the functions point-wise, and then at each point, we're adding them coordinate-wise.

If both f and g are differentiable, then so is $f + g$, and $(f + g)' = f' + g'$, and if a is a real number, then $(af)' = a(f')$.

This rule extends to higher derivatives as well (secret reason: a composition of linear maps is linear), so we get that if f and g are k times differentiable, then $(f + g)^{(k)} = f^{(k)} + g^{(k)}$ for all natural numbers k, and $(af)^{(k)} = af^{(k)}$ for all real numbers a.

5.4. **Rule for scalar multiplication.** If f is a differentiable scalar-valued function and g is a differentiable vector-valued function, then fg is a differentiable vector-valued function, and:

$$(fg)' = (f')(g) + (f)(g')$$

This is the scalar-vector multiplication version of the product rule. Is it a coincidence that it looks just like the product rule for scalars? No. In fact, there are two far-reaching reasons why any product rule should look like the product rule you are familiar with. We will not cover either right now, but they will both become clear to you later in life.

5.5. **Rule for dot product.** If f and g are both differentiable *as vector-valued functions*, then $f \cdot g$ is differentiable as a scalar-valued function, and we have:

$$(f \cdot g)' = ((f') \cdot g) + (f \cdot (g'))$$

where the addition on the right side is pointwise addition of scalar-valued functions.

5.6. **Rule for cross product.** It turns out that we have a product rule for this, just as we would expect:

$$(f \times g)' = (f' \times g) + (f \times g')$$

However, you need to be careful here, because the cross product is *not commutative*. For the product rule with scalars (and even for the dot product) it is not critical to remember the order in which we write the terms within each product. But with the cross product, it is. *Remember that the order in which the functions being crossed is the same* on the left side and in each of the things being added on the right side.

6. Tangent vectors

6.1. **Tangent vector in the context of a parameterization.** For a curve with parametric description:

$$t \mapsto (f_1(t), f_2(t), \ldots, f_n(t))$$

The tangent vector to the curve is the derivative of the corresponding vector-valued function, i.e., the tangent vector at a point t_0 is (as a free vector) $\langle f_1'(t_0), f_2'(t_0), \ldots, f_n'(t_0)\rangle$. As a *localized* vector, it is the vector from the point $\langle f_1(t_0), f_2(t_0), \ldots, f_n(t_0)\rangle$ to the point $\langle f_1(t_0) + f_1'(t_0), f_2(t_0) + f_2'(t_0), \ldots, f_n(t_0) + f_n'(t_0)\rangle$. In other words, it starts at the point where we are taking the tangent vector, and goes the tangent vector.

6.2. **Unit tangent vector to a curve.** It's worth noting that the tangent vector depends on the parameterization, but the *line* along which that vector appears does not. If we choose a different parameterization, then the length of the tangent vector at a point might change. If our parameterization traverses the same curve in reverse, then the direction may become opposite to what it originally was. However, the line of the tangent vector remains the same, i.e., any two tangent vectors for different parameterizations are parallel.

Given this, it makes sense, just given the curve, to talk of the unit tangent vectors (two of them, opposite to each other) and these are independent of the parameterization. Given a direction of traversal (without a concrete parameterization) we can pick the unit tangent vector in the forward direction for that traversal.

7. Integration of vector-valued functions

7.1. **Indefinite integration.** Suppose f is a vector-valued function. We say that F is an *antiderivative* or *indefinite integral* of f if $F' = f$. To find an antiderivative of a vector-valued function, we simply find an antiderivative of each of the coordinate functions.

7.2. **Definite integration.** The *definite integral* of a vector-valued function of t from $t = a$ to $t = b$ is defined coordinate-wise: we perform definite integrations for each of the coordinate function. Specifically if we have:

$$f(t) := \langle f_1(t), f_2(t), \ldots, f_n(t) \rangle$$

then the integral is:

$$\int_a^b f(t)\,dt = \langle \int_a^b f_1(t)\,dt, \int_a^b f_2(t)\,dt, \ldots, \int_a^b f_n(t)\,dt \rangle$$

This definite integral is thus a n-dimensional vector. Recall that ordinarily, the definite integral of a scalar-valued function between fixed limits is a *number*, i.e., a *scalar*.

7.3. **Fundamental theorem of calculus.** The fundamental theorem of calculus applies for vector integration. This isn't deep – the basic reason here is that it applies in each coordinate.

7.4. **Sums and scalar multiples.** Both indefinite and definite integrals are *linear*, in the sense that they split across sums and scalar multiples can be pulled out of expressions. In particular, we have the following for vector-valued functions f and g, real numbers a and b, and scalars λ:

$$\begin{aligned} \int_a^b ((f(t) + g(t))\,dt &= \int_a^b f(t)\,dt + \int_a^b g(t)\,dt \\ \int_a^b \lambda f(t)\,dt &= \lambda \int_a^b f(t)\,dt \end{aligned}$$

7.5. **Products and chains.** It is possible to write down vector versions of the integration by parts, u-substitution rules, etc. In practice, though, it is much easier, and more general as well as more powerful, to simply do the integrations in each coordinate separately and bring to bear in each coordinate the entire arsenal of techniques for integration in one variable.

EQUATIONS OF LINES AND PLANES

MATH 195, SECTION 59 (VIPUL NAIK)

Corresponding material in the book: Section 12.5.

What students should definitely get: Parametric equation of line given in point-direction and two-point form, symmetric equations of line, degenerate cases where direction vector has one or more coordinate zero, intersecting lines, equation of plane, angle between planes, line of intersection of planes, distance of point and plane.

What students should hopefully get: How the equation setup relates to the general setup for curves and surfaces. Understanding of the degenerate cases. Role of parameter restrictions in defining a line segment. Deeper understanding of relationship of direction vector and direction cosines.

Executive summary

0.1. **Direction cosines.**

(1) For a nonzero vector v, there are two unit vectors parallel to v, namely $v/|v|$ and $-v/|v|$.
(2) The direction cosines of v are the coordinates of $v/|v|$. if $v/|v| = \langle \ell, m, n \rangle$, then the direction cosines are ℓ, m, and n. We have the relation $\ell^2 + m^2 + n^2 = 1$. Further, if α, β, and γ are the angles made by v with the positive x, y, and z axes, then $\ell = \cos\alpha$, $m = \cos\beta$, and $n = \cos\gamma$.

0.2. **Lines.** Words ...

(1) A line in $\mathbb{R}^3$ has dimension 1 and codimension 2. A parametric description of a line thus requires 1 parameter. A top-down equational description requires two equations.
(2) Given a point with radial vector $\mathbf{r}_0$ and a direction vector $\mathbf{v}$ along a line, the parametric description of the line is given by $\mathbf{r}(t) = \mathbf{r}_0 + t\mathbf{v}$. If $\mathbf{r}_0 = \langle x_0, y_0, z_0 \rangle$ and $\mathbf{v} = \langle a, b, c \rangle$, this is more explicitly described as $x = x_0 + ta$, $y = y_0 + tb$, $z = z_0 + tc$.
(3) Given two points with radial vectors $\mathbf{r_0}$ and $\mathbf{r_1}$, we obtain a vector equation for the line as $\mathbf{r}(t) = t\mathbf{r_1} + (1-t)\mathbf{r_0}$. If we restrict t to the interval $[0,1]$, then we get the line segment joining the points with these radial vectors.
(4) If the line is not parallel to any of the coordinate planes, this parametric description can be converted to symmetric equations by eliminating the parameter t. With the above notation, we get:

$$\frac{x - x_0}{a} = \frac{y - y_0}{b} = \frac{z - z_0}{c}$$

This is actually *two* equations rolled into one.
(5) If $c = 0$ and $ab \neq 0$, the line is parallel to the xy-plane, and we get the equations:

$$\frac{x - x_0}{a} = \frac{y - y_0}{b}, \qquad z = z_0$$

Similarly for the other cases where precisely one coordinate is zero.
(6) If $a = b = 0$ and $c \neq 0$, the line is parallel to the z-axis, and we get the equations:

$$x = x_0, \qquad y = y_0$$

Actions ...

(1) To intersect two lines both given parametrically: Choose different letters for parameters, equate coordinates, solve 3 equations in 2 variables. *Note: Expected dimension of solution space is* $2 - 3 = -1$.

(2) To intersect a line given parametrically and a line given by equations: Plug in the coordinates as functions of parameters into both equations, solve. Solve 2 equations in 1 variable. *Note: Expected dimension of solution space is* $1 - 2 = -1$.
(3) To intersect two lines given by equations: Combine equations, solve 4 equations in 3 variables. *Note: Expected dimension of solution space is* $3 - 4 = -1$.

0.3. **Planes.** Words ...

(1) Vector equation of a plane (for the radial vector) is $\mathbf{n} \cdot (\mathbf{r} - \mathbf{r_0}) = 0$ where $\mathbf{n}$ is a normal vector to the plane and $\mathbf{r_0}$ is the radial vector of any fixed point in the plane. This can be rewritten as $\mathbf{n} \cdot \mathbf{r} = \mathbf{n} \cdot \mathbf{r_0}$. Using $\mathbf{n} = \langle a, b, c \rangle$, $\mathbf{r} = \langle x, y, z \rangle$, and $\mathbf{r}_0 = \langle x_0, y_0, z_0 \rangle$, we get the corresponding scalar equation $ax + by + cz = ax_0 + by_0 + cz_0$. Set $d = -(ax_0 + by_0 + cz_0)$ and we get $ax + by + cz + d = 0$.
(2) The "direction" or "parallel family" of a plane is determined by its normal vector. The angle between planes is the angle between their normal vectors. Two planes are parallel if their normal vecors are parallel. And so on.

Actions ...

(1) Given three non-collinear points, we find the equation of the unique plane containing them as follows: first we find a normal vector by taking the cross product of two of the difference vectors. Then we use any of the three points to calculate the dot product with the normal vector in the above vector equation or the corresponding scalar equation.

    Note that if the points are collinear, there is no unique plane through them – any plane containing their line is a plane containing them.
(2) We can compute the angle of intersection of two planes by computing the angle of intersection of their normal vectors.
(3) The line of intersection of two planes that are not parallel can be computed by simply taking the equations for *both* planes. This, however, is not a standard form for a line in $\mathbb{R}^3$. To find a standard form, either find two points by inspection and join them, or find one point by inspection and another point by taking the cross product of the normal vectors to the plane.
(4) To intersect a plane and a line, plug in parametric expressions for the coordinates arising from the line into the equation of the plane. We get one equation in the one parameter variable. In general, this is expected to have a unique solution for the parameter. Plug in the value of the parameter into the parametric expressions for the line and get the coordinates of the point of intersection.
(5) For a point with coordinates (x_1, y_1, z_1) and a plane $ax + by + cz + d = 0$, the distance of the point from the plane is given by $|ax_1 + by_1 + cz_1 + d|/\sqrt{a^2 + b^2 + c^2}$.

1. Lines and planes

1.1. **Lines: dimension and codimension.** A line in $\mathbb{R}^n$ has dimension one and codimension $n - 1$. In particular, a line in Euclidean space $\mathbb{R}^3$ has dimension 1 and codimension $3 - 1 = 2$. In particular, based on what we know of dimension and codimension, we expect that:

- In a top-down or relational description, we should need *two* independent equations to define a line.
- In a bottom-up or parametric description, we should need *one* parameter to define a line.

1.2. **Planes: dimension and codimension.** A plane in $\mathbb{R}^3$ is 2-dimensional, and it has codimension $3 - 2 = 1$. In particular, based on what we know of dimension and codimension, we expect that:

- In a top-down or relational description, we should need *one* equation to define a plane.
- In a bottom-up or parametric description, we should need *two* parameters to define a plane. *This gets into the realm of functions of two variables, so we will defer the actual* 2*-parameter description of planes for now.*

1.3. **Intersection theory.** We have the following basic intersection facts:

Intersect	Generic case	Special case 1	Special case 2	Special case 3
Plane, plane	Line	Empty (parallel planes)	Plane (equal planes)	
Plane, line	Point	Empty (line parallel to, not on plane)	Line (line on plane)	
Line, line	Empty (skew lines)	Point (intersecting lines)	Empty (parallel lines)	Line (equal lines)

The *generic case* here represents the case that is most likely, i.e., the case that would arise if the things being intersected were chosen randomly. There are mathematical ways of making this precise, but these are beyond the current scope.

In particular, it is worth pointing out that the generic case is exactly as intersection theory predicts. Let's consider the three generic cases:

- *Generic intersection of plane and plane*: A plane has codimension 1, so the intersection of two planes (generically) has codimension $1+1 = 2$. We know that a line has codimension 2, so this makes sense.
- *Generic intersection of plane and line*: A plane has codimension 1 and a line has codimension 2, so the intersection of a plane and a line (generically) has codimension $1+2 = 3$, so it is zero-dimensional. A point is zero-dimensional.
- *Generic intersection of line and line*: A line has codimension 2, so the intersection of two lines (generically) has codimension $2 + 2 = 4$, so it has dimension $3 - 4 = -1$. Negative dimension indicates that the intersection is generically empty.

After we study the intersection theory in detail for lines and planes, we will be in a position to acquire a better understanding of the *general principles* of intersection theory. Specifically, we will acquire a better grasp of the *non-generic* cases where the intersections don't work out as they generically do.

2. Equations of lines

2.1. **The point-direction form.** The general principle behind this is the same as it is with the *point-slope form*. Basically, to describe a line, it suffices to specify a point on the line, and the *direction* of the line.

The *direction* is specified by specifying any vector parallel to the line. Specifically, given a line with points A and B on it, the direction of the line is given by taking the vector AB. Note that any two vectors that are scalar multiples of each other (i.e., parallel to each other) specify the same direction.

Suppose $\mathbf{r_0}$ is the radial vector for one point on the line, and $\mathbf{v}$ is any nonzero vector along the line. Then the radial vector (i.e., vector from the origin to a point) for points on the line can be defined by the parametric equation:

$$\mathbf{r}(t) = \mathbf{r_0} + t\mathbf{v}$$

where t varies over the real numbers. For each value of t, we get a radial vector for some point on the line, and every point on the line is covered this way.

Suppose $\mathbf{r_0} = \langle x_0, y_0, z_0 \rangle$ and $\mathbf{v} = \langle a, b, c \rangle$. Then $\mathbf{r_0} + t\mathbf{v}$ is the vector:

$$\langle x_0 + ta, y_0 + tb, z_0 + tc \rangle$$

The corresponding parametric description of a curve is:

$$\{(x_0 + ta, y_0 + tb, z_0 + tc) : t \in \mathbb{R}\}$$

Note that the *choice* of parametric description depends on the choice of basepoint in the line and the choice of vector (which can be varied up to scalar multiples).

By the way, here is some terminology (which we overlooked earlier). The *direction cosines* for a particular direction are defined as the coordinates of the *unit vector* in that direction. The direction cosines of a particular direction are denoted ℓ, m, and n. For instance, if a direction vector is $\langle 1, 2, 3 \rangle$, then the corresponding unit vector is $\langle 1/\sqrt{14}, 2/\sqrt{14}, 3/\sqrt{14} \rangle$, so the direction cosines are $\ell = 1/\sqrt{14}$, $m = 2/\sqrt{14}$, and $n = 3/\sqrt{14}$.

The direction cosines are also the cosines of the angles made by the vectors with the x-axis, y-axis, and z-axis. They satisfy the relation:

$$\ell^2 + m^2 + n^2 = 1$$

2.2. **The two-point form.** Suppose $\mathbf{r_0}$ and $\mathbf{r_1}$ are the radial vectors of two points on a line. Then, we can get a line in the point-direction form by setting $\mathbf{v} = \mathbf{r_1} - \mathbf{r_0}$. We thus get the form:

$$\mathbf{r}(t) = \mathbf{r_0} + t(\mathbf{r_1} - \mathbf{r_0})$$

This simplifies to:

$$\mathbf{r}(t) = t\mathbf{r_1} + (1-t)\mathbf{r_0}$$

As t varies over all of $\mathbb{R}$, this gives the whole line. When $t = 0$, we get the point with radial vector $\mathbf{r_0}$ and when $t = 1$, we get the point with radial vector $\mathbf{r_1}$. If we allow only $0 \leq t \leq 1$, we get the *line segment* joining the two points.

2.3. **Top-down description: symmmetric equations.** To obtain the symmetric equations, we start with the parametric equations and then eliminate the parameter. In other words, with the parametric description:

$$\{(x_0 + ta, y_0 + tb, z_0 + tc) : t \in \mathbb{R}\}$$

We note that:

$$x = x_0 + ta, \qquad \implies \qquad t = \frac{x - x_0}{a}$$

Similarly, we get $t = (y - y_0)/b$ and $t = (z - z_0)/c$. Eliminating t, we get:

$$\frac{x - x_0}{a} = \frac{y - y_0}{b} = \frac{z - z_0}{c}$$

Note that while this looks like a single long equation, it is actually *two* equations:

$$\frac{x - x_0}{a} = \frac{y - y_0}{b}$$

and

$$\frac{y - y_0}{b} = \frac{z - z_0}{c}$$

This is in keeping with what we expect/hope – that to describe a 1-dimensional subset in 3-dimensional space, we need $3 - 1 = 2$ equations.

Intuitively, what these equations are saying is that the coordinate changes are in the ratio $a : b : c$.

2.4. **Exceptional case of lines parallel to one of the coordinate planes.** The symmetric equations formulation breaks down if one of the coordinates of the direction vector $\langle a, b, c\rangle$ is zero. In this case, the line is parallel to one of the three coordinate planes, with the third coordinate being unchanged (e.g., if $c = 0$, then the line is parallel to the xy-plane, because its z-coordinate is unchanged).

They break down even more when two coordinates of the direction vector are zero, which means that the line is parallel to one of the axes.

In this case, the symmetric equations given above do not work, and we instead do the following.

- If only one coordinate of the direction vector is zero: If $c = 0$ and $a, b \neq 0$, then we get the two equations:

  $$\frac{x - x_0}{a} = \frac{y - y_0}{b}, \qquad z = z_0$$

  Similarly for the other cases.
- If two coordinates are zero: If, say $a = b = 0$, then we get the two equations:

  $$x = x_0, \qquad y = y_0$$

  z does not appear in the equations because it can vary freely. This line is parallel to the z-axis.

2.5. **Pairs of lines: questions about intersection.** As we noted earlier, lines in $\mathbb{R}^3$ have codimension 2, so the intersection of two lines is expected to be empty. There are qualitatively four possibilities:

(1) The lines are skew lines: This is the most "independent" case possible. Here, the equations describing the two lines are as independent of each other as possible and the two lines thus do not lie in the same plane. They do not intersect.
(2) The lines are intersecting lines in the same plane: This is a somewhat less independent case. Here, there is a plane (not necessarily containing the origin) that contains both lines, and the lines are not parallel, so they intersect at a point.
(3) The lines are parallel lines in the same plane: Here, the equations for the line are inconsistent in a specific way, so they lie in the same plane but are parallel. They do not intersect. *Although the conclusion about intersection is the same both for pairs of parallel lines and for pairs of skew lines, the reasons behind this conclusion are different.*
(4) The two lines are actually the same line: In this case, their intersection is the same line. This is the most dependent case possible.

We now examine how to find the intersection of two lines. The approach is simply a special case of finding the intersection of two curves. Since the equations are all linear, we can actually devise specific procedures to solve the equations.

- *Both lines are given parametrically*: In this case, we first make sure we have different letters for the parameters for each line. Then we equate coordinate-wise and solve the system of 3 linear equations in 2 variables (the parameter variables for the two lines). Note that the number of equations is more than the number of variables – unsurprising since the generic case is one of skew lines.

  After finding solutions for the two parameters, plug back to find the points.
- *One line is given parametrically in terms of t, the other using symmetric equations*: We substitute the parametric expressions into the values of x, y, and z in the symmetric equations and solve the system of two equations in the one (parameter) variable t. After finding solutions for t, plug back to find the points.
- *Both lines are given by symmetric equations*: We solve all the four symmetric equations.

3. Planes

3.1. **Vector description in terms of dot product.** For a given plane in $\mathbb{R}^3$, it either already passes through the origin, or there is a unique plane parallel to it that passes through the origin. We say that two planes are *parallel* if they either coincide or they do not intersect – equivalently, if for every line in one plane, there is a line in the other plane parallel to it.

A family of parallel planes can be thought of as sharing a direction. But how do we specify the direction of a plane, which is a two-dimensional object? The idea is to look at the *complement*, or the *codimension*, of the plane. Specifically, we look at the direction that is *orthogonal* to the plane.

There is a unique direction vector (up to scalar multiples) orthogonal to a family of parallel planes. Further, the dot product of this vector with the radial vector in any fixed plane in the family is a constant, and this constant differs for each plane in the family. This allows us to give equations for planes as follows.

Let $\mathbf{n}$ be a normal vector (orthogonal vector) to a plane and let $\mathbf{r_0}$ be the radial vector for a fixed point in the plane. Then, if $\mathbf{r}$ is the radial vector for an arbitrary point in the plane, we have:

$$\mathbf{n} \cdot (\mathbf{r} - \mathbf{r_0}) = 0$$

Rearranging, we get:

$$\mathbf{n} \cdot \mathbf{r} = \mathbf{n} \cdot \mathbf{r_0}$$

Note that the right side is an actual real number.

If $\mathbf{n} = \langle a, b, c \rangle$ and $\mathbf{r_0} = \langle x_0, y_0, z_0 \rangle$, we get the scalar equation:

$$ax + by + cz = ax_0 + by_0 + cz_0$$

If we define $d = -(ax_0 + by_0 + cz_0)$, we can rewrite the above as:

$$ax + by + cz + d = 0$$

Conversely, any equation of the above sort, where at least one of the numbers a, b, and c is nonzero, gives a plane.

3.2. **Plane parallel to the coordinate axes and planes.** We say that a plane and a line are parallel if either the line lies on the plane or they do not intersect at all.

If $a = 0$, the plane is parallel to the x-axis. If $b = 0$, the plane is parallel to the y-axis. If $c = 0$, the plane is parallel to the z-axis.

If $a = b = 0$, the plane is parallel to the xy-plane. If $b = c = 0$, the plane is parallel to the yz-plane. If $a = c = 0$, the plane is parallel to the xz-plane.

3.3. **Finding the equation of a plane given three points.** To specify a plane, we need to provide at least three points on the plane. Given these three points, we can find the equation of the plane as follows:

- We first take two difference vectors and take their cross product to find a normal vector to the plane: If the points are P, Q, and R, we take the difference vectors PQ and PR and compute their cross product.
- We now use the vector equation, and hence from that the scalar equation, taking any of of the three points P, Q, or R as the basepoint.

Note that if the three points given are *collinear*, then they do not define a unique plane. Rather, any plane through the line joining these three points works. It is no surprise that the above procedure fails at the stage where we need to take cross product, because the cross product turns out to be the zero vector.

3.4. **Intersecting two planes: line of intersection.** Given two planes, the typical case is that they intersect in a line. If we have scalar equations for both planes, then the intersection line can be described by taking the two equations together.

Unfortunately, this pair of two equations together, while it does define a line, is not directly one of the *standard* descriptions of a line.

There are many ways of obtaining the line in standard form. One of these is as follows: first, find normal vectors to the planes. For instance, if the equations for the planes are:

$$\begin{aligned} a_1x + b_1y + c_1z + d_1 &= 0 \\ a_2x + b_2y + c_2z + d_2 &= 0 \end{aligned}$$

Then the normal vectors to these planes are $\langle a_1, b_1, c_1 \rangle$ and $\langle a_2, b_2, c_2 \rangle$. A direction vector along the line of intersection must be perpendicular to *both* these normal vectors, hence, it must be in the line of the *cross product*. Hence, we take the cross product $\langle a_1, b_1, c_1 \rangle \times \langle a_2, b_2, c_2 \rangle$.

Now that we've found the direction vector along the intersection of these planes, we need to find just one point along the intersection and we can then use the point-direction form. One way of finding a point is to set $z = 0$ in both equations and solve the system for x and y (this is assuming that neither is parallel to the xy-plane; otherwise choose some other coordinate).

Note that if the planes are parallel or coincide, then their normal vectors are parallel and thus the cross product of the normal vectors becomes zero. Conversely, the cross product becoming zero means the planes are parallel, so there is a good reason for the line of intersection to not make sense.

3.5. **Intersecting two planes: angle of intersection.** The *angle of intersection* between two planes is the angle of intersection between their normal vectors. As for the line of intersection, we can extract the normal vector from the scalar equation of the planes. To compute the angle of intersection, we use the formula as arc cosine of the quotient of the dot product by the product of the lengths.

3.6. **Intersecting a plane and a line.** Given a plane and a line, we can intersect them as follows: If the plane is given by a scalar equation and the line is given parametrically using a parameter t, then to compute the intersection, we plug in all coordinates as functions of the parameter into the scalar equation for the plane, and solve one equation in the one variable t. After finding the solution t, we plug this into the parametric equation of the line to find the coordinates of the point of intersection.

There are three possibilities:

- The typical case is that we have a linear equation in one variable, and it has a unique solution. In other words, the plane and line intersect at a point.
- Another case is that the equation simplifies to something nonsensical, such as $0 = 1$. In this case, there is no intersection. Geometrically, this means the line is parallel to but not on the plane.
- The final case is that the equation simplifies to a tautology, such as $0 = 0$. In this case, all real t give solutions. Geometrically, this means that the line is on the plane.

3.7. **Distance of a point from a plane.** We will not have much occasion to use this formula, but we note it briefly nonetheless. Given a point with coordinates (x_1, y_1, z_1) and a plane $ax + by + cz + d = 0$, the distance from the point to the plane is given by the formula:

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

FUNCTIONS OF SEVERAL VARIABLES: DEFINITION, EXAMPLES

MATH 195, SECTION 59 (VIPUL NAIK)

Corresponding material in the book: Section 14.1.

What students should definitely get: Definition of function of two variables, concept of graph of such a function, level curves, function of three variables, level surfaces, determining the domain of a function, functions of n variables.

Executive summary

Words ...

(1) A function of n variables is a function on a subset of $\mathbb{R}^n$. We can think of it in three ways: as a function with n real inputs, as a function with input a point in (a subset of) $\mathbb{R}^n$, and as a function with n-dimensional vector inputs. We often write the inputs with numerical subscripts, so a function f of n inputs is written as $f(x_1, x_2, \ldots, x_n)$.
(2) In the case $n = 2$, we often write the inputs as x, y so we write $f(x, y)$. This may be concretely described as an expression in terms of x and y.
(3) The graph of a function $f(x, y)$ of the two variables x and y is the surface $z = f(x, y)$. The xy-plane plays the role of the independent variable plane and the z-axis is the dependent variable axis. Any such graph satisfies the "vertical" line test where vertical means parallel to the z-axis.
(4) The level curves of a function $f(x, y)$ are curves satisfying $f(x, y) = z_0$ for some fixed z_0. These are curves in the xy-plane.
(5) The level surfaces of a function $f(x, y, z)$ of three variables are the surfaces satisfying $f(x, y, z) = c$ for some fixed c.
(6) Domain convention: If nothing else is specified, the domain of a function in n variables given by an expression is defined as the largest subset of $\mathbb{R}^n$ on which that expression makes sense.
(7) We can also define vector-valued functions of many variables, e.g., a function from a subset of $\mathbb{R}^m$ to a subset of $\mathbb{R}^n$.
(8) We can do various pointwise combination operations on functions of many variables, similar to what we do for functions of one variable (both the scalar and vector cases).
(9) To compose functions, we need that the number of outputs of the inner/right function equals the number of inputs of the outer/left function.

Actions ...

(1) To find the domain, we first apply the usual conditions on denominators, things under square roots, and inputs to logarithms and inverse trigonometric functions. For functions of two variables, each such condition usually gives a region of $\mathbb{R}^2$ bounded by a line or curve.
(2) After getting a bunch of conditions that need to be satisfied, we try to find the common solution set for all of these. This involves intersecting the regions in $\mathbb{R}^2$ obtained previously.

1. Functions of two variables

1.1. **Definition.** A *function of two variables* is a function whose domain is a subset of the plane $\mathbb{R}^2$ and whose range is a subset of $\mathbb{R}$. If we denote the domain set by D, then a function f is a rule that assigns to every point $(x, y) \in D$ a real number $f(x, y) \in \mathbb{R}$.

The typical way of describing a function in two variables is by means of an expression that may include both variables. For instance $f(x, y) := \frac{xy}{\sqrt{1+x^2+y^2}}$ is a function of the two variables x and y. The usual remarks from functions of one variable apply:

- The key fact about a function is that *equal inputs give equal outputs*. This remains true for functions of two variables: the value of $f(x_0, y_0)$ depends only on (x_0, y_0).
- We typically define a function using an algebraic expression. When the expression is given without a specified domain, we take the domain to be the largest possible subset of $\mathbb{R}^2$ where the expression makes sense.
- The same function can be given by multiple different-looking expressions (which turn out to be the same upon algebraic simplification). Also, the same expression can give different functions if the domains are taken to be different.

Instead of a function of two variables, we can also consider a function of three variables, four variables, or more. For instance, a function of three variables is defined as a function whose domain is a subset of $\mathbb{R}^3$ and whose range is a subset of $\mathbb{R}$.

Some of the visualization techniques we have start breaking down for more than two input variables, because we're constrained to live in a three-dimensional world, and $2+1=3$. To overcome these limitations, we need to resort to techniques like color, texture, and pattern to visualize such functions.

1.2. **Graphs of functions of two variables.** The convention we'll follow here is that, by default, the dependent variable (the output of the function) will be denoted z and the independent variables will be noted x and y, so we have $z = f(x, y)$.

The *graph* of such a function is thus the *surface* given by the equation $z = f(x, y)$. Comparing with the way we usually think of functions in one variable as $y = f(x)$, we have the following analogies:

- The xy-plane plays the role that the x-axis played in the past, as the space of possible *inputs* (the actual domain may be a subset rather than the whole plane).
- The z-axis plays the role that the y-axis played earlier, as the set of all possible *outputs*.
- To read the function value at a point in the xy-plane, we note the corresponding value of the z-coordinate so that we hit a point in the surface.
- Graphs of functions in this sense satisfy the *vertical line test* where "vertical" here means parallel to the z-axis.

1.3. **Two inputs, one output, much repetition, level curves.** In the midst of one-variable calculus, we introduced a notion of *one-to-one function*. A one-to-one function is a function with the property that no two different input values give the same output, i.e., *equal outputs must have arisen from equal inputs*.

When we're dealing with functions where the number of input variables is greater than the number of outputs, however, being one-to-one is highly unlikely (in fact, impossible for continuous functions defined on any big enough subset). Rather, for any given output, it is highly likely that the number of inputs isn't just more than one, it is itself a curve.

Let's think about this. To find all the input pairs (x, y) which map to a fixed value $z = z_0$, we need to solve:

$$f(x, y) = z_0$$

In other words, we need to find the (x, y)-coordinate of the intersection of the surface $z = f(x, y)$ and the plane $z = z_0$.

This is an intersection of two surfaces, and is expected to be a curve. Thus, the general collection of points which map to a given z-value is a *curve* in the xy-plane (the actual intersection of surfaces will be the same curve translated to the $z = z_0$ plane).

These curves are called the *level curves* of the function. We could choose to depict all the level curves in the xy-plane. *The picture of the level curves, along with the labels of the levels by the z-value*, together describe the function completely.

Note that for z-values which are outside the range, the corresponding level curve is empty. For z-values in the boundary of the range, the level curve may not be a curve at all but may be a bunch of points.

For instance, consider the function:

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

The graph of this is the surface $z = x^2 + y^2$. It lies in the $z \geq 0$ region. For a fixed value $z_0 > 0$, the corresponding level curve is a circle centered at the origin and with radius $\sqrt{z_0}$. For $z_0 = 0$, the corresponding level "curve" is the single point $(0, 0)$, and for $z_0 < 0$, the corresponding level curve is empty.

1.4. **Color, a new dimension.** Level curves suggest a new, fascinating approach to depict functions with two inputs and one output using *pictures in the plane*. The key idea is to use a dimension such as color or texture to code for the function value. Specifically, we first map the output values to a color spectrum. For instance, if the range of a function f is $[0, 1]$, we may map $[0, 1]$ linearly to the wavelengths for the visible color spectrum, setting 0 for the violet color and 1 for the red color.

Next, for any point (x, y), we "color" that point with the color value associated with $f(x, y)$. In particular, this means that the "single color curves" are precisely the "level curves" for the function. The more reddish the color, the larger the function value at the point. The more violetish the color, the smaller the function value at that point.

In other words, instead of choosing a physical or height axis for the z-axis, we choose a *color axis*.

1.5. **Finding the domain of a function.** As indicated earlier, the largest possible domain of a function of two variables is the set of all possible input pairs on which the function makes sense. In particular, this means things like:

- Any denominator should be nonzero
- Any thing under a square root should be nonnegative
- Any thing under a square root sign *in the denominator* should be positive
- Any thing inside a logarithm should be positive

and more similar stuff.

What each of these conditions do typically is remove subsets from the plane, and what's left is the domain. For instance, consider the function:

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

Here, the denominator factors as $(y + 2x)(y - 2x) = 0$, giving the lines $y = -2x$ and $y = 2x$. These are the two forbidden lines – everything else is the domain. The domain is thus the *complement of the union of these lines in the plane*.

Similarly, consider:

$$f(x, y) := \frac{x + \ln(x - y)}{\sqrt{x^2 + y}}$$

The domain here must satisfy both the conditions $x > y$ and $x^2 + y > 0$, so it is the intersection of the subsets satisfying these two conditions. The condition $x > y$ is the half-plane region to the left/down of the $y = x$ line. The condition $x^2 + y > 0$ is the region above the parabola $y = -x^2$. The intersection of these regions gives a particular subset of the plane.

See more examples of this sort in the book.

2. Functions of more than two variables

2.1. **Functions of three variables: depiction.** Most of the ideas mentioned above have analogues for functions of three variables. The key difference is that there is no fourth dimension in our three-dimensionally constrained visualization capacities to make the graph of the function. Nonetheless, we can still study the behavior in a similar manner.

First, we can talk of *level surfaces* (in place of level curves). For a function $f(x, y, z)$ of three variables, the level surface corresponding to a function value c is the set of all input points which get mapped to c under f. This typically looks like a surface – we have one equation in three variables, so the solution space is expected to be two-dimensional. In degenerate cases, it may be a curve or even a point or bunch of points.

Second, in order to depict such functions, we can again use the *color axis* trick – determine the function value by the choice of color at the point. This allows us to *depict functions of three variables in three-dimensional space*. Of course, we cannot see all of the function at the same time, because we can only see

the outer regions of the domain (they block the sight to the inner regions). But with sophisticated slicing and visualization tools, we can get to see all parts of the function.

2.2. **Functions of n variables.** A function of n variables is a function from a subset of $\mathbb{R}^n$ to $\mathbb{R}$. In other words, the input of this function is a point in $\mathbb{R}^n$ (possibly restricted to a subset) and the output is a real number.

There are three ways of thinking of a function of n variables:

(1) As a function of n real variables $x_1, x_2, \ldots, x_n$.
(2) As a function of a single point variable $(x_1, x_2, \ldots, x_n)$.
(3) As a function of a single vector variable $\mathbf{x} = \langle x_1, x_2, \ldots, x_n \rangle$.

All these perspectives turn out to be useful.

2.3. **Multiple inputs – what's the big deal?** We earlier dealt with functions that take one real input and give more than one real output – we called these *vector-valued functions*. We found that vector-valued functions are no big deal – to handle a vector-valued function, we simply look at each of the outputs separately.

The story is more complicated for multiple inputs. It is *not* possible to look at each input separately, because the inputs interact with each other in msyterious ways and it is hard to disentangle them.

Here's an analogy that may help you understand this. If you write a press release, it is easy to send it to a hundred different news wires, put it up on your website, and print copies to put on loads of bulletin boards. It may require more effort than simply putting it up on your website, but the complexity of the task does not increase.

On the other hand, if, in order to write the press release, you need to coordinate the activities of a hundred people, bring them together in a meeting, and have them interchange ideas to come up with the wording, that's a fundamentally more challenging task.

That's because multiple *inputs* are a lot harder than multiple *outputs*.

3. Operations on functions of many variables

3.1. **Pointwise combination.** It is possible to do pointwise addition, subtraction, and multiplication on real-valued functions of more than one variable. For instance, if f and g are functions of two variables, then $f + g$ is also a function of two variables, defined as:

$$(f + g)(x, y) = f(x, y) + g(x, y)$$

Similarly, we can define the pointwise difference, pointwise product, and pointwise quotient.

The usual comments on domain apply: the domain for any pointwise sum, difference, or product is the intersection of the domains of the functions. With pointwise quotient, we also need to exclude points where the denominator function takes the value zero.

3.2. **Vector-valued functions: multiple inputs and multiple outputs.** Before moving to the next topic, i.e., composition, we consider the concept of a *vector-valued function of multiple variables*. Here, we simultaneously have multiple inputs (i.e., the domain is a subset of $\mathbb{R}^m$) and multiple outputs (i.e., the range is a subset of $\mathbb{R}^n$).

For vector-valued functions of multiple variables, we can do the same types of pointwise combinations as for vector-valued functions of one variable: pointwise addition, subtraction, scalar-vector multiplication, dot product, and (when the output vector sare three-dimensional) cross product.

3.3. **Composition.** If f and g are both real-valued functions of two variables, it does not make direct sense to talk of $f \circ g$. The problem is that the output of g is a single real number, and this cannot be fed as the input to f. Please remember: *with pointwise combination, like combines with like. With composition, on the other hand, what matters is that the output of one should feed into the input of the other.*

Here is the kind of composition we can do: If g is a vector-valued function with m inputs and n outputs (i.e., the output is a n-dimensional vector), and f is a vector-valued function with n inputs and p outputs, then $f \circ g$ makes sense as a vector-valued function with m inputs and p outputs. If this reminds you of matrix multiplication, that is for good reason! Unfortunately, the details are beyond the current scope.

4. Real world examples

4.1. **Quantity demanded: in the microeconomic realm.** We consider an example from microeconomics to illustrate the notion of functions of many variables. The book has a number of other examples that you can also review.

According to standard microeconomic theory, the "quantity demanded" for a good by a household is determined by six kinds of things:

- The unit price of the good
- The unit prices of substitute goods
- The unit prices of complementary goods
- The income/wealth of the household
- The tastes and preferences of the household
- Expectations regarding future prices

Each of these things in turn may involve multiple sub-items. The unit price of the good is a single real number, but the unit prices of substitute and complementary goods could be as many real numbers as the number of such substitute and complementary goods. For instance, if we are looking at the quantity demanded for whole wheat bread, we may identify substitute goods such as pita bread and baguettes, and complementary goods such as butter and jam. Thus, we already have five input variables: the price of whole wheat bread, the price of pita bread, the price of baguettes, the price of butter, and the price of jam.

The income of the household is a single variable, but there may be other income-related variables of interest. The tastes and preferences may be modeled using many measurements, such as a scale for the taste for novelty, a scale for the preference for "healthy" food, a scale for the taste of sweetness, etc. Expectations regarding future prices may again be modeled using many real inputs.

Thus, we see that the quantity demanded for a good is a function of a potentially large number of input variables, even in an ideal microeconomic world of rational decision making. The kinds of questions that we are interested in, and that will motivate further mathematical analysis, include:

- When these "determinants of demand" are perturbed slightly, does the quantity demanded also respond by changing only slightly, or does it experience sudden, massive shifts? The "changing only slightly" is the key motivation behind the concept of continuity.
- If we fix all the determinants of demand except one, how does the quantity demanded vary with changes in that one parameter? Does it increase or decrease? Can we differentiate this function? In economics, keeping everything else fixed is termed *ceteris paribus*. The related ideas in multivariable calculus are *separate continuity in the variables* and *partial derivatives*.

4.2. **Feels like temperature: of wind and water.** Page 892, Section 15.1, Example 2 in the book talks of the *wind chill index* and the "feels like" temperature. Basically, when it is cold and windy, it appears colder than it actually is, and weather services have constructed tables that compute the "feels like" temperature as a function of the actual temperature and the wind speed.

There is also a *temperature-humidity index* or humidex, which measures the perceived air temperature as a function of actual temperature and humidity. In general, the more the humidity, the higher the perceived temperature. See Exercise 2, Section 15.1, Page 902 and also the discussion at the beginning of Section 15.3, Page 914 of the book.

These examples are not of direct relevance for us, but they may be fun to read about (and you can learn more on the Internet) since they'll give you a better perspective on interpreting weather forecasts.

4.3. **Functions on physical locations/geographic areas.** We consider another related concept where the book offers some examples: functions whose inputs are physical locations, and which output quantities that can be measured at those locations. For instance, we can think of temperature as a function defined at every point in physical space. It can thus be modeled as a function from a subset of $\mathbb{R}^3$ to $\mathbb{R}$. If we are interested in measuring the "surface temperature" on the earth, then this can be modeled as a function from the subset of $\mathbb{R}^3$ comprising the surface of the earth, to $\mathbb{R}$.

The surface of the earth is an interesting domain upon which to define a function. Although it lives in a three-dimensional world, the surface itself is a two-dimensional construct. This means that it is possible to parameterize the surface of the earth using a *two-parameter formulation* (something we will explore in

more detail later). For those of you who remember high school geography, one way of parameterizing the surface of the earth using two coordinates is via the *latitude* and *longitude*. Once we choose this kind of parameterization, then a function whose domain is the surface of the earth can be viewed as a function of two variables (the latitude and longitude coordinates) and hence can be viewed as a function on a subset of $\mathbb{R}^2$ (with some caveats).

With this in mind, here is some terminology:

- If the function we are considering is temperature, then the level curves (i.e., the curves corresponding to fixed temperature values) are called *isotherms* or *isothermal curves*.
- If the function we are considering is atmospheric pressure, then the level curves are called *isobars*.

4.4. **Cobb-Douglas production function.** This is an important idea but we will postpone discussion of this until later, once we have seen partial derivatives.

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 $x = \langle x_1, x_2, \ldots, x_n \rangle$ 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 $|f(x) - L| < \varepsilon$.

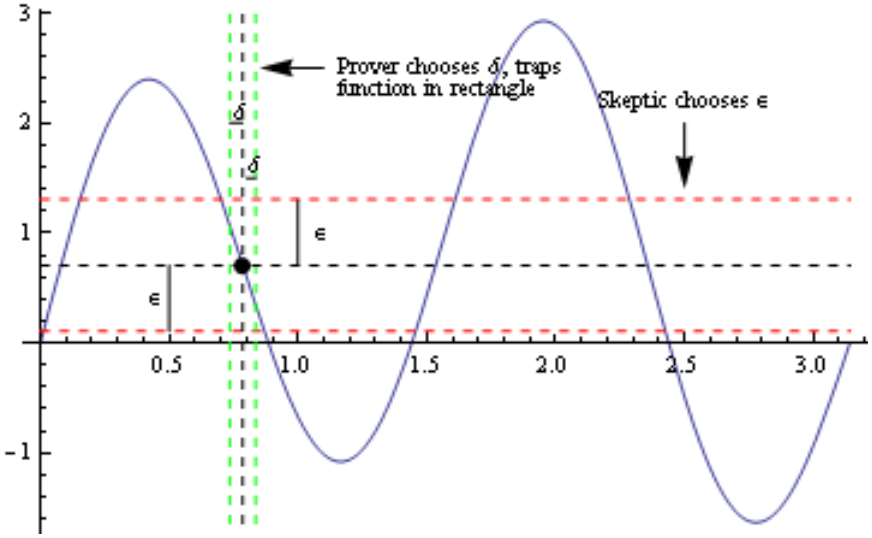
The way I think about the definition is in terms of a *cage* (or *trap*). And the reason why we need this notion of a cage or trap is precisely to avoid these kinds of oscillations that give rise to multiple limits. So, here is the formal definition:

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 $x = \langle x_1, x_2, \ldots, x_n \rangle$ *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 $x_1, x_2, \ldots, x_n$ 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 $x_1, x_2, \ldots, x_n$ 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:

$$f(x,y) := \begin{cases} \frac{xy}{x^2+y^2}, & (x,y) \neq (0,0) \\ 0, & (x,y) = (0,0) \end{cases}$$

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, $\ldots, x_n$ is a function $F(x_1, x_2, \ldots, x_n)$ with the property that the *total degree* of x_1, x_2, $\ldots, x_n$ 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, $\ldots x_n$ 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):

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

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:

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

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

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

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:

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

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.

PARTIAL DERIVATIVES

MATH 195, SECTION 59 (VIPUL NAIK)

Corresponding material in the book: Section 14.3.

What students should definitely get: Definition and computation techniques for first and higher partials, both at a specific point and as general expressions, statement of Clairaut's theorem.

What students should hopefully get: Cases of particular interest in partial derivative computation, interpretation of signs of partial derivatives.

Executive summary

Words ...

(1) The partial derivative of a function of many variables with respect to any one variable is the derivative with respect to that variable, keeping others constant. It can be written as a limit of a difference quotient, using variable letter subscript (such as $f_x(x,y)$), numerical subscript based on input position (such as $f_2(x_1,x_2,x_3)$), Leibniz notation (such as $\partial/\partial x$).

(2) In the separate continuity-joint continuity paradigm, partial derivatives correspond to the "separate" side. The corresponding "joint" side notion requires linear algebra and we will therefore defer it.

(3) The expression for the partial derivative of a function of many variables with respect to any one of them involves all the variables, not just the one being differentiated against (the exception is additively separable functions). In particular, the *value* of the partial derivative (as a number) depends on the values of all the inputs.

(4) The procedure for partial derivatives differs from the procedure used for implicit differentiation: in partial derivatives, we assume that the other variable is independent and constant, while in implicit differentiation, we treat the other variable as an unknown (implicit) function of the variable.

(5) We can combine partial derivatives and implicit differentiation, for instance, $G(x,y,z) = 0$ may be a description of z as an implicit function of x and y, and we can compute $\partial z/\partial x$ by implicit differentiation, differentiate G, treat z as an implicit function of x and treat y as a constant.

(6) By iterating partial differentiation, we can define higher order partial derivatives. For instance f_{xx} is the derivative of f_x with respect to x. For a function of two variables x and y, we have four second order partials: f_{xx}, f_{yy}, f_{xy} and f_{yx}.

(7) Clairaut's theorem states that if f is defined in an open disk surrounding a point, and both mixed partials f_{xy} and f_{yx} are jointly continuous in the open disk, then $f_{xy} = f_{yx}$ at the point.

(8) We can take higher order partial derivatives. By iterated application of Clairaut's theorem, we can conclude that under suitable continuity assumptions, the mixed partials having the same number of differentiations with respect to each variable are equal in value.

(9) We can consider a partial differential equation for functions of many variables. This is an equation involving the function and its partial derivatives (first or higher order) all at one point. A solution is a function of many variables that, when plugged in, satisfies the partial differential equation.

(10) Unlike the case of ordinary differential equations, the solution spaces to partial differential equations are huge, usually infinite-dimensional, and there is often no neat description of the general solution.

Pictures ...

(1) The partial derivatives can be interpreted as slopes of tangent lines to graphs of functions of the one variable being differentiated with respect to, once we fix the value of the other variable.

Actions ...

(1) To compute the first partials, differentiate with respect to the relevant variable, treating other variables as constants.

(2) Implicit differentiation for first partial of implicit function of two variables, e.g., z as a function of x and y given via $G(x, y, z) = 0$.
(3) In cases where differentiation formulas do not apply directly, use the limit of difference quotient idea.
(4) To calculate partial derivative at a point, it may be helpful to first fix the values of the other coordinates and then differentiate the function of one variable rather than trying to compute the general expression for the derivative using partial differentiation and then plugging in values. On the other hand, it might not.
(5) Two cases of particular note for computing partial derivatives are the cases of additively and multiplicatively separable functions.
(6) To find whether a function satisfies a partial differential equation, plug it in and check. Don't try to find a general solution to the partial differential equation.

Econ-speak ...

(1) Partial derivatives = marginal analysis. Positive = increasing, negative = decreasing
(2) Second partial derivatives = nature of returns to scale. Positive = increasing returns (concave up), zero = constant returns (linear), negative = decreasing returns (concave down)
(3) Mixed partial derivatives = interaction analysis; positive mixed partial derivative means complementary, negative mixed partial derivative means substitution
(4) The signs of the first partials are invariant under monotone transformations, not true for signs of second partials, pure or mixed.
(5) Examples of quantity demanded, production functions.
(6) Cobb-Douglas production functions (see section of lecture notes and corresponding discussion in the book)

1. Partial derivatives: introduction

If there's one concept that is really unique to multivariable calculus, and really important at a conceptual level in applications of mathematics to the social sciences, it is the concept of partial derivatives. This is particularly true in the case of economics, since a key ingredient of economic thinking is *marginal analysis*, which is just another way of saying *partial derivatives*. The truly revolutionary idea of *mixed partial derivatives* is fairly important, and serves as a conceptual lens for understanding the interaction of multiple variables.

1.1. **Separate versus joint, and partial derivatives.** Recall, from our discussion of limits and continuity for functions of many variables, the concept of *separate* versus *joint*. Separate continuity referes to continuity in each of the variables in isolation, where we are allowed to move only one variable at a time. Joint continuity refers to continuity where we are allowed to simultaneously move more than one variable, i.e., it refers to robustness under simultaneous perturbations of all the variables together.

Analogous to that, we have the notion of *separate differentiation* and *joint differentiation*. Partial derivatives corresponds to the notion of *separate differentiation*. The corresponding notion in joint differentiation is called the *total derivative*. The notion of total derivative, however, requires some knowledge/understanding of linear algebra, which we cannot currently assume. Hence, we restrict our attention/analysis to partial derivatives.

1.2. **Definition of partial derivative.** For simplicity, we restrict formulations and notation to functions of two variables, where we denote the two input variables as x and y. The same ideas apply to functions of more variables.

For a function f with input variables (x, y), we define the *partial derivative* of f with respect to x at the point (a, b), denoted $f_x(a, b)$, as the number:

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

In other words, it is the derivative with respect to the *first coordinate* keeping the second coordinate value fixed at b. In other words, it is the value $g'(a)$ where $g(x) := f(x, b)$.

Simiarly, we define the partial derivative of f with respect to y at the point (a, b), denoted $f_y(a, b)$, as the number:

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

In other words, it is the derivative with respect to the *second coordinate* keeping the first coordinate value fixed at a. In other words, it is the value $g'(b)$ where $g(y) := f(a,y)$.

Note that since the above are calcuating partial derivatives at a *fixed* point, they give actual numbers. We could, however, now make a and b variable, and relabel them x and y. In this case $f_x(x,y)$ and $f_y(x,y)$ are now *both* functions of *both* variables x and y. The formulas look as follows:

$$\begin{aligned} f_x(x,y) &:= \lim_{h\to 0} \frac{f(x+h,y)-f(x,y)}{h} \\ f_y(x,y) &:= \lim_{h\to 0} \frac{f(x,y+h)-f(x,y)}{h} \end{aligned}$$

1.3. **The expression and value of partial derivative depend on both variable values.** One possible misconception is that the partial derivative with respect to a particular variable depends only on that variable. This is *not true*. The expression for the partial derivative with respect to x potentially depends on both x and y. What this means is that the value of the partial derivative depends on the location of the point, even the *other* coordinate.

The exception is the case of *additively separable functions*. In other words, if we can write $F(x,y)$ as $f(x)+g(y)$ where f is a function of one variable and g is a function of one variable. Then, $F_x(x,y) = f'(x)$ and is independent of y and $F_y(x,y) = g'(y)$ and is independent of x.

1.4. **Numerical subscripts for partials.** For functions of two variables, we can use the letter x for the first input and the letter y for the second input. This does not naturally generalize to functions with more inputs. Hence, there is an alternate convention: we use the subscript i to denote partial derivative with respect to the i^{th} input coordinate. Thus, $f_1(x,y)$ stands for $f_x(x,y)$ and $f_2(x,y)$ stands for $f_y(x,y)$.

Note that subscripts are often used in other contexts, so just because you see a subscript being used, do not blindly assume that it refers to a partial derivative. Context is everything.

1.5. **Leibniz-like notation for partial derivatives.** Recall that the Leibniz notation for ordinary differentiation uses the d/dx operator. For partial differentiation, we replace the English letter d by a letter ∂, so $f_x(x,y)$ is also denoted as $\frac{\partial}{\partial x}(f(x,y))$ and $f_y(x,y)$ is also denoted as $\frac{\partial}{\partial y}(f(x,y))$. In particular, if $z = f(x,y)$, we can write these partial derivatives as $\partial z/\partial x$ and $\partial z/\partial y$ respectively.

1.6. **Rule for computing partial derivatives.** Partial derivatives are computed just like ordinary derivatives – we just treat all the other input variables as constant. So, for instance:

$$\frac{\partial}{\partial x}(x^2+xy+y^2) = \frac{\partial}{\partial x}(x^2) + y\frac{\partial}{\partial x}(x) + 0 = 2x+y$$

1.7. **Partial versus implicit differentiation.** In single variable calculus, you came across a concept called *implicit differentiation*, for which we used the letter d. With implicit differentiation, we start with an expression that involves both x and y, and then differentiate with respect to x. *However, for implicit differentiation, we do not assume that y is a constant.* Rather, we assume that y is an unknown implicit function of x, so our final expression involves dy/dx, which we *do not convert to zero*.

Partial differentiation with respect to x is different in that it does not assume y to be dependent on x – rather it assumes y is a constant, and treats y as such. However, if we have already done a calculation of the implicit derivative of an expression $f(x,y)$ with respect to x, we can calculate the partial derivative by simply setting $dy/dx = 0$ wherever it appears in the expression for the implicit derivative.

For instance, under implicit differentiation:

$$\frac{d}{dx}\sin(x+y+y^2) = \left[1+(1+2y)\frac{dy}{dx}\right]\cos(x+y+y^2)$$

To compute the partial derivative, we simply set $dy/dx = 0$ in the above, and get:

$$\frac{\partial}{\partial x}\sin(x+y+y^2) = \cos(x+y+y^2)$$

1.8. **Partial derivatives plus implicit differentiation.** In the previous subsection, we contrasted partial differentiation of $F(x,y)$ with respect to x and implicit differentiation. The partial derivative can be obtained from the implicit derivative by setting $dy/dx = 0$, i.e., assuming that the variables have no dependence on each other.

When we have more than two variables, however, we may combine the ideas of partial and implicit differentiation. For instance, we may have an expression $G(x,y,z) = 0$ to describe z as an *implicit* function of the variables x and y. We now want to determine the partial derivatives $\partial z/\partial x$ and $\partial z/\partial y$. In order to do this, we start with:

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

To find $\partial z/\partial x$, we differentiate $G(x,y,z)$ partially with respect to x, assuming the following: (i) y is treated as a constant, (ii) z is treated as an implicit function of x and y, so its partial derivative with respect to x is denoted $\partial z/\partial x$.

For instance, consider z as an implicit function of x and y:

$$xy + \sin(xz) = \cos(y+z)$$

Doing implicit differentiation, we get:

$$y + \cos(xz)\left(x\frac{\partial z}{\partial x} + z\right) = -\sin(y+z)\frac{\partial z}{\partial x}$$

We can rearrange, collect terms for $\partial z/\partial x$, and get an expression for it in terms of the three variables x, y, and z.

2. Conceptual and geometric interpretations

2.1. **What's a partial derivative, conceptually?** A given output typically depends on multiple inputs. For instance, in classical microeconomic theory, the quantity of a commodity demanded by a household is considered a function of six types of variables: the unit price, the tastes and preferences of the household, the income/wealth of the household, the prices of substitute goods, the prices of complementary goods, and expectations regarding future prices. Each of these "types" of variables may itself comprise multiple variables (for instance, there may be many different complementary and substitute goods), so the actual quantity demanded may be modeled as a function of a much larger number of variables. We may wish to study the effect of *just one of these* variables on the quantity demanded, keeping the other variables fixed. This is known as *ceteris paribus* in economics. The partial derivative is a key tool in the study of this kind of relationship.

If you have studied classical microeconomics in a quantitative sense, you may have encountered the concept of *price elasticity*. The price elasticity of demand is sort of like the partial derivative of the quantity demanded with respect to the price. However, in order to achieve dimensionlessness, we do not simply take the partial derivative but instead divide the partial derivative by the quantity-price ratio, i.e., we take the partial derivative of the *logarithm* of the quantity demanded with respect to the *logarithm* of the unit price of the good, keeping all the other determinants of demand constant. In symbols:

$$\text{Price elasticity of demand} = \frac{\partial q/\partial p}{q/p} = \frac{\partial(\ln q)}{\partial(\ln p)}$$

The quantity as computed here turns out to be negative if the good satisfies the law of demand, and it is customary to take the absolute value when giving numerical values.

Similarly, we can define the *cross price elasticity* with respect to a complementary or substitute good as the derivative of the logarithm of quantity demanded for a particular good with respect to the unit price of a particular complementary or substitute good. There is also a related notion of *income elasticity of demand*.

Note that the price elasticity and cross price elasticity values depend not only on the price, but also on the values of other determinants of demand. In other words, if we want to know the value of the price

elasticity of demand for bread at a particular price of bread, we *also* need to specify all the other values of determinants of demands and only then can we compute the price elasticity. This is the same observation we made earlier: *the value of the partial derivative with respect to one input depends on the values of all inputs at the point of evaluation.*

To take an example that makes this context clear: we know that the price elasticity of demand is fairly high at prices close to those of substitutes, because the substitution effect operates most strongly. At prices much higher than the price of the substitute, demand is unfiformly lower, and at prices much lower than the price of the substitute, demand is uniformly higher.

Thus, if we were to change the price of a substitute good, that would affect the price ranges for which price elasticity of demand is high.

The *extent* to which changes in the values of one input affect the partial derivative with respect to another variable can be captured using *second-order mixed partial derivatives*, something we will see in a little while.

Derivatives cut both ways! Many "laws of economics" such as the law of demand and law of supply basically make an assertion about the sign of a partial derivative. For instance, the *law of demand* states that, ceteris paribus, as the unit price falls, the quantity demanded rises (or at least stays the same). This can be reframed as saying that the partial derivative of quantity demanded with respect to price is negative (or non-positive).

Surprisingly, people who lack calculus skills often believe that derivatives cut one way. For instance, they may agree with the statement that "as the price rises, the quantity falls" but disagree with the statement "as the price falls, the quantity rises." For instance, people are more likely to agree with the statement "as the price of accidents rises, people will be more careful" rather than the equivalent statement "as the price of accidents falls, people will be less careful." Now, there are ways of finding elusive wisdom in such affronts to calculus (for instance, differing values of price elasticity across the demand curve, exceptional demand curves, or long-run technological progress/secular trends) but at face value, if you agree with one statement, you should agree with the other.

2.2. **What's a partial derivative, geometrically?** Consider a function $z = f(x, y)$ of two variables. Consider the graph of this function, which is the surface $z = f(x, y)$. At a point (a, b):

- The value $f_x(a, b)$ is the slope (z/x-type) of the tangent line to the curve we obtain by intersecting the surface with the plane $y = b$ (which is parallel to the xz-plane). This intersection is basically the graph of the function $x \mapsto f(x, b)$.
- The value $f_y(a, b)$ is the slope (z/y-type) of the tangent line to the curve we obtain by intersecting the surface with the plane $x = a$. This intersection is basically the graph of the function $y \mapsto f(a, y)$.

We will explore this geometric interpretation in the next lecture.

3. Higher partial derivatives

3.1. **Basic definitions.** Any of the partial derivatives of a function of n variables is itself a function of the same n variables, so it can be differentiated again. We use the following basic notation for second partial derivatives of $z = f(x, y)$:

$$\begin{aligned}
(f_x)_x = f_{xx} = f_{11} &= \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial x^2} = \frac{\partial^2 z}{\partial x^2} \\
(f_x)_y = f_{xy} = f_{12} &= \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial y \partial x} = \frac{\partial^2 z}{\partial y \partial x} \\
(f_y)_x = f_{yx} = f_{21} &= \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 z}{\partial x \partial y} \\
(f_y)_y = f_{yy} = f_{22} &= \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial^2 f}{\partial y^2} = \frac{\partial^2 z}{\partial y^2}
\end{aligned}$$

These are all called *second order partial derivatives* or *second partial derivatives* or (in short) *second partials* because they involve two differentiations. The second partials f_{xy} and f_{yx} are termed *mixed partial derivatives* or *mixed partials.*

3.2. **Clairaut's theorem: equality of mixed partials.** Clairaut's theorem states that if f is defined around a point (a, b) (i.e., in an open disk containing the point (a, b)) and if the mixed partials f_{xy} and f_{yx} are both continuous around (a, b), then:

$$f_{xy} = f_{yx}$$

We will get a better conceptual feel of this in a short while.

3.3. **Additively and multiplicatively separable functions.** Suppose $F(x, y)$ can be written as a sum $f(x) + g(y)$, i.e., it is *additively separable* in terms of functions of x and y. Then, we have $F_x(x, y) = f'(x)$ and $F_y(x, y) = g'(y)$. Moreover:

- All pure higher order derivatives with respect to x are the corresponding ordinary derivatives of f.
- All pure higher order derivatives with respect to y are the corresponding ordinary derivatives of g.
- All mixed partial derivatives are zero.

Another case of interest is where $F(x, y)$ can be written as a product $f(x)g(y)$, i.e., it is *multiplicatively separable* in terms of functions of x and y. Then, we have $F_x(x, y) = f'(x)g(y)$, $F_y(x, y) = f(x)g'(y)$, and $F_{xy}(x, y) = f'(x)g'(y)$. More generally, any mixed partial in which there are a many x's and b many y's is $f^{(a)}(x)g^{(b)}(y)$.

3.4. **Higher partials.** We can take higher order partials for functions in two variables. For instance, the third order partials are f_{xxx}, f_{xxy}, f_{xyx}, f_{xyy}, f_{yxx}, f_{yxy}, f_{yyx}, and f_{yyy}. This is a total of 8 possibilities. Clairaut's theorem, however, can be used to show that if the partials are all continuous, then what matters is only the number of x's and the number of y's – the sequence of differentiation does not matter. After this identification, there are four partials: f_{xxx}, f_{xxy}, f_{xyy}, and f_{yyy}.

3.5. **Functions of more than two variables.** The same notation and ideas apply. In particular, Clairaut's theorem reveals that, if the mixed partials are continuous around a point, then mixed partials that involve differentiation in each variable the same number of times must be equal. Thus, for instance, if $w = f(x, y, z)$ has all sixth order mixed partials continuous, then $f_{xxxyyz} = f_{yxyxzx}$.

3.6. **First versus higher partials: invariance under monotone transformations.** The first partial derivatives satisfy an important property: the sign of the first partial derivative is invariant under monotone transformations of the variables, such as replacing a variable by its logarithm, or its square (for positive variables) or its cube, etc. This means that when we say that A is increasing relative to B, the statement would still be true if we replaced A by $\ln A$ and B by B^3. In other words, the signs of the first partials depend only on the *ordinal scale* and not on the actual distance ratios involved.

On the other hand, the sign of the second and higher partials is not invariant under monotone transformations. If we replace the output function by the logarithm of the output function, the sign of the mixed partial may go from negative to positive or to zero. *While first partials do not depend on the choice of a correct measurement scale, higher partials are highly sensitive to the choice of measurement scale.*

4. Real world applications

4.1. **Higher partials, conceptually.** The mixed partial derivative of a function is an extremely important idea. It measures the sensitivity of the nature of how the function changes with respect to one variable as we change another variable. Let's consider some examples.

Suppose that, in order to achieve some output, you need two types of inputs: labor L and capital K. Your output is given by a function $f(L, K)$. The *marginal productivity* of labor is defined (roughly) as the partial derivative of output with respect to labor, i.e., the partial derivative $f_L(L, K)$ (again, for some definitions, we may choose to take logarithms before taking partial derivatives in order to obtain a dimensionless quantity, however, here for convenience we do not use logarithms). The marginal productivity of capital is defined as the partial derivative $f_K(L, K)$. Marginal productivity basically answers the question: if I increase the given factor of production ever so slightly, then to what extent is output affected? Or equivalently, if I decrease the given factor of production ever so slightly, then to what extent is output affected?

We now note the significance of the three second partials:

- The second partial f_{LL} measures whether the marginal product of labor is increasing or decreasing, and by how much. If $f_{LL} > 0$, that implies that labor is subject to *increasing returns*. This means that the more labor you put in, the more attractive it is to put in each additional unit of labor. On the other hand, if $f_{LL} < 0$, that implies that labor is subject to *diminishing returns*, i.e., the gains from adding additional units of labor becomes less (though they may still be positive) as we add more labor. Generally, we see that for a fixed value of K, labor is eventually subjected to diminishing returns, and possibly eventually even negative returns.

  For instance, if each worker sitting on a machine can produce 2 units (but workers take coffee breaks, so machines can be used to a slight extent by more than one worker), and if there are 90 computers, increasing the number of workers from 80 to 90 might increase output from 160 units to 180 units, but increasing the number of workers from 90 to 100 may produce very little increase, because the only way to squeeze more output is to have some workers use the machines while the others are on coffee break. This may, for instance, boost output only to 185 units. In this case, we see diminishing returns once the capital utilization starts getting complete.

  If $f_{LL} = 0$, we talk of constant returns to labor.
- We can similarly study f_{KK}, and talk of increasing returns to capital, decreasing returns to capital, and constant returns to capital.
- Finally, consider the mixed partial derivative $f_{LK} = f_{KL}$ (under the assumptions of Clairaut's theorem). This is basically addressing the question: how does the marginal productivity of labor get affected if we add more capital? Or put another way, how does the marginal productivity of capital get affected if we add more labor?

  If $f_{LK} > 0$, we say that *labor and capital are complementary inputs*. This is the typical situation, and it is not hard to think of examples. For instance, adding more machines increases the number of people who can be productively employed to operate the machines. This is particularly true if existing capital is close to being fully exploited. Put another way, adding more people makes it more worthwhile to obtain new machines for these people to operate.

  If $f_{LK} < 0$, we say that *labor and capital substitute for each other*. This means that adding more workers *reduces* the marginal product of capital. To think of this kind of situation, imagine that some of the tasks can be performed *either* by people or by machines. For instance, buying a machine that can generate calculus lectures reduces the need for, and the marginal product of, hiring a calculus lecturer. Or, buying a machine with payroll software reduces the amount of labor that the payroll manager needs to put in, and perhaps renders his job redundant.

Recall from what we learned of functions of one variable that increasing returns means the graph studying dependency only on that variable is concave up, constant returns means the graph studying dependency only on that variable is linear, and decreasing returns means the graph studying dependency only on that variable is concave down.

Note that the expressions f_{LL}, f_{LK}, and f_{KK} are not fixed numbers – they are themselves functions of L and K. In particular, this means that they need not have constant signs – the sign of f_{LL} may be positive for some values of (L, K) but negative for others. Similarly, the value of f_{LK} may be positive for some values of (L, K) and negative for others.

Interestingly, in the short run, the story is largely one of diminishing returns and a mix of complementary and substitution effects. In the somewhat longer run, on the other hand, the story is one of increasing returns and complementary effects. The chief reason is that over the longer run, it is possible to reconfigure the modes of production (through technological innovation, both low-tech and hi-tech) in order to better exploit the synergies between different resources. In the short run, for instance, machines may put people out of work by substituting for them, but in the longer run, people acquire new skills that complement those of the machines.[1]

Clairaut's theorem and intuition. Clairaut's theorem, although not counter-intuitive, is not entirely intuitive either. However, to get a good understanding of the interaction of multiple variables, you should try to make this a part of your intution.

[1]For instance, farmers put out of work due to greater mechanization of agriculture may end up having to settle for being computer engineers.

In our context, for instance, it says that the effect on the marginal product of labor of adding one unit of capital is the same as the effect on the marginal product of capital of adding one unit of labor. In other words, the extent to which labor makes capital more (or less) valuable is the same as the extent to which capital makes labor more (or less) valuable.

4.2. **A marriage of derivatives: like begets like, or opposites attract?** Turning from production to personal life, let's consider the question of marriage. Simplifying from the realities of the messy world, assume that there are two sexes (male and female), each person from one sex wants to marry a person from the other sex, and there is an equal number of people of each sex. Each person has a "quality score" and all males can be ranked by quality score, while all females can be ranked by quality score. A marriage creates a "household" whose goal is to maximize some kind of domestic production, which is a function of the quality scores of the two partners being married. If (following chromosome conventions) we denote the female's quality score by x and the male's quality score by y, then this is a function $f(x, y)$, and is increasing in both x and y.

Question: What way of matching males and females maximizes production? There are two extreme possibilities. The first is *assortative mating*, where the highest quality males join hands with the highest quality females, and the lowest quality males walk the aisle with the lowest quality females. We might call this *like begets like theory of mating*. The other is the reverse, where the highest quality males marry the lowest quality females, and the lowest quality males marry the highest quality females. We might call this the *opposites attract theory of mating*.

Which story maximizes domestic production depends on *mixed partials*. Specifically, if the domestic production function has a positive mixed partial with respect to male and female quality, i.e., if $f_{xy} > 0$, that means that the higher the male quality, the *more* beneficial it is to marry a female of higher quality. Similarly, the higher the female quality, the more beneficial it is to marry a male of higher quality. In other words, *a positive mixed partial derivative bodes well for assortative mating*.

On the other hand, if $f_{xy} < 0$, then the higher the male quality, the less beneficial it is to marry a female of higher quality. This is a subtle point so it's worth pondering a bit. For any given male, if he has the choice, it always makes sense to marry the highest quality female he can get. But with a negative mixed derivative, the *margin of difference between high quality females and low quality females* is *lower for high quality males*. On the other hand, low quality males see a significant boost from attracting high quality females. Thus, the utility-maximizing arrangement would be one where the lowest quality males pair up with the highest quality females, and the highest quality males pair up with the lowest quality females.

One could also set up a bidding/auction scenario to see how, under an open bidding process for mates, this utility-maximizing outcome can be achieved. Suppose females are bidding for males. Basically, in the positive mixed partial derivative case, the high quality females are more desperate for the high quality males than the low quality females are, so they outbid the low quality females, leaving the low quality females to make do with the low quality males. On the other hand, with negative mixed partials, the low quality females, despite being less attractive than high quality females, are still able to get the high quality males because they are more desperate to win over the males and are willing to offer better terms. The disgruntled high quality females settle for the low quality males left in the pool, who are glad at their catch.

For what it's worth, most of the evidence suggests that mating is highly assortative. This doesn't quite prove that domestic production enjoys positive mixed partials, but it suggests that if we are trying to fit the real world into our highly restrictive model, then positive mixed partials would generate more realistic predictions.

4.3. **Cobb-Douglas production function.** The Cobb-Douglas production function is a particular form of a production function describing output in terms of two or more inputs. We consider the case of two inputs, labor L and capital K. Suppose the output is given by a function:

$$f(L, K) := CL^a K^b$$

Note that the numbers L and K denote the financial expenditures on labor and capital.

Here, C, a, and b are all positive and L and K are restricted to positive inputs. Logarithmically, we get:

$$\ln(f(L, K)) = \ln C + a \ln L + b \ln K$$

The partial derivatives are as follows:

$$\begin{aligned}
\frac{\partial}{\partial L} f(L,K) &= CaL^{a-1}K^b \\
\frac{\partial}{\partial K} f(L,K) &= CbL^{a}K^{b-1} \\
\frac{\partial^2}{\partial L^2} f(L,K) &= Ca(a-1)L^{a-2}K^b \\
\frac{\partial^2}{\partial K^2} f(L,K) &= Cb(b-1)L^{a}K^{b-2} \\
\frac{\partial^2}{\partial L \partial K}(f(L,K)) &= CabL^{a-1}K^{b-1} \\
\frac{\partial}{\partial(\ln L)}(\ln(f(L,K))) &= a \\
\frac{\partial}{\partial(\ln K)}(\ln(f(L,K))) &= b \\
\frac{\partial^2}{\partial(\ln L)\partial(\ln K)}(\ln(f(L,K))) &= 0
\end{aligned}$$

The partial derivatives involving logarithms are the dimensionless versions of the parital derivatives, in the same sense that price elasticity is a dimensionless version of the partial derivative of quantity demanded with respect to price.

4.4. **Returns to scale.** We note that:

- If labor alone is multiplied by a factor of λ, then output gets multiplied by a factor of λ^a. From this, or by looking at the second derivative, we see that if $0 < a < 1$ we have positive but decreasing returns on labor holding capital fixed. If $a = 1$, we have constant returns on labor holding capital fixed. If $a > 1$, we have positive returns on labor holding capital fixed.
- If capital alone is multiplied by a factor of λ, then output gets multiplied by a factor of λ^b. From this, or by looking at the second derivative, we see that if $0 < b < 1$, then there are positive but decreasing returns on capital holding labor fixed. For $b = 1$, constant returns on capital holding labor fixed, and for $b > 1$, increasing returns on capital holding labor fixed.
- If we multiply labor and capital both simultaneously by the same factor of λ, output is multiplied by a factor of λ^{a+b}. Thus, we get decreasing, constant, or increasing returns to scale depending on whether $a + b < 1$, $a + b = 1$, or $a + b > 1$.
- Note that in any Cobb-Douglas production model, labor and capital always play complementary roles, because the mixed partial derivative is positive.

4.5. **Maximizing production: fixed total of labor and capital.** If the total investment in labor and capital is fixed, i.e., $L + K$ is fixed, then, in order to maximize the production, we must allocate resources of L and K in the proportion $a : b$. In other words, if $L + K = M$ for some constant M, then the maximum production occurs when $L = Ma/(a + b)$ and $K = Mb/(a + b)$.

To see this, note that if x is the fraction allocated to labor, and $1 - x$ to capital, then the output is $CM^{a+b}x^a(1 - x)^b$. Differentiating with respect to x and then computing the local maximum by setting the derivative to zero gives the answer.

Roughly speaking, this means that the proportion of investment in labor should be based on the exponents on labor and capital. To determine whether investment is being done the way it "should" we need to find out:

- The exponents a and b in the Cobb-Douglas production function.
- The proportion of investment used for labor.

Historically, this was the motivation for the creation of the Cobb-Douglas production function. Douglas assembled historical data and estimated that production could be modeled roughly by a Cobb-Douglas

production function with the exponents a and b in the ratio $3:1$. He also found that the expenditures on labor and capital were in the ratio $3:1$. Thus, he concluded that things were being done the way they "should" be.

Although a production function need not be a Cobb-Douglas production function, trying to use a Cobb-Douglas model is a good first pass.

5. Partial differential equations

This is a fairly tricky topic and we consider it in a very superficial way.

5.1. **Ordinary differential equations: reminder.** Recall the notion of (*ordinary*) differential equation. Here, there is one dependent variable (typically denoted y) and one independent variable (typically denoted x). The ordinary differential equation is of the form:

$$F(x, y, y', y'', \ldots, y^{(k)}) = 0$$

In other words, some expression involving x, y, and the derivatives (first and higher) of y with respect to x is zero. A "solution" to this equation is a function (given either explicitly or via an implicit/relational description) $y = f(x)$ such that the above equation holds *identically* for all x in the domain.

Note that not every equation involving derivatives of functions of one variable is a differential equation. For instance, the equation $f'(x) = f(1-x)f(2-x)$ is a functional equation involving derivatives, but it is *not* a differential equation because it involves evaluating the function f at multiple points. A differential equation, as we use the term, refers to a situation that is restricted to the local behavior around a *single* point.

5.2. **Partial differential equations.** Partial differential equations are to functions of many variables what ordinary differential equations are to functions of one variable. Specifically, a partial differential equation involves some expression in terms of a bunch of variables, a function of those variables, and partial derivatives of that function.

Just like the case of ordinary differential equations, it remains true that a partial differential equation must involve *only* the behavior at and around a particular point in the domain. For instance, if we have a function $u = f(x, y)$ then the equation $f(x, y) = f_y(x^2, x+y) - f_{xx}(x^2, y^2)$ is a functional equation involving derivatives but it is *not* a partial differential equation.

The *order* of a partial differential equation is defined as the highest of the orders of the partial derivatives that appear in the partial differential equation.

5.3. **Key examples of partial differential equations.** The *Laplace equation* is an example of a partial differential equation. For a function u of variables x and y, the equation is:

$$u_{xx} + u_{yy} = 0$$

In Leibniz notation, this becomes:

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

Another example is the wave equation:

$$\frac{\partial^2 u}{\partial t^2} = a^2 \frac{\partial^2 u}{\partial x^2}$$

5.4. **How big is the space of solutions functions?** Recall that for an ordinary differential equation, the solution function need not be unique, but the space of solution functions can typically be described as a r-dimensional space (in the sense of there being r free parameters) where r is the order of the differential equation.

For a partial differential equation, the solution space is *much much larger*. Usually it is infinite-dimensional if we are dealing with functions of more than one variable. It is usually impossible to give a description of the general solution.

TANGENT PLANES AND LINEAR APPROXIMATIONS

MATH 195, SECTION 59 (VIPUL NAIK)

Corresponding material in the book: Section 14.4. *Note: We are, for now, omitting the topic of differentials, which is the second half of this section (Pages 931–934). We may return to it later in the course, if we get time after completing the rest of Chapters 14 and 15.*

What students should definitely get: Finding the tangent plane at a point, the concept of best linear approximation.

Executive summary

Words ...

(1) For a d-dimensional subset of $\mathbb{R}^n$, it (occasionally) makes sense to talk of the tangent space and the normal space at a point. The tangent space is a linear/affine d-dimensional space and the normal space is a linear/affine $(n-d)$-dimensional space. Both pass through the point and are mutually orthogonal.

(2) For a function $z = f(x,y)$, the tangent plane to the graph of this function (a surface in $\mathbb{R}^3$) at the point $(x_0, y_0, f(x_0,y_0))$ *such that f is differentiable at the point* (x_0,y_0) is the plane:

$$z - f(x_0,y_0) = f_x(x_0,y_0)(x-x_0) + f_y(x_0,y_0)(y-y_0)$$

The corresponding linear function we get is:

$$L(x,y) = f(x_0,y_0) + f_x(x_0,y_0)(x-x_0) + f_y(x_0,y_0)(y-y_0)$$

This provides a linear approximation to the function near the point where we are computing the tangent plane.

(3) It may be the case that a function f of two variables is not differentiable at a point in its domain but the partial derivatives exist. In this case, although the *above formula makes sense as a formula*, the plane it gives *is not the tangent plane* – in fact, *no tangent plane exists*. Similarly, *no linearization exists*, and the linear function given by the above formula *is not a close approximation to the function near the point*.

1. Approximation theory: recall of one variable

1.1. **Taylor polynomials.** For a function f of one variable, we can define some Taylor polynomials of f. If f is n times differentiable around a point $x = c$, we can define the n^{th} Taylor polynomial of f about c as the polynomial:

$$P_n(f,c)(x) := \sum_{k=0}^{n} \frac{f^{(k)}(c)}{k!}(x-c)^k$$

If f is $(n+1)$ times differentiable, it turns out that $f - P_n(f,c)$ is a function with a zero of order at least $n+1$ at c. This means that we can approximate f by a *polynomial function* of degree at most n such that the error term is very zeroey (order at least $n+1$). Note that the Taylor polynomial has degree at most n, but the degree may be less than n if the n^{th} derivative takes the value 0 at c.

1.2. **Special cases: degrees zero and one.** The zeroth degree Taylor polynomial is a constant function with the same value as $f(c)$. This is a very crude description of the function around the point and ignores any change in the value.

The first degree Taylor polynomial is the function whose graph gives the tangent line, i.e., the line:

$$y = f(c) + f'(c)(x - c)$$

Note that this is the tangent line in point-slope form.

The first degree Taylor polynomial, or the tangent line, captures the rate of change of the function *at* the point. However, it fails to capture second derivative and higher derivative behavior, i.e., how this rate of change itself is changing.

Geometrically, the tangent line to a point on a curve is the "best linear approximation" to the curve locally around the point, i.e., the line that comes closest to describing the curve near the point. Note that it is *not* true that the tangent line intersects the curve at a unique point, or that no other line has this property.

2. The multivariable situation

2.1. **Geometric notion of tangent space and normal space.** Given a subset of dimension d in n-dimensional space, we can try talking of a tangent space at a point on the subset to the subset. This attempt at talk succeeds only when that subset has some particularly nice properties. Anyway, the point is that this tangent space looks like a *flat* d-dimensional space, i.e., a line, or a plane, or a higher-dimensional analogue thereof.

For instance, the "tangent" to a curve (which is a one-dimensional construct) is a line (the linear one-dimensional construct). Similarly, the tangent to a surface in $\mathbb{R}^3$ is a plane (a linear two-dimensional construct). To take the example of a sphere, think of a sphere resting on a floor. The floor is the tangent plane to the sphere through the point of contact.

The "normal space" at a point to a d-dimensional subset in n-dimensional space is a $(n-d)$-dimensional linear space through the same point, such that the two spaces intersect orthogonally, i.e., every direction in one space is orthogonal to every direction in the other space. Thus, for instance:

- For a curve in $\mathbb{R}^2$, the tangent space and normal space are both one-dimensional.
- For a curve in $\mathbb{R}^3$, the tangent space is one-dimensional and the normal space is two-dimensional.
- For a surface in $\mathbb{R}^3$, the tangent space is two-dimensional and the normal space is one-dimensional.

Note that for geometric subsets, we can only make sense of tangent and normal spaces, rather than specific tangent and normal vectors. However, given a *parametric* description of a curve, we can make sense of the tangent *vector*, with the length and direction of the vector determined by the speed and direction of motion along the curve as per the parameterization.

2.2. **Non-existent tangent planes.** There are many possible reasons why the tangent plane to a surface at a point on the surface may not exist. First, the surface may be broken at that point, i.e., it may not locally look like a plane in a small neighborhood of the point. In such a case, it might be an abuse of language to call it a surface.

Second, the surface may have a lot of variation around the point – too many hills and valleys to make sense of a meaningful tangent plane. This is a surface analogue of the function $x\sin(1/x)$ which, if extended to the value 0 at 0, becomes continuous but not differentiable at 0.

Third, the surface may be sharp and pointy at the point. Think, for instance, of the curved surface of a right circular cone. At most points on this curved surface, the tangent plane exists. However, at the vertex of the point, the tangent plane does not exist. This is a surface analogue of a function that has one-sided derivatives that are not equal, i.e., it takes a sharp turn.

2.3. **How should the tangent plane relate geometrically to the surface?** In the one-dimensional situation, we recall that *generally speaking*, the tangent line *locally* lies entirely to one side of the curve, i.e., the curve rests against the tangent line. In fact, it is the only line that does not *cut through* the curve.

But this is not always true. Two notable kinds of exceptions are:

- *Points of inflection*: Here, the tangent line *cuts through* the curve, i.e., the curve and the tangent line cross each other. An example is $y = x^3$ at $(0,0)$. The tangent line is the x-axis, and it crosses the curve.
- Points where the curve keeps crossing above and below the tangent line arbitrarily close to the point of tangency: For this to occur, the second derivative must change sign infinitely often close to the point. Examples include functions such as $x^2 \sin(1/x)$ (with the value 0 at 0) about the origin $(0,0)$. The tangent line is the x-axis, and it intersects the curve at points arbitrarily close to $(0,0)$, hence fails to be *on one side* of the curve.

Does something similar happen for planes? Yes. *Generically*, we expect that *locally*, the tangent plane lies to one side of the surface, and this can be a reasonable characterization of the tangent plane. That's the reason why for a sphere "resting" on a floor, the floor is the tangent plane to the sphere at the point of contact.

However, there is an analogue of point of inflection, where the tangent plane *cuts through* the surface at the point. This type of two-dimensional analogue of point of inflection is termed a *saddle point*. We will deal with saddle points when we cover the topic of maxima and minima for functions of many variables later in the course.

2.4. **Relation between tangents for curve on surface.** If a curve is in/on a surface in $\mathbb{R}^3$, then the tangent *line* to the curve at a point on the curve lies in the tangent *plane* to the surface at the point. Similarly, the normal plane to the curve at a point contains the normal line to the surface at that point.

In particular, if we have two different curves on a surface intersecting at a point, and the tangent lines to these curves at the point do not coincide, then these lines together determine the tangent plane to the surface, if it exists: it is the unique plane containing both the tangent lines.

If, on the other hand, we find a situation where there are three curves intersecting at a point, all in a surface, and the tangent lines at the point to these three curves do not lie in the same plane, then the tangent plane at the point does not exist.

2.5. **Partial derivatives of functions of many variables.** Consider a function $z = f(x,y)$, a function of two variables. The graph of this function is a surface in $\mathbb{R}^3$. Recall that the partial derivatives can be interpreted as slopes of tangent lines as follows:

- $f_x(x_0,y_0)$: Consider the plane $y = y_0$. This is a plane parallel to the xz-plane, and the intersection of the surface with this plane can be thought of as the graph of the function $z = f(x,y_0)$ of one variable. The partial derivative is the slope of the tangent line in this plane to the graph at the point $(x_0,y_0,f(x_0,y_0))$. A free vector along this tangent line, viewed in $\mathbb{R}^3$, is $\langle 1,0,f_x(x_0,y_0)\rangle$.
- $f_y(x_0,y_0)$: Here, we fix the plane $x = x_0$, parallel to the yz-plane, and a similar interpretation follows. A free vector along this tangent line, viewed in $\mathbb{R}^3$, is $\langle 0,1,f_y(x_0,y_0)\rangle$.

Thus, computing the partial derivatives allows us to compute tangent vectors to two curves in the surface that's the graph of this function. Since we know the basepoint, we can also compute parametric descriptions of the corresponding tangent lines.

If both the partial derivative f_x and f_y are continuous, then we can make sense of the notion of tangent plane, to which we now turn.

These two vectors are both parallel to the tangent plane, so we can take their cross product and find a normal vector. A quick computation of the cross product shows that $\langle f_x(x_0,y_0), f_y(x_0,y_0), -1\rangle$ is a normal vector. Thus, the tangent plane passes through the point $(x_0,y_0,f(x_0,y_0))$ and has normal vector $\langle f_x(x_0,y_0), f_y(x_0,y_0), -1\rangle$. Working out the scalar equation from the vector equation, we get:

$$(x-x_0)f_x(x_0,y_0) + (y-y_0)f_y(x_0,y_0) + (z-f(x_0,y_0))(-1) = 0$$

Rearranging, we get:

$$z - f(x_0,y_0) = f_x(x_0,y_0)(x-x_0) + f_y(x_0,y_0)(y-y_0)$$

We can rewrite this as:

$$z = f(x_0,y_0) + f_x(x_0,y_0)(x-x_0) + f_y(x_0,y_0)(y-y_0)$$

This is the equation of the tangent plane, and is the most convenient form for applications.

2.6. **Tangent plane as good linear approximation.** Just as the tangent line is a good linear approximation to the graph of a function in one variable, the tangent plane is a good linear approximation to the graph of a function in two variables. Roughly, it can be thought of as a first-order approximation, so that any "error term" will be zeroey of order two or higher. In particular, for points close to the point at which we are computing the tangent plane, the function value arising from the linear approximation is pretty close to the actual function value.

More concretely, with the above setup, we have the plane:

$$z = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0)$$

The corresponding linear function is the right side of this equation:

$$L(x, y) = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0)$$

2.7. **Continuity of partial derivatives necessary.** Recall the earlier example we studied of a function that is separately continuous in each variable but is not jointly continuous. The function is $f(x, y) = (xy)/(x^2 + y^2)$ except at the origin, and $f(0, 0) = 0$. This can also be described as $(1/2)\sin(2\theta)$ with respect to polar coordinates. At the origin, the function is separately continuous in each variable, but not jointly continuous.

We can further note that in fact, since the function is a constant zero function along both the axes, all its partial derivatives exist and equal zero at the origin. However, the function is not jointly continuous, and hence, we should not expect it to have anything like a tangent plane. In fact, it does not. Although we can blindly apply the above formula to obtain the equation of a plane, this is not a "tangent plane" to the surface that is the graph of the function. (See Page 930 of the book).

2.8. **Tangent plane and total derivative.** Recall an earlier conundrum:

Separate continuity:Joint continuity::Partial derivatives:?

In other words, what is the "joint" equivalent of partial derivatives? Unfortunately, the proper way of thinking of this joint equivalent requires the use of linear algebra, and we will therefore not be able to cover it. I will simply state the following result:

> If all the first-order partial derivatives of a function exist around a point, and each first-order partial derivative is jointly continuous at the point, then the function is differentiable at the point. However, the converse is not true.

In other words:

Partials exist around the point, are continuous $\implies$ Function is differentiable (total derivative exists) $\implies$ Partials exist *at* the point

CHAIN RULE

MATH 195, SECTION 59 (VIPUL NAIK)

Corresponding material in the book: Section 14.5.
What students should definitely get: The generic formulation of the chain rule, the particular cases of $1 \to 2 \to 1$ and $2 \to 2 \to 1$.

Executive summary

Words ...

(1) The general formulation of chain rule: consider a function with m inputs and n outputs, and another function with n inputs and p outputs. Composing these, we get a function with m inputs and p outputs. The m original inputs are termed *independent variables*, the n in-between things are termed *intermediate variables*, and the p final outputs are termed *dependent variables*.

For a given triple of independent variable t, intermediate variable x, and dependent variable u, the partial derivative of u with respect to t via x is defined as:

$$\frac{\partial u}{\partial x}\frac{\partial x}{\partial t}$$

The chain rule says that the partial derivative of u with respect to t is the sum, over all intermediate variables, over the partial derivatives via each intermediate variable.

(2) The $1 \to 2 \to 1$ and $2 \to 2 \to 1$ versions (see the lecture notes or the book).

(3) There is also a tree interpretation of this, where we make pathways based on the directions/paths of dependence. This is discussed in the book, not the lecture notes.

(4) The product rule for scalar functions can be proved using the chain rule. Other variants of the product rule can be proved using generalized formulations of the chain rule, which are beyond our current scope.

(5) Implicit differentiation can be understood in terms of the chain rule and partial derivatives.

1. The chain rule

1.1. **1 to 2 to 1 chain rule.** This simplest nontrivial chain rule is as follows: Consider two functions $x = x(t)$ and $y = y(t)$ of a single variable t, and consider a function $z = f(x, y)$ of two variables. We can compose these to get a function with one input and one output: $z = f(x(t), y(t))$. In other words, we have the composition:

$$t \stackrel{\langle x,y \rangle}{\mapsto} \langle x(t), y(t) \rangle \stackrel{f}{\mapsto} f(x(t), y(t))$$

We are composing a function from 1 variable to 2 variables and a function from 2 variables to 1 variable. Overall, we get a function from 1 variable to 1 variable. The chain rule states that:

$$\frac{d(f(x(t), y(t)))}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}$$

We can think of this as follows: the first product is measuring the contribution to the derivative via changes in x, keeping y constant, and the second product is measuring the contribution to the derivative via changes in y, keeping x constant. (More on this interpretation a little later).

1.2. 1 **to** n **to** 1 **chain rule.** For the $1 \to n \to 1$ chain rule, we have n functions on one variable, and then compose them with a function of n variables to get a scalar function of one variable. The chain rule adds up the contributions of each variable.

In symbols: suppose we have function $x_1(t)$, $x_2(t)$, ..., $x_n(t)$ and a function f of n variables. We can consider the function $t \mapsto f(x_1(t), x_2(t), \ldots, x_n(t))$ and its derivative is as follows:

$$\frac{d}{dt}[f(x_1(t), x_2(t), \ldots, x_n(t))] = \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i} \frac{dx_i}{dt} \right)$$

1.3. 1 **to** n **to** m**.** If we are composing a $1 \to n$ function and a $n \to m$ function, we can reduce the chain rule to the previous chain rule, by simply looking at each coordinate of the final m-dimensional output and writing the corresponding $1 \to n \to 1$ rule.

1.4. m **to** n **to** 1**,** m **to** n **to** p**.** Suppose we have n functions, each having m inputs, and then we have a function of n inputs. Then, we can compose these and get a function with m inputs and 1 output. The chain rule for this looks the same as for $1 \to n \to 1$, except that now we have partial derivatives everywhere.

We explicitly write out the $2 \to 2 \to 1$ case. Suppose $z = f(x, y)$ is a differentiable function of x and y, where $z = g(s, t)$ and $y = h(s, t)$. Then, we have the following formulas:

$$\begin{aligned} \frac{\partial z}{\partial s} &= \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s} \\ \frac{\partial z}{\partial t} &= \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t} \end{aligned}$$

We can combine all the ideas above to get the $m \to n \to p$ chain rule – the most generic version available.

The book has an interesting explanation in terms of tree diagrams. Please review this if you would like to improve upon your understanding of the above.

1.5. **Terminology and conceptual formulation.** For a $m \to n \to p$ chain rule, the starting m variables are termed the *independent variables*, the n variables in the middle are termed the *intermediate variables*, and the p variables at the end are termed the *dependent variables*.

Conceptually, the chain rule says that:

> The partial derivative of any dependent variable with respect to any independent variable is the sum over all intermediate variables of the product of (partial derivative of dependent variable with respect to intermediate variable) and (partial derivative of intermediate variable with respect to dependent variable).

Here's a longer version of the same explanation: given an independent variable t , an intermediate variable x , and a dependent variable u, the *derivative of u with respect to t via x* is the product:

$$\frac{\partial u}{\partial x}\frac{\partial x}{\partial t}$$

The partial derivative $\partial u/\partial x$ is to be understood as the partial derivative of u with respect to x *keeping all other intermediate variables constant* in the $n \to p$ function. The partial derivative $\partial x/\partial t$ is to be understood as the partial derivative of x with respect to t *keeping all the other independent variables constant* in the $m \to n$ function.

The derivative of u with respect to t is the *sum of all possible intermediates* of the derivative via each intermediate.

2. Deriving the product rule from the chain rule

2.1. **Product rule for two scalar functions.** Recall that the product rule says that:

$$\frac{d}{dt}[x(t)y(t)] = x(t)y'(t) + x'(t)y(t)$$

We now see how the product rule can be deduced using the chain rule and the fact that *constants can be pulled out of products* (in other words, the derivative of a constant times a function is the constant times the derivative of a function). Think of the function:

$$f(x, y) = xy$$

Then, the left side of the product rule is $(d/dt)[f(x(t), y(t))]$, and can thus be written as:

$$\frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}$$

Now, $\partial(xy)/\partial x = y$ because the "constant" y can be pulled out of the derivative. Similarly, $\partial(xy)/\partial y = x$ because the "constant" x can be pulled out of the derivative. Plugging thesse in, we get the formula for the product rule.

Not only does this establish the product rule, it also gives a rough explanation for rules like the product rule which we've seen: the variant of the product rule for dot products, and the variant of the product rule for the product of a scalar and a vector function. The variant for the cross product is somewhat subtler and needs an even more generic perspective on the chain rule that is beyond our current scope.

2.2. **Product rule for more than two functions.** The general product rule says that:

$$\frac{d}{dt}[x_1(t)x_2(t)\dots x_n(t)] = x_1'(t)x_2(t)\dots x_n(t) + x_1(t)x_2'(t)\dots x_n(t) + \dots + x_1(t)x_2(t)\dots x_n'(t)$$

We can deduce this from the $1 \to n \to 1$ chain rule. The i^{th} of the summands on the right side is the partial derivative with respect to t via the i^{th} intermediate variable x_i.

3. Implicit differentiation explained

We now turn to unraveling the mystery of implicit differentiation, a topic that we learned way back in single variable calculus.

Here is how we thought of implicit differentiation. Suppose y is an implicit function of x given by a relational description of the form $F(x, y) = 0$, where it is not obvious how to isolate an expression for y in terms of x.

To find the derivative, we differentiate F with respect to x, treating y as an implicit function of x. This means that wherever we have to differentiate y, we just write dy/dx and leave it at that. After doing this differentiation, we regroup terms and compute dy/dx in terms of x and y.

We can now think of implicit differentiations as a special case of partial derivatives in the following sense. We treat x as the parameter and view x and y both as functions of x (with x being the identity function of itself). In this case, we have:

$$\frac{dF}{dx} = \frac{\partial F}{\partial x}\frac{dx}{dx} + \frac{\partial F}{\partial y}\frac{dy}{dx}$$

Here the x on the left side is the original x (viewed as the *independent variable*) and the x of partial differentiation on the right side is the *intermediate variable* x of the (x, y) pair, i.e., $\partial F/\partial x$ basically means we are differentiating with respect to the intermediate variable x treating the intermediate variable y constant in the $2 \to 1$ function, which differs from the actual differentiation with respect to x in the *composite* $1 \to 2 \to 1$ function.

Thus, if we start with $F(x, y) = 0$ and differentiate, we get $dF/dx = 0$, which gives:

$$\frac{\partial F}{\partial x} + \frac{\partial F}{\partial y}\frac{dy}{dx} = 0$$

We then rearrange and calculate:

$$\frac{dy}{dx} = -\frac{F_x}{F_y}$$

CHAIN RULE AND SECOND DERIVATIVES

MATH 195, SECTION 59 (VIPUL NAIK)

The homework question is as follows: Suppose $z = f(x, y)$ where $x = g(s, t)$ and $y = h(s, t)$.

(1) Show the formula for $\frac{\partial^2 z}{\partial t^2}$ as given in the book. It is a somewhat long formula.
(2) Find a similar formula for $\partial^2 z/(\partial s \partial t)$.

Last year, many people got it correct, but some people didn't clearly understand how to proceed. In anticipation of students facing similar difficulties this year, I include a detailed outline of the solution below. Even those students who are able to arrive at the correct answer by themselves may benefit from reading through this.

Similar one-variable question. Let's consider a simple situation, where we have $x = g(t)$ and $z = f(x)$, with no y and s appearing. We want a formula for d^2z/dt^2. In other words, we want to compute $(f \circ g)''$

You've all seen this, by the way, in a past quiz (January 9), where almost all of you got the question correct. The idea now is to carefully understand what we're doing so that it can readily be replicated in the multiple variable case.

We have, by the chain rule:

$$\frac{dz}{dt} = \frac{dz}{dx}\frac{dx}{dt} = f'(x)g'(t) = f'(g(t))g'(t)$$

We now want to differentiate this with respect to t again:

$$\frac{d^2z}{dt^2} = \frac{d}{dt}[f'(g(t))g'(t)]$$

Note that the expression to be differentiated on the right side is a product, so we use the product rule, and obtain:

$$\frac{d^2z}{dt^2} = \frac{d}{dt}[f'(g(t))]g'(t) + f'(g(t))g''(t)$$

For the derivative of $f'(g(t))$ with respect to t, we *again* use the chain rule, and get $f''(g(t))g'(t)$, so overall:

$$\frac{d^2z}{dt^2} = f''(g(t))(g'(t))^2 + f'(g(t))g''(t)$$

Let's understand this step by step:

(1) We calculated the first derivative using the chain rule, and got a product of a derivative with respect to the intermediate value x (which became $f'(g(t))$) and a derivative with respect to the initial variable t (which became $g'(t)$).
(2) To differentiate this, we use the product rule.
(3) For one of the pieces in the product rule, we are differentiating $g'(t)$, which becomes $g''(t)$. This is the straightforward piece (second summand in our description above).
(4) For the other piece, we need to differentiate the composite $f'(g(t))$ with respect to t. For this, we *use the chain rule again.*

Second partial with a single variable. Return to the original question: Suppose $z = f(x, y)$ where $x = g(s, t)$ and $y = h(s, t)$.

We have:

$$\frac{\partial z}{\partial t} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t}$$

With the subscript notation, this becomes:

$$\frac{\partial z}{\partial t} = f_x g_t + f_y h_t$$

So far, we are done with the equivalent of Step (1) for one variable. Note that instead of just a single produce, we have a sum of two products, indicating the two paths of dependence (via x and via y).

Now, we want to differentiate both sides with respect to t:

$$\frac{\partial^2 z}{\partial t^2} = \frac{\partial}{\partial t}\left(f_x g_t + f_y h_t\right)$$

We break additively and use the product rule on each piece (analogous to Step (2)), and get:

$$\frac{\partial^2 z}{\partial t^2} = \frac{\partial (f_x)}{\partial t} g_t + f_x \frac{\partial g_t}{\partial t} + \frac{\partial f_y}{\partial t} h_t + f_y \frac{\partial h_t}{\partial t}$$

Within each product rule, the second summand is easy to rewrite: $\partial g_t/\partial t$ becomes g_{tt} and $\partial h_t/\partial t = h_{tt}$. These simpler parts are analogous to Step (3) in the one-variable scenario.

The harder pieces are $\partial f_x/\partial t$ and $\partial f_y/\partial t$. These are analogous to Step (4). Let's look at these more carefully.

$f_x(x, y)$ is a function of the variables x and y, each of which in turn depends on s and t. Thus:

$$\frac{\partial f_x}{\partial t} = \frac{\partial f_x}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f_x}{\partial y}\frac{\partial y}{\partial t} = f_{xx} g_t + f_{xy} h_t$$

Similarly:

$$\frac{\partial f_y}{\partial t} = \frac{\partial f_y}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f_y}{\partial y}\frac{\partial y}{\partial t} = f_{yx} g_t + f_{yy} h_t$$

Plugging these all back in the original expression, we get:

$$\frac{\partial^2 z}{\partial t^2} = (f_{xx} g_t + f_{xy} h_t) g_t + f_x g_{tt} + (f_{yx} g_t + f_{yy} h_t) h_t + f_y h_{tt}$$

There is a total of six terms, but using Clairaut's theorem, we see that the term $f_{xy} g_t h_t$ appears twice, so combining, we get a sum of five terms with a coefficient of 2 appearing on one of them.

The key difference relative to the situation with one variable: *dependence on two intermediate variables means that at Step (1), we have a sum of two products, and in Step (4), again each of the chain derivatives is a sum of two products. Hence we get a total of* 6 *terms.*

Mixed partial with both variables. Let's try to compute $\partial^2 z/(\partial s \partial t)$. We already have:

$$\frac{\partial z}{\partial t} = f_x g_t + f_y h_t$$

We differentiate both sides with respect to s, and get:

$$\frac{\partial^2 z}{\partial s \partial t} = \frac{\partial}{\partial s}[f_x g_t + f_y h_t]$$

Using additive splitting and the product rule (analogous to Step (2)), we get:

$$\frac{\partial^2 z}{\partial s \partial t} = \frac{\partial f_x}{\partial s} g_t + f_x \frac{\partial g_t}{\partial s} + \frac{\partial f_y}{\partial s} h_t + f_y \frac{\partial h_t}{\partial s}$$

The second and fourth summand are easy to simply: $\partial g_t/\partial s = g_{ts}$ and $\partial h_t/\partial s = h_{ts}$. The harder ones are $\partial f_x/\partial s$ and $\partial f_y/\partial s$. As before, we use the chain rule:

$$\frac{\partial f_x}{\partial s} = \frac{\partial f_x}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial f_x}{\partial y}\frac{\partial y}{\partial s} = f_{xx}g_s + f_{xy}h_s$$

Similarly:

$$\frac{\partial f_y}{\partial s} = \frac{\partial f_y}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial f_y}{\partial y}\frac{\partial y}{\partial s} = f_{yx}g_s + f_{yy}h_s$$

Plugging these in, we get:

$$\frac{\partial^2 z}{\partial s \partial t} = (f_{xx}g_s + f_{xy}h_s)g_t + f_x g_{ts} + (f_{yx}g_s + f_{yy}h_s)h_t + f_y h_{ts}$$

As with the previous case, this is a sum of six terms.

DOUBLE INTEGRALS AND ITERATED INTEGRALS

MATH 195, SECTION 59 (VIPUL NAIK)

Corresponding material in the book: Section 15.2, 15.3. *Note: We are omitting the question types from the book that require three-dimensional visualization, i.e., those that require sketching figures in three dimensions to compute volumes.*

What students should definitely get: The procedure for computing double integrals over rectangles using iterated integrals, the procedure for computing double integrals over other regions using iterated integrals, the idea of Fubini's theorem and its use in interchanging the order of integration. Use of symmetry and inequality-based bounding/estimation techniques.

What students should hopefully get: Relation between single and double integrals, dealing with piecewise cases, breaking up domain into smaller pieces when direct integration over entire domain is infeasible.

Note: The lecture notes contain only a few examples. For more examples, please refer to worked examples in Sections 15.2 and 15.3.

Executive summary

Words ...

(1) The double integral of a function f of two variables, over a domain D in $\mathbb{R}^2$, is denoted $\int\int_D f(x,y)\,dA$ and measures an infinite analogue of the sum of f-values at all points in D.

(2) Fubini's theorem for rectangles states that if F is a function of two variables on a rectangle $R = [a,b] \times [p,q]$, such that F is continuous except possibly at the union of finitely many smooth curves, then the integral equals either of these iterated integrals:

$$\int\int_R F(x,y)\,dA = \int_a^b \int_p^q F(x,y)\,dy\,dx = \int_p^q \int_a^b F(x,y)\,dx\,dy$$

(3) For a function f defined on a closed connected bounded domain D with a smooth boundary, we can make sense of $\int\int_D f(x,y)\,dA$ as being $\int\int_R F(x,y)\,dA$ where R is a rectangular region containing D and F is a function that equals f on D and is 0 on the rest of R.

(4) Suppose D is a Type I region, i.e., its intersection with every vertical line is either empty or a point or a line segment. Then, we can describe D as $a \leq x \leq b$, $g_1(x) \leq y \leq g_2(x)$, where g_1 and g_2 are continuous functions. The integral $\int\int_D f(x,y)\,dA$ becomes:

$$\int_a^b \int_{g_1(x)}^{g_2(x)} f(x,y)\,dy\,dx$$

(5) Suppose D is a Type II region, i.e., its intersection with every horizontal line is either empty or a point or a line segment. Then, we can describe D as $p \leq y \leq q$, $g_1(y) \leq x \leq g_2(y)$, where g_1 and g_2 are continuous functions. The integral $\int\int_D f(x,y)\,dA$ becomes:

$$\int_p^q \int_{g_1(y)}^{g_2(y)} f(x,y)\,dx\,dy$$

(6) The double integral of $f+g$ over D is the sum of the double integral of f over D and the double integral of g over D. Similarly, scalars can be pulled out of double integrals.

(7) The integral of the function 1 over a domain is the area of the domain.

(8) If $f(x,y) \geq 0$ on a domain D, the integral of f over D is also ≥ 0.

(9) If $f(x,y) \geq g(x,y)$ on a domain D, the integral of f over D is $\geq$ the integral of g over D.

(10) If $m \leq f(x,y) \leq M$ over a domain D, then $\int\int_D f(x,y)\,dA$ is betweem mA and MA where A is the area of D.
(11) If $f(x,y)$ is odd in x and the domain of integration is symmetric about the y-axis, the integral is zero. If $f(x,y)$ is odd in y and the domain is symmetric about the x-axis, the integral is zero.

Actions ...

(1) To compute a double integral, compute it as an iterated integral. For a rectangle, we can choose either order of integration, as long as the integration is feasible. For other types of regions, we need to first determine whether the region is Type I or Type II, and break it up into pieces of those types.
(2) For a multiplicatively separable function over a rectangular region (or for a sum of such multiplicatively separable functions), things are particularly easy.
(3) Sometimes, an integral cannot be computed using a particular order of integration – we might get stuck on the inner or the outer stage. However, it may be computable using the other order of integration.
(4) We can often use symmetry-based techniques to argue that certain parts of the integrand integrate to zero.
(5) Even in cases where the integral cannot be computed, we can bound it between limits using maximum or minimum values of function and/or using bigger or smaller regions on which the integral can be computed.

1. Double integral and iterated integral

1.1. **What's a double integral?** We will study the theory of double integrals (Section 16.1 of the book) a little later in the course. For now, we provide an intuitive idea of a double integral. Suppose f is a function of two variables (x,y). The *double integral* of f over a subset D of $\mathbb{R}^2$ on which f is defined is the *total contribution* of the f-values at all points in the domain. One way of thinking of it is as follows: we divide D up into a lot of small regions, we pick a point in each region, multiply the f-value at that point with the area of the region and add up. This total gives the integral of f over the region D.

The notation for the double integral of a function f over a region D is:

$$\int\int_D f(x,y)\,dA$$

The dA here represents an area element or area differential, and is the two-dimensional analogue of dx in one dimension. A detailed exploration of the meaning is possible, but beyond our current scope.

The double integral does integration over a *region* in the same way that the ordinary (single) integral does integration over an *interval*. The region over which integration is being done is termed the *region of integration* or *domain of integration* and the function being integrated is termed the *integrand*.

For a function with nonnegative values, the double integral over a region can also be interpreted as a volume. We will see this interpretation a little later here.

1.2. **Linearity.** The double integral of a sum of two functions is the sum of their double integrals:

$$\int\int_D [f(x,y) + g(x,y)]\,dA = \int\int_D f(x,y)\,dA + \int\int_D g(x,y)\,dA$$

Also, scalars can be pulled out of double integrals:

$$\int\int_D cf(x,y)\,dA = c\int\int_D f(x,y)\,dA$$

1.3. **What's an iterated integral?** An iterated integral is an expression that involves an integral inside another integral (and possibly even more integrals. For instance:

$$\int_a^b \left(\int_{p(x)}^{q(x)} f(x,y)\,dy \right) dx$$

What this means is:

- We first compute the inner integral by integrating with respect to y, treating x as a constant. If $F(x, y)$ is an antiderivative, then the definite integral is $F(x, q(x)) - F(x, p(x))$.
- The final answer computed above now depends only on x, the variable y has been integrated over and thus discarded. We now integrate this function of x between the limits a and b.

A special case of this kind of iterated integral is one where the limits for the inner function do not depend on the outer variable, i.e., an integration of the form:

$$\int_a^b \left(\int_p^q F(x, y)\, dy \right) dx$$

Note that we could also consider an iterated integral where the inner variable of integration is x and the outer variable of integration is y.

When things are reasonably clear, we can drop the parenthesization for iterated integrals, so the above can be written as:

$$\int_a^b \int_p^q F(x, y)\, dy\, dx$$

1.4. **Fubini's theorem relating double and iterated integrals on rectangles.** Consider the filled rectangle $R = [a, b] \times [p, q]$ in the xy-plane. This is a rectangle with vertices (a, p), (b, p), (a, q), and (b, q). The region can be described as $\{(x, y) : x \in [a, b], y \in [p, q]\}$. Fubini's theorem for rectangles says that if F is a continuous function of two variables defined on this filled rectangle, then:

$$\int\int_R F(x, y) dA = \int_a^b \int_p^q F(x, y)\, dy\, dx = \int_p^q \int_a^b F(x, y)\, dx\, dy$$

In other words, the double integral equals the iterated integral computed in either order.

The assumption of continuity can be weakened somewhat: we only need to assume that f is bounded on R, and the set of points where it is discontinuous is contained in a union of a finite number of smooth curves. This generalization will help us deduce an important corollary for functions whose domains are not rectangular.

1.5. **Intuitive explanation of Fubini's theorem.** Recall that the double integral of a function $F(x, y)$ can be thought of as follows: $F(x, y)$ denotes the value at point (x, y), and the double integral is the total contribution of all points. For instance, $F(x, y)$ could denote the pressure at the point (x, y), and the double integral over the rectangle/region is the total force exerted on the region.

Iterated integration serves to break this integration up by *slicing* horizontally or vertically. Let's be more specific:

- The iterated integral $\int_a^b \left(\int_p^q F(x, y)\, dy \right) dx$ can be interpreted as follows: the inner integral $\int_p^q F(x, y)\, dy$ is integrating *along a vertical slice* for a fixed value of x (i.e., along a line parallel to the y-axis). The outer integral is then adding up the contributions of all the vertical slices.
- The iterated integral $\int_p^q \left(\int_a^b F(x, y)\, dx \right) dy$ can be interpreted as follows: the inner integral $\int_a^b F(x, y)\, dx$ is integrating *along a horizontal slice* for a fixed value of y (i.e., along a line parallel to the x-axis). The outer integral is then adding up the contributions of all the horizontal slices.

That all these values are the same is some infinite version of the idea that addition is commutative and associative, i.e., we can regroup summations by collecting all things with one common coordinate and then adding up over that coordinate.

1.6. **The special case of multiplicatively separable functions.** A case worth noting is where $F(x, y)$ is of the form $F(x, y) = f(x)g(y)$, i.e., we can separate it as the product of a function purely of x and a function purely of y.

Using the notation established above, if f is continuous on $[a, b]$ and g is continuous on $[p, q]$, then F is continuous on $R = [a, b] \times [p, q]$ and:

$$\int\int_R F(x,y)\,dA = \left(\int_a^b f(x)\,dx\right)\left(\int_p^q g(y)\,dy\right)$$

This is a corollary of Fubini's theorem, and can be deduced by using either of the iterated integrals.

In particular, this means that if F can be written as a *sum* of multiplicatively separable functions, then its integral is a sum of the products of integrals of these functions. In fancy notation, if $f = \langle f_1, f_2, \ldots, f_n\rangle$ and $g = \langle g_1, g_2, \ldots, g_n\rangle$, with all the f_is continuous on $[a,b]$ and all the g_is continuous on $[p,q]$, and if $F(x,y) = \sum_{i=1}^n f_i(x)g_i(y)$, then:

$$\int\int_R F(x,y)\,dA = \sum_{i=1}^n \left[\left(\int_a^b f_i(x)\,dx\right)\left(\int_p^q g_i(y)\,dy\right)\right]$$

Note also that when calculating the integral of a multiplicatively separable function, if either of the integrals of the pieces is zero, the other one does not need to be computed and the product is zero. We will see related ideas a little later when we cover symmetry.

1.7. **A concept of antiderivative.** Suppose G is a function with the property that $G_{xy} = F$, i.e., F is the mixed second-order partial derivative of G. Then, the integral of F over a rectangle $[a,b]\times[p,q]$ is:

$$G(b,q) - G(a,q) - G(b,p) + G(a,p)$$

Basically, the top right and bottom left values get added and the bottom right and top left values get subtracted.

This is sort of like an antiderivative. But the approach is rarely used for explicit computations and we usually try to find definite integrals.

2. Double integrals over regions other than rectangles

2.1. **Defining such a double integral using a rectangle.** Suppose D is a closed bounded region in the plane. In particular, this means that D can be enclosed inside a rectangular region. Suppose R is such a rectangular region. Then, we define the double integral $\int\int_D f(x,y)\,dA$ as $\int\int_R F(x,y)\,dA$ where $F(x,y)$ is defined as:

$$F(x,y) := \{ \begin{array}{ll} f(x,y), & (x,y)\in D \\ 0, & (x,y)\notin D \end{array}$$

In other words, we integrate the function that's f on D and 0 outside. Note that F need not be continuous, even if f is. So, we might be skeptical of applying results such as Fubini's theorem to F. If, however, the boundary of D is a piecewise smooth curve, then by the slightly more general formulation of Fubini's theorem, it turns out that the continuity of f within D allows us to apply Fubini's theorem to F. This is great news because it means that we can compute double integrals as iterated integrals.

2.2. **Type I and Type II regions.** For simplicity, we assume that the regions we are dealing with are all connected, closed, bounded regions and their boundary curves are piecewise smooth.

We call a region D in the xy-plane a *Type I region* if its intersection with every line parallel to the y-axis is either empty, or a point, or a line segment, i.e., the intersection is always connected. Such a region can be described as the region enclosed by the graphs of two continuous functions $y = g_1(x)$ and $y = g_2(x)$, with $g_1(x) \le g_2(x)$, for x in an interval $[a,b]$. The function g_2 is simply the y-coordinate value of the upper endpoint of the line segment and the function g_1 is the y-coordinate value of the lower endpoint of the line segnment. In other words:

$$D = \{(x,y) : a \le x \le b, g_1(x) \le y \le g_2(x)\}$$

We can compute double integrals over Type I regions using iterated integration. To integrate $f(x,y)$ over the Type I region of the kind given above:

$$\int\int_D f(x,y)\,dA = \int_a^b \int_{g_1(x)}^{g_2(x)} f(x,y)\,dy\,dx$$

A *Type II region* is a region whose intersection with every horizontal line is either empty or a point or a line segment. Such a region can be described as the region enclosed by the the graphs of continuous functions with x expressed in terms of y, i.e., something of the form:

$$D = \{(x, y) : p \leq y \leq q, g_1(y) \leq x \leq g_2(y)\}$$

To integrate over the Type II region of the kind given above, we can do the integration:

$$\int\int_D f(x, y)\, dA = \int_p^q \int_{g_1(y)}^{g_2(y)} f(x, y)\, dx\, dy$$

Note that both these results follow from the general version of Fubini's theorem for rectangles, using the trick of transitioning to $F(x, y)$ from $f(x, y)$.

2.3. **Convex regions.** A *convex region* is a region with the property that for any two points in the region, the line segment joining those two points lies completely inside the region. Convex regions are both Type I and Type II. In particular, this means we can use either of the integration methods to compute integrals over convex regions.

Circular disks, triangular regions, and rectangular regions are all examples of convex regions. A heart-shaped region is *not* a convex region.

2.4. **Breaking up a region into Type I and Type II regions.** If a region D is closed, connected, and bounded with a smooth bounding curve, and f is a continuous function of D, it may still happen that D is neither Type I nor Type II. There are still some ways out. The first is to partition D into finitely many pieces (chambers) such that:

- Each piece is Type I or Type II
- The intersection of any two of the pieces is one-dimensional and hence the restriction of the double integral over that intersection is zero.
- The double integral on D is now the sum of the values of double integrals on each piece, and each of the individual double integrals can be computed as an iterated integral by Fubini's theorem.

This is a two-dimensional analogue of chopping up an interval into sub-intervals using a partition. Here, instead of sub-intervals, we use subregions.

In the one-dimensional case, the slight overlap (isolated points) between the partitioned pieces does not result in any double-counting, i.e., the integral on the whole interval is the sum of the integrals on the parts. In the two-dimensional cases, the slight overlap at boundary curves (which are one-dimensional) does not result in any double-counting, because the boundary curves are infinitesimal/negligible.

3. In practice: computing iterated and double integrals

3.1. **Theory versus practice: the one-variable nightmare.** Let's recall the situation in one variable first and then we'll discuss how the situation changes with more variables. We know that any continuous function in one variable is integrable. This knowledge does not always translate to actually being able to find expressions for the integrals. There are three levels of difficulty:

- First, there are many functions expressible in terms of elementary functions but which do not have antiderivatives expressible in terms of elementary functions. To give names to the antiderivatives, we need to invent new branches of mathematics. For instance, logarithms were invented to integrate $1/x$, and trigonometry was invented to integrate $1/(x^2 + 1)$. But there's a lot more work to do – some functions slip through the cracks and integrating them requires us to invent more branches of mathematics.

  Examples of elementarily expressible functions that do not have elementarily expressible antiderivatives are e^{-x^2}, $\sin(x^2)$, $(\sin x)/x$, $(e^x - 1)/x$, $1/\sqrt{x^4 + 1}$, and many others.
- Second, the procedure for integrating a function does not break down into a bunch of deterministic rules. This is in sharp contrast with differentiation, where if we know how to differentiate a bunch of functions, we know how to differentiate all functions generated from them using the processes of pointwise combination, composition, inverses, and piecewise combination. For integration, all we have are heuristics. Thus, even if a neat antiderivative does exist, it can be hard to find.

- Third, even if we are able to find antiderivatives, computing their values between limits can be difficult. Even to integrate a rational function, we need ln and arctan and computing the values of these is hard.

Each of these challenges continues to operate in many variables. With multiple variables, there is some further bad news and some mitigating good news. We turn to these.

3.2. **The further bad news.** The inner-most step of an iterated integral is something like:

$$\int_a^b f(x,y)\,dx$$

Here, we are treating y as a constant temporarily while doing this integration. However, we cannot put an actual value on y – it's an *unknown known* for now, and in fact, when we have completed this integration and are willing to move on outward, it will become a variable again. Thus, this integration really is not integrating a plain vanilla function but rather trying to do a large number of integrations – one for each fixed value of y – *simultaneously* by getting a generic expression.

Now, it may turn out that there is no uniform general expression for y. Consider the example:

$$\int_2^3 \frac{dx}{x^2+\sin y}$$

When $\sin y > 0$, the integral becomes:

$$\frac{1}{\sqrt{\sin y}}\left[\arctan(x/\sqrt{\sin y})\right]_2^3$$

When $\sin y = 0$, the integral becomes:

$$[-1/x]_2^3 = (1/2) - (1/3) = 1/6$$

When $\sin y < 0$, the integral becomes:

$$\frac{1}{2\sqrt{-\sin y}}\left[\ln((x-\sqrt{-\sin y})/(x+\sqrt{-\sin y}))\right]_2^3$$

So, even though the original function had a single piece description, the new function we get after integrating has a *piecewise description*.

This will occur only rarely, and not in the routine examples that we will see. Also, although it complicates matters, it does not make the task any more impossible. To perform the outer integration for the resultant piecewise function, we simply break the domain (for the outer integration) into the various pieces and perform the integration separately in each piece.

3.3. **More bad news for non-rectangular regions.** Another piece of bad news, that applies particularly to non-rectangular regions, is that complications could arise not only from the nature of the integrand, but also from the shape of the region. For Type I or Type II regions, the nature of the bounding functions that determine how the inner variable varies in terms of the outer variable determine the expression to be integrated on the outside. Thus, even for very easy functions $f(x,y)$, the actual integration procedure may become difficult because of the complexity arising from the shape of the region.

3.4. **The good news: use Fubini's to change order of integration.** The good news is that sometimes, an integral is impossible to do when written as an iterated integral with a particular ordering of x and y, but can be done if the ordering of x and y were reversed. Luckily, by Fubini's theorem, the answers have the same value.

Let's consider a couple of examples.

Our first example is the function x^y on the interval $[0,1]\times[0,1]$. The domain is a square region with vertices $(0,0)$, $(0,1)$, $(1,0)$ and $(1,1)$. Note that the function is undefined at the bottom left vertex $(0,0)$. It takes the value 1 on the lower edge, 0 on the left edge, 1 on the right edge, and is equal to the function x on the top edge. Note that everywhere in the square where it is defined, the function takes a value in $[0,1]$. We want to integrate it over the square.

We could set up the integral as an iterated integral in either of these two ways:

$$\int_0^1 \int_0^1 x^y \, dy \, dx, \qquad \int_0^1 \int_0^1 x^y \, dx \, dy$$

Let's consider the first formulation of the integral:

$$\int_0^1 \int_0^1 x^y \, dy \, dx$$

The inner integral is:

$$\int_0^1 x^y \, dy$$

This simplifies to:

$$\left[\frac{x^y}{\ln x}\right]_0^1 = \frac{x-1}{\ln x}$$

The new integral that we need to compute is thus:

$$\int_0^1 \frac{x-1}{\ln x} \, dx$$

The *indefinite integral* of the integrand is not possible to compute. So we're basically stuck.

On the other hand, if we use the other formulation (*FIXED ERROR BELOW!*:

$$\int_0^1 \int_0^1 x^y \, dx \, dy$$

The inner integral is:

$$\int_0^1 x^y \, dx$$

This simplifies to:

$$\left[\frac{x^{y+1}}{y+1}\right]_0^1 = \frac{1}{y+1}$$

We can now integrate this:

$$\int_0^1 \frac{1}{y+1} \, dy = [\ln(y+1)]_0^1 = \ln 2$$

Note that in this example, integrating in the wrong order got us into problems at the *outer stage*, not at the inner stage. In some cases, integrating in the wrong order can prevent us from getting started. Here is an example:

$$\int_0^1 \int_x^1 \exp(-y^2) \, dy \, dx$$

This is the integral of the function $\exp(-y^2)$ over the triangular region for the triangle with vertices $(0,0)$, $(1,1)$, and $(0,1)$, i.e., the upper left half triangle in the unit square $[0,1] \times [0,1]$. Unfortunately, as written here, the inner integral cannot be computed in elementary terms.

Note that the region here is both a Type I and a Type II region. This means that it can be sliced either vertically or horizontally. If we slice horizontally instead, then for any fixed y, the constraint on x is $0 \leq x \leq y$, and we get:

$$\int_0^1 \int_0^y \exp(-y^2) \, dx \, dy$$

The inner integral is now:

$$\int_0^y \exp(-y^2)\,dx = y\exp(-y^2)$$

The outer integral now becomes:

$$\int_0^1 y\exp(-y^2)\,dy = \left[\frac{-1}{2}\exp(-y^2)\right]_0^1 = \frac{1}{2}\left(1-\frac{1}{e}\right)$$

This is similar to Example 3 in the book.

3.5. **Integrating polynomials.** Polynomials are very easy to integrate *over rectangular regions* because every polynomial is a sum of monomials, every monomial is multiplicatively separable as a product of power functions, and each power function can be integrated.

For instance, to integrate the polynomial $xy + 2x^5y^3$ over the interval $[1,3]\times[4,6]$, we do:

$$\int_1^3 x\,dx\int_4^6 y\,dy + 2\int_1^3 x^5\,dx\int_4^6 y^3\,dy$$

In other words, it is a sum of products of integrals of power functions of one variable. The rest is just straightforward arithmetic.

To integrate a polynomial over a non-rectangular region is a little trickier, and may not be feasible for all regions. First, note that we can still additively separate the polynomial as a sum of monomials, so it suffices to integrate each monomial, i.e., each expression of the form x^ay^b. However, because the region is no longer rectangular, we cannot use multiplicative separability.

Here's an example. Consider integrating x^2y^2 on the circular disk $x^2+y^2\leq 1$. This is both Type I and Type II. If we go for horizontal slicing, then for $x\in[-1,1]$, we have $-\sqrt{1-x^2}\leq y\leq\sqrt{1-x^2}$. The integral thus becomes:

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

The inner integral becomes $x^2y^3/3$ which between limits is $2x^2(1-x^2)^{3/2}/3$. This needs to be integrated on $[-1,1]$. Note how, even though we started only with polynomials, the integrand for the outer integration involves fractional powers. The fractional powers are arising from the shape of the domain of integration.

It is possible to complete the question in the case of circular disks, but this is best done using double integrals in polar coordinates, covered in Section 16.4 of the book. We will, however, not cover this topic formally as part of the syllabus, although I will explain it in class and give a few examples.

3.6. **Integrating rational functions.** We first consider integrating rational functions over rectangular regions. If the denominator of the rational function is of the form cx^ay^b (i.e., it is a monomial) then the rational function is a sum of multiplicatively separable functions and can be integrated using the same idea discussed above for polynomials.

More generally, if the denominator of the rational function can be factorized as the product of a polynomial in x and a polynomial in y, we can use the multiplicatively separable approach.

For instance, for the rational function:

$$\frac{x^2+y^2-2xy+3}{x^2y^2+2x^2+y^2+2}$$

The denominator can be factored as $(x^2+1)(y^2+2)$ and hence the rational function can be written as:

$$\frac{x^2}{x^2+1}\frac{1}{y^2+1} + \frac{1}{x^2+1}\frac{y^2}{y^2+2} - 2\frac{x}{x^2+1}\frac{y}{y^2+2} + 3\frac{1}{x^2+1}\frac{1}{y^2+2}$$

This is a multiplicatively separable form, and can be integrated over a rectangular region. Note: We are assuming knowledge of how to integrate rational functions of one variable, something you saw in one variable calculus.

In other cases, it is not completely obvious how to do the integration, so we just try iterated integration and see how it works out. For instance, consider:

$$\int_1^2 \int_1^2 \frac{1}{x+y}\, dy\, dx$$

The inner integral is $\ln(x+2) - \ln(x+1)$. The outer integral thus becomes:

$$\int_1^2 \ln(x+2) - \ln(x+1)\, dx$$

After some integration by parts (we skip steps) this becomes:

$$[(x+2)\ln(x+2) - (x+1)\ln(x+1)]_1^2 = 4\ln 4 - 3\ln 3 - 3\ln 3 + 2\ln 2 = 10\ln 2 - 6\ln 3$$

In fact, it is possible to give a sketch for why this kind of integration procedure will work for a wide variety of (all?) rational functions.

3.7. **Exponential and trigonometric functions.** Again, for these, one thing to look for is multiplicative separability, or expressibility as a sum of multiplicatively separable functions, and hope that each of the constituent functions of one variable can be integrated.

Consider $f(x,y) = \sin(x+y)$. We want to calculate:

$$\int_a^b \int_p^q \sin(x+y)\, dy\, dx$$

We could do this directly, integrating first with respect to y, to get $-\cos(x+q) + \cos(x+p)$ and then integrating with respect to x to get $-\sin(b+q) + \sin(a+q) + \sin(b+p) - \sin(a+p)$.

Alternatively, we could rewrite $\sin(x+y) = \sin x \cos y + \cos x \sin y$ and integrate by additively and then multiplicatively splitting, to get:

$$(\cos a - \cos b)(\sin q - \sin p) + (\sin b - \sin a)(\cos p - \cos q)$$

It's possible to work out that both of these are the same. (Note: This is easier to see if we use the "antiderivative" concept mentioned earlier: the antiderivative by iterated integration is $-\sin(x+y)$ and the antiderivative by multiplicative separation is $-\cos x \sin y - \sin x \cos y$ which becomes the same thing.

For exponential functions, note that $\exp(f(x) + g(y)) = \exp(f(x))\exp(g(y))$ and is hence multiplicatively separable.

4. Area and volume interpretations

To make this course simpler, we will refrain from complicated volume computations for the surfaces that are graphs of functions, but we will go over the theoretical facts just in case you need them for the future.

4.1. **Double integral equals volume.** Consider a function $z = f(x,y)$ on a closed connected bounded domain D such that $z \geq 0$ for all $(x,y) \in D$. Then, the integral $\int \int_D f(x,y)\, dA$ equals the volume of the region between the surface $z = f(x,y)$, the xy-plane. On the sides, this region is bounded by line segments joining points in the boundary of D and the corresponding points on the graph of the surface above them.

The three-dimensional region can also be described as follows: it is the union of all the line segments obtained by joining each point $(x,y,0)$ with the point $(x,y,f(x,y))$ where $(x,y) \in D$.

The fact that the double integral value equals the volume is the three-dimensional analogue of the fact that the single integral value equals the area under the graph of the function.

4.2. **Interpretation of slicing and iterated integration.** We can now interpret the horizontal and vertical slicing.

Computing the integral along a horizontal slice, i.e., a line parallel to the xy-plane, correspondings to computing the area of the intersection of the region with a plane parallel to the xz-plane through that line. Specifically, computing the integral:

$$\int_a^b f(x, y_0)\, dx$$

means computing the area of the intersection of the region with the plane $y = y_0$, or equivalently, computing the area under the graph of the function $x \mapsto f(x, y_0)$ between $x = a$ and $x = b$.

The outer part of the integration then integrates this area function along the other axis, to give the total volume.

If we perform the integration in the other order, we are computing the areas of intersection with planes parallel to the yz-plane, and then integrating this area function along the x-axis.

Note that all this fits in with the cross sectional method of determining volume as the integral of the areas of the cross sections along planes as we move along an axis perpendicular to these planes.

5. Properties of double integrals

5.1. **Inequalities that can be used for estimation.** These inequalities are a lot like those of single integrals:

- If $f(x, y) \geq 0$ on a domain D, then $\int\int_D f(x, y)\, dA \geq 0$.
- If $f(x, y) \geq g(x, y)$ on a domain D, then $\int\int_D f(x, y)\, dA \geq \int\int_D g(x, y)\, dA$.
- If $D_1 \subseteq D_2$, and f is a *nonnegative function* defined on D_2, then $\int\int_{D_1} f(x, y)\, dA \leq \int\int_{D_2} f(x, y)\, dA$: This last one is important because it means that to calculate integrals over an irregularly shaped region, we can bound from above and below by calculating integrals over a region contained inside it and over a region containing it.
- If $D = D_1 \cup D_2$ and $D_1 \cap D_2$ is one-dimensional, then $\int\int_D f(x, y) dA = \int\int_{D_1} f(x, y)\, dA + \int\int_{D_2} f(x, y)\, dA$ (this was already discussed earlier).
- The integral of the function 1 over a domain D is the area of D.
- If $f(x, y)$ on a domain D is bounded from above and below by M and m respectively, and D has area A, then the integral of $f(x, y)$ over D is between mA and MA.

5.2. **Symmetry-based ideas.** These all build on the corresponding symmetry-based ideas for functions of one variable:

- If f is odd in the variable x, and the domain of integration is symmetric about the y-axis, then the integral is zero: If we do integration using horizontal slices, we see that each horizontal slice integrates to zero, so the overall integral is zero.

  Note that for a multiplicatively separable function, what matters is that the part depending on x be odd, and the part depending on y does not matter.
- If f is odd in the variable y, and the domain of integration is symmetric about the x-axis, then the integral is zero: If we do integration using vertical slices, we see that each vertical slice integrates to zero, so the overall integral is zero.

  Note that for a multiplicatively separable function, what matters is that the part depending on y be odd, and the part depending on x does not matter.

Often, a given function can be expressed as a sum of functions some of which are odd in x or in y, and hence, using symmetry of domain, can be declared to be zero. Others may need to be computed.

Consider, for instance, the case $f(x, y) = x^3y^2 + \ln(x^2 + x + 1)\sin(y^3)$, being integrated over the circular disk $x^2 + y^2 \leq 1$. Note that f as given is not odd in either variable. However, it is the sum of the functions x^3y^2 (which is odd in x) and $\ln(x^2 + x + 1)\sin(y^3)$ (which is odd in y). Moreover, the domain is symmetric about both axes. Thus, the integral for both these functions is zero, hence the overall integral for f is zero.

DIRECTIONAL DERIVATIVES AND GRADIENT VECTORS

MATH 195, SECTION 59 (VIPUL NAIK)

Corresponding material in the book: Section 14.6.

What students should definitely get: Definition of directional derivative and gradient vector, gradient vector as direction with maximum magnitude of directional derivative, directional derivative as dot product of gradient vector and unit vector in the direction. Formulas for tangent plane and normal line at a point to a surface with a relational description.

Executive summary

Words ...

(1) The directional derivative of a scalar function f of two variables along a *unit* vector $\mathbf{u} = a\mathbf{i} + b\mathbf{j}$ at a point (x_0, y_0) is defined as the following limit of difference quotient, if the limit exists:

$$\lim_{h \to 0} \frac{f(x_0 + ah, y_0 + bh) - f(x_0, y_0)}{h}$$

(2) The directional derivative of a differentiable scalar function f of two variables along a *unit* vector $\mathbf{u} = a\mathbf{i} + b\mathbf{j}$ at a point (x_0, y_0) is $D_{\mathbf{u}}(f) = af_x(x_0, y_0) + bf_y(x_0, y_0)$.

(3) The gradient vector for a differentiable scalar function f of two variables at a point (x_0, y_0) is $\nabla f(x_0, y_0) = f_x(x_0, y_0)\mathbf{i} + f_y(x_0, y_0)\mathbf{j}$.

(4) The directional derivative of f is the dot product of the gradient vector of ∇f and the unit vector $\mathbf{u}$.

(5) Suppose ∇f is nonzero. Then, if $\mathbf{u}$ makes an angle θ with ∇f, then $D_{\mathbf{u}}(f)$ is $|\nabla_f| \cos\theta$. The directional derivative is maximum in the direction of the gradient vector, zero in directions orthogonal to the gradient vector, and minimum in the direction opposite to the gradient vector.

(6) The level curves are orthogonal to the gradient vector.

(7) Similar formulas for gradient vector and directional derivative work in three dimensions.

(8) The level surfaces are orthogonal to the gradient vector for a function of three variables.

(9) For a surface given by $F(x, y, z) = 0$, if (x_0, y_0, z_0) is a point on the surface, and $F_x(x_0, y_0, z_0)$, $F_y(x_0, y_0, z_0)$, and $F_z(x_0, y_0, z_0)$ all exist and are nonzero, then the normal line is given by:

$$\frac{x - x_0}{F_x(x_0, y_0, z_0)} = \frac{y - y_0}{F_y(x_0, y_0, z_0)} = \frac{z - z_0}{F_z(x_0, y_0, z_0)}$$

The tangent plane is given by:

$$F_x(x_0, y_0, z_0)(x - x_0) + F_y(x_0, y_0, z_0)(y - y_0) + F_z(x_0, y_0, z_0)(z - z_0) = 0$$

1. Directional derivatives: definition and key facts

1.1. **Partial derivatives as derivatives along coordinate directions.** The partial derivative $f_x(x, y)$ is defined as the derivative of f with respect to x, keeping y constant. Thinking of the domain of f geometrically, this is the same as the derivative of f along a unit vector in the x-direction (which we archaically denote by $\mathbf{i}$). Similarly, the partial derivative $f_y(x, y)$ is defined as the derivative of f with respect to y, keeping x constant. This can be viewed as the derivative of f with respect to a unit vector in the y-direction (which we archaically denote by $\mathbf{j}$).

We may be interested in derivatives with respect to mixed directions, i.e., we may be interested in the question: if we move along the direction of the vector $\mathbf{i} + \mathbf{j}$, how does the function value change?

The correct notion is that of *directional derivative*. After defining this notion, we consider its implication both for functions that have physical significance and for the more abstract functions in economics (such as production and demand functions).

1.2. **Definition of directional derivative.** For a function f, the directional derivative of f at (x_0, y_0) in the direction of a unit vector $\mathbf{u} = \langle a, b \rangle$ (i.e., $a^2 + b^2 = 1$) is denoted $D_{\mathbf{u}} f(x_0, y_0)$ and is defined as:

$$D_{\mathbf{u}} f(x_0, y_0) = \lim_{h \to 0} \frac{f(x_0 + ha, y_0 + hb) - f(x_0, y_0)}{h}$$

if the limit exists.

The special case of $\mathbf{u} = \mathbf{i} = \langle 1, 0 \rangle$ gives the partial derivative $f_x(x, y)$ and the case $\mathbf{u} = \mathbf{j} = \langle 0, 1 \rangle$ gives the partial derivative $f_y(x, y)$.

1.3. **Existence and computation of directional derivative.** It turns out that there is a notion of "differentiable" for a function of two variables (which are are avoiding discussion of) and if a function is differentiable at a point in that sense, then it has well-defined directional derivatives along all unit vectors.

Further, then the directional derivative in the direction of $\mathbf{u} = \langle a, b \rangle$ is given by:

$$D_{\mathbf{u}} f(x_0, y_0) = a f_x(x_0, y_0) + b f_y(x_0, y_0)$$

A *sufficient* condition for a function to be differentiable is that both the partial derivative f_x and f_y exist and are continuous *around* the point.

1.4. **Geometric sense of directional derivatives.** Directional derivatives make direct geometric sense in cases where the functions actually have physical significance. For instance, consider a function whose input domain is a flat surface and whose output value at any point on the surface is the temperature at that point. The *directional* derivative at a point with respect to a direction at that point can be thought of as the rate at which the temperature changes if you move along that direction (physically) starting at that point.

More generally, suppose you move along a curve in the surface that's the domain of the function. You want to find the rate at which the temperature is changing along the curve that you are moving along. This rate of temperature change is the product of the (directional derivative along the unit vector tangent direction to the curve at the point) times the (length of the tangent vector, i.e., the speed of motion).

1.5. **Sense of directional derivatives in non-physical contexts.** Consider the case of a production function $f(L, K)$ with inputs L (the expenditure on labor) and K (the expenditure on capital). We already made sense of the partial derivatives $f_L(L, K)$ and $f_K(L, K)$. The partial derivative $f_L(L, K)$ is the marginal change in output for a marginal change in the labor input (i.e., it is the marginal product of labor). Similarly, the partial derivative $f_K(L, K)$ is the marginal chane in output for a marginal change in the capital input (i.e., it is the marginal product of capital).

Let's think of what it means to take the directional derivative along the vector $\langle a, b \rangle$ with $a^2 + b^2 = 1$. What this basically means is the following: we want to measure the marginal change in output if the inputs L and K are changed marginally in the ratio $a : b$. In other words, it measures the impact on output of a particular *combined trajectory of change* in the values of L and K.

2. The gradient vector

2.1. **The direction of change, the direction of no change.** The *gradient vector* for a differentiable function f of two variables is defined as follows:

$$\nabla f(x, y) = \langle f_x(x, y), f_y(x, y) \rangle = \frac{\partial f}{\partial x}\mathbf{i} + \frac{\partial f}{\partial y}\mathbf{j}$$

Note that the gradient vector, thus viewed, is a *vector-valued function of two variables*, i.e., it has type $2 \to 2$. The gradient vector at a particular point is, however, an actual vector, i.e., an actual tuple of two real numbers.

If the gradient vector is nonzero, the *unit vector* in this direction can be computed by dividing this vector by its length.

Here are some key observations that hold if the gradient vector is nonzero:

- Of all the unit vector directions, the directional derivative is maximum (and positive) along the unit vector in the same direction as the gradient vector, and is minimum (and negative) along the unit vector in the opposite direction to the gradient vector.
- The directional derivative is zero along the directions *perpendicular* to the gradient vector.

Intuitively the gradient vector is telling us the direction along which the change/action is happening, and also telling us that there's no action happening orthogonal to it.

2.2. **Special case of function depending on only one variable.** If the function f depends only on the variable x and has no dependence on the variable y, then the gradient vector, where nonzero, will always point parallel to the x-axis (in a positive or negative direction, depending on whether the function is increasing or decreasing).

2.3. **Writing directional derivative in terms of gradient vector.** The directional derivative along a unit vector $\mathbf{u}$ can be defined as the dot product $(\nabla f) \cdot \mathbf{u}$. Since $\mathbf{u}$ is a unit vector, this can be interpreted as the scalar projection of ∇f along $\mathbf{u}$.

Intuitively, what this means is that the extent to which the function is changing in a particular direction depends on the component of the gradient vector that falls in that direction.

Another way of thinking of the directional derivative $D_{\mathbf{u}}(f)$ for a unit vector $\mathbf{u}$ is as $|\nabla f| \cos\theta$ where θ is the angle between ∇f and $\mathbf{u}$. Note that this is equal to $|\nabla f|$ when $\mathbf{u}$ is in the direction of ∇f, and it is $-|\nabla f|$ when $\mathbf{u}$ is opposite to ∇f. It is 0 when $\mathbf{u}$ is orthogonal to ∇f.

Note that if the gradient vector is zero, then there is no direction to it, and the directional derivative along *every* direction is zero.

2.4. **Relationship of gradient vector and level curves.** Recall that the gradient vector represents the direction along which all the change in the function value is happening. It should thus come as no surprise that at any point, the level curve through that point is orthogonal to the gradient vector at that point (note that the statement is trivially true if the gradient vector is zero, but gives no geometric information in that case). Further, if we consider the picture of level curves along with the function values for each curve, then the gradient vector points in the direction of increasing function values.

3. Case of functions of three or more variables

3.1. **Description.** The same expressions for gradient vector and directional derivative apply to functions of three or more variables.

Basically:

- The directional derivative along a unit vector can be defined as a limit of a difference quotient. In three variables: For a function $f(x, y, z)$ at a point (x_0, y_0, z_0) with unit vector $\mathbf{u} = \langle a, b, c \rangle$, this becomes:

$$\lim_{h \to 0} \frac{f(x_0 + ah, y_0 + bh, z_0 + ch)}{h}$$

- For a differentiable function, the gradient vector is defined as the vector obtained by adding, for each coordinate direction, the partial derivative in that direction times the unit vector in that direction. In three variables: For a function $f(x, y, z)$, this becomes:

$$\nabla f(x, y, z) = f_x(x, y, z)\mathbf{i} + f_y(x, y, z)\mathbf{j} + f_z(x, y, z)\mathbf{k}$$

  This is a vector-valued function, i.e., it has type $3 \to 3$ (and more generally $n \to n$). At a particular point in the domain, it gives an actual vector, i.e., a tuple of 3 (respectively n) real numbers.
- For a differentiable function, and in particular for a function where all the first partials exist and are continuous around the point, the directional derivative along a unit vector is the dot product of the gradient vector and that unit vector.
- The gradient vector at a point (if nonzero) is orthogonal to the level surface for the function at the point, or equivalently, it is orthogonal to the tangent plane for the level surface.

3.2. **Application to finding normal direction and tangent plane.** The fact that the gradient vector points in the normal direction and is hence orthogonal to the tangent plane provides a strategy to compute the tangent plane for any point on a surface in $\mathbb{R}^3$ given by a top-down (relational) description of the form $F(x, y, z) = 0$. Namely, to find the normal vector at a point (x_0, y_0, z_0), we compute $\nabla F(x_0, y_0, z_0)$. If this is nonzero, it is a normal vector. We can convert it to a unit vector if desired. Next, we use the technique for finding the scalar equation of a plane to obtain the scalar equation of the tangent plane.

Instead of working things out in each case using vectors, we can also directly determine the scalar version and apply these directly. For a relational description $F(x, y, z) = 0$ and a point (x_0, y_0, z_0), the symmetric equations of the normal line are:

$$\frac{x - x_0}{F_x(x_0, y_0, z_0)} = \frac{y - y_0}{F_y(x_0, y_0, z_0)} = \frac{z - z_0}{F_z(x_0, y_0, z_0)}$$

The equation of the tangent plane is:

$$F_x(x_0, y_0, z_0)(x - x_0) + F_y(x_0, y_0, z_0)(y - y_0) + F_z(x_0, y_0, z_0)(z - z_0) = 0$$

Note that this method does not work if ∇F *is* 0. In other words, it does not work if all the three first partials $F_x(x_0, y_0, z_0)$, $F_y(x_0, y_0, z_0)$, and $F_z(x_0, y_0, z_0)$ equal 0. In this case, it may happen either that there is no tangent plane, or it may happen that the tangent plane exists but cannot be found through this procedure.

Recall that we had earlier determined the equation to the tangent plane for $z = f(x, y)$, which is a special case of the above with $F(x, y, z) = f(x, y) - z$. It can be verified (See the book) that the earlier formula is consistent with the formula obtained above.

Note: application to temperature. We consider the example of temperature.

Consider the temperature function defined on the surface of the earth that sends a point to the surface of the earth to the surface temperature at that point.

The level curves for this are the *isothermal lines* and represent curves of constant temperature. The gradient vector at any point is a vector tangential to the sphere and represents the direction in which temperature is changing fastest. The positive direction along the gradient vector is the direction of fastest temperature increase. The negative direction is the direction of fastest temperature decrease.

The directional derivative along a direction tangential to the surface of the earth describes the rate at which temperature changes if we move along that direction.

The same example of temperature can be adapted to a three-dimensional setting instead of the surface of the earth. For instance, in a three-dimensional container that is not in thermal equilibrium (so different parts have different temperatures) we can consider the temperature as a function of the location in space. Then, the level *surfaces* are the isothermal surfaces, and the gradient vector at a point is orthogonal to the level surface and describes the direction along which temperature is changing fastest. The directional derivative along a particular direction is the rate at which temperature changes if we move along that direction.

MAXIMUM AND MINIMUM VALUES: ONE VARIABLE AND TWO

MATH 195, SECTION 59 (VIPUL NAIK)

Corresponding material in the book: Section 14.7.

What students should definitely get: The definition of critical point in terms of gradient vector being zero, the second derivative test, the use of these to find local and absolute extreme values.

What students should hopefully get: The similarities and differences between the situations for one and two variables.

Executive summary

Words ...

(1) For a directional local minimum, the directional derivative (in the outward direction from the point) is greater than or equal to zero. For a directional local maximum, the directional derivative (in the outward direction from the point) is less than or equal to zero.

Note that even for *strict* directional local maximum or minimum, the possibility of the directional derivative being zero cannot be ruled out.

(2) If a point is a point of directional local minimum from two opposite directions (i.e., it is a local minimum along a line through the point, from both directions on the line) then the directional derivative along the line, if it exists, must equal zero.

(3) If a function of two variables is differentiable at a point of local minimum or local maximum, then the directional derivative of the function is zero at the point in every direction. Equivalently, the gradient vector of the function at the point is the zero vector. Equivalently, both the first partial derivatives at the point are zero.

Points where the gradient vector is zero are termed *critical points*.

(4) If the directional derivatives along some directions are positive and the directional derivatives along other directions are negative, the point is likely to be a *saddle point*. A saddle point is a point for which the tangent plane to the surface that's the graph of the function slides through the graph, i.e., it is not completely on one side.

(5) For a function f of two variables with continuous second partials, and a critical point (a, b) in the domain (so $f_x(a, b) = f_y(a, b) = 0$) we compute the Hessian determinant:

$$D(a, b) = f_{xx}(a, b)f_{yy}(a, b) - [f_{xy}(a, b)]^2$$

If $D(a, b) > 0$ and $f_{xx}(a, b) > 0$, the function has a local *minimum* at the point (a, b). If $D(a, b) > 0$ and $f_{xx}(a, b) < 0$, the function has a local *maximum* at the point (a, b). If $D(a, b) < 0$, we get a saddle point at the point. If $D(a, b) = 0$, the situation is inconclusive, i.e., the test is indecisive.

(6) For a closed bounded subset of $\mathbb{R}^n$ (and specifically $\mathbb{R}^2$) any continuous function with domain that subset attains its absolute maximum and minimum values. These values are attained either at interior points (in which case they are local extreme values and must be attained at critical points) or at boundary points.

(7) *Relation with level curves*: *Typically*, local extreme values correspond to isolated single point level curves. However, this is not always the case, and there are some counterexamples. To be more precise, any *isolated* or *strict* local extreme value corresponds to a (locally) single point level curve.

Actions ...

(1) Strategy for finding local extreme values: First, find all the critical points by solving $f_x(a, b) = 0$ and $f_y(a, b) = 0$ as a pair of simultaneous equations. Next, use the second derivative test for each critical point, and if feasible, try to figure out if this is a point of local maximum, or local minimum, or a saddle point.

(2) To find absolute extreme values of a function on a closed bounded subset of $\mathbb{R}^2$, first find critical points, then find critical points for a parameterization of the boundary, and then compute values at all of these and see which is largest and smallest. *If the list of critical points is finite, and we need to find absolute maximum and minimum, it is not necessary to do the second derivative test to figure out which points give local maximum, local minimum, or neither, we just need to evaluate at all points and find the maximum/minimum.*
(3) When the domain of the function is bounded but not closed, we must consider the possibility of extreme values occurring as we approach boundary points not in the domain. If the domain is not bounded, we must consider directions of approach to infinity.

1. Local increase and decrease behavior: one variable recall

1.1. **Larger than stuff on the left.** Suppose c is a point and $a < c$ such that $f(x) \leq f(c)$ for all $x \in (a, c)$. In other words, c is a *local maximum from the left*. What do I mean by that? I mean that $f(c)$ is larger than or equal to f of the stuff on the *immediate* left of it. That doesn't mean that $f(c)$ is a maximum over the entire domain of f – it just means it is greater than or equal to stuff on the immediate left.

Now, we claim that, if the left-hand derivative of f at c exists, then it is greater than or equal to 0. How do we work that out? The left-hand derivative is the limit of the difference quotient:

$$\frac{f(x) - f(c)}{x - c}$$

where $x \to c^-$. Note that for x close enough to c, (i.e., $a < x < c$), the numerator is negative or zero, and the denominator is negative, so the difference quotient is zero or positive. Thus, the limit of this, if it exists, is zero or positive.

There are three other cases. Let's just summarize the four cases:

(1) If c is a point that is a local maximum from the left for f, then the left-hand derivative of f at c, if it exists, is zero or positive.
(2) If c is a point that is a local maximum from the right for f, then the right-hand derivative of f at c, if it exists, is zero or negative.
(3) If c is a point that is a local minimum from the left for f, then the left-hand derivative of f at c, if it exists, is zero or negative.
(4) If c is a point that is a local minimum from the right for f, then the right-hand derivative of f at c, if it exists, is zero or positive.

1.2. **Strict maxima and minima.** We said that for a function f, a point c is a *local maximum from the left* if there exists $a < c$ such that $f(x) \leq f(c)$ for all $x \in (a, c)$. Now, this definition also includes the possibility that the function is constant just before c.

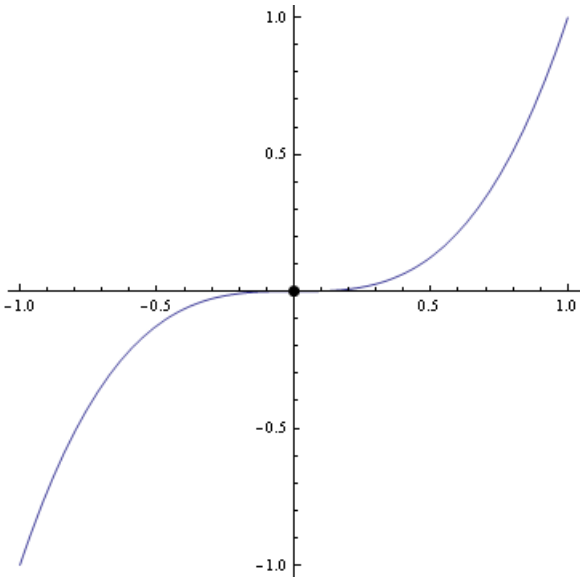
A related notion is that of a *strict local maximum from the left*, which means that there exists $a < c$ such that $f(x) < f(c)$ for all $x \in (a, c)$. In other words, $f(c)$ is *strictly bigger* than $f(x)$ for x to the immediate left of c.

Similarly, we can define the notions of strict local maximum from the right, strict local minimum from the left, and strict local minimum from the right.

1.3. **Does strict maximum/minimum from the left/right tell us more?** Recall that if c is a point that is a local maximum from the left for f, then the left-hand derivative of f at c, if it exists, is greater than or equal to zero. What if c is a point that is a strict local maximum from the left for f? Can we say something more about the left-hand derivative of f at c?

The first thing you might intuitively expect is that that left-hand derivative of f at c should now not just be greater than or equal to zero, it should be strictly greater than zero. But you would be wrong.

It *is* true that if c is a strict local maximum from the left for f, then the difference quotients, as $x \to c^-$, are all positive. However, the *limit* of these difference quotients could still be zero. Another way of thinking about this is that even if the function is increasing up to the point c, it may happen that the rate of increase is leveling off to 0. An example is the function x^3 at the point 0: 0 is a strict local maximum from the left, but the derivative at 0 is 0. Here's a picture:



As you know by now, the phenomenon we are dealing with in this case is called a *point of inflection*. However, there are many other cases where a similar phenomenon occurs, as will be clear soon.

1.4. **Minimum, maximum from both sides.** So we have some sign information about the derivative closely related to how the function at the point compares with the value of the function at nearby points. Maximum from the left means left-hand derivative is nonnegative, maximum from the right means right-hand derivative is nonpositive, minimum from the left means left-hand derivative is nonpositive, minimum from the right means right-hand derivative is nonnegative.

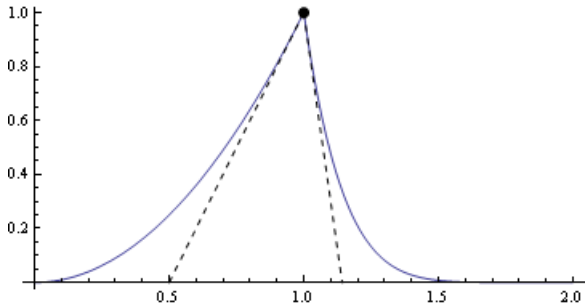
So, let's piece these together:

(1) A *local maximum* for the function f is a point c such that $f(c)$ is the maximum possible value for $f(x)$ in an open interval containing c. Thus, a point of local maximum for f is a point that is both a local maximum from the left and a local maximum from the right. A *strict local maximum* for the function f is a point c such that $f(c)$ is strictly greater than $f(x)$ for all x in some open interval containing c.
(2) A *local minimum* for the function f is a point c such that $f(c)$ is the minimum possible value for $f(x)$ in an open interval containing c. Thus, a point of local minimum for f is a point that is both a local minimum from the left and a local minimum from the right. A *strict local minimum* for the function f is a point c such that $f(c)$ is strictly smaller than $f(x)$ for all x in some open interval containing c.

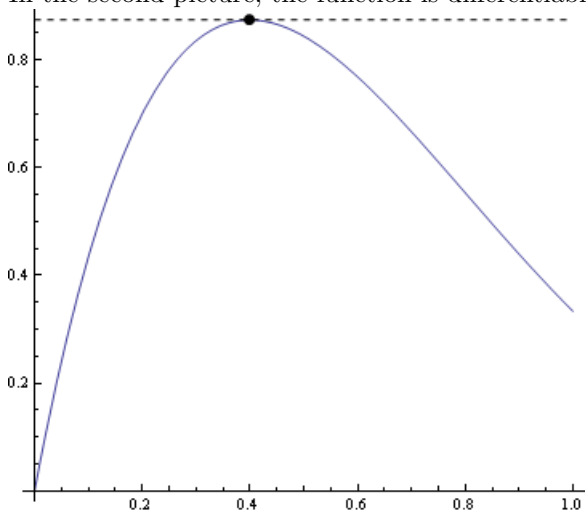
What can we say about local maxima and local minima? We can say the following:

(1) At a local maximum, the left-hand derivative (if it exists) is greater than or equal to zero, and the right-hand derivative (if it exists) is less than or equal to zero. Thus, *if* the derivative exists at a point of local maximum, it *equals zero*. The same applies to strict local maxima.
(2) At a local minimum, the left-hand derivative (if it exists) is less than or equal to zero, and the right-hand derivative (if it exists) is greater than or equal to zero. Thus, *if* the derivative exists at a point of local minimum, it *equals zero*. The same applies to strict local minima.

Below are two pictures depicting points of local maximum. In the first picture, the left-hand derivative is positive, the right-hand derivative is negative, and the function is not differentiable at the point of local maximum.



In the second picture, the function is differentiable, and the derivative is zero.



1.5. **Maximum from the left, minimum from the right.** Suppose c is a point such that it is a local maximum from the left for f and is a local minimum from the right for f. This means that $f(c)$ is greater than or equal to $f(x)$ for x to the immediate left of c, and $f(c)$ is less than or equal to $f(x)$ for x to the immediate right of c. In this case, we say that f is *non-decreasing* at the point c.

In other words, f at c is bigger than or equal to what it is on the left and smaller than or equal to what it is on the right. Well, in this case, the left-hand derivative is greater than or equal to zero and the right-hand derivative is greater than or equal to zero. Thus, if $f'(c)$ exists, we have $f'(c) \geq 0$.

Now consider the case where c is a point that is a local minimum from the left for f and is a local maximum from the right for f. This means that $f(c)$ is less than or equal to $f(x)$ for x to the immediate left of c and greater than or equal to $f(x)$ for x to the immediate right of c. In this case, we say that f is *non-increasing* at the point c.

In other words, f at c is smaller than what it is on the right and larger than what it is on the left. Well, in this case, the left-hand derivative is less than or equal to zero and the right-hand derivative is less than or equal to zero. Thus, if $f'(c)$ exists, we have $f'(c) \leq 0$.

1.6. **Introducing strictness.** We said that f is *non-decreasing* at the point c if $f(c) \geq f(x)$ for x just to the left of c and $f(c) \leq f(x)$ for x just to the right of c. We now consider the *strict* version of this concept. We say that f is *increasing* at the point c if there is an open interval (a, b) containing c such that, for $x \in (a, b)$, $f(x) < f(c)$ if $x < c$ and $f(x) > f(c)$ if $x > c$. In other words, c is a strict local maximum from the left and a strict local minimum from the right.

Well, what can we say about the derivative at a point where the function is increasing, rather than just non-decreasing? We already know that $f'(c)$, if it exists, is greater than or equal to zero, but we might hope to say that the derivative $f'(c)$ is strictly greater than zero. Unfortunately, that is not true.

In other words, a function could be increasing at the point c, in the sense that it is strictly increasing, but still have derivative 0. For instance, consider the function $f(x) := x^3$. This is increasing everywhere, but at the point zero, its derivative is zero.

How can a function be increasing at a point even though its derivative is zero? Well, what happens is that the derivative was positive before the point, is positive just after the point, and becomes zero just momentarily. Alternatively, if you think in terms of the derivative as a limit of difference quotients, all the difference quotients are positive, but the limit is still zero because they get smaller and smaller in magnitude as you come closer and closer to the point. Another way of thinking of this is that you reduce your car's speed to zero for the split second that you cross the STOP line, so as to comply with the letter of the law without actually stopping for any interval of time.

Similarly, we can define the notion of a function f being *decreasing* at a point c. This means that $f(c) < f(x)$ for x to the immediate left of c and $f(c) > f(x)$ for x to the immediate right of c. As in the previous case, we can deduce that $f'(c)$, if it exists, is less than or equal to zero, but it could very well happen that $f'(c) = 0$. An example is $f(x) := -x^3$, at the point $x = 0$.

1.7. **Increasing functions and sign of derivative.** Here's what we did. We first did separate analyses for what we can conclude about the left-hand derivative and the right-hand derivative of a function based on how the value of the function at the point compares with the value of the function at points to its immediate left. We used this to come to some conclusions about the nature of the derivative of a function (if it exists) at points of local maxima, local minima, and points where the function is nondecreasing and nonincreasing. Let's now discuss a converse result.

So far, we have used information about the nature of changes of the function to deduce information about the sign of the derivative. Now, we want to go the other way around: use information about the sign of the derivative to deduce information about the behavior of the function. And this is particularly useful because now that we have a huge toolkit, we can differentiate practically any function that we can write down. This means that even for functions that we have no idea how to visualize, we can formally differentiate them and work with the derivative. Thus, if we can relate information about the derivative to information about the function, we are in good shape.

Remember what we said: if a function is increasing, it is nondecreasing, and if it is nondecreasing, then the derivative is greater than or equal to zero. Now, a converse for this would mean some condition on the derivative telling us whether the function is increasing.

Unfortunately, the derivative being zero is very inconclusive. The function could be constant, it could be a local maximum, it could be a local minimum, it could be increasing, or it could be decreasing. However, it turns out that if the derivative is *strictly* positive, then we can conclude that the function is increasing.

Specifically, we have the following chain of implications for a function f defined around a point c and differentiable at c:

$f'(c) > 0 \implies f$ is increasing at $c \implies f$ is nondecreasing at $c \implies f'(c) \geq 0$

And each of these implications is strict, in the sense that you cannot proceed backwards with any of them, becausethere are counterexamples to each possible reverse implication.

Similarly, for a function f defined around a point c and differentiable at c:

$f'(c) < 0 \implies f$ is decreasing at $c \implies f$ is nonincreasing at $c \implies f'(c) \leq 0$

1.8. **Increasing and decreasing functions.** A function f is said to be increasing an an interval I (which may be open, closed, half-open, half-closed, or stretching to infinity) if for any $x_1 < x_2$, with both x_1 and x_2 in I, we have $f(x_1) < f(x_2)$. In other words, the larger the input, the larger the output.

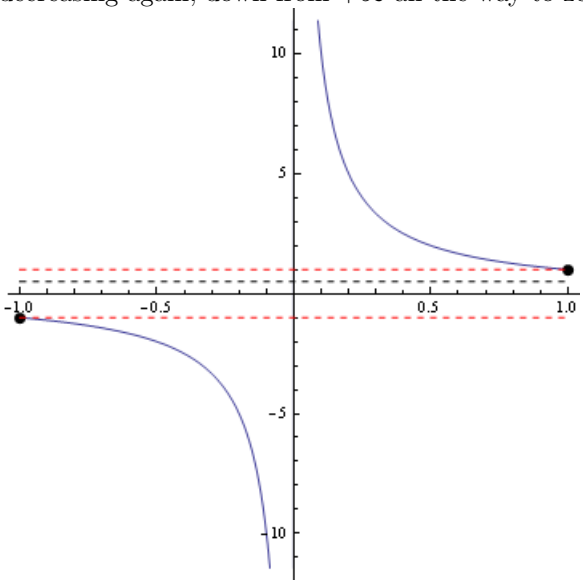
A little while ago, we talked of the notion of a function that is increasing at a point, and that was basically something similar, except that there one of the comparison points was fixed and the other one was restricted to somewhere close by. For a function to be increasing on an interval means that it is increasing at every point in the interior of the interval. If the interval has endpoints, then the function attains a strict local minimum at the left endpoint and a strict local maximum at the right endpoint.

Similarly, we say that f is *decreasing* on an interval I if, for any $x_1, x_2 \in I$, with $x_1 < x_2$, we have $f(x_1) > f(x_2)$. In other words, the larger the input, the smaller the output.

When I do not specify the interval and simply say that a function is increasing (respectively, decreasing), I mean that the function is increasing (respectively, decreasing) over its entire domain. For functions whose domain is the set of all real numbers, this means that the function is increasing (respectively, decreasing) over the set of all real numbers.

An example of an increasing function is a function $f(x) := ax + b$ with $a > 0$. An example of a decreasing function is a function $f(x) := ax + b$ with $a < 0$.

By the way, here's an interesting and weird example. Consider the function $f(x) := 1/x$. This function is not defined at $x = 0$. So, its domain is a union of two disjoint open intervals: the interval $(-\infty, 0)$ and the interval $(0, \infty)$. Now, we see that on each of these intervals, the function is decreasing. In fact, on the interval $(-\infty, 0)$, the function starts out from something close to 0 and then becomes more and more negative, approaching $-\infty$ as x tends to zero from the left. And then, on the interval $(0, \infty)$, the function is decreasing again, down from $+\infty$ all the way to zero.



But, taken together, is the function decreasing? No, and the reason is that at the point 0, where the function is undefined, it is undergoing this *huge* shift – from $-\infty$ to ∞. This fact – that points where the function is undefined can be points where it jumps from $-\infty$ to $+\infty$ or $+\infty$ to $-\infty$ – is a fact that keeps coming up. If you remember, this same fact haunted us when we were trying to apply the intermediate-value theorem to the function $1/x$ on an interval containing 0.

1.9. **The derivative sign condition for increasing/decreasing.** We first state the result for open intervals, where it is fairly straightforward. Suppose f is a function defined on an open interval (a, b). Suppose, further, that f is continuous and differentiable on (a, b), and for every point $x \in (a, b)$, $f'(x) > 0$. Then, f is an increasing function on (a, b).

A similar statement for decreasing: If f is a function defined on an interval (a, b). Suppose, further, that f is continuous and differentiable on (a, b), and for every point $x \in (a, b)$, $f'(x) < 0$. Then, f is a decreasing function on (a, b).

The result also holds for open intervals that stretch to ∞ or $-\infty$.

Note that it is important that f should be defined for all values in the interval (a, b), that it should be continuous on the interval, and that it should be differentiable on the interval. Here are some counterexamples:

(1) Consider the function $f(x) := 1/x$, defined and differentiable for $x \neq 0$. Its derivative is $f'(x) := -1/x^2$, which is negative wherever defined. Hence, f is decreasing on any open interval not containing 0. However, it is *not* decreasing on any open interval containing 0.

(2) Consider the function $f(x) := \tan x$. The derivative of the function is $f'(x) := \sec^2 x$. Note that f is defined for all x that are not odd multiples of $\pi/2$, and the same holds for f'. Also, note that $f'(x) > 0$ wherever defined, because $|\sec x| \geq 1$ wherever defined. Thus, the tan function is increasing on any interval not containing an odd multiple of $\pi/2$. But at each odd multiple of $\pi/2$, it slips from $+\infty$ to $-\infty$.

Let us now look at the version for a closed interval.

Suppose f is a function defined on a closed interval $[a,b]$, which is continuous on $[a,b]$ and differentiable on (a,b). Then, if $f'(x) > 0$ for $x \in (a,b)$, then f is increasing on all of $[a,b]$. Similarly, if $f'(x) < 0$ for $x \in (a,b)$, then f is decreasing on all of $[a,b]$.

In other words, we do *not* need to impose conditions on one-sided derivatives at the endpoints in order to guarantee that the function is increasing on the entire closed interval.

Finally, if $f'(x) = 0$ on the interval (a,b), then f is constant on $[a,b]$.

Some other versions:

(1) The result also applies to half-closed, half-open intervals. So, it may happen that a function f is continuous on $[a,b)$, differentiable on (a,b), and $f'(x) > 0$ for $x \in (a,b)$. In this case, f is increasing on $[a,b)$.
(2) The result also applies to intervals that stretch to infinity in either or both directions.

1.10. **Finding where a function is increasing and decreasing.** Let's consider a function f that, for simplicity, is continuously differentiable on its domain. So, f' is a continuous function. We now note that, in order to find out where f is increasing and decreasing, we need to find out where f' is positive, negative and zero.

Here's an example, Consider the function $f(x) := x^3 - 3x^2 - 9x + 7$. Where is f increasing and where is it decreasing? In order to find out, we need to differentiate f. The function $f'(x)$ is equal to $3x^2 - 6x - 9 = 3(x-3)(x+1)$. By the usual methods, we know that f' is positive on $(-\infty,-1) \cup (3,\infty)$, negative on $(-1,3)$, and zero at -1 and 3. Thus, the function f is increasing on the intervals $(-\infty,-1]$ and $[3,\infty)$ and decreasing on the interval $[-1,3]$.

Note that it is *not* correct to conclude from the above that f is increasing on the set $(-\infty,-1] \cup [3,\infty)$, although it is increasing on each of the intervals $(-\infty,-1]$ and $[3,\infty)$ *separately*. This is because the two pieces $(-\infty,-1]$ and $[3,\infty)$ are in some sense independent of each other. In general, the positive derivative implies increasing conclusions hold on intervals because they are what mathematicians call *connected sets*, and not for disjoint unions of intervals. In the case of this specific function, we note that $f(-1) = 12$ and $f(3) = -20$, so the value of the function at the point 3 is smaller than it is at -1. Thus, it is not correct to think of the function as being increasing on the union of the two intervals.

Similarly, if f is a rational function $x^2/(x^3-1)$, then we get $f'(x) = (-2x - x^4)/(x^3-1)^2$. Now, in order to find out where this is positive and where this is negative, we need to factor the numerator and the denominator. The factorization is:

$$\frac{-x(x+2^{1/3})(x^2 - 2^{1/3}x + 2^{2/3})}{(x-1)^2(x^2+x+1)^2}$$

The zeros of the numerator are 0 and $-2^{1/3}$ and the zero of the denominator is 1. The quadratic factors in both the numerator and the denominator are always positive. Also note that there is a minus sign on the outside.

Hence, f' is negative on $(1,\infty)$, $(0,1)$, and $(-\infty,-2^{1/3})$, positive on $(-2^{1/3},0)$, zero on 0 and $-2^{1/3}$, and undefined at 1. Thus, f is decreasing on $[0,1)$, $(1,\infty)$, and $(-\infty,-2^{1/3}]$, increasing on $[-2^{1/3},0]$.

Now, let's combine this with the information we have about f itself. Note that f is undefined at 1, it is positive on $(1,\infty)$, it is zero at 0, and it is negative on $(-\infty,0) \cup (0,1)$. How do we combine this with information about what's happening with the derivative?

On the interval $(-\infty,-2^{1/3})$, f is negative *and* decreasing. What's happening as $x \to -\infty$? f tends to zero (we'll see why a little later). So, as x goes from $-\infty$ to $-2^{1/3}$, f goes down from 0 to $-2^{2/3}/3$. Then, as x goes from $-2^{1/3}$ to 0, f is still negative but starts going up from $-2^{2/3}/3$ and reaches 0. In the interval from 0 to 1, f goes back in the negative direction, from 0 down to $-\infty$. Then, in the interval $(1,\infty)$, f goes emerges from $+\infty$ and goes down to 0 as $x \to +\infty$.

So we see that information about the sign of the derivative helps us get a better picture of how the function behaves, and allows us to better draw the graph of the function – something that we will try to do more of a short while from now.

Point-value distinction. We use the term *point of local maximum* or *point of local minimum* (or simply *local maximum* or *local minimum*) for the point in the domain, and the term *local maximum value* for the value of the function at the point.

2. Strategies for local and absolute maxima/minima

2.1. **Local extreme values and critical points.** If f is a function and c is a point in the interior of the domain of f, then f is said to have a *local maximum* at c if $f(x) \leq f(c)$ for all x sufficiently close to c. Here, *sufficiently close* means that there exists $a < c$ and $b > c$ such that the statement holds for all $x \in (a, b)$.

Similarly, we have the concept of *local minimum* at c.

The points in the domain at which local maxima and local minima occur are termed the *points of local extrema* and the values of the function at these points are termed the *local extreme values*.

As we discussed last time, if f is differentiable at a point c of local maximum or local minimum, the derivative of f at c is zero. This suggests that we define a notion.

An interior point c in the domain of a function f is termed a *critical point* if either $f'(c) = 0$ or $f'(c)$ does not exist. Thus, all the local extreme values occur at critical points – because at a local maximum or minimum, either the derivative does not exist, or the derivative equals zero.

Note that not all critical points are points of local maxima and minima. For instance, for the function $f(x) := x^3$, the point $x = 0$ is a critical point, but the function does not attain a local maximum or local minimum at that point. However, critical points give us a small set of points that we need to check against. Once we have this small set, we can use other methods to determine what precisely is happening at these points.

2.2. **First-derivative test.** The first-derivative test basically tries to determine whether something is a local maximum by looking, not just at the value of the derivative *at* the point, but also the value of the derivative *close* to the point.

Basically, we want to combine the idea of *increasing on the left, decreasing on the right* to show that something is a local maximum, and similarly, we combine the idea of *decreasing on the left, increasing on the right* to show that something is a local minimum.

The first-derivative test says that if c is a critical point for f and f is continuous at c (Note that f need not be differentiable at c). if there is a positive number δ such that:

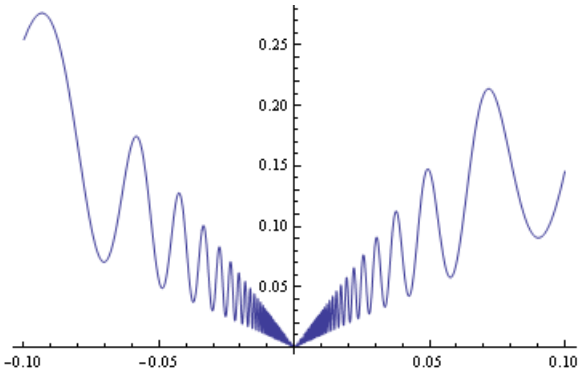
(1) $f'(x) > 0$ for all $x \in (c - \delta, c)$ and $f'(x) < 0$ for all $x \in (c, c + \delta)$, then $f(c)$ is a local maximum, i.e., c is a point of local maximum for f.
(2) $f'(x) < 0$ for all $x \in (c - \delta, c)$ and $f'(x) > 0$ for all $x \in (c, c + \delta)$, then $f(c)$ is a local minimum, i.e., c is a point of local minimum for f.
(3) $f'(x)$ keeps constant sign on $(c - \delta, c) \cup (c, c + \delta)$, then c is not a point of local maximum/minimum for f.

Thus, for the function $f(x) := x^2/(x^3 - 1)$, there is a local minimum at $-2^{1/3}$ and a local maximum at 0.

Recall that for the function $f(x) := x^3$, the derivative at zero is zero, so it is a critical point but it is not a point of local extremum, because the derivative is positive everywhere else.

2.3. **What are we essentially doing with the first-derivative test?** Why does the first-derivative test work? Essentially it is an application of the results on increasing and decreasing functions for closed intervals. What we're doing is using the information about the derivative from the left to conclude that the point is a strict local maximum from the left, because the function is increasing up to the point, and is a strict local maximum from the right, because the function is decreasing down from the point.

2.4. **The first-derivative test is sufficient but not necessary.** For most of the function that you'll see, the first-derivative test will give you a good way of figuring out whether a given critical point is a local maximum or local minimum. There are, however, situations where the first-derivative test fails to work. These are situations where the derivative changes sign infinitely often, close to the critical point, so does not have a constant sign near the critical point. For instance, consider the function $f(x) := |x|(2 + \sin(1/x))$. This attains a local minimum at the point $x = 0$, which is a critical point. However, the derivative of the function oscillates between the positive and negative sign close to zero and doesn't settle into a single sign on either side of zero.



2.5. **Second-derivative test.** One problem with the first-derivative test is that it requires us to make two local sign computations over *intervals*, rather than *at points*. Discussed here is a variant of the first-derivative test, called the second-derivative test, that is sometimes easier to use.

Suppose c is a critical point in the interior of the domain of a function f, and f is twice differentiable at c. Then, if $f''(c) > 0$, c is a point of local minimum, whereas if $f''(c) < 0$, then c is a point of local minimum.

The way this works is as follows: if $f''(c) > 0$, that means that f' is (strictly) increasing at c. This means that f' is negative to the immediate left of c and is positive to the immediate right of c. Thus, f attains a local minimum at c.

Note that the second-derivative test works for critical points where the function is twice-differentiable. In particular, it does not work for the kind of sharp peak points where the function suddenly changes direction. On the other hand, since the second-derivative test involves evaluation of the second derivative at only one point, it may be easier to apply in certain situations than the first-derivative test, which requires reasoning about the sign of a function over an interval.

2.6. **Endpoint maxima and minima.** An *endpoint maximum* is something like a local maximum, except that it occurs at the endpoint of the domain, so the value of the function at the point needs to be compared only with the values of the function at points sufficiently close to it on one side (the side that the domain is in). Similarly, an *endpoint minimum* is like a local minimum, except that it occurs at the endpoint of the domain, so the value of the function at the point needs to be compared only with the values of the function at points sufficiently close to it on one side.

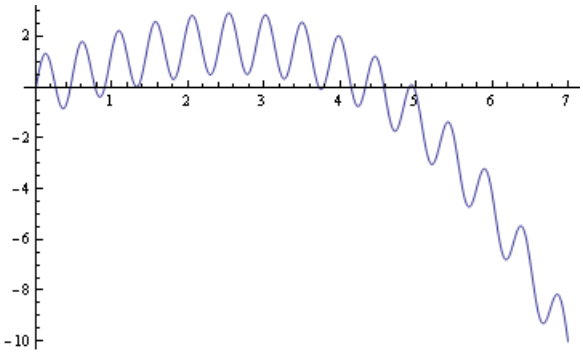
If the endpoint is a left endpoint, then being an endpoint maximum (respectively, minimum) means being a local maximum (respectively, minimum) from the right. If the endpoint is a right endpoint, then being an endpoint maximum (respectively, minimum) means being a local maximum (respectively, minimum) from the left.

2.7. **Absolute maxima and minima.** We say that a function f has an absolute maximum at a point d in the domain if $f(d) \geq f(x)$ for all x in the domain. We say that f has an absolute minimum at a point d in the domain if $f(d) \leq f(x)$ for all x in the domain. The corresponding value $f(d)$ is termed the absolute maximum (respectively, minimum) of f on its domain.

Notice the following very important fact about absolute maxima and minima, which distinguishes them from local maxima and minima. If an absolute maximum value exists, then the value is unique, even though it may be attained at multiple points on the domain. Similarly, if an absolute minimum value exists, then the value is unique, even though it may be attained at multiple points of the domain. Further, assuming the function to be continuous through the domain, and assuming the domain to be connected (i.e., not fragmented into intervals) the range of the function is the interval between the absolute minimum value and the absolute maximum value. This follows from the intermediate value theorem.

For instance, for the cos function, absolute maxima occur at multiples of 2π and absolute minima occur at odd multiples of π. The absolute maximum value is 1 and the absolute minimum value is -1.

Just for fun, here's a picture of a function having lots of local maxima and minima, but all at different levels. Note that some of the local maximum values are less than some of the local minimum values. This highlights the extremely local nature of local maxima/minima.



2.8. **Where and when do absolute maxima/minima exist?** Recall the *extreme value theorem* from some time ago. It said that for a continuous function on a closed interval, the function attains its maximum and minimum. This was basically asserting the existence of absolute maxima and minima for a continuous function on a closed interval.

Notice that any point of absolute maximum (respectively, minimum) is either an endpoint or is a point of local maximum (respectively, minimum). We further know that any point of local maximum or minimum is a critical point. Thus, in order to find all the absolute maxima and minima, a good first step is to find critical points and endpoints.

Another thing needs to be noted. For some funny functions, it turns out that there is no maximum or minimum. This could happen for two reasons: first, the function approaches $+\infty$ or $-\infty$, i.e., it gets arbitrarily large in one direction, somewhere. Second, it might happen that the function approaches some maximum value but does not attain it on the domain. For instance, the function $f(x) = x$ on the interval $(0, 1)$ does not attain a maximum or minimum, since these occur at the endpoints, which by design are not included in the domain.

Thus, the absolute maxima and minima, *if they occur*, occur at critical points and endpoints. But we need to further tackle the question of existence. In order to deal with this issue clearly, we need to face up to something we have avoided so far: limits to infinity.

3. Moving to multiple variables

3.1. **Directions and neighborhoods.** The first key difference between functions of one variable and functions of more than one variable is that in the latter case, there are a lot more directions. For a function of 2 variables, there are four directions parallel to the axes: a left and a right approach in each coordinate keeping the other fixed. But there are many other straight line approaches along other directions, which are not parallel to either axes. There could also be curved directions of approach.

Thus, although it is possible to talk of directional maxima and minima (and we'll do this in a moment) we need another approach to defining local maxima and minima. The key idea is that of *neighborhood*, which works by simultaneously covering all directions. We could use as neighborhoods either circular disks or rectangular regions.

3.2. **The technical definition of local maximum and local minimum.** Consider a function f of two variables with domain D. We say that a point (a, b) in the interior of D is a point of *local minimum* for f if there exists an open disk U around (a, b) contained in D such that $f(a, b) \leq f(x, y)$ for all $(x, y) \in U$. Equivalently, there exists a positive number r such that $f(a, b) \leq f(x, y)$ whenever the distance between the points (x, y) and (a, b) is less than r.

Note that with the open disk formulation, we typically use circular disks. However, we could also use square disks centered at (a, b) – these are sometimes easier to work with.

Analogously, we can define *strict local minimum*, *local maximum*, and *strict local maximum*.

3.3. **The technical definition of absolute maximum and absolute minimum.** A point in the domain of a function is termed a point of *absolute maximum* if the value of the function at the point is greater than or equal to the value of the function at all other points in the domain of the function. Similarly, we talk of a point of *absolute minimum* for a function.

3.4. **Directional maximum and directional derivative.** Suppose f is a function of two variables and (a,b) is a point in the domain. For a particular direction of approach to the point (a,b) , e.g., a half-line (or curve) ending at (a,b), we can ask whether (a,b) is a *directional local maximum* or *directional local minimum* from that direction.

We say that (a,b) is a point of directional local maximum from that direction if $f(a,b)$ is greater than or equal to the f-value for all points in the half-line sufficiently close to (a,b). Similarly, we define directional local minimum. We can also define strict directional local maximum and strict directional local minimum.

For functions of one variable, there were only two directions worth pondering: left and right. Ergo, the idea of local maximum/minimum from the left and the right.

In the case of functions of one variable, we related being a local maximum/minimum from the left with the sign of the left-hand derivative (if it exists) and being a local maximum/minimum from the right with the sign of the right-hand derivative (if it exists). We can do the same thing now, and we obtain that:

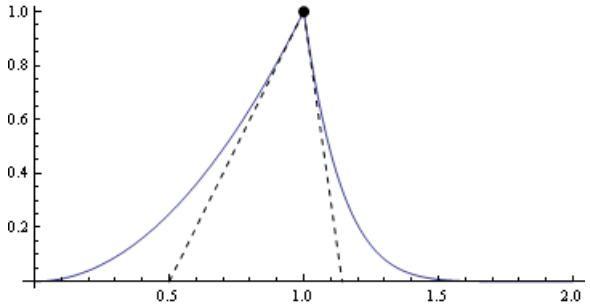
> For a directional local minimum, the directional derivative (in the outward direction from the point) is greater than or equal to zero. For a directional local maximum, the directional derivative (in the outward direction from the point) is less than or equal to zero.

Further, it remains true that *even for strict directional local maxima/minima*, the directional derivative *may well be zero*, i.e., we cannot rule out the possibility of the directional derivative being zero. We saw that for one variable, this proved crucial, because it meant that being a local maximum both from left and right forced the derivative (if it exists) to *equal zero*. The same idea works here:

> If a point is a point of directional local minimum from two opposite directions (i.e., it is a local minimum along a line through the point, from both directions on the line) then the directional derivative along the line, if it exists, must equal zero.

The reasoning is the same as for one variable: being a local minimum from one direction means the directional derivative in that direction is positive or zero. Being a local minimum from the opposite direction means the directional derivative in the opposite direction is positive or zero. *We know that directional derivatives in opposite directions are negatives of each other*, so overall we conclude that both directional derivatives must be zero.

On the other hand, it is possible to have a two variable situation analogous to the function of one variable that has a local maximum with a strictly positive left hand derivative and strictly negative right hand derivative:



For functions of two variables, this would correspond to the graph (which is a surface) not having a well defined tangent plane at the point of local maximum or minimum.

3.5. **Relating directional with the all-directions ideas.** A point of local minimum (respectively strict local minimum, local maximum, strict local maximum) for a function is also a point of local minimum (respectively strict local minimum, local maximum, strict local maximum) if we restrict attention to behavior on any curve/line containing the point. In particular, it is a directional local minimum (respectively, directional local maximum, strict directional local minimum, strict directional local maximum) for all directions.

Thus, we obtain that:

> If a function of two variables is differentiable at a point of local minimum or local maximum, then the directional derivative of the function is zero at the point in every direction. Equivalently, the gradient vector of the function at the point is the zero vector. Equivalently, both the first partial derivatives at the point are zero.

In symbols, if $f(x,y)$ is a function of two variables, and (a,b) is a point of local minimum for f, then, if f is differentiable at (a,b), we have $f_x(a,b) = f_y(a,b) = 0$. *Note that for a differentiable function, both partial derivatives being equal to zero implies that all directional derivatives are equal to zero.*

If we graph the surface $z = f(x,y)$, then pictorially, this means that the tangent plane at a point of local minimum, if it exists, is parallel to the xy-plane, and forms a kind of flat floor for the curve around the point. For a point of local maximum, the tangent plane exists, and forms a kind of flat ceiling around the point.

Points where the gradient vector is zero, i.e., the directional derivative is zero in all directions (which for continuous first partials just means that both partial derivatives are zero) are termed *critical points*. In particular, this means that all candidates for local extreme values occur at critical points.

3.6. **Local maximum, minimum, or neither?** For a function of one variable, recall that the derivative being zero could imply one of many things:

- Point of local maximum, such as x^2 at 0.
- Point of local minimum, such as $-x^2$ at 0.
- Point of increase/constancy – local maximum from left, local minimum from right. The typical examples are points of inflection as with x^3 at 0.
- Point of decrease/constancy – local minimum from left, local maximum from right, such as $-x^3$ at 0.
- None of the above: this includes oscillatory type examples, such as $x^2 \sin(1/x)$, $x \neq 0$ and defined to be 0 at 0. This is oscillatory aroun the point from both sides, though the derivative is still zero.

The same holds with functions of more than one variable – we could very well have the "point of inflection" type situation. However, things become more interesting in two variables, both easier and more complicated. Namely, we have not two, but many, directions, and more importantly, all the directions are connected to each other. By this I mean that I can smoothly rotate from one direction to any other direction, a feat that is impossible in one variable because the left and the right are discrete and opposite directions with no way of bridging across them.[1]

So, suppose we have a point (a,b) in the domain of f such that $f_x(a,b) = 0$ and $f_y(a,b) = 0$. This means that all the directional derivatives are zero.

Now, suppose f is a directional local minimum from some specific direction, e.g., from the positive x-direction. If we now smoothly rotate the direction, either f remains a directional local minimum as we keep rotating, or at some stage, it flips over into not being a directional local minimum at the point, perhaps being a directional local maximum. But *at the direction where the transition occurs, something funny is going on*. For instance, it may be that at the direction where the transition occurs, the function is actually *constant* along the direction. At any rate, the transition from being a local maximum to a local minimum is an interesting and nontrivial phenomenon.

This suggests that the right analogue of point of inflection with increase through – i.e., the right analogue of the type of thinking behind x^3 – is something where there are directions where there is a transition from directional maximum to directional minimum. It will be tricky to go into this in too much detail, but I'll note that the term for a point at which this kind of phenomenon occurs is *saddle point*.

3.7. **In search of a second derivative test.** Obviously, if $f(x,y)$ has a local minimum at (a,b), and f has second derivatives, then it should have a local minimum at (a,b) purely viewed as a function of x (keeping y fixed at b) and also purely viewed as a function of y (keeping x fixed at a). This suggests a second derivative test: we would like to see $f_{xx}(a,b) > 0$ (second derivative test on x) and $f_{yy}(a,b) > 0$ (second derivative test on y). Unfortunately, just having the conditions of both pure second partials greater than 0 at the point is not good enough. There are many other directions.

Luckily, there is an easy fix, and that involves looking at the *mixed partials*. Unfortunately, the logic of this explanation is beyond the scope of the course, so you just need to take it on faith.

Define $D(a,b)$ as the quantity:

$$D(a,b) := f_{xx}(a,b)f_{yy}(a,b) - [f_{xy}(a,b)]^2$$

[1]This is basically the fact that a plane minus a point is still connected, whereas a line minus a point is disconnected.

Then, the second derivative test says the following for a point (a, b) such that f is *twice continuously differentiable* around (a, b) and $f_x(a, b) = f_y(a, b) = 0$:

- If $D(a, b) > 0$ and $f_{xx}(a, b) > 0$ then f attains a local minimum at (a, b). Note that the conditions $D > 0$ and $f_{xx}(a, b) > 0$ together not only imply that $f_{yy}(a, b) > 0$, they also imply that the second partial derivative along *any* direction is positive.
- If $D(a, b) > 0$ and $f_{xx}(a, b) < 0$ then f attains a local maximum at (a, b). Note that the conditions $D > 0$ and $f_{xx}(a, b) < 0$ together not only imply that $f_{yy}(a, b) < 0$, they also imply that the second partial derivative along *any* direction is negative.
- If $D(a, b) < 0$, then f does not attain a local maximum or a local minimum at (a, b). In this case, the point $(a, b, f(a, b))$ is a saddle point of the graph of f: the tangent plane through this point cuts through the surface that's the graph of this function.

  Note that this case definitely occurs if $f_{xx}(a, b)$ and $f_{yy}(a, b)$ have opposite signs, which makes sense because that would mean that the function is a local minimum along one axis and a local maximum along the other. But it could occur even if both of them have the same sign, if the magnitude of $(f_{xy}(a, b))^2$ is bigger than the product of their magnitudes. What that basically means is that there are diagonal directions along which the function's behavior is opposite to what it is along the x and y directions.
- If $D(a, b) = 0$, the second derivative test is inconclusive.

The value $D(a, b)$ can be thought of as the determinant of this 2×2 matrix:

$$\begin{pmatrix} f_{xx}(a, b) & f_{xy}(a, b) \\ f_{yx}(a, b) & f_{yy}(a, b) \end{pmatrix}$$

Note that the two off-diagonal entries $f_{xy}(a, b)$ and $f_{yx}(a, b)$ are equal by Clairaut's theorem and the assumption of continuity of partials of f around (a, b). If you ever see multivariable calculus in its proper form, you will learn that this matrix is called a *Hessian* and the conditions of the second derivative test are basically conditions to ensure that the Hessian is a positive definite (respectively, negative definite) matrix.

3.8. **Putting things together: the technique for a function of two variables.** Here now is the overall procedure for finding local and absolute maxima/minima for a twice continuously differentiable function of two variables:

- First, find all the *critical points*. A critical point is a point in the domain at which all directional derivatives are zero, or equivalently, the two first partials are zero. In symbols, (a, b) is a critical point if $f_x(a, b) = 0$ and $f_y(a, b) = 0$.
- Next, for each critical point, use the second derivative test, if feasible, to find out whether it is a point of local maximum, point of local minimum, or a saddle point. If the second derivative test is inconclusive, see if there are other ways of figuring things out.

3.9. **Absolute maxima and minima: boundary issues.** Recall that for functions of one variable, in addition to finding the local maxima/minima, we also needed to consider the endpoint maxima/minima and also the limits to boundary points not in the domain (and to infinity). A similar kind of complication arises for functions of two variables. For simplicity, we restrict attention to the following case: a function f with domain D a closed, bounded subset of $\mathbb{R}^2$ and with the property that f is twice continuously differentiable on the interior of D and its restriction to the boundary of D is "differentiable" under some smooth parameterization of the boundary.

Explanation of terminology:

- The boundary of a subset D of $\mathbb{R}^n$ is the set of points with the property that any open ball centered at the ball intersects D but is not completely contained in D. In other words, it's the points that are in close contact with D and with the complement of D.
- A closed subset of $\mathbb{R}^n$ is a subset that contains all its boundary points. A bounded subset of $\mathbb{R}^n$ is a subset that can be enclosed inside an open unit disk (or equivalently, in a rectangular region). Closed bounded subsets in $\mathbb{R}^n$ are what's called *compact* which makes a lot of facts about them true. (You don't have to worry what this means).

- The interior of a subset D of $\mathbb{R}^n$ is the subset of D comprising those points not in the boundary, i.e., those points for which we can make open balls centered at them lying completely inside D.

In the case of functions of one variable, the subsets live in $\mathbb{R}^1 = \mathbb{R}$ and are one-dimensional, and their boundaries are typically zero-dimensional, i.e., usually, sets of isolated points. In the case of functions of two variables, the subsets live in $\mathbb{R}^2$ and are two-dimensional, and the boundaries are typically one-dimensional, i.e., unions of lines and curves. In general, for n-dimensional subsets of $\mathbb{R}^n$, the boundary sets are expected to be $(n-1)$-dimensional.

The first result of relevance is the *extreme value theorem*. It states that for a closed, bounded subset D of $\mathbb{R}^2$ and a continuous function f on D, f attains its absolute maximum and minimum values at points in D.

The procedure for finding the absolute extreme values, if f is twice continuously differentiable on the interior of D and differentiable under a smooth parameterization of the boundary of D, is as follows:

(1) Find the critical points of f in the interior of D, and hence all the candidates for local extreme values in there.
(2) Find the extreme values of f on the boundary of D.
(3) Compare all these values and use these to find the absolute extremes.

3.10. **Stretching off to infinity.** Things get more complicated when the domain of the function is not a closed bounded region but instead stretches off to infinity in one or more than one direction. In this case, we need to figure out the "limits to infinity" of the function, if any, in order to find the absolute maxima and minima.

Unfortunately, there is more to this than meets the eye, because there are many different directions in which one can go off to infinity, and the limit may be different in each direction.

So, instead of finding these limits, a more useful approach may be to foreclose options, i.e., figure out that, say, if we go off to infinity in any direction, the function value is going to become too large, and thus the absolute minimum will not be attained outside of a certain close dbounded interval.

3.11. **Understanding the relationship with level curves.** It's worth pondering the relationship between local extreme values and level curves.

Level curves denote curves along which the function takes constant values. Local extreme values are typically level curves that are *single point* level curves. For instance, for the function x^2+y^2, the level curves are the circles centered at the origin. The local (and absolute) minimum is the unique single point level curve, which occurs at the origin.

In a subsequent lecture discussion, we will look in detail at various classes of examples of maximum and minimum value computations for functions of two variables.

LAGRANGE MULTIPLIERS

MATH 195, SECTION 59 (VIPUL NAIK)

Corresponding material in the book: Section 14.8

What students should definitely get: The Lagrange multiplier condition (one constraint, two constraints and in principle more than two constraints), the application to finding absolute extreme values.

What students should hopefully get: Situations where Lagrange multipliers fail, the underlying logic behind Lagrange multipliers, how to use Lagrange multipliers for piecewise smooth situations.

Executive summary

Words ...

(1) Two of the reasons why the derivative of a function may be zero: the function is constant around the point, or the function has a local extreme value at the point.

Version for many variables: two of the reasons why the gradient vector of a function of many variables may be zero: the function is constant around the point, or the function has a local extreme value at the point.

Version for function restricted to a subset smooth around a point: two of the reasons why the gradient vector may be *orthogonal* to the subset at the point: the function is constant on the subset around the point, or the function has a local extreme value (relative to the subset) at the point.

(2) For a function f defined on a subset smoooth around a point (i.e., with a well defined tangent and normal space), if f has a local extreme value at the point when restricted to the subset, then ∇f lives in the normal direction to the subset (this includes the possibility of it being zero).

(3) For a codimension one subset of $\mathbb{R}^n$ defined by a condition $g(x_1, x_2, \ldots, x_n) = k$, if a point $(a_1, a_2, \ldots, a_n)$ gives a local extreme value for a function f of n variables, and if ∇g is well defined and nonzero at the point, then there exists a real number λ such that $\nabla f(a_1, a_2, \ldots, a_n) = \lambda \nabla g(a_1, a_2, \ldots, a_n)$. Note that λ may be zero.

(4) Suppose a codimension r subset of $\mathbb{R}^n$ is given by r independent constraints $g_1(x_1, x_2, \ldots, x_n) = k_1$, $g_2(x_1, x_2, \ldots, x_n) = k_2$, and so on till $g_r(x_1, x_2, \ldots, x_n) = k_r$. Suppose ∇g_i are nonzero for all i at a point $(a_1, a_2, \ldots, a_n)$ of local extreme value for a function f relative to this subset. Suppose further that all the ∇g_i are linearly independent. Then $\nabla f(a_1, a_2, \ldots, a_n)$ is a linear combination of the vectors $\nabla g_1(a_1, a_2, \ldots, a_n)$, $\nabla g_2(a_1, a_2, \ldots, a_n)$, $\ldots$, $\nabla g_r(a_1, a_2, \ldots, a_n)$. In other words, there exist real numbers λ_1, λ_2, $\ldots$, λ_r such that:

$$\nabla f(a_1, a_2, \ldots, a_n) = \lambda_1 \nabla g_1(a_1, a_2, \ldots, a_n) + \lambda_2 \nabla g_2(a_1, a_2, \ldots, a_n) + \cdots + \lambda_r \nabla g_r(a_1, a_2, \ldots, a_n)$$

(5) The Lagrange condition may be violated at points of local extremum where ∇g is zero, or more generally, where the ∇g_i fail to be linearly independent. This may occur either because the tangent and normal space are not well defined or because the functions fail to capture it well.

Actions ...

(1) Suppose we want to maximize and minimize f on the set $g(x_1, x_2, \ldots, x_n) = k$. Assume $\nabla g(x_1, x_2, \ldots, x_n)$ is defined everywhere on the set and never zero. Suppose ∇f is also defined. Then, all local maxima and local minima are attained at points where $\nabla f = \lambda \nabla g$ for some real number λ. To find these, we solve the system of $n + 1$ equations in the $n + 1$ variables x_1, x_2, $\ldots$, x_n, namely the n scalar equations from the Lagrange condition and the equation $g(x_1, x_2, \ldots, x_n) = k$.

To find the actual extreme values, once we've collected all candidate points from the above procedure, we evaluate the function at all these and find the largest and smallest value to find the absolute maximum and minimum.

(2) If there are points in the domain where ∇g takes the value 0, these may also be candidates for local extreme values, and the function should additionally be evaluated at these as well to find the absolute maximum and minimum.
(3) A similar procedure works for a subset given by r constraints. In this case, we have the equation:

$$\nabla f(a_1, a_2, \ldots, a_n) = \lambda_1 \nabla g_1(a_1, a_2, \ldots, a_n) + \lambda_2 \nabla g_2(a_1, a_2, \ldots, a_n) + \cdots + \lambda_r \nabla g_r(a_1, a_2, \ldots, a_n)$$

as well as the r equations $g_1(x_1, x_2, \ldots, x_n) = k_1$, $g_2(x_1, x_2, \ldots, x_n) = k_2$, and so on. In total, we have $n + r$ equations in $n + r$ variables: the x_1, x_2, ..., x_n and the λ_1, λ_2, ..., λ_r.

1. Lagrange multipliers: basic formulation with one constraint

1.1. **The two key ideas.** We summarize two key ideas behind Lagrange multipliers:

- If a scalar function is constant on a subset of $\mathbb{R}^n$, its directional derivative along any direction tangent to the subset at a point on the subset is zero. Thus, the gradient of the function at any point in the subset (if nonzero) is orthogonal to the subset.
- Consider a scalar function and a subset of $\mathbb{R}^n$. At any point in the subset where the function attains a local extreme value relative to the subset, the directional derivative along any direction tangent to the subset at the point is zero. Thus, the gradient of the function at the point (if nonzero) is orthogonal to the subset.

Roughly speaking, we're saying that there are two reasons (among many!) why the directional derivative along all tangents at a point should be zero: one, the function is constant around the point, and the other, the function attains a local extreme value at the point.

The key insight behind Lagrange multipliers is to combine these insights and ask: *on a subset defined by one function being constant, how do we find the local extreme values of another function*? The idea is to use the fact that both for the function that's constant and the function that is attaining an extreme value, the gradient is normal (orthogonal) to the subset. If the subset is given by a single constraint, then it has codimension one, so the normal space is one-dimensional, and this forces the gradient vectors for the two functions to be scalar multiples of each other (with suitable assumptions of being nonzero).

1.2. **Getting started.** We know that for a differentiable function f defined in an open domain in $\mathbb{R}^n$, if the function has a local extreme value at a point $(a_1, a_2, \ldots, a_n)$, then the directional derivative of f along *every* direction in $\mathbb{R}^n$ is zero. More specifically, if f has a local extreme value *along* a particular line (both directions on that line) then the directional derivative along that particular direction is zero.

Now, suppose we want to maximize f, not on an open domain in $\mathbb{R}^n$, but on a subset of smaller dimension that has a well defined tangent space at the point. Then, the key idea is that a *necessary* condition for a point to have a local extreme value is that the directional derivative along all tangent directions to that subset are zero. *However, the directional derivative along non-tangential directions may well be nonzero.*

In other words, the gradient of the function f does not have any component tangential to the subset. Thus, the gradient of f, ∇f, is *either zero* or is a vector *perpendicular* to the tangent space at the point, i.e., a vecor *in* the normal space at the point.

We now turn to some specific cases for $n = 2$ and $n = 3$.

1.3. **In two dimensions: setup.** Suppose we have a function f in two variables, and a smooth curve γ. We want to find the extreme values of f along the curve γ. Suppose f attains a local maximum at (x_0, y_0) in the curve γ. This just means that f has a local maximum *relative* to the curve γ, i.e., if we take points in the curve γ close to the point (x_0, y_0), then the f-value at those points in γ is less than or equal to the value $f(x_0, y_0)$.

This implies that if we move slightly *along* γ, or tangential to γ, then the directional derivative of f is zero. The reason: the directional derivative along one direction on γ is less than or equal to zero, because the function is smaller if we move a bit in that direction. Similarly, the directioal derivative along the reverse direction is less than or equal to zero. Since these directional derivatives are negatives of each other, this forces both of them to be zero.

Thus, the directional derivative of f along the tangent direction to γ is zero. This means that the dot product of the gradient of f and the unit tangent vector to γ is zero, so ∇f is either equal to zero or points in a direction perpendicular to the tangent direction to γ.

1.4. **γ as the level curve of g.** We now continue with the same setup as above, now setting up γ as a level curve of another differentiable function g, i.e., γ is defined as the set $g(x,y) = k$ for some constant k. We want to find the tangent and normal vectors at a point (x_0, y_0) to this curve.

Since g is not changing along γ, ∇g has a component of zero along γ *at every point of* γ. In particular, if ∇g is nonzero, it is along the normal direction to γ.

Thus, the upshot is that if (x_0, y_0) is a point of extreme value for f on the curve γ defined as $g(x,y) = k$, then $\nabla g(x_0, y_0)$, if nonzero, is normal to the curve at the point, and $\nabla f(x_0, y_0)$, if nonzero, is normal to the curve at the point. Thus, $\nabla f(x_0, y_0)$ is a scalar multiple of $\nabla g(x_0, y_0)$, i.e., there is a constant λ such that:

$$\nabla f(x_0, y_0) = \lambda \nabla g(x_0, y_0)$$

This constant λ is termed a *Lagrange multiplier*.

1.5. **Case of $n = 3$.** We quickly state the similar result for 3 variables. We want to find extreme values for a function $f(x,y,z)$ on the *surface* (codimension one subset) $g(x,y,z) = k$. Then, if ∇g is nonzero, it is in the normal direction to the surface, and thus, at a point (x_0, y_0, z_0) where f has a local extreme, ∇f is a scalar multiple of ∇g, i.e., we have:

$$\nabla f(x_0, y_0, z_0) = \lambda \nabla g(x_0, y_0, z_0)$$

1.6. **General statement.** Suppose we have two functions f and g, both of n variables. In other words, $f(x_1, x_2, \ldots, x_n)$ is a function of n variables and $g(x_1, x_2, \ldots, x_n)$ is also a function of n variables. Suppose further that k is a real number, and suppose that ∇g is nonzero everywhere on the codimension one subset $g(x_1, x_2, \ldots, x_n) = k$. Suppose $(a_1, a_2, \ldots, a_n)$ is a point satisfying $g(a_1, a_2, \ldots, a_n) = k$, and such that f has a local extreme value at $(a_1, a_2, \ldots, a_n)$ *when restricted to the subset* $g(x_1, x_2, \ldots, x_n) = k$. Then, there exists a scalar λ such that:

$$(\nabla f)(a_1, a_2, \ldots, a_n) = \lambda \nabla g(a_1, a_2, \ldots, a_n)$$

In other words, the directional derivative of f is a scalar multiple of the directional derivative of g at the point. Another way of thinking of this is that the directional derivative of f has no component tangential to the subset $g(x_1, x_2, \ldots, x_n) = k$.

Note that if f has a local extreme value at the point $(a_1, a_2, \ldots, a_n)$ with respect to the whole space (and not just the codimension one subset $g(x_1, x_2, \ldots, x_n) = k$) then in fact $\nabla f(a_1, a_2, \ldots, a_n)$ is the zero vector, so $\lambda = 0$ in this case. This is a much stronger condition.

2. Lagrange multipliers: multiple constraints

The Lagrange multiplier applications we have seen so far concentrate on codimension one subsets, i.e., subsets that are given as solutions to $g(x,y) = k$ for a single constraint g. However, the ideas generalize a little further to multiple constraints. The key difference is that the normal space is more than one-dimensional.

2.1. **Arithmetic of dimension and codimension.** Here's a quick recall of the arithmetic of dimension and codimension. If a subset of $\mathbb{R}^n$ is specified in a top-down fashion by r independent scalar equality constraints, then the subset has *codimension* r and *dimension* $n - r$. The way to think of this is that we start with the whole n-dimensional space and each new constraint reduces the dimension by 1, provided it is independent of all the previous constraints.

Let's recall what this means for tangent and normal spaces. If the subset is sufficiently smooth around a point in the subset, then we can define the tangent space to the subset about the point. The tangent space is a flat (linear) space at the point, and it has the same dimension as the subset, which in our case is $n - r$. The *normal space* is a space of dimension r orthogonal to the tangent space at the point, i.e., every vector in the normal space is orthogonal to every vector in the tangent space.

We now turn to a new aspect: *actually describing the normal space*. We know that for a function given by $g(x_1, x_2, \ldots, x_n) = k$, then the normal vector (unique up to scaling) at a point $(a_1, a_2, \ldots, a_n)$ is given by $\nabla g(a_1, a_2, \ldots, a_n)$, if that vector is nonzero.

Suppose we consider the subset of $\mathbb{R}^n$ satisfying this collection of r constraints, where all the functions g_i are differentiable functions:

$$\begin{aligned} g_1(x_1, x_2, \ldots, x_n) &= k_1 \\ g_2(x_1, x_2, \ldots, x_n) &= k_2 \\ \cdot &= \cdot \\ g_r(x_1, x_2, \ldots, x_n) &= k_r \end{aligned}$$

Then, at a point $(a_1, a_2, \ldots, a_n)$ in the subset, each of the gradient vectors $\nabla g_1(a_1, a_2, \ldots, a_n)$, $\nabla g_2(a_1, a_2, \ldots, a_n)$, $\ldots$, $\nabla g_r(a_1, a_2, \ldots, a_n)$, if nonzero, is orthogonal to the subset at the point $(a_1, a_2, \ldots, a_n)$. If the constraints are all independent at the point, then we get a bunch of *linearly independent* (whatever that means) vectors that *span* (whatever that means) the normal space.

For a function f on $\mathbb{R}^n$, if $(a_1, a_2, \ldots, a_n)$ is a point in the $(n-r)$-dimensional subset where f attains an extreme value, then the directional derivative of f along any direction tangent to the subset is zero. Thus, the gradient vector ∇f is in the normal space. So, we can find constants λ_1, λ_2, $\ldots$, λ_r such that:

$$(\nabla f)(a_1, a_2, \ldots, a_n) = \lambda_1 \nabla g_1(a_1, a_2, \ldots, a_n) + \lambda_2 \nabla g_2(a_1, a_2, \ldots, a_n) + \cdots + \lambda_r \nabla g_r(a_1, a_2, \ldots, a_n)$$

Unfortunately, a deeper understanding of the ideas here requires a rudimentary understanding of linear algebra, which very few of you have had.

2.2. **Curves in $\mathbb{R}^3$.** Consider the case $n = 3$ and $r = 2$, i.e., we have a curve in $\mathbb{R}^3$ given by a pair of independent scalar equality constraints. Let's say the constraints are as follows:

$$\begin{aligned} g_1(x, y, z) &= k_1 \\ g_2(x, y, z) &= k_2 \end{aligned}$$

Suppose further that ∇g_1 is not the zero vector anywhere on the curve and ∇g_2 is also not the zero vector anywhere on the curve. Suppose further that ∇g_1 and ∇g_2 are linearly independent everywhere on the curve, i.e., it is never the case that the vectors are scalar multiples of each other.[1]

Then, at a point (x_0, y_0, z_0) on the curve, the normal space is spanned by the vectors $\nabla g_1(x_0, y_0, z_0)$ and $\nabla g_2(x_0, y_0, z_0)$. If (x_0, y_0, z_0) is a point of local extreme for a function f relative to the curve, then the theory of Lagrange multipliers tells us that:

$$\nabla f(x_0, y_0, z_0) = \lambda_1 \nabla g_1(x_0, y_0, z_0) + \lambda_2 \nabla g_2(x_0, y_0, z_0)$$

3. Using Lagrange multipliers

3.1. **Finding critical points: equation setup in codimension one.** We consider optimization for a codimension one subset in $\mathbb{R}^n$, of a function $f(x_1, x_2, \ldots, x_n)$ on the subset of $\mathbb{R}^n$ given by the equation $g(x_1, x_2, \ldots, x_n) = k$, with ∇g not a zero vector anywhere on the subset. Then, we need to solve the equation:

$$\begin{aligned} \nabla f(x_1, x_2, \ldots, x_n) &= \lambda \nabla g(x_1, x_2, \ldots, x_n) \\ g(x_1, x_2, \ldots, x_n) &= k \end{aligned}$$

[1]For two vectors, linear independence just means that neither is a scalar multiple of the other. The situation is considerably more complicated for more than two vectors.

The first of these is a vector equation. In particular, when we consider it coordinate-wise, we get n scalar equations. Counting the second equation as well, we get a total of $n+1$ scalar equations. There are $n+1$ variables: x_1, x_2, ..., x_n and λ. We thus have a system of $n+1$ equations in $n+1$ variables. The solution space is thus expected to be zero-dimensional, i.e., we *expect* that the set of solutions is a discrete collection of isolated points. These are the *critical points*.

3.2. **Finding critical points: equation setup in codimension r.** In the codimension r setup in $\mathbb{R}^n$ discussed earlier, we get the following equations:

$$\begin{aligned}
\nabla f(x_1, x_2, \ldots, x_n) &= \lambda_1 \nabla g_1(x_1, x_2, \ldots, x_n) + \lambda_2 \nabla g_2(x_1, x_2, \ldots, x_n) + \cdots + \lambda_r \nabla g_r(x_1, x_2, \ldots, x_n) \\
g_1(x_1, x_2, \ldots, x_n) &= k_1 \\
g_2(x_1, x_2, \ldots, x_n) &= k_2 \\
\cdot &= \cdot \\
g_r(x_1, x_2, \ldots, x_n) &= k_r
\end{aligned}$$

The first equation is n scalar equations, so we have a total of $n+r$ scalar equations. The number of variables is also $n+r$: the n variables x_1, x_2, ..., x_n, and the r variables λ_1, λ_2, ..., λ_r. The number of equations equals the number of variables, so we expect the solution set to be a bunch of isolated points.

3.3. **Absolute maxima and minima.** After we have computed all the critical points, we need to figure out which of them give rise to local maxima, which of them give rise to local minima, which of them give neither, and what the absolute maximum and minimum are. For local maxima and minima, we need an analogue of the second derivative test, which is too hard to develop and conceptually justify here. So, we simply avoid that question and only concentrate on finding the absolute maxima/minima.

As was the case with our earlier discussion of maxima/minima, we simplify matters and only consider the case where the space $g(x_1, x_2, \ldots, x_n) = k$ is a closed bounded set. Fortunately, since g is assumed to be continuous, the set $g(x_1, x_2, \ldots, x_n) = k$ is automatically a closed subset. Boundedness is something we can check for separately. Once we have established this, we can use the extreme value theorem, and conclude that the absolute maximum/minimum are attained. The strategy for finding them is as follows in a closed bounded subset without any boundary:

> Set up equations using Lagrange multipliers (as discussed above) and solve to find all critical points, that are candidates for the absolute maximum/minimum. Then, compare the function values at all these points. The smallest among these gives the absolute minimum, and the largest among these gives the absolute maximum.

Note that the situation becomes a little more complicated for a closed bounded subset that has a boundary, because the boundary is a smaller dimension subset. In this case, we need to separately find critical points relative to the boundary. Fortunately, this does not happen for typical subsets defined by conditions of the form $g(x_1, x_2, \ldots, x_n) = k$.

3.4. **Piecewise smooth curves, curves where the gradient vector becomes zero.** Recall that a necessary condition for the Lagrange condition for codimension one to hold at a point of local extremum is that ∇g be nonzero, i.e., the gradient of g have a well defined direction. If $\nabla g = 0$, the Lagrange condition may be violated at a local extreme value.

Similarly, for the Lagrange condition to hold in higher codimension, what we need is that ∇g_i be nonzero for each i, and further, that all the ∇g_i be linearly independent (whatever that means). Since the higher codimension case requires some knowledge of linear algebra, we'll skip it for now and stick to the codimension one case.

In the codimension one case, the condition $\nabla g = 0$ could arise for either of two reasons. First, there is no well defined normal direction to the codimension one subset at the point. This may happen because of a sharp cusp-like appearance or sudden direction change near the point, like the vertex of a cone. Second, the normal direction may be well defined but the function g may simply have been chosen poorly. For instance, consider $g(x, y) = (x - y)^3$ in two dimensions. The level curves for this are lines parallel to $y = x$. For each

such line, there is a well defined normal direction. However, for the line $y = x$ itself, although a normal direction does exist, ∇g takes the value zero.

To handle these kinds of situations we add in the following caveat to the Lagrange method:

> In addition to testing all points where the Lagrange condition holds, *also* test all points where $\nabla g = 0$ (if such points exist), when trying to find absolute maximum and minimum.

The technique does not work if we end up with infinitely many points satisfying $\nabla g = 0$. In this case, an alternative approach might work.

> Find a new function h and a constant l such that the set $g(x_1, x_2, \ldots, x_n) = k$ is the same as the set $h(x_1, x_2, \ldots, x_n) = l$, but such that ∇h is never zero on the set $h(x_1, x_2, \ldots, x_n) = l$.

This technique allows us to deal with piecewise smooth curves and the analogous surfaces. In the examples we gave of $g(x, y) = (x - y)^3$, the set $g(x, y) = 0$ can also be described as the set $h(x, y) = 0$ where $h(x, y) = x - y$. The function h has the advantage over g that its gradient is never zero, so it always provides a nonzero vector in the normal direction.

3.5. **An application to a polygon.** Suppose we want to find the maximum and minimum of a differentiable function f of two variables on a triangle with three vertices (here, triangle refers only to the boundary, not the interior region). The triangle is piecewise linear, and *on the interior of any side, all normal vectors point in the same direction.* The points we need to test are as follows:

- The three vertices of the triangles
- The solutions to the Lagrange condition on each side: Note that on any fixed side, all normal vectors can be taken to be the same, so we just need to solve $\nabla f = \lambda$ times a constant vector, along with the equation constraining the point to be on the line for that side of the triangle and the inequality constraining it to be between the vertices.

After finding a hopefully finite list of points, we evaluate the function at each of them and determine the local extreme values.

MAXIMUM AND MINIMUM VALUES: EXAMPLES

MATH 195, SECTION 59 (VIPUL NAIK)

What students should hopefully get: The description of critical points, local extreme values, and absolute extreme values for additively separable functions and (the more complicated version for) multiplicatively separable functions. The special nature of extreme values for quasiconvex and strictly quasiconvex functions and the notion of extreme points. The nature of extreme values for linear, quadratic, and homogeneous polynomials. The use of Lagrange multipliers to find extrema on the boundary.

Executive summary

(1) *Additively separable, critical points*: For an additively separable function $F(x,y) := f(x) + g(y)$, the critical points of F are the points whose x-coordinate gives a critical point for f and y-coordinate gives a critical point for g.

(2) *Additively separable, local extreme values*: The local maxima occur at points whose x-coordinate gives a local maximum for f and y-coordinates gives a local maximum for g. Similarly for local minima. If one coordinate gives a local maximum and the other coordinate gives a local minimum, we get a saddle point.

(3) *Additively separable, absolute extreme values*: If the domain is a rectangular region, rectangular strip, or the whole plane, then the absolute maximum occurs at the point for which each coordinate gives the absolute maximum for that coordinate, and analogously for absolute minimum. This does *not* work for non-rectangular regions in general.

(4) *Multiplicatively separable, critical points*: For a multiplicatively separable function $F(x,y) := f(x)g(y)$ with f, g, differentiable, there are four kinds of critical points (x_0, y_0): (1) $f'(x_0) = g'(y_0) = 0$, (2) $f(x_0) = f'(x_0) = 0$, (3) $g(y_0) = g'(y_0) = 0$, (4) $f(x_0) = g(y_0) = 0$.

(5) *Multiplicatively separable, local extreme values*: At a critical point of Type (1), the nature of local extreme value for F depends on the signs of f and g *and* on the nature of local extreme values for each. See the table. Critical points of Type (4) alone do not give local extreme values. The situation with critical points of Types (2) and (3) is more ambiguous and too complicated for discussion.

(6) *Multiplicatively separable, absolute extreme values*: Often, these don't exist, if one function takes arbitrarily large magnitude values and the other one takes nonzero values (details based on sign). If both functions are everywhere positive, and we are on a rectangular region, then the absolute maximum/minimum for the product occur at points whose coordinates give respective absolute maximum/minimum for f and g. (See notes)

(7) For a continuous quasiconvex function on a convex domain, the maximum must occur at one of the extreme points, in particular on the boundary. If the function is strictly quasiconvex, the maximum can occur only at a boundary point.

(8) For a continuous quasiconvex function on a convex domain, the minimum must occur on a convex subset. If the function is strictly quasiconvex, it must occur at a unique point.

(9) Linear functions are quasiconvex but not strictly so. The negative of a linear function is also quasiconvex. The maximum and minimum for linear functions on convex domains must occur at extreme points.

(10) To find maxima/minima on the boundary, we can use the method of Lagrange multipliers.

See also: tables, discussion for linear, quadratic, and homogeneous functions (hard to summarize).

1. Additively separable functions

1.1. **Partial derivatives, critical points and Hessian.** Consider a function of the form $F(x,y) := f(x) + g(y)$, i.e., F is additively separable. Then, $F_x(x,y) = f'(x)$, $F_y(x,y) = g'(y)$, $F_{xx}(x,y) = f''(x)$, $F_{yy}(x,y) = g''(y)$, and $F_{xy}(x,y) = 0$.

Note that in this case, the Hessian (the determinant used to determine the nature of extreme value at a critical point) is simply the product $F_{xx}F_{yy} = f''(x)g''(y)$.

We have the following:

> The critical points for F are precisely the points whose x-coordinate gives a critical point for f and y-coordinate gives a critical point for g. In other words, (x_0, y_0) in the domain of F gives a critical point if and only if x_0 gives a critical point for f and y_0 gives a critical point for g.

1.2. **Local extreme values.** We can say the following about local extreme values:

- The points of local maximum for F are precisely the points whose x-coordinate gives a local maximum for f and whose y-coordinate gives a local maximum for g. In other words, (x_0, y_0) in the domain of F gives a local maximum if and only if x_0 gives a local maximum for f and y_0 gives a local maximum for g.
- The points of local minimum for F are precisely the points whose x-coordinate gives a local minimum for f and whose y-coordinate gives a local minimum for g. In other words, (x_0, y_0) in the domain of F gives a local minimum if and only if x_0 gives a local minimum for f and y_0 gives a local minimum for g.
- In particular, a critical point gives a saddle point if any of these conditions hold: it is not a point of local extremum in one of the variables (e.g., a point of inflection type), *or* it is a point of local extrema of opposite kinds in the two variables. For instance, for the function $(x-1)^2 - (y-2)^2$, the point $(1,2)$ is a point of local minimum for the first coordinate but local maximum for the second, so we get a saddle point overall.
- For an additively separable function, the second derivative test simply boils down to checking whether $f''(x)$ and $g''(y)$ have the same sign. This is because the *interaction term* arising as a *mixed partial* is *absent*.

1.3. **Absolute extreme values.** Continuing notation from above, we note that:

- If the domain of F is rectangular (or the whole plane or a rectangular infinite strip) then the absolute maximum value for F occurs at a point whose x-coordinate maximizes f and whose y-coordinate maximizes g.
- If the domain of F is rectangular (or the whole plane or a rectangular infinite strip) then the absolute minimum value for F occurs at a point whose x-coordinate minimizes f and whose y-coordinate minimizes g.

These results don't hold for non-rectangular domains because we cannot carry out separate analysis of the variables. For instance, consider the function $x+y$ on the circular disk $x^2 + y^2 \leq 1$. The maximum for x occurs at $x = 1$, and the maximum for y occurs at $y = 1$. However, the point $(1,1)$ lies outside the domain of the function.

We will deal with non-rectangular regions in more detail a little later.

1.4. **Examples.** Consider the function:

$$F(x,y) := x^2 - 3x + \sin^2 y$$

This is the sum of the functions $f(x) := x^2 - 3x$ and $g(y) := \sin^2 y$. f attains its local and absolute minimum at $x_0 = 3/2$ with value $-9/4$, and it has no local or absolute maximum. g attains its local and absolute minima at multiples of π with value 0, and its local and absolute maxima at odd multiples of $\pi/2$, with value 1.

The upshot is that:

- F attains its local and absolute minima at points of the form $(3/2, n\pi)$, n an integer. This is because f is minimum on the x-coordinate and g is minimum on the y-coordinate.

- F has saddle points at $(3/2, n\pi + \pi/2)$. This is because f is minimum on the x-coordinate and g is maximum on the y-coordinate.

1.5. **Key observation: cases where second derivative test doesn't work.** Consider the additively separable function:

$$F(x, y) = (x - 1)^3 - (y - 2)^2$$

The function has a unique critical point for the point $(1, 2)$ in the domain. If we didn't notice additive separability, and directly tried to compute the Hessian, we'd get 0, indicating that the second derivative test is inconclusive.

We note that this is the sum of the functions $f(x) := (x - 1)^3$ and $g(y) := -(y - 2)^2$. Since we are now dealing with functions of *one* variable, we have methods other than the second derivative test (for instance, the first derivative test or higher derivative tests) to find out whether a given critical point gives a local extreme value. In this case we figure that $x = 1$ gives a point of inflection and *not* a local extreme value for f, whereas $y = 2$ gives a local maximum for g. Thus, overall, we conclude that $(1, 2)$ gives a saddle point.

In other words, for additively separable functions, we can go beyond the second derivative test using our knowledge of functions of one variable, despite our ignorance of analogous results for functions of two variables.

2. Multiplicatively separable functions

2.1. **Partial derivatives, critical points, and Hessian.** Consider a function of the form $F(x, y) := f(x)g(y)$, i.e., F is multiplicatively separable. Then, $F_x(x, y) = f'(x)g(y)$, $F_y(x, y) = f(x)g'(y)$, $F_{xx}(x, y) = f''(x)g(y)$, $F_{xy}(x, y) = f'(x)g'(y)$, and $F_{yy}(x, y) = f(x)g''(y)$.

The Hessian determinant (used for the second derivative test) at a point (x_0, y_0) thus becomes:

$$f(x_0)g(y_0)f''(x_0)g''(y_0) - [f'(x_0)g'(y_0)]^2$$

We first try to figure out the necessary and sufficient conditions for a point (x_0, y_0) to be a critical point for F. This happens iff $f'(x_0)g(y_0) = 0$ and $f(x_0)g'(y_0) = 0$. This could occur for four different reasons. We provide each reason along with an interpretation.

(1) $f'(x_0) = g'(y_0) = 0$: This means that x_0 is a critical point for f and y_0 is a critical point for g.
(2) $f(x_0) = f'(x_0) = 0$: Note that in this case the partial with respect to y is 0 at the point, not because of y_0, but because of x_0. What's happening is that on the line $x = x_0$, the function is identically zero, so changes in g do not matter.
(3) $g(y_0) = g'(y_0) = 0$: Note that in this case the partial with respect to x is 0 at the point, not because of x_0, but because of y_0. What's happening is that on the line $y = y_0$, the function is identically zero, so changes in f do not matter.
(4) $f(x_0) = 0$ and $g(y_0) = 0$: In this case, the function is identically zero along both the vertical and the horizontal line containing (x_0, y_0).

Note that any critical point that is of Type (4) above but not any of the preceding types must *fail* the second derivative test. For a critical point of Type (2) or (3) above, the second derivative test is inconclusive because we get 0 (more is discussed in the next subsection). For a critical point of Type (1), the second derivative test is most useful. Note that for such a critical point, the Hessian determinant simply becomes $f(x_0)g(y_0)f''(x_0)g''(y_0)$, so its sign depends not only on the signs of the second derivatives of f and g but *also* on the signs of the functions f and g themselves.

2.2. **Local extreme values: Type 1 case.** The following table gives conclusions for the nature of local extreme values of $F(x, y) = f(x)g(y)$ at (x_0, y_0) if x_0 gives a local extreme value for f and y_0 gives a local extreme value for g.

$f(x_0)$ sign	$g(y_0)$ sign	$f(x_0)$ (local max/min)	$g(y_0)$ (local max/min)	$F(x_0, y_0)$ (local max/min/saddle)
positive	positive	local max	local max	local max
positive	positive	local max	local min	saddle point
positive	positive	local min	local max	saddle point
positive	positive	local min	local min	local min
positive	negative	local max	local max	saddle point
positive	negative	local max	local min	local min
positive	negative	local min	local max	local max
positive	negative	local min	local min	saddle point
negative	positive	local max	local max	saddle point
negative	positive	local max	local min	local max
negative	positive	local min	local max	local min
negative	positive	local min	local min	saddle point
negative	negative	local max	local max	local min
negative	negative	local max	local min	saddle point
negative	negative	local min	local max	saddle point
negative	negative	local min	local min	local max

Note that the conclusion about F depends not merely on whether f and g have local max/min but also on the sign of the local max/min for f. The *saddle point cases* arise when f and g are pulling (multiplicatively) in opposite directions. Here, the function is a local maximum along one of the x- and y-directions and a local minimum along the other.

In cases where the second derivative test is conclusive for both f and g as functions of one variable, the above observations can be cross-checked by looking at the sign of the Hessian, which is $f(x_0)g(y_0)f''(x_0)g''(y_0)$, and of $F_{xx} = g(y_0)f''(x_0)$ and $F_{yy} = f(x_0)g''(y_0)$.

We do two of the sixteen examples for illustration:

- If f has a positive local maximum at x_0 and g has a negative local maximum at y_0, then F has a saddle point: We get $f(x_0) > 0$, $g(y_0) < 0$, $f''(x_0) < 0$, $g''(y_0) < 0$. So, the Hessian is negative (product of one positive and three negatives). Thus, by the second derivative test, F has a saddle point.
- If f and g both have negative local minima, then F has a local maximum. Here's how we see this: we get $f(x_0) < 0$, $g(y_0) < 0$, $f''(x_0) > 0$ (local minimum), $g''(y_0) > 0$ (local minimum), so multiplying all the signs, we see that the Hessian is positive. Thus, the function does attain a local extreme value. Next, we look at the sign of $F_{xx}(x_0, y_0)$, which is $g(y_0)f''(x_0)$. This is negative, since it is the product of a negative and a positive number. Thus, F has a local minimum.

It's important to keep in mind that the statements in the table *are more general and apply even when the second derivative tests are inconclusive*. We'll be looking at some examples shortly.

2.3. **Critical points of Types 2 and 3.** We now turn to the situation $F(x, y) = f(x)g(y)$ where there are points x_0 satisfying $f(x_0) = f'(x_0) = 0$. In this case, the second derivative test is inconclusive because the Hessian determinant takes the value 0.

Let's try to examine what's happening near the point. On the line $x = x_0$, the function is constant at 0. That explains why $F_y(x_0, y_0) = 0$ – the function is not changing along the y-direction because it's the product of $f(x_0)$, which is zero, and a changing number. On the line $y = y_0$, the function has derivative zero because $f'(x_0) = 0$, so that is why $F_x(x_0, y_0) = 0$.

Now, the first condition we need to obtain a local minimum is that the function be a local minimum under slight perturbations in the x-direction. So, we would like that the function $x \mapsto f(x)g(y_0)$ have a local minimum at x_0. If $g(y_0) > 0$, this is equivalent to wanting f to have a local minimum at x_0. If $g(y_0) < 0$, this is equivalent to wanting f to have a local maximum at x_0. In fact, as long as $g(y_0) \neq 0$, these are necessary and sufficient conditions to impose. Let's make this explicit in a table:

$g(y_0)$ sign	$f(x_0)$ (local max/min) at point with $f(x_0) = f'(x_0) = 0$	$F(x_0, y_0)$ (local max/min)
positive	local max	local max
negative	local max	local min
positive	local min	local min
negative	local min	local max

2.4. **Absolute extreme values.** Here are some results on the *non-existence* of absolute extreme values:

- If f takes a positive value anywhere on its domain, and g takes arbitrarily large positive values, then F takes arbitrarily large positive values, and hence has no absolute maximum.
- If f takes a negative value anywhere on its domain, and g takes arbitrarily large positive values, then F takes arbitrarily large magnitude negative values, and hence has no absolute minimum.
- If f takes a positive value anywhere on its domain, and g takes arbitrarily large magntiude negative values, then F takes arbitrarily large magnitude negative values, and hence has no absolute minimum.
- If f takes a negative value anywhere on its domain, and g takes arbitrarily large magnitude negative values, then F takes arbitrarily large positive values, and hence has no absolute maximum.

To each of the above, an analogous statement holds if we interchange the roles of f and g.

On the other hand, the following is true: if both f and g are everywhere positive, and the domain is a rectangular region, then the absolute minimum for F occurs at the point whose x-coordinate gives the absolute minimum for f and whose y-coordinate gives the absolute minimum for g.

2.5. **Examples (Type 1 critical points only).** Consider the function:

$$F(x, y) := (x^2 - x + 2)(3 + \cos y)$$

F is multiplicatively separable and can be written as $f(x)g(y)$ where $f(x) = x^2 - x + 2$ and $g(y) = 3 + \cos y$.

f has a unique critical point with a local and absolute *minimum* at $x = 1/2$, and the value of the minimum is $7/4$.

As for g, it attains its local and absolute maxima at multiples of 2π (with value 4) and its local and absolute minima at odd multiples of π (with value 2).

Note that there are no critical points of types (2), (3), and (4). This is because f is never 0 and further, g and g' are never simultaneously 0.

We can see from this that:

- F has local and absolute minimum attained at points of the form $(1/2, (2n + 1)\pi)$ with value $7/2$. This is because f has a local and absolute *positive* minimum at the point $1/2$ and g has a local and absolute *positive* minimum at the point $(2n + 1)\pi$.
- F has saddle points at points of the form $(1/2, 2n\pi)$ with value 7. This is because f has a local minimum at $1/2$ and g has a local maximum at $2n\pi$.
- F has no absolute maximum. To see this, note that f is unbounded from above and g takes values in $[2, 4]$.

Consider a very similar example:

$$F(x, y) := (x^2 - x + 2)\cos y$$

This is similar to the previous example except that the $3 + \cos y$ is replaced by $\cos y$. We take $f(x) = x^2 - x + 2$ and $g(y) = \cos y$. f has a unique local and absolute minimum at $x_0 = 1/2$ with value $7/4$. g has local and absolute maxima at even multiples of π with value 1, and local and absolute minima at odd multiples of π with value -1.

From this, we conclude that:

- F has *no* local maxima or minima. To see this, note that both the $(1/2, 2n\pi)$ and the $(1/2, (2n+1)\pi)$ cases give saddle points. For $(1/2, 2n\pi)$, we get minimum and maximum, all positive, which gives saddle points. For $(1/2, (2n+1)\pi)$, we get minimum and minimum, positive and negative respectively, which again gives saddle points.
- There are no absolute maxima and minima either. To see this, note that f is unbounded from above, and g takes values in $[-1, 1]$, so F can take arbitrarily large positive and negative values.

2.6. **Example (Type 1 and Type 4 critical points).** Consider the function:

$$F(x,y) := (x-2)(x-4)(y+1)(y-5)$$

This is the product of the function $f(x) := (x-2)(x-4)$ and $g(y) := (y+1)(y-5)$.

f has a unique critical point at $x_0 = 3$ and g has a unique critical point at $y_0 = 2$. So F has a unique critical point of Type 1 (with $f'(x_0) = g'(y_0) = 0$), namely the point $(3,2)$ in the domain. Since both f and g have *negative* local *minima* at the point, we conclude that F has a *positive* local *maximum* at the point.

However, there are also other kinds of critical points, specifically Type 4 critical points where $f(x_0) = g(y_0) = 0$. There are in fact four such critical points: $(2,-1)$, $(2,5)$, $(4,-1)$, and $(4,5)$.

As already mentioned earlier, the critical points that are only Type 4 critical points *cannot* must give saddle points, so we obtain saddle points at all these four points. The upshot:

- There is a unique positive local maximum at $(3,2)$.
- There are four saddle points: $(2,-1)$, $(2,5)$, $(4,-1)$, and $(4,5)$.
- There is *no* absolute maximum or minimum. To see this, note that f takes both a positive and a negative value, and g takes arbitrarily large positive values, so we can arrange the product to be a positive or a negative number of arbitrarily large magnitude.

2.7. **We don't need no second derivative test.** Consider a function such as:

$$F(x,y) := ((x-1)^4 - 1)((y+3)^6 - 2)$$

If we directly tried to use the second derivative test on this at the unique critical point $(1,-3)$, we'd get 0, i.e., the inconclusive case. However, thinking of the function as multiplicatively separable allows us to do a little better.

We consider F as a product of the function $f(x) := (x-1)^4 - 1$ and $g(y) := (y+3)^6 - 2$. Using the first derivative test (or the higher derivative tests) we can conclude that f has its unique local and absolute minimum with value -1 at $x_0 = 1$. g has its unique local and absolute minimum with value -2 at $y_0 = -3$. Consulting the table on various combinations, we note that F has a local *maximum* at $(1,-3)$ with value 2. However, F has no absolute maxima or minima because f takes both positive and negative values and g takes arbitrarily large positive values.

Incidentally, this also gives an example of a function with a unique local maximum and no local minimum but where the local maximum is not an absolute maximum.

The saddle points arising as Type 4 critical points in this case are $(0, \pm 2^{1/6} - 3)$ and $(2, \pm 2^{1/6} - 3)$.

3. Polynomials

3.1. **Linear polynomials.** A linear polynomial is a polynomial of the form:

$$f(x,y) := ax + by + c$$

where at least one of the values a and b is nonzero. For a linear polynomial, there are no critical points and hence no local extreme values. We may still have boundary extreme values, discussed later when we talk of maximization on closed bounded subsets.

3.2. **Homogeneous quadratic case.** We begin by looking at a homogeneous quadratic polynomial:

$$f(x,y) := ax^2 + bxy + cy^2$$

where at least one of the coefficients a, b, and c is nonzero.

First, we calculate the partials:

$$\begin{aligned} f_x(x,y) &= 2ax + by \\ f_y(x,y) &= bx + 2cy \\ f_{xy}(x,y) &= b \\ f_{xx}(x,y) &= 2a \\ f_{yy}(x,y) &= 2c \end{aligned}$$

The Hessian determinant in this case becomes the constant $4ac - b^2$, which is the *negative* of the *discriminant* of the quadratic polynomial.

If $4ac - b^2 \neq 0$, then the equations $2ax + by = 0$ and $bx + 2cy = 0$ are independent linear equations, so their solution set is the unique point $(0,0)$, so this is the only critical point. We note that:

- If $4ac - b^2 > 0$, i.e., the discriminant is negative, then this is a unique local extreme value for the function with value 0 attained at the origin $(0,0)$. Whether it is a maximum or a minimum depends on whether a and c are positive or negative. If $a > 0$, then the local extreme value of 0 at the origin is the unique minimum for the function. This also turns out to be the absolute maximum/minimum.
- If $4ac - b^2 < 0$, i.e., the discriminant is positive, then the function has no local extreme values. In fact, in this case, there are two lines through the origin on which the function takes the value 0. The two lines divide the plane into four regions. In two of these regions, the function can take negative values of arbitrarily large magnitude. In the other two regions, the function can take positive values of arbitrarily large magnitude.

Finally, if $b^2 = 4ac$, then the function attains its extreme value of 0 along a single line through the origin, obtained as the line $2ax + by = 0$. This is a minimum or maximum again depending on the sign of a.

These cases are also summarized in the table:

$4ac - b^2$ sign	$b^2 - 4ac$ sign	Extreme value	Points where it is attained	Nature (max or min?)
> 0	< 0	0	$(0,0)$	min if $a > 0$, max if $a < 0$
< 0	> 0	–	–	–
$= 0$	$= 0$	0	all points on the line $2ax + by = 0$	min if $a > 0$, max if $a < 0$

What's happening here will become clearer with our general analysis of homogeneous functions.

3.3. **Homogeneous polynomials and functions.** Suppose $F(x,y)$ is a homogeneous function of degree d. Then, we know by definition that:

$$F(ax, ay) = a^d F(x,y)$$

We note the following:

- For any line through the origin, either F is identically zero along the line, or the gradient of F at any point on the line other than the origin is a nonzero vector.
- In particular, this means that the only possibility for a local extreme value is 0, and this must be attained either only at the origin or at a union of lines through the origin.
- To determine what lines through the origin work, rewrite $F(x,y) = x^d g(m)$ where $m = y/x$. Now, find the values of m for which this function of one variable attains a local extreme value of 0. The lines $y = mx$ for these values of m are the relevant ones. The y-axis needs to be checked separately.

  To check whether the origin works, check whether g has a uniform sign (excluding points where it is zero).

In the light of this, the discussion of quadratics makes extra sense. Rewriting the quadratic $ax^2 + bxy + cy^2$, we get $x^2(a + bm + cm^2)$. The function is $g(m) = cm^2 + bm + a$. If the discriminant $b^2 - 4ac$ is less than 0, that means the quadratic has uniform sign, so the origin is a local extreme value, but there are no lines on which the function is zero.

If the discriminant is positive, then the quadratic g does not have uniform sign, and 0 is not a local extreme value, so there are no local extreme values for the function F of two variables.

If the discriminant is zero, then the quadratic g has a uniform sign except at an isolated point where it attains an extreme value of 0, so that corresponds to a line on which F attains its local extreme value of 0.

3.4. **Non-homogeneous quadratics: general discussion and specific example.** We now turn to the situation of a non-homogeneous quadratic polynomial. This is of the general form:

$$f(x,y) := ax^2 + bxy + cy^2 + px + qy + l$$

where at least one of the a, b, and c is nonzero.

Note that the value of l does not affect the points of local extrema, though it affects their values. The partial derivatives are:

$$\begin{aligned} f_x(x,y) &= 2ax + by + p \\ f_y(x,y) &= bx + 2cy + q \\ f_{xx}(x,y) &= 2a \\ f_{xy}(x,y) &= b \\ f_{yy}(x,y) &= 2c \end{aligned}$$

The Hessian determinant at any point is thus $4ac - b^2$. In other words, it is a *constant independent of the point* and is the negative of the discriminant of the homogeneous quadratic part.

Thus, we note the following cases:

- If $4ac - b^2 > 0$, or $b^2 - 4ac < 0$, then we get a unique critical point solving the simultaneous system of linear equations $f_x(x,y) = 0$ and $f_y(x,y) = 0$ and this gives a local extreme value. It is a local minimum if $a > 0$ and a local maximum if $a < 0$.
- If $4ac - b^2 < 0$ or $b^2 - 4ac > 0$, then we get a unique critical point solving the simultaneous system of linear equations $f_x(x,y) = 0$ and $f_y(x,y) = 0$, but this gives a saddle point and not a local extremum.
- If $4ac - b^2 = 0$, then, depending on the values p and q, we either have *no* critical points or *a line's worth* of critical points. In the latter case, the line's worth of critical points gives local extrema.

4. Boundary issues

4.1. **The concept of a quasiconvex function.** Suppose f is a function defined on a domain D in $\mathbb{R}^n$. Suppose the domain D is a convex subset of $\mathbb{R}^n$, i.e., given any two points in D, the line segment joining them lies completely in D.

We say that f is a *quasiconvex function* if given any two points P and Q in D, the maximum value of f on the line segment joining P and Q is attained at one (or possibly both) of the endpoints P and Q. In other words, the value in the interior of a line segment is less than or equal to the value at one (or possibly both) of the endpoints.

We say that f is *strictly quasiconvex* if the maximum can occur *only* at endpoints, i.e., it is not possible for the maximum value to also be attained at an interior point.

We note that:

- All linear functions are quasiconvex but not strictly quasiconvex.
- All convex functions are quasiconvex and all strictly convex functions are strictly quasiconvex. We're not going to go into the meaning of convex and strictly convex here (the definition is a non-calculus definition and is fairly simple but will take us a little off track). For functions of one variable, strictly convex simply means *concave up* if the function is continuously differentiable. For functions of two variables, if the Hessian determinant is strictly positive everywhere (except possibly at isolated points where it is zero) and the second pure partials are positive everywhere, the function is strictly convex.
- Quadratic functions of two variables with negative discriminant of the homogeneous part (i.e., positive Hessian determinant) and with positive coefficients on the square terms are strictly convex and hence strictly quasiconvex.

There are two reasons quasiconvex functions are significant:

- The *maximum* of a continuous quasiconvex function on a closed bounded convex domain *must* be attained somewhere on its boundary. In fact, we can go further and note that the maximum must

be attained at an *extreme point* of the domain: a point not in the interior of any line segment within the domain.

For a strictly quasiconvex function, the maximum *cannot* be attained at any point other than an extreme point. For a function that's quasiconvex but not strictly so, the maximum may also be attained at other points.

- The *minimum* of a continuous quasiconvex function on a closed bounded convex domain is attained either at a unique point or on a convex subset of the convex domain, i.e., if the minimum occurs at two distinct points in the domain, it also occurs at all points in the line segment joining them.

  For a strictly quasiconvex function, the minimum *must* be attained at a *unique* point.
- For a continuous function that is the negative of a quasiconvex function, the same observations as above hold but with the roles of maximum and minimum interchanged.

We now consider some examples of maximization for functions that are strictly quasiconvex.

4.2. **Linear examples.** Note that *for linear functions*, both the function and its negative are quasiconvex, so both the maximum and the minimum of a linear function on a closed bounded convex domain must occur at extreme points.

Consider a function $f(x, y) = 2x - 3y$ on the square region $[-1, 1] \times [-1, 1]$.

Since f is a linear function, it is quasiconvex (though not strictly so). This means that the maximum value, if it occurs, must occur at one of the extreme points. The extreme points of a square are its four vertices, i.e., the vertices $(-1, -1)$, $(-1, 1)$, $(1, 1)$, and $(1, -1)$. We simply need to evaluate f at all these points and see which is the largest. We have $f(-1, -1) = 1$, $f(-1, 1) = -5$, $f(1, 1) = -1$, $f(1, -1) = 5$. The largest occurs at $(1, -1)$ with value 5, so this is the maximum.

In this case, the minimum occurs at $(-1, 1)$ with value -5.

Now consider another example:

$$f(x, y) := 3x + 4y$$

on the circular disk $x^2 + y^2 \leq 1$.

Here, the maximum and minimum both occur on the boundary circle $x^2 + y^2 = 1$. However, all points on the boundary circle are extreme points, so the minimum may be attained at any of them – we cannot rule any point offhand. The problem thus reduces to maximizing a function on the circle. There are many ways of doing this, including Lagrange multipliers (which we'll see shortly) but one approach is to put $x = \cos\theta$, $y = \sin\theta$, and thus convert the problem to a maximization/minimization in one variable of the function $g(\theta) = 3\cos\theta + 4\sin\theta$.

4.3. **Strictly convex quadratic example.**

$$f(x, y) := 2x^2 - 2xy + y^2 - x + 3$$

Suppose we want to calculate the maximum and minimum values of f on the square region $[0, 1] \times [0, 1]$.

We compute the first and second partials:

$$\begin{aligned} f_x(x, y) &= 4x - 2y - 1 \\ f_y(x, y) &= 2y - 2x \\ f_{xx}(x, y) &= 4 \\ f_{yy}(x, y) &= 2 \\ f_{xy}(x, y) &= -2 \end{aligned}$$

We note that the Hessian determinant is 4 *everywhere* and $f_{xx} = 4$ is positive everywhere. So this is an example of a strictly convex, and hence strictly quasiconvex, quadratic function. As also seen from the discussion of quadratics, we should get a unique critical point that gives a local extreme value. We first find the critical point by solving:

$$\begin{aligned} 4x - 2y - 1 &= 0 \\ 2y - 2x &= 0 \end{aligned}$$

Solving, we get $x = 1/2$, $y = 1/2$. This gives the point $(1/2, 1/2)$, which lies within the domain. Evaluating the function at this point gives $f(1/2, 1/2) = 11/4$. This is the unique local minimum of the function on the whole plane and hence on the domain. In this case, it also turns out to be the absolute minimum. *Note: In general, there are examples of continuous functions having a unique local minimum that is not an absolute minimum. But those examples don't include strictly quasiconvex functions.*

Since the function is strictly convex and hence strictly quasiconvex, the maximum must occur at one of the four corner points. It remains to evaluate f at the four boundary points $(0, 0)$, $(1, 0)$, $(1, 1)$, and $(0, 1)$. We get $f(0, 0) = 3$, $f(1, 0) = 4$, $f(1, 1) = 3$, $f(0, 1) = 4$. Of these, we see that the maximum occurs at the points $(1, 0)$ and $(0, 1)$, with a value of 4.

5. Combining max-min and Lagrange

For quasiconvex functions, we already noted that the maximum must occur somewhere on the boundary of the domain. For strictly quasiconvex functions, it can occur *only* at boundary points.

Even for functions that aren't quasiconvex, maximization over a certain domain requires us to find not just local maxima/minima in the interior but also boundary maxima/minima and then compare them all. Lagrange multipliers can offer a method to do so.

Let's return to our earlier example:

$$f(x, y) := 3x + 4y$$

on the disk $x^2 + y^2 \leq 1$.

Since this is a linear function, both the function and its negative are quasiconvex, and since the domain of consideration is a convex domain, the maximum and minimum must both be attained at the boundary circle $x^2 + y^2 = 1$.

We can now use the method of Lagrange multipliers to carry out the maximization and minimization relative to the boundary. Here, $g(x, y) = x^2 + y^2$, so $\nabla g(x, y) = \langle 2x, 2y \rangle$. Also, $\nabla f(x, y) = \langle 3, 4 \rangle$. We get:

$$\langle 3, 4 \rangle = \lambda \langle 2x, 2y \rangle, \qquad x^2 + y^2 = 1$$

Simplifying, we get:

$$\begin{aligned} 3 &= 2\lambda x \\ 4 &= 2\lambda y \\ x^2 + y^2 &= 1 \end{aligned}$$

Plug in $x = 3/(2\lambda)$ and $y = 2/\lambda$ into the third equation and we get:

$$\frac{9}{4\lambda^2} + \frac{4}{\lambda^2} = 1$$

Simplifying, we get:

$$\lambda = \pm 5/2$$

Plugging back, we obtain that the two candidate critical points are $\langle 3/5, 4/5 \rangle$ and $\langle -3/5, -4/5 \rangle$. In this case, it is easy to see from inspection that $\langle 3/5, 4/5 \rangle$ is a point of local maximum with value 5 and $\langle -3/5, -4/5 \rangle$ is a point of local minimum with value -5.