Math 291-3: Intensive Linear Algebra & Multivariable Calculus

Northwestern University, Lecture Notes

Written by Santiago Cañez

These are notes which provide a basic summary of each lecture for Math 291-3, the third quarter of "MENU: Intensive Linear Algebra & Multivariable Calculus", taught by the author at Northwestern University. The book used as a reference is the 4th edition of *Vector Calculus* by Colley. Watch out for typos! Comments and suggestions are welcome.

These notes will focus on material covered in Math 291 which is not normally covered in Math 290, and should thus be used in conjunction with my notes for Math 290, which are available at http://www.math.northwestern.edu/∼scanez/courses/290/notes.php.

Contents

Lecture 1: Local Extrema

Definition. Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ be a real-valued function defined on some open subset U of $\mathbb{R}^n$. We say that a point $\mathbf{a} \in U$ is a *local maximum* of f if there exists an open ball $B_r(\mathbf{a})$ around $\mathbf{a}$ such that

$$f(\mathbf{p}) \leq f(\mathbf{a}) \text{ for all } \mathbf{p} \in B_r(\mathbf{a}),$$

and we say $\mathbf{a} \in U$ is a *local minimum* of f if there exists an open ball $B_s(\mathbf{a})$ around $\mathbf{a}$ such that

$$f(\mathbf{a}) \leq f(\mathbf{q}) \text{ for all } \mathbf{q} \in B_s(\mathbf{a}).$$

Thus, a local maximum is a point at which the value of f is greater than or equal to its value at all points nearby, and a local minimum is a point at which the value of f is less than or equal to its value at all points nearby. A *local extremum* of f is a point which is either a local maximum or a local minimum.

Theorem. The first key property of local extrema is that, just as in the single-variable case, the derivative of the function in question is zero at such points. To be precise, the claim is that if $\mathbf{a} \in U$ is a local extremum of $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$, then $Df(\mathbf{a}) = \mathbf{0}$. We will give a proof below, and the upshot is that if we are trying to find the local extrema of a function, we should first find points at which the derivative is zero.

Proof. We take as given the analogous fact about single-variable functions: the derivative of a single-variable function at a local maximum or local minimum is zero. The idea is that if f has a local maximum (or minimum) at $\mathbf{a}$, then the single-variable function obtained by only varying one coordinate at a time still has a local maximum (or minimum) at $\mathbf{a}$. That is, fix $1 \leq i \leq n$ and consider the single-variable function

$$g(x_i) = f(a_1, \ldots, x_i, \ldots, a_n)$$

obtained by holding all variables except for x_i constant. (Here, $a_1, \ldots, a_n$ are the coordinates of $\mathbf{a}$.) If f has a local maximum or minimum at $\mathbf{a}$, then $g(x_i)$ also has a local maximum or minimum at $\mathbf{a}$. Thus the derivative of g with respect to x_i at $\mathbf{a}$ is zero; this derivative is precisely the partial derivative of f with respect to x_i at $\mathbf{a}$, so we conclude that

$$\frac{\partial f}{\partial x_i}(\mathbf{a}) = 0 \text{ for all } i,$$

and hence $Df(\mathbf{a}) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(\mathbf{a}) & \cdots & \frac{\partial f}{\partial x_n}(\mathbf{a}) \end{pmatrix} = \mathbf{0}$ as claimed. □

Definition. A point $\mathbf{a} \in U$ at which $Df(\mathbf{a}) = \mathbf{0}$ is called a *critical point* of f, just as in the single-variable case. The result above thus says that local extrema are always critical points, but it is not necessarily true that a critical point must be a local extrema.

In particular, we single out one more special type of critical point: a critical point $\mathbf{a} \in U$ of f is called a *saddle point* of f if it has the property that for *any* ball $B_r(\mathbf{a})$ around it, there always exist $\mathbf{p} \in B_r(\mathbf{a})$ and $\mathbf{q} \in B_r(\mathbf{a})$ such that

$$f(\mathbf{p}) < f(\mathbf{a}) \quad \text{and} \quad f(\mathbf{a}) < f(\mathbf{q}).$$

Thus a saddle point is a critical point such that no matter how close we get to it, we can always find points at which f gives something greater than and something smaller than the value at that point.

In particular, this implies that a saddle point is neither a local minimum nor a local maximum, and so really is a new type of critical point.

Example. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be the function $f(x, y) = 4x + 6y - 12 - x^2 - y^2$. We have

$$Df(x, y) = \begin{pmatrix} 4 - 2x & 6 - 2y \end{pmatrix},$$

so the only critical point of f is $(2, 3)$. Moreover, we can rewrite f as

$$f(x, y) = -(x - 2)^2 - (y - 3)^2 + 1,$$

which implies that $(2, 3)$ is a local maximum of f: the expression $-(x - 2)^2 - (y - 3)^2$ is always negative for a point which is not $(2, 3)$, so the value at such a point is always less than the value at $(2, 3)$. The graph of f in this case is a downward-opening paraboloid, which is the prototypical example of a local maximum in the two variable case.

Let $g : \mathbb{R}^2 \to \mathbb{R}$ be the function $g(x, y) = x^2 - 2y^2 + 2x + 3$. The only critical point of this function is $(-1, 0)$, and after rewriting g as

$$g(x, y) = (x + 1)^2 - 2y^2 + 2,$$

we can see that $(-1, 0)$ is a saddle point of g. Indeed, the graph of g is a hyperbolic paraboloid (i.e. a saddle, which is where the name "saddle point" comes from), which implies that we can always find points close to $(-1, 0)$ at which g is either greater than or less than its value at $(-1, 0)$.

Classifying critical points. The examples above where special in that the function involved were polynomials of degree 2, but for more general functions classifying critical points isn't as simple as rewriting the function in a nice way. Instead, in more general settings we make use a fact we finished with last quarter, namely that a C^2 function is well-approximated near a point by its second-order Taylor polynomial at that point.

To be clear, the idea is that if f is C^2, then

$$f(\mathbf{a} + \mathbf{h}) \approx f(\mathbf{a}) + Df(\mathbf{a})\mathbf{h} + \frac{1}{2}\mathbf{h}^T D^2 f(\mathbf{a})\mathbf{h}$$

for points $\mathbf{a} + \mathbf{h}$ "close enough" to $\mathbf{a}$, or equivalently for points $\mathbf{a} + \mathbf{h}$ where $\mathbf{h}$ is "close enough" to $\mathbf{0}$. When $\mathbf{a}$ is a critical point of f, this simplifies to

$$f(\mathbf{a} + \mathbf{h}) \approx f(\mathbf{a}) + \frac{1}{2}\mathbf{h}^T D^2 f(\mathbf{a})\mathbf{h},$$

and the point is that the "definiteness" (i.e. whether it is positive definite, negative definite, or indefinite) of the symmetric matrix $D^2 f(\mathbf{a})$ describes whether the right side is always larger than $f(\mathbf{a})$, always smaller than $f(\mathbf{a})$, or can be both larger and smaller than $f(\mathbf{a})$ depending on $\mathbf{h}$. Since this right side well-approximates $f(\mathbf{a} + \mathbf{h})$, this in turn should tell us whether $f(\mathbf{a} + \mathbf{h})$ is larger or smaller than $f(\mathbf{a})$, as we would need to know in order to say what type of critical point $\mathbf{a}$ is.

To make this precise, we need to recall what it means to say that the given expression "well-approximates" f near $\mathbf{a}$. The correct statement is that

$$f(\mathbf{a} + \mathbf{h}) = f(\mathbf{a}) + Df(\mathbf{a})\mathbf{h} + \frac{1}{2}\mathbf{h}^T D^2 f(\mathbf{a})\mathbf{h} + R(\mathbf{h})$$

where the "remainder" term $R(\mathbf{h})$ (which is just $f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) - Df(\mathbf{a})\mathbf{h} - \frac{1}{2}\mathbf{h}^T D^2 f(\mathbf{a})\mathbf{h}$) has the property that

$$\frac{R(\mathbf{h})}{\|\mathbf{h}\|^2} \to 0 \text{ as } \mathbf{h} \to \mathbf{0}.$$

As we'll see, this property of the remainder term gives us a way to control how large or how small it can be, which will then give us a way to say something concrete about $f(\mathbf{a}+\mathbf{h})$ in relation to $f(\mathbf{a})$. Note again that when $\mathbf{a}$ is a critical point of f, the equality above can be rewritten as

$$f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) = \frac{1}{2}\mathbf{h}^T D^2 f(\mathbf{a})\mathbf{h} + R(\mathbf{h})$$

since the Jacobian term $Df(\mathbf{a})\mathbf{h}$ vanishes. The idea, as mentioned above, is that the eigenvalues of the Hessian determine the behavior of $f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a})$.

Definition. Before moving on we need one more definition. We say that a critical point $\mathbf{a}$ of f is *nondegenerate* if the Hessian matrix $D^2 f(\mathbf{a})$ of f at $\mathbf{a}$ is invertible, or equivalently if 0 is not an eigenvalue of the Hessian. A critical point is *degenerate* if $D^2 f(\mathbf{a})$ is not invertible, or equivalently has 0 as an eigenvalue.

Second Derivative Test. Here then is the multivariable version of the second derivative test. Suppose $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ is C^2 and that $\mathbf{a} \in U$ is a nondegenerate critical point of f. Then:

- if $D^2 f(\mathbf{a})$ is positive definite, $\mathbf{a}$ is a local minimum of f,
- if $D^2 f(\mathbf{a})$ is negative definite, $\mathbf{a}$ is a local maximum of f, and
- if $D^2 f(\mathbf{a})$ is indefinite, $\mathbf{a}$ is a saddle point of f.

When $\mathbf{a}$ is a degenerate critical point, the second derivative test gives no information as to what type of critical point $\mathbf{a}$ actually is.

Proof. We'll prove the second derivative test in the positive definite case, which illustrates the main ideas needed in the other cases as well. The reason to go through this proof is that it is a nice blend of both linear algebraic tools and analytic (i.e. calculus-based) tools.

Before giving the proof we need one linear algebraic fact: if $q(\mathbf{h}) = \mathbf{h}^T A\mathbf{h}$ is a positive definite quadratic form (meaning that the symmetric matrix A is positive definite), then there exists a positive constant M such that

$$q(\mathbf{h}) \geq M \|\mathbf{h}\|^2 \text{ for all } \mathbf{h}.$$

The book derives this as a consequence of the Extreme Value Theorem without using linear algebra, but we'll instead give a linear algebraic reason for this, which will also tell us precisely what M is. Orthogonally diagonalizing A gives an orthonormal eigenbasis $\mathbf{u}_1, \dots, \mathbf{u}_n$ of $\mathbb{R}^n$ relative to which q takes the form

$$q(\mathbf{h}) = \lambda_1 c_1^2 + \cdots + \lambda_n c_n^2$$

where $\lambda_1, \dots, \lambda_n$ are the eigenvalues of A and $c_1, \dots, c_n$ the coordinates of $\mathbf{h}$ taken with respect to $\mathbf{u}_1, \dots, \mathbf{u}_n$. Letting M denote the smallest eigenvalue of A, which is strictly positive since we are assuming A is positive definite, we get:

$$q(\mathbf{h}) = \lambda_1 c_1^2 + \cdots + \lambda_n c_n^2 \geq M c_1^2 + \cdots + M c_n^2 = M(c_1^2 + \cdots + c_n^2).$$

We can check that if $\mathbf{h} = c_1\mathbf{u}_1 + \cdots + c_n\mathbf{u}_n$, then $\|\mathbf{h}\|^2 = c_1^2 + \cdots + c_n^2$, so we get $q(\mathbf{h}) \geq M\|\mathbf{h}\|^2$ as claimed. (Again, check the book for a proof of this using the Extreme Value Theorem instead.)

Now, assuming $D^2 f(\mathbf{a})$ is positive definite, our goal is to show that $\mathbf{a}$ is a local minimum of f, meaning that

$$f(\mathbf{a}+\mathbf{h}) \geq f(\mathbf{a}) \text{ for } \mathbf{a}+\mathbf{h} \text{ close enough to } \mathbf{a}.$$

This is the same as showing that

$$f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) = \frac{1}{2}\mathbf{h}^T D^2 f(\mathbf{a})\mathbf{h} + R(\mathbf{h})$$

is greater than or equal to 0, where $R(\mathbf{h})$ is the remainder term described above. The first term on the right is greater than or equal to $M\|\mathbf{h}\|^2$ where M is the smallest eigenvalue of the positive definite symmetric matrix $\frac{1}{2}D^2 f(\mathbf{a})$. Since

$$\frac{R(\mathbf{h})}{\|\mathbf{h}\|^2} \to 0 \text{ as } \mathbf{h} \to \mathbf{0},$$

we know that there exists $B_r(\mathbf{0})$ such that

$$\left|\frac{R(\mathbf{h})}{\|\mathbf{h}\|^2}\right| < M \text{ for } \mathbf{h} \in B_r(\mathbf{0})$$

from the rigorous definition of limits. (Indeed, since $R(\mathbf{h})/\|\mathbf{h}\|^2$ approaches 0, eventually its absolute value must be smaller than whatever positive number M happens to be once $\mathbf{h}$ is close enough to $\mathbf{0}$.) This inequality can be rewritten as

$$|R(\mathbf{h})| < M\|\mathbf{h}\|^2 \text{ for } \mathbf{h} \in B_r(\mathbf{0}),$$

which is the same as

$$-M\|\mathbf{h}\|^2 < R(\mathbf{h}) < M\|\mathbf{h}\|^2$$

if we get rid of the absolute value. Thus, in the expression

$$f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) = \frac{1}{2}\mathbf{h}^T D^2 f(\mathbf{a})\mathbf{h} + R(\mathbf{h}),$$

the first term on the right is greater than or equal to $M\|\mathbf{h}\|^2$ while the second term on the right is greater than $-M\|\mathbf{h}\|^2$, so we get that

$$f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) = \frac{1}{2}\mathbf{h}^T D^2 f(\mathbf{a})\mathbf{h} + R(\mathbf{h}) \geq M\|\mathbf{h}\|^2 + (-M\|\mathbf{h}\|^2) = 0$$

for $\mathbf{h} \in B_r(\mathbf{0})$. Saying that $\mathbf{h} \in B_r(\mathbf{0})$ is the same as saying that $\mathbf{a}+\mathbf{h} \in B_r(\mathbf{a})$, so this inequality implies that

$$f(\mathbf{a}+\mathbf{h}) \geq f(\mathbf{a}) \text{ for } \mathbf{a}+\mathbf{h} \in B_r(\mathbf{a}),$$

which says that $\mathbf{a}$ is a local minimum of f as claimed.

The proof that $\mathbf{a}$ is a local maximum in the case where $D^2 f(\mathbf{a})$ is negative definite is very similar, the only difference between that various inequalities are flipped around, and the proof that $\mathbf{a}$ is a saddle point when $D^2 f(\mathbf{a})$ is indefinite is a hybrid of the proofs in the other two cases. Again, check the book if you're interested in the details, but the main ideas are the same as the ones we used above. □

Example. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be the function $f(x,y) = x^2 - y^3 - x^2 y + y$. This has two critical points: $(0, 1/\sqrt{3})$ and $(0, -1/\sqrt{3})$. The Hessians at these two points are:

$$D^2 f(0,1/\sqrt{3}) = \begin{pmatrix} 2 - \frac{2}{\sqrt{3}} & 0 \\ 0 & -\frac{6}{\sqrt{3}} \end{pmatrix} \quad \text{and} \quad D^2 f(0,-1/\sqrt{3}) = \begin{pmatrix} 2 + \frac{2}{\sqrt{3}} & 0 \\ 0 & \frac{6}{\sqrt{3}} \end{pmatrix}.$$

The first Hessian is indefinite, so $(0, 1/\sqrt{3})$ is a saddle point of f, and the second Hessian is positive definite, so $(0, -1/\sqrt{3})$ is a local minimum of f. If you look at the graphs of these two functions on a computer, you'll see that near $(0, 1/\sqrt{3})$ the graph looks like a hyperbolic paraboloid and that near $(0, -1/\sqrt{3})$ is looks like an upward-opening paraboloid.

Lecture 2: Global Extrema

Warm-Up. We show that the function

$$f(x, y) = 3ye^x - e^{3x} - y^3$$

has a unique critical point, which is local maximum but not a global maximum. (A global maximum is a point $\mathbf{a}$ such that $f(\mathbf{a}) \geq f(\mathbf{x})$ for all $\mathbf{x}$ in the domain of the function. Contrast this with what it means to say that $\mathbf{a}$ is a local maximum, where in that case such an inequality is only required to hold on some open ball around $\mathbf{a}$.) The point is that this is not something which can happen in the single-variable case: if a single-variable function has a local maximum, which is the only critical point of the function, then that local maximum must be a global maximum. Thus, this example illustrates an important different between single-variable and multivariable optimization problems.

The gradient of the given function is

$$\nabla f(x, y) = (3ye^x - 3e^{3x}, 3e^x - 3y^2),$$

which is $\mathbf{0}$ when

$$3ye^x = 3e^{3x} \text{ and } 3e^x = 3y^2.$$

The first condition gives $y = e^{2x}$, which combined with the second gives $e^x = e^{4x}$. Thus we must have $x = 0$, and in turn $y = 1$. Hence $(0, 1)$ is the only critical point of f. The Hessian of f is:

$$D^2 f(x, y) = \begin{pmatrix} 3ye^x - 9e^{3x} & 3e^x \\ 3e^x & -6y \end{pmatrix}, \text{ so } D^2 f(0, 1) = \begin{pmatrix} -6 & 3 \\ 3 & -6 \end{pmatrix}.$$

Hence $D^2 f(0, 1)$ is negative definite (it has positive determinant, so both eigenvalues are of the same sign, and it has negative trace, so the eigenvalues must be negative), so $(0, 1)$ is a local maximum of f as claimed.

Now, along the y-axis the value of f are given by:

$$f(0, y) = 3y - 1 - y^3.$$

As $y \to -\infty$, this expression goes to $+\infty$, so f grows without bound and hence cannot have a global maximum. Thus, the $(0, 1)$ is a local maximum which is not a global maximum.

To mention another example of something which can happen in the multivariable setting but not the single-variable setting: a multivariable function can have two critical points, both of which are local maximums, but not have a local minimum. In the single-variable case, if a function has two local maximums it must have a local minimum "between" them, but not so in higher dimensions. You'll see an example of this on the homework.

Morse Theory. At this point we went on a bit of a tangent to briefly discuss what's called *Morse Theory*, which is a subject devoted to using critical points of functions $f : X \to \mathbb{R}$ to derive geometric information about X itself. This is also outlined in my Math 290-2 lecture notes, so I'll just refer you to those if interested.

Global extrema. We mentioned the idea of a "global maximum" above, but let us spell it out precisely again. Given $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$, a point $\mathbf{a} \in U$ is a *global maximum* of f if $f(\mathbf{a}) \geq f(\mathbf{x})$ for all $\mathbf{x} \in U$. Similarly, $\mathbf{a}$ is a *global minimum* of f if $f(\mathbf{a}) \leq f(\mathbf{x})$ for all $\mathbf{x} \in U$. Thus, global maximum describe the largest value a function can have over a region, and global minimums give the smallest value. Note that the Extreme Value Theorem says that continuous functions always

have global maximums and minimums over compact regions, so searching for such extrema is not done in vain.

Finding such global extrema is usually a multi-step process which works as follows. First we determine the critical points of the function, which gives us possible local maximums and minimums, which might be the global maximum and minimums we are looking for. Next we determine the possible maximum and minimum points of our given function along the *boundary* of the region in question. There are various ways of doing this, for now the main way being to use the equation of the boundary or turn a multivariable function into a single-variable one along the boundary. Finally, we test all points we've found (including any "corners" or points occurring at the endpoints of any intervals used throughout) by plugging them into the given function to see which gives the largest value and which gives the smallest value.

Examples. In class we worked through a couple of examples illustrating the process described (vaguely) above: optimizing $f(x, y) = x^2 + xy + y^2 - 6y$ over the rectangle $[-3, 3] \times [0, 5]$, and then optimizing $f(x, y) = x^2 y$ over the region described by $3x^2 + 4y^2 \leq 12$, which is the region enclosed by the ellipse $3x^2 + 4y^2 = 12$. Both of these examples are worked out in my Math 290-2 lecture notes, so I won't reproduce them here and instead will direct you to those notes. Note, as mentioned previously, that in each of these examples the Extreme Value Theorem guarantees the global extrema we are after actually exist.

Lecture 3: Lagrange Multipliers

Warm-Up. Say we want to determine the maximum and minimum values of $f(x, y) = xy$ over the unit disk $x^2 + y^2 \leq 1$. (Since the unit disk is compact and f is continuous, the Extreme Value Theorem guarantees that such extrema exist.) The only critical point of f is the origin, at which the value of f is zero. Since f definitely attains positive values, say in the first quadrant, and negative values, say in the second quadrant, we know that the origin does not give a global maximum nor minimum, so we ignore it. (The origin is in fact a saddle point.)

Thus in order to determine the absolute maximum and minimum of f over the unit disk, we are left checking for maxima and minima along the boundary circle $x^2 + y^2 = 1$. One way to do this is to solve for either x or y in the equation of the circle, and use the resulting expression to write f as a function of one variable over the boundary. For instance, $x = \pm\sqrt{1 - y^2}$ on the boundary, so

$$f(x, y) = \pm y\sqrt{1 - y^2} \text{ where } -1 \leq y \leq 1,$$

where the positive square root gives f over the right half of the circle and the negative square root gives f over the left half. We could then optimize this single variable function, remembering to check the points where $y = \pm 1$ at the end.

Alternatively, to avoid working with square roots we can notice that maximizing $f(x, y)$ is the same as maximize $f(x, y)^2$, since xy is at a maximum (knowing that already that this expression can be positive) when $x^2 y^2$ is at a maximum. Thus we instead consider the problem of maximizing $g(x, y) = x^2 y^2$ over the unit circle, which we can approach by again using the equation of the circle to write g as a function of one variable; the benefit, as mentioned, is that we avoid introducing square roots this way. As for the minimum value of f, we can note a local minimum (which we know to be a negative value) for f also corresponds to a local maximum value for g since "large" negatives become large positives after squaring.

One final approach to notice is that we can interpret this question in terms of polar coordinates instead. In polar coordinates, f becomes

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

which after restricting to the boundary circle $r = 1$ becomes a function of one variable

$$f(1, \theta) = \cos\theta \sin\theta.$$

We can then optimize this single-variable function as normal. You can see how all this work out in detail in my Math 290-2 lecture notes, and the point of mentioning this problem here was really just to illustrate different ways of optimizing a function, in particular the idea of optimizing its square instead or of converting to different coordinates.

Constrained extrema. The problem of optimizing the function in the Warm-Up among points lying on the unit circle is an example of a *constrained optimization* problem. The setup is as follows.

Say we are giving a function $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ we wish to optimize subject to a *constraint* $g(\mathbf{x}) = c$, where $g : U \subseteq \mathbb{R}^n \to \mathbb{R}$ is another function. This means we want to determine the maxima/minima of f but only among points satisfying the constraint equation $g(\mathbf{x}) = c$. The constraint equation defines a set S

$$S := \{\mathbf{x} \in U \mid g(\mathbf{x}) = c\}$$

of points satisfying the constraint, so the goal is to optimize the function $f : S \to \mathbb{R}$ obtained by restricting f to S. It turns out that there is a nice method for doing so, called the method of *Lagrange multipliers*, whose validity depends on the following theorem.

Theorem (Lagrange multipliers). With the same notation as above, suppose $\mathbf{a} \in S$ is a local maximum or minimum of f subject to the constraint $g(\mathbf{x}) = c$, and suppose that $\nabla g(\mathbf{a}) \neq \mathbf{0}$. Then there exists $\lambda \in \mathbb{R}$ such that $\nabla f(\mathbf{a}) = \lambda \nabla g(\mathbf{a})$.

The point is that when wanting to optimize f subject to the constraint $g(\mathbf{x}) = c$, we should be looking among the points satisfying an equation of the form $\nabla f(\mathbf{a}) = \lambda \nabla g(\mathbf{a})$ since this theorem says that the maximum/minimum we are looking for must be among such points. So, the strategy to first find the possible points at which this type of equation can hold, and then do something else to determine the nature of these critical points. The scalar λ satisfying the given equation is called a *Lagrange multiplier* for f subject to $g(\mathbf{x}) = c$, and next time we'll say a bit about what this scalar represents. For now, here is the proof of the theorem, which depends on the chain rule and the fact that the derivative of a single variable function at a maximum or minimum is zero. You can find a more "hand-wavy" reason as to why this works in my Math 290-2 lecture notes, which is not really a proof but does give a different source of geometric intuition.

Proof. Let $\mathbf{x}(t)$ be any curve in S passing through $\mathbf{a}$, say $\mathbf{x}(0) = \mathbf{a}$. (Concretely, this means $\mathbf{x} : I \to S$ is a differentiable function defined on an interval I in $\mathbb{R}$, and its image is the curve at which we are looking.) The expression $f(\mathbf{x}(t))$ then gives the value of f at points along this curve. Since $\mathbf{x}(0) = \mathbf{a}$ is a local maximum or minimum of f among all points of S, it is still a local maximum or minimum of f among points on the given curve, so the single variable function $f(\mathbf{x}(t))$ must have derivative zero at $t = 0$. Viewing $t \mapsto f(\mathbf{x}(t))$ as the composition of $\mathbf{x}$ with f, the chain rule gives

$$0 = \frac{d}{dt} f(\mathbf{x}(t))\bigg|_{t=0} = \nabla f(\mathbf{x}(0)) \cdot \mathbf{x}'(0).$$

Now, $\mathbf{x}'(0)$ gives the vector tangent to the given curve at $\mathbf{x}(0) = \mathbf{a}$, so this says that $\nabla f(\mathbf{a})$ is orthogonal to this tangent vector. Since this is true for any curve in S (and hence any tangent vector to S), $\nabla f(\mathbf{a})$ is orthogonal to S itself at $\mathbf{a}$. But we also know that $\nabla g(\mathbf{a})$ is orthogonal to S at $\mathbf{a}$ since S is a level set of the function g. Hence $\nabla f(\mathbf{a})$ and $\nabla g(\mathbf{a})$ both lie on the line perpendicular to S through $\mathbf{a}$, and so must be multiples of each. In particular, since $\nabla g(\mathbf{a}) \neq \mathbf{0}$, $\nabla f(\mathbf{a})$ can be written as a multiple of $\nabla g(\mathbf{a})$ so there exists λ such that $\nabla f(\mathbf{a}) = \lambda \nabla g(\mathbf{a})$ as claimed. □

Back to Warm-Up. Going back to the Warm-Up, the method of Lagrange multipliers says that the point(s) at which $f(x,y) = xy$ attains a maximum or minimum subject to $x^2 + y^2 = 1$ must be among the points satisfying

$$\nabla f(x,y) = \lambda \nabla g(x,y)$$

where $g(x,y) = x^2 + y^2$. This equation is

$$(y,x) = \lambda(2x, 2y),$$

and working with the equations $y = 2\lambda x$ and $x = 2\lambda y$ shows that $x^2 = y^2$, which together with the constraint gives the points $(\pm 1/\sqrt{2}, \pm 1/\sqrt{2})$, which are indeed the points at which the required maxima/minima occur as given in the solution to the Warm-Up problem in my Math 290-2 notes.

Example. Here is a standard type of example. Fix $A > 0$ and say we consider all possible 3-dimenasional rectangular boxes of surface area A. We claim that the largest such box (i.e. the box of surface area A which maximizes volume) is a cube, meaning a rectangular box where the length, width, and height are all the same.

The goal, then, is to maximize the volume function $V(x,y,z) = xyz$ subject to the constraint

$$2(xy + yz + xz) = A,$$

where the left side is the surface area of a rectangular box with dimensions x, y, z. (Assume one corner of the box is at the origin.) In addition, we'll assume x, y, z are all positive since otherwise the volume is zero and we wouldn't really have a 3-dimensional box after all. The method of Lagrange multipliers says that the dimensions which maximize volume are among those satisfying

$$\nabla V(x,y,z) = \lambda \nabla g(x,y,z)$$

where $g(x,y,z) = 2(xy + yz + xz)$. This equality becomes

$$(yz, xz, xy) = \lambda(2y + 2z, 2x + 2z, 2x + 2y),$$

which together with the constraint gives the equations

$$\begin{aligned} yz &= 2\lambda(y+z) \\ xz &= 2\lambda(x+z) \\ xy &= 2\lambda(x+y) \\ 2(xy + yz + xz) &= A. \end{aligned}$$

We now must solve for x, y, z, which usually involves some trial and error. In this case, multiplying the first equation through by x and the second by y gives the same left-hand side, so the resulting right hand sides must be the same:

$$2\lambda x(y+z) = 2\lambda y(x+z).$$

This implies $2\lambda xz = 2\lambda yz$.

Now, $\lambda \neq 0$ since otherwise the first equation above would imply $yz = 0$, which is not possible. Hence since λ and z are nonzero, we get that $x = y$. Similarly, multiplying the third equation above by z and manipulating the result will show that $x = z$ and $y = z$ as well. Thus the dimensions x, y, z which satisfy the Lagrange multiplier equation are the ones where $x = y = z$, and the constraint then gives the explicit values $x = y = z = \sqrt{A/6}$. In particular, this describes a cube as claimed. (Actually, there is one subtle point remaining, which we'll come back to next time. The point is that we cannot yet definitively conclude that the box we have found is one which *maximizes* volume.)

Lecture 4: More on Multipliers

Back to box example. Before getting started we will finish off the final example from last time, which dealt with showing that a rectangular box of fixed surface area and maximal volume must be a cube. Technically what we showed using Lagrange multipliers last time was that *if* there is a box of surface area A with maximal volume, then it must be a cube. Indeed, in general the method of Lagrange multipliers only gives us points which are *candidates* for ones where f is optimized subject to a constraint, but we have to do more to determine the nature of the critical points found. In this case, we have to show that there is actually a box of fixed surface area which maximizes volume; once we have done so, the computation from last time shows that it must be a cube.

Usually we would want to apply something like the Extreme Value Theorem in order to show maximums exist, but the issue here is that the set of points we are considering, i.e. those satisfying the constraint

$$2(xy + yz + xz) = A,$$

is *not* compact (it is closed but not bounded), and so the Extreme Value Theorem does not apply. To get around this fact we proceed as follows. First, from the constraint we get

$$z = \frac{A - 2xy}{2(x + y)}.$$

(Recall that we are assuming x, y, z are all positive, so this denominator is not zero.) Note that we can see from this that the constraint set is unbounded: z can get arbitrarily large as long as x, y are chosen to satisfy the equation above accordingly. Since x, y are positive, the numerator is less than A, so

$$z < \frac{A}{2(x + y)},$$

which says that as $z \to \infty$ we must have $x + y \to 0$ since the fraction on the right must also go to ∞. Since x, y are positive, this means that each of x and y must approach 0 as $z \to \infty$. A similar argument where we start off by solving for x in the constraint equation, or y, shows that as *any* of x, y, z go to ∞ among points satisfying the constraint, the remaining variables must approach 0.

Now, from the constraint we get

$$xy^2 + y^2 z + xyz = \frac{Ay}{2}, \text{ so } xyz = \frac{Ay}{2} - xy^2 - y^2 z.$$

Since x, y, z are all positive, this says that the volume $V(x, y, z) = xyz$ function satisfies

$$0 \leq V(x, y, z) < \frac{Ay}{2}.$$

As either $x \to \infty$ or $z \to \infty$, we saw previously that $y \to 0$, so the term on the right in these inequalities goes to 0 and hence the volume must as well. A similar argument shows that $y \to \infty$ also implies the volume goes to 0. Thus we have that as (x, y, z) gets further away from the origin in $\mathbb{R}^3$ among points satisfying the constraint, the corresponding volume smaller and approaches 0. There there is some large enough ball $B_r(\mathbf{0})$ outside of which the corresponding volume is however small we like, say:

$$V(x, y, z) < \frac{\sqrt{A/6}^3}{2} \text{ for } (x, y, z) \notin B_r(\mathbf{0}).$$

Consider the portion of the constraint set which lies within the ball $B_r(\mathbf{0})$:

$$\{(x, y, z) \in \mathbb{R}^3 \mid 2(xy + yz + xz) = A \text{ and } (x, y, z) \in B_r(\mathbf{0})\}.$$

This set is still closed but now bounded as well, so it is compact. By the Extreme Value Theorem the volume function $V(x, y, z)$ has a maximum value among these points, and the final claim is that this maximum value is the maximum value of V over the entire constraint set. Indeed, the point $(\sqrt{A/6}, \sqrt{A/6}, \sqrt{A/6})$ is in the constraint set and gives a volume of $\sqrt{A/6}^3$, which is larger than the volume given by any point in the constraint set outside the ball $B_r(\mathbf{0})$ since, as we described above, this ball was chosen so that any point outside gave a volume no larger than $\frac{\sqrt{A/6}^3}{2}$. (This is why we used this exact value when describing $B_r(\mathbf{0})$.) Thus, no point outside $B_r(\mathbf{0})$ can possibly give a global maximum for the volume function among points satisfying the constraint, so the maximum attained within the restricted constraint set given above (the portion of the constraint set within $B_r(\mathbf{0})$) must in fact be the global maximum. Hence there is a box of maximal volume and fixed surface area A, which the Lagrange multipliers argument from last time shows must be a cube as required.

No doubt this argument was quite involved, but it is important to realize why it is necessary: Lagrange multipliers does not guarantee maximums/minimums exist, it only characterizes the points at which such extrema can occur *if* they happen to exist.

Warm-Up. Let A be a symmetric $n \times n$ matrix. We wish to find the maximum and minimum values of the quadratic form determined by A among vectors of norm 1, or more concretely, we want to find the global extrema of

$$f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} \text{ subject to the constraint } \|\mathbf{x}\| = 1.$$

Note that such global extrema exist since f is continuous and the set of vectors in $\mathbb{R}^n$ of norm 1 is compact. Actually, we can find these global extrema using linear algebra alone, but first we'll do it using the method of Lagrange multipliers.

The given constraint is the same as $\|\mathbf{x}\|^2 = 1$, or $\mathbf{x}^T\mathbf{x} = 1$. Set $g(x) = \mathbf{x}^T\mathbf{x}$ to be our constraint function. The gradient of f can be found using one of the versions of the product rule we briefly saw last quarter (for instance on the final practice problems), or using a brute-force computation. Denoting the entries of A by a_{ij}, we have

$$\mathbf{x}^T A\mathbf{x} = \sum_{i,j} a_{ij}x_i x_j$$

where $\mathbf{x} = (x_1, \dots, x_n)$. Note that in this summation, each $x_\ell x_k$ and $x_k x_\ell$ are written as separate terms—the first with coefficient $a_{\ell k}$ and the second with coefficient $a_{k\ell}$. We get:

$$\frac{\partial f}{\partial x_k}(\mathbf{x}) = 2a_{kk}x_k + \sum_{\ell \neq k}(a_{k\ell} + a_{\ell k})x_\ell,$$

which since A is symmetric (so $a_{k\ell} = a_{\ell k}$) can be written as

$$\frac{\partial f}{\partial x_k}(\mathbf{x}) = 2\sum_{\ell} a_{k\ell}x_\ell.$$

Thus the gradient of $f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ at $\mathbf{x}$ is:

$$\begin{pmatrix} 2(a_{11}x_1 + \cdots + a_{1n}x_n) \\ \vdots \\ 2(a_{n1}x_1 + \cdots + a_{nn}x_n) \end{pmatrix} = 2\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = 2A\mathbf{x}.$$

(Of course, this is just the higher-dimensional analog of the fact that the single-variable function $f(x) = ax^2$ has derivative $f'(x) = 2ax$.) Viewing $g(\mathbf{x}) = \mathbf{x}^T\mathbf{x}$ as the special case of f where $A = I$, we immediately get that

$$\nabla g(\mathbf{x}) = 2I\mathbf{x} = 2\mathbf{x}.$$

According to the method of Lagrange multipliers, the global extrema of f subject to the given constraint thus satisfy

$$\nabla f(\mathbf{x}) = \lambda \nabla g(\mathbf{x}), \text{ or } 2A\mathbf{x} = \lambda(2\mathbf{x}).$$

This simplifies to $A\mathbf{x} = \lambda\mathbf{x}$, so the conclusion is that the maxima and minima of f among vectors of norm 1 are attained at *eigenvectors* of A. If the maximum occurs at the eigenvector $\mathbf{x}$ of norm 1 with corresponding eigenvalue λ, the maximum value is then

$$f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} = \mathbf{x}^T(\lambda\mathbf{x}) = \lambda(\mathbf{x}^T\mathbf{x}) = \lambda,$$

and if the minimum occurs at the eigenvector $\mathbf{y}$ of norm 1 with corresponding eigenvalue μ, the minimum value is

$$f(\mathbf{y}) = \mathbf{y}^T A\mathbf{y} = \mathbf{y}^T(\mu\mathbf{y}) = \mu(\mathbf{y}^T\mathbf{y}) = \mu.$$

Thus the global maximum value of f among vectors of norm 1 is the largest eigenvalue of A and the minimum value is the smallest eigenvalue of A.

Spectral Theorem revisited. As mentioned previously, the fact that the maximum and minimum values of a quadratic form among vectors of norm 1 are eigenvalues of the associated symmetric matrix and occur at eigenvectors is something we could have derived using linear algebra alone. Orthogonally diagonalizing A leads to the expression

$$f(\mathbf{c}) = \lambda_1 c_1^2 + \cdots + \lambda_n c_n^2,$$

and if $\lambda_{max}, \lambda_{min}$ denote the largest and smallest eigenvalues of A respectively, we have

$$\lambda_{min} = \lambda_{min}(c_1^2 + \cdots + c_n^2) \leq \lambda_1 c_1^2 + \cdots + \lambda_n c_n^2 \leq \lambda_{max}(c_1^2 + \cdots + c_n^2) = \lambda_{max}$$

where $c_1^2 + \cdots + c_n^2$ due to the constraint of considering vectors of norm 1. This shows that any value of f is between λ_{max} and λ_{min}, and evaluating f at corresponding eigenvectors shows that these candidates for global extrema are actually attained.

So, what's the point of doing this using Lagrange multipliers instead of linear algebra? The answer is that the linear-algebraic approach depended on the Spectral Theorem, whereas the Lagrange multipliers approach did not, and in fact this approach actually leads to a *proof* of the Spectral Theorem which is different than how we proved it last quarter. If you go back and review, the proof we gave last quarter depended on knowing that any symmetric matrix has a real eigenvalue, and we argued this by noting that any matrix has a *complex* eigenvalue and then showing that this eigenvalue must in fact be real when the matrix is symmetric. The fact that any matrix has a complex eigenvalue depends on the fact that any polynomial with complex coefficients has a complex root, which is a highly nontrivial fact known as the Fundamental Theorem of Algebra.

Rather than depending on this deep fact in order to prove the Spectral Theorem, here we have shown that any symmetric matrix has a real eigenvalue using calculus instead and avoiding any mention of complex numbers. Once we know that any symmetric matrix has a real eigenvalue, the proof of the rest of the Spectral Theorem proceeds as the one we gave last quarter. In infinite-dimensional settings there is no analog of the fact that polynomials have complex roots, and so proving infinite-dimensional analogs of the Spectral Theorem requires a different way of showing

that eigenvalues exist; using something similar to the method of Lagrange multipliers is the way these infinite-dimensional proofs usually proceed.

Rephrasing Lagrange. The method of Lagrange multipliers can be rephrased as follows. Given the function $f : \mathbb{R}^n \to \mathbb{R}$ to optimize subject to the constraint $g(\mathbf{x}) = c$, define the function $L : \mathbb{R}^{1+n} \to \mathbb{R}$ by

$$L(\lambda, \mathbf{x}) = f(\mathbf{x}) - \lambda(g(\mathbf{x}) - c).$$

Note what we get when determining the critical points of L. We have:

$$\frac{\partial L}{\partial \lambda} = -(g(\mathbf{x}) - c) = -g(\mathbf{x}) + c \quad \text{and} \quad \frac{\partial L}{\partial x_i} = \frac{\partial f}{\partial x_i} - \lambda \frac{\partial g}{\partial x_i},$$

which gives

$$\nabla L(\lambda, \mathbf{x}) = (-g(\mathbf{x}) + c, \nabla f(\mathbf{x}) - \lambda \nabla g(\mathbf{x})).$$

Thus the critical points of L satisfy the equations

$$-g(\mathbf{x}) + c = 0 \quad \text{and} \quad \nabla f(\mathbf{x}) - \lambda \nabla g(\mathbf{x}) = \mathbf{0},$$

which are precisely the constraint equation $g(\mathbf{x}) = c$ and Lagrange multipliers equation $\nabla f = \lambda \nabla g(\mathbf{x})$. The conclusion is that the method of Lagrange multipliers is equivalent to finding the critical points of L.

What do multipliers mean? The above rephrasing of the method of Lagrange multipliers gives a way to interpret what the so-called Lagrange multiplier λ actually represents. Imagine that in the definition of L we consider the scalar c to also be a variable. Then we get something like

$$\frac{\partial L}{\partial c} = \lambda,$$

meaning that λ is some type of rate of change. Indeed, λ can be interpreted as the rate of change of the extrema in the optimization problem in question with respect to the *constraint*, meaning that λ measures how the extrema change as the constraint changes. For instance, in the box problem, λ measures how the dimensions of the box of maximal volume change as the surface area being fixed changes.

This leads to many useful applications, which we don't have time to look at in this class, but which you'll likely see at some point. In particular, anytime you learn about various *marginal* quantities (marginal price, marginal utility, etc.) in economics, there's usually some Lagrange multipliers explanation hiding in the background.

Multiple constraints. Finally, we consider the problem of optimizing a function subject to not only one but *multiple* constraints. This is not something we'll do much with, so I'm only outlining this to illustrate the linear algebra involved. The goal is to optimize $f : \mathbb{R}^n \to \mathbb{R}$ among points satisfying the constraints

$$g_1(\mathbf{x}) = c_1, \ \ldots, \ g_k(\mathbf{x}) = c_k$$

where each g_i is a function $\mathbb{R}^n \to \mathbb{R}$ and each c_i is in $\mathbb{R}$.

To see what to do, I'll have to make some assumptions about various geometric notions, such as the notion of a "tangent space" to a geometric object. (For instance, the tangent space to a curve is a tangent line, a tangent space to a surface is a tangent plane, and so on.) That is to say, this explanation will be quite hand-wavy and a lot of details will be swept under the rug, but this

won't be a big deal for us since we're not going to focus on this method. To make everything we do here completely precise would require knowing more analysis and differential geometry.

Suppose that the extrema of f subject to the given constraints occurs at a point $\mathbf{a}$. We will make technical assumption that $\nabla g_1(\mathbf{a}), \dots, \nabla g_k(\mathbf{a})$ are linearly independent. (We'll soon see why we need this.) The first fact we'll take for granted is that the set of points satisfying the first constraint:

$$\{\mathbf{x} \in \mathbb{R}^n \mid g_1(\mathbf{x}) = c_1\}$$

is an $(n-1)$-dimensional object in $\mathbb{R}^n$. (The technical term is $(n-1)$-dimensional *manifold*.) For instant, the set of points satisfying a single constraint in $\mathbb{R}^2$ is a curve, and the set of points satisfying a single constraint in $\mathbb{R}^3$ is a surface. (This uses the fact that $\nabla g_1(\mathbf{a}) \neq \mathbf{0}$.) Now, it turns out that each additional constraint *reduces* the dimension of the resulting geometric object, so that the set of points satisfying both of the first two constraint is an $(n-2)$-dimensional object, and so on until we get that the set of points satisfying all k constraints:

$$S = \{\mathbf{x} \in \mathbb{R}^n \mid g_i(\mathbf{x}) = c_i \text{ for all } i\}$$

is an $(n-k)$-dimensional object. (This is where the fact that the gradients are linearly independent comes in.) A similar argument as to the proof of Lagrange multipliers in the single constraint case shows that $\nabla f(\mathbf{a})$ is orthogonal to S at $\mathbf{a}$, and so lies in the orthogonal complement to the "tangent space" of S at $\mathbf{a}$:

$$\nabla f(\mathbf{a}) \in (\text{tangent space})^{\perp}.$$

It is a fact that if S is $(n-k)$-dimensional, its tangent space is also $(n-k)$-dimensional (in the linear algebraic sense of dimension), so its orthogonal complement is k-dimensional.

Now, due to properties of gradients we've seen, each $\nabla g_i(\mathbf{a})$ is orthogonal to the constraint set for $g_i(\mathbf{x}) = c_i$. Since S lies in each such constraint set, each $\nabla g_i(\mathbf{a})$ is orthogonal to S at $\mathbf{a}$, and so each $\nabla g_i(\mathbf{a})$ is also orthogonal to the tangent space of S at $\mathbf{a}$:

$$\nabla g_1(\mathbf{a}), \dots, \nabla g_k(\mathbf{a}) \in (\text{tangent space})^{\perp}.$$

However, these here are k linearly independent vectors in a k-dimensional space, so they must span this entire orthogonal complement. We conclude that $\nabla f(\mathbf{a})$, being a vector in this orthogonal complement, must a linear combination of the $\nabla g_i(\mathbf{a})$, so there exist $\lambda_1, \dots, \lambda_k$ such that

$$\nabla f(\mathbf{a}) = \lambda_1 \nabla g_1(\mathbf{a}) + \cdots + \lambda_k \nabla g_k(\mathbf{a}).$$

Thus, the extrema of f subject to the given multiple constraints must be among the points satisfying this equation, which is the analog of $\nabla f = \lambda \nabla g$ in the case of multiple constraints.

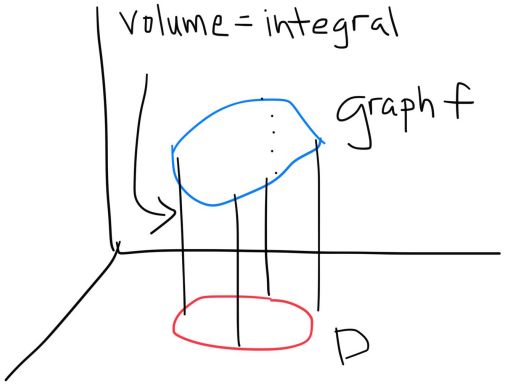
Example. You can find a multiple constraint example of Lagrange multipliers in my Math 290-2 notes. The example deals with finding the point on the line of intersection of two planes which is closest to some given point.

Lecture 5: Riemann Sums

Integration. We now switch gears towards studying multivariable integration, which will be our focus the rest of the quarter. We'll see that, for the most part, multivariable integrals are computed using so-called *iterated integrals*, which essentially amount to computing various single-variable integrals in succession. Thus, all integration techniques from single-variable calculus will

be important. However, before we get to computations, we should have a sense as to what it is we are actually computing, meaning we will spend some time talking about the precise definition of an integral and when such integrals actually exist.

Geometrically multivariable integrals have a simple geometric interpretation, as they do in the single variable case in terms of area. For instance, if $f : \mathbb{R}^2 \to \mathbb{R}$ is a function of two-variables and D is a region in $\mathbb{R}^2$, the integral of f over D will give the volume of the region lying between the graph of f and D:



More generally, for a function $f : \mathbb{R}^n \to \mathbb{R}$ of n variables and a region D in $\mathbb{R}^n$, the integral of f over D will give the n-dimensional volume (higher-dimensional analog of ordinary volume) of the region of $\mathbb{R}^{n+1}$ lying "between" the graph of f and D. Certainly, once we are beyond two variables, "volume" is harder to interpret since we can no longer visualize the corresponding objects, but we'll come to other ways of interpreting higher-dimensional integrals later on, where the main idea is that we should interpret an integral as an analog of summation.

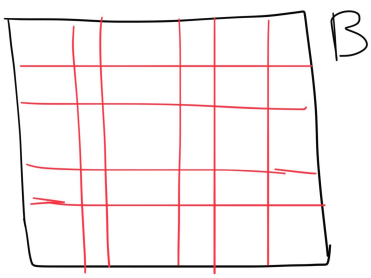
Boxes. We first discuss integration over boxes. (We'll discuss integration over more general regions soon enough.) Given n intervals $[a_1, b_1], \dots, [a_n, b_n]$, the notation $[a_1, b_1] \times \cdots \times [a_n, b_n]$ denotes the *(rectangular) box* in $\mathbb{R}^n$ consisting of points whose i-th coordinate lies in the i-th interval:

$$[a_1, b_1] \times \cdots \times [a_n, b_n] = \{(x_1, \dots, x_n) \mid x_i \in [a_i, b_i]\}.$$

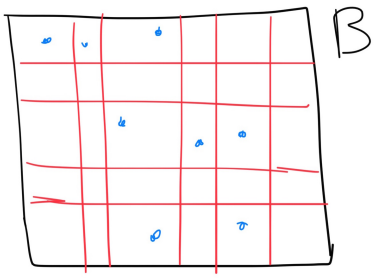
For instance, $[a, b] \times [c, d]$ denotes the rectangle in $\mathbb{R}^2$ consisting of points (x, y) where $a \leq x \leq b$ and $c \leq y \leq d$, and $[a, b] \times [c, d] \times [g, h]$ denotes what you normally think of as being a rectangular box in $\mathbb{R}^3$. We use the term "box" to refer to the analogous object in $\mathbb{R}^n$ as well.

Riemann sums. As in the single-variable case, the idea we use to compute the higher-dimensional volume an integral is meant to represent is that of approximating this volume using boxes (rectangles in the single-variable case), and then taking some kind of limit.

Let B be a box. A *partition* P of B is a division of B into smaller boxes B_i:



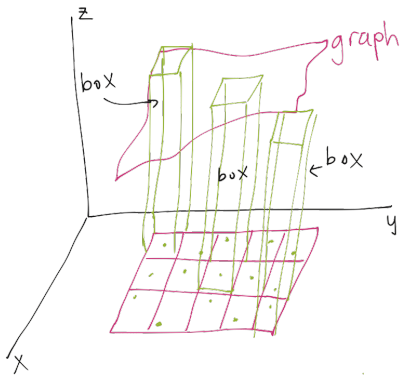
Given such a partition, we choose *sample points* $\mathbf{c}_i$ from each smaller box:



We define the *Riemann sum* of $f : B \to \mathbb{R}$ corresponding to the partition P and sample points $\mathbf{c}_i$ to be the sum:

$$R(f, P, \mathbf{c}_i) = \sum_i f(\mathbf{c}_i) \operatorname{Vol}(B_i)$$

which ranges over all smaller boxes. Visually, the quantity $f(\mathbf{c}_i) \operatorname{Vol}(B_i)$ gives the $(n+1)$-dimensional volume of the $(n+1)$-dimensional box whose base is the n-dimensional box B_i and whose height is given by the value of f at the sample point $\mathbf{c}_i$:



The Riemann sum in question adds together all of these individual volumes, and such a Riemann sum thus provides an approximation to the volume of the region between the graph of f and B. The idea then is that, as the sizes of the boxes B_i get smaller and smaller, the corresponding Riemann sums provide a better and better approximation to the volume we want.

Integrability. Thus, we define the *integral* of f over B as the limit of these Riemann sums as the sizes of the boxes B_i get smaller and smaller:

$$\int_B f(\mathbf{x})\, d\mathbf{x} := \lim_{\text{all } \Delta x_i \to 0} R(f, P, \mathbf{c}_i),$$

at least in the case where this limit actually exists; when it does, we say that f is *integrable* over B. The notation Δx_i denotes the length of one of the sides of the smaller box B_i, and so saying "all $\Delta x_i \to 0$" in this definition is saying that all of these lengths are getting smaller and smaller, so the volumes of the corresponding small boxes are also getting smaller and smaller.

Now, this is often how integrals are indeed presented, but in actuality this definition is far from being precise. For one thing, even though we might have intuitive notion as to what it means to take a "limit" as the sizes of the boxes get "smaller and smaller", this approach is quite vague and

not very rigorous at all. Even worse, the fact is that this limit should approach a fixed value *no matter* the sample points we use, and this subtlely isn't really reflected in this limit "definition" we gave. Indeed, the ability to change our sample points at will and still get the same limit is really what allows integration to actually work as we expect it to.

For our purposes, the "definition" we gave above will be good enough, and indeed this is the way in which integrals are defined in our book. However, I would be remiss if I didn't state the actual precise definition of the integral in this class, as it is something everyone should see at least once in their lives. The issue is that this definition is hard to work with, and indeed is better dealt with in a course in real analysis. For this reason we won't use this definition at all and will stick with the vague "limit of Riemann sums" definition above, but nonetheless here you go:

> Given a box B in $\mathbb{R}^n$ and a function $f : B \to \mathbb{R}$, we say that f is *integrable* over B if there exists a number I such that for all $\epsilon > 0$, there exists $\delta > 0$ such that for any partition P of B with $\text{mesh}\, P < \delta$ and any collection of sample points $\mathbf{c}_i$, we have $|R(f, P, \mathbf{c}_i) - I| < \epsilon$. When f is integrable over B, we call the number I the *integral* of f over B and denote it by $\int_B f(\mathbf{x})\, d\mathbf{x}$.

Let's try to understand this definition a bit. First, the *mesh* of a partition is the largest length of any edge among the smaller boxes determined by that partition, so in the notation used above this is the largest of the Δx_i quantities. If we think of δ as being small, then to say that $\text{mesh}\, P < \delta$ means that we only consider partitions where the small boxes are "small enough", which is a precise way of phrasing the idea that these small boxes are getting "smaller and smaller". The number I in the definition is the actual numerical value the integral of f over B should have, so the definition says that given some (small) measure ϵ for how close we want a Riemann sum to be to the actual value of the integral, there is a small enough δ which guarantees that for any "fine" enough partition P (i.e. ones which mesh smaller than δ), any corresponding Riemann sum for any possible sample points will indeed be within ϵ away from I, as stated by the inequality $|R(f, P, \mathbf{c}_i) - I| < \epsilon$. Thus, as ϵ gets smaller and smaller, and in turn δ gets smaller and smaller, the Riemann sums in question (with smaller and smaller mesh) are providing better and better approximations to I.

Riemann-Lebesgue. Thus, the definition given above makes precise the idea of taking a "limit of Riemann sums", but as stated earlier for us the informal notion of "limit of Riemann sums" will be good enough. Nonetheless, we will highlight one fact, namely that the notion of being "integrable" can be rephrased in yet another way, which is often much simpler to work with than the precise definition given above. In particular, a main observation is that not all functions are integrable, meaning that the integral of f over B may not even exist depending on what f actually is. In general there are functions so that the region "under" their graphs does not have a well-defined volume, so for such functions integrals do not exist. This might seem strange coming from a single-variable calculus course, where the existence of integrals was likely never an issue, but is something we should say more about in this course. (The reason why this was likely not an issue in your single-variable calculus course is that you very probably only dealt with continuous functions in that course, and it turns out, as we'll see, that continuous functions are always integrable.)

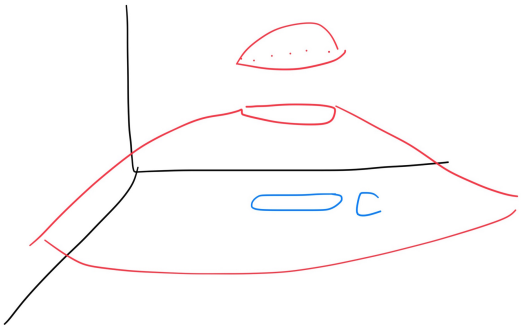
The nice thing is that there is an equivalent characterization as to when an integral actually exists, which is not-so-hard to work with in practice. Here is the statement of the so-called *Riemann-Lebesgue Theorem*

> A function $f : B \to \mathbb{R}$ is integrable if and only if it is bounded and the set of points in B at which f is discontinuous has measure zero in $\mathbb{R}^n$.

Thus, to know that a function is integrable, all we have to know is that it is bounded and that the set of points where it is not continuous has measure zero, whatever that means. We'll come back to the notion of "measure zero" next time, but intuitively this just means that the set of points at which f is not continuous should have "zero volume" in $\mathbb{R}^n$.

Note that, in the end, this theorem only gives us an alternate way of testing whether or not an integral exists, but says nothing about what value an integral should actually have. This is fine, since as stated in the beginning, we'll come to talking about how to compute integrals using "iterated integrals" later on; the point now is that, before we can even talk about how to compute integrals, we should know that these integrals actually exist, and the Riemann-Lebesgue Theorem gives us the simplest way of showing this.

Example. Consider a function $f : \mathbb{R}^2 \to \mathbb{R}$ whose graph looks something like:



Thus, the graph has the shape of a mountain, only that the top portion of the mountain has been shifted up so that mountain is not "continuous" throughout. To be precise, f is continuous everywhere except for at points along the curve C drawn in the xy-plane, which are the points at which the graph "jumps" up to a higher height. The function as drawn is clearly bounded, and so whether or not the integral of f exists over a box depends on the set of points at which f is not continuous. If B is a rectangle large enough to contain all of C, then C itself is the set of points at which f is not continuous. Intuitively, the curve C has zero area in $\mathbb{R}^2$, so the Riemann-Lebesgue Theorem will say that the integral of f over B indeed exists.

Note that to say C has zero area is different from saying that the region it encloses has zero area; indeed, the region enclosed by C has positive area, but the point is that C itself only has a length but not an area, essentially since C, being a curve, is only a 1-dimensional object in $\mathbb{R}^2$. Again, we'll give some more precise statements next time. Also note that intuitively the integral of f over B *should* exist, since the region under the graph of f and above B seems to have a well-defined volume, which comes from taking the volume under the "mountain", including the volume of the entire region underneath the "snow cap" which is shifted up.

Lecture 6: Integrability

Warm-Up. We show that *constant* functions are always integrable over any box B in $\mathbb{R}^n$ using the Riemann sum definition. Fix a constant M and let $f : B \to \mathbb{R}$ be the function defined by $f(\mathbf{x}) = M$ for all $\mathbf{x} \in B$. Given any partition P of B and any collection $\mathbf{c}_i$ of sample points, the corresponding Riemann sum looks like:

$$R(f, P, \mathbf{c}_i) = \sum_i f(\mathbf{c}_i) \operatorname{Vol}(B_i).$$

In this case, since f is constant, the value $f(\mathbf{c}_i)$ is always M, so the Riemann sum becomes:

$$R(f, P, \mathbf{c}_i) = \sum_i M \operatorname{Vol}(B_i) = M \sum_i \operatorname{Vol}(B_i).$$

But the boxes B_i all together fill out the original box B, so the volumes of the B_i add up to the volume of B and hence

$$R(f, P, \mathbf{c}_i) = M \operatorname{Vol}(B).$$

Thus *every* possible Riemann sum is equal to $M \operatorname{Vol}(B)$, so the limit of these Riemann sums as the sides of the boxes approach 0 is also equal to $M \operatorname{Vol}(B)$. Since this limit exists, f is integrable over B and

$$\int_B f(\mathbf{x})\, d\mathbf{x} = M \operatorname{Vol}(B).$$

This is the higher-dimensional analog of the fact that $\int_a^b c\, dx = c(b-a)$ in single variable calculus.

Measure zero. Recall the statement of the Riemann-Lebesgue Theorem from last time: $f : B \to \mathbb{R}$ is integrable if and only if f is bounded over B and the set of points in B at which f is not continuous has measure zero. We now define what "measure zero" means:

> Let $S \subseteq \mathbb{R}^n$. We say that S has *measure zero* in $\mathbb{R}^n$ if for every $\epsilon > 0$ there exists a (possibly infinite) collection of boxes $B_1, B_2, B_3, \dots$ such that
>
> $$S \subseteq B_1 \cup B_2 \cup B_3 \cup \cdots \quad \text{and} \quad \sum_i \operatorname{Vol}(B_i) < \epsilon.$$
>
> Often times we phrase the first condition as saying that the boxes $B_1, B_2, \dots$ *cover* S.

Let us digest this definition. The intuition is that sets of measure zero are ones which have "volume zero" in $\mathbb{R}^n$. Indeed, if boxes $B_1, B_2, \dots$ cover S, it should be true that the volume of S (if such a thing is defined) should be less than or equal to the sum of the volumes of the B_i:

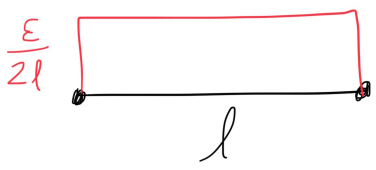
$$\operatorname{Vol}(S) \leq \sum_i \operatorname{Vol}(B_i),$$

since S is "smaller" than the union of all the B_i, and this union cannot have larger volume than the sum of the individual volumes of the B_i. The definition above says that given any positive number whatsoever, no matter how small, we can always find boxes which cover S and which all together have "total volume" smaller than ϵ. This then implies that

$$\text{for any } \epsilon > 0,\ \operatorname{Vol}(S) < \epsilon,$$

so $\operatorname{Vol}(S)$ should be a nonnegative number which is smaller than every positive number; the only such number is zero, so $\operatorname{Vol}(S)$ should be zero.

Example. Here is a simple example just to see how this definition works. Take any horizontal line segment in $\mathbb{R}^2$, say of length ℓ. We claim that this line segment has measure zero in $\mathbb{R}^2$. Indeed, for any $\epsilon > 0$, let B be the box which has base equal to this line segment and height $\frac{\epsilon}{2\ell}$:

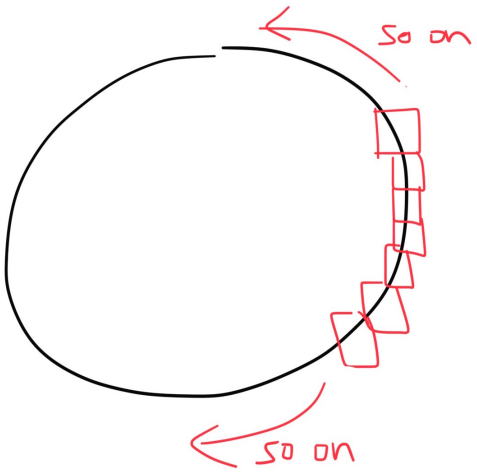


This single box alone covers the line segment in question, and

$$\text{Vol}(B) = \ell \frac{\epsilon}{2\ell} = \frac{\epsilon}{2} < \epsilon.$$

(In this case, Vol really means "area".) Hence, for any $\epsilon > 0$ we have found boxes (just one in this case) covering the line segment with total volume (i.e. area) smaller than ϵ, so the line segment has measure zero. This makes sense, since a line segment in $\mathbb{R}^2$ should indeed have no area.

C^1 curves have measure zero. Consider the unit circle in $\mathbb{R}^2$. We claim that this also measure zero. As mentioned last time, note that this means the circle itself has zero area, *not* that the region *enclosed* by the circle has zero area. (Of course, this region has area π.) Visually it makes sense that you can cover the circle with boxes of ever decreasing total area:



which intuitively implies the circle should have no area.

Giving a precise proof of this using the definition of measure zero requires some work and is something you might do in an analysis course. Instead, we make use of the following fact, which is also something you might prove in an analysis course. First a definition:

> A C^1 curve in $\mathbb{R}^n$ is a C^1 function $\gamma : I \to \mathbb{R}^n$ defined on some interval I in $\mathbb{R}$. Such a function has components $\gamma(t) = (\gamma_1(t), \dots, \gamma_n(t))$, and the point is that as t varies the points described by $\gamma(t)$ trace out the curve in question. (Normally, the image of γ in $\mathbb{R}^n$ is though of as being *the* curve.)

The fact is that a C^1 curve in $\mathbb{R}^n$ always has measure zero. In the case of the unit circle, $\gamma(t) = (\cos t, \sin t)$ for $t \in [0, 2\pi]$, and since $\cos t, \sin t$ are both C^1, the unit circle is a C^1 curve, so it has measure zero in $\mathbb{R}^2$.

Measure zero is a "relative" notion. Just one point of clarification: to say that a set has measure zero depends on the space we are considering that set to sit inside of. For instance, a square does *not* have measure zero in $\mathbb{R}^2$ since it has positive area, but it *does* have measure zero in $\mathbb{R}^3$ (or $\mathbb{R}^n$ for $n \geq 3$) since it has zero volume. Similarly, a solid region in $\mathbb{R}^3$ might have positive 3-dimensional volume, but considered as a subset of $\mathbb{R}^4$ it would have measure zero. This won't be a major issue for us, but is worth mentioning.

Integrability and measure zero. Finally, we give some sense as to why this measure zero condition on the set of discontinuity points is related to integrability. The intuition is that functions which are discontinuous only on a set of measure zero are pretty "close" to being continuous, since the set of points at which they are not continuous are in a sense negligible.

Here is a true fact. Suppose $f, g : B \to \mathbb{R}$ are both integrable and agree everywhere except for on a set of measure zero, meaning that there exists $S \subseteq B$ of measure zero such that

$$f(\mathbf{x}) = g(\mathbf{x}) \text{ for all } \mathbf{x} \notin S.$$

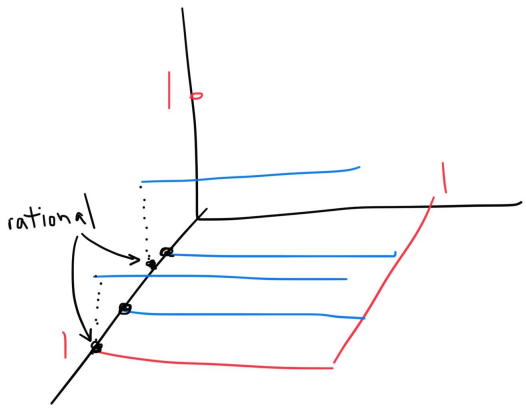
Then $\int_B f(\mathbf{x})\, d\mathbf{x} = \int_B g(\mathbf{x})\, d\mathbf{x}$. Thus, what happens on a set of measure zero can never affect the value of an integral, which is the sense in which sets of measure zero are "negligible".

Now, if there is any justice in the world, continuous functions should always be integrable, essentially because the regions between their graphs and the region of integration should always have a well-defined volume. Thus, a function which agrees with a continuous function except for on a set of measure zero "should" itself be integrable, since what happens on this measure zero set where the functions differ cannot affect the value of the integral. (This is not literally true as stated for the integral we have defined, but is true for a more general type of integral known as the Lebesgue integral.) The Riemann-Lebesgue Theorem says indeed that an integrable function is precisely one which fails to be continuous only on a set of measure zero. The "bounded" condition in the Riemann-Lebesgue Theorem is expected since continuous functions on boxes are always bounded by the Extreme Value Theorem.

Example. Define $f : \mathbb{R}^2 \to \mathbb{R}$ by

$$f(x, y) = \begin{cases} 1 & x \text{ is rational} \\ 0 & x \text{ is irrational.} \end{cases}$$

To be clear, a rational number is a real number which can be expressed as the quotient of integers, and an irrational number is one which cannot. We show that f is not integrable over $[0, 1] \times [0, 1]$. Indeed, the graph of f consists of lines at a height of 1 at any x which is rational, and lines at a height of 0 at any x which is irrational:



It is a basic property of rational and irrational numbers (basic meaning standard, but not as in something you should already know; indeed, this is likely the first you've heard this fact) that between any real numbers there exist infinitely many rationals and infinitely many irrationals. This leads to the fact that f is discontinuous *everywhere*, because its graph constantly "jumps"

up and down between height 1 and height 0. Thus the discontinuity set of f in $[0,1] \times [0,1]$ is $[0,1] \times [0,1]$ itself, which does not have measure zero. Hence f is not integrable over this square.

Alternatively, we can also see this using Riemann sums. Given any partition of $[0,1] \times [0,1]$, no matter how fine, picking samples points which always have rational x coordinate give a Riemann sum value of

$$\sum_i f(\mathbf{c}_i) \operatorname{Vol}(B_i) = \sum_i \operatorname{Vol}(B_i) = \operatorname{Vol}([0,1] \times [0,1]) = 1$$

since f evaluated at any sample point is 1 in this case. (As usual, Vol here really means area.) However, picking sample points which always have irrational x coordinate give

$$\sum_i f(\mathbf{c}_i) \operatorname{Vol}(B_i) = \sum_i 0 \operatorname{Vol}(B_i) = 0$$

since f evaluated at any sample point is 0 now. Thus the limit of these Riemann sums does not exist since the answer depends on the type of sample point chosen, so f is not integrable.

Lecture 7: Iterated Integrals

Warm-Up. Define $f : \mathbb{R}^3 \to \mathbb{R}$ by

$$f(x,y,z) = \begin{cases} xyz & x^2 + y^2 \neq z^2 \\ 10 & x^2 + y^2 = z^2. \end{cases}$$

We show that f is integrable over the 3-dimensional box $[0,1] \times [0,2] \times [0,3]$. First, the function $g(x,y,z) = xyz$ is continuous over this box, so since this box is compact g is bounded by the Extreme Value Theorem. The function f in question only differs from this bounded function by changing its value at some points to be 10 instead, but this change does not affect boundedness since 10 is a constant. Thus f is itself bounded over this box. (Note that f is not continuous over this box, which is why we cannot simply apply the Extreme Value Theorem to f alone.)

Now, since $g(x,y,z) = xyz$ is continuous everywhere, f will be discontinuous only at the points where it differs from g, which are the points on the surface $x^2 + y^2 = z^2$, which describes a double cone. Thus the discontinuity set of f consists of the portion of this cone within the box $[0,1] \times [0,2] \times [0,3]$. This portion is still a 2-dimensional surface, so it has zero volume and hence measure zero. (To be precise you can draw small boxes covering the cone of every decreasing volume. More generally, any C^1 *surface* in $\mathbb{R}^3$ has measure zero; we'll define what this means later when we talk about surfaces in more detail.) We conclude that f is integrable over this box as claimed. Moreover, since what happens over a set of measure zero does not affect the value of an integral, and f agrees with $g(x,y,z) = xyz$ everywhere except for on a set of measure zero, the integral of f over this box is the same as that of $g(x,y,z) = xyz$, which we'll compute below.

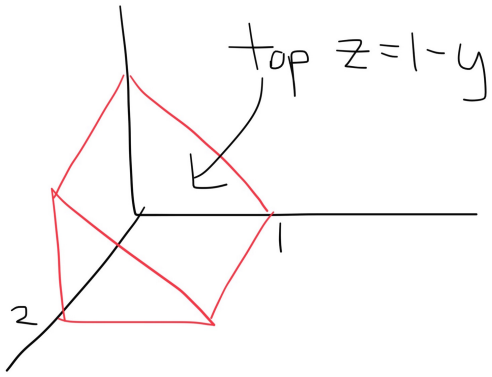
Example. We have spent some time talking about deciding whether or not an integral exists, and now we come to the question of actually computing integrals. Say we want to compute

$$\iint_{[0,2]\times[0,1]} (1-y)\, dA.$$

In our previous notation, this would have been denoted

$$\int_{[0,2]\times[0,1]} (1-y)\, d\mathbf{x},$$

but in the special case where we are integrating a function of two variables over a region in $\mathbb{R}^2$, we use the *double integral* notation above. (The "dA" is purely a notational device, although it helps to think of it as an "infinitesimal area"; we'll come back to this perspective later.) This integral should give the volume of the region in $\mathbb{R}^3$ between the graph of f and the rectangle $[0,2] \times [0,1]$ in the xy-plane. The graph of f is a plane, so this region looks like:



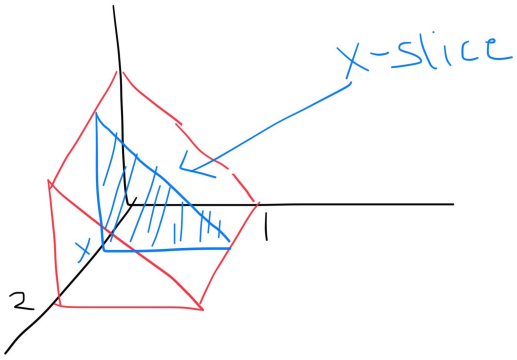
Viewing this region has making up half of the rectangular box $[0,2] \times [0,1] \times [0,1]$, we see that its volume should be half the volume of this box, so it is 1. Thus

$$\iint_{[0,2]\times[0,1]} (1-y)\, dA = 1.$$

Computing integrals via slices. Continuing with the same example, we can alternatively argue as follows. The region whose volume we want is obtained by taking the triangle forming the back side in the yz-plane and sliding it out forward a distance of 2 in the x-direction. Thus, the volume should be

$$(\text{area of the triangle})(\text{length in } x\text{-direction}) = \frac{1}{2} \cdot = 1,$$

where the area of the triangle is $\frac{1}{2}$ times base times height. This works because the *x-slice* of this region at any value of x ("slice" means the portion of the region which lies at a given value, or the intersection of the region with a plane) is always a triangle of the same area as the one in back:



We can interpret this computation as "adding" up the areas of all such x-slices as x varies from 0 to 2, where this "addition" is carried out by the single-variable integral:

$$\int_0^2 (\text{area of } x\text{-slice})\, dx.$$

Now, the area of any x-slice can in turn be computed via a single variable integral, only now taken with respect to y. Indeed, at a fixed value of x, the corresponding x-slice is the 2-dimensional region under the graph of $f(x, y) = 1 - y$ and above the segment $[0, 1]$ in the y-direction, so

$$\text{area of } x\text{-slice } = \int_0^1 (1 - y)\, dy.$$

The fact that this is independent of x reflects the fact that all x-slices in this example have the same area. Thus making this substitution into the previous formula gives:

$$\iint_{[0,1]\times[0,2]} (1 - y)\, dA = \int_0^2 \left(\int_0^1 (1 - y)\, dy \right) dx,$$

which says that the given double integral can be computed via a succession of two separate integrations: the "inner" one taken with respect to y and the "outer" one taken with respect to x. Normally, the parentheses are dropped from the notation and we would simply write this as :

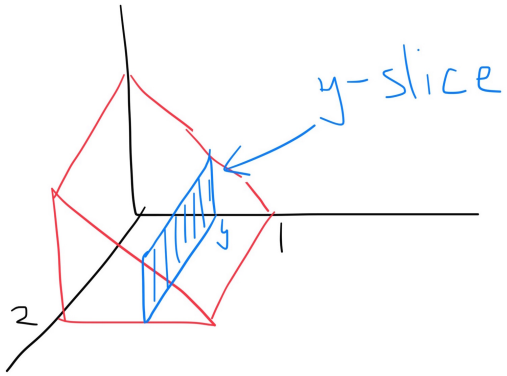
$$\iint_{[0,1]\times[0,2]} (1 - y)\, dA = \int_0^2 \int_0^1 (1 - y)\, dy\, dx.$$

The expression on the right is called an *iterated integral* since it consists of multiple single-variable integrals, one taken after another. To compute such an iterated integral, we simple perform each integration, starting with the most inner one and working our way out. The only thing to keep in mind is that when integrating with respect to a certain variable, all other variables are treated as constants. Thus:

$$\int_0^2 \int_0^1 (1 - y)\, dy\, dx = \int_0^2 \left(y - \frac{1}{2}y^2 \right)\Big|_0^1 dx = \int_0^2 \frac{1}{2}\, dx = 1,$$

which agrees with the previous volume we found for the double integral in question.

Of course, there was nothing special about adding up the areas of the x-slices, and we should get the same result by instead considering slices in the y direction. The slice at a specific y looks like a rectangle



whose height is given by $1 - y$, which remains the same as x varies form 0 to 2. Thus, the volume in question should equal

$$\iint_{[0,1]\times[0,2]} (1 - y)\, dA = \int_0^2 (\text{area of } y\text{-slice})\, dy = \int_0^2 \int_0^1 (1 - y)\, dx\, dy,$$

where the inner integral $\int_0^1 (1-y)\,dx$ gives the area of the slice at a fixed y. Computing this iterated integral gives

$$\int_0^2 \int_0^1 (1-y)\,dx\,dy = \int_0^2 x(1-y)\Big|_0^1 dy = \int_0^2 (1-y)\,dy = 1,$$

which also agrees with the values found previously. The idea that you can compute volumes by "adding" up areas of slices is known as *Cavalieri's principle*.

Iterated integrals. Given a function $f : B \to \mathbb{R}$ defined over a box $B = [a_1, b_1] \times \cdots \times [a_n, b_n]$ in $\mathbb{R}^n$, the *iterated integrals* of f over B are the expressions

$$\int_{a_{i_n}}^{b_{i_n}} \int_{a_{i_{n-1}}}^{b_{i_{n-1}}} \cdots \int_{a_{i_1}}^{b_{i_1}} f(x_1, \ldots, x_n)\,dx_{i_1} \ldots dx_{i_{n-1}}\,dx_{i_n},$$

where $x_{i_1}, \ldots, x_{i_n}$ correspond to some ordering of the variables $x_1, \ldots, x_n$. In total, an n-variable function will have $n!$ possible iterated integrals. To compute an iterated integral, we first compute the innermost integral, then the next one, then the next one, and so on, at each step treating as constant all variables not being integrated with respect to. Note that the bounds on each single-variable integral in this expression matches up with the variable being integrated with respect to at that step.

Fubini's Theorem. As mentioned previously, Cavalieri's principle suggests that we can compute multivariable integrals by adding up lower-dimensional volumes of slices, which makes sense intuitively in the example we went through above. However, not all functions are as simple or nice as the one we used there, and so we really have to think about whether all types of integrals can indeed be computed via iterated integrals. A similar geometric picture as the one above suggests this should be true for continuous (i.e. "nice") functions at least, but in general there is no reason to automatically suspect this is so; after all, n-dimensional integrals are technically defined as limits of n-dimensional Riemann sums, whereas iterated integrals—if you dig deep into their definition—are defined as "limits of 1-dimensional Riemann sums of limits of 1-dimensional Riemann sums of limits of 1-dimensional Riemann sums..." and so on, which is different type of beast.

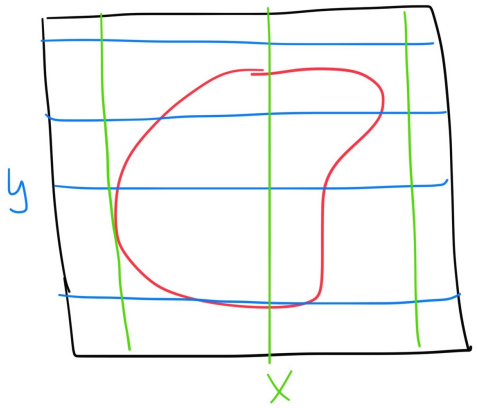
Nonetheless, multivariable integrals can for most purposes indeed be computed via iterated integrals, as long as certain technical assumptions are met. This is the content of *Fubini's Theorem*:

> Suppose $f : B \to \mathbb{R}^n$ is integrable over a box $B = [a_1, b_1] \times \cdots \times [a_n, b_n]$ in $\mathbb{R}^n$, and suppose further that for each $(p_1, \ldots, p_n) \in B$ and each i, the single-variable function $f(p_1, \ldots, x_i, \ldots, p_n)$ obtained by only varying the i-th coordinate while holding all other coordinates fixed is integrable over the interval $[a_i, b_i]$. Then all iterated integrals of f over B exist and
>
> $$\int_B f(\mathbf{x})\,d\mathbf{x} = \int_{a_{i_n}}^{b_{i_n}} \int_{a_{i_{n-1}}}^{b_{i_{n-1}}} \cdots \int_{a_{i_1}}^{b_{i_1}} f(x_1, \ldots, x_n)\,dx_{i_1} \ldots dx_{i_{n-1}}\,dx_{i_n}$$
>
> for any ordering $x_{i_1}, \ldots, x_{i_n}$ of $x_1, \ldots, x_n$.

In other words, when the integral of f over B exists and all iterated integrals exist, then any iterated integral gives the same value and this value is equal to the integral of f over B. The point is that, in general, it is possible for multivariable integrals and iterated integrals to behave rather differently: it's possible for a multivariable integral to exist where none of the iterated integrals exist, it's possible for iterated integrals to exist but give different values depending on the order of

integration, it's possible for some iterated integrals to exist but not others, etc. So, we need Fubini's Theorem to guarantee that iterated integrals give a valid way of computing multivariable integrals, as long as all integrals in question exist. When f is continuous throughout B, all the assumptions in Fubini's Theorem automatically hold, and so whether or not iterated integrals give the value of the multivariable integral is not something we'll have to worry about when f is continuous.

The version of Fubini's Theorem stated above is more general than the version stated in the book. For one thing, the book only discusses the two-variable version of Fubini's Theorem, but even in that case our version is more general. The book's version uses the assumption that "any line in B parallel to one of the axes intersects the discontinuity set of f only finitely often". To why this is a special case of our version, consider a function f over a rectangle B which is discontinuous over a curve like the one below:



For a fixed $y \in [c, d]$, the horizontal line at this value intersects the discontinuity curve in either 0 points, 1 point, or 2 points depending on where y actually is, but in case this intersection is always finite. This means that the single variable function $f(x, y)$ obtained by fixing y and only varying x through the interval $[a, b]$ is discontinuous only at finitely many points in this interval; since a finite subset of $[a, b]$ has measure zero in $\mathbb{R}$, this means that this single variable function is integrable over $[a, b]$, so that

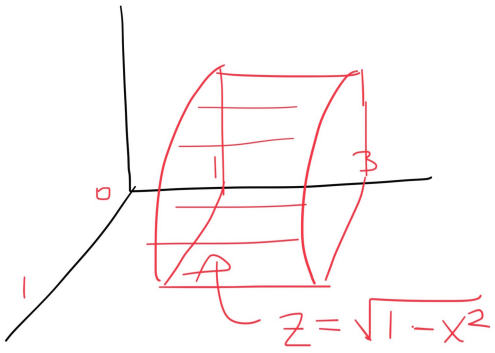
$$\int_a^b f(x, y)\, dx$$

exists for any $y \in [c, d]$. Similarly, for a fixed $x \in [a, b]$, the vertical line at this value intersects the discontinuity set of f in only finitely many points, so the single variable function $f(x, y)$ obtained by fixing x and varying y throughout $[c, d]$ only has finitely many discontinuities, and so is integrable over $[c, d]$ since finite many points in $[c, d]$ have measure zero in $\mathbb{R}$. This says that such a function satisfies the assumptions in our version of Fubini's Theorem. The upshot is that the assumptions in the book's version of Fubini's Theorem are simple ways of implying that the iterated integrals in question exist, but not the only way of implying they exist. Still, for most purposes, the book's version will be enough; in particular, when working with continuous functions both versions of Fubini's Theorem say the same thing.

Example. The double integral

$$\iint_{[0,1]\times[1,3]} \sqrt{1 - x^2}\, dA$$

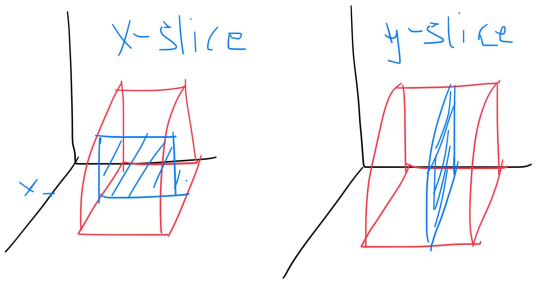
gives the volume of the region under the surface $z = \sqrt{1 - x^2}$ and above the rectangle $[0, 1] \times [1, 3]$, which looks like a portion of a cylinder:



This volume is the area of the left side (which is a quarter of a unit disk) times the length, so it is $\frac{\pi}{2}$. The function $f(x,y) = \sqrt{1-x^2}$ is continuous, so Fubini's Theorem applies and gives that the integral in question is equal to either iterated integral:

$$\iint_{[0,1]\times[1,3]} \sqrt{1-x^2}\,dA = \int_1^3 \int_0^1 \sqrt{1-x^2}\,dx\,dy = \int_0^1 \int_1^3 \sqrt{1-x^2}\,dy\,dx.$$

To be clear, the inner integral $\int_0^1 \sqrt{1-x^2}\,dx$ of the first iterated integral gives the area of the y-slice at a fixed y, and the inner integral $\int_1^3 \sqrt{1-x^2}\,dy$ of the second gives the area of the x-slice at a fixed x:



Computing either of these iterated integrals would require trigonometric substitutions, but this is unnecessary since we know that the values should be $\frac{\pi}{2}$.

Back to Warm-Up. Finally we compute the integral in the Warm-Up. As mentioned there, since the given function f and $g(x,y,z) = xyz$ agree except for on a set of measure zero, the integral of f over the given box should be the same as that of g. Since g is continuous, Fubini's Theorem applies to give:

$$\iiint_{[0,1]\times[0,2]\times[0,3]} g(x,y,z)\,dV = \int_0^3 \int_0^2 \int_0^1 xyz\,dx\,dy\,dz.$$

Computing this iterated integral involves three separate single-variable integral computations:

$$\int_0^3 \int_0^2 \int_0^1 xyz\,dx\,dy\,dz = \int_0^3 \int_0^2 \frac{1}{2}yz\,dy\,dz = \int_0^3 z\,dz = \frac{9}{2},$$

so the integral of f over $[0,1]\times[0,2]\times[0,3]4$ is $\frac{9}{2}$ as well.

Lecture 8: Double Integrals

Warm-Up (in-class version). Define $f : \mathbb{R}^2 \to \mathbb{R}$ by

$$f(x,y) = \begin{cases} x + xy & x \leq 1 \\ 0 & x > 1. \end{cases}$$

We compute the integral of f over the rectangle $[0,2] \times [0,1]$. (If you recall, this was the function we looked at in class, where it was pointed out that the book's version of Fubini's Theorem does not actually apply. However, the more general version of Fubini's Theorem stated in the previous lecture does apply, and so we'll make use of that. Afterwards we'll look at a modified version of this function, and show how to use the book's version of Fubini's Theorem in that case.)

First, f is continuous everywhere except for on the line $x = 1$. Since this has measure zero and f is bounded on $[0,2] \times [0,1]$ (indeed, the function $g(x,y) = x + xy$ is bounded on this rectangle by the Extreme Value Theorem, and changing the value of g to be zero along $x = 1$ does not affect boundedness), f is indeed integrable over the given rectangle. Now we check that (the more general version of) Fubini's Theorem applies. For any $b \in [0,1]$, the single-variable function $f(x,b)$ obtained by fixing $y = b$ and varying x is discontinuous at only one point (namely the point where the line segment $0 \leq x \leq 2$ at $y = b$ intersects the discontinuity set $x = 1$ of the two-variable function f), so since this intersection has measure zero in the line segment $0 \leq x \leq 2$ at $y = b$, the single variable function $f(x,b)$ is integrable on $0 \leq x \leq 2$. For $0 \leq a \leq 1$, the single-variable function $f(a,y)$ looks like

$$f(a,y) = a + ay \text{ for all } 0 \leq y \leq 1,$$

which is continuous on $0 \leq y \leq 1$ and hence integrable on this interval as well. For $1 < a \leq 2$, the single-variable function $f(a,y)$ is identically zero and hence continuous on $0 \leq y \leq 1$, so this is also integrable. Hence the single-variable functions obtained from f by holding one variable constant are always integrable, so our version of Fubini's Theorem applies. (The book's version of Fubini's Theorem does not apply since the line segment at $x = 1$ intersects the discontinuity set of the two-variable function f infinitely often since $x = 1$ *is* the discontinuity set of f.)

Since Fubini's Theorem applies, we can compute the integral of f over $[0,2] \times [0,1]$ using iterated integrals. Here are two approaches. First, using a general result similar to that of Problem 2 on Homework 2, we can split up the region of integration in question and integrate over each piece separately; that is, splitting the rectangle $[0,2] \times [0,1]$ into the rectangles $[0,1] \times [0,1]$ and $[1,2] \times [0,1]$ gives:

$$\iint_{[0,2]\times[0,1]} f(x,y)\, dA = \iint_{[0,1]\times[0,1]} f(x,y)\, dA + \iint_{[1,2]\times[0,1]} f(x,y)\, dA.$$

Over $[1,2] \times [0,1]$ the function f is identically zero except for on a set of measure zero (at $x = 1$), so the integral of f over this portion is the same as the integral of the zero function, so is zero. Thus:

$$\iint_{[0,2]\times[0,1]} f(x,y)\, dA = \iint_{[0,1]\times[0,1]} (x + xy)\, dA = \int_0^1 \int_0^1 (x + xy)\, dx\, dy,$$

which is now straightforward to compute. (You can also integrate with respect to the order $dy\, dx$ instead and get the same value.)

Instead of splitting up the entire rectangle, we can proceed as follows, which eventually ends up at the same place. Fubini's Theorem gives

$$\iint_{[0,2]\times[0,1]} f(x,y)\, dA = \int_0^1 \int_0^2 f(x,y)\, dx\, dy.$$

Now, the inner integral over $[0,2]$ can be broken up into the integral over $[0,1]$ plus the integral over $[1,2]$:

$$\int_0^2 f(x,y)\,dx = \int_0^1 f(x,y)\,dx + \int_1^2 f(x,y)\,dx$$

Again, the second integral here is zero since f is zero when $x > 1$, so we are left with:

$$\int_0^2 f(x,y)\,dx = \int_0^1 (x+xy)\,dx.$$

Thus

$$\iint_{[0,2]\times[0,1]} f(x,y)\,dA = \int_0^1 \int_0^2 f(x,y)\,dx\,dy = \int_0^1 \int_0^1 (x+xy)\,dx\,dy$$

just as before. The upshot is that, just as in the single-variable case, regions of integration can be split up into pieces, which in turn splits up a given integral into a sum of integrals.

Warm-Up (alternate version). Define $g : \mathbb{R}^2 \to \mathbb{R}$ by

$$g(x,y) \begin{cases} x+xy & x \le y \\ 0 & x > y, \end{cases}$$

so that now the discontinuities occur along the line $y = x$. In this case, any line segment in $[0,2]\times[0,1]$ parallel to one of the axes intersects the set of discontinuities of g only once, so the book's version of Fubini's Theorem applies. (Of course, our more general version would apply as well.) We get

$$\iint_{[0,2]\times[0,1]} g(x,y)\,dA = \int_0^1 \int_0^2 g(x,y)\,dx\,dy.$$

Now, for a fixed $0 \le y \le 1$ we can split up the inner integral as:

$$\int_0^2 g(x,y)\,dx = \int_0^y g(x,y)\,dx + \int_y^2 g(x,y)\,dx.$$

For $0 \le x \le y$ we have $g(x,y) = x + xy$, while for $y < x \le 2$ we have $g(x,y) = 0$, so the expression above becomes:

$$\int_0^2 g(x,y)\,dx = \int_0^y g(x,y)\,dx + \int_y^2 g(x,y)\,dx = \int_0^y (x+xy)\,dx + \int_y^2 0\,dx = \int_0^y (x+xy)\,dx.$$

Hence our original iterated integral becomes

$$\iint_{[0,2]\times[0,1]} g(x,y)\,dA = \int_0^1 \int_0^2 g(x,y)\,dx\,dy = \int_0^1 \int_0^y (x+xy)\,dx\,dy,$$

which is now straightforward to compute. The new realization is that we have an inner bound which is not constant, but depends on one of the variables.

Integrating over general regions. Suppose D is a compact region in $\mathbb{R}^n$ and that $f : D \to \mathbb{R}$ is a function we wish to integrate over D. (There are technical assumptions we should make about D in order to guarantee that it makes to integrate functions *over* D. Essentially, what's required is that D have a well-defined "volume" in $\mathbb{R}^n$, but we won't make this precise in this course. The "Type I", "Type II", etc. regions that the book defines have this property, as will any region whose

boundary can be described by piecewise continuous equations; these are essentially the *elementary regions* to which the book refers. We'll assume from now on that we are always dealing with regions on which "integration" makes sense, so we won't use the "Type I", etc. terminology. You would learn more about these restrictions in an analysis course.) To define what it means to integrate f over D we use what we've already developed about integrating over boxes.

Pick a box B in $\mathbb{R}^n$ containing D. *Extend* the function f to all of B be defining it to be zero outside of D; that is, define $f^{ext} : B \to \mathbb{R}$ by

$$f^{ext}(\mathbf{x}) = \begin{cases} f(\mathbf{x}) & \mathbf{x} \in D \\ 0 & \mathbf{x} \notin D, \end{cases}$$

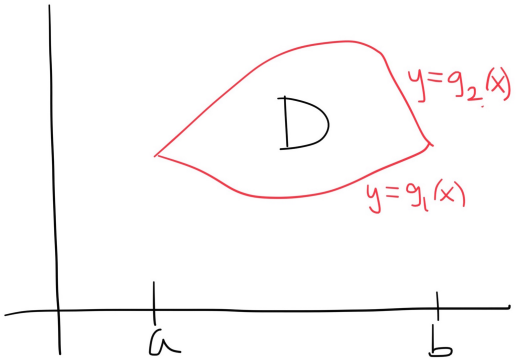
which agrees with f throughout D. Since f^{ext} is now defined on a box, it makes sense to talk about integrating f^{ext} over B. We say that f is *integrable* over D if f^{ext} is integrable over B, and we define the *integral* of f over D to be that of f^{ext} over B:

$$\int_D f(\mathbf{x})\, d\mathbf{x} := \int_B f^{ext}(\mathbf{x})\, d\mathbf{x}.$$

The idea is that, since f^{ext} equals 0 outside of D, the integral above in the end only depends on how f^{ext} behaves on D, which is given by how the original f behaves on D. Thus, even though we are technically integrating the extended function over a larger box, the value only depends on the original given information. It is a fact that the value obtained does not depend on which specific box B was chosen to enclose D, essentially because the extended function is zero outside of D.

Computing via iterated integrals. Now that we have spoken about how to *define* integrals of functions over more general regions, we move to the question of how to *compute* such integrals. The answer, as in the box case, is given by Fubini's Theorem. We assume the the extended function f^{ext} satisfies the hypotheses of Fubini's Theorem. Then we get that we can express our original integral over D as an iterated integral over the larger box B. The new observation is that the bounds on these integrals are no longer necessarily constant, but rather are used to describe the region of integration D itself.

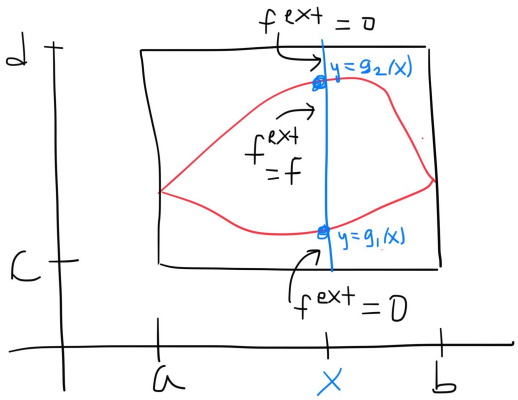
Let's see what this looks like in a 2-dimensional example. Say D is the region in $\mathbb{R}^2$ below, bounded on the bottom by a curve with equation $y = g_1(x)$ and on the top by a curve equation $y = g_2(x)$:



Pick some box $[a, b] \times [c, d]$ which contains D. Then Fubini's Theorem gives:

$$\iint_D f(x, y)\, dA := \iint_B f^{ext}(x, y)\, dA = \int_a^b \int_c^d f^{ext}(x, y)\, dy\, dx.$$

Now, at a fixed x, the interval $[c, d]$ over which the inner integral is computed over can be split up into the intervals $[c, g_1(x)]$, $[g_1(x), g_2(x)]$, and $[g_2(x), d]$:



so

$$\int_c^d f^{ext}(x,y)\,dy = \int_c^{g_1(x)} f^{ext}(x,y)\,dy + \int_{g_1(x)}^{g_2(x)} f^{ext}(x,y)\,dy + \int_{g_2(x)}^d f^{ext}(x,y)\,dy.$$

Over the first and third intervals the extended function is zero, and over the middle interval the extended function is the same as f, so

$$\int_c^d f^{ext}(x,y)\,dy = \int_{g_1(x)}^{g_2(x)} f(x,y)\,dy$$

and thus we get finally that

$$\iint_D f(x,y)\,dA = \int_a^b \int_{g_1(x)}^{g_2(x)} f(x,y)\,dy\,dx.$$

A similar reasoning works to express integrals over arbitrary regions in terms of iterated integrals. Rather than go through the process of picking a box, extending our function, and breaking up inner integrals into pieces every single time, we will from now on jump directly to the final iterated integral expression. The bounds in general come from the equations characterizing the boundaries of the region in question. For instance, in the 2-dimensional example above, the outer bounds on x came from the varying x through all possible values it could take on throughout the given region, so a to b in this case, and the inner bounds on y came from determining the values which y can range over at a *fixed* value of the outer variable x: at a fixed x, the values of y characterizing points in our region D can only range from $y = g_1(x)$ at the bottom to $y = g_2(x)$ at the top.

Note that we can also write this as an iterated integral with respect to $dx\,dy$ instead, only that now a subtlety arises. The inner bounds on x in this case should come from looking at the values x rangers over at a fixed y; since x increases from left to right in the usual way we draw the xy-plane, this means that the lower bound on x comes from the "leftmost" boundary of D and the upper bound comes from the "rightmost" boundary. However, the subtlety is that these left and right boundaries are NOT described by single equations, but rather each is described by two separate equations: in the picture above, the leftmost boundary consists of a portion of the curve $y = g_2(x)$ together with a portion of the curve $y = g_1(x)$, and similar for the rightmost boundary. Thus, to write this as an iterated integral with respect to $dx\,dy$ requires that we split up the region of integration into pieces, so that on each piece the left and right boundaries are indeed each given

by single equations. We'll see plenty of examples of this occurring, and is at the core of why the book describes regions in terms of the "Type I", "Type II", etc. terminology.

Examples. The examples we looked at in class can be found in my Math 290-3 lecture notes, and indeed many more examples beyond what we looked at in class can be found there as well. I encourage you to look through those notes to see as many examples as possible. The lectures titled "More on Double Integrals" and "Changing Order of Integration" are the relevant ones as far as double integrals are concerned. In particular, Example 4 from "More on Double Integrals" is one which requires splitting a region up into pieces, regardless of which order we choose to integrate with respect to. I'll assume going forward that you've worked through enough examples of setting up double integrals to understand how to do it well, but again, check my old notes for explicit computations worked out in detail.

Lecture 9: Triple Integrals

Warm-Up. As a Warm-Up we computed the value of

$$\int_0^1 \int_{-\sqrt[5]{x}}^{\sqrt{x}} \sin y^3 \, dy \, dx.$$

This is worked out in detail in my Math 290-3 notes, in particular the Warm-Up of the first "Triple Integrals" lecture, so I'll leave it to you to check there. The key points were recognizing that we must switch the order of integration, determining the correct region of integration, determining the correct bounds after switching the order, and using symmetry to compute the resulting iterated integral.

Triple integrals via iterated integrals. The integral of a function f of three variables over a general compact region E in $\mathbb{R}^3$ is defined similarly to the 2-dimensional case we saw last time: we pick a box B containing E, extend f to be zero outside of E, and then define the integral of f over E to be the integral of the extended function over B. In the end, when Fubini's Theorem applies, such integrals can also be computed using iterated integrals.

As in the two-variable case, the bounds on a 3-dimensional iterated integral over a general region are no longer required to be constant, but are used to describe the region of integration itself. The key things to note are: the "middle" and "outer" bounds are used to describe the *shadow* (or projection) of the 3-dimensional region of integration in the plane corresponding to the middle and outer variables, and the inner bounds indicate the range of values the inner variable can take on at fixed values of the middle and outer variables.

Examples. The examples we looked at in class and many more can be found in my Math 290-3 notes, in particular the days titled "Triple Integrals", "More on Triple Integrals", and "Yet More on Triple Integrals". I strongly encourage you to go through *all* examples given there since it definitely takes practice to understand how to setup triple integrals correctly. This is something I will assume you are comfortable doing, so ask if the process is still unclear after checking my 290-3 notes. As part of this, you really have to get comfortable with visualizing solid regions in 3-dimensions—there's just no way around this.

Lecture 10: More on Integration

Warm-Up. Continuing with an example from the previous lecture, as a Warm-Up in class we rewrote the integral

$$\int_0^1 \int_0^{1-x} \int_0^{1-x^2} f(x,y,z)\,dz\,dy\,dx$$

with respect to the order $dx\,dz\,dy$ instead. This is done in Example 1 of the "Yet More on Triple Integrals" lecture of my Math 290-3 notes, so check there for the details. The key observation is that when integrating with respect to $dx\,dz\,dy$, we are forced to split the 3-dimensional region up into pieces since the "upper" (i.e. frontmost) bound on x depends on where in the region we are.

What do integrals compute? After having spent much discussing how to setup and compute integrals, we should return to the question as to what integrals actually compute. The answer we've given is that: the integral of $f : E \to \mathbb{R}$ over $E \subseteq \mathbb{R}^n$ computes the n-dimensional "volume" of the region in $\mathbb{R}^{n+1}$ lying between the graph of f and E. Certainly, this is simple to visualize in the $n = 1$ and $n = 2$ cases, where integrals give areas and ordinary 3-dimensional volumes accordingly. However, this point of view does not work well when $n \geq 3$ since four- and higher-dimensional things are difficult to visualize.

However, in a sense this geometric perspective on what integration is isn't the most useful and isn't how mathematicians and others actually use integration in practice. Instead, we should focus on the idea that an integral is meant to some type of analog of "summation". Indeed, a single-variable integral

$$\int_a^b f(x)\,dx$$

should be interpreted as "adding" up all the values of $f(x)$ as x ranges over $a \leq x \leq b$, and a two-variable integral

$$\iint_D f(x,y)\,dA$$

should be thought of as "adding" all values $f(x,y)$ as (x,y) ranges throughout D. More generally then, we think of

$$\int_E f(\mathbf{x})\,d\mathbf{x}$$

for $E \subseteq \mathbb{R}^n$ and $f : E \to \mathbb{R}$ as adding up all values of $f(\mathbf{x})$ as $\mathbf{x}$ ranges throughout E. The point is that it does not literal sense to add up all such infinitely many values of $f(\mathbf{x})$, but the Riemann sum definition of an integral gives a way to make some sense of this. Then, $d\mathbf{x}$ should be interpreted as some type of "infinitesimal volume" (dA is an infinitesimal area in the two-variable case) which is used to try to give some meaning to "adding up infinitely many things".

The fact that

$$\int_E 1\,d\mathbf{x} = \text{Vol}(E)$$

where $E \subseteq \mathbb{R}^n$ and Vol denotes a general n-dimensional volume makes sense from this perspective, since we are "adding" up a bunch of 1's, one for each point of E, so that this process keeps track of how much "stuff" (i.e. volume) is actually in E. The "integration = summation" point of view will be incredibly useful in the coming weeks when we cover *line* and *surface* integrals.

Example. Define E to be the subset of $\mathbb{R}^n$ defined by

$$E := \{\mathbf{x} \in \mathbb{R}^n \mid \text{all } x_i \geq 0 \text{ and } x_1 + \cdots + x_n \leq 1\}.$$

So, this is the region in $\mathbb{R}^n$ bounded by all coordinate hyperplanes $x_i = 0$ and the hyperplane $x_1 + \cdots + x_n = 1$. When $n = 2$ this looks like a triangle, and when $n = 3$ this looks like a tetrahedron, so in general this is a higher-dimensional analog of these objects. (The formal name for this region is the *standard n-simplex*.)

We setup an iterated integral which computes the volume of E, which is equal to the n-dimensional integral

$$\int_E 1 \, d\mathbf{x}$$

where $\mathbf{x} = (x_1, \ldots, x_n)$. Since the constant function 1 is continuous, Fubini's Theorem applies without concern. Saw we want to integrate with respect to the order

$$dx_1 \, dx_2 \, \ldots \, dx_{n-1} \, dx_n.$$

The outer bounds on x_n come from the smallest and largest values x_n can take on throughout E, which are 0 and 1 respectively. Now, the next set of bounds on x_{n-1} come from the "slice" of E occurring at a fixed value of x_n. At a fixed x_n, the remaining variables must satisfy

$$x_1 + \cdots + x_{n-1} \leq 1 - x_n.$$

Thus, the smallest and largest values which x_{n-1} can take on throughout this slice are 0 and $1 - x_n$, so these give the bounds on x_{n-1}. At a fixed (x_{n-1}, x_n), the remaining variables satisfy

$$x_1 + \cdots + x_{n-2} \leq 1 - x_n - x_{n-1},$$

so the smallest and largest values of x_{n-2} are 0 and $1 - x_n - x_{n-1}$, and so on. Once we get down to x_2 the bounds on 0 to $1 - x_n - \cdots - x_3$, and the finally the innermost bounds on x_1 are 0 to $1 - x_1 - \cdots - x_2$ since at this stage the allowed values of x_1 satisfy

$$0 \leq x_1 \leq 1 - x_n - x_{n-1} - \cdots - x_2$$

at a fixed $(x_2, \ldots, x_n)$. The point is that at each stage, the bounds on the corresponding integral come from the slice of E occurring at fixed values of all previous variables. We end up with:

$$\int_E 1 \, d\mathbf{x} = \int_0^1 \int_0^{1-x_n} \cdots \int_0^{1-x_n-\cdots-x_3} \int_0^{1-x_n-\cdots-x_3-x_2} dx_1 \, dx_2 \, \ldots \, dx_{n-1} \, dx_n$$

as the iterated integral giving the volume of E. Computing these would involve n total integrations, working our way from the innermost integral to the outermose.

Regions enclosed by spheres. The set S^n, called the *n-sphere*, is the set of all points in $\mathbb{R}^{n+1}$ at a distance 1 from the origin:

$$S^n := \{\mathbf{x} \in \mathbb{R}^{n+1} \mid \|\mathbf{x}\| = 1\}.$$

In particular, S^1 is the usual unit circle centered at the origin $\mathbb{R}^2$, and S^2 is the usual unit sphere in $\mathbb{R}^3$. (In higher dimensions, S^n is also often referred to as a *hypersphere*.) To describe higher-dimensional spheres of other radii, a subscript S^n_R is usually used:

$$S^n_R := \{\mathbf{x} \in \mathbb{R}^{n+1} \mid \|\mathbf{x}\| = R\}.$$

The region enclosed by S_R^{n-1} is often denoted by D_R^n and consists of all points in $\mathbb{R}^n$ whose distance to the origin is smaller than or equal to R:

$$D_R^n := \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\| \leq R\}.$$

This is often called the *closed ball* of radius R centered at the origin, and is denoted by $\overline{B_R(\mathbf{0})}$ where $B_R(\mathbf{0})$ is the notation for open balls we used last quarter. Another name for D_R^n is the *disk* (or *hyperdisk*) of radius R centered at the origin, which comes from the fact that D_R^2 in $\mathbb{R}^2$ is an ordinary disk.

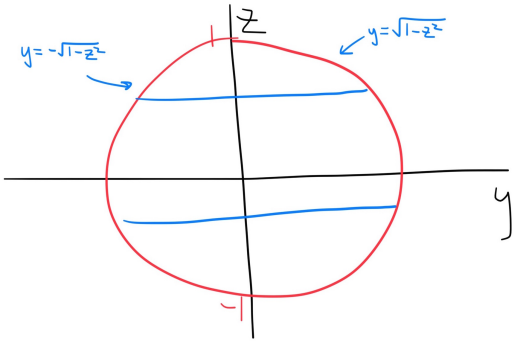
We are interested in finding the volume of D_R^n, which is the volume of the region enclosed by a sphere of radius R in $\mathbb{R}^n$. The computation for general n and R is on Homework 3, but here is how to proceed in the case D_1^3 of the region enclosed by the ordinary unit sphere in $\mathbb{R}^3$. The answer is that

$$\iiint_{D_1^3} 1\, dV = \text{Vol}(D_1^3) = \frac{4\pi}{3},$$

which is a well-known formula for the volume of the region enclosed by the unit sphere in $\mathbb{R}^3$. However, here we instead *derive* this fact. Say we want to setup this triple integral as an iterated integral with respect to the order $dx\, dy\, dz$:

$$\iiint_{D_1^3} dV = \int_?^? \int_?^? \int_?^? dx\, dy\, dz.$$

The outermost bounds on z are -1 and 1, which are the smallest and largest values of z throughout D_1^3. Now, the shadow of this region in the yz-plane is the unit disk $y^2 + z^2 \leq 1$ in the yz-plane:

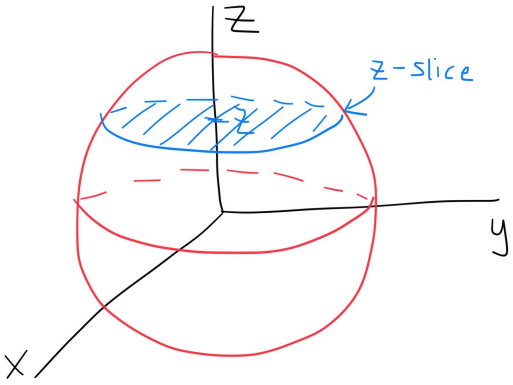


so the bounds on y are $y = -\sqrt{1-z^2}$ on the left side and $y = \sqrt{1-z^2}$ on the right. Finally, at a fixed (y, z), x (moving back to front) goes from the back half of the unit sphere $x = -\sqrt{1-y^2-z^2}$ to the front half $x = \sqrt{1-y^2-z^2}$, so we end up with:

$$\iiint_{D_1^3} dV = \int_{-1}^{1} \int_{-\sqrt{1-z^2}}^{\sqrt{1-z^2}} \int_{-\sqrt{1-y^2-z^2}}^{\sqrt{1-y^2-z^2}} dx\, dy\, dz.$$

This iterated integral is possible to compute by hand, but will require some trig substitutions and will be a little messy. (We'll see later than in spherical coordinates this becomes much simpler to compute by hand.)

However, there is another simple way to compute this integral, by interpreting it in terms of slices instead. Indeed, we recognize that at a fixed z, the middle and inner bounds are meant to describe to the slice of D_1^3 occurring at that value of z. Such a slice looks like a disk at a certain height:



Thus, since we are integrating the function 1, the middle and inner integrals together should give the area of this slice:

$$\iiint_{D_1^3} dV = \int_{-1}^{1} \underbrace{\left(\iint_{z\text{-slice}} 1 \, dA \right)}_{\text{area of } z\text{-slice}} dz$$

In other words, we are saying that the volume of D_1^3 should be obtained by "adding" up the areas of all horizontal slices as z varies. Since the slice at a fixed z has equation

$$x^2 + y^2 \leq 1 - z^2,$$

this slice is a disk of radius $\sqrt{1-z^2}$ in a plane parallel to the xy-plane, so the area of this slice is

$$\pi(\text{radius})^2 = \pi(1 - z^2).$$

Thus we get that

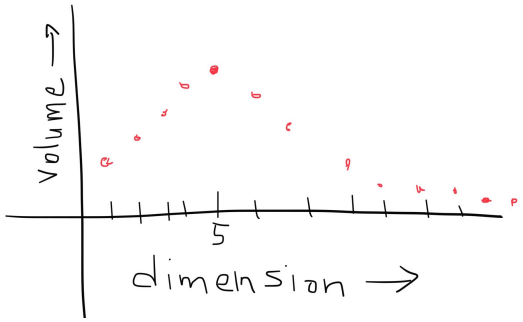
$$\iiint_{D_1^3} dV = \int_{-1}^{1} \underbrace{\left(\iint_{z\text{-slice}} 1 \, dA \right)}_{\text{area of } z\text{-slice}} dz = \int_{-1}^{1} \pi(1 - z^2) \, dz = \frac{4\pi}{3}$$

as expected. A similar argument gives the volume of D_R^3 for any radius, and a similar reasoning using higher-dimensional slices can be used to compute the volume of D_R^n in general, as you'll do on the homework.

Fun fact about spheres. Just for fun, here is an interesting fact about spheres which follows from the general volume computation on the homework. The observation is that the volumes of the regions enclosed by hyperspheres of a fixed radius have an unexpected behavior as the dimension increases. To start with, consider the hyperspheres of radius 1. The first few volumes of the regions they enclosed are:

$$\text{Vol}(D_1^1) = 2, \quad \text{Vol}(D_1^2) = 2\pi, \quad \text{Vol}(D_1^3) = \frac{4\pi}{3},$$

where we keep in mind that "volume" in $\mathbb{R}^1$ really means length, and "volume" in $\mathbb{R}^2$ really means area. If you plot these values in a graph showing what happens as the dimension n increases you get something like:



The amazing behavior this suggests, and which is in fact true, is that the volume of unit hyperspheres is maximized in dimension 5, and then approaches 0 as n increases! Thus, higher-dimensional spheres actually start getting *smaller* past a certain point, which runs counter to our intuition based on what we know about circles in $\mathbb{R}^2$ and spheres in $\mathbb{R}^3$.

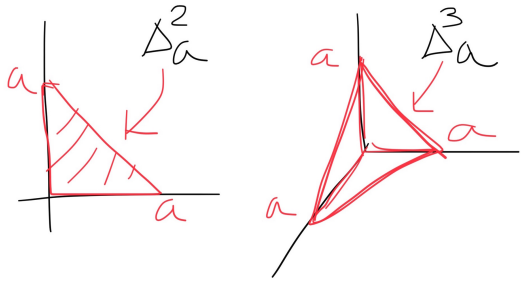
For a general radius R, the maximum volume isn't attained in dimension 5, but nonetheless there *is* a dimension in which the volume of a hypersphere of radius R is maximized, *and* after this dimension the volumes start decreasing towards zero. I'll give an explanation as to why this happens in the solutions to Homework 3 after we derive the general expression for these volumes in any dimension. The moral is: crazy things can happen in higher dimensions.

Lecture 11: Change of Variables

Warm-Up. Fix $a > 0$ and define Δ_a^n to be the region in $\mathbb{R}^n$ bounded by the coordinate hyperplanes and the hyperplane $x_1 + \cdots + x_n = a$:

$$\Delta_a^n := \{\mathbf{x} \in \mathbb{R}^n \mid \text{all } x_i \geq 0 \text{ and } x_1 + \cdots + x_n \leq a\}.$$

This is called an *n-simplex*; the standard n-simplex is the one we looked at last time where $a = 1$. As mentioned last time, when $n = 2$ this looks like a triangle and when $n = 3$ this looks like a tetrahedron:



We now describe a recursive approach to computing the volume of Δ_a^n, where "recursive" means that we will express this volume in terms of the volumes of lower-dimensional simplices.

We wrote out the iterated integral giving the volume of the standard n-simplex last time, and the idea we use here is the one we used last time when considering volumes of regions enclosed by hyperspheres, namely we use slices to characterizes volumes. We have

$$\int_{\Delta_a^n} 1 \, d\mathbf{x} = \int_0^a \underbrace{\int_0^{a-x_n} \cdots \int_0^{a-x_n-\cdots-x_3-x_2} dx_1 \, \ldots \, dx_{n-1}}_{\text{integrates over an } x_n\text{-slice}} \, dx_n.$$

For a fixed x_n, the remaining variables satisfy

$$x_1 + \cdots + x_{n-1} \leq a - x_n,$$

which is the inequality defining the $(n-1)$-simplex $\Delta^{n-1}_{a-x_n}$. Thus, the iterated integrals corresponding to $x_1, \ldots, x_{n-1}$ describe the integration of the function 1 over this x_n-slice, meaning that these integrals compute the volume of this x_n-slice. Hence

$$\operatorname{Vol}(\Delta^n_a) = \int_{\Delta^n_a} d\mathbf{x} = \int_0^a \operatorname{Vol}(\Delta^{n-1}_{a-x_n})\, dx_n,$$

which says that the volume of Δ^n_a is obtained by adding up the lower dimensional volumes of all its x_n-slices as x_n varies. Thus, if we know all of these lower-dimensional volumes, we can use this to recursively compute all higher-dimensional volumes as well.

But we can do a bit better, since we can directly relate the volume of $\Delta^{n-1}_{a-x_n}$ to that of Δ^{n-1}_1. Indeed, note that $\Delta^{n-1}_{a-x_n}$ can be obtained from Δ^{n-1}_1 by scaling each coordinate by $a - x_n$, since if $(x_1, \ldots, x_{n-1})$ satisfy

$$x_1 + \cdots + x_{n-1} \leq 1,$$

then $(x'_1, \ldots, x'_{n-1}) = ((a-x_n)x_1, \ldots, (a-x_n)x_{n-1})$ satisfy

$$x'_1 + \cdots + x'_{n-1} \leq a - x_n.$$

Thus, $\Delta^{n-1}_{a-x_n}$ is the image of Δ^{n-1}_1 under the linear transformation given by $(a-x_n)I_{n-1}$, so based on what we know about the expansion factor interpretation of determinants we get:

$$\operatorname{Vol}(\Delta^{n-1}_{a-x_n}) = |\det(a-x_n)I_{n-1}|\operatorname{Vol}(\Delta^{n-1}_1) = (a-x_n)^{n-1}\operatorname{Vol}(\Delta^{n-1}_1).$$

This gives

$$\operatorname{Vol}(\Delta^n_a) = \int_0^a (a-x_n)^{n-1}\operatorname{Vol}(\Delta^{n-1}_1)\, dx_n,$$

so that all these volumes computations can be reduced to ones involving the standard n-simplices alone. Using the starting facts that

$$\operatorname{Vol}(\Delta^1_1) = 1, \quad \operatorname{Vol}(\Delta^2_1) = \frac{1}{2}, \quad \operatorname{Vol}(\Delta^3_1) = \frac{1}{6},$$

we can then inductively determine the volume of Δ^n_a for all n and a. The answer, it turns out, is $\operatorname{Vol}(\Delta^n_a) = \frac{a^n}{n!}$. A nice exercise in induction indeed.

Review of substitution. We now move towards developing a general "change of variables" formula for multiple integrals, meaning a technique which will let us rewrite a given integral expression in terms of a new set of coordinates—think polar or spherical coordinates. On the one hand, the immediate practical benefit of doing so is that changing variables in this way will often times make integral computations simpler, either because the region in question is simpler to describe in terms of new coordinates, or because the function to integrate is simpler to describe. But also, we'll see that the general change of variables formula isn't only useful for the sake of practical computations, it also gives us a way to *define* other types of integrals we'll be interested in (i.e. line and surface integrals) in a way which makes use of the types of integrals we've already defined without having to dig down into a Riemann sum type of definition all over again.

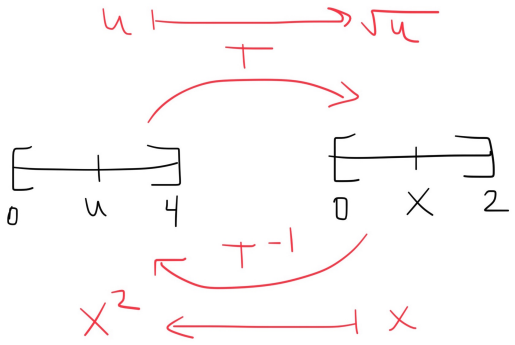
Before moving to the general setting, we first give a new perspective on the single-variable integration method of substitution, of which the change of variables formula we'll give is meant to be generalization. For instance, the consider the integral

$$\int_0^2 xe^{x^2}\,dx.$$

Normally you would compute this using the substitution $u = x^2$, which gives $du = 2x\,dx$ and turns the given integral into

$$\int_0^4 \frac{1}{2}e^u\,du.$$

However, now we emphasize what the substitution $u = x^2$ is actually doing by interpreting it as a *coordinate transformation*. That is, consider the function $T : [0, 4] \to [0, 2]$ which sends the variable u in the domain to the variable x in the codomain:



Since $u = x^2$, $x = \sqrt{u}$ so T is given by $T(u) = \sqrt{u}$. (The map $x \mapsto x^2$ actually describes the *inverse* of T.) The original integral in terms of x occurs on the "right" side of the diagram above, and the substitution we make allows us to rewrite this as an integral on the "left" side in terms of u instead. The function xe^{x^2} on the right side becomes the function $\sqrt{u}e^u$ on the left since $x = \sqrt{u}$.

Now, the Jacobian matrix (which is of size 1×1 in this case) of T is given by $T'(u) = \frac{1}{2\sqrt{u}}$. This matrix describes how to transform the dx term on the right into the du term on the left:

$$dx = T'(u)\,du = \frac{1}{2\sqrt{u}}\,du,$$

which is equivalent to $du = 2x\,dx$ under the substituton $x = \sqrt{u}$. Thus

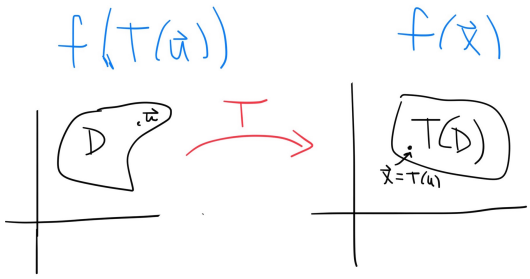
$$xe^{x^2}\,dx \text{ becomes } \sqrt{u}e^u\left(\frac{1}{2\sqrt{u}}\,du\right) = \frac{1}{2}e^u\,du,$$

which is the integrand we end up with after the substitution. The bounds on the resulting interval come from recognizing that the image of $[0, 4]$ under the transformation $T(u) = \sqrt{u}$ is $T([0, 4]) = [0, 2]$, so the interval of integration $[0, 2]$ on the right corresponds to the interval of integration $[0, 4]$ on the right. To summarize, the substitution $x = T(u) = \sqrt{u}$ (whose inverse is $u = x^2$), gives

$$\int_{[0,4]} \underbrace{\sqrt{u}e^u}_{\text{function}}\ \underbrace{\frac{1}{2\sqrt{u}}}_{\text{Jacobian}}\ du = \int_{T([0,4])=[0,2]} xe^{x^2}\,dx,$$

where we have pointed out where each term on the left explicitly comes from. This type of expression, involving a Jacobian, is the form the general change of variables formula will take.

Coordinate transformations. More generally then, we consider other *coordinate transformations* $T : \mathbb{R}^n \to \mathbb{R}^n$. We denote the variables in the domain by $\mathbf{u} = (u_1, \dots, u_n)$ and in the codomain by $\mathbf{x} = (x_1, \dots, x_n)$. The problem is to relate integrals in terms of the $\mathbf{x}$'s to integrals in terms of the $\mathbf{u}$'s. Given some region D we wish to integrate over on the $\mathbf{u}$-side, we in turn get some region $T(D)$ (i.e. the image of D under T) we can integrate over on the $\mathbf{x}$-side. Given some function f we want to integrate on the $\mathbf{x}$-side, the composition $f \circ T$ describes the corresponding function on the left side, which is just the same function as f only written in terms of $u_1, \dots, u_n$ instead of $x_1, \dots, x_n$ using the substitution $\mathbf{x} = T(\mathbf{u})$:



With this notation, the change of variables formula we are after will look something like

$$\int_D f(T(\mathbf{u}))(\text{something})\, d\mathbf{u} = \int_{T(D)} f(\mathbf{x})\, d\mathbf{x},$$

where "something", which describes what happens to $d\mathbf{x}$ under this change of variables, is still to-be-determined. In practice, we would be given, say, the integral on the right in terms of standard rectangular coordinates, and after finding the appropriate coordinate transformation to use we would get the integral on the left, which might be simpler to compute directly. As an example, converting into polar coordinates corresponds to the transformation

$$T(r, \theta) = (r \cos\theta, r \sin\theta),$$

which tells us how to describe x, y on the right side in terms of r, θ on the left side.

We make some technical assumptions on the types of coordinate transformations T which are allowed; namely, T should be C^1, injective, and should have invertible Jacobian matrix at each point. We'll see later that we can relax these conditions a bit by only requiring that they hold everywhere except for on a set of measure zero. We'll see what can go wrong without these assumptions later, but the point is that these are the requirements needed to ensure that volumes transform appropriately, in that nonzero volumes shouldn't be sent to zero volumes and volumes should not be "over counted".

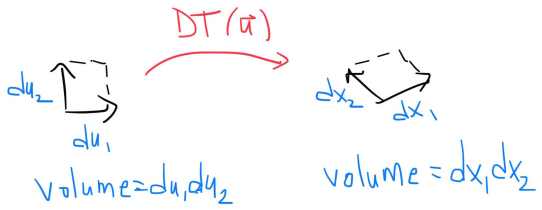
Infinitesimal volumes. The key thing in all of this comes in understanding what happens to $d\mathbf{x}$ under our change of variables. For now we focus solely on an intuitive approach using "infinitesimals", but we'll say something a little more concrete next time in terms of Riemann sums. The intuitive idea is that

$$d\mathbf{x} = dx_1 \dots dx_n$$

is meant to represent an infinitesimal volume in the codomain and

$$d\mathbf{u} = du_1 \dots, du_n$$

represents an infinitesimal volume in the domain:



Thinking of each side of these infinitesimal parallelopipeds as infinitesimal vectors, we know from one interpretation of Jacobians we gave last quarter that $DT(\mathbf{u})$ is the matrix which tells us how infinitesimal vectors in the domain get transformed into infinitesimal vectors in the codomain. Thus, based on the expansion factor interpretation of determinants, we find that:

$$d\mathbf{x} = |\det DT(\mathbf{u})|\, d\mathbf{u},$$

since $d\mathbf{u}$ is the infinitesimal volume of the original infinitesimal parallelopiped, and $d\mathbf{x}$ is the infinitesimal volume of the image infinitesimal parallelopiped. Thus, the so-called *Jacobian determinant* $\det DT(\mathbf{u})$ gives the expansion factor by which infinitesimal volumes change under the coordinate transformation T. This is the final ingredient we need in our change of variables formula.

Change of variables. Here is the statement.

> Let D be a region of integration in the domain $\mathbb{R}^n$. Suppose $T : \mathbb{R}^n \to \mathbb{R}^n$ is a coordinate transformation, meaning that T is C^1, injective, and has invertible Jacobian throughout D. (Again, we'll see later that we can relax these conditions a bit so that they only hold away from a set of measure zero.) Let $T(D)$ be the image of D in the codomain $\mathbb{R}^n$ under the coordinate transformation T. If $f : T(D) \to \mathbb{R}$ is integrable, then $f \circ T : D \to \mathbb{R}$ is integrable and
>
> $$\int_D f(T(\mathbf{u}))|\det DT(\mathbf{u})|\, d\mathbf{u} = \int_{T(D)} f(\mathbf{x})\, d\mathbf{x}.$$
>
> To be clear, the left side is what is obtained from the right side after making the change of variables $\mathbf{x} = T(\mathbf{u})$.

The book only states this in the 2- and 3-dimensional cases, but the formula for general n is the same. Also, the book is a little sloppy in the assumptions it makes, in that it doesn't require that $DT(\mathbf{u})$ be invertible, which is actually important. Finally, note that the book writes the resulting formula in the opposite order:

$$\int_{T(D)} f(\mathbf{x})\, d\mathbf{x} = \int_D f(T(\mathbf{u}))|\det DT(\mathbf{u})|\, d\mathbf{u},$$

to emphasize that the integral on the left is being converted into the one on the right. I prefer the order I used above since it "matches" the direction the coordinate transformation T goes in, namely that it goes *from* $\mathbf{u}$'s *to* $\mathbf{x}$'s. With respect to the order in which I wrote it, we'll think of the integrand

$$f(T(\mathbf{u}))|\det DT(\mathbf{u})|\, d\mathbf{u}$$

on the left as what you get when you "pullback" the integrand

$$f(\mathbf{x})\, d\mathbf{x}$$

on the right using T. We'll give a more formal meaning to the term *pullback* later when discussing differential forms, which we'll make this even clearer.

Example. To give one concrete example for now, suppose we wish to compute:

$$\iint_{D^*} xy^2 \, dA$$

where D^* is the region in the first quadrant of $\mathbb{R}^2$ bounded by the curves $xy = 1, xy = 4, xy^2 = 1$, and $xy^2 = 4$. (You can find a picture of this in Example 3 of the "Change of Variables" lecture in my Math 290-3 notes.) This is possible to compute as is, but would require splitting the region D^* up into pieces when setting up the appropriate iterated integrals.

Instead, we make a change of variables as follows. Set

$$u = xy \quad \text{and} \quad v = xy^2.$$

In terms of coordinate transformations, these equations describe a transformation from (x, y) to (u, v). To match up with the way we described coordinate transformation above, $T : \mathbb{R}^2 \to \mathbb{R}^2$ should really describe how to transform (u, v) into (x, y), meaning that the equations above actually describe the *inverse* of T:

$$T^{-1}(x, y) = (xy, xy^2).$$

Using the given equations to solve for x, y in terms of u, v gives

$$x = \frac{u^2}{v} \quad \text{and} \quad y = \frac{v}{u},$$

which describe T itself:

$$T(u, v) = \left(\frac{u^2}{v}, \frac{v}{u} \right).$$

Under this transformation, $D^* = T(D)$ as given is the image of the square $D = [1, 4] \times [1, 4]$, which we get from writing the equations of the curves bounding D^* in terms of u, v instead. Indeed, the form that these equations take is the main reason why we chose to make the change of variables we did. Two things to note: one, this expression is not defined when $v = 0$ or $u = 0$, but this is okay since the resulting integral we get will take place over D and T *is* defined throughout D. Second, actually we did not really have to solve or x, y in terms of u, v to make this work—we'll see that we could have avoided doing this explicitly.

Now, T is C^1 and injective on D, and has invertible Jacobian matrix throughout D as well, as you can check. The change of variables formula gives

$$\iint_D v |\det DT(u, v)| \, d(u, v) = \iint_{T(D)} xy^2 \, d(x, y),$$

where the v on the left comes from the expression for $f(x, y) = xy^2$ in terms of u, v. (Here I'm using $d(u, v)$ and $d(x, y)$ instead of dA in the notation to avoid the ambiguity as to whether "dA" means taken with respect to x, y or with respect to u, v.) The final thing to determine is the Jacobian determinant, which, given the expression for T we derived above, is:

$$\det DT(u, v) = \det \begin{pmatrix} \frac{2u}{v} & -\frac{u^2}{v^2} \\ -\frac{v}{u^2} & \frac{1}{u} \end{pmatrix} = \frac{2}{v} - \frac{1}{v} = \frac{1}{v}.$$

This is what the book denote as

$$\frac{\partial(x,y)}{\partial(u,v)},$$

which, as the notation suggests, is the determinant of the matrix obtained by differentiating x, y with respect to u, v.

But, there is another way we can find this Jacobian determinant which avoids having to express x, y in terms of u, v directly. At the start, we described our change of variables by setting $u = xy$ and $v = xy^2$, which as mentioned previously actually gives the inverse transformation

$$T^{-1}(x, y) = (xy, xy^2).$$

The Jacobian matrix of this inverse transformation is

$$DT^{-1}(x,y) = \begin{pmatrix} y & x \\ y^2 & 2xy \end{pmatrix}, \text{ so } \det DT^{-1}(x,y) = 2xy^2 - xy^2 = xy^2.$$

The point is that the matrices DT and DT^{-1} are actually inverses of one another, since the chain rule (in matrix form) gives:

$$DT^{-1}DT = D(T^{-1} \circ T) = D(\text{identity}) = I.$$

Thus, we have that

$$\det DT = \frac{1}{\det DT^{-1}} = \frac{1}{xy^2},$$

which agrees with the direct computation we gave for DT after we express xy^2 back in terms of u, v. The upshot is that we didn't actually need to solve for x, y in terms of u, v after setting $u = xy$ and $v = xy^2$, since the required Jacobian determinant can also be derived using the inverse transformation instead. Using the book's notation again, the determinant of DT^{-1} is

$$\frac{\partial(u,v)}{\partial(x,y)},$$

which indicates differentiating u, v with respect to x, y, and the fact is that this expression and the one we actually want $\frac{\partial(x,y)}{\partial(u,v)}$ should be inverses of one another:

$$\frac{\partial(x,y)}{\partial(u,v)} = \frac{\partial(u,v)}{\partial(x,y)}^{-1}.$$

This type of observation simplifies a lot of these computations in practice, since often it is simpler to describe the $\mathbf{u}$'s in terms of the $\mathbf{x}$'s, and this says that doing so will still give a way to compute the Jacobian determinant we actually need.

To finish up, the Jacobian expansion factor in our case is thus:

$$|\det DT(u,v)| = \left|\frac{1}{v}\right| = \frac{1}{v}$$

since v is positive throughout our region D. The change of variables formula thus gives:

$$\iint_D \underbrace{v}_{\text{function}} \underbrace{\left(\frac{1}{v}\right)}_{\text{Jacobian}} d(u,v) = \iint_{T(D)} xy^2 \, d(x,y)$$

where $D = [1,4] \times [1,4]$. The integral on the left is now just

$$\int_1^4 \int_1^4 du\, dv,$$

which is much simpler to compute than the original integral in terms of x, y.

Lectures 12 and 13: More on Change of Variables

These two days focused mainly on applying the general change of variables method to the case of polar, cylindrical, and spherical coordinates. The main take aways are that

$$dA = r\,dr\,d\theta$$

in polar coordinates (so the polar expansion factor is r), and

$$dV = r\,dr\,d\theta\,dz = \rho^2 \sin\phi\,d\rho\,d\phi\,d\theta$$

in cylindrical and spherical coordinates respectively (so the cylindrical expansion factor is r and the spherical expansion factor is $\rho^2 \sin\phi$.) All of this material is included in my Math 290-3 notes, as are all of the examples we looked at and then some. (The Warm-Up we did for Lecture 12 can also be found in those notes.) So, check those notes to review what this is all about and to see plenty of examples, and in these notes I'll just focus on the few concepts we covered which aren't mentioned in my Math 290-3 notes.

Change of variables proof. Last time we gave some intuition behind the change of variables formula in terms of "infinitesimal volumes", but a more formal proof would require working with Riemann sums. We'll outline how this works here, but a full proof is still beyond our grasp and requires a better understanding of analysis. We'll point out precisely where the assumptions in the change of variables formula actually come into play.

In the interest of saving time, I'll finish this portion later since it is not really relevant for the midterm.

Relaxing the assumptions. As mentioned in the statement of the change of variables formula from last time, the conditions that the coordinate transformation T be C^1, injective, and have invertible Jacobian matrix are actually only required to hold off of a set of measure zero. For instance, consider the usual polar coordinate transformation $T : [0, \infty) \times [0, 2\pi]$ given by

$$T(r, \theta) = (r\cos\theta, r\sin\theta).$$

This is certainly C^1, but it is not injective since

$$T(r, 0) = T(r, 2\pi)$$

for any r, and it does not have invertible Jacobian matrix at $r = 0$ since the determinant of this Jacobian is precisely r itself. So, T would not satisfy the strict requirements that it be injective and have invertible Jacobian everywhere throughout its domain.

Nonetheless, the point is that T only fails to be injective on the lines $\theta = 0$ and $\theta = 2\pi$, and the union of these two lines has measure zero in $\mathbb{R}^2$, and similarly DT fails to be invertible only at the origin where $r = 0$, which also has measure zero in $\mathbb{R}^2$. So, T *does* satisfy the assumptions needed in the change of variables formula if we only require that they hold off a set of measure zero, which means concretely that the set of points where these assumptions do not hold has measure zero. The intuition is that since, as we've mentioned previously, what happens over a set of measure zero can never affect the value of an integral, coordinate transformations which fail to be C^1, injective, or have invertible Jacobian only on such sets will still give a valid change of variables.

The same thing happens when converting to spherical coordinates. The determinant of the Jacobian is $\rho^2 \sin\phi$, which is actually zero for certain values within the ranges

$$0 \leq \rho < \infty, \quad 0 \leq \phi \leq \pi, \quad 0 \leq \theta \leq 2\pi$$

so this Jacobian is not invertible everywhere, but the set of points at which it fails to be invertible has measure zero in $\mathbb{R}^3$. Also, the spherical coordinate transformation over these ranges is not injective since cosine and sine give the same values at 0 and at 2π, but again these only fail to be injective on a set of measure zero. Thus, the spherical coordinate transformation is also valid when changing variables.

Why injectivity is needed. Consider the transformation $T : [0, 1] \times [0, 4\pi] \to \mathbb{R}^2$ given by

$$T(r, \theta) = (r \cos \theta, r \sin \theta),$$

which is just the usual polar coordinate transformation only that we allow θ to take on values from 0 to 4π as opposed to 0 to 2π as normal. This transformation is C^1 and has invertible Jacobian except for on a set of measure zero. If this were a valid change of variables formula, we would get

$$\int_0^{4\pi} \int_0^1 |\det DT(r, \theta)| \, dr \, d\theta = \iint_D dA$$

where $D = T([0, 1] \times [0, 4\pi])$ is the unit disk. The integral on the right just gives the area of D, which is π, but the integral on left, using the fact that $|\det DT| = r$, gives 2π. Thus this equality is not valid.

The issue is that T now fails to be injective everywhere on $[0, 1] \times [0, 4\pi]$ except for off a set of measure zero. This is because now for any $0 < \theta < 2\pi$, we have

$$T(r, \theta) = T(r, \theta + 2\pi)$$

and $\theta + 2\pi$ is still within the range $[0, 4\pi]$, whereas previously if we restrict the range of θ values to be between 0 and 2π this non-injectivity only happens at $\theta = 0$ and $\theta = 2\pi$. The problem is that with the range $[0, 4\pi]$, the transformation T actually traces out the unit disk *twice*, since the values $0 \leq \theta \leq 2\pi$ give the unit disk once and then $2\pi \leq \theta \leq 4\pi$ gives it again. Thus, the integral on the left "overcounts" the unit disk, which is why it gives twice the expected area. This is at the core of why T is required to be injective (off a set of measure zero) in the change of variables formula: injectivity guarantees that volumes aren't "overcounted".

Gaussian integrals. Finally, the one example we did which is not covered fully in my Math 290-3 notes is the computation of

$$\int_{-\infty}^{\infty} e^{-x^2} \, dx.$$

To be clear, this *is* mentioned in my Math 290-3 notes, but the explanation given there is actually not quite accurate, since it doesn't talk about the subtlety arising when converting improper integrals to polar coordinates. So, here we'll do it fully rigorously and correctly. The function $f(x) = e^{-x^2}$ is known as a *Gaussian* function, and the integrals of such functions are of crucial importance in probability and statistics.

Denote the required integral value by I. The observation is that to compute I we will first compute I^2 instead. To be clear, I is given by an improper integral, which means that the integral in question is really defined by the limit:

$$I = \lim_{b \to \infty} \int_{-b}^{b} e^{-x^2} \, dx,$$

and so we'll have to work with such a limit eventually. First, we write I^2 as follows:

$$I^2 = \left(\int_{-\infty}^{\infty} e^{-x^2}\,dx\right)\left(\int_{-\infty}^{\infty} e^{-y^2}\,dy\right)$$

where all we are doing is denoting the variable of integration by y in the second copy of I. The first integral is a constant with respect to the second, so we can bring it inside the second expression to get the double integral:

$$I^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-x^2}e^{-y^2}\,dx\,dy.$$

Again, to be precise this is an improper double integral, so that we are really claiming is:

$$I^2 = \lim_{b\to\infty}\int_{-b}^{b}\int_{-b}^{b} e^{-(x^2+y^2)}\,dx\,dy.$$

The form which the integrand now takes suggests that converting to polar coordinates might be nice, but the issue is that the region of integration (before taking the limit) is a square $[-b,b]\times[-b,b]$, which is not so nice to describe in terms of polar coordinates. (This is what my Math 290-3 notes do not mention at all.) We get around this as follows. Let D_b be the disk of radius b centered at the origin and $D_{b\sqrt{2}}$ the disk of radius $b\sqrt{2}$ centered at the origin. The point is that the square in question is sandwiched between these two disks, which implies that

$$\iint_{D_b} e^{-(x^2+y^2)}\,dA \leq \iint_{[-b,b]\times[-b,b]} e^{-(x^2+y^2)}\,dA \leq \iint_{D_{b\sqrt{2}}} e^{-(x^2+y^2)}\,dA.$$

Indeed, since the function we are integrating is always positive, integrating over a larger region can only make the integral itself larger and never smaller. Now, in polar coordinates the integral on the left is:

$$\int_0^{2\pi}\int_0^{b} re^{-r^2}\,dr\,d\theta = -\pi\left(e^{-b^2}-1\right)$$

and the integral on the right is

$$\int_0^{2\pi}\int_0^{b\sqrt{2}} re^{-r^2}\,dr\,d\theta = -\pi\left(e^{-2b^2}-1\right).$$

Thus

$$-\pi\left(e^{-b^2}-1\right) \leq \iint_{[-b,b]\times[-b,b]} e^{-(x^2+y^2)}\,dA \leq -\pi\left(e^{-2b^2}-1\right).$$

Now, when taking the limit as $b\to\infty$, the left side converges to π and the right side converges to π, so the squeeze theorem implies that

$$I^2 = \lim_{b\to\infty}\int_{-b}^{b}\int_{-b}^{b} e^{-(x^2+y^2)}\,dx\,dy = \pi$$

as well. Hence $I=\sqrt{\pi}$, so

$$\int_{-\infty}^{\infty} e^{-x^2}\,dx = \sqrt{\pi}$$

is our desired value. Tada!

Lectures 14 and 15: Curves and Surfaces

All the material on curves and surfaces we looked at is included in my Math 290-3 lecture notes, except for one thing which I include below. Otherwise, check those old notes for explanations and examples. Also, note that we're jumping around in the book a bit: curves are covered in Chapter 3 and surfaces are covered in Chapter 7. Curves and surfaces will serve as more general regions of integration than what we've considered up until this point.

Arclength is well-defined. Recall that for a smooth C^1 curve C with parametrization $\mathbf{x} : [a, b] \to \mathbb{R}^n$, we defined the *arclength* of C as the integral

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

In order for this to make sense, we have to know that the value obtained does not depend on the specific parametric equations chosen to describe C, or in other words that *any* choice of parametrization will give the same value as the one given by $\mathbf{x}$.

Suppose that $\mathbf{y} : [c, d] \to \mathbb{R}^n$ is another parametrization of C which is related to the parametrization above by $\mathbf{y} = \mathbf{x} \circ \tau$ for some coordinate transformation $\tau : [c, d] \to [a, b]$. If we use u to denote the parameter in the parametric equations given by $\mathbf{y}$, the idea is that τ is telling us how to express the parameter t for the $\mathbf{x}$-equations in terms of u via $t = \tau(u)$. Think of τ then as a "change of variables" transformation, and the goal is to rewrite the integral defining arclength via $\mathbf{x}$ in terms of t as the integral defining arclength via $\mathbf{y}$ in terms of u.

By the change of variables formula we have:

$$\int_{[c,d]} \left\|\mathbf{x}'(\tau(u))\right\| \, |\det D\tau(u)| \, du = \int_{\tau([c,d])} \left\|\mathbf{x}'(t)\right\| \, dt,$$

where, since $\tau([c, d]) = [a, b]$, the integral on the right is just an integral from a to b and so gives the arclength of C as determined by $\mathbf{x}$. Now, $D\tau(u)$ is a 1×1 matrix whose only entry is $\tau'(u)$, so the above expression becomes

$$\int_c^d \left\|\mathbf{x}'(\tau(u))\right\| \, |\tau'(u)| \, du = \int_a^b \left\|\mathbf{x}'(t)\right\| \, dt.$$

Since $\mathbf{y} = \mathbf{x} \circ \tau$, the chain rule gives

$$\mathbf{y}'(u) = \mathbf{x}'(\tau(u))\tau'(u).$$

Taking norms of both sides and using the fact that $\tau'(u)$ is just a scalar gives

$$\left\|\mathbf{y}'(u)\right\| = \left\|\mathbf{x}'(\tau(u))\right\| \, |\tau'(u)|.$$

Thus the integral on the left in the expression above is

$$\int_c^d \left\|\mathbf{y}'(u)\right\| \, du,$$

which is precisely the arclength of C as determined by $\mathbf{y}$. Hence

$$\int_c^d \left\|\mathbf{y}'(u)\right\| \, du = \int_a^b \left\|\mathbf{x}'(t)\right\| \, dt$$

so the value obtained for arclength is the same for $\mathbf{x}$ as it is for $\mathbf{y}$, so arclength is independent of parametrization.

Surface area is well-defined. Check my 290-3 notes for the definition of surface area. The fact that surface area is independent of parametrization is on Homework 5.

Scalar line and surface integrals. The definition of what it means to integrate a *function* over a curve or surface is included in my old notes and in the book. Similar arguments to those above show that these notions are also independent of parametrization.

Lecture 16 and 17: Vector Fields, Curl, and Divergence

A C^1 (or C^2) *vector field* on $\mathbb{R}^n$ is a C^1 (or C^2) function $\mathbb{R}^n \to \mathbb{R}^n$. Vector fields will serve as the *integrands* in the final types of integrals we'll consider. All of this material (including curl and divergence) can also be found in Math 290-3 notes, so check there for definitions, explanations, and examples. In the book this can be found in Chapter 3. The reason as to why curl and divergence indeed have the geometric interpretations described in my old notes is something we'll derive in this course later on as a consequence of Stokes' Theorem and Gauss's Theorem.

Non-conservative field of curl zero. The standard example of a vector field which is not conservative but nonetheless has curl zero is the field

$$\mathbf{F} = \frac{-y\,\mathbf{i} + x\,\mathbf{j}}{x^2 + y^2}$$

on the *punctured plane* $U := \{(x, y) \in \mathbb{R}^2 \mid (x, y) \neq (0, 0)\}$. The check that this field has curl zero is in my 290-3 notes, and is something you should work out on your own as well.

Now, you can also find in those notes the fact that the gradient of the function $\tan^{-1}\left(\frac{y}{x}\right)$ is precisely $\mathbf{F}$. This would seem to contradict the fact that $\mathbf{F}$ is not conservative on U, but the point is that the candidate potential function $\tan^{-1}\left(\frac{y}{x}\right)$ is NOT defined on all of U since it is undefined on the y-axis. So, this does not count as a valid potential function for $\mathbf{F}$ on U. It *would* count as a valid potential function over any region in $\mathbb{R}^2$ which does not intersect the y-axis.

On the other hand, this field still has curl zero on the *upper-half plane*, which is the set of points with positive y-coordinate. Since the upper-half plane is simply connected, a fact mentioned in my 290-3 notes says that $\mathbf{F}$ *should* be conservative over this region. But, $\tan^{-1}\left(\frac{y}{x}\right)$ is not defined on the entire upper-half plane, so how do we reconcile these two seemingly contradictory statements? The answer is that the function

$$-\tan^{-1}\left(\frac{x}{y}\right)$$

also has gradient equal to $\mathbf{F}$, which you should check. This function is defined on the entire upper-half plane, and so *does* serve as a potential function for $\mathbf{F}$ over this region. The upshot is that a vector field can be conservative over different regions which require different potential functions. This observation is relevant to Problem 10 on Homework 5.

Two key properties. Finally, I'll highlight two key properties of curl and divergence, which you should be able to verify yourselves as we did in class and can be found in my 290-3 notes or the book. They are the facts that:

$$\text{if } f \text{ is a } C^2 \text{ function, then } \operatorname{curl}(\nabla f) = \mathbf{0},$$

and

$$\text{if } \mathbf{F} \text{ is a } C^2 \text{ vector field, then } \operatorname{div}(\operatorname{curl} \mathbf{F}) = 0.$$

Both of these facts are reflections of Clairaut's Theorem, which if you recall was the fact that the second-order mixed partial derivatives of C^2 functions are equal.

Lecture 18: Differential Forms

Remark. Differential forms will give us a way to rephrase many properties of vector fields in a way which will make the final integration theorems we have remaining much simpler to state. Note that this topic is not covered in Math 290-3, and so will not be found in my old notes. The book covers differential forms in Chapter 8, so you can use that as a reference as well, although I'll do it with a slightly different emphasis than the book does. Everything you need to know about differential forms can be found in these current notes as they are updated.

We'll give more precise definitions of differential forms later on, but the point is that *what* differential forms exactly are is not as important as *how* to manipulation and work with them in order to make concepts in integration simpler to state. This will be our emphasis. The moral in general is that "differential k-forms are the things we integrate over k-dimensional objects".

0-forms. A *differential 0-form* on $\mathbb{R}^n$ is simply a function $f : \mathbb{R}^n \to \mathbb{R}$. We can put various adjectives in front of these terms, so that for instance a C^1 0-form is a C^1 function, a C^2 0-form is a C^2 function, and so on.

1-forms. Denote the coordinates of $\mathbb{R}^n$ by $x_1, \dots, x_n$. A *differential 1-form* on $\mathbb{R}^n$ is an expression of the form

$$f_1(x_1, \dots, x_n)\, dx_1 + \cdots + f_n(x_1, \dots, x_n)\, dx_n$$

where $f_1, \dots, f_n$ are functions $\mathbb{R}^n \to \mathbb{R}$. Requiring that these functions be C^1 (or C^2) gives the notion of a C^1 (or C^2) 1-form. For instance,

$$xye^z\, dx + \sin(xyz)\, dy - xyz^3\, dz$$

is a C^2 1-form on $\mathbb{R}^3$.

One particularly important type of 1-form is the *differential* of a function: if $f : \mathbb{R}^n \to \mathbb{R}$ is a C^2 function, the differential of f is the C^1 1-form defined by

$$df := \frac{\partial f}{\partial x_1}\, dx_1 + \cdots + \frac{\partial f}{\partial x_n}\, dx_n.$$

Note that the coefficients which are used here are the same as the components of the gradient of f, which is an observation we'll come back to.

2-forms. A *differential 2-form* on $\mathbb{R}^n$ is an expression of the form

$$\sum_{i,j} f_{ij}(\mathbf{x})\, dx_i \wedge dx_j.$$

For instance,

$$4xy\, dx \wedge dy - xz\, dx \wedge dz + (x + y)\, dy \wedge dz$$

is a 2-form on $\mathbb{R}^3$. The expression $dx_i \wedge dx_j$ is called the *wedge* product of the 1-forms dx_i and dx_j. The key algebraic property to remember is that the wedge product is *anti-commutative*, meaning

that switching the order of the differentials of which we are taking the wedge product changes the overall sign:

$$dx_j \wedge dx_i = -dx_i \wedge dx_j.$$

This implies in particular that the wedge product $dx_i \wedge dx_i$ of any differential with itself is zero:

$$dx_i \wedge dx_i = 0 \text{ since } dx_i \wedge dx_i = -dx_i \wedge dx_i.$$

Thus, anytime we see the wedge product of a differential with itself we should immediately treat that as zero.

For instance, at first glance a 2-form on $\mathbb{R}^2$ is something which looks like:

$$\omega = A\,dx \wedge dx + B\,dx \wedge dy + C\,dy \wedge dx + D\,dy \wedge dy,$$

where each term comes from one of the four possible expressions $dx_i \wedge dx_j$ when each x_i, x_j is either x or y. However, $dx \wedge dx$ and $dy \wedge dy$ are both zero, so that our expression simplifies to

$$\omega = B\,dx \wedge dy + C\,dy \wedge dx.$$

Moreover, since $dy \wedge dx = -\,dx \wedge dy$, we can further rewrite this as

$$\omega = (B - C)\,dx \wedge dy.$$

Thus the conclusion is that any 2-form on $\mathbb{R}^2$ can be written as

$$(\text{some function})\,dx \wedge dy.$$

In a similar way, any 2-form on $\mathbb{R}^3$ can be written solely in terms of $dx \wedge dy, dy \wedge dz$, and $dz \wedge dx$:

$$A\,dx \wedge dy + B\,dy \wedge dz + C\,dz \wedge dx.$$

Note that any 2-form on $\mathbb{R}^3$ is thus completely characterized by the three coefficients A, B, C—this too will play a role in some things we'll do later on.

Higher-order forms. In general, a *differential k-form* on $\mathbb{R}^n$ is an expression of the form

$$\sum_{i_1,\dots,i_k} f_{i_1,\dots,i_k}(\mathbf{x})\,dx_{i_1} \wedge \cdots \wedge dx_{i_k}$$

where each of $x_{i_1}, \dots, x_{i_k}$ are among the variables $x_1, \dots, x_n$. In other words, a k-form is an expression involving sums of wedge products of k differentials. For instance,

$$e^{xyz}\,dx \wedge dy \wedge dz$$

is a 3-form on $\mathbb{R}^3$. A 3-form on $\mathbb{R}^2$ would involve wedge products of three differentials using dx and dy alone, such as something like

$$x\,dx \wedge dy \wedge dx + y\,dy \wedge dx \wedge dy.$$

However, note that any such expression must always have a repeated dx or dy, meaning that any such expression must be zero since $dx \wedge dx = 0$ and $dy \wedge dy = 0$. To be clear, using the anti-commutative property we have:

$$dx \wedge dy \wedge dx = -dx \wedge dx \wedge dy = -0 \wedge dy = 0,$$

where in the first step we flipped dx and dy in order to get the two dx terms adjacent, and such a flip changes the overall sign. Thus, the only 3-form on $\mathbb{R}^2$ is zero.

More generally, any k-form on $\mathbb{R}^n$ where $k > n$ is always zero since such a k-form is made up of only the n differentials $dx_1, \ldots, dx_n$, so that at least one these differentials will be repeated. Also, any n-form on $\mathbb{R}^n$ can always be written as

$$f(x_1, \ldots, x_n)\, dx_1 \wedge dx_2 \wedge \cdots dx_n$$

since any other ordering of these differentials in this wedge product can also be put into this specific ordering using the anti-commutative properties. For instance, any 3-form on $\mathbb{R}^3$ looks like

$$f(x, y, z)\, dx \wedge dy \wedge dz,$$

and so is completely characterized by a single coefficient function.

Changing variables. One of the main reasons why differential forms are useful is that they make change of variables operations very simple to state. For instance, consider the 2-form

$$dx \wedge dy$$

on $\mathbb{R}^2$. We want to rewrite this in terms of polar coordinates, so in terms of the differentials dr and $d\theta$. This is very simple to do: all we do is take the polar change of coordinates

$$x = r\cos\theta \quad \text{and} \quad y = r\sin\theta$$

and substitute them into $dx \wedge dy$:

$$dx \wedge dy = d(r\cos\theta) \wedge d(r\sin\theta).$$

Now, each of $d(r\cos\theta)$ and $d(r\sin\theta)$ is the differential of a function, which we described how to compute previously:

$$d(r\cos\theta) = \frac{\partial(r\cos\theta)}{\partial r}\, dr + \frac{\partial(r\cos\theta)}{\partial\theta}\, d\theta = \cos\theta\, dr - r\sin\theta\, d\theta$$

and

$$d(r\sin\theta) = \frac{\partial(r\sin\theta)}{\partial r}\, dr + \frac{\partial(r\sin\theta)}{\partial\theta}\, d\theta = \sin\theta\, dr + r\cos\theta\, d\theta.$$

This just comes from the fact that the coefficient of dr comes from the partial derivative with respect to r and the coefficient of $d\theta$ comes from the partial with respect to θ.

We make these substitutions and "distribute" the wedge product:

$$\begin{aligned} d(r\cos\theta) \wedge d(r\sin\theta) &= (\cos\theta\, dr - r\sin\theta\, d\theta) \wedge (\sin\theta\, dr + r\cos\theta\, d\theta) \\ &= \cos\theta\sin\theta\, dr \wedge dr + r\cos^2\theta\, dr \wedge d\theta - r\sin^2\theta\, d\theta \wedge dr - r^2\sin\theta\cos\theta\, d\theta \wedge d\theta. \end{aligned}$$

Now, in this resulting expression, $dr \wedge dr = 0$ and $d\theta \wedge d\theta = 0$, so we omit these terms:

$$dx \wedge dy = r\cos^2\theta\, dr \wedge d\theta - r\sin^2\theta\, d\theta \wedge dr.$$

Finally, since $d\theta \wedge dr = -dr \wedge d\theta$, we can rewrite this final expression as:

$$dx \wedge dy = (r\cos^2\theta + r\sin^2\theta)\, dr \wedge d\theta = r\, dr \wedge d\theta,$$

which is the required polar form of $dx \wedge dy$. In general, the same procedure applies when writing a differential form expressed in terms of one set of coordinates as a form expressed in another set: we simply make the substitutions for the variables, compute the resulting differentials, and use properties of wedge products to simplify what we get.

The thing which should jump out at you is that the coefficient r obtained in

$$dx \wedge dy = r\, dr \wedge d\theta$$

is precisely the Jacobian expansion factor we get when converting from rectangular to polar coordinates in double integrals. This is no accident, and will in fact be true in general. Indeed, the upshot is that the change of variables formula in integration is more succinctly statement in terms of differential forms since this approach has the Jacobian factors "built" in, meaning that they automatically pop-out simply from the definitions of differentials and wedge products. We'll talk more about this next time.

Lecture 19: More on Differential Forms

Warm-Up. We rewrite the 3-form $dx \wedge dy \wedge dz$ on $\mathbb{R}^3$ in terms of spherical coordinates. Recall that

$$x = \rho \sin\phi \cos\theta,\ y = \rho \sin\phi \sin\theta,\ z = \rho \cos\phi.$$

Thus:

$$\begin{aligned}
dx &= \frac{\partial x}{\partial \rho}\, d\rho + \frac{\partial x}{\partial \phi}\, d\phi + \frac{\partial x}{\partial \theta}\, d\theta = \sin\phi\cos\theta\, d\rho + \rho\cos\phi\cos\theta\, d\phi - \rho\sin\phi\sin\theta\, d\theta \\
dy &= \frac{\partial y}{\partial \rho}\, d\rho + \frac{\partial y}{\partial \phi}\, d\phi + \frac{\partial y}{\partial \theta}\, d\theta = \sin\phi\sin\theta\, d\rho + \rho\cos\phi\sin\theta\, d\phi + \rho\sin\phi\cos\theta\, d\theta \\
dz &= \frac{\partial z}{\partial \rho}\, d\rho + \frac{\partial z}{\partial \phi}\, d\phi + \frac{\partial x}{\partial \theta}\, d\theta = \cos\phi\, d\rho - \rho\sin\phi\, d\phi.
\end{aligned}$$

Now we take the wedge product of the three resulting expressions and use the distributive property to expand what we get. But recall that any such triple wedge product expression which involves the same $d\rho, d\phi, d\theta$ repeated more than once will automatically be zero. For instance, the term

$$\sin^2\phi\cos\theta\sin\theta\cos\phi\, d\rho \wedge d\rho \wedge d\rho$$

we get by wedging together the first piece of each of dx, dy, dz is zero. The only nonzero terms we get in the expression for $dx \wedge dy \wedge dz$ come from taking one term from dx, a term from dy corresponding to a *different* differential than the one we used from dx, and then a term from dz corresponding to a differential differential than the one we used for dx and dy. We get:

$$\begin{aligned}
dx \wedge dy \wedge dz = &-\rho^2 \sin^3\phi\cos^2\theta\, d\rho \wedge d\theta \wedge d\phi + \rho^2 \sin\phi\cos^2\phi\cos^2\theta\, d\phi \wedge d\theta \wedge d\rho \\
&+ \rho^2 \sin^3\phi\sin^2\theta\, d\theta \wedge d\rho \wedge d\phi - \rho^2 \sin\phi\cos^2\phi\sin^2\theta\, d\theta \wedge d\phi \wedge d\rho.
\end{aligned}$$

as the only nonzero terms.

Finally, using the anti-commutative property of wedge products (namely that switching adjacent differentials changes the sign), each of the resulting 3-forms can be written in terms of $d\rho \wedge d\phi \wedge d\theta$:

$$\begin{aligned}
d\rho \wedge d\theta \wedge d\phi &= -d\rho \wedge d\phi \wedge d\theta \\
d\rho \wedge d\theta \wedge d\rho &= d\rho \wedge d\phi \wedge d\theta
\end{aligned}$$

$$d\theta \wedge d\rho \wedge d\phi = d\rho \wedge d\phi \wedge d\theta$$
$$d\theta \wedge d\phi \wedge d\rho = -d\rho \wedge d\phi \wedge d\theta.$$

With this in mind, we have:

$$\begin{aligned} &dx \wedge dy \wedge dz \\ &= (\rho^2 \sin^3\phi \cos^2\theta + \rho^2 \sin\phi \cos^2\phi \cos^2\theta + \rho^2 \sin^3\phi \sin^2\theta + \rho^2 \sin\phi \cos^2\phi \sin^2\theta)\, d\rho \wedge d\phi \wedge d\theta, \end{aligned}$$

which after using $\sin^2 + \cos^2 = 1$ a few times simplifies to

$$dx \wedge dy \wedge dz = \rho^2 \sin\phi \, d\rho \wedge d\phi \wedge d\theta.$$

Thus, as claimed last time, the coefficient $\rho^2 \sin\phi$ ends up being precisely the Jacobian expansion factor we get when converting from rectangular to spherical coordinates in triple integrals.

Pullbacks. In general, let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a coordinate transformation, denote the coordinates of the domain by $u_1, \dots, u_n$ and those of the codomain by $x_1, \dots, x_n$. The components of T tell us how to express the x_i in terms of the u_j:

$$T(u_1, \dots, u_n) = (x_1(u_1, \dots, u_n), \dots, x_n(u_1, \dots, u_n)).$$

The result is that this change of variables gives:

$$dx_1 \wedge \cdots \wedge dx_n = (\det DT(\mathbf{u}))\, du_1 \wedge \cdots \wedge du_n,$$

generalizing what we saw for polar coordinates last time and for spherical coordinates in the Warm-Up. You'll verify this on the homework.

A general n-form on $\mathbb{R}^n$ looks like

$$\omega = f(\mathbf{x})\, dx_1 \wedge \cdots \wedge dx_n.$$

The n-form obtained by rewriting this in terms of $u_1, \dots, u_n$ using T is called the *pullback* of ω by T and is denoted by $T^*\omega$. Thus, the result above implies that the expression for the pullback is

$$T^*(f(\mathbf{x})\, dx_1 \wedge \cdots \wedge dx_n) = f(T(\mathbf{u}))(\det DT(\mathbf{u}))\, du_1 \wedge \cdots \wedge du_n.$$

Lower-order forms can be pulled-back as well; for instance:

$$T^*(dx_i) = d(x_i(u_1, \dots, u_n)) = \frac{\partial x_i}{\partial u_1}\, du_1 + \cdots + \frac{\partial x_i}{\partial u_n}\, du_n.$$

Again, all that pulling-back a form does is rewrite it in terms of some new coordinates.

Integration of forms. For an n-dimensional region $E \subseteq \mathbb{R}^n$, we define the *integral* of the n-form $\omega = f(\mathbf{x})\, dx_1 \wedge \cdots \wedge dx_n$ over E to be:

$$\int_E \omega := \int_E f(\mathbf{x})\, d\mathbf{x}.$$

In other words, the integral of an n-form is simply the integral of the coefficient function $f(\mathbf{x})$ as we've already defined earlier this quarter. To be clear, to integrate an n-form, we first write

the n-form in terms of $dx_1 \wedge \cdots \wedge dx_n$, in that specific ordering, and then integrate the resulting coefficient function.

With this in mind, we now note what the change of variables formula looks like in terms of differential forms. Recall that this formula says:

$$\int_D f(T(\mathbf{u}))|\det DT(\mathbf{u})|\,d\mathbf{u} = \int_{T(D)} f(\mathbf{x})\,d\mathbf{x}$$

where T is some coordinate transformation. The right side is simply the integral of the differential form

$$\omega := f(\mathbf{x})\,dx_1 \wedge \cdots \wedge dx_n$$

over $T(D)$. The integrand on the left side is *almost* the expression for the pullback $T^*\omega$, except that this pullback uses $\det DT(\mathbf{u})$ itself instead of its absolute value. Thus, the left side is

$$\pm \int_D T^*\omega,$$

where we have a $+$ when $\det DT(\mathbf{u})$ is always positive, and we have a $-$ when $\det DT(\mathbf{u})$ is always negative. (Note that saying T is a "coordinate transformation" means that it is C^1, injective, and has invertible Jacobian everywhere, so that $\det DT(\mathbf{u})$ is never zero and hence the C^1 condition implies that $\det DT(\mathbf{u})$ is either always positive or always negative since it is a continuous expression.)

Thus, the change of variables formula looks like:

$$\pm \int_D T^*\omega = \int_{T(D)} \omega.$$

For some better terminology, we say that T is *orientation-preserving* when $\det DT(\mathbf{u})$ is always positive, and T is *orientation-reversing* when $\det DT(\mathbf{u})$ is always negative. The upshot is that when integrating forms, integrating a form over the image of a region is the same as integrating its pullback over the original region, with either a $+$ or $-$ sign depending on whether T preserves or reverses orietnation.

Exterior derivatives. The final operation on differential forms we need is that of *exterior differentiation*. In general, taking the exterior derivative of a k-form results in a $(k+1)$-form, so exterior differentiation increases order by 1.

To start with, the exterior derivative of a 0-form (i.e. function) f is the differential df. In general, the exterior derivative of the k-form

$$\omega := f(\mathbf{x})\,dx_{i_1} \wedge \cdots \wedge dx_{i_k}$$

is the $(k+1)$-form

$$d\omega := df \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_k}.$$

Thus, we take the differential of the coefficient function and wedge the result with the rest of the form. For forms expressed as sums of individual wedge products, the exterior derivative operator d acts linearly, so that it satisfies

$$d(\omega + \eta) = d\omega + d\eta.$$

For example, let us work out the exterior derivative of the 2-form on $\mathbb{R}^4$ given by

$$\omega = x_1 x_3^2 x_4\,dx_1 \wedge dx_2 + x_1^3 x_2 x_3^3\,dx_3 \wedge dx_4.$$

First, by linearity:

$$d\omega = d(x_1x_3^2x_4\,dx_1\wedge dx_2 + x_1^3x_2x_3^3\,dx_3\wedge dx_4) = d(x_1x_3^2x_4\,dx_1\wedge dx_2) + d(x_1^3x_2x_3^3\,dx_3\wedge dx_4).$$

Then:

$$d(x_1x_3^2x_4\,dx_1\wedge dx_2) = d(x_1x_3^2x_4)\wedge dx_1\wedge dx_2$$

and

$$d(x_1^3x_2x_3^3\,dx_3\wedge dx_4) = d(x_1^3x_2x_3^3)\wedge dx_3\wedge dx_4.$$

The differential of $x_1x_3^2x_4$ involves a dx_1 term, a dx_3 term, and a dx_4 term, but no dx_2 term since the partial of $x_1x_3^2x_4$ with respect to x_2 is zero. However, the dx_1 term after wedging with $dx_1\wedge dx_2$ becomes zero, so only the dx_3 and dx_4 terms actually matter. We get:

$$\begin{aligned} d(x_1x_3^2x_4)\wedge dx_1\wedge dx_2 &= (x_3^2x_3\,dx_1 + 2x_1x_3x_4\,dx_3 + x_1x_3^2\,dx_4)\wedge dx_1\wedge dx_2 \\ &= 2x_1x_3x_4\,dx_3\wedge dx_1\wedge dx_2 + x_1x_3^2\,dx_4\wedge dx_1\wedge dx_2. \end{aligned}$$

Similarly, we get:

$$\begin{aligned} d(x_1^3x_2x_3^3)\wedge dx_3\wedge dx_4 &= (3x_1^2x_2x_3^3\,dx_1 + x_1^3x_3^3\,dx_2 + 3x_1^3x_2x_3^2\,dx_3)\wedge dx_3\wedge dx_4 \\ &= 3x_1^2x_2x_3^3\,dx_1\wedge dx_3\wedge dx_4 + x_1^3x_3^3\,dx_2\wedge dx_3\wedge dx_4. \end{aligned}$$

Thus, to summarize, the exterior derivative of $\omega = x_1x_3^2x_4\,dx_1\wedge dx_2 + x_1^3x_2x_3^3\,dx_3\wedge dx_4$ is

$$\begin{aligned} d\omega = {} & 2x_1x_3x_4\,dx_3\wedge dx_1\wedge dx_2 + x_1x_3^2\,dx_4\wedge dx_1\wedge dx_2 \\ & + 3x_1^2x_2x_3^3\,dx_1\wedge dx_3\wedge dx_4 + x_1^3x_3^3\,dx_2\wedge dx_3\wedge dx_4. \end{aligned}$$

Note that the result is indeed a 3-form, and that this can be rewritten in various ways using the anti-commutativity of the wedge product.

Curl and divergence recast. To see why we care about exterior derivatives, let us work out explicitly the exterior derivative of any 1-form and any 2-form on $\mathbb{R}^3$. First, take a 1-form

$$\omega = P\,dx + Q\,dy + R\,dz$$

where P, Q, R are functions. Then

$$d\omega = dP\wedge dx + dQ\wedge dy + dR\wedge dz.$$

Now, only the dy and dz components of dP will give something nonzero after wedging with dx, so

$$dP\wedge dx = \frac{\partial P}{\partial y}\,dy\wedge dx + \frac{\partial P}{\partial z}\,dz\wedge dx.$$

Similarly, only the dx and dz components of dQ matter, and only the dx and dy components of dR matters. Overall we get:

$$\begin{aligned} d\omega &= dP\wedge dx + dQ\wedge dy + dR\wedge dz \\ &= \frac{\partial P}{\partial y}\,dy\wedge dx + \frac{\partial P}{\partial z}\,dz\wedge dx + \frac{\partial Q}{\partial x}\,dx\wedge dy + \frac{\partial Q}{\partial z}\,dz\wedge dy + \frac{\partial R}{\partial x}\,dx\wedge dz + \frac{\partial R}{\partial y}\,dy\wedge dz \\ &= \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right)dy\wedge dz + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right)dz\wedge dx + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)dx\wedge dy. \end{aligned}$$

The observation is that the coefficients showing up here are precisely the components in the curl of the vector field $P\,\mathbf{i} + Q\,\mathbf{j} + R\,\mathbf{k}$. Indeed, as we'll make clear next time, there is a way to translate differential 1-forms into vector fields and vice-versa, and a way to translate 2-forms on $\mathbb{R}^3$ into vector fields and vice-versa, and under these correspondences the exterior derivative of a 1-form to a 2-form *is* the same thing as taking the curl!

Now, take a 2-form

$$\eta = A\,dy \wedge dz + B\,dz \wedge dx + C\,dx \wedge dy.$$

Then

$$d\eta = dA \wedge dy \wedge dz + dB \wedge dz \wedge dx + dC \wedge dx \wedge dy.$$

Only the dx component of dA gives something nonzero after wedging with $dy \wedge dz$, and similarly only the dy component of dB and only the dz component of dC matter. We get:

$$d\eta = \frac{\partial A}{\partial x}\,dx \wedge dy \wedge dz + \frac{\partial B}{\partial y}\,dy \wedge dz \wedge dx + \frac{\partial C}{\partial z}\,dz \wedge dx \wedge dy.$$

Using the anti-commutative properties, this can all be written as

$$d\eta = \left(\frac{\partial A}{\partial x} + \frac{\partial B}{\partial y} + \frac{\partial C}{\partial z}\right)\,dx \wedge dy \wedge dz.$$

The observation now is that the coefficient of $d\eta$ is precisely the divergence of the vector field $A\,\mathbf{i} + B\,\mathbf{j} + C\,\mathbf{k}$. Thus, under the correspondence between 2-forms on $\mathbb{R}^3$ and vector fields and between 3-forms on $\mathbb{R}^3$ and functions, the exterior derivative of a 2-form to a 3-form *is* the same thing as taking the divergence!

The upshot is that the vector field operations of curl and divergence, and of gradient, are all instances of the *same* operation on differential forms, namely that of taking the exterior derivative. The difference comes in what order form we apply this operation to. Again, we'll elaborate on this more next time, and together with the succinct rephrasing of the change of variables formula in terms of differential forms is why differential forms will give an elegant way of phrasing the final BIG THEOREMS we'll soon see.

Lecture 20: Line Integrals

Warm-Up. We compute $d^2 f := d(df)$ for a C^2 0-form f (i.e. function) on $\mathbb{R}^3$, and $d^2\alpha := d(d\alpha)$ for a C^2 1-form α on $\mathbb{R}^3$. First,

$$df = \frac{\partial f}{\partial x}\,dx + \frac{\partial f}{\partial y}\,dy + \frac{\partial f}{\partial z}\,dz.$$

Then:

$$\begin{aligned}
d^2 f &= d(df) \\
&= d\left(\frac{\partial f}{\partial x}\,dx\right) + d\left(\frac{\partial f}{\partial y}\,dy\right) + d\left(\frac{\partial f}{\partial z}\,dz\right) \\
&= d\left(\frac{\partial f}{\partial x}\right) \wedge dx + d\left(\frac{\partial f}{\partial y}\right) \wedge dy + d\left(\frac{\partial f}{\partial z}\right) \wedge dz \\
&= \left(\frac{\partial^2 f}{\partial y \partial x}\,dy + \frac{\partial^2 f}{\partial z \partial x}\,dz\right) \wedge dx + \left(\frac{\partial^2 f}{\partial x \partial y}\,dx + \frac{\partial^2 f}{\partial z \partial y}\,dz\right) \wedge dy + \left(\frac{\partial^2 f}{\partial x \partial z}\,dx + \frac{\partial^2 f}{\partial y \partial z}\,dy\right) \wedge dz
\end{aligned}$$

$$= \left(\frac{\partial^2 f}{\partial y \partial z} - \frac{\partial^2 f}{\partial z \partial y}\right) dy \wedge dz + \left(\frac{\partial^2 f}{\partial z \partial x} - \frac{\partial^2 f}{\partial x \partial z}\right) dz \wedge dx + \left(\frac{\partial^2 f}{\partial x \partial y} - \frac{\partial^2 f}{\partial y \partial x}\right) dx \wedge dy$$
$$= 0,$$

where in the fourth line we omit the dx term from $d\left(\frac{\partial f}{\partial x}\right)$ since after wedging with dx this will give zero anyway and similarly we omit the dy term from $d\left(\frac{\partial f}{\partial y}\right)$ and the dz term from $d\left(\frac{\partial f}{\partial z}\right)$, in the fifth line we use anti-commutativity of the wedge product to combine terms, and in the final line we use Clairaut's Theorem to say that the mixed second-order partial derivatives of f are zero since f is C^2. Hence applying d twice to a C^2 0-form results in zero.

Now suppose that

$$\alpha = P\,dx + Q\,dy + R\,dz$$

where each of P, Q, R are C^2. We have:

$$d\alpha = \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right) dy \wedge dz + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial z}\right) dz \wedge dx + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx \wedge dy$$

using a computation we went through last time, so:

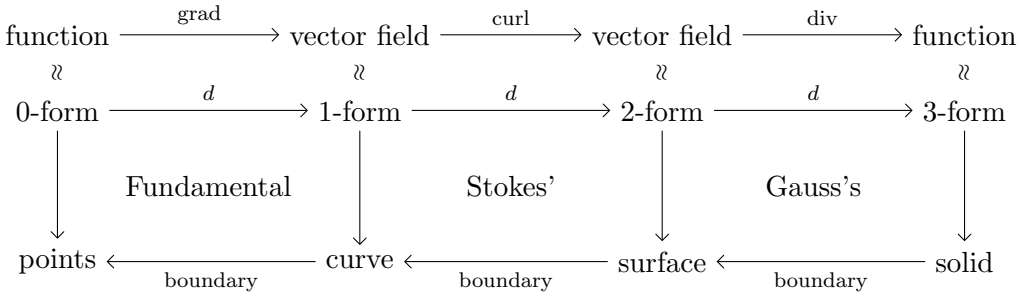
$$\begin{aligned} d^2\alpha &= d(d\alpha) \\ &= d\left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right) \wedge dy \wedge dz + d\left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right) \wedge dz \wedge dx + d\left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) \wedge dx \wedge dy \\ &= (R_{yx} - Q_{zx})\,dx \wedge dy \wedge dz + (P_{zy} - R_{xy})\,dy \wedge dz \wedge dx + (Q_{xz} - P_{yz})\,dz \wedge dx \wedge dy \\ &= (R_{yx} - Q_{zx} + P_{zy} - Rxy + Q_{xz} - P_{yz})\,dx \wedge dy \wedge dz \\ &= 0, \end{aligned}$$

where again we use Clairaut's Theorem at the end to say that the second-order mixed partials of P, Q, R are equal. Hence applying d twice to a C^2 1-form also results in zero.

This is no accident: you'll show on the homework that the same is true for any C^2 k-form in general. In the $\mathbb{R}^3$ case, recalling the relation between d and gradient, curl, and divergence we alluded to last time, the fact that $d^2 f = 0$ for a 0-form is a reflection of the fact that $\text{curl}(\nabla f) = \mathbf{0}$, and the fact that $d^2\alpha = 0$ for a 1-form is a reflection of the face that $\text{div}(\text{curl}\,\mathbf{F}) = 0$. The point is that these seemingly-different properties of gradient, curl, and divergence are really to seen to be the *same* property when phrased in terms of differential forms.

THE BIG PICTURE. Before moving on, I want to describe what I will call the BIG PICTURE of vector calculus, which gives the correct framework from which to view everything we are about to do. *Vector calculus* is essentially the study of the calculus of vector fields, where the main important concepts are those dealing with integrals of vector fields over curves and surfaces, which are *line* and *surface* integrals respectively. Here we will see some far-reaching generalizations of the single-variable Fundamental Theorem of Calculus, including the Fundamental Theorem of Line Integrals, Green's Theorem, Stokes' Theorem, and Gauss's Theorem (also known as the Divergence Theorem). I will refer to these as the BIG THEOREMS of vector calculus. The point is that, although these theorems might at first glance appear to say different things, they are in reality a reflection of the *same* underlying concept, which is that how a vector field behaves over the boundary of some object is directly related to how the "derivative" of the field behaves throughout the object itself. The idea that these BIG THEOREMS are really the same will become even clearer once we see how to phrase them in terms of differential forms.

Here, then, is the BIG PICTURE to have in mind:



The two rows on top describe the types of things we integrate, and the bottom row describes the types of things we integrate over. The horizontal arrows in the two rows on top describe a type of derivative, and the horizontal arrows in the bottom row describe the operation of taking a boundary. Finally, the vertical arrows describe a type of integral, and each of the squares describe one of our BIG THEOREMS. Let us unpack all this.

The things we integrate. First, we have alluded to the idea previously that there is a way translate back and forth between functions and vector fields on the one hand and differential forms on the other. To be sure, everything we say here takes place in $\mathbb{R}^3$. A 0-form on $\mathbb{R}^3$ is simply a function by definition. Now, a 1-form on $\mathbb{R}^3$ looks like

$$P\,dx + Q\,dy + R\,dz.$$

The key observation is that a 1-form is determined by three component functions P, Q, R, which is the same type of data needed to specify a vector field on $\mathbb{R}^3$. Thus, to this 1-form we can associate the vector field $P\,\mathbf{i} + Q\,\mathbf{j} + R\,\mathbf{k}$, and vice-versa given a vector field we can construct from it a 1-form using the same coefficient functions:

$$P\,dx + Q\,dy + R\,dz \longleftrightarrow P\,\mathbf{i} + Q\,\mathbf{j} + R\,\mathbf{k}.$$

This gives a 1-to-1 correspondence between 1-forms and vector fields, which is what the vertical $\approx$ symbol denotes in the BIG PICTURE relating "vector field" in the first row to "1-form" in the second. The first vertical $\approx$ symbol gives the correspondence between 0-forms and functions, which again in this is just by the definition of a 0-form.

Now, a 2-form on $\mathbb{R}^3$ looks like

$$A\,dy \wedge dz + B\,dz \wedge dx + C\,dx \wedge dy.$$

Again, this is characterized by three coefficient functions A, B, C, so to this we can also associate a vector field, and vice-versa:

$$A\,dy \wedge dz + B\,dz \wedge dx + C\,dx \wedge dy \longleftrightarrow A\,\mathbf{i} + B\,\mathbf{j} + C\,\mathbf{k}.$$

This correspondence between 2-forms on $\mathbb{R}^3$ and vector fields accounts for the third vertical $\approx$ symbol in the BIG PICTURE. Now well how this works: the coefficient of $dy \wedge dz$ gives the $\mathbf{i}$-component of the corresponding vector field, which we can remember by noting that it is the dx term which is missing from $dy \wedge dz$; similarly, dy is missing from $dz \wedge dx$ so its coefficient gives the $\mathbf{j}$-component of the associated vector field, and dz is missing from $dx \wedge dy$ so its coefficient gives the $\mathbf{k}$-component. Moreover, the specific ordering we used in these wedge products is important, since for instance $A\,dz \wedge dy$ would give $-A\mathbf{i}$ instead of $A\mathbf{i}$ since $dz \wedge dy = -dy \wedge dz$. To remember the ordering, use the usual ordering of x, y, z, only once you get to the end start back at the beginning;

thus, we have $dx \wedge dy$ since y comes after x, we have $dy \wedge dz$ since z comes after y, and we have $dz \wedge dx$ instead of $dx \wedge dz$ since x comes "after" z after we cycle from z at the end back to the beginning. The reason why we use this specific ordering will become clear after we talk about surface integrals.

Finally, a 3-form on $\mathbb{R}^3$ looks like

$$f(x, y, z)\, dx \wedge dy \wedge dz,$$

so this is characterized by the single function $f(x, y, z)$ alone. Vice-versa, given a function we can associate to it a unique 3-form:

$$f(x, y, z)\, dx \wedge dy \wedge dz \longleftrightarrow f(x, y, z),$$

and this is the correspondence between 3-forms and functions given by the final vertical $\approx$ symbol.

Next we explain the horizontal arrows in the first two rows. Moving from left to right, "grad" denotes the operation of taking the gradient which sends a function to a vector field, "curl" denotes the operation of taking a curl which sends a vector field to a vector field, and "div" denotes the operation of taking a divergence which sends a vector field to a function. In the differential forms row, "d" denotes the exterior derivative operation, which sends a 0-from to a 1-form, a 1-form to a 2-form, and a 2-form to a 3-form. The point here is that each "d" occurring in this row is nothing but a way to rephrase the corresponding grad,curl,div operation above it. Indeed, we worked this out last time: under the correspondence between functions/vector fields and 0- and 1-forms, the operation of taking the gradient becomes the operation of taking a differential; under the correspondence between vector fields and 1- and 2-forms, the operation of taking a curl becomes the exterior operation applied to 1-forms; and under the corresponding between vector fields/functions and 2- and 3-forms, the operation of taking a divergence becomes the exterior derivative applied to a 2-form. As we mentioned last the time, the fact that these three operations (grade,curl,div) can be viewed as reflections of the *same* operation on differential forms is one of the main reasons why differential forms will make various things simpler to state. Note that following two successive horizontal arrows in either of these first two rows always results in zero: in the differential form row these is the statement from the Warm-Up that $d^2 = 0$, while in the first row this is the pair of facts that $\operatorname{curl}(\nabla f) = \mathbf{0}$ and $\operatorname{div}(\operatorname{curl} \mathbf{F}) = 0$.

The things we integrate over. Now on to the bottom row, which denote the types of things we will integrate over: collections of points, curves, surfaces, and 3-dimensional solid regions in $\mathbb{R}^3$. The horizontal arrow here, which you should notice moves from right to left, denotes the operation of taking a *boundary*. The rightmost horizontal arrow

$$\text{surface} \longleftarrow \text{solid}$$

says that taking the boundary of a 3-dimensional solid gives a surface; for instance, the boundary of a solid ball $x^2 + y^2 + z^2 \leq R^2$ is the sphere $x^2 + y^2 + z^2 = R^2$, and in general the boundary of the solid enclosed by a surface S gives S itself. Next, the horizontal arrow

$$\text{curve} \longleftarrow \text{surface}$$

says that taking the boundary of a surface gives a curve.

In the simplest case, the boundary of a 2-dimensional region in $\mathbb{R}^2$ (i.e. a "flat" surface) is a curve in $\mathbb{R}^2$; for instance, the boundary of a disk is a circle. However, for other types of surfaces we have to be more careful about what we mean by boundary. The point is that the term "boundary"

used here does NOT necessarily mean boundary in the sense we defined last quarter, but rather means *manifold boundary*. (The type of "boundary" we defined last quarter is often called the *topological boundary* in order to distinguish it from the manifold boundary.) For solids in $\mathbb{R}^3$, the topological and manifold boundaries are the same, and for flat 2-dimensional surfaces in $\mathbb{R}^2$, these two notions of boundaries are also the same. However, for more general surfaces there is a difference. We will define the notion of "manifold boundary" later when we discuss Stokes' Theorem. The leftmost horizontal arrow

$$\text{points} \longleftarrow \text{curve}$$

says that taking the boundary of a curve gives some points. Again, the notion of boundary being used here is really that of the manifold boundary, but in the case of curves this is simple to define: the boundary of a curve is simply the set consisting of its endpoints. A *closed* curve which begins and ends at the same place has empty boundary in this sense.

I'll point out one more special thing about this bottom row: following two successive horizontal arrows results in nothing! This says that if we start with a solid and take its boundary, the resulting boundary surface itself has no boundary, and if we start with a surface, take its boundary to get a curve, this resulting boundary curve is always closed and so has no boundary. We will denote these boundaries using the same ∂ notation we used last quarter, only keeping in mind that now this really means manifold boundary. The fact that taking two boundaries in a row gives an empty set is then summarized by saying that

$$\partial^2 = \emptyset,$$

and it is NO accident that this looks eerily similar to the fact that $d^2 = 0$ for differential forms. Indeed, in general, the operation of taking a boundary can be viewed as a type of "derivative" operation, and we'll see glimpses of why this is in our BIG THEOREMS.

The types of integrals. The vertical arrows in the BIG PICTURE denote integrals. Indeed, the point in general is that k-forms are the things we integrate over k-dimensional objects. In the $\mathbb{R}^3$ case, this says that 3-forms (i.e. functions) are the things we integrate over solids, 2-forms (i.e. vector fields) are the things we integrate over surfaces, 1-forms (i.e. vector fields) are the things we integrate over curves, and 0-forms (i.e. functions) are the things we integrate over collections of points. The rightmost vertical arrow denotes the ordinary triple integral of a function over a solid; the next vertical arrow denotes a *surface integral*, which is what we get when we integrate a vector field (or 2-form) over surface; the next vertical arrow denotes a *line integral*, which is what we get when we integrate a vector field (or 1-form) over a curve. Finally, the leftmost vertical arrow is actually something we've seen all of our lives, and simply denotes *evaluating* a function (or 0-form) on a collection of points. But, from the point of view of the BIG PICTURE, we will think of this evaluation as also being a type of "integral", and we'll see precisely what we mean by this when we discuss the Fundamental Theorem of Line Integrals.

Finally, the squares. Each square in the BIG PICTURE describes the ingredients involved in one of the BIG THEOREMS. The corresponding theorem relates the two types of integrals showing up on its vertical sides, where the relation between the "integrands" in these integrals is described by the horizontal edge on top of the square and the relation between the "regions of integration" is described by the horizontal edge on the bottom of the square. So, the Fundamental Theorem of Line Integrals relates evaluating a function on the boundary points of a curve to the integral of the gradient of the function over the curve, Stokes' Theorem (of which Green's Theorem is a special case) relates the integral of a vector field over the boundary curve of a surface to the integral of its curl over the surface itself, and Gauss's (Divergence) Theorem relates the integral of vector field over the boundary surface of a solid to the integral of its divergence over that solid itself. The BIG

PICTURE makes explicit what major ingredients (integrands, regions of integration, derivatives) contribute to each BIG THEOREM. Expressed in terms of differential forms, I claim that the equality

$$\int_{\partial M} \omega = \int_M d\omega,$$

whatever it means, encapsulates all BIG THEOREMS at once. Our goal is to make this all clear.

Line integrals. We begin our journey by first considering *vector line integrals*, which are integrals of vector fields over curves. Suppose C is a smooth, C^1 curve in $\mathbb{R}^n$ and $\mathbf{F} : C \to \mathbb{R}^n$ a C^1 vector field on C. We assume in addition that C is *oriented*, meaning that we have chosen a specific direction for the tangent vectors along C. (Or, in other words, we have chosen a specific direction in which C is traced out.) The *line integral* of $\mathbf{F}$ over C is defined to be:

$$\int_C \mathbf{F} \cdot d\mathbf{s} := \int_C (\mathbf{F} \cdot \mathbf{T})\, ds$$

where $\mathbf{T}$ denotes the *unit tangent vector* along C in the direction of the specified orientation. To clarify the notation, the left hand side is simply the notation we use for vector line integrals, and the right hand side is the *scalar* line integral (as defined in a previous lecture or on Problem 10 of Homework 4) of the *function* $\mathbf{F} \cdot \mathbf{T}$, namely the function which assigns to each point $\mathbf{p}$ on C the value of the dot product $\mathbf{F}(\mathbf{p}) \cdot \mathbf{T}(\mathbf{p})$ of the vector field value at $\mathbf{p}$ with the unit tangent vector at $\mathbf{p}$. The idea is that we are adding up the value of all these dot products as we vary along C.

The point is that this line integrals gives a way to measure the extent to which C moves "with" or "against" the flow of $\mathbf{F}$. To be precise, at each $\mathbf{p} \in C$, $\mathbf{F}(\mathbf{p}) \cdot \mathbf{T}(\mathbf{p})$ measures the extent to which $\mathbf{F}$ and $\mathbf{T}$ point in the same general direction: this dot product is positive when the angle between $\mathbf{F}(\mathbf{p})$ and $\mathbf{T}(\mathbf{p})$ is less than $\frac{\pi}{2}$ (so $\mathbf{F}$ and $\mathbf{T}$ point in "similar directions") while this dot product is negative when the angle between these vectors is greater than $\frac{\pi}{2}$ (so $\mathbf{F}$ and $\mathbf{T}$ points in "opposite" directions.) Check my Math 290-3 notes for some pictures which illustrate this dot product, and the sense in which this line integrals measures what I claimed it measures above. Also check my 290-3 notes for plenty of examples.

Next time we'll see how to compute such line integrals using parametric equations, and how to rephrase such line integrals in terms of differential forms instead.

Lecture 21: More on Line Integrals

Today we spoke more about line integrals and how to compute them. The examples we looked at, including the Warm-Up, and more can be found in my Math 290-3 notes, so I encourage to look there for explicit computations. Here I'll just summarize the key points and describe how to rephrase all this in terms of differential forms.

Line integrals via parametrizations. Suppose C is a smooth C^1 oriented curve in $\mathbb{R}^n$ with parametrization $\mathbf{x} : [a, b] \to \mathbb{R}^n$, and let $\mathbf{F} : C \to \mathbb{R}^n$ be a C^1 vector field on C. Recall that the line integral of $\mathbf{F}$ over $\mathbb{C}$ is

$$\int_C \mathbf{F} \cdot d\mathbf{s} := \int_C (\mathbf{F} \cdot \mathbf{T})\, ds$$

where $\mathbf{T}$ is the unit tangent vector along C. One thing to clarify is why we require taking a unit tangent vector in this definition: this guarantees that the value we get depends only on $\mathbf{F}$ and the curve C, but not on how fast we our traveling along this curve. In other words, if we had just said to take any tangent vector, then this does not give a well-defined expression since describing

tangent vectors concretely requires the use of a parametrization, and different parametrizations can give different tangent vectors. Using the unit tangent vector guarantees that our answer will be independent of parametrization, so that it really only depends on C itself.

Now, in terms of our parametrization, we have

$$\mathbf{T}(\mathbf{x}(t)) = \frac{\mathbf{x}'(t)}{\|\mathbf{x}'(t)\|}.$$

Thus:

$$\int_C (\mathbf{F} \cdot \mathbf{T})\, ds = \int_a^b \left(\mathbf{F}(\mathbf{x}(t)) \cdot \frac{\mathbf{x}'(t)}{\|\mathbf{x}'(t)\|} \right) \|\mathbf{x}'(t)\|\, dt = \int_a^b \mathbf{F}(\mathbf{x}(t)) \cdot \mathbf{x}'(t)\, dt,$$

where the extra $\|\mathbf{x}'(t)\|$ term comes from the expression for $ds = \|\mathbf{x}'(t)\|\, dt$. This final integral gives us a way to compute vector line integrals concretely in terms of parametrizations. Again, we know that the value we get is actually independent of parametrization via the way in which we defined vector line integrals in terms of scalar line integrals and unit tangent vectors.

Integrating 1-forms. Now we see how to rephrase vector line integrals in terms of differential forms. For simplicity, suppose $\mathbf{F} = (P, Q, R)$ is a C^1 vector field on $\mathbb{R}^3$ and that

$$\omega = P\, dx + Q\, dy + R\, dz$$

is the corresponding 1-form. Let $\mathbf{x} : [a, b] \to \mathbb{R}^3$ be a parametrization of a smooth C^1 oriented curve C in $\mathbb{R}^n$. We define the integral of ω over C to be:

$$\int_C \omega := \int_a^b (P(\mathbf{x}(t))x'(t) + Q(\mathbf{x}(t))y'(t) + R(\mathbf{x}(t))z'(t))\, dt,$$

which is *precisely* the same as

$$\int_C \mathbf{F} \cdot d\mathbf{s} = \int_a^b \mathbf{F}(\mathbf{x}(t)) \cdot \mathbf{x}'(t)\, dt = \int_a^b (P(\mathbf{x}(t)), Q(\mathbf{x}(t)), R(\mathbf{x}(t))) \cdot (x'(t), y'(t), z'(t))\, dt.$$

Thus, the upshot is that the line integral of a 1-form over a curve is nothing but the vector line integral of the associated vector field over that curve.

Let us understand this better. The parametric equations

$$x = x(t),\ y = y(t),\ z = z(t)$$

give us a way to write the differentials dx, dy, dz in the expression for ω in terms of dt instead:

$$dx = x'(t)\, dt \quad dy = y'(t)\, dt \quad dz = z'(t)\, dt.$$

With these expressions, we get

$$\begin{aligned} P\, dx + Q\, dy + R\, dz &= P(\mathbf{x}(t))x'(t)\, dt + Q(\mathbf{x}(t))y'(t)\, dt + R(\mathbf{x}(t))z'(t)\, dt \\ &= (P(\mathbf{x}(t))x'(t) + Q(\mathbf{x}(t))y'(t) + R(\mathbf{x}(t))z'(t))\, dt, \end{aligned}$$

and this final expression is precisely what we integrate in the definition of $\int_C \omega$. Thus, the notation

$$\int_C P\, dx + Q\, dy + R\, dz$$

suggests to simply use given parametric equations to express the integrand $P\,dx + Q\,dy + R\,dz$ solely in terms of t by taking P times the derivative of the x-equation, Q times the derivative of the y-equation, and R times the derivative of the z-equation, and then integrate the result over the values of t in our parametrization.

Previously we referred to the process of substituting in for dx, dy, dz the corresponding expressions in terms of dt using $x = x(t), y = y(t), z = z(t)$ as *pulling back* ω by the function $\mathbf{x} : [a, b] \to \mathbb{R}^n$. Thus, we can summarize the discussion above by saying that the integral of ω over C is defined simply by pulling ω back to $[a, b]$ using $\mathbf{x}$, and then integrating the result over $[a, b]$ instead:

$$\int_{[a,b]} \mathbf{x}^*\omega = \int_{C=\mathbf{x}([a,b])} \omega,$$

where we interpret $\mathbf{x}$ as a "change of variables" between $C \subseteq \mathbb{R}^n$ and $[a, b]$. The fact that this is well-defined, meaning independent of parametrization, comes down to the ways in which differential forms behave under change of variables, which we'll elaborate on next time. Alternatively, you can go through the vector field definition to argue this is indeed well-defined.

What are 1-forms? We can now finally give some meaning as to what a 1-form is. Here is a definition: a differential 1-form is a linear mapping which takes vector fields as inputs and outputs functions, meaning a type of linear transformation from the real vector space of vector fields on $\mathbb{R}^n$ to the real vector space of real-valued functions on $\mathbb{R}^n$. In the case of $\mathbb{R}^3$, dx denotes the mapping which sends a vector field to its **i**-component function, dy the mapping which sends a vector field to its **j**-component, and dz the mapping which sends a vector field to its **k**-component:

$$dx(A, B, C) = A \quad dy(A, B, C) = B \quad dz(A, B, C) = C.$$

Thus, $P\,dx$ is the mapping which sends a vector field to P times its **i**-component, and similarly for $Q\,dy$ and $R\,dz$:

$$(P\,dx + Q\,dy + R\,dz)(A, B, C) = P\,dx(A, B, C) + Q\,dy(A, B, C) + R\,dz(A, B, C) = PA + QB + RC.$$

In particular, given a parametrization $\mathbf{x}(t)$ of a curve C, $\mathbf{x}'(t)$ gives an example of a vector field along C, in which case for $\omega = P\,dx + Q\,dy + R\,dz$ we have:

$$\omega(\mathbf{x}'(t)) = (P\,dx + Q\,dy + R\,dz)(x'(t), y'(t), z'(t)) = P(\mathbf{x}(t))x'(t) + Q(\mathbf{x}(t))y'(t) + R(\mathbf{x}(t))z'(t),$$

which is precisely the thing we integrate in the definition of $\int_C \omega$. Thus, this integral can be written as:

$$\int_C \omega = \int_a^b \omega(\mathbf{x}'(t))\,dt,$$

meaning $\int_C \omega$ is defined by taking the function $\omega(\mathbf{x}'(t))$ obtained by evaluating ω on the tangent vector field $\mathbf{x}(t)$ itself, and integrating *that* over $[a, b]$. In this notation, the pullback of ω by $\mathbf{x}$ is

$$\mathbf{x}^*\omega = \omega(\mathbf{x}'(t))\,dt.$$

So, this all gives meaning to what it means to integrate a 1-form over a curve, and meaning to what a 1-form itself actually is. However, note two things: first, when doing explicit computations, we always fall back to using parametric equations; and second, the definition of line integrals via differential forms says nothing about what they mean geometrically. Indeed, this is the point of view we will take going forward: differential forms will give us good ways of phrasing and remembering

definitions and properties of line integrals, but the geometric insight into what they mean comes from the vector field approach. Thus, both the vector field approach and differential forms approach to line (and later surface) integrals are important to keep in mind for different reasons. Also, even though we have now given some meaning as to what 1-forms actually are, note that still what they *are* is not as important as how we *work* with them. Said another way, the definition for what differential forms are are motivated by wanting to give a definition that would fit the types of manipulations we want to be able to do with them.

Lecture 22: The Fundamental Theorem of Line Integrals

Line integrals are well-defined. Of course, we have shown previously that line integrals are well-defined, meaning independent of parametrization, using the vector field approach. Let us now see what this would look like in terms of differential forms. The key point is that pulling a form back via a function already has the necessary "change of variables" expansion factors baked into it. So, suppose ω is a C^1 form and $\mathbf{x} : [a, b] \to \mathbb{R}^n$ and $\mathbf{y} : [c, d] \to \mathbb{R}^n$ are two parametrizations of a smooth C^1 oriented curve C, related by $\mathbf{y} = \mathbf{x} \circ \tau$ via some coordinate transformation $\tau : [c, d] \to [a, b]$. The claim is that

$$\int_{[c,d]} \mathbf{y}^*\omega = \int_{[a,b]} \mathbf{x}^*\omega,$$

where the left side is the definition of $\int_C \omega$ as determined by $\mathbf{y}$ and the right side is the definition of $\int_C \omega$ as determined by $\mathbf{x}$. This equality says that the value of $\int_C \omega$ is independent of parametrization as desired.

Now, the integral

$$\int_{[a,b]=\tau([c,d])} \mathbf{x}^*\omega$$

is the single-variable integral of the single-variable function $\mathbf{x}^*\omega = \omega(\mathbf{x}'(t))$. The change of variables formula for integrals expressed in terms of differential forms gives

$$\pm \int_{[c,d]} \tau^*(\mathbf{x}^*\omega) = \int_{[a,b]=\tau([c,d])} \mathbf{x}^*\omega,$$

where the sign on the left depends on whether or not τ is orientation-preserving (meaning $\tau' > 0$) or orientation-reversing (meaning $\tau' < 0$). However, $\mathbf{x}$ and $\mathbf{y}$ are meant to give the *same* orientation on C, meaning that the tangent vectors determined by either always point in the same direction. Problem 9 of Homework 4 shows that this happens if and only if $\tau'(u) > 0$ for all $u \in [c, d]$, we indeed get that τ is orientation-preserving and thus

$$\int_{[c,d]} \tau^*(\mathbf{x}^*\omega) = \int_{[a,b]=\tau([c,d])} \mathbf{x}^*\omega.$$

Something I won't explain in full detail but should seem plausible is that the result of pulling ω back via $\mathbf{x}$ and then pulling the result back via τ is the same as pulling ω back by the composition $\mathbf{x} \circ \tau$; this is essentially a phrasing of the chain rule expressed in terms of differential forms. Thus we have

$$\int_{[c,d]} (\mathbf{x} \circ \tau)^*\omega = \int_{[a,b]} \mathbf{x}^*\omega.$$

Since $\mathbf{x} \circ \tau = \mathbf{y}$, we get the desired equality.

These manipulations seem a little "magical", and it might still not be clear exactly what they all mean. Don't worry about the full details: all this is meant to do is point out how "clean" the proof that line integrals are well-defined looks like when phrased in terms of differential forms.

Fundamental Theorem. Before looking at the Warm-Up, let us give the statement of the so-called *Fundamental Theorem of Line Integrals*, which will be useful in the Warm-Up. The claim is as follows:

> Suppose f is a C^1 function defined on some open set containing the smooth C^1 oriented curve C. Then
>
> $$\int_C \nabla f \cdot d\mathbf{s} = f(\text{end point of } C) - f(\text{start point of } C).$$

Thus, line integrals of conservative fields are easy to compute: we simply evaluate the potential function at the endpoints of the curve and subtract those values. This is meant to be a direct analog of the ordinary Fundamental Theorem of Calculus:

$$\int_a^b f'(x)\,dx = f(b) - f(a).$$

Indeed, if you take C to be a curve in $\mathbb{R}$ (i.e. a line segment $[a, b]$) and f to be a single-variable function, the gradient of f looks like $\nabla f = f'(x)\,i$, so that in this special the Fundamental Theorem of Line Integrals *is* a statement of the Fundamental Theorem of Calculus. We'll give a proof the Fundamental Theorem after the Warm-Up.

Warm-Up. We compute

$$\int_C (e^y + y^2 + 1)\,dx + (xe^y + 2xy + \cos y + x)\,dy$$

where C is the left half of the circle $(x-1)^2 + y^2 = 1$ oriented clockwise from $(1, -1)$ to $(1, 1)$. Although this is possible to setup using parametric equations, the resulting integral is not so simple to compute directly. Instead, we can find a way to compute this using the statement of the Fundamental Theorem.

The key observation is that, although the vector field

$$\mathbf{F} = (e^y + y^2 + 1)\,\mathbf{i} + (xe^y + 2xy + \cos y + x)\,\mathbf{j}$$

is not conservative (or equivalently the 1-form $(e^y + y^2 + 1)\,dx + (xe^y + 2xy + \cos y + x)\,dy$ is not exact), it is *almost* conservative in the sense that we can write is as something conservative plus something simpler:

$$\mathbf{F} = \nabla f + x\,\mathbf{j}.$$

where

$$f(x, y) = xe^y + xy^2 + x + \sin y.$$

In differential form notation this says

$$(e^y + y^2 + 1)\,dx + (xe^y + 2xy + \cos y + x)\,dy = df + x\,dy.$$

The point is that the integral in question becomes

$$\int_C \nabla f \cdot d\mathbf{s} + \int_C x\mathbf{j} \cdot d\mathbf{s}, \text{ or } \int_C df + \int_C x\,dy,$$

where the first term in either expression can be computed using the Fundamental Theorem and the second term using parametric equations.

We get

$$\int_C df = f(1,1) - f(1,-1) = e + \frac{1}{e} + 2\sin 1$$

as a consequence of the Fundamental Theorem of Line Integrals. Using the parametrization

$$x = 1 + \cos t, \quad y = -\sin t, \quad \frac{\pi}{2} \le t \le \frac{3\pi}{2}$$

for C, we get:

$$\int_C x\,dy = \int_{\pi/2}^{3\pi/2} (1+\cos t)(-\cos t)\,dt = -\int_{\pi/2}^{3\pi/2} (\cos t + \cos^2 t)\,dt = \frac{1}{2}(4-\pi).$$

Thus all together we get

$$\int_C (e^y + y^2 + 1)\,dx + (xe^y + 2xy + \cos y + x)\,dy = e + \frac{1}{e} + 2\sin 1 + 2 - \frac{\pi}{2}$$

as the desired value.

Proof of Fundamental Theorem. Pick a parametrization $\mathbf{x} : [a,b] \to \mathbb{R}^n$ of C. Then

$$\int_C \nabla f \cdot d\mathbf{s} = \int_a^b \nabla f(\mathbf{x}(t)) \cdot \mathbf{x}'(t)\,dt.$$

By the multivariable chain rule, we have

$$\frac{d}{dt} f(\mathbf{x}(t)) = \nabla f(\mathbf{x}(t)) \cdot \mathbf{x}'(t),$$

so

$$\int_a^b f(\mathbf{x}(t)) \cdot \mathbf{x}'(t)\,dt = \int_a^b \frac{d}{dt} f(\mathbf{x}(t))\,dt = f(\mathbf{x}(b)) - f(\mathbf{x}(a))$$

where in the second equality we use the single-variable Fundamental Theorem of Calculus. Since $\mathbf{x}(b)$ gives the end point of C and $\mathbf{x}(a)$ the starting point, this gives

$$\int_C \nabla f \cdot d\mathbf{s} = f(\text{end point of } C) - f(\text{start point of } C)$$

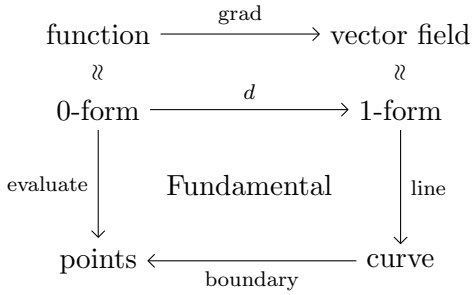
as required.

In differential form notation, this proof would like:

$$\begin{aligned}
\int_C df &= \int_C \frac{\partial f}{\partial x_1}\,dx_1 + \cdots + \frac{\partial f}{\partial x_n}\,dx_n \\
&= \int_a^b \left[\frac{\partial f}{\partial x_1}(\mathbf{x}(t)) x_1'(t) + \cdots \frac{\partial f}{\partial x_n}(\mathbf{x}(t)) x_n'(t) \right] dt \\
&= \int_a^b Df(\mathbf{x}(t))\mathbf{x}'(t)\,dt \\
&= \int_a^b \frac{d}{dt} f(\mathbf{x}(t))\,dt \\
&= f(\mathbf{x}(b)) - f(\mathbf{x}(a)),
\end{aligned}$$

where again we use the chain rule and single-variable Fundamental Theorem of Calculus. □

Back to BIG PICTURE. I claim that the Fundamental Theorem explains the first square of the BIG PICTURE, which looked like:



Indeed, the Fundamental Theorem relates the integrals characterizing both vertical arrows to one another. To be clear, take a 0-form f, apply the horizontal arrow d to get df, and then follow the vertical arrow to get the expression

$$\int_C df.$$

The square above says this should be equal to the "integral" expression defined by the remaining arrows: the vertical arrow on the left and the horizontal arrow on the bottom. The horizontal arrow on the bottom turns C into its boundary ∂C, which is the set containing the end point and starting point of C:

$$\partial C = \{\text{end point}, \text{start point}\}.$$

Thus, the integral characterizing the left vertical arrow should be an "integral" taking place over this set of points.

We *define* the integral of f over $\{\text{end point}, \text{start point}\}$ to be the expression:

$$\int_{\partial C} f = f(\text{end point}) - f(\text{start point}).$$

In other words, we are defining what it means to integrate a function over a 0-dimensional collection of points to be simply what you get when you evaluate that function on those points and subtract the resulting values. Although it might seem to call such an expression an "integral", it makes from the point of view in general that integration should be viewed as a type of summation. With this notation, the Fundamental Theorem says that

$$\int_{\partial C} f = \int_C df,$$

which, as claimed, relates the integrals showing up on both vertical edges of the first square in the BIG PICTURE. Thus, how f behaves over the boundary of C is intimately related to how df behaves over C. Again, view this a generalization of the single-variable Fundamental Theorem of Calculus: f is an "antiderivative" of df, and the left side is what you get when you "evaluate" this antiderivative on ∂C.

The point, to jump ahead a bit, is that the remaining squares in the BIG PICTURE follow the exact same formula. In general, if ω is a k-form and M a k-dimensional region in $\mathbb{R}^n$, it will be true that

$$\int_{\partial M} \omega = \int_M d\omega,$$

whatever this notation means. We won't look at this in the full generality of $\mathbb{R}^n$, but will instead focus on the case of $\mathbb{R}^3$. In that case, when ω is a 1-form and M is a surface, the equality above

is precisely the statement of Stokes' Theorem, and when ω is a 2-form and M a solid, it is the statement of Gauss's Theorem. Although we'll at first phrase these BIG THEOREMS in terms of vector fields, it is the formulation in terms of differential forms which makes it clear that they are all the "same" theorem. We'll get to all this soon enough.

Properties of Conservative Fields. We now point out some key properties which conservative fields have as a consequence of the Fundamental Theorem. They are the following:

- if C is a *closed* curve, meaning one which starts and ends at the same point, then

$$\oint_C \nabla f \cdot d\mathbf{s} = 0.$$

  (The notation $\oint$ means the same thing as $\int$, and is used simply to denote the fact that we are taking the line integral over a closed curve.)

- if C_1 and C_2 are curves which start at the same point and end at the same point, then

$$\int_{C_1} \nabla f \cdot d\mathbf{s} = \int_{C_2} \nabla f \cdot d\mathbf{s}.$$

  We refer to this say saying that line integrals of ∇f are *path independent* in the sense that the value of a line integral depends only on the endpoints of a path and not on the particular curve we choose to connect those two points.

The first property is a consequence of the fact that $f(\text{end point}) = f(\text{start point})$ for a closed curve, while the second is a consequence of the fact that $f(\text{end point}) = f(\text{start point})$ is the *same* expression for C_1 as it is for C_2.

In fact, these two properties are equivalent to one another: a vector field $\mathbf{F}$ has the property that its line integral over any closed curve is zero if and only if its line integrals are path independent, as we'll show in the Warm-Up next time. Moreover, these properties in fact *imply* that $\mathbf{F}$ is conservative, so conservative fields are the only ones which these properties. We'll say something about this next time as well, and it will also be on the homework.

Curl zero but not conservative. Finally, we go back to something we mentioned previously and for which we can now give adequate justification: the vector field

$$\mathbf{F} = \frac{-y\,\mathbf{i} + x\,\mathbf{j}}{x^2 + y^2}$$

defined on the punctured plane ($\mathbb{R}^2$ with the origin removed) has curl zero everywhere but is not conservative. We showed earlier that $\operatorname{curl} \mathbf{F}$ is indeed zero, but up until now we did not have a way to show that $\mathbf{F}$ is not conservative on the punctured plane. The key observation is that the line integral of $\mathbf{F}$ over the unit circle is $\pm 2\pi$:

$$\int_{\text{unit circle}} \mathbf{F} \cdot d\mathbf{s} = \pm 2\pi$$

where the $\pm$ depends on which orientation we use. This comes from a direct computation using parametric equations $\mathbf{x}(t) = (\cos t, \sin t), 0 \leq t \leq 2\pi$. The point is that, if $\mathbf{F}$ were conservative, its line integral over the unit circle would have to be zero (since the line integral of a conservative field over any closed curve is zero), so since this specific line integral is not zero, $\mathbf{F}$ cannot be conservative. We'll come back to this example when discussing Green's Theorem, where the fact that $\mathbf{F}$ is not defined everywhere on the region enclosed by the unit circle will have crucial consequences.

Lecture 23: Green's Theorem

Warm-Up. We showed in the Warm-Up that a vector field $\mathbf{F}$ has the property that its line integral over any closed curve is zero if and only if its line integrals are path-independent. This can be found in my Math 290-3 notes as the Warm-Up for the day covering "Parametric Surfaces". The key observation is that we can break up a closed curve into two curves which start at the same point and end at the same point. Check the notes (or the book) for the details.

Path-independence implies conservative. We have seen that conservative fields have the property that their line integrals are path-independent, or equivalently that their line integrals over closed curves are always zero. In fact, it is true that *only* conservative fields have these properties, meaning that if a field has either of these (equivalent) properties, then it must be conservative.

Here is the idea. Suppose for simplicity that $\mathbf{F} = P\,\mathbf{i} + +Q\,\mathbf{j}$ is a C^1 vector field on $\mathbb{R}^2$ with the property that its line integrals are path-independent. Define the function $f : \mathbb{R}^2 \to \mathbb{R}$ by setting, for any $(x, y) \in \mathbb{R}^2$, $f(x, y)$ to be the value

$$f(x, y) := \int_{(0,0)}^{(x,y)} P\,dx + Q\,dy,$$

where the notation on the right is the line integral taken over *any* curve which starts at $(0, 0)$ and ends at (x, y). The fact that line integrals of $\mathbf{F}$ are path-independent guarantees that any such curve will give the same value for the line integral in question, so that $f(x, y)$ is well-defined, meaning that its value does not depend on which curve we take from $(0, 0)$ to (x, y).

The claim is that f is a potential function for $\mathbf{F}$, meaning that $\nabla f = \mathbf{F}$ and thus showing that $\mathbf{F}$ is indeed conservative. In other words, the claim is that the partial derivatives of f are given by:

$$\frac{\partial f}{\partial x} = P \quad \text{and} \quad \frac{\partial f}{\partial y} = Q.$$

The idea is to come up with expressions for $f(x, y)$ by choosing especially nice curves from $(0, 0)$ to (x, y), in a way which will let us easily see what the partial derivatives of f are. For instance, consider first the curve consisting of the vertical line segment C_1 from $(0, 0)$ to $(0, y)$ followed by the horizontal line segment C_2 from $(0, y)$ to (x, y). With these we get

$$f(x, y) = \int_{C_1} (P\,dx + Q\,dy) + \int_{C_2} (P\,dx + Q\,dy).$$

Now, we can parametrize C_1 and C_2 using

$$\mathbf{x}_1(t) = (0, t),\ 0 \le t \le y \quad \text{and} \quad \mathbf{x}_2(t) = (t, y),\ 0 \le t \le x$$

respectively. With these, we get

$$\int_{C_1} P\,dx + Q\,dy = \int_0^y Q(0, t)\,dy \quad \text{and} \quad \int_{C_2} P\,dx + Q\,dy = \int_0^x P(t, y)\,dt,$$

so

$$f(x, y) = \int_0^y Q(0, t)\,dy + \int_0^x P(t, y)\,dt.$$

Now we can differentiate: the first term does not depend on x, so its derivative with respect to x is zero, while the derivative of the second term with respect to x is $P(x, y)$ by the Fundamental Theorem of Calculus. Thus

$$\frac{\partial f}{\partial x}(x, y) = P(x, y)$$

as desired.

Computing $\frac{\partial f}{\partial y}$ will require the use of a different curve connecting $(0,0)$ to (x,y), and the details are left to the homework. Note that if we differentiate the expression for $f(x,y)$ we derived above:

$$f(x,y) = \int_0^y Q(0,t)\,dy + \int_0^x P(t,y)\,dt$$

with respect to y would give

$$\frac{\partial f}{\partial y}(x,y) = Q(0,y) + \frac{\partial}{\partial y}\left(\int_0^x P(t,y)\,dt\right).$$

This is not so simple to manipulate further to see that it equals $Q(x,y)$, which is why we need to derive a different expression for $f(x,y)$ using a different curve if we want to have any hope of showing that $\frac{\partial f}{\partial y} = Q$.

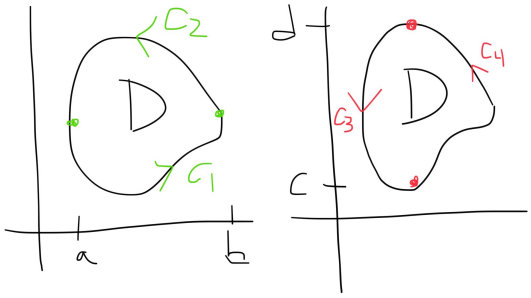
Green's Theorem. In the most basic sense, Green's Theorem gives a way to write certain line integrals in terms of double integrals instead. This relation between line and double integrals may seem to come out of nowhere, but it is really at the core of all BIG THEOREMS we'll see, and when viewed in the right light is just a generalization of the Fundamental Theorem of Calculus. Here is the statement:

> Suppose $D \subseteq \mathbb{R}^2$ is a compact region whose boundary ∂D consists of finitely many, closed, piecewise C^1 curves. Orient the boundary ∂D so that, when walking along it, the region D is on left side. Then if $\mathbf{F} = P\,\mathbf{i} + Q\,\mathbf{j}$ is a C^1 vector field on D, we have
>
> $$\oint_{\partial D} P\,dx + Q\,dy = \iint_D (Q_x - P_y)\,dA.$$

Thus, as stated earlier, this expresses a certain line integral in terms of a double integral, and vice-versa. Often times we want the compute the thing on the left and it turns out to be simpler to compute the thing on the right, but sometimes we'll want to compute the thing on the right and it will be simpler to compute the thing on the left, as we'll see.

Some remarks are in order. First, by boundary here we just mean boundary in the sense of last quarter, or what we'll now call the topological boundary. The assumption is that this boundary consists of closed curves. Note that, the "finitely many" is needed, since although the boundary of, say, a unit disk is a single circle, the boundary of the region lying *between* two circles (what is called an *annulus*) consists of *both* of those circles and so has two "pieces". Second, the orientation needed on this boundary in order so that Green's Theorem holds is usually called the *positive* or *induced* orientation. We'll see in examples what this orientation really looks like. If ∂D has the wrong orientation, Green's Theorem still applies by changing the sign of the double integral on the right. Finally, it is crucial that $\mathbf{F}$ be C^1 on all of D and not only on the boundary of D. This subtlety is at the core of one of the most important uses of Green's Theorem, which we'll look at next time. The $Q_x - P_y$ showing up in the double integral will pop out naturally from the proof, and it is no accident that this looks like the type of thing which shows up in $\operatorname{curl}\mathbf{F}$; we'll get a sense as to why next time.

Proof in a special case. We will prove Green's Theorem only in a special case, namely the case where D is a region with the property that its upper and lower boundaries can each be described using single equations of the form $y = g(x)$, and its leftmost and rightmost boundaries can also each be described using single equations of the form $x = h(y)$:



(This is what book calls a Type III region back when discussing double integrals.) The line integral in question is then:

$$\oint_{\partial D} P\,dx + Q\,dy = \oint_{\partial D} P\,dx + \oint_{\partial D} Q\,dy.$$

To compute the integral of $P\,dx$, we parametrize the bottom and top pieces C_1 and C_2 of ∂D respectively using

$$\mathbf{x}_1(x) = (x, g_1(x)),\ a \le x \le b \quad \text{and} \quad \mathbf{x}_2(x) = (x, g_2(x)),\ a \le x \le b.$$

Note that, actually, $\mathbf{x}_2$ parametrizes C_2 with the wrong orientation, so we correct for this by including an extra negative sign:

$$\int_{\partial D} P\,dx = \int_{C_1} P\,dx - \int_{C_2} P\,dx.$$

With the given parametric equations we get:

$$\int_{\partial D} P\,dx = \int_a^b P(x, g_1(x))\,dx - \int_a^b P(x, g_2(x))\,dx = \int_a^b [P(x, g_1(x)) - P(x, g_2(x))]\,dx.$$

Now, here comes the magic: by the Fundamental Theorem of Calculus, we can rewrite the integrand in the resulting integral as:

$$P(x, g_1(x)) - P(x, g_2(x)) = -\int_{g_1(x)}^{g_2(x)} P_y(x, y)\,dy.$$

(Indeed, we are just using

$$f(\alpha) - f(\beta) = \int_\beta^\alpha f'(y)\,dy$$

in the case where $f(y) = P(x, y)$ with a fixed x where only y varies.) Thus, with this substitution we get:

$$\int_{\partial D} P\,dx = \int_a^b \int_{g_1(x)}^{g_2(x)} -P_y(x, y)\,dy\,dx = \iint_D -P_y(x, y)\,dy\,dx.$$

Thus we have written the line integral of $P\,dx$ over ∂D as a double integral over D.

To compute the integral of $Q\,dy$ we instead break up ∂D into the curves making up the leftmost and rightmost boundaries. The left and right pieces C_3 and C_4 of ∂D can be parametrized respectively using

$$\mathbf{x}_3(y) = (h_1(y), y),\ c \le y \le d \quad \text{and} \quad \mathbf{x}_4(y) = (h_2(y), y),\ c \le y \le d.$$

Note that this gives the wrong orientation on C_3, so we correct with a negative sign:

$$\int_{\partial D} Q\,dy = -\int_{C_3} Q\,dy + \int_{C_4} Q\,dy$$

. With the given parametrizations we get:

$$\int_{\partial D} Q\,dy = \int_c^d -Q(h_1(y), y)\,dy + \int_c^d Q(h_2(y), y)\,dy = \int_c^d [Q(h_2(y), y) - Q(h_1(y), y)]\,dy.$$

By the Fundamental Theorem of calculus we have:

$$Q(h_2(y), y) - Q(h_1(y), y) = \int_{h_1(y)}^{h_2(y)} Q_x(x, y)\,dx,$$

so

$$\int_{\partial D} Q\,dy = \int_c^d \int_{h_1(y)}^{h_2(y)} Q_x(x, y)]\,dx\,dy = \iint_D Q_x(x, y)\,dA.$$

Hence all together we get:

$$\oint_{\partial D} P\,dx + Q\,dy = \iint_D -P_y(x, y)\,dA + \iint_D Q_x(x, y)\,dA = \iint_D (Q_x - P_y)\,dA$$

as claimed by Green's Theorem. □

Observations. The proof above only works in the special case we mentioned, where D is particularly "nice". The idea for the proof in the more general setting is to break up a general compact region into such "nice" pieces, apply the special case to each of these pieces, and then add everything up together. The details are not overly complicated, but I'll omit them here. The book says something about this if you're interested.

A more important observation is the use of the Fundamental Theorem of Calculus as the thing which allows us to introduce an extra integration along the way by rewriting a difference as an integral. This, of course, is why we are able to turn a (single) line integral into a double integral, and is the source of where the Q_x and P_y terms come from. This same exact idea is at the heart of the proof of Stokes' Theorem and of Gauss's Theorem, which we have yet to look at. We'll see next time that phrasing Green's Theorem in terms of differential forms will it clearer as to where the $Q_x - P_y$ comes from.

Examples. Check the book and my 290-3 notes for plenty of examples of how to use Green's Theorem, both to convert line integrals is more easily computable double integrals, and for converting double integrals into more easily computable line integrals, so that Green's Theorem really works "both ways". One thing to note in particular is that line integrals such as

$$\int_{\partial D} x\,dy, \quad \int_{\partial D} -y\,dx, \quad \frac{1}{2}\int_{\partial D} -y\,dx + x\,dy$$

give a way to compute *areas* using line integrals, which is surprising: why should the vector fields $x\,\mathbf{j}, -y\,\mathbf{i}$, or $\frac{1}{2}(-y\,\mathbf{i} + x\,\mathbf{j})$ have anything to do with somehow "detecting" areas? The answer, of course, is Green's Theorem.

Lecture 24: More on Green's Theorem

Warm-Ups. We looked at two Warm-Ups, highlighting certain non-obvious uses of Green's Theorem. These were: using Green's Theorem for a non-closed curve by "closing" it off, and using Green's Theorem to replace a complicated curve by a simpler curve in the case of a field of curl zero. Both of these uses are outlined in my "Notes on Green's Theorem" available here: http://math.northwestern.edu/∼scanez/courses/290/notes/greens-thrm.pdf. This latter usage also showed up in Problem 9 of Homework 7, and is really quite an important technique to know.

Rephrase in terms of forms. Consider the differential 1-form $\omega = P\,dx + Q\,dy$ on $\mathbb{R}^2$. Then

$$d\omega = \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx \wedge dy.$$

To integrate this 2-form over a region D of $\mathbb{R}^2$ means simply to integrate the function $\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}$ over this region, so in this notation Green's Theorem looks like:

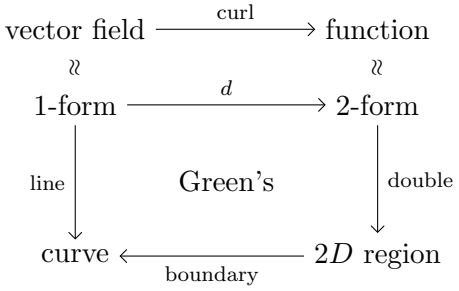
$$\int_{\partial D} \omega = \int_D d\omega.$$

(Again, to be clear, the integral on the right is really a double integral over D.) The point of rewriting Green's Theorem in this way is to make clear the relation between Green's Theorem and the Fundamental Theorem of Line Integrals; we saw previously that the statement Fundamental Theorem also looks like

$$\int_{\partial C} \alpha = \int_C d\alpha,$$

only in this case α is a 0-form and C a curve. In this way, Green's Theorem and the Fundamental Theorem (and our remaining BIG THEOREMS) say the "same" thing.

Back to BIG PICTURE. The middle square of our BIG PICTURE in the 2-dimensional case looks like:



The point is that Green's Theorem is what relates the different sides of this square to one another: start with a 1-form ω, follow the horizontal arrow on top to get a 2-form $d\alpha$, then follow the vertical arrow to get an integral of $d\alpha$ over a 2-dimensional region in the xy-plane, and the claim is that the result is the same as the one obtained by following the vertical arrow on the left to integrate α over the curve obtained by taking the boundary (i.e. following the bottom horizontal arrow) of the given region.

Some points of clarification are in order. The type of vector field being considered here is a 2-dimensional vector field of the form $\mathbf{F} = P(x, y)\,\mathbf{i} + Q(x, y)\,\mathbf{j}$. The curl of such a field looks like

$$\operatorname{curl} \mathbf{F} = (Q_x - P_y)\,\mathbf{k},$$

and so this curl is completely characterized by the coefficient function $Q_x - P_y$. This explains what is going on in the topmost horizontal arrow from vector fields (in $\mathbb{R}^2$) to functions. The correspondence between 2-forms and functions given at the upper right comes from the fact that any 2-form on $\mathbb{R}^2$ looks like

$$f(x, y)\, dx \wedge dy,$$

and so is completely characterized by the coefficient function $f(x, y)$.

More generally, in $\mathbb{R}^3$, this middle square will explain Stokes' Theorem, which is the generalization of Green's Theorem to arbitrary surfaces which aren't necessarily "flat" and constrained to lie in the xy-plane. We will see that Green's Theorem is just a special case of Stokes' Theorem, namely the case where the surface in question is a flat surface in the xy-plane

Geometric meaning of curl. As pointed out above, the integrand $Q_x - P_y$ showing up in Green's Theorem is the $\mathbf{k}$-component of the curl of the 2-dimensional vector field $\mathbf{F} = P\,\mathbf{i} + Q\,\mathbf{j}$. to be clear, to compute the curl of this we must interpret it as a 3-dimensional field, which we do by taking the $\mathbf{k}$-component to be zero: $\mathbf{F} = (P, Q, 0)$ where P, Q only depend on x and y. Take the product of $\operatorname{curl} \mathbf{F} = (Q_x - P_y)\,\mathbf{k}$ with $\mathbf{k}$ itself gives $Q_x - P_y$, so that we can rewrite Green's Theorem in vector field notation as:

$$\int_{\partial D} \mathbf{F} \cdot d\mathbf{s} = \iint_D \operatorname{curl} \mathbf{F} \cdot \mathbf{k}\, dA.$$

Thus, Green's Theorem relates on one hand the integral of $\mathbf{F}$ over the boundary of a region to the integral of its curl over the region itself. This will be made clearer after we talk about Stokes' Theorem.

With Green's Theorem at hand, we can now justify the geometric interpretation of curl (at least in the 2-dimensional case) we described earlier. The claim was that for a 2-dimensional field $\mathbf{F} = (P, Q)$, its curl, or more precisely the $\mathbf{k}$-component $Q_x - P_y$ of its curl, measures the circulation of $\mathbf{F}$ around a given point. Indeed, fix $\mathbf{x}_0 \in \mathbb{R}^2$ and take a ball $B_r(\mathbf{x}_0)$ around $\mathbf{x}_0$. The value of the limit

$$\lim_{r \to 0^+} \frac{1}{\pi r^2} \int_{\partial B_r(\mathbf{x}_0)} (P, Q) \cdot d\mathbf{s}$$

is what we interpret as measuring the circulation of $\mathbf{F}$ around $\mathbf{x}_0$; indeed, the line integral on the right side measures the circulation of $\mathbf{F}$ along a circle of radius r centered at $\mathbf{x}_0$, and when taking the limit in question we shrink this circles down towards $\mathbf{x}_0$ (the πr^2 factor balances out the effect the areas of these circles have on the resulting circulation), so that what we obtained is the circulation occurring "around" $\mathbf{x}_0$ itself.

By Green's Theorem,

$$\int_{\partial B_r(\mathbf{x}_0)} (P, Q) \cdot d\mathbf{s} = \iint_{B_r(\mathbf{x}_0)} (Q_x - P_y)\, dA,$$

so the limit in question becomes

$$\lim_{r \to 0^+} \frac{1}{\operatorname{area}(B_r(\mathbf{x}_0))} \iint_{B_r(\mathbf{x}_0)} (Q_x - P_y)\, dA.$$

We saw a limit of this type back on Problem 4 of Homework 4, where the equality

$$\lim_{r \to 0^+} \frac{1}{\operatorname{Vol}(B_r(\mathbf{x}_0))} \int_{B_r(\mathbf{x}_0)} f(\mathbf{x})\, d\mathbf{x} = f(\mathbf{x}_0)$$

as claimed to be an analog of the Fundamental Theorem of Calculus. In our case, the role of f is played by $Q_x - P_y$, so the conclusion is that

$$\lim_{r\to 0^+} \frac{1}{\text{area}(B_r(\mathbf{x}_0))} \iint_{B_r(\mathbf{x}_0)} (Q_x - P_y)\, dA = (Q_x - P_y)(\mathbf{x}_0).$$

Thus

$$(Q_x - P_y)(\mathbf{x}_0) = \lim_{r\to 0^+} \frac{1}{\pi r^2} \int_{\partial B_r(\mathbf{x}_0)} (P, Q) \cdot d\mathbf{s},$$

which says that $(Q_x - P_y)(\mathbf{x}_0)$ measures the circulation of $\mathbf{F} = (P, Q)$ around $\mathbf{x}_0$ as claimed. We'll see that this same type of argument is what justifies the geometric meaning of curl more generally in three dimensions as a consequence of Stokes' Theorem, and is what justifies the geometric meaning of divergence as a consequence of Gauss's Theorem.

Green's Theorem intuitively. With the interpretation of the quantity $Q_x - P_y$ as the thing which measures circulation around points, we can now go back and say what Green's Theorem says intuitively. The integral

$$\int_{\partial D} (P, Q) \cdot d\mathbf{s}$$

on the left side of Green's Theorem gives the circulation of $\mathbf{F}$ along the boundary of D, while the integral

$$\iint_D (Q_x - P_y)\, dA$$

on the right gives the "sum" of the circulation of (P, Q) among all points of D. Thus, the statement is that: adding up the circulations of a vector field at all points of a region gives the circulation of that field along the boundary. This point of view is made clearer in the "Notes on Green's Theorem" I referred to above, and really gets at the heart of, not only Green's Theorem, but all the BIG THEOREMS we're developing.

Lecture 25: Surface Integrals

Warm-Up. The Warm-Up we looked at was proving the so-called *Divergence Theorem in the Plane* as a consequence of Green's Theorem. This theorem was Problem 3 on the Discussion 6 Problems, so I'll refer you to those solutions for the proof. The resulting equality was

$$\int_{\partial D} (P, Q) \cdot \mathbf{n}\, ds = \iint_D (P_x + Q_y)\, dA.$$

The integral on the left is not a vector line integral, but rather a different type of integral arising when integrating a vector field over a curve. In this case, the dot product $(P, Q) \cdot \mathbf{n}$ measures the extent to which $\mathbf{F} = (P, Q)$ flows *through* ∂D at a given point, as opposed to the expression $(P, Q) \cdot \mathbf{T}$ in the vector line integral which measure the extent to which $\mathbf{F}$ flows *along* ∂D. Thus, the integral on the left in this Divergence Theorem gives a net measure of the flow of $\mathbf{F}$ through ∂D. The equality above can be used to justify the fact that $P_x + Q_y$, the divergence of the 2-dimensional field $\mathbf{F} = (P, Q)$, measures the "inward/outward" flow of $\mathbf{F}$ *through* a point; this just involves the same type of limiting procedure we looked at last time when deriving the geometric interpretation of 2-dimensional curl.

We won't go through this here, since we'll derive the more general interpretation of divergence in three dimensions using a similar limiting procedure and Gauss's Theorem. Indeed, the Divergence

Theorem in the Plane *is* just a 2-dimensional version of Gauss's Theorem, as we'll see. However, even though in two dimensions the Divergence Theorem is just a consequence of Gauss's Theorem, it will not be true in three dimensions that Gauss's Theorem (i.e. the three dimensional Divergence Theorem) is not a consequence of Stokes' Theorem (i.e. the three dimensional analog of Green's Theorem), but is really something new.

Surface integrals. Surface integrals are to surfaces what line integrals are to curves. Suppose S is a smooth C^1 surface and $\mathbf{F}$ a C^1 vector field defined on S. We equip S with an *orientation*, which is a choice of a direction for normal vectors on S. The *vector surface integral* of $\mathbf{F}$ over S is defined to be:

$$\iint_S \mathbf{F} \cdot d\mathbf{S} := \iint_S (\mathbf{F} \cdot \mathbf{n})\, dS.$$

In other words, we integrate over S the function whose value at a point $\mathbf{p} \in S$ is the dot product of $\mathbf{F}(\mathbf{p})$ and the normal vector to S at $\mathbf{p}$. This dot product measures the extent to which $\mathbf{F}$ aligns itself with $\mathbf{n}$ at various points, where a closer alignment indicates a greater "flow" of $\mathbf{F}$ through S at that point. Thus, vector surface integrals measure the net flow (or what is often called *flux*) of $\mathbf{F}$ *through* S in the direction of the given orientation. (Contrast this with vector line integrals, which measure the flow of vector fields *along* a curve.)

Suppose S is given by a parametrization

$$\mathbf{X}(u, v),\ (u, v) \in D,$$

where we assume that the normal vectors given by $\mathbf{X}_u \times \mathbf{X}_v$ agree with the given orientation on S. Then the unit normal $\mathbf{n}$ is given by

$$\mathbf{n} = \frac{\mathbf{X}_u \times \mathbf{X}_v}{\|\mathbf{X}_u \times \mathbf{X}_v\|}.$$

Recalling that the infinitesimal area dS is given by

$$dS = \|\mathbf{X}_u \times \mathbf{X}_v\|\, du\, dv,$$

we get that in terms of our parametrization, the value of the vector surface integral is given by

$$\begin{aligned}\iint_S \mathbf{F} \cdot d\mathbf{S} &= \iint_D \mathbf{F}(\mathbf{X}(u, v)) \cdot \frac{\mathbf{X}_u \times \mathbf{X}_v}{\|\mathbf{X}_u \times \mathbf{X}_v\|}\ \|\mathbf{X}_u \times \mathbf{X}_v\|\ d(u, v) \\ &= \iint_D \mathbf{F}(\mathbf{X}(u, v)) \cdot (\mathbf{X}_u \times \mathbf{X}_v)\, d(u, v).\end{aligned}$$

Thus, to compute vector surface integrals via a parametrization we simply plug in the equations for x, y, z into $\mathbf{F}$, compute the normal vector $\mathbf{X}_u \times \mathbf{X}_v$, check that it gives the correct orientation (if not we multiply through by a negative), take the dot product of $\mathbf{F}$ with this normal vector, and integrate the resulting function over the allowed values of the parameters. Of course, the value obtained is independent of parametrization, which essentially comes from the fact that we used *unit* normal vectors in the definition.

Examples. The examples we did in class and plenty more can be found in my Math 290-3 notes.

Integrating 2-forms. Finally, we recast vector surface integrals in terms of differential forms. Recall that a 2-form on $\mathbb{R}^3$ looks like

$$P\, dy \wedge dz + Q\, dz \wedge dy + R\, dx \wedge dy.$$

We *define* the integral of such a 2-form over a surface S to be the vector surface integral of $\mathbf{F} = (P, Q, R)$ over S:

$$\iint_S P\,dy \wedge dz + Q\,dz \wedge dx + R\,dx \wedge dy := \iint_S (P, Q, R) \cdot d\mathbf{S}.$$

Let us see why this is a good definition. Given a parametrization $\mathbf{X}(u, v),\ (u, v) \in D$ of S, we can rewrite the given 2-form in terms of $du \wedge dv$ instead. (In fancier language, we look at the pullback of this 2-form under the map $\mathbf{X} : D \to \mathbb{R}^3$ defining our parametrization.) With

$$\mathbf{X}(u, v) = (x(u, v), y(u, v), z(u, v)),$$

we have:

$$\begin{aligned}
dy \wedge dz &= d(y(u, v)) \wedge d(z(u, v)) \\
&= \left(\frac{\partial y}{\partial u}\,du + \frac{\partial y}{\partial v}\,dv\right) \wedge \left(\frac{\partial z}{\partial u}\,du + \frac{\partial z}{\partial v}\,dv\right) \\
&= \left(\frac{\partial y}{\partial u}\frac{\partial z}{\partial v} - \frac{\partial y}{\partial v}\frac{\partial z}{\partial u}\right)\,du \wedge dv \\
dz \wedge dx &= d(z(u, v)) \wedge d(x(u, v)) \\
&= \left(\frac{\partial z}{\partial u}\,du + \frac{\partial z}{\partial v}\,dv\right) \wedge \left(\frac{\partial x}{\partial u}\,du + \frac{\partial x}{\partial v}\,dv\right) \\
&= \left(\frac{\partial z}{\partial u}\frac{\partial x}{\partial v} - \frac{\partial z}{\partial v}\frac{\partial x}{\partial u}\right)\,du \wedge dv \\
dx \wedge dy &= d(x(u, v)) \wedge d(y(u, v)) \\
&= \left(\frac{\partial x}{\partial u}\,du + \frac{\partial x}{\partial v}\,dv\right) \wedge \left(\frac{\partial y}{\partial u}\,du + \frac{\partial y}{\partial v}\,dv\right) \\
&= \left(\frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}\right)\,du \wedge dv.
\end{aligned}$$

Notice that the coefficients we get in these three expressions are *precisely* the components of the normal vector $\mathbf{X}_u \times \mathbf{X}_v$:

$$\mathbf{X}_u \times \mathbf{X}_v = \left(\frac{\partial y}{\partial u}\frac{\partial z}{\partial v} - \frac{\partial y}{\partial v}\frac{\partial z}{\partial u}, \frac{\partial z}{\partial u}\frac{\partial x}{\partial v} - \frac{\partial z}{\partial v}\frac{\partial x}{\partial u}, \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}\right).$$

Thus the expression for

$$P\,dy \wedge dz + Q\,dz \wedge dx + R\,dx \wedge dy$$

in terms of $du \wedge dv$ is $du \wedge dv$ with coefficient:

$$P(\mathbf{X}(u, v))\left(\frac{\partial y}{\partial u}\frac{\partial z}{\partial v} - \frac{\partial y}{\partial v}\frac{\partial z}{\partial u}\right) + Q(\mathbf{X}(u, v))\left(\frac{\partial z}{\partial u}\frac{\partial x}{\partial v} - \frac{\partial z}{\partial v}\frac{\partial x}{\partial u}\right) + R(\mathbf{X}(u, v))\left(\frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}\right).$$

This coefficient is exactly the function

$$(P, Q, R) \cdot (\mathbf{X}_u \times \mathbf{X}_v)$$

we integrate in the expression for vector surface integrals in terms of parametric equations. The upshot is that if we attempt to compute

$$\iint_S P\,dy \wedge dz + Q\,dz \wedge dx + R\,dx \wedge dy$$

simply by using equations to rewrite the 2-form in question in terms of $du \wedge dv$ instead, the expression we end up with *is* the expression for the vector surface integral of $\mathbf{F} = (P, Q, R)$ over S. If you think of this 2-form as coming from the "dot product"

$$P\,dy \wedge dz + Q\,dz \wedge dx + R\,dx \wedge dy = (P, Q, R) \cdot (dy \wedge dz, dz \wedge dx, dx \wedge dy),$$

which is not really a thing since it doesn't make sense to consider a vector whose entries are 2-forms (a subtlety we ignore), the idea is that

$$(dy \wedge dz, dz \wedge dx, dx \wedge dy)$$

essentially encodes normal vectors along the surface over which we are integrating.

If you go back to when we first outlined the BIG PICTURE, we described that 2-forms on $\mathbb{R}^3$ can be associated to vector fields via the correspondence

$$P\,dy \wedge dz + Q\,dz \wedge dx + R\,dx \wedge dy \leftrightarrow P\,\mathbf{i} + Q\,\mathbf{j} + R\,\mathbf{k}.$$

The reason why we used this particular correspondence was in order to guarantee eventually that integrating a 2-form over a surface S is the *same* as integrating its corresponding vector field over that surface. The upshot is that writing vector surface integrals in terms of differential forms instead will make clear what is going on in the BIG PICTURE.

Lectures 26 and 27: Stokes' Theorem

What are 2-forms? Before moving on, we finally give an answer as to what a 2-form actually is. Previously we said that 1-forms are mappings which take as input vector fields and output functions. Similarly, a 2-form is simply a mapping which takes as input *two* vector fields and outputs a function. For instance, suppose

$$\mathbf{F} = (P_1, Q_1, R_1) \quad \text{and} \quad \mathbf{G} = (P_2, Q_2, R_2)$$

are two vector fields. The 2-form $dx \wedge dy$ acting on this pair of vector fields is defined to give the following output:

$$dx \wedge dy\,(\mathbf{F}, \mathbf{G}) = \det \begin{pmatrix} dx(\mathbf{F}) & dx(\mathbf{G}) \\ dy(\mathbf{F}) & dy(\mathbf{G}) \end{pmatrix} = \det \begin{pmatrix} P_1 & P_2 \\ Q_1 & Q_2 \end{pmatrix} = P_1Q_2 - P_2Q_1,$$

where the entries in the 2×2 matrix in the second step are the results of applying the 1-forms dx, dy to the vector fields $\mathbf{F}, \mathbf{G}$. In general, for 1 forms α and β, the wedge product $\alpha \wedge \beta$ is defined to be the 2-form whose value at a pair $(\mathbf{F}, \mathbf{G})$ is

$$\alpha \wedge \beta\,(\mathbf{F}, \mathbf{G}) = \det \begin{pmatrix} \alpha(\mathbf{F}) & \alpha(\mathbf{G}) \\ \beta(\mathbf{F}) & \beta(\mathbf{G}) \end{pmatrix} = \alpha(\mathbf{F})\beta(\mathbf{G}) - \alpha(\mathbf{G})\beta(\mathbf{F}).$$

More generally, k-forms are mappings which take k vector fields as inputs and output a function, and higher-dimensional wedge products are defined via higher-dimensional determinants.

If you work it out in all its glory, it turns out that the value of the 2-form

$$\omega = P\,dy \wedge dz + Q\,dz \wedge dx + R\,dx \wedge dy$$

on the pair of tangent vector fields $\mathbf{X}_u, \mathbf{X}_v$ obtained by differentiating the parametric equations of a surface with respect to the two parameters, is

$$\omega(\mathbf{X}_u, \mathbf{X}_v) = P\left(\frac{\partial y}{\partial u}\frac{\partial z}{\partial v} - \frac{\partial y}{\partial v}\frac{\partial z}{\partial u}\right) + Q\left(\frac{\partial z}{\partial u}\frac{\partial x}{\partial v} - \frac{\partial z}{\partial v}\frac{\partial x}{\partial u}\right) + R\left(\frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}\right),$$

which is precisely the integrand obtained in a vector surface integral when computing the dot product of the vector field with the normal vector. The upshot is that the integral of a 2-form over a surface S can be characterized as

$$\int_S \omega := \iint_D \omega(\mathbf{X}_u, \mathbf{X}_v)\, d(u, v)$$

where $\mathbf{X}(u, v)$, $(x, y) \in D$ is a choice of parametrization of S. That is, to integrate a 2-form we *evaluate* that 2-form on the pair of vector fields $\mathbf{X}_u, \mathbf{X}_v$ to obtain a function $\omega(\mathbf{X}_u, \mathbf{X}_v)$, and then we integrate that function over the possible values of the parameters. This is NOT the way in which you probably want to think about such integral computations, but is only meant to give a better perspective on what 2-forms and their integrals actually are.

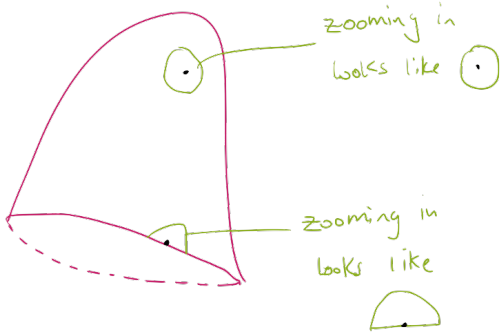
The integrand here can be thought of as the pullback of ω under the map $\mathbf{X} : D \to \mathbb{R}^3$ characterizing the parametrization:

$$\mathbf{X}^*\omega = \omega(\mathbf{X}_u, \mathbf{X}_v)\, du \wedge dv,$$

so the definition of the integral of a 2-form is just a reflection of the change of variables formula when phrased in terms of differential forms and pullbacks. We won't work it out here, but it is possible to show that such integrals are independent of parametrization from this point of view as well, in a manner similar to how we previously argued that line integrals of 1-forms are independent of parametrization.

Examples and what not. The Warm-Ups we did and other examples we looked at (for both days), and more, can be found in my Math 290-3 notes. There you will also see the statement of Stokes' Theorem. The few things we did which are not explicitly spelled out in those notes are included below.

Manifold boundary. My 290-3 notes only give a vague (i.e. visual) description of the type of boundary used in Stokes' Theorem, which we called the *manifold* boundary of a surface. To be clear, the manifold boundary of a compact surface consists of the set of points on that surface near which the surface looks like a half-disk as opposed to a full disk:



In this picture, the manifold boundary is the curve (which looks almost like a circle) at the bottom where the "opening" into the surface is. With this in mind, the description of the induced orientation on the manifold boundary given in my 290-3 notes is perfectly fine.

Stokes' Theorem in terms of differential forms. The vector field version of Stokes' Theorem says

$$\int_{\partial S} \mathbf{F} \cdot d\mathbf{S} = \iint_S \operatorname{curl} \mathbf{F} \cdot d\mathbf{S}.$$

If $\mathbf{F} = (P, Q, R)$, consider the corresponding 1-form:

$$\omega = P\,dx + Q\,dy + R\,dz$$

whose integral over ∂S is the same as the integral on the left side of Stokes' Theorem. A computation we did previously gives:

$$d\omega = (R_y - Q_z)\,dy \wedge dz + (P_z - R_x)\,dz \wedge dx + (Q_x - P_y)\,dx \wedge dy,$$

so by the way in which we defined what it means to integrate a 2-form over a surface, we see that the integral of $d\omega$ is precisely the same as the integral on the right side of Stokes' Theorem. Thus in the language of differential forms, Stokes' Theorem says:

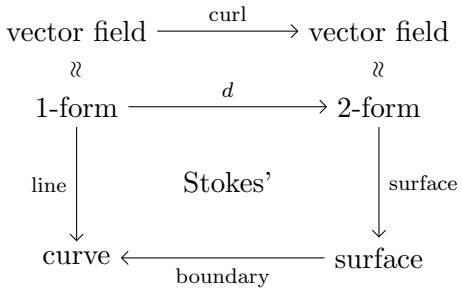
$$\int_{\partial S} P\,dx + Q\,dy + R\,dz = \iint_S (R_y - Q_z)\,dy \wedge dz + (P_z - R_x)\,dz \wedge dx + (Q_x - P_y)\,dx \wedge dy,$$

which can be more compactly written as

$$\int_{\partial S} \omega = \int_S d\omega.$$

The point is that this is the *same* formula we saw when phrasing the Fundamental Theorem of Line Integrals and Green's Theorem in terms of differential forms.

Back to BIG PICTURE. The middle square of the BIG PICTURE was:



Thus, Stokes' Theorem is the thing which tells us (in either vector field or differential form notation) how the different edges of this square relate to one another. That is, starting with a 1-form ω (or vector field $\mathbf{F}$) in the upper left, applying the horizontal arrow to get a 2-form $d\omega$ (or vector field $\operatorname{curl} \mathbf{F}$), and then following the vertical arrow on the right to get an integral over a surface, gives the *same* value as integrating the original ω (or $\mathbf{F}$) over the curve obtained from S by applying the horizontal (boundary) arrow on the bottom. The point again, of course, is notice the similarity between this square and the square characterizing the Fundamental Theorem of Line Integrals and the square characterizing Green's Theorem.

Curl as infinitesimal circulation. With Stokes' Theorem at hand we can now justify the geometric meaning of curl we gave when first introducing curl. ***FINISH***

Lecture 28: Gauss's Theorem