MATH GRE BOOTCAMP: LECTURE NOTES

IAN COLEY

These are lecture notes I wrote for a Math Subject GRE preparation course I ran at UCLA in 2016 and 2018, then at Rutgers in 2021. They have been updated a number of times with feedback from many mathematicians over the years noticing small (or large) typos. As it stands, they are as complete as I would hope. The formatting is not exactly consistent throughout, but in here is definitely everything I want to say.

One final note to all the budding mathematicians reading these notes: please don't hesitate to reach out to me with any questions or comments. I might not reply to every email immediately, but I'm here for you in your GRE prep journey and welcome any suggestions to improve the content.

Last Updated: May 5, 2025.

Contents

1.	Day 1: Single variable calculus	4
1.1.	Basics	4
1.2.	Derivatives	6
1.3.	Applications of Derivatives	10
1.4.	Graphical Analysis	12
1.5.	L'Hôpital's Rule	13
2.	Day 2: Single variable calculus	15
2.1.	Integrals	15
2.2.	Applications of the integral	18
2.3.	Integration techniques	20
2.4.	Sequences and series	24
2.5.	Convergence tests	25
2.6.	Taylor polynomials and series	26
3.	Day 3: Multivariable calculus	29
3.1.	Vectors in $\mathbb{R}^3$	29
3.2.	Planes	30
3.3.	Parametric curves	31
3.4.	Multivariable functions	32
3.5.	Partial derivatives	33
3.6.	Gradient and directional derivatives	35
3.7.	Local extrema	38
3.8.	Optimisation and Lagrange Multipliers	41
4.	Day 4: Multivariable calculus	45
4.1.	Double integrals	45
4.2.	Triple Integrals	46
4.3.	Change of coordinates	48
4.4.	Quadric surfaces	50
4.5.	Vector fields and fancier integration	52
4.6.	Conservative vector fields	54
4.7.	Surface integrals	55
4.8.	Fundamental Theorems of Vector Calculus	58
5.	Day 5: Linear algebra	64
5.1.	Vector spaces	64
5.2.	Bases	66
5.3.	Linear transformations	68
5.4.	Matrices	69
5.5.	Eigenvalues and eigenvectors	72
6.	Day 6: Differential equations and complex analysis	75

6.1. Basic differential equations 75
6.2. Higher order differential equations 77
6.3. Nonhomogeneous differential equations 79
6.4. Complex analysis 80
6.5. Cauchy integral formula 81
7. Day 7: Algebra 83
7.1. Groups 83
7.2. Examples of groups 85
7.3. Abelian groups 86
7.4. Rings 87
7.5. Modular arithmetic 90
7.6. Fields 90
8. Day 8: Analysis and topology 92
8.1. The real numbers 92
8.2. Topological interlude 93
8.3. Continuity revisited 94
8.4. Convergence of functions 97
8.5. Integrals 98
8.6. Topology 99
8.7. Metric spaces 101
9. Day 9: Miscellaneous 103
9.1. Combinatorics 103
9.2. Probability via area 104
9.3. General probability 104
9.4. Statistics 105
9.5. Geometry 106
9.6. Set theory 107
9.7. Graph theory 108
9.8. Algorithms 108

1. Day 1: Single variable calculus

Topics covered: limits, derivatives, implicit differentiation, related rates, the intermediate value theorem, the mean value theorem, optimisation, L'Hôpital's rule, inverse functions and their derivatives, logarithms and exponential functions and their derivatives.

1.1. Basics.

Definition 1.1. Let $f\colon \mathbb{R} \to \mathbb{R}$ be a function. We say that $\lim_{x\to a} f(x) = L$ if for every $\varepsilon > 0$, there exists $\delta > 0$ such that whenever $0 < |x-a| < \delta$, $|f(x) - L| < \varepsilon$.

We will worry about limits of sequences on another day. To set some notation: a *neighbourhood* of $x = a$ will be the set $\{x \in \mathbb{R} : |x-a| < \delta\}$ for some $\delta > 0$; if we specify the radius $\varepsilon > 0$, we will call it an *ε-neighbourhood*. A *punctured neighbourhood* of $x = a$ will be the set $\{x \in \mathbb{R} : 0 < |x-a| < \delta\}$ for some $\delta > 0$. Thus we could rephrase the above definition in words: for every $\varepsilon > 0$, there exists a punctured neighbourhood of $x = a$ that maps entirely to the ε-neighbourhood of L.

Definition 1.2. We say that $f\colon \mathbb{R} \to \mathbb{R}$ is continuous at $x = a$ if $\lim_{x\to a} f(x) = f(a)$.

The condition $|f(a) - f(a)| < \varepsilon$ is automatically satisfied for all $\varepsilon > 0$, so we can extend the demand of a punctured neighbourhood of $x = a$ to a non-punctured one in the case that $L = f(a)$.

Problem 1.3. At what points is the following function continuous?

$$f(x) = \begin{cases} x^2 & x \in \mathbb{Q} \\ x/5 & x \in \mathbb{R} \setminus \mathbb{Q} \end{cases}$$

This sort of problem (coupled with having taken real analysis) should remind you that the heuristic 'continuous means you don't pick up your pen' doesn't work in sophisticated situations.

There is also a notion of right continuous and left continuous, where we use only one-sided limits. This is more easily drawn than defined (and the analogue in multivariable calculus isn't really helpful, so we will omit the full definition). If this were a blackboard, there would be a better example here.

When are functions discontinuous? Jump discontinuities (almost always in piecewise functions), infinite discontinuities, and removable discontinuities (holes). One could also consider a function discontinuous at the points where it ceases to exist, e.g. $f(x) = \sqrt{x}$ is discontinuous at $x \leq 0$, but this isn't really necessary. For examples of continuous functions, think of almost literally any function: polynomials, trigonometric functions, logarithms, exponential functions, etc.

It is likely unimportant for the GRE, but it might be nice to recall the squeeze theorem just in case:

Theorem 1.4 (Squeeze Theorem)**.** Suppose that $f, g, h\colon \mathbb{R} \to \mathbb{R}$ are three functions. If $f(x) \leq g(x) \leq h(x)$ in a punctured neighbourhood of $x = a$, then

$$\lim_{x\to a} f(x) \leq \lim_{x\to a} g(x) \leq \lim_{x\to a} h(x).$$

This is particularly useful when $f(x)$ and $h(x)$ are continuous at $x = a$ (or one is even constant) but $g(x)$ isn't.

Problem 1.5. Compute $\lim\limits_{x\to 0} x \cdot \sin\left(\dfrac{1}{x}\right)$.

The second main definition we need for calculus is that of differentiable functions.

Definition 1.6. We say that $f\colon \mathbb{R} \to \mathbb{R}$ is differentiable at $x = a$ if $\lim\limits_{x\to a} \dfrac{f(x) - f(a)}{x - a}$ exists. Alternatively, we can ask that $\lim\limits_{h\to 0} \dfrac{f(a+h) - f(a)}{h}$ exists. We write $f'(a)$ for this value.

Problem 1.7. Prove that these two definitions agree.

Problem 1.8. Prove that if $f(x)$ is differentiable at $x = a$, it is continuous at $x = a$.

When are functions non-differentiable? By the previous exercise, when they're discontinuous. It's important to remember to always check continuity first, as in the following:

Problem 1.9. Describe the set of solutions $(a, b, c) \in \mathbb{R}^3$ such that the following function is continuous and differentiable (everywhere):

$$f(x) = \begin{cases} ax^2 + bx + c & x \leq 1 \\ x \log x & x > 1 \end{cases}$$

Solution. The pair (a, b) determines the equality of derivatives at $x = 1$ and c determines the continuity. To check continuity, we have that

$$f(1) = a + b + c = 1 \cdot \log(1)$$

so that $a + b + c = 0$. Second, by the product rule we have $(x \log x)' = 1 + \log x$, so for differentiability we need $f'(1) = 2a + b = 1 + \log(1) = 1$, so that $2a + b = 1$. We can solve $b = 1 - 2a$, so $a + b + c = a + (1 - 2a) + c = 0$ hence $-a + c = -1$ so $c = a - 1$ and $b = 1 - 2a$. Thus the set of solutions can be written $(a, 1 - 2a, a - 1)$.

There are two basic theorems about continuous and differentiable functions we will state now, and a third later.

Theorem 1.10 (Intermediate Value Theorem)**.** Let $f\colon [a,b] \to \mathbb{R}$ be a continuous function. Assume that $f(a) < f(b)$. Then for any $y \in [f(a), f(b)]$, there exists $c \in [a,b]$ such that $f(c) = y$.

The IVT is used primarily to prove that a given function has a root in a certain interval, because you know that zero always lies between any positive and negative number.

Problem 1.11. Prove that $f(\theta) = \cos(\theta) + \theta\sin(\theta)$ has a root in the interval $[0, \pi]$.

In the above problem, the particular value of θ is pretty annoying to compute; with the IVT you don't need to do anything fiddly like rewriting $\sin(\theta)$ and $\cos(\theta)$ in terms of $e^{i\theta}$!

Theorem 1.12 (Mean Value Theorem)**.** Let $f\colon [a,b] \to \mathbb{R}$ be a differentiable function (except at the endpoints). Then there exists a point $c \in [a,b]$ such that

$$f'(c) \cdot (b-a) = f(b) - f(a).$$

Problem 1.13. Find a point $x = c$ satisfying the conclusion of the MVT for $g(x) = x^2 + 2x + 2$ on $[0,2]$. Is there more than one?

I would call these theorems 'fundamental' but we all know that's reserved for a bit later.

1.2. **Derivatives.** Of course we all remember the basic rules for doing derivatives, but let's recall them anyway: let $f, g\colon \mathbb{R} \to \mathbb{R}$ be two functions.

- $(f \pm g)'(x) = f'(x) \pm g'(x)$
- $(c \cdot f)'(x) = c \cdot f'(x)$ for all $c \in \mathbb{R}$
- $(f \cdot g)'(x) = f(x) \cdot g'(x) + f'(x) \cdot g(x)$
- $\left(\dfrac{f}{g}\right)'(x) = \dfrac{g(x) \cdot f'(x) - f(x) \cdot g'(x)}{g(x)^2}$ for $g(x) \neq 0$
- $(f \circ g)'(x) = f'(g(x)) \cdot g'(x)$

There are a few other functions we should memorise the derivative of:

- $\dfrac{d}{dx}c = 0$ for $c \in \mathbb{R}$
- $\dfrac{d}{dx}x^n = n \cdot x^{n-1}$
- $\dfrac{d}{dx}\sin(x) = \cos(x)$

- $\frac{d}{dx}\cos(x) = -\sin(x)$
- $\frac{d}{dx}\sinh(x) = \cosh(x)$
- $\frac{d}{dx}\cosh(x) = \sinh(x)$
- $\frac{d}{dx}e^x = e^x$
- $\frac{d}{dx}\log(x) = \frac{1}{x}$

Other trigonometric functions can be computed using the quotient rule.

Perhaps it is also worth remembering the less-used but GRE-noteworthy formula for the second derivative of a function:

Problem 1.14. Prove that, if $f'(x)$ is differentiable, then

$$f''(x) = \lim_{h\to 0} \frac{f(x+h) + f(x-h) - 2f(x)}{h^2}$$

Solution. We can rewrite this expression as follows:

$$\frac{f(x+h) + f(x-h) - 2f(x)}{h^2} = \frac{\frac{f(x+h)-f(x)}{h} - \frac{f(x)-f(x-h)}{h}}{h}$$

If we take $\lim_{h\to 0}$ of the top, we obtain $f'(x) - f'(x-h)$, and we can certainly agree that

$$f''(x) = \lim_{h\to 0} \frac{f'(x) - f'(x-h)}{h}.$$

It is a fun fact that the righthand side (the *symmetric second derivative*) of the above can exist even when $f''(x)$ itself does not exist – the example given on Wikipedia is the sign function

$$\sigma(x) = \begin{cases} -1 & x < 0 \\ 0 & x = 0 \\ 1 & x > 0 \end{cases}$$

What other techniques do we use for computing derivatives? Computing the derivatives of inverse functions can be difficult, specifically when we don't have a closed formula for the inverse. What circumstances are those?

Definition 1.15. Let $f: \mathbb{R} \to \mathbb{R}$ be a function, and suppose that $X \subset \mathbb{R}$ is a set on which f is one-to-one. Then we say that f is invertible on X, and write $f^{-1}(y)$ for the inverse, which is defined by $f^{-1}(y) = x$ if and only if $f(x) = y$.

Problem 1.16. Suppose that $f: \mathbb{R} \to \mathbb{R}$ is a function and that for all $x \in X \subset \mathbb{R}$, $f'(x) > 0$. Then f is invertible on X.

Problem 1.17. Suppose that $f: \mathbb{R} \to \mathbb{R}$ is a differentiable invertible function. Then $f^{-1}: \mathbb{R} \to \mathbb{R}$ is also differentiable, except at those $y = f(x) \in \mathbb{R}$ such that when $f'(x) = 0$.

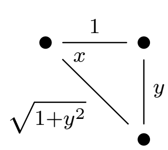
How do we compute the derivative of the inverse? Write $f^{-1} = g$ for simplicity, Then $f(g(y)) = y$, so taking derivatives and using the chain rule we have

$$(f \circ g)'(y) = f'(g(y)) \cdot g'(y) = 1 \implies g'(y) = \frac{1}{f'(g(y))}$$

So as long as we can figure out $g(y)$ and $f'(x)$, we can figure out $g'(y)$.

Problem 1.18. Compute the derivative of $\tan^{-1}(y)$.

Solution. By the formula, we have that the inverse should be the derivative of $\tan(x)$ evaluated on $\tan^{-1}(y)$. The derivative of $\tan(x)$ is $\sec^2(x)$, so we need to figure out $\sec^2(\tan^{-1}(y))$. This is done with a technique I personally call 'draw the triangle'. We know that $\tan^{-1}(y) = x$ for some x, so we need to draw the triangle in which x is one of the angles. We know only that $\tan(x) = y$, so we may draw a right triangle:



Therefore $\sec^2(x)$ we can compute as the hypotenuse squared over the adjacent side squared, that is $\sec^2(\tan^{-1}(y)) = 1 + y^2$. Therefore the derivative of $\tan^{-1}(y)$ is $\dfrac{1}{1+y^2}$.

This method can be used to compute the derivatives of the other inverse trigonometric and hyperbolic trigonometric functions on the fly, so you don't need to necessarily memorise all of them. That said, you should definitely memorise the above example.

Logarithmic differentiation is a useful technique, and it also recalls implicit differentiation.

Definition 1.19. Let $F\colon \mathbb{R}^2 \to \mathbb{R}$ be a function and let $F(x,y) = c$ implicitly define a function of one variable $f\colon (a,b) \to \mathbb{R}$. Then

$$\frac{dy}{dx} = -\frac{F_x}{F_y}$$

where F_x, F_y are the partial derivatives of F.

In practice, life is easier than this. We will illustrate with an example.

Problem 1.20. Find the tangent line to the circle $x^2 + y^2 = 25$ at $(3,4)$.

Solution. In practice, we just take the derivative of everything with respect to x and recall that $y' = dy/dx$. Hence

$$0 = 2x + 2y \cdot y' \implies y' = \frac{-2x}{2y} = -\frac{x}{y}.$$

This is exactly what we get when we use the definition as well. To finish the actual problem, we have

$$\left.\frac{dy}{dx}\right|_{(x,y)=(3,4)} = -\frac{3}{4}$$

so that the tangent line is $y - 4 = -\frac{3}{4}(x-3)$.

Now, what is logarithmic differentiation? Let $y = f(x)$. Using implicit differentiation (or the chain rule, depending on your perspective), we have

$$(\log y)' = \frac{y'}{y} \implies y' = y \cdot (\log y)'.$$

In the case that $\log y$ is easier to differentiate than y, this is a helpful trick. For example, consider $f(x) = x^x$ or

$$f(x) = \sqrt{\frac{(x+11)^2(x-4)}{(x-1)^2(x+4)}}.$$

Problem 1.21. Compute the second derivative of $f(x) = x^{x^x}$.

The last type of basic derivative is for parametric functions. Suppose that we define a graph using two functions $x(t)$ and $y(t)$ rather than $y = f(x)$. It's still straightforward to compute the change in y with respect to the change in x:

$$\frac{dy}{dx} = \frac{dy/dt}{dx/dt}.$$

This process can be iterated to find the second derivative of y with respect to x, but it's a bit more difficult.

Problem 1.22. Using the quotient rule, compute the formula for $\dfrac{d^2y}{dx^2} = \dfrac{d}{dx}\dfrac{dy}{dx}$.

1.3. **Applications of Derivatives.** Related rates problems, which involve parametric functions as discussed above, show up quite often in the single-variable curriculum. Classic examples include balloons filling up or deflating and basins filling up or emptying of water. Ladders sliding down a wall or shadows lengthening are also common.

Problem 1.23. Suppose we have a right conical coffee filter which is 8cm tall with a radius of 4cm. The water drips through at a constant rate of 2 cubic centimetres per second. When there is one eighth of the original water remaining, how fast is the water level dropping?

Solution. We begin by noting that $V(t) = \dfrac{\pi}{3} \cdot h(t) \cdot r(t)^2$. But because our cone is conical, we also know that the height and radius of the cone are in a fixed ratio: $h(t) = 2 \cdot r(t)$. Because we are solving for $h'(t)$, we will substitute in $r(t) = h(t)/2$. Hence

$$V(t) = \frac{\pi}{3} \cdot h(t) \cdot \frac{h(t)^2}{4} = \frac{\pi}{12} \cdot h(t)^3.$$

Taking the time derivative,

$$V'(t) = \frac{\pi}{4} h(t)^2 \cdot h'(t)$$

Letting $t = t_0$ be the time at which we would like to find the change in height, we know that $V'(t_0) = -2$ no matter what. Therefore we need to find $h(t_0)$ to complete the problem. We know that $V(t_0) = V(0)/8$, and that $V(0) = \dfrac{\pi}{3} \cdot 8 \cdot 4^2 = \dfrac{128\pi}{3}$. So as $V(t_0) = \frac{\pi}{12} \cdot h(t_0)^3 = \frac{16\pi}{3}$, it's pretty clear that $h(t_0) = 4$. Therefore

$$-2 = V'(t) = \frac{\pi}{4} \cdot 4^2 \cdot h'(t_0) \implies -\frac{1}{2\pi} = h'(t_0).$$

We can now turn to optimisation. This is certainly a favourite in the undergraduate curriculum and appears sometimes in the GRE. Why does it work?

Theorem 1.24 (Extreme Value Theorem)**.** Let $f\colon [a, b] \to \mathbb{R}$ be a continuous function. Then the set $f([a, b])$ has both a maximum value and a minimum value.

Note that it's necessary that $[a, b]$ be a closed interval. The image of an open interval under a continuous function may have neither a maximum nor a minimum, e.g. $f(x) = 1/x$ and the interval $(0, \infty)$.

We can even say more:

Definition 1.25. Let $f\colon \mathbb{R} \to \mathbb{R}$ be a differentiable function. We say that $a \in \mathbb{R}$ is a critical point if $f'(a) = 0$.

We might also include the case that $f\colon \mathbb{R} \to \mathbb{R}$ is a function which is differentiable except at finitely many points and call those critical points as well. Their utility is the following theorem, credited to Fermat by various sources.

Theorem 1.26 (Fermat's Boring Theorem)**.** Let $f\colon [a,b] \to \mathbb{R}$ be a differentiable function (except at the endpoints). Then the maxima and minima of $f(x)$ occur at critical points and end points.

There are many different kinds of optimisation problems (as there are related rates problems), but all have the same basic process: we are given both a function to optimise and a constraint. The function to optimise is going to be some $F(x, y)$ in two variables, and the constraint is going to be some equation $g(x, y) = c$. After substituting, we'll have a one-variable problem. Sometimes we will have endpoints, and sometimes it will be implicit that as your one variables gets too big or small that there is no extremum to be found. We then find the critical points and exhaustively check which is the largest and which is the smallest.

Problem 1.27. Suppose we are constructing a window comprised of a semicircle sitting atop a rectangle. Given that the perimeter of the window must be 4 meters, what is the maximum area?

Solution. Let us call x the width and y the height of the rectangle. We know that the area is given by the sum $xy + \frac{\pi}{2} \cdot \left(\frac{x}{2}\right)^2$, the first for the rectangle and the second for the semicircle. Our constraint is $2y + x + \frac{\pi x}{2} = 4$, the bottom three sides of the rectangle and the arc of the semicircle. It looks like making x our sole variable will be the best path to success, so we will substitute $y = 2 - \frac{2+\pi}{4} \cdot x$. Our function is thus

$$A(x) = x \cdot \left(2 - \frac{2+\pi}{4} \cdot x\right) + \frac{\pi}{2} \cdot \left(\frac{x}{2}\right)^2 = 2x - \frac{4+\pi}{8} \cdot x^2.$$

Note that this is the equation of a downward-facing parabola, so if x is too big or too small we have $A(x) < 0$. This is obviously a nonsense answer to the question, so what we're looking for is the critical point giving the vertex of the parabola – its maximum.

We now need to solve $A'(x) = 2 - \frac{4+\pi}{4} \cdot x = 0$, so $x = \frac{8}{4+\pi}$. We can now compute the maximum area:

$$A\left(\frac{8}{4+\pi}\right) = \frac{8}{4+\pi} \quad \text{(not a typo).}$$

Problem 1.28. What is the minimum distance between the curve $y = 1/x$ and the origin?

Distance questions seem more difficult, as the optimisation equation we are dealing with is of the form $d(x) = \sqrt{x^2 + f(x)^2}$. But an important observation is that the minimum of $d(x)$ is also achieved by $d(x)^2$, because the distance function is always positive. Moreover, critical points of $d(x)$ and $d(x)^2$ are the same, as

$$\frac{d}{dx}(d(x)^2) = 2 \cdot d(x) \cdot d'(x)$$

and $d(x) > 0$ as long as $x \neq 0$ or $f(x) \neq 0$. The bright side is that $d(x)^2 = x^2 + f(x)^2$ is much easier to differentiate than $d(x)$.

1.4. **Graphical Analysis.** We already know that $f'(x) > 0$ means increasing and $f'(x) < 0$ means decreasing, but now let's recall what the second derivative means. If $f''(x) > 0$, the graph $y = f(x)$ is concave up, so that the graph is open to $+\infty$, and $f''(x) < 0$ is concave down.

Definition 1.29. If $f''(a) = 0$, we say that $x = a$ is a point of inflection.

In the circumstances of optimisation, we can use the second derivative to test whether a critical point is a maximum or minimum.

Theorem 1.30 (Second Derivative Test)**.** Suppose that $x = a$ is a critical point of $f\colon \mathbb{R} \to \mathbb{R}$. Suppose that $f''(x)$ exists in a neighbourhood of a. If $f''(a) < 0$, then $f(a)$ is a local maximum. If $f''(a) > 0$, then $f(a)$ is a local minimum. If $f''(a) = 0$, the test is inconclusive.

There is also the first derivative test: suppose $f'(x)$ exists and is continuous near $x = a$. If $f'(x) < 0$ to the left of a and $f'(x) > 0$ to the right of a, then $f(a)$ is a local minimum. If $f'(x) > 0$ to the left of a and $f'(x) < 0$ to the right of a, then $f(a)$ is a local maximum. This can often be more useful for optimisation problems when the second derivative is too difficult to compute. However, as in the above example of the window (Problem 1.27), we see that $A''(x) = -\dfrac{3\pi}{4}$ constantly, so any extreme values that exist must be maxima.

Problem 1.31. Suppose that $y = f(x)$ is smooth (i.e. has continuous derivatives of all orders) and that $f(1) = 2$ is a local maximum. Order the values $f(1)$, $f'(1)$, $f''(1)$.

Solution. We know that $f(1) = 2$, so that settles that. Because $x = 1$ is a local extremum, we must have $f'(1) = 0$. Finally, because this extreme point is a maximum, we must be concave down so $f''(1) < 0$. Hence $f''(1) < f'(1) < f(1)$.

1.5. **L'Hôpital's Rule.** The last topic worth remembering L'Hôpital's rule, which comes surprisingly in the second quarter of calculus at UCLA but we can recall now.

Theorem 1.32. Let $f, g\colon \mathbb{R} \to \mathbb{R}$ be two functions and $a \in \mathbb{R} \cup \{\pm\infty\}$. If $\lim_{x\to a} f(x) = \lim_{x\to a} g(x) = c$ where $c \in \{0, \pm\infty\}$, then

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)}$$

Expressions of the form $\dfrac{0}{0}$ and $\pm\dfrac{\infty}{\infty}$ are called indeterminate forms. L'Hôpital's rule can also be used to solve problems which are not immediately in an indeterminate form. The easiest case is expressions of the form $f(x) \cdot g(x)$ which give rise to the indeterminate form $0 \cdot \infty$. By rearranging to

$$\frac{f(x)}{1/g(x)} \text{ or } \frac{g(x)}{1/f(x)}$$

we will obtain an actual indeterminate form. Which one to choose depends on whether $1/f(x)$ or $1/g(x)$ is easier to differentiate.

Problem 1.33. Compute the following limit:

$$\lim_{x\to 2} \frac{1}{x-2} - \frac{2x}{x^2-4}$$

Solution. If you plug in $x = 2$, we don't obtain an indeterminate form, but we obtain something that looks like $\infty - \infty$. This is our clue to combine the fractions into an indeterminate form. With a common denominator,

$$\frac{x+2}{x^2-4} - \frac{2x}{x^2-4} = \frac{-x+2}{x^2-4} = \frac{-(x-2)}{(x-2)(x+2)} = \frac{-1}{x+2}.$$

In this case we don't even need to use L'Hôpital's rule to finish the problem since we had some nice cancellation. In other cases we might not be so lucky.

Problem 1.34. Compute the following limit:

$$\lim_{x\to\infty} x^{1/x}$$

Solution. These problems are also related to L'Hôpital's rule as well. Plugging in, we obtain ∞^0. We notice that if we took the log of this expression, $\log\left(x^{1/x}\right) = \frac{1}{x} \cdot \log x$ yields the form $0 \cdot \infty$. We can now rearrange it to $\dfrac{\log x}{x}$ and finish the problem:

$$\lim_{x\to\infty} \log\left(x^{1/x}\right) = \lim_{x\to\infty} \frac{\log x}{x} \overset{\text{L'H}}{=} \lim_{x\to\infty} \frac{1}{x} = 0.$$

But of course this solves the wrong question. If we say $L = \lim_{x\to\infty} x^{1/x}$, then $\log L = 0$ (as log is a continuous function so commutes with limits). Thus $L = 1$.

This same type of solution works if we have the form 1^∞, as $\log(1^\infty) = \infty \cdot \log 1$ yields $\infty \cdot 0$. This concludes the differential side of single-variable calculus.

2. Day 2: Single variable calculus

Topics covered: the integral, area between curves, volumes of revolution, the fundamental theorem of calculus, u-substitution, integration by parts, trigonometric integration, partial fractions, arc length and surface area, sequences and series, convergence tests, Taylor polynomials and power series, root and ratio tests.

2.1. **Integrals.** Let us start with the definition of the Riemann integral. To do so, we need to recall the limit of a sequence.

Definition 2.1. Let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence. We say that $\lim\limits_{n\to\infty} x_n = L$ if for all $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that whenever $n > N$, $|x_n - L| < \varepsilon$.

The difference here is that there is no function floating around. We will revisit the intricacies of sequences and series later in this lecture.

Definition 2.2. Let $f\colon [a,b] \to \mathbb{R}$ be a function. We define the left Riemann integral of f on the interval $[a,b]$ to be the limit

$$\lim_{n\to\infty} \frac{b-a}{n} \cdot \sum_{i=0}^{n-1} f\left(a + \frac{b-a}{n} \cdot i\right)$$

We define the right Riemann integral of f on the interval $[a,b]$ to be the limit

$$\lim_{n\to\infty} \frac{b-a}{n} \cdot \sum_{i=1}^{n} f\left(a + \frac{b-a}{n} \cdot i\right)$$

We say that the Riemann integral of f on the interval $[a,b]$ if the above exist and agree. In that case, we write

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

(Well, as we know from real analysis, that technically isn't true. We should check that every possible Riemann sum (for any partitions with width shrinking to zero) actually gives us the same value. But we're never going to do that – we're pretty much going to be integrating continuous functions, whose integrability we never need question.)

We also call the individual terms of these limits the left and right Riemann sums, denoted $L_n f$ or $R_n f$. There are a few other approximations for integrals, including the midpoint and trapezoid approximations. The trapezoid approximation is the average of the left and right, and the midpoint rule uses $\left(a + \frac{b-a}{2n} \cdot (2i+1)\right)$ in its argument.

We will almost never use the right and left Riemann sums, as these limits are not calculable in practice, but it's important to know a few things. If a function is increasing, then the left Riemann sum will always underestimate the actual value of the integral, and the right Riemann sum will always overestimate it. This definitely comes up on the GRE.

Problem 2.3. Let $f(x) = 1/x^2$. For the interval $[1, 2]$, order the following terms: $L_5 f$, $\int_1^2 f(x)\,dx$, $R_5 f$.

Solution. We know that $1/x^2$ is *decreasing* on the interval $[1, 2]$, so the left Riemann sum will always overestimate it and the right Riemann sum will always underestimate it. The actual integral will land in the middle. So while the hasty student is still putting numbers into their calculator (which isn't allowed on the test anyway) to try to crunch these Riemann sums, we can finish the problem: $L_5 f > \int_1^2 f(x)\,dx > R_5 f$.

The Riemann integral is not guaranteed to exist for an arbitrary function, but it must exist for our favourite functions.

Proposition 2.4. If $f\colon [a, b] \to \mathbb{R}$ is bounded and continuous at all but finitely many points, then $\int_a^b f(x)\,dx$ exists.

We could assign this as an exercise, but it's a bit difficult and not necessary at all for the GRE. This isn't a necessary and sufficient condition, but it is certainly good enough for almost all purposes.

The integral is linear, just as the derivative was. In particular, this means that

$$\int_a^b c \cdot f(x)\,dx = c \cdot \int_a^b f(x)\,dx$$

and

$$\int_a^b f(x) \pm g(x)\,dx = \int_a^b f(x)\,dx \pm \int_a^b g(x)\,dx$$

Moreover, recalling the definition via Riemann sums, we can always split an integral into intermediate chunks. For any $c \in [a, b]$, we have

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx$$

By definition we will say $\int_a^a f(x)\,dx = 0$, so using the additivity above

$$\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx$$

This way we can make sense of integrals where our interval $[a, b]$ happens to be oriented the wrong way (i.e. $a > b$).

There's a useful computational trick that is best explained using Riemann sums, so we include it here.

Definition 2.5. A function $f : \mathbb{R} \to \mathbb{R}$ is called *odd* if $f(x) = -f(-x)$ for all $x \in \mathbb{R}$. A function $g : \mathbb{R} \to \mathbb{R}$ is called *even* if $g(x) = g(-x)$ for all $x \in \mathbb{R}$.

Odd functions include polynomials with only odd-degree terms, $\sin(x)$, and $\tan(x)$. Even functions include polynomials with only even-degree terms (including the constant term) and $\cos(x)$. Compositions of even and odd functions work like multiplication – odd composed with odd is odd, even composed with odd is even, etc. Products of even and odd functions work like addition: odd times odd is even, even times odd is odd, etc.

We mention all this because of the following trick:

Proposition 2.6. Let $f : \mathbb{R} \to \mathbb{R}$ be an odd, integrable function. Then

$$\int_{-R}^{R} f(x)\, dx = 0$$

for any $R > 0$.

Proof. Using the midpoint Riemann sum, we can give a computation:

$$\int_{-R}^{R} f(x)\, dx = \lim_{n\to\infty} \frac{R-(-R)}{n} \sum_{i=0}^{n-1} f\left(-R + \frac{R-(-R)}{2n} \cdot (2i+1)\right)$$

The list of values $-R + \frac{R(2i+1)}{n}$ is completely symmetric about the y-axis and if n is odd we get 0 in the middle. Since $f(x)$ is odd, we have $f(x_i) = -f(-x_i)$, so these terms of the Riemann sum will cancel each other out. Thus

$$\frac{R-(-R)}{n} \sum_{i=0}^{n-1} f\left(-R + \frac{R-(-R)}{2n} \cdot (2i+1)\right) = 0$$

so the integral will be zero as well. $\square$

Problem 2.7. Compute the integral

$$\int_{-1}^{1} \sin(x^3) + \sin^3(x)\, dx$$

Solution. There are both odd functions (since they are compositions of the odd functions $\sin(x)$ and x^3), so they both integrate to zero.

Besides the above tricks, we will always compute the integral using the Fundamental Theorem of Calculus.

Theorem 2.8 (FTC I)**.** Assume that $f\colon [a,b] \to \mathbb{R}$ is a continuous function. If $F\colon [a,b] \to \mathbb{R}$ is a function such that $F'(x) = f(x)$, then

$$\int_a^b f(x)\,dx = F(b) - F(a).$$

Hence the calculation of integrals will amount to the calculation of antiderivatives (i.e. the function $F(x)$ above).

There's a companion theorem that we state now.

Theorem 2.9 (FTC II)**.** Let $f\colon [a,b] \to \mathbb{R}$ be a continuous function, and define $F(x) = \displaystyle\int_a^x f(t)\,dt$. Then $F'(x) = f(x)$.

We can extend this theorem using the chain rule:

Problem 2.10. Prove that

$$\frac{d}{dx}\int_{a(x)}^{b(x)} f(t)\,dt = f(b(x)) \cdot b'(x) - f(a(x)) \cdot a'(x).$$

One last thing to say at this point is the definition of an improper integral. Suppose that we are trying to integrate $f(x)$ on an unbounded region, say $[0,\infty)$, or over a region $[a,b]$ on which $g(x)$ has an infinite discontinuity (say at $x = a$). Then we can define the integral (should it exist) as a limit of integrals as defined above, e.g.

$$\int_0^\infty f(x)\,dx := \lim_{R\to\infty}\int_0^R f(x)\,dx \qquad \int_a^b g(x)\,dx := \lim_{h\to 0^+}\int_{a+h}^b g(x)\,dx$$

These limits are not guaranteed to exist. The most common types of integrals we consider in this situation are functions $f(x)$ such that $\displaystyle\lim_{x\to\infty} f(x) = 0$ so that the integral at least has a chance of converging. We will readdress this issue in the section on series.

2.2. **Applications of the integral.** But before we get into techniques, why bother computing integrals at all? For one, the integral of $f(x)$ on $[a,b]$ gives the signed area of the region under the graph of $f(x)$. This can even be used in reverse: integrals can be calculated using geometry.

Problem 2.11. Compute the integral

$$\int_{-2}^2 \sqrt{4 - x^2}\,dx$$

Solution. The equation $y = \sqrt{4-x^2}$ corresponds to $x^2 + y^2 = 4$, i.e. a circle of radius 2. Our graph is the top half of this circle. Therefore the area under the curve is half the area of the circle, 2π. This problem can be solved using trigonometric substitution as well, but it's much more annoying (and the GRE is all about saving time where you can).

In the same way, we can compute the area between curves with integrals. We need only to compute the area under the upper curve and subtract the area under the lower curve. The main question in situations like this is which curve is on top and which is on bottom.

Problem 2.12. Compute the area between the curves $y = \sqrt{x}$ and $y = x^2$ in the first quadrant.

Solution. Thinking on the graphs, we can recall that $\sqrt{x}$ is above x^2 in the region $[0, 1]$ where these curves intersect. Therefore the integral we need to compute is

$$\int_0^1 \sqrt{x} - x^2\, dx.$$

Luckily, we can get antiderivatives of these functions very easily:

$$\int_0^1 \sqrt{x} - x^2\, dx = \frac{2x^{3/2}}{3} - \frac{x^3}{3}\Big|_0^1 = \frac{2}{3} - \frac{1}{3} = \frac{1}{3}.$$

The other application is to surfaces and volumes of revolution. Supposing that we rotate a curve $y = f(x)$ around the x-axis, the height of the curve becomes the radius of a disc, which we then need to integrate along the interval $[a, b]$ in question, whatever it is:

$$V = \pi \int_a^b f(x)^2\, dx$$

Problem 2.13. What is the volume of the region created by rotating $y = \log x$ around the x-axis between $x = 1$ and $x = e^2$?

Note that you'll need to use integration by parts for this one (see below).

It's harder to compute the volume when we rotate $y = f(x)$ around the y-axis. Rather than use the method of discs, we use the method of cylindrical shells. In this case, we are computing the area of a cylinder of radius r and height h, which is given by $2\pi \cdot r \cdot h$. In our case, the radius is x and the height is $f(x)$, hence

$$V = 2\pi \int_a^b x \cdot f(x)\, dx$$

Problem 2.14. Compute the volume of the region created by rotating $y = 1 - 2x + 3x^2 - 2x^3$ from $[0, 1]$ around the y-axis.

We can also repeat the above problems when asked to revolve the area between two curves around either the x- or the y-axis. These follow the general formulas, for

$f(x) \geq g(x)$:

$$V = \pi \int_a^b f(x)^2 - g(x)^2 \, dx$$

and for the method of cylindrical shells:

$$V = 2\pi \int_a^b x(f(x) - g(x)) \, dx$$

Arc length of a curve is another application. If we are looking at the infinitesimal change in the length of a curve, it travels dx in the x direction and dy in the y direction. Therefore its total length is $ds = \sqrt{(dx)^2 + (dy)^2}$. In the case that $y = f(x)$, we have

$$s = \int ds = \int \sqrt{(dx)^2 + \left(\frac{dy}{dx}\right)^2} = \int_a^b \sqrt{1 + f'(x)^2} \, dx$$

Problem 2.15. Compute the arc length of $y = \cosh x$ over the interval $[0, 2]$.

Sometimes we also want to calculate the surface areas of regions of revolution, not just the volumes. In this case, we again have a formula: the infinitesimal amount of surface area is given by the arc length along the surface times the circumference of the disc, which in the case of rotation around the x-axis is $2\pi f(x)$. Thus

$$S = 2\pi \int f(x)\sqrt{1 + f'(x)^2} \, dx$$

Problem 2.16. Compute the surface area of a sphere of radius R, using the curve $y = \sqrt{R^2 - x^2}$.

2.3. **Integration techniques.** Now besides guessing at antiderivatives, what are our other integration techniques? The first is u-substitution, which is is our answer to the chain rule.

Theorem 2.17 (u-substitution)**.** Suppose that $h(x)$ is a continuous function and we can write $h(x) = f(g(x))g'(x)$. Then

$$\int_a^b h(x) \, dx = \int_{g(a)}^{g(b)} f(u) \, du.$$

Problem 2.18. Evaluate the integral of $f(x) = 2x \cdot \sin(x^2)$ on $[0, \sqrt{\pi/2}]$.

Solution. We see that $g(x) = x^2$ and $f(u) = \sin u$ is a good choice. Hence

$$\int_0^{\sqrt{\pi/2}} 2x \cdot \sin(x^2) \, dx = \int_0^{\pi/2} \sin u \, du = -\cos u \Big|_0^{\pi/2} = 1.$$

The second is integration by parts, which is our answer to the product rule. It steps from the following observation: if we consider the product $(f \cdot g)' = f' \cdot g + f \cdot g'$ and integrate both sides, we obtain $f \cdot g = \displaystyle\int g\, df + \int f\, dg$.

Theorem 2.19 (Integration by Parts)**.** Suppose that $f(x)$ has the form $u(x) \cdot v'(x)$. Then

$$\int f(x)\, dx = \int u(x) \cdot v'(x)\, dx = u(x) \cdot v(x) - \int v(x) \cdot u'(x)\, dx.$$

This is useful when your function is a product of a part which is easy to integrate and a part which is difficult to integrate. Another situation is when part of your function will differentiate to zero and the other part will not differentiate, as in $x \cdot e^x$. A general mnemonic for what functions should be chosen for $u(x)$ is LIATE: logarithms, inverse trigonometric functions, algebraic (e.g. polynomials), trigonometric functions, and finally exponential functions. Note that these latter two types in particular are almost never ideal.

Problem 2.20. Compute the integral of $f(x) = \log x$.

Solution. If we follow our mnemonic, we will choose $dv = \log x\, dx$. This means that $u = 1$. Therefore $v = \dfrac{1}{x}$ and $du = dx$. Hence:

$$\int \log x\, dx = x \log x - \int x \cdot \frac{1}{x}\, dx = x \log x - \int 1\, dx = x \log x - x.$$

It doesn't seem like it would work, and then it does.

Problem 2.21. Compute the integral of $f(x) = x^2 e^x$.

The next method is the method of trigonometric substitution. There are some integrals that do not lend themselves to either of the above methods, but require a special kind of u-substitution. We recognise this situation in the case that one of the Pythagorean identities holds, usually the following:

Problem 2.22. Compute the integral of $f(x) = \sqrt{1 + x^2}$.

Solution. The trigonometric identity we are looking for in this situation is that $1 + \tan^2\theta = \sec^2\theta$. Therefore if we substitute $x = \tan\theta$, we will only need to integrate $\sec\theta$, which is much more tractable than the current issue.

However, we have to consider the integration term. If $x = \tan\theta$, then $dx = \sec^2\theta\, d\theta$. Thus

$$\int \sqrt{1 + x^2}\, dx = \int \sec\theta \cdot \sec^2\theta\, d\theta.$$

We leave it in this form because we can now use integration by parts to solve this problem: let $dv = \sec^2\theta\, d\theta$ and $u = \sec\theta$. Then $v = \tan\theta$ and $du = \sec\theta\tan\theta$, so

$$\int \sec^3\theta\, d\theta = \sec\theta\cdot\tan\theta - \int \sec\theta\tan^2\theta\, d\theta.$$

We now use that $\tan^2\theta = \sec^2\theta - 1$ as we did above so that

$$\int \sec^3\theta\, d\theta = \sec\theta\tan\theta - \int \sec\theta(\sec^2\theta - 1)\, d\theta$$

which rearranged gives us

$$2\int \sec^3\theta\, d\theta = \sec\theta\tan\theta - \int \sec\theta\, d\theta$$

Therefore the last thing to compute is $\int \sec\theta\, d\theta$. This is a hard calculation that one must memorise. The key is that we can apply u-substitution if we decide (as if by magic) to multiply by $\dfrac{\sec\theta + \tan\theta}{\sec\theta + \tan\theta}$. I will leave it to the reader to verify

$$\int \sec\theta\, d\theta = \log(\tan\theta + \sec\theta).$$

Putting this all together completes the problem, sort of. Though we know that

$$\int \sec^3\theta\, d\theta = \frac{\sec\theta\tan\theta}{2} - \frac{1}{2}\log(\tan\theta + \sec\theta)$$

we were asked a question about $f(x)$, not $f(\theta)$. We know that $x = \tan\theta$, so we can make that substitution. In order to determine what $\sec\theta$ is in terms of x, we draw the triangle as in yesterday's lecture. The Pythagorean theorem will tell us that $\sec\theta = \sqrt{1+x^2}$. If we then put it all together,

$$\int \sqrt{1+x^2}\, dx = \frac{x\sqrt{1+x^2}}{2} - \frac{1}{2}\log(x + \sqrt{1+x^2}).$$

There is a little bit more to say about trigonometric techniques. Recall the double angle formulas for sine and cosine:

$$\sin(2x) = 2\sin(x)\cos(x), \qquad \cos(2x) = 2\cos^2(x) - 1 = 1 - 2\sin^2(x) = \cos^2(x) - \sin^2(x)$$

These become relevant when we are faced with a problem like

$$\int_0^{2\pi} \cos^2(x)\, dx$$

which it seems like we have to attack with integration by parts. But by rearranging one of the above formulas, we have

$$\cos(2x) = 2\cos^2 x - 1 \implies \cos^2 x = \frac{\cos(2x) + 1}{2}$$

and the righthand side is easy to integrate with a u-substitution. There is an additional trick we can use: because sine and cosine are periodic we know that

$$\int_0^{2\pi} \sin(x)\,dx = \int_t^{t+2\pi} \sin(x)\,dx$$

for any $t \in \mathbb{R}$. In addition, when integrating sine or cosine for a whole period the value of the integral is zero; we can make this argument using the area under the curve interpretation of the integral, as the positive part of each period cancels out the negative part exactly. Another argument is the following: $\sin(x)$ is an odd function, so we can make the choice $t = -\pi$ above to show that $\displaystyle\int_{-\pi}^{\pi} \sin(x)\,dx = 0$. Since $\cos(x) = \sin(\pi/2 - x)$, we can manipulate the integral $\displaystyle\int_0^{2\pi} \cos(x)\,dx$ to make the same conclusion.

Since the period of $\cos(2x)$ is π, we can also conclude that

$$\int_0^{2\pi} \cos(2x)\,dx = 0$$

since we are integrating over two complete periods. This lets us compute the integral:

$$\int_0^{2\pi} \cos^2(x)\,dx = \int_0^{2\pi} \frac{\cos(2x)+1}{2}\,dx = \int_0^{2\pi} \frac{1}{2}\,dx = \pi.$$

Note that the periodicity argument work for $\cos^2 x$ because there is no negative part of this function anymore.

The final integration technique is the method of partial fractions. We are only capable with our techniques to integrate fairly simple rational functions, so the method of partial fractions allows us to break up complicated expressions into integrable ones. The idea is that any quotient $\dfrac{p(x)}{q(x)}$ of polynomials comes from a sum of polynomials whose denominators are the irreducible factors of $q(x)$. We can integrate expressions of the form $\dfrac{a}{x-r}$ or $\dfrac{ax+b}{x^2+r}$.

Problem 2.23. Compute the integral of $\dfrac{4}{x^4-1}$.

Solution. We factor the denominator as $(x+1)(x-1)(x^2+1)$, and so set up

$$\frac{4}{x^4-1} = \frac{A}{x+1} + \frac{B}{x-1} + \frac{Cx+D}{x^2+1}$$

Multiplying through by the denominator,

$$4 = A(x-1)(x^2+1) + B(x+1)(x^2+1) + (Cx+D)(x+1)(x-1)$$

We can plug in particular values for x to compute A and B. If $x = -1$,

$$4 = A(-2)(2) \implies A = -1$$

If $x = 1$,

$$4 = B(2)(2) \implies B = 1$$

For the last two variables, we need to do some brute multiplication. Omitting the details,

$$4 = Cx^3 + (2 + D)x^2 - Cx + (2 - D)$$

which implies that $C = 0$ and $D = -2$. Hence

$$\frac{4}{x^4 - 1} = \frac{-1}{x + 1} + \frac{1}{x - 1} + \frac{-2}{x^2 + 1}$$

which we can now integrate:

$$\begin{aligned} \int \frac{4}{x^4 - 1}\,dx &= \int \frac{-1}{x + 1}\,dx + \int \frac{1}{x - 1}\,dx + \int \frac{-2}{x^2 + 1}\,dx \\ &= -\log|x + 1| + \log|x - 1| - 2\tan^{-1}(x) \end{aligned}$$

There's a bit of a complication if the irreducible factor of $q(x)$ is not exactly of the form $x^2 + r$, but completing the square and some further manipulation can always get us to that form.

Problem 2.24. Prove that

$$\int \frac{dx}{x^2 + r} = \frac{1}{\sqrt{r}} \tan^{-1}\left(\frac{x}{\sqrt{r}}\right)$$

2.4. **Sequences and series.** We gave above the definition of a sequence and the definition of a convergent sequence. We can now recall what a series is.

Definition 2.25. A series is a sequence $\{S_n\}_{n\in\mathbb{N}}$ defined by a sequence $\{a_i\}_{i\in\mathbb{N}}$ such that $S_n = \displaystyle\sum_{i=0}^{n} a_i$. A series converges if the sequence $\{S_n\}$ converges as above. We write $\displaystyle\sum_{i=0}^{\infty} a_i$ for $\displaystyle\lim_{n\to\infty} S_n$.

Sometimes we will call $\{a_i\}$ the series and leave the fact that we are taking sums implicit.

The simplest type of sequence/series that we encounter is the geometric series, which is defined by $a_i = a_0 \cdot r^i$ for some $r \in \mathbb{R}$. In this circumstance, there is an easy criterion for when the series $\{a_i\}$ converges.

Problem 2.26. Prove that the series $\{a_i = a_0 \cdot r^i\}$ converges if and only if $|r| < 1$.

In this situation, the infinite sum has the formula $\dfrac{a_0}{1 - r}$.

2.5. **Convergence tests.** Unfortunately, most series aren't geometric, so we have convergence tests to decide whether they converge. If a series does not converge, we say it diverges. The first test is the following easy check:

Theorem 2.27 (Divergence Test)**.** The series $\{a_n\}$ diverges if $\lim_{n\to\infty} a_n \neq 0$.

Theorem 2.28 (p-test)**.** Suppose we have the series $a_n = \dfrac{1}{n^p}$ for some $p \in \mathbb{R}$. The series $\{a_n\}$ converges if and only if $p > 1$.

This situation is not so common, but is a useful tool (as we will soon see).

Theorem 2.29 (Limit Comparison Test)**.** Suppose that $\{a_n\}, \{b_n\}$ are nonnegative series. Consider the quantity

$$\lim_{n\to\infty} \frac{a_n}{b_n} = L$$

where $L \in [0, \infty]$. Then we have the following three conclusions:

- If $L = 0$, the series $\{a_n\}$ converges if $\{b_n\}$ converges
- If $L = \infty$, the series $\{b_n\}$ converges if $\{a_n\}$ converges
- If $L \in (0, \infty)$, the series $\{a_n\}$ converges if and only if $\{b_n\}$ converges

It is sometimes useful to bear in mind the contrapositives of these statements. Therefore while many series do not take the form of the p-test, they can be limit-compared to a 'p-series' and thus we can determine their convergence. These two also combine to answer the standing question about improper integrals:

Theorem 2.30 (Integral Test)**.** Suppose that $\{a_n\}$ is a positive series such that there exists a continuous, decreasing function $f(x)$ satisfying $f(n) = a_n$. Then the series $\{a_n\}$ converges if and only if the improper integral $\displaystyle\int_0^\infty f(x)\,dx$ exists (converges).

Therefore the same types of tests that check for the convergence of infinite series can be used to check the convergence of improper integrals when the limits in question are too difficult to compute. Here, however, is one case that can be computed directly.

Problem 2.31. Show that $\displaystyle\int_0^1 \frac{1}{x^p}\,dx$ converges if and only if $p < 1$.

Problem 2.32. When does the integral $\displaystyle\int_0^\infty \frac{1}{x^p}$ converge?

Problem 2.33. Does $\displaystyle\sum_{n=3}^\infty \frac{1}{n\log(n)}$ converge or diverge?

Thus far we have spoken of positive series, but there is specific test for alternating series that is occasionally useful:

Theorem 2.34 (Alternating Series Test)**.** Suppose that $\{a_n\}$ is a positive, decreasing series. Then as long as $\lim_{n\to\infty} a_n = 0$, the series $\sum_{n=0}^{\infty}(-1)^n a_n$ converges.

For example, we saw above that the series $\left\{\dfrac{1}{n}\right\}$ does not converge (the harmonic series), but the alternating series $\left\{\dfrac{(-1)^n}{n}\right\}$ does. As a remark, the factor $(-1)^n$ is the easiest way to see that a series is alternating, but $\cos(\pi \cdot n)$ does the job as well.

2.6. **Taylor polynomials and series.** Before getting into the last two tests, we need to recall the definition of a Taylor polynomial and a Taylor series. To begin, we can define a Taylor polynomial and then explain its utility.

Definition 2.35. Let $f\colon \mathbb{R} \to \mathbb{R}$ be a function of class C^k, i.e. , it has continuous derivatives up to degree k. The kth Taylor polynomial of f centred at $x = a$ is defined by

$$T_k f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)(x-a)^2}{2} + \cdots + \frac{f^{(k)}(a)(x-a)^k}{k!}$$

This is the best polynomial approximation to $f(x)$ with the property that the first k derivatives at $x = a$ agree. For example, the first Taylor polynomial is just the tangent line to $x = a$, i.e. the linear approximation. If $a = 0$, then usually we use the name Maclaurin instead of Taylor.

The error of a Taylor polynomial can be approximated using what would be the next term. In particular,

$$|T_k f(b) - f(b)| \leq \frac{M \cdot |b-a|^{k+1}}{(k+1)!}$$

where M is the maximum of $|f^{(k+1)}(x)|$ on the interval $[a, b]$ or $[b, a]$ (whichever makes sense). In nice circumstances, M takes its maximum at a or b and so a more complicated calculation is not necessary.

Problem 2.36. Compute $\sqrt{10}$ using the third Taylor polynomial of an appropriate function centred at an appropriate value. What is the maximum error?

Supposing that f is smooth, we can take this definition to infinity.

Definition 2.37. The Taylor series of $f(x)$ centred at $x = a$ is the infinite sum

$$T(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)(x-a)^k}{k!}$$

In ideal circumstances, we have $T(x) = f(x)$ for all $x \in \mathbb{R}$, but this is not guaranteed. This infinite sum may not even converge for some values of x, or for most values of x for that matter. This is where our last two tests come in handy.

Theorem 2.38 (Ratio Test)**.** Consider a series $\{a_n\}$ which may or may not be positive, and consider the limit

$$\rho = \lim_{n\to\infty} \left| \frac{a_{n+1}}{a_n} \right|$$

Then $\{a_n\}$ converges absolutely if $\rho < 1$, diverges if $\rho > 1$, and is inconclusive if $\rho = 1$.

Recall that a series is said to converge absolutely if $\{|a_n|\}$ converges. A series converges conditionally if $\{a_n\}$ converges but $\{|a_n|\}$ does not. It is an easy exercise in the triangle inequality that if $\sum_{n=0}^{\infty} |a_n|$ converges, so does $\sum_{n=0}^{\infty} a_n$.

In the case that the series depends on a parameter x, this gives us a function $\rho(x)$ that we demand be ≤ 1 to have a chance at convergence. Those x for which $\rho(x) < 1$ is called the radius of convergence. If we also determine whether the cases $\rho(x) = 1$ converge (using other techniques), we obtain the interval of convergence.

Problem 2.39. Compute the interval of convergence of

$$f(x) = \sum_{n=1}^{\infty} \frac{2x^n}{n^2}$$

There is another test that is sometimes more helpful (though rarely).

Theorem 2.40 (Root Test)**.** Consider a series $\{a_n\}$ which may or may not be positive, and consider the limit

$$\rho = \lim_{n\to\infty} |a_n|^{1/n}$$

Then $\{a_n\}$ converges absolutely if $\rho < 1$, diverges if $\rho > 1$, and is inconclusive if $\rho = 1$.

Both these tests are checking to what extent the series we are considering is geometric with common ratio ρ. Hence series that are mostly polynomial (and not geometric) will yield an inconclusive root/ratio test. One should attempt to apply a p-test to these (perhaps via a limit comparison). Series that include factorial or exponential terms are the target for the root and ratio tests.

We end by recalling some useful Taylor expansions and their radius of convergence.

- $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}, \quad x \in \mathbb{R}$

- $\dfrac{1}{1-x} = \displaystyle\sum_{n=0}^{\infty} x^n, \quad |x| < 1$
- $\sin(x) = \displaystyle\sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!}, \quad x \in \mathbb{R}$
- $\cos(x) = \displaystyle\sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!}, \quad x \in \mathbb{R}$

We can use these building blocks to construct other Taylor series using substitution, differentiation, and integration.

Problem 2.41. Compute the Taylor series for $\tan^{-1}(x)$ centred at $a = 0$. What is its radius of convergence?

Solution. This looks like a dubious prospect, but we recall that

$$\frac{d}{dx}\tan^{-1}(x) = \frac{1}{1+x^2}$$

and the righthand side looks a lot like a Taylor series we already know. In particular,

$$\frac{1}{1-u} = \sum_{n=0}^{\infty} u^n \implies \frac{1}{1-(-x^2)} = \sum_{n=0}^{\infty} (-x^2)^n = \sum_{n=0}^{\infty} (-1)^n x^{2n}$$

We now need to integrate this Taylor series to obtain the one for $\tan^{-1}(x)$. We do so term-by-term,

$$\tan^{-1}(x) = \sum_{n=0}^{\infty} \int (-1)^n x^{2n}\, dx = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{2n+1}$$

but need to recall that we might need to add a constant term. The constant term of the Taylor series is $\tan^{-1}(0) = 0$, so we don't need to add anything in this case

What happens to the radius of convergence? Integrating or differentiating doesn't do anything, but substituting does. We know that the series converges for $|u| < 1$, but now $u = -x^2$, so

$$|-x^2| < 1 \implies |x|^2 < 1 \implies |x| < \sqrt{1} = 1$$

In general substitutions will change the radius of convergence, but in this case it happens to stay the same.

3. Day 3: Multivariable calculus

Topics covered: basics on vectors in 3 dimensions, planes, parametric equations, arc length and speed, limits and continuity in multiple variables, partial derivatives, differentiability and tangent planes, gradient and directional derivatives, multivariable chain rule, optimisation.

3.1. Vectors in $\mathbb{R}^3$.

Since multivariable calculus takes place with two variables (at least in general and at most for our purposes), graphs will occur in $2+1=3$ dimensions. Thus we should learn a little bit about $\mathbb{R}^3$. In particular, we need to familiarise ourselves with vector operations (that will reappear in generality for our linear algebra section).

A vector in $\mathbb{R}^3$ is a triple $\langle x, y, z\rangle$ which we think of as an arrow from $(0,0,0)$ to (x,y,z). Between any two points in $\mathbb{R}^3$ we can obtain another vector, namely

$$(a_1,a_2,a_3)\to(b_1,b_2,b_3)\sim\langle b_1-a_1, b_2-a_2, b_3-a_3\rangle$$

We'll need to recall the two main operations on vectors in $\mathbb{R}^3$. The first is common to every vector space, which is the dot product (or scalar product):

$$\langle a_1,a_2,a_3\rangle\cdot\langle b_1,b_2,b_3\rangle = a_1b_1+a_2b_2+a_3b_3$$

The dot product is linear in each argument, that is,

$$(c\vec{v})\cdot\vec{w} = c(\vec{v}\cdot\vec{w}) = \vec{v}\cdot(c\vec{w})$$

(at least in $\mathbb{R}$ – no complex conjugation here!) and

$$(\vec{u}+\vec{v})\cdot\vec{w} = \vec{u}\cdot\vec{w}+\vec{v}\cdot\vec{w}$$

It's also commutative, which is pretty obvious from its definition. We also define the norm of a vector using this:

$$\|\vec{v}\|^2 = \vec{v}\cdot\vec{v}$$

We say that $\vec{v}$ and $\vec{w}$ are *orthogonal* if $\vec{v}\cdot\vec{w}=0$. This comes from actual geometry: we can measure the angle between two vectors via the following relationship:

$$\vec{v}\cdot\vec{w} = \|\vec{v}\|\cdot\|\vec{w}\|\cdot\cos(\theta)$$

where θ is the planar angle between the vectors.

The second is the cross product, which only exists on $\mathbb{R}^3$ (at least in this form):

$$\langle a_1,a_2,a_3\rangle\times\langle b_1,b_2,b_3\rangle = \det\begin{bmatrix}\hat{\imath} & \hat{\jmath} & \hat{k}\\ a_1 & a_2 & a_3\\ b_1 & b_2 & b_3\end{bmatrix}$$

where $\hat{\imath}=\langle 1,0,0\rangle$, $\hat{\jmath}=\langle 0,1,0\rangle$, and $\hat{k}=\langle 0,0,1\rangle$ are the three unit basis vectors in $\mathbb{R}^3$. The cross product is linear in each variable as well, but it is anticommutative:

$$\vec{v}\times\vec{w} = -\vec{w}\times\vec{v}$$

The cross product is actually defined by a universal property: it is a bilinear operation such that $\vec{v} \times \vec{w}$ is orthogonal to both $\vec{v}$ and $\vec{w}$, that the ordered set $\{\vec{v}, \vec{w}, \vec{v} \times \vec{w}\}$ obeys the right-hand rule, and that

$$\|\vec{v} \times \vec{w}\| = \|\vec{v}\| \cdot \|\vec{w}\| \cdot \sin(\theta)$$

where θ is the planar angle between the two vectors. In particular, $\vec{v} \times \vec{w}$ gives a normal vector to the plane spanned by $\vec{v}$ and $\vec{w}$. But before we investigate that, let's look at an interesting (and useful) formula:

Proposition 3.1. Let $\vec{u}, \vec{v}, \vec{w}$ be three vectors in $\mathbb{R}^3$. Then we can form a 3D parallelogram (the *parallelopiped*) using these three vectors. Then its (signed) volume is given by

$$\vec{u} \cdot (\vec{v} \times \vec{w}) = \det \begin{bmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{bmatrix}$$

This could also be computed via $(\vec{u} \times \vec{v}) \cdot \vec{w}$, or by any cyclic permutation of the above formula. If any of the vectors $\vec{u}, \vec{v}, \vec{w}$ are parallel, the volume of the paralellopiped is zero.

3.2. **Planes.** There are two main ways to define a plane (in $\mathbb{R}^3$ at least). A plane is the solutions to a linear function in $\mathbb{R}^3$, so the equation looks like

$$ax + by + cz = d$$

for fixed $a, b, c, d \in \mathbb{R}$. The better way to think about it is to consider a plane as a set of vectors that are orthogonal to the normal vector to the plane:

$$\vec{n} \cdot \vec{v} = 0$$

But this defines a plane that passes through the origin. To move the plane to a point elsewhere, we move it by a fixed amount:

$$\vec{n} \cdot (\vec{v} - \vec{v}_0) = 0$$

But now $\vec{n} \cdot \vec{v}_0 = d$ for some $d \in \mathbb{R}$, so we obtain

$$\vec{n} \cdot \vec{v} = d$$

and letting $\vec{n} = \langle a, b, c \rangle$ and $\vec{v} = \langle x, y, z \rangle$ recovers the above formula.

The cross product is particularly convenient when solving the following problems:

Problem 3.2. Find the equation of the plane in $\mathbb{R}^3$ passing through $P = (0, 0, 1)$, $Q = (1, 0, 0)$ and $R = (1, 1, 1)$.

Solution. Any three non-colinear points defines a unique plane in $\mathbb{R}^3$, and we can take for granted that these points are non-colinear (or check quickly). Recall above we said that if we know that a plane is spanned by two vectors $\vec{v}$ and $\vec{w}$, then $\vec{v} \times \vec{w}$ is normal to the plane. If we know three points, we can come up with two vectors:

$$\vec{v} = \overrightarrow{PQ} = \langle 1, 0, -1 \rangle, \quad \vec{w} = \overrightarrow{PR} = \langle 1, 1, 0 \rangle$$

The cross product is

$$\vec{n} = \vec{v} \times \vec{w} = \langle 1, -1, 1 \rangle$$

Thus the equation of the plane is $x - y + z = d$, where d is some constant. We can compute it by plugging in any point that is already on the plane, say $(0, 0, 1)$. Hence $d = 0 - 0 + 1 = 1$, so the plane has the equation $x - y + z = 1$.

3.3. **Parametric curves.** We now need to look at parametrised curves in $\mathbb{R}^3$. These are functions $\vec{\gamma}r \colon \mathbb{R} \to \mathbb{R}^3$ which we write as $\vec{\gamma}(t) = \langle x(t), y(t), z(t) \rangle$. Skipping over some details, limits, continuity, and differentiability of these parametric functions is determined precisely by the component functions.

But what is the meaning of the derivative in this case? If it's done componentwise, then $\vec{\gamma}\,'(t) = \langle x'(t), y'(t), z'(t) \rangle$, so every time we plug in a time t we get a whole vector in $\mathbb{R}^3$ instead of a number. How are we supposed to use this to find the tangent line? Luckily, we can shed the chains of $\mathbb{R}^2$ and define lines parametrically instead. The most appropriate form of the linear approximation to a curve at $x = a$ in $\mathbb{R}^2$ looks like

$$L(x) = f(a) + f'(a)(x - a)$$

We can do something similar. We can define a parametric curve in $\mathbb{R}^3$ by

$$\begin{aligned} \vec{\ell}(t) &= \langle x(t_0) + x'(t_0)(t - t_0), y(t_0) + y'(t_0)(t - t_0), z(t_0) + z'(t_0)(t - t_0) \rangle \\ &= \vec{\gamma}(t_0) + \vec{\gamma}\,'(t_0) \cdot (t - t_0) \end{aligned}$$

So what do we have? We have the point $\vec{\gamma}(t_0)$ at which we are taking this linear approximation (and this is a line), and we have a slope $\vec{\gamma}\,'(t_0)$ which determines the line's direction.

Problem 3.3. Suppose we have an ordinary curve $y = f(x)$. Prove that the tangent line to the parametrised version of the curve $\vec{\gamma}(t) = \langle t, f(t) \rangle$ is the same as the usual tangent line.

Problem 3.4. Find the linear approximation to the curve $\vec{\gamma}(t) = \langle t^2, t^3, 2t - 1 \rangle$ at time $t = 2$.

What about arc length? We went over how to do this in $\mathbb{R}^2$ earlier: the arc length of the curve $y = f(x)$ is

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

which we obtained by trying to integrate $ds = \sqrt{(dx)^2 + (dy)^2}$. Now, we aren't going to want to integrate with respect to x, because these curves are functions of time t. Moreover, we have three components, so that

$$ds = \sqrt{(dx)^2 + (dy)^2 + (dz)^2}$$

This is the length of the diagonal of the infinitesimal cube with sides dx, dy, dz. Thus by 'factoring out' a dt from all these terms,

$$\begin{aligned} \int ds &= \int \sqrt{(dx)^2 + (dy)^2 + (dz)^2} \\ &= \int \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2}\, dt \\ & \int \|\vec{r}\,'(t)\|\, dt \end{aligned}$$

No surprise: we are integrating the length of the velocity, i.e. the (directionless) speed of the curve at every point. This is, in fact, the same formula as we were dealing with before. The planar curve $y = f(x)$ can be parametrised as $\langle t, f(t)\rangle$, which has derivative $\langle 1, f'(t)\rangle$ and thus speed $\sqrt{1 + f'(t)^2}$. Thus we can forget the old formula and stick with the new.

Problem 3.5. Compute the arc length of the helix $\vec{\gamma}(t) = \langle \sin(t), -\cos(t), t\rangle$ from $t = 0$ to $t = 2\pi$.

One may recall studying curvature or other horrible topics, but these don't seem to appear on the GRE so we will not revisit them.

3.4. **Multivariable functions.** We will give everything in terms of two variables for now, but the same could be done for 3 or n variables without changing definitions very much.

Suppose now that $f\colon \mathbb{R}^2 \to \mathbb{R}$ is a function. The definition of a limit is still the same.

Definition 3.6. We say that $\lim\limits_{(x,y)\to(a,b)} f(x,y) = L$ if for every $\varepsilon > 0$, there exists $\delta > 0$ such that whenever $\|(x,y) - (a,b)\| < \delta$, $\|f(x,y) - L\| < \varepsilon$.

Just as we don't really compute limits in $\mathbb{R}$ very explicitly, we also don't compute them in $\mathbb{R}^2$ (or higher). Mainly we just hope that we're working with continuous functions and move on with our day. But in case there's any doubt, we have a nice criterion that lets us (dis)prove that a limit exists.

Theorem 3.7. Prove that $\lim_{(x,y)\to(a,b)} f(x,y) = L$ if and only if for all $\vec{\gamma}\colon [-1,1] \to \mathbb{R}^2$ such that $\vec{\gamma}(0) = (a,b)$, then $\lim_{t\to 0} f(\vec{\gamma}(t)) = L$.

You can try to prove that theorem if you like; it's pretty standard in (many) real analysis courses. A consequence of the above theorem is that if you can find two paths $\vec{\gamma}_1$ and $\vec{\gamma}_2$ such that $\lim_{t\to 0} f(\vec{\gamma}_1(t)) \neq \lim_{t\to 0} f(\vec{\gamma}_2(t))$, then the 2D limit cannot exist. Another consequence is that, if you can find any single path $\vec{\gamma}$ such that $\lim_{t\to 0} f(\vec{\gamma}(t))$ does not exist, neither does the 2D limit.

Problem 3.8. Prove that $\lim_{(x,y)\to(0,0)} \frac{xy}{x^2+y^2}$ does not exist. Hint: find two paths that give different limits.

Proving that limits *do* exist is a lot harder and is accomplished by reducing to a 1-dimensional theorem.

Problem 3.9. Prove that $\lim_{(x,y)\to(0,0)} \frac{xy^2}{x^2+y^2} = 0$.

Solution. For this, we will recall polar coordinates, which we will use more tomorrow. If we convert the point (x,y) into polar coordinates, then we are instead taking the limit $r \to 0$ and under the substitution $x = r\cos(\theta)$, $y = r\sin(\theta)$, we need to solve

$$\lim_{r\to 0} \frac{r^3 \cos(\theta)\sin^2\theta}{r^2} = \lim_{r\to 0} r(\cos(\theta)\sin^2(\theta))$$

But now we can apply the squeeze theorem to the problem:

$$0 \le \lim_{r\to 0} |r(\cos(\theta)\sin^2(\theta))| \le \lim_{r\to 0} |r| = 0$$

and conclude that the middle limit must be zero as well.

In general, these types of limits exist when the numerator is a higher degree than the denominator and do not otherwise. In the case that we believe the limit to exist, polar coordinates and the squeeze theorem is usually the way to prove it.

Continuity is defined the same way using these limits. A function is continuous if the limit at all points in its domain is the actual value of the function. All functions from single variable calculus are still continuous in multivariable calculus, except that now we allow both x and y to appear.

3.5. **Partial derivatives.** Differentiability is defined slightly differently for multivariable functions. Instead of having one derivative, we have several.

Definition 3.10. The partial derivative of $f\colon \mathbb{R}^2 \to \mathbb{R}$ at (a,b) with respect to x is (if it exists) the limit

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

The partial derivative with respect to y is

$$\partial_y f(a,b) = \lim_{k\to 0} \frac{f(a,b+k)-f(a,b)}{k}$$

These definitions are well and good, but how do we actually compute partial derivatives? To compute $\partial_x f(x,y)$ we compute what is essentially the implicit derivative of $f(x,y)$ with respect to x but under the assumption that $\partial y/\partial x = 0$ – since y is now an independent variable. For example, let $f(x,y) = \sin(xy)$. Then

$$\partial_x f(x,y) = \cos(xy)\cdot(x\cdot \partial y/\partial x + y) = \cos(xy)\cdot y$$

Definition 3.11. We say that $f\colon \mathbb{R}^2 \to \mathbb{R}$ is differentiable at (a,b) if both partial derivatives $\partial_x f(a,b)$ and $\partial_y f(a,b)$ exist *and* $f(x,y)$ is locally linear at (a,b).

Local linearity is actually a very important information that makes sure that the specific choices of the x and y-partial derivatives (i.e. the directional derivatives along the x- and y-axes) don't just accidentally exist. This will lead to the appropriate definition of general directional derivatives below. Unfortunately, we won't ever be able to check local linearity, but we don't have to in general.

Proposition 3.12. If $\partial_x f(a,b)$ and $\partial_y f(a,b)$ exist and are continuous in a neighbourhood of (a,b), then $f(x,y)$ is differentiable at (a,b).

In this case, we can define the tangent plane to $f(x,y)$ at (a,b) and it is actually the linear approximation: the tangent plane is spanned by the partial derivative in the x-direction and the one in the y-direction. In general we can say that the tangent plane is spanned by any two directional derivatives so long as those directions are linearly independent. We didn't go over above how to parametrise a plane using two variables, but we can do so now:

$$P(s,t) = s\cdot\langle 1,0,\partial_x f(a,b)\rangle + t\cdot\langle 0,1,\partial_y f(a,b)\rangle + \langle a,b,f(a,b)\rangle$$

But this isn't particularly helpful for us, since we would like a form in terms of (x,y,z). Instead, we will define the tangent plane using its normal vector. Because we know two vectors on the plane, which moreover are linearly independent, we take their cross product:

$$\begin{aligned}\langle 1, 0, \partial_x f(a,b)\rangle \times \langle 0, 1, \partial_y f(a,b)\rangle &= \det \begin{bmatrix} \hat{i} & \hat{j} & \hat{k} \\ 1 & 0 & \partial_x f(a,b) \\ 0 & 1 & \partial_y f(a,b) \end{bmatrix} \\ &= \langle -\partial_x f(a,b), -\partial_y f(a,b), 1\rangle\end{aligned}$$

Hence our equation is $\langle -\partial_x f(a,b), -\partial_y f(a,b), 1\rangle \cdot \langle x-a, y-b, z-f(a,b)\rangle = 0$. If we work this out and rearrange it, it becomes

$$z = \partial_x f(a,b)(x-a) + \partial_y f(a,b)(y-b) + f(a,b)$$

which looks a lot like the equation for the tangent line, except now there are two slopes and two variables that need to be taken into account.

Problem 3.13. Compute the equation of the tangent plane at $(1,1,2)$ to the graph of $f(x,y) = x^2 + y^2$.

Now, what about taking multiple partial derivatives? In principle one can take both $\partial_x \partial_y f(x,y)$ and $\partial_y \partial_x f(x,y)$. Are these the same? Are we detecting the same change in both x and y in both cases? In general, no we are not, but in every case that we'll run into during the GRE, yes. The reason is Clairaut's theorem:

Theorem 3.14 (Clairaut's Theorem)**.** Suppose that $f\colon \mathbb{R}^2 \to \mathbb{R}$ is a function, and suppose that the second-order partial derivatives of f exist and are continuous in a neighbourhood of (a,b). Then $\partial_x \partial_y f(a,b) = \partial_y \partial_x f(a,b)$.

This is convenient and not necessarily expected, but it does make a particular technique in optimisation a whole lot more convenient later on. This also works in more than 2 variables when taking partial derivatives with respect to any two different independent variables.

Problem 3.15. Consider the function

$$g(x,y) = \begin{cases} xy\dfrac{x^2-y^2}{x^2+y^2} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0) \end{cases}$$

Prove that both $\partial_x \partial_y g(0,0)$ and $\partial_y \partial_x g(0,0)$ exist, but they aren't equal. Conclude that the mixed second-order partial derivatives can't be continuous. Note that you'll have to compute these derivatives using the limit definition.

3.6. **Gradient and directional derivatives.** Having done partial derivatives, it's fair to ask if there's anything resembling a 'total derivative' of the function, something that takes all the variables into account. There is.

Definition 3.16. For a differentiable function $f\colon \mathbb{R}^2 \to \mathbb{R}$, define the gradient of f at (a,b) to be

$$\nabla f(a,b) = \langle \partial_x f(a,b), \partial_y f(a,b)\rangle.$$

We can make the obvious modification for a function $F\colon \mathbb{R}^3 \to \mathbb{R}$.

What good is the gradient for us? It is a vector quantity instead of a scalar quantity, which is interesting, and it still satisfies some nice properties. Because it is made of partial derivatives, it is still linear, and it satisfies a product rule:

$$\nabla(f(x,y)\cdot g(x,y)) = f(x,y)\cdot \nabla g(x,y) + g(x,y)\cdot \nabla f(x,y)$$

where we think of f and g as scalar multiples (though depending on (x,y)). There is also a chain rule for the types of functions that we can actually compose at this point: suppose that $\varphi\colon \mathbb{R}\to\mathbb{R}$, so that $\varphi\circ f\colon \mathbb{R}^2 \to \mathbb{R}$ is still a multivariable function. Then

$$\nabla(\varphi\circ f) = \varphi'(f(x,y))\cdot \nabla f(x,y)$$

where again we think of the function $\varphi'\colon \mathbb{R}\to\mathbb{R}$ as acting by scalar multiplication.

Problem 3.17. Compute the gradient of

$$g(x,y,z) = (x^2+y^2+z^2)^8$$

We can now talk about directional derivatives in other directions. The partial derivatives are the derivatives in the direction $\langle 1,0\rangle$ or $\langle 0,1\rangle$, but we could have used any other unit vector. Recall that a unit vector is $\vec{u}$ such that $\|\vec{u}\| = 1$.

Definition 3.18. The directional derivative of $f\colon \mathbb{R}^2\to\mathbb{R}$ at (a,b) in the direction $\vec{u} = \langle h,k\rangle$ is the limit

$$\partial_{\vec{u}} f(a,b) = \lim_{t\to 0} \frac{f(a+th, b+tk) - f(a,b)}{t}$$

Note that these may not exist if the function is not differentiable at (a,b). In particular, the existence of x and y partial derivatives alone is not sufficient to conclude these exist. But in the case $f(x,y)$ is differentiable, we have the following:

$$\partial_{\vec{u}} f(a,b) = \nabla f(a,b)\cdot \vec{u}$$

which means that the above limit needs to be computed only rarely.

Problem 3.19. Prove that the directional derivatives of

$$f(x,y) = \begin{cases} \dfrac{xy^4}{x^2+y^8} & (x,y)\neq(0,0) \\ 0 & (x,y) = (0,0)\end{cases}$$

at $(0,0)$ exist and depend linearly on the gradient, but that $f(x,y)$ is not differentiable at $(0,0)$.

Problem 3.20. Prove that there is no function $f(x,y)$ such that $\nabla f(x,y) = \langle y^2, x\rangle$. Hint: Clairaut's theorem.

We have another version of the chain rule, where we compose a curve and a multivariable function to obtain a function $\mathbb{R} \to \mathbb{R}$.

Theorem 3.21 (Chain Rule II)**.** Let $\vec{\gamma}\colon \mathbb{R} \to \mathbb{R}^2$ be a differentiable curve and let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be a differentiable function. Then

$$\frac{d}{dt} f(\vec{\gamma}(t)) = \nabla f(\vec{\gamma}(t)) \cdot \vec{\gamma}\,'(t)$$

where this is now the dot product of the two vector-valued functions.

Problem 3.22. Prove this theorem from the definition of a single variable derivative.

The gradient is in the direction of greatest change on the graph $z = f(x,y)$. How do we see this? The directional derivatives of $f(x,y)$ tell us the rate of change in each direction. The direction $\vec{u}$ which makes the quantity $\nabla f(a,b)\cdot\vec{u}$ the most is the unit vector in the direction of $\nabla f(a,b)$ itself. Similarly, $-\nabla f(a,b)$ is the direction of greatest decrease. The quantity of the greatest rate of change requires normalising the gradient, which gives the computation

$$\nabla f(a,b) \cdot \frac{\nabla f(a,b)}{\|\nabla f(a,b)\|} = \|\nabla f(a,b)\|.$$

For another perspective on the gradient, suppose that we look at the level curves of the graph $z = f(x,y)$. These are the specific subsets $f(x,y) = c$ for a fixed constant $c \in \mathbb{R}$. Let $\vec{\gamma}_c(t)$ parametrise the curve, and consider a point $(a,b) = \vec{\gamma}_c(t_0)$ on this curve. Then the tangent vector to the curve is $\vec{\gamma}\,'_c(t)$, and we can examine $\nabla f(a,b) \cdot \vec{\gamma}\,'_c(t_0)$. By the chain rule,

$$\nabla f(a,b) \cdot \vec{\gamma}\,'_c(t_0) = \frac{d}{dt} f(\vec{\gamma}_c(t))\Big|_{t=t_0}$$

But on the curve $\vec{\gamma}_c(t)$, f is the constant c. Thus the above derivative is zero. This means that the gradient is orthogonal to the tangent vector to the level curve, i.e. it is normal to the level curve. This is also true in higher dimensions, though it's a bit more complicated to prove.

Problem 3.23. What is the greatest rate of change of $f(x,y) = x^4y^{-2}$ at the point $(a,b) = (2,1)$?

Solution. We can compute that the gradient is

$$\nabla f(x,y) = \langle 4x^3y^{-2}, -2x^4y^{-3}\rangle$$

so that $\nabla f(2,1) = \langle 32, -32\rangle$. The rate of greatest change is the norm of this vector, i.e. $32\sqrt{2}$.

The last thing to say is on the subject of surfaces on $\mathbb{R}^3$ which are defined using 3-variable functions. Consider a function $F\colon \mathbb{R}^3 \to \mathbb{R}$ and consider the set of points (x, y, z) such that $F(x, y, z) = c$. Assuming that F is a nice function (say, F is C^2 and ∇F is nowhere zero in all components), this defines a surface in $\mathbb{R}^3$, but usually one that isn't the graph of a function. If we consider the easiest example,

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

then we get a sphere, which we know isn't the graph of a function, but is the graph of two functions glued together.

Now, how do we find the tangent plane to such a surface? We clearly can't take the same approach because we don't have a function $f(x, y) = z$ to deal with. Instead, we need to figure out how to use $F(x, y, z)$. Suppose that $\vec{\gamma}(t)$ is a curve on the surface $F(x, y, z) = c$. Then by the chain rule,

$$\frac{d}{dt} F(\vec{\gamma}(t))\Big|_{t=t_0} = \nabla F(\vec{\gamma}(t_0)) \cdot \vec{\gamma}\,'(t_0).$$

But $F(\vec{\gamma}(t)) = c$ is a constant function, so its gradient is the zero vector. Thus the above dot product is also zero, so that $\nabla F(\vec{\gamma}(t))$ is orthogonal to to the curve $\vec{\gamma}(t)$ at any point. We conclude that ∇F is orthogonal to the surface $F(x, y, z) = c$ and can use it as the normal vector to the tangent plane. We see now the reason that ∇F should not be identically zero anywhere – it would mean that the 'normal vector' to a tangent plane is the zero vector (somewhere), implying something is wrong with the geometry of the situation.

Problem 3.24. What is the tangent plane to the surface $x^2 + y^2 + z^2 = 3$ at the point $(1, 1, 1)$?

Solution. This is defined by $F(x, y, z) = x^2 + y^2 + z^2$ and $c = 3$. We also have $\nabla F(x, y, z) = \langle 2x, 2y, 2z\rangle$. So at the point in question, we have

$$\nabla F(1, 1, 1) = \langle 2, 2, 2\rangle$$

so the tangent plane has the formula

$$2x + 2y + 2z = d$$

for some d. To find d, we just plug in a point that we know is on the plane, namely $(1, 1, 1)$. Once we note that $2 + 2 + 2 = 6$, we have

$$2x + 2y + 2z = 6 \text{ or } x + y + z = 3.$$

3.7. **Local extrema.** First, let's talk about finding local minima and maxima. Just as in single-variable calculus, these occur at critical points.

Definition 3.25. Let $f: \mathbb{R}^2 \to \mathbb{R}$ be a differentiable function. Then we say that $(a, b) \in \mathbb{R}^2$ is a critical point of $f(x, y)$ if $\nabla f(a, b) = \vec{0}$. That is, $\partial_x f(a, b) = 0$ and $\partial_y f(a, b) = 0$.

Of course, just like in single-variable calculus, while every extremum occurs at a critical point, not every critical point gives rise to an extremum. There are two methods in single-variable calculus to give us more information: the first derivative test and the second derivative test. Neither has an immediate analogue in multivariable calculus, but the second derivative test will turn out to be the solution.

But, as discussed above, there are four different 'second derivatives' of a given function. What we do is assemble them into a matrix called the Hessian of f as follows:

$$Hf = \begin{pmatrix} \partial_x \partial_x f & \partial_x \partial_y f \\ \partial_y \partial_x f & \partial_y \partial_y f \end{pmatrix}.$$

Then the second derivative test says the following:

Theorem 3.26 (Second Derivative Test)**.** Let $f: \mathbb{R}^2 \to \mathbb{R}$ be a function of class C^2. Suppose that (a, b) is a critical point of $f(x, y)$. Let $d = \det Hf(a, b)$ be the determinant of the Hessian matrix and $T = \operatorname{tr} Hf(a, b)$ be its trace. The following conclusions hold:

- If $d < 0$, then the point (a, b) is a saddle point.
- If $d > 0$ and $T < 0$, then the point (a, b) is a local maximum.
- If $d > 0$ and $T > 0$, then the point (a, b) is a local minimum.
- If $d = 0$, the test is inconclusive.

If it's hard to remember which condition corresponds to maximum and which to minimum, then just remember single-variable: if $f''(x) < 0$, we have a local maximum and if $f''(x) > 0$, we have a local minimum. The trace follows the same convention. It's also fun fact that, in the case that $d > 0$, $\partial_x^2 f$ and $\partial_y^2 f$ must have the same sign, so you can use one of those instead of the trace.

Remark 3.27. *Why* the second derivative test works requires some linear algebra to understand. Since f is of class C^2, Clairaut's theorem applies and thus the Hessian Hf is symmetric. A real symmetric matrix is diagonalisable by the spectral theorem (see below), and thus we have

$$\begin{pmatrix} \partial_x \partial_x f(a, b) & \partial_x \partial_y f(a, b) \\ \partial_y \partial_x f(a, b) & \partial_y \partial_y f(a, b) \end{pmatrix} \sim \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

when we plug in a (critical) point (a, b).

This matrix corresponds to second derivatives in the two essential directions which are describing the behaviour of $f(x, y)$ near (a, b). Thus we would want both directions to agree on what we're seeing. If $\lambda_1, \lambda_2 > 0$, then both directions think we are concave up and thus we should be at a minimum. If $\lambda_1, \lambda_2 < 0$, then both directions think we are concave down and thus we should be at a maximum. However, if λ_1 and λ_2 have different signs, then this means that we are concave up in one direction and concave down in another – a saddle point. We have a similar problem if λ_1 or λ_2 are equal to zero.

How does this reasoning apply to the second derivative test? The determinant of the Hessian is equal to $\lambda_1\lambda_2$. If λ_1 and λ_2 have the same sign, then $d > 0$. Otherwise, $d \leq 0$. The trace of the Hessian is equal to $\lambda_1 + \lambda_2$, which lets us figure out if both are positive or both are negative (in the case that $d > 0$).

The second derivative test for $\mathbb{R}^2$ takes advantage of a particular quirk: the product of two numbers is positive if and only if the numbers have the same sign. If we were to discuss local extrema in $\mathbb{R}^3$ or higher, we would end up needing to analyse *three* eigenvalues $\lambda_1, \lambda_2, \lambda_3$. It's impossible to tell if three numbers are all positive just from their product and sum: the triple $(3, -1, -1)$ has positive determinant and trace, but corresponds to a saddle point.

Let's have one example before moving on:

Problem 3.28. Find and classify all critical points of $f(x, y) = (x^2 + y^2)e^{-x}$.

Solution. First we need the gradient:

$$\begin{aligned}
\partial_x f(x, y) &= (x^2 + y^2)(-e^{-x}) + (2x)e^{-x} = (2x - x^2 - y^2)e^{-x} \\
\partial_y f(x, y) &= 2ye^{-x}
\end{aligned}$$

Starting with the y-derivative, we must have $y = 0$. Plugging that into the x-derivative,

$$\partial_x f(x, 0) = (2x - x^2)e^{-x} = (2 - x) \cdot x \cdot e^{-x}$$

giving us two solutions: $(2, 0)$ and $(0, 0)$, as e^{-x} will never equal zero. We now need to compute the Hessian. It's useful here to take advantage that Clairaut's theorem applies, so that $\partial_y\partial_x f = \partial_x\partial_y f$:

$$\begin{aligned}
\partial_x^2 f(x, y) &= (2x - x^2 - y^2)(-e^{-x}) + (2 - 2x)e^{-x} \\
&= (2 - 4x + x^2 + y^2)e^{-x} \\
\partial_x\partial_y f(x, y) &= -2ye^{-x} \\
\partial_y^2 f(x, y) &= 2e^{-x}
\end{aligned}$$

Thus we can compute some Hessians:

$$Hf(0, 0) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

We're already diagonal, so we can see that this corresponds to a local minimum.

$$Hf(2,0) = \begin{pmatrix} -2e^{-2} & 0 \\ 0 & 2e^{-2} \end{pmatrix}$$

Now the eigenvalues have opposite signs, so this corresponds to a saddle point.

3.8. **Optimisation and Lagrange Multipliers.** Let's now turn to global maxima. Optimisation in multivariable calculus works about the same as in single-variable calculus: first, figure out the region on which you are trying to optimise. Check for critical points of your function on the inside of your region, then check the boundary. It's not necessary to classify the critical points because, at the end of the day, we're just going to write down a list of values and pick the biggest and smallest one.

In single-variable, the boundary of a compact region (i.e. closed and bounded) is always a discrete set of points which you can check individually. In multivariable calculus, regions are two-dimensional so boundaries are one-dimensional. This means that the 'check the boundary' step in multivariable calculus is just an ordinary optimisation problem in single-variable calculus, which (having gotten this far in the course) you already know how to do.

Enough talk – let's have an example.

Problem 3.29. Find the maximum of the function $f(x,y) = x + y - x^2 - y^2 - xy$ on $[0,2] \times [0,2] \subset \mathbb{R}^2$.

Solution. We are working on a compact region so we are guaranteed a maximum, so that's a relief. First, find the gradient of the function:

$$\nabla f(x,y) = \langle 1 - 2x - y, 1 - 2y - x \rangle$$

We need to simultaneously solve $1 - 2x - y = 0$ and $1 - 2y - x = 0$. We've known how to do this since Algebra I, so we skip the step and find the critical point is at $(1/3, 1/3)$. If we were very bold, we would conclude that this must be maximum because it's the only critical point, but we need to check the boundary.

The boundary of this region is made up of four lines, which we need to parametrise using a single variable. The first edge is the bottom edge $(0,0) \to (2,0)$, for which we have the parametrisation $\vec{\gamma}_1(t) = \langle t, 0 \rangle$ for $t \in [0,2]$. This gives us a single-variable problem:

$$f_1(t) = f(\vec{\gamma}_1(t)) = t - t^2 \implies f_1'(t) = 1 - 2t$$

giving us a critical point (on this line) of $t = 1/2$, so the point $(1/2, 0)$ all in all. Noticing that $f(x,y) = f(y,x)$, we will obtain a critical point $(0, 1/2)$ on the left edge of the square.

Moving to the top edge, we have $\vec{\gamma}_2(t) = \langle t, 2 \rangle$ for $t \in [0,2]$. Solving as above,

$$f_2(t) = f(\vec{\gamma}_2(t)) = t + 2 - t^2 - 4 - 2t = -2 - t - t^2 \implies f_2'(t) = -1 - 2t$$

giving us a critical point at $-1/2$. This is outside our region, so we ignore it. By symmetry, we won't get anything on the right edge either.

The last step is to the check the boundaries of our boundary edges, which are the corners of the square: $(0,0)$, $(0,2)$, $(2,0)$, and $(2,2)$. Having assembled all our points, we now get a list of values:

$$\begin{aligned} f(0,0) &= 0 \\ f(0,2) = f(2,0) &= -2 \\ f(2,2) &= -8 \\ f(0,1/2) = f(1/2,0) &= 1/4 \\ f(1/3,1/3) &= 1/3 \end{aligned}$$

which proves that, indeed, the maximum was at the critical point $(1/3,1/3)$ all along. However, we now know that the minimum of the function occurs at $(-2,-2)$.

Problem 3.30. Find the global extrema of the function $f(x,y) = x^2 - x \cdot y$ on the ellipse $x^2 + 4y^2 \leq 4$.

Solution. The first step is to find the critical points of the function on the interior of the ellipse, which I will leave as an exercise. The problem comes with checking the boundary – it is certainly one-dimensional, but how do we parametrise it? Here is a sub-exercise for you to do:

Problem 3.31. The ellipse $x^2/a^2 + y^2/b^2 = 1$ is parametrised by $\vec{\gamma}(\theta) = \langle a\cos(\theta), b\sin(\theta)\rangle$ for $\theta \in [0, 2\pi]$.

Once you've done this problem, you'll be equipped to parametrise the boundary and complete the problem. Unlike the case of the square, we do not have a 'boundary of the boundary' in this case since the ellipse doesn't have endpoints.

We now turn to the special case of Lagrange multipliers. This applies to optimisation of functions $g\colon \mathbb{R}^3 \to \mathbb{R}$ on closed surfaces (i.e. compact surfaces without boundary) in $\mathbb{R}^3$ defined implicitly by $F(x,y,z) = 0$ (or $F(x,y,z) = c$ for any c, but by modifying F we can assume $c = 0$). It can also apply to optimisation on ellipses or circles in $\mathbb{R}^2$, but we will only demonstrate in the more difficult case.

Suppose that we are trying to find a maximum of $g(x,y,z)$ on $S = F^{-1}(0)$ for a C^2 function $F\colon \mathbb{R}^3 \to \mathbb{R}$. Let's pick a random point (a,b,c) (where we do not mean c as above). We can examine the gradient $\nabla g(a,b,c)$, which points in the direction of greatest change of g. We can use this direction to move along S to another point (a',b',c') near to our starting point such that $g(a',b',c') > g(a,b,c)$. There is one circumstance when that fails – if ∇g is pointing directly away from the surface S, we cannot travel in that direction at all.

But we already know what direction is directly away from S – it is $\nabla F(a,b,c)$. Thus:

Theorem 3.32 (Lagrange Multipliers)**.** Let F be a C^2 function so that $F(x,y,z)=0$ define a closed surface S in $\mathbb{R}^3$, and let $g\colon S\to\mathbb{R}$ be a C^1 function. Then g has its local extrema at those points (a,b,c) so that $\nabla F(a,b,c)$ and $\nabla g(a,b,c)$ are parallel, i.e. there exists $\lambda\in\mathbb{R}$ such that $\lambda\nabla F(a,b,c)=\nabla g(a,b,c)$.

Note that this includes the case $\nabla g=\vec{0}$ identically, which would correspond to a local maximum or minimum of $g(x,y,z)$ without constraining ourselves to S.

Problem 3.33. Find the point on the plane

$$\frac{x}{2}+\frac{y}{4}+\frac{z}{4}=1$$

closest to the origin in $\mathbb{R}^3$, then compute the distance.

Solution. As always, we need a constraint function and a function to optimise. The function to optimise is $d(x,y,z)=\sqrt{x^2+y^2+z^2}$, which is a bit messy. As we argued above, it suffices to minimise $d^2=g(x,y,z)=x^2+y^2+z^2$, which will have a much nicer gradient. Our constraint is $F(x,y,z)=x/2+y/4+z/4-1=0$. Computing gradients, we have

$$\nabla F(x,y,z)=\langle 1/2,1/4,1/4\rangle,\quad \nabla g(x,y,z)=\langle 2x,2y,2z\rangle.$$

Thus we are looking for a simultaneous solution to

$$\lambda/2=2x,\quad \lambda/4=2y,\quad \lambda/4=2z.$$

A key point to the theory of Lagrange multipliers is that we never need to compute λ, but we can use it symbolically to arrange all that we have. Solving each of those equations for λ tells us that

$$\lambda=4x=8y=8z\implies x=2y=2z$$

so we can use the one-variable substitution $y=x/2$ and $z=x/2$ to compute an actual point on this plane:

$$\frac{x}{2}+\frac{x/2}{4}+\frac{x/2}{4}=1\implies \frac{3}{4}x=1\implies x=\frac{4}{3},\quad y=z=\frac{2}{3}.$$

Answering the question, we have to plug all this in to the original distance function

$$d(4/3,2/3,2/3)=\sqrt{16/9+4/9+4/9}=\frac{2\sqrt{6}}{3}.$$

But wait! We never determined that this was a minimum! Fortunately, we can appeal to our other senses: it's very easy for a point on a plane to get far away from the origin, but it's difficult for it to be close. We should expect a minimum but no maximum. As such, any extremum we encounter should be a minimum.

If we're being extra fancy, we can compute the (three-dimensional!) Hessian for $g(x, y, z)$. Most of the second partial derivatives are zero, and the Hessian is diagonal with entries $(2, 2, 2)$.Thus we are in a permanent state of concave up, i.e. all local extrema are minima.

I leave you with a classical practice problem:

Problem 3.34. What is the maximum of the function $g(x, y, z) = xyz$ on the unit sphere?

You perhaps know intuitively what the answer should be, but see how the method of Lagrange multipliers bears out your intuition.

4. Day 4: Multivariable calculus

Topics covered: double and triple integrals, change of coordinates, surfaces in $\mathbb{R}^3$, vector fields, integration with vector fields, the fundamental theorems of vector calculus.

4.1. Double integrals.

How do you integrate in two variables? First, learn how to integrate boxes. You can do that using Riemann sums, but I really don't want to do that. It's conceptually important but not worth typing up in the grand scheme of things. Again, they are linear and you can separate them up and so on. There's a technical definition for when functions are integrable, but in all cases we care about it will suffice to know that continuous functions on bounded domains (with non-ridiculous boundaries) are differentiable.

For boxes, it doesn't matter whether you integrate over x or y first.

Theorem 4.1 (Fubini's Theorem)**.** Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be continuous and let $R = [a,b]\times[c,d]$ be a rectangle. Then

$$\iint_R f(x,y)\,dA = \int_a^b \int_c^d f(x,y)\,dy\,dx = \int_c^d \int_a^b f(x,y)\,dx\,dy$$

Then, for integrating functions over more complicated regions D which are not rectangles, you want to parametrise the region in terms of x in some range, then y as a function of x, then integrate y first then x. Or you can do it the other way, depending on exactly how your region looks. We call these *vertically-simple* and *horizontally-simple* respectively.

Problem 4.2. Integrate $f(x,y) = xy$ over the region bounded by $y = 4$ and $y = x^2$ in the first quadrant.

Solution. We can easily describe this region as $x \in [a,b]$ with $\varphi(x) \le y \le \psi(x)$. Since we are in the first quadrant, we must start at $x = 0$. The end point is where these two curves intersect, which occurs at $(2,4)$. Thus we would like to integrate:

$$\int_0^2 \int_{x^2}^4 xy\,dy\,dx$$

Why have we ordered the y-integral like this? Drawing out the region shows that $y = x^2$ is on bottom and $y = 4$ is on top. When performing multivariable integrals, if we are integrating with respect to y we just pretend that x is a constant (because it is for our purposes):

$$\int_{x^2}^4 xy\,dy = \frac{xy^2}{2}\Big|_{x^2}^4 = 8x - \frac{x^5}{2}.$$

This shows why we are integrating with respect to y first. If we were to do this integral second, our final answer would still have variables, which is suboptimal for a definite integral. But now we integrate with respect to x and all our variables will vanish:

$$\int_0^2 8x - \frac{x^5}{2}\,dx = 4x^2 - \frac{x^6}{12}\Big|_0^2 = 16 - \frac{64}{12} = \frac{32}{3}.$$

If our regions are oriented in the other fashion, we should integrate first with respect to x then with respect to y. As an example,

Problem 4.3. Compute the area between the curves $x = y^2$ and $x = 2y$ in the first quadrant.

To find the area of the region, just integrate the function $f(x, y) = 1$. The hard part is setting up the bounds, which I leave to you.

When our function is not just $f(x, y) = 1$, the integral over R is the volume under the surface $z = f(x, y)$ in $\mathbb{R}^3$ which lies over R in the xy-plane. This means that instead of calculating the area between curves, we can calculate the volume of a region between surfaces. In particular, when a certain region has a nice boundary with respect to the xy-plane and its upper and lower boundaries are nice functions of x, y, we're in business. We can also do this with triple integrals.

4.2. **Triple Integrals.** Really, there's not a whole lot different here, except that we have three variables instead of two. Riemann sums are Riemann sums, except now they're one dimension more annoying. Supposing that we can parametrise our region $W \subset \mathbb{R}^3$ anagolously, so that its boundary is of the form $z_1(x, y) \leq z \leq z_2(x, y)$ on a region D in the xy-plane, then

$$\iiint_W f(x, y, z)\,dV = \iint_D \left(\int_{z_1(x,y)}^{z_2(x,y)} f(x, y, z)\,dz \right) dA$$

Moreover, if we want the volume of W, we just integrate the function 1.

Problem 4.4. Evaluate $\displaystyle\iiint_W z\,dV$ where W is the region between the planes $z = x + y$ and $z = 3x + 5y$ over the rectangle $[0, 3] \times [0, 2]$. Then compute the volume of W.

Solution. Because $x, y \geq 0$, it's clear that the plane $z = 3x + 5y$ is on top. Thus we have $x + y \leq z \leq 3x + 5y$ for our z-boundary. This sets up the triple integral.

$$\begin{aligned}\iint_R \int_{x+y}^{3x+5y} z\,dz\,dA &= \iint_R \frac{(3x+5y)^2}{2} - \frac{(x+y)^2}{2}\,dA \\ &= \iint_R 4x^2 + 14xy + 12y^2\,dA\end{aligned}$$

and now we just have to integrate over the rectangle, which is pretty straightforward, thus left as an exercise.

To find the volume, we have two conceptual choices that amount to the same integral. First, it's the region between two surfaces, so by the brief comment above, we could solve

$$\iint_R (3x + 5y) - (x + y)\, dA$$

That is, we want to integrate over the region R the difference in the heights of these functions. Alternatively, we perform the same integral by plugging in 1 instead of z:

$$\iint_R \int_{x+y}^{3x+5y} 1\, dz\, dA = \iint_R (3x + 5y) - (x + y)\, dA$$

which amounts to the same thing. This is an even easier computation that I will not do.

The next problem is slightly more confusing.

Problem 4.5. Integrate $f(x, y, z) = x$ over the region W bounded above by $z = 4 - x^2 - y^2$ and below by $z = x^2 + 3y^2$ in the first octant.

Solution. In order to parametrise the region in the xy-plane over which W lies we need to compute the intersection of the surfaces. This turns out to be an ellipse:

$$4 - x^2 - y^2 = x^2 + 3y^2 \implies 4 = 2x^2 + 4y^2$$

Call the quarter of this ellipse we care about E. Our integral is thus

$$\iint_E \int_{x^2+3y^2}^{4-x^2-y^2} x\, dz\, dA$$

We need to solve the ellipse in terms of x or y, and we might as well pick x. We have $x = \pm\sqrt{2 - 2y^2}$. We also know that $x \geq 0$ and $y \geq 0$ in the part we care about, so we will pick $0 \leq x \leq \sqrt{2 - 2y^2}$. The bounds of y are $0 \leq y \leq 1$. Therefore we can set up our integral and go:

$$\int_0^1 \int_0^{\sqrt{2-2y^2}} \int_{x^2+3y^2}^{4-x^2-y^2} x\, dz\, dx\, dy.$$

The answer is 16/15, and the computation is left as practice.

Alternatively, if you want to read the next section first, it's less messy to compute this integral in polar/cylindrical coordinates.

4.3. **Change of coordinates.** There are other coordinate systems that we greatly prefer in the case of roundness. In two dimensions, we already remembered polar coordinates to do some limit computations. We even recalled how to parametrise an ellipse in the last section. Discs, annuli, and their sections are the 'rectangles' of polar coordinates. We can recall that

$$x = r\cos(\theta), y = r\sin(\theta) \iff x^2 + y^2 = r^2, \tan\theta = \frac{y}{x}$$

converts between the two. But is doing an integral like $\iint_R f(x,y)\,dx\,dy$ as easy as $\iint_R f(r,\theta)\,dr\,d\theta$?

No, it's not. The problem is that $dr\,d\theta$ is not the same area as $dx\,dy$. In fact, we can draw the usual picture and prove that

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

Thus swapping your integral into polar coordinates is almost as easy as posited.

Problem 4.6. Compute the area of the unit disk using polar coordinates.

Solution. The unit circle is described as $\theta \in [0, 2\pi]$ and $r \in [0,1]$. Thus its area is

$$\int_0^{2\pi}\int_0^1 r\,dr\,d\theta = 2\pi \cdot \frac{r^2}{2}\Big|_0^1 = \pi.$$

Note that if we forget to include that r, we get

$$\int_0^{2\pi}\int_0^1 1\,dr\,d\theta = 2\pi$$

which is a wrong answer.

So that's for two dimensions; what about three? There's an analogue of polar coordinates called cylindrical coordinates, which just adds z as the third variable. It's another easy computation that $dx\,dy\,dz = r\,dr\,d\theta\,dz$. These coordinates are best used with surfaces or regions that have nice symmetry when rotating around the z-axis but not for any other types of rotation, for example cones, cylinders, and hyperboloids or paraboloids.

Spherical coordinates are must useful for spheres, and can occasionally be useful in other situations. The conversions are as follows:

$$x^2 + y^2 + z^2 = \rho^2, \quad \cos(\varphi) = \frac{z}{\rho}, \quad \tan(\theta) = \frac{y}{x}.$$

Conversely (and more usefully),

$$x = \rho\sin(\varphi)\cos(\theta), \quad y = \rho\sin(\varphi)\sin(\theta), \quad z = \rho\cos(\varphi)$$

We give the answer first and describe its computation in the subsequent remark:

$$dx\,dy\,dz = \rho^2 \sin(\varphi)\,d\rho\,d\varphi\,d\theta$$

Remark 4.7. Suppose that we have a change of coordinates $x = f(a,b,c)$, $y = g(a,b,c)$, and $z = h(a,b,c)$. Then this defines a function $\Phi\colon \mathbb{R}^3 \to \mathbb{R}^3$ given by $\Phi(a,b,c) = (f(a,b,c), g(a,b,c), h(a,b,c))$. The first partial derivatives of Φ give us a 3×3 matrix called the Jacobian:

$$J_\Phi = \begin{bmatrix} \partial_a f & \partial_b f & \partial_c f \\ \partial_a g & \partial_b g & \partial_c g \\ \partial_a h & \partial_b h & \partial_c h \end{bmatrix}$$

Then the determinant of the Jacobian is exactly the scaling factor we need to convert from (x,y,z) to (a,b,c):

$$\iiint_W f(x,y,z)\,dx\,dy\,dz = \iiint_W f(\Phi(a,b,c)) \cdot \det(J_\Phi)\,da\,db\,dc$$

To demonstrate with cylindrical coordinates, we let $a = \rho$, $b = \varphi$, and $c = \theta$ for the sake of ordering. Then

$$J_\Phi = \begin{bmatrix} \sin(\varphi)\cos(\theta) & \rho\cos(\varphi)\cos(\theta) & -\rho\sin(\varphi)\sin(\theta) \\ \sin(\varphi)\sin(\theta) & \rho\cos(\varphi)\sin(\theta) & \rho\sin(\varphi)\cos(\theta) \\ \cos(\varphi) & -\rho\sin(\varphi) & 0 \end{bmatrix}$$

It's easiest to compute the determinant if we expand by minors along the bottom row.

$$\begin{aligned} \det(J_\Phi) =& \cos(\varphi)(\rho^2\sin(\varphi)\cos(\varphi)\cos^2\theta + \rho^2\sin(\varphi)\cos(\varphi)\sin^2\theta) \\ &+ \rho\sin(\varphi)(\rho\sin^2(\varphi)\cos^2\theta + \rho\sin^2(\varphi)\sin^2\theta) + 0 \end{aligned}$$

Both terms have a $\sin^2 + \cos^2$ factor in them, which we can factor out to get 1. So our first simplification is

$$\begin{aligned} \det(J_\Phi) &= \cos(\varphi)(\rho^2\sin(\varphi)\cos(\varphi)) + \rho\sin(\varphi)(\rho\sin^2(\varphi)) \\ &= \rho^2\sin(\varphi)\cos^2(\varphi) + \rho^2\sin^3(\varphi) \end{aligned}$$

This also has a $\sin^2(\varphi)+\cos^2(\varphi)$ we can factor out, which leaves $\rho^2\sin(\varphi)$ as promised.

Problem 4.8. Compute the volume of the region between the surfaces $z = x^2 + y^2$ and $z = 8 - x^2 - y^2$.

Solution. The first surface is on bottom and the second is on top. The intersection between these surfaces is

$$x^2 + y^2 = 8 - x^2 - y^2 \implies x^2 + y^2 = 4$$

which is the circle of radius 2. Thus we can compute this volume as a cylindrical integral over the disc D given by $r \leq 2$. The first thing is rephrasing the integrand in terms of polar coordinates.

$$8 - x^2 - y^2 - (x^2 - y^2) = 8 - 2(x^2 + y^2) = 8 - 2r^2$$

Thus:

$$\begin{aligned}\iint_D 8 - 2r^2\, dA &= \int_0^{2\pi} \int_0^2 (8 - 2r^2) r\, dr\, d\theta \\ &= 2\pi \cdot \int_0^2 8r - 2r^3\, dr \\ &= 2\pi \cdot \left(4r^2 - \frac{r^4}{2} \Big|_0^2 \right) = 2\pi \cdot (16 - 8) = 16\pi.\end{aligned}$$

4.4. **Quadric surfaces.** Now would probably be a good time to go over quadric surfaces, i.e. the basic surfaces we will encounter in $\mathbb{R}^3$.

Ellipsoid. These are the analogue of ellipses, and look basically the same: for positive numbers $a, b, c \in \mathbb{R}$, we have

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$$

the ellipse with radii a, b, c in the x, y, z directions respectively. The volume of such an ellipsoid is $(4/3)\pi abc$, as one might expect.

Elliptic paraboloid These can be written as the graph of a function, namely

$$z = ax^2 + by^2 + c$$

for any $c \in \mathbb{R}$ and $a, b \in \mathbb{R}$ with the same sign. If a, b have different signs, we have a different situation which is covered below.

Why is this called elliptic? Assume $a, b > 0$. At each fixed value $z = d$, we get one of two options: if $d \geq c$, then our level curve has the equation $ax^2 + bx^2 = d - c$ which gives an ellipse (potentially a degenerate one at $d = c$. If $d < c$, then the equation $ax^2 + bx^2 = d - c$ has no solution because $d - c$ is negative. Therefore we only get elliptical slices *above* a certain z-value. If $a, b < 0$, we only get elliptic slices *below* a certain z value.

To justify the term paraboloid, we can look at the coordinate traces, i.e. the projections of this surface onto the xz- and yz-planes. If we fix $y = 0$, we get a parabola $z = ax^2$ in the xz-plane, and similarly if we fix $x = 0$ we get $z = by^2$. This explains both parts of the naming convention.

Of course, in this version the elliptical slices are all parallel to the xy-plane and enclose the z-axis. It's also possible to permute the variables for other options, e.g.

$$y = x^2 + z^2, \quad x = 2y^2 + 3z^2.$$

Hyperbolic paraboloid. If a, b have different signs, we no longer get ellipses for our level curves. Suppose that we take $z = x^2 - y^2$ as a simple example. Then the slices $z = d$ are of the form $d = x^2 - y^2$, which we can rearrange to obtain $y = \pm\sqrt{x^2 - d}$. This is the formula of a hyperbola. If we again look at the coordinate traces by setting $x = 0$ or $y = 0$, we obtain two parabolas, except that one is facing up and one is facing down – hence hyperbolic paraboloid.

This type of shape is incredibly difficult to draw, but in our example $(0, 0)$ is a saddle point. In fact, every saddle point on a surface locally looks like $z = x^2 - y^2$ or $z = y^2 - x^2$. Thus hyperbolic paraboloids are these Pringle-shaped graphs that we learn about when studying the second derivative test.

Hyperboloid of one sheet. What if we have an ellipsoid but then flip one of the signs? Then we can arrange it to obtain

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = \frac{z^2}{c^2} + 1$$

up to permutation of variables. Then if we set $x = 0$ or $y = 0$ we obtain again the formula of a hyperbola, and now the slices at fixed values of z are ellipses. This is not a combination we have seen before and we baptise it hyperboloid. You'll want to Google what these look like. If z is the isolated variable on the other side of the equation, then we see that the elliptical slices are again parallel to the xy-plane.

Hyperboloid of two sheets. Suppose that flip two of the signs on an ellipsoid. Then up to permuting variables, we obtain

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = \frac{z^2}{c^2} - 1$$

A fair question is how this differs from the last example. It doesn't really – we still obtain hyperbolas if $x = 0$ or $y = 0$ and the horizontal slices are ellipses. But now what if we plug in $z = 0$? Then we have to solve

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = -1$$

but this has no solutions. In fact, we need $|z| \geq c$ for there to be any points in x, y that satisfy the equation. Thus the two halves of the hyperboloid are separated from each other, i.e. there are two separate 'sheets'.

Cone. A special case is the intermediate point between the two kinds of hyperboloids:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = z^2$$

where we imagine we have multiplied through by c^2 and reorganised our constants a, b. Then when $z = 0$, there is only one point $(0, 0, 0)$ on the level set. Thus our two sheets are joined at a single point, and it's not hard to see that we have a cone. If we set $x = 0$ or $y = 0$, we get (for example)

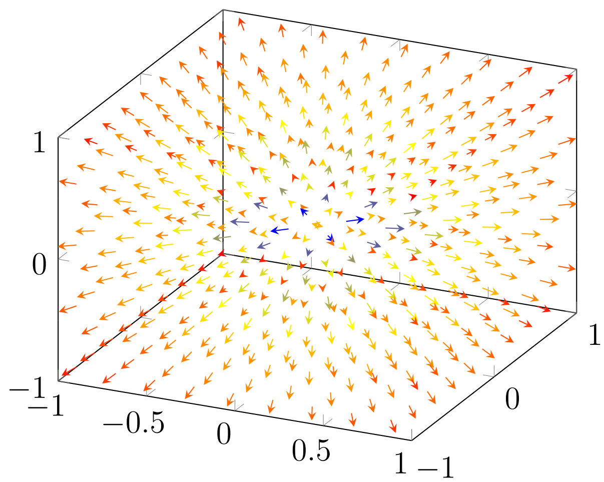
$$\frac{x^2}{a^2} = z^2 \implies \pm\frac{x}{a} = z$$

which is a pair of lines intersecting at the origin. This certainly feels like the slice of a cone (as it's nice and pointy).

4.5. **Vector fields and fancier integration.** We now turn to the second kind of integration in multivariable calculus, namely those involving vector-valued functions.

Definition 4.9. A vector field is a function $\vec{F} \colon \mathbb{R}^n \to \mathbb{R}^n$, which we think of as assigning to each point in $\mathbb{R}^n$ a vector in $\mathbb{R}^n$ beginning at that point.

At this point, I would draw a picture, or steal one from StackExchange[1]



There are many vector fields in real life, two easily coming to mind are the gravitational vector field which expresses the force (and direction) due to gravity on any object in space, and on each we can talk about the vector field of wind – to each point on each we can assign the vector of which direction (and speed) the wind is blowing.

The most boring example is when we are still working with real-valued functions. If we want to integrate a real-valued function $f \colon \mathbb{R}^3 \to \mathbb{R}$ over some curve C in $\mathbb{R}^3$ (or a surface, but let's stick with curve), then we think of C as being the image of some $\gamma \colon [a, b] \to \mathbb{R}^3$). There are a few technical assumptions we want to make on the function γ: it should be C^1, it should be injective (so that the curve C is traced out exactly once), and the derivative $\gamma'(t)$ should vanish nowhere. For this last condition,

[1]https://tex.stackexchange.com/questions/328036/velocity-field-3d-vector-fields-in-tikz-or-pgfplots

the justification is that if we are actually traversing the curve C, there's no reason that $\gamma(t)$ should grind to a complete halt at any point.

Plugging in this parametrization, the function we are considering is

$$f(\gamma(t))\colon \mathbb{R} \to \mathbb{R}$$

so we might think the integral is just $\displaystyle\int_a^b f(\gamma(t))\,dt$. But this isn't quite enough, because the parametrisation matters. If it didn't then two different γ which describe the same C would give us different integrals; this is a problem.

Problem 4.10. The unit circle S^1 in $\mathbb{R}^2$ (oriented counterclockwise) can be parametrised by $\gamma_\alpha(t) = \langle \cos(\alpha t), \sin(\alpha t)\rangle$ for any $\alpha > 0$ on the interval $t \in [0, 2\pi/\alpha]$. Using the function $f(x,y) = \sqrt{x^2+y^2}$, conclude that

$$\int_0^{2\pi/\alpha} f(\gamma_\alpha(t))\,dt$$

is the wrong version of the integral, as it depends on α.

The integral should actually be taken ds rather than dt, i.e. it's with respect to an infinitesimal amount of arc length. Put another way, we need to make sure that the speed at which we are traversing this curve C is taken into account, which is solved by

$$\int_C f(x,y,z)\,ds = \int_a^b f(\gamma(t))\|\gamma'(t)\|\,dt$$

Note that for this equation to be without problems, we want to assume that $\gamma'(t)$ and $f(x,y,z)$ are continuous. This type of scalar integral is actually not very relevant for the GRE; vector integrals are more important.

Problem 4.11. Repeating the preceding problem with the correct formula to show that the integral does not actually depend on α.

Now, suppose we are thinking physics, and we want to know something like 'how much energy does it take to fight gravity or the wind'? This would involve a vector field $\vec{F}\colon \mathbb{R}^3 \to \mathbb{R}^3$ and a curve C. In this case, whenever the curve travels in the same direction as the vector field, we would like to value that positively (going with the flow), and negative when the curve travels against it. Remember that C we cannot think of as just a 1-dimensional object in $\mathbb{R}^3$, it comes with an orientation – it has a back and a front, and the function $\gamma\colon [a,b] \to \mathbb{R}^3$ we use needs to take this into account.

What vector operation determines whether things go in the same direction? The dot product. What gives the (linear) direction the curve is going? Its tangent vectors $\gamma'(t)$. Thus:

Definition 4.12. The line integral of a vector field $\vec{F}$ along a curve C, parametrised by $\gamma\colon [a,b] \to \mathbb{R}^3$, is

$$\int_C \vec{F} \cdot dr = \int_a^b F(\gamma(t)) \cdot \gamma'(t)\, dt$$

Both of these quantities are vectors, so it makes sense to dot them. Another way the expression $\vec{F} \cdot dr$ is sometimes written is $F_1\, dx + F_2\, dy + F_3\, dz$, where these are the component functions of $\vec{F}$. This will come up later.

Problem 4.13. Compute the line integral of $\vec{F} = \langle z, y^2, x\rangle$ along the curve $\gamma(t) = (t+1, e^t, t^2)$ for $t \in [0,2]$.

Solution. We compute that

$$\gamma'(t) = \langle 1, e^t, 2t\rangle$$

So we need to compute

$$\begin{aligned}\int_0^2 \vec{F}(\gamma(t)) \cdot \gamma'(t)\, dt &= \int_0^2 \langle t^2, e^{2t}, t+1\rangle \cdot \langle 1, e^t, 2t\rangle\, dt \\ &= \int_0^2 t^2 + e^{3t} + 2t^2 + 2t\, dt = \int_0^2 2t + 3t^2 + e^{3t}\, dt \\ &= t^2 + t^3 + e^{3t}/3\Big|_0^2 = 12 + e^6/3 - 1/3\end{aligned}$$

What are some basic properties of the line integral? They are still linear, and now if one reverses the orientation of the curve C, this is like swapping a, b on the righthand side of the above equation, hence negates the integral. The last thing is that stringing together multiple curves end-to-end gives a sum of integrals.

4.6. **Conservative vector fields.** These are just the best. Our prototype here is a vector field F that arises as ∇f for some $f\colon \mathbb{R}^3 \to \mathbb{R}$. Such a vector field is called conservative. Then we can use the fundamental theorem of calculus to evaluate

$$\int_C \vec{F} \cdot dr = \int_a^b \nabla f(\gamma(t)) \cdot \gamma'(t)\, dt = \int_a^b \frac{d}{dt} f(\gamma(t))\, dt = f(\gamma(b)) - f(\gamma(a))$$

But $\gamma(b)$ and $\gamma(a)$ are just the endpoints of the curve, so the actual curve C doesn't matter in this case. Any vector field for which this happens is called path-independent.

Similarly, if C is a closed curve, its endpoints are the same, so

$$\oint_C \vec{F} \cdot dr = 0$$

for any conservative vector field $\vec{F}$. In this case, we don't even need to figure out the function f such that $\vec{F} = \nabla f$. But you might say: how do we know F is conservative if we don't have an explicit description $\vec{F} = \nabla f$? The relevant theorem is in a few paragraphs.

Problem 4.14. Verify that $\vec{F}(x, y, z) = \langle 2xy + z, x^2, x\rangle$ is the gradient of a function, then evaluate the line integral over the curve $\gamma(t) = (\sin(t)\cos(\pi t), e^t, 4t^3 - 1)$ for $t \in [0, 1/2]$.

One might ask if there are path-independent vector fields that do not arise as the gradient of some function. The answer is, essentially, no.

Theorem 4.15. A vector field F on an open, connected domain D is path-independent if and only if it is conservative.

Now, let us go over some of the other vector derivatives that will become useful shortly. The first is the divergence of a vector field,

$$\operatorname{div} \vec{F}(x, y, z) = \nabla \cdot \vec{F} = \partial_x F_1 + \partial_y F_2 + \partial_z F_3$$

and the second is the curl of a vector field, which only makes sense in $\mathbb{R}^3$:

$$\operatorname{curl} \vec{F}(x, y, z) = \nabla \times \vec{F} = \det \begin{bmatrix} \hat{i} & \hat{j} & \hat{k} \\ \partial_x & \partial_y & \partial_z \\ F_1 & F_2 & F_3 \end{bmatrix}$$

Problem 4.16. Prove that if $\vec{F} = \nabla f$ is conservative for a C^2 function $f(x, y, z)$, then $\operatorname{curl} \vec{F} = \vec{0}$.

The converse to this problem is not always true, but it is true in a great many cases.

Theorem 4.17. Suppose that D is an open simply-connected domain. A vector field F on D is conservative if and only if $\operatorname{curl} F = \vec{0}$.

Hence one should know if the domains they are working on are simply-connected, which (we remember) means that all loops in D be contracted to a point. That means that something like $x^2 + y^2 < c$ is okay but punctured regions like $\mathbb{R}^2 \setminus \{(a, b)\}$ are not.

4.7. **Surface integrals.** In order to integrate using surfaces in $\mathbb{R}^3$, we need to be able to parametrise them like

$$G(u, v) = (x(u, v), y(u, v), z(u, v))$$

where (u, v) are in some region D in $\mathbb{R}^2$. For example, the graph of the function $z = f(x, y)$ in $\mathbb{R}^3$ is easily parametrised by $(x, y, f(x, y))$. This is our prototype, but of course not all surfaces in $\mathbb{R}^3$ are graphs.

Suppose we want to parametrise the cylinder $x^2 + y^2 = 1$ in $\mathbb{R}^3$. This is best done with cylindrical coordinates (of course), yielding an easy parametrisation $(1, \theta, z)$ for $\theta \in [0, 2\pi]$ and $z \in \mathbb{R}$. Since the radius is fixed, we only get two variables.

We can parametrise spheres similarly using spherical coordinates. There's only a slight problem with this picture, as we get kind of an overlap at $\theta = 0$ and $\theta = 2\pi$, but we will not concern ourselves overmuch with this.

Okay, what are we doing with surfaces? Given a vector field, we want to measure how much the vector field is flowing through the surface. But what does flowing through the surface mean? 'Through' in this case will mean a direction normal to the surface (i.e. orthogonal to tangent). If $G(u, v)$ parametrises the surface S, we have two vectors $\partial_u G(a, b)$ and $\partial_v G(a, b)$ which are always tangent to S at the point $(a, b, G(a, b))$. As discussed earlier, in $\mathbb{R}^3$ we have a natural operation that gives a direction normal to the plane spanned by two vectors: the cross product $N(a, b) = \partial_u G(a, b) \times \partial_v G(a, b)$. Note that if the partial derivatives are parallel, the cross product becomes zero (which we consider pathological), so we want to choose our parametrisations wisely so that this never happens.

Since we can find the normal vector to the tangent plane, this gives us a nice way to describe the tangent plane to a surface that is defined parametrically.

Problem 4.18. Compute the tangent plane to the surface $G(\theta, z) = (2\cos(\theta), 2\sin(\theta), z)$ at $P = (\pi/4, 5, G(\pi/4, 5))$.

Now, the choice of N in the argument above is not quite canonical. We could easily reverse the orientation on the parametrisation by using the equation $G(v, u)$ so that the order in which we take the cross product is swapped. This results in flipping the normal vector completely around, i.e. multiplying it by -1. We therefore make the following choices once and for all: for the surface given by the graph of a function, we will take the upwards direction to the canonical orientation for the normal vector, i.e. the one going in the positive z-direction $\langle -\partial_x, -\partial_y, 1\rangle$. For a surface defined parametrically, we will make the choice so that the normal vector points outward.

Problem 4.19. Compute the normal vector to the unit sphere $S^2 \subset \mathbb{R}^3$ using the spherical parametrisation.

Solution. The unit sphere is not the graph of a function, but it is easily described in spherical coordinates; from the formula

$$(\rho \sin(\varphi)\cos(\theta), \rho \sin(\varphi)\sin(\theta), \rho\cos(\varphi)) = (x, y, z)$$

we fix the radius $\rho = 1$ to obtain a two-variable parametrisation. The sphere S^2 is parametrised by

$$G(\varphi, \theta) = (\sin(\varphi)\cos(\theta), \sin(\varphi)\sin(\theta), \cos(\varphi))$$

with $\varphi \in [0, \pi]$ and $\theta \in [0, 2\pi]$. Therefore to compute the normal vector, we have

$$\partial_\varphi G = \langle \cos(\varphi)\cos(\theta), \cos(\varphi)\sin(\theta), -\sin(\varphi)\rangle, \quad \partial_\theta G = \langle -\sin(\varphi)\sin(\theta), \sin(\varphi)\cos(\theta), 0\rangle$$

An unpleasant but straightforward computation shows that

$$\partial_\varphi G \times \partial_\theta G = \langle \sin^2(\varphi)\cos(\theta), \sin^2(\varphi)\sin(\theta), \sin(\varphi)\cos(\varphi)\rangle = \sin(\varphi) \cdot G(\varphi, \theta)$$

This is definitely an outward pointing vector, as it is in the same direction as $G(\varphi, \theta)$ as $\sin(\varphi)$ is always non-negative. We should be a little bit cautious about this because when $\varphi = 0, \pi$, this normal vector vanishes. But this only happens at two points on the sphere, and as far as surface integrals are concerned the behaviour at discrete points doesn't matter.

Now, how do we perform scalar surface integrals, i.e. ones that ignore the vector field for the moment? Just like for scalar line integrals, we need to worry about how the parametrisation interacts with dA, i.e. how $dx\,dy$ compares to $du\,dv$. It turns out that, $\vec{N} = \partial_u G \times \partial_v G$ also encodes the area of the parallelogram with sides du, dv: it's $\|\vec{N}\|$. This is an easy computation using the $\sin(\theta)$ interpretation of the cross product. Therefore if we want to do our integral, we need to scale by this amount, just like we had to account for the speed in the case of line integrals.

Definition 4.20. Let $G(u, v)$ be a parametrisation of a surface $S \subset \mathbb{R}^3$ with domain D. Assume that G is C^1, one-to-one, and regular (i.e. the normal vector is nondegenerate). Then for a function $f\colon \mathbb{R}^3 \to \mathbb{R}$,

$$\iint_S f(x, y, z)\, dS = \iint f(G(u, v))\|\partial_u G(u, v) \times \partial_v G(u, v)\|\, du\, dv$$

Now, once we bring in the actual flow and a vector field $\vec{F}\colon \mathbb{R}^3 \to \mathbb{R}^3$, we need to use a dot product:

Theorem 4.21.

$$\iint_S \vec{F} \cdot dS = \iint_D \vec{F}(G(u, v)) \cdot \vec{N}(u, v)\, du\, dv$$

where again $\vec{N}(u, v) = \partial_u G(u, v) \times \partial_v G(u, v)$.

Problem 4.22. Compute the flux through the surface $G(u, v) = (u^2, v, u^3 - v^2)$ over $D = [0, 1]^2$ of the vector field $\vec{F} = \langle 0, 0, x\rangle$.

Solution. We need to compute first the normal vector, and so need $\partial_u G$ and $\partial_v G$:

$$\partial_u G(u, v) = \langle 2u, 0, 3u^2\rangle, \quad \partial_v G(u, v) = \langle 0, 1, -2v\rangle.$$

The zeroes make the cross product slightly nicer (details omitted):

$$\vec{N}(u, v) = \langle 2u, 0, 3u^2\rangle \times \langle 0, 1, -2v\rangle = \langle -3u^2, 4uv, 2u\rangle.$$

Is this upward pointing? Looking at the z-coordinate, it's always positive when $u \in [0,1]$, so we're in business.

We now need to compute the other part of our integrand:

$$\vec{F}(G(u,v)) = \langle 0, 0, u^2 \rangle \implies \vec{F}(G(u,v)) \cdot \vec{N}(u,v) = 2u^3.$$

The final computation is thus

$$\int_0^1 \int_0^1 2u^3 \, du \, dv = 1 \cdot \frac{u^4}{2}\Big|_0^1 = \frac{1}{2}.$$

4.8. **Fundamental Theorems of Vector Calculus.** These all unify nicely in differential topology, but not many of my readers will have that perspective before graduate school (I certainly didn't). Thus we will proceed one at a time and try to use whatever intuition is accessible to us. The first is Green's Theorem.

Theorem 4.23 (Green's Theorem)**.** Let D be a closed domain with ∂D a simple closed curve, oriented counterclockwise. Then

$$\oint_{\partial D} F_1 dx + F_2 dy = \iint_D (\partial_x F_2 - \partial_y F_1) \, dA$$

Use this when the line integral of the closed curve would be way too confusing to compute.

Problem 4.24. Verify Green's Theorem by computing the line integral over the unit circle C of $\vec{F}(x,y) = \langle xy^2, x \rangle$.

Solution. On the one hand, we may parametrise the unit circle by $\vec{\gamma}(t) = \langle \cos(t), \sin(t) \rangle$. Note that this is the correct counterclockwise orientation. Also note that $\vec{\gamma}\,'(t) = \langle -\sin(t), \cos(t) \rangle$. Thus

$$\begin{aligned} \oint_C \vec{F} \cdot dr &= \int_0^{2\pi} \vec{F}(\vec{\gamma}(t)) \cdot \vec{\gamma}\,'(t) \, dt \\ &= \int_0^{2\pi} \langle \cos(t)\sin^2(t), \cos(t) \rangle \cdot \langle -\sin(t), \cos(t) \rangle \, dt \\ &= \int_0^{2\pi} -\cos(t)\sin^3(t) + \cos^2(t) \, dt \end{aligned}$$

The integral of this first term is zero for the following reason: $\cos(t)\sin^3(t)$ is an odd function, as it is the product of an odd function and an even function. Because of the periodicity of trigonometric functions, we have

$$\int_0^{2\pi} \cos(t)\sin^3(t) \, dt = \int_{-\pi}^{\pi} \cos(t)\sin^3(t) \, dt.$$

Finally, the integral of an odd function over a symmetric domain is zero. However, the integral of $\cos^2(t)$ is not zero – as we computed earlier, it is π.

Using Green's theorem instead, we have

$$\oint_C \vec{F} \cdot dr = \iint_D 1 - 2xy \, dA$$

It would be better to convert to polar for this integral. Giving some of the steps (and reminding the reader that $2\sin(\theta)\cos(\theta) = \sin(2\theta)$),

$$\begin{aligned}\iint_D 1 - 2xy \, dA &= \int_0^{2\pi} \int_0^1 (1 - 2(r\cos(\theta))(r\sin(\theta)))r \, dr \, d\theta \\ &= \int_0^{2\pi} \int_0^1 r - 2r^3 \cos(\theta)\sin(\theta) \, dr \, d\theta \\ &= \int_0^{2\pi} \frac{1}{2} - \frac{1}{2}\sin(2\theta) \, d\theta \\ &= \pi\end{aligned}$$

In this case, neither integral was particularly nice. However, in the case that the integrand of the double integral is a constant, life is much better. For example, consider the vector field $\vec{F}(x, y) = \langle -y, x \rangle$. Then

$$\partial_x F_2 - \partial_y F_1 = 1 - (-1) = 2.$$

Thus integrating a closed curve along this vector field gives you twice the area it encloses. Look for phenomena like this when it seems that Green's theorem might be in play.

One key feature is that, even if your curve is not closed, you can close it up and appeal to a simpler area calculation, i.e. if you have to compute a line integral over half a circle, complete it to a whole circle.

Problem 4.25. Compute the line integral over the 'curve' of straight lines connecting $(1, 1)$ to $(0, 1)$ to $(0, 0)$ to $(1, 0)$ of the vector field $\vec{F}(x, y) = \langle x^2 - y^2, 2xy \rangle$.

Solution. Now, this isn't a closed curve, so we can't use Green's theorem. However, it is oriented counterclockwise and it's just a little bit off being closed. Let's call the curve in the problem C_1 and let C_2 denote the straight line between $(1, 0)$ and $(1, 1)$. If we let C be the closed curve, then we have

$$\oint_C \vec{F} \cdot dr = \int_{C_1} \vec{F} \cdot dr + \int_{C_2} \vec{F} \cdot dr.$$

But now the lefthand integral can be computed using Stokes' theorem:

$$\oint_C \vec{F} \cdot dr = \iint_{[0,1]^2} 2y - (-2y) \, dA = \int_0^1 \int_0^1 4y \, dy \, dx = 2y^2 \Big|_0^1 = 2$$

Hence we can solve the integral we want a little more easily: we parametrise C_2 by $\gamma(t) = \langle 1, t\rangle$ for $t \in [0,1]$, and so

$$\begin{aligned}\int_{C_1} \vec{F}\cdot dr &= 2 - \int_{C_2} \vec{F}\cdot dr = 2 - \int_{-1}^{1} \vec{F}(\gamma(t))\cdot \gamma'(t)\,dt\\ &= 2 - \int_0^1 \langle 1-t^2, 2t\rangle \cdot \langle 0,1\rangle\,dt = 2 - \int_0^1 2t\,dt\\ &= 2-1 = 1.\end{aligned}$$

This is much easier than the alternative: breaking the curve C_1 into three line segments and doing three different line integrals.

Next up: Stokes' Theorem. The previous theorem told us how to compute a line integral around a closed curve as a double integral. Stokes' theorem will tell us how to compute a line integral as a surface integral (and sometimes vice versa).

Theorem 4.26 (Stokes' Theorem)**.** Let S be an oriented surface in $\mathbb{R}^3$ with boundary oriented so that the surface is always on your left (assuming outward pointing normal vectors). Assume that $\vec{F}\colon \mathbb{R}^3 \to \mathbb{R}^3$ is a C^1 vector field. Then

$$\oint_{\partial S} \vec{F}\cdot dr = \iint_S \operatorname{curl}\vec{F}\cdot dS$$

In particular, if $\partial S = \varnothing$, then the integral is zero.

It doesn't look like this is a particularly helpful theorem, in that surface integrals are usually pretty nasty. But again, a vector field $\vec{F}$ might be nasty but have a curl which is not – it's hard to tell without computing it though. Supposing that we start with the righthand side, something of the form $\iint_S \vec{G}\cdot dS$, how can we tell if $\vec{G} = \operatorname{curl}\vec{F}$ for some $\vec{F}$?

Proposition 4.27. Suppose that $\vec{G}$ is a vector field in a simply-connected domain. Then $\vec{G} = \operatorname{curl}\vec{F}$ if and only if $\operatorname{div}\vec{G} = 0$.

At least the forwards direction of this proposition is easy, and the backwards direction is done by actually constructing an $\vec{F}$ which has curl $\vec{G}$. I will not be doing this. Thus I haven't actually told you how to find that $\vec{F}$, but we usually don't need to in practice.

Problem 4.28. Let S be the unit sphere in $\mathbb{R}^3$, and let

$$\vec{G}(x,y,z) = \langle 2xyz, 4 - y^2 z, x^3 + y^2\rangle.$$

Compute the flux of $\vec{G}$ through S.

Solution. This looks nigh-impossible. This vector field $\vec{G}$ is defined on all of $\mathbb{R}^3$, which is simply-connected. Moreover, we see that

$$\operatorname{div} \vec{G} = 2yz - 2yz + 0 = 0$$

therefore it's the curl of some vector field $\vec{F}$. Thus:

$$\iint_S \vec{G} \cdot dS = \oint_{\partial S} \vec{F} \cdot dr.$$

But hey, $\partial S = \varnothing$, so we don't even need to know $\vec{F}$ to conclude that the righthand integral is zero.

Another use for Stokes' theorem is the observation that many different surfaces S have the same boundary ∂S. As an illustrating example:

Problem 4.29. Consider the vector field

$$\vec{G} = \langle 2ye^x + z, \log(x+z) - y^2 e^x, x + y + 1\rangle.$$

Compute the flux of $\vec{G}$ through the upper hemisphere of the unit circle S, with counterclockwise oriented boundary and upward pointing normal vector.

Solution. Again, the brute force method would take ages. First, note that $\vec{G}$ is defined as long as $x + z > 0$, which is a simply-connected domain in $\mathbb{R}^3$. We also notice

$$\operatorname{div} \vec{G} = 2yze^x - 2ye^x + 0 = 0$$

So by the recognition theorem, $\vec{G}$ is the curl of something. But hang on, our surface now has a boundary! It's the unit circle in the xy-plane, and doing the line integral of something unknown over that circle seems really bad.

But let's consider the unit disk D, which has the same boundary as S but has a much more straightforward normal vector. Using Stokes' Theorem twice,

$$\iint_S \vec{G} \cdot dS = \oint_{\partial S} ?? \cdot dr = \iint_D \vec{G} \cdot dS.$$

Let's now try to find this right-most integral. The normal vector to D is given everywhere by $\langle 0, 0, 1\rangle$, so we just need to compute the double integral

$$\iint_D \vec{G} \cdot \langle 0,0,1\rangle \, dA = \iint_D x + y + 1 \, dA$$

where we think of D in two dimensions as parametrising D in $\mathbb{R}^3$ via $f(x, y) = (x, y, 0)$. But now this integral is easy: the region D is symmetric in both x and y, so

$$\iint_D x + y \, dA = 0 \implies \iint_D x + y + 1 \, dA = \operatorname{area}(D) = \pi.$$

The last theorem to discuss is the divergence theorem, which will tell us how to compute (certain) triple integrals in terms of surface integrals, and more helpfully vice versa.

Theorem 4.30 (Divergence Theorem)**.** Let S be a closed surface, i.e. one that has no boundary, enclosing a region $W \subset \mathbb{R}^3$. Let S be oriented by outward pointing normal vectors, and suppose that $\vec{F}$ is a C^1 vector field defined on open domain in $\mathbb{R}^3$ that contains W. Then

$$\iint_S \vec{F} \cdot dS = \iiint_W \operatorname{div} \vec{F}\, dV$$

This is related to an observation we made earlier: if $\vec{F}$ is a vector field with $\operatorname{div} \vec{F} = 0$ and if $\partial S = \varnothing$, then the surface integral on the lefthand side vanishes.

Remark 4.31. We have discussed the following operations in $\mathbb{R}^3$:

$$f\colon \mathbb{R}^3 \to \mathbb{R} \xmapsto{\nabla} \vec{F}\colon \mathbb{R}^3 \to \mathbb{R}^3 \xmapsto{\text{curl}} \vec{G}\colon \mathbb{R}^3 \to \mathbb{R}^3 \xmapsto{\text{div}} g\colon \mathbb{R}^3 \to \mathbb{R}$$

The composition of any two of these operations yields the zero function or vector field. Also, as long as we're defining everything over a simply-connected domain, if we know that $\operatorname{curl} \vec{F} = 0$, then $\vec{F} = \nabla f$ and if $\operatorname{div} \vec{G} = 0$, then $\vec{G} = \operatorname{curl} \vec{F}$. That is, if something goes to zero, then it comes as a result of the previous operation in the chain.

There's a nice way to discuss this from the point of differential topology, but that's a bit beyond the scope of the Math GRE. Indeed, I didn't learn any of that until graduate school, at which point I understood most of this well for the first time.

Let's see it in action.

Problem 4.32. Let $\vec{F}(x, y, z) = \langle y, yz, z^2 \rangle$, and let S be the hollow cylinder of radius 2 and height 5 with its base on the xy-plane (with outward pointing normal vector). Compute the flux of $\vec{F}$ through S.

Solution. To actually do this computation, we would need to decompose the cylinder into its top, bottom, and body. That gives us three different flavors of normal vector, which we can use to compute the surface integrals.

But let's not. The divergence of this vector field is beautiful:

$$\operatorname{div} \vec{F} = 0 + z + 2z = 3z.$$

So we need to compute the integral of $3z$ over the solid cylinder. Luckily, since we have access to cylindrical coordinates, it's very easy to rephrase the triple integral we need to perform:

$$\iiint_{\text{cylinder}} 3z\, dV = \int_0^{2\pi} \int_0^2 \int_0^5 3z \cdot r\, dz\, dr\, d\theta = 150\pi$$

Actually doing this triple integral is left as an exercise.

Thus: never ever compute the flux through a closed surface. You have access to the divergence theorem, and it'll almost certainly be a simpler calculation.

Remark 4.33. The divergence theorem has an application to physics which most people learn in their introductory E&M course. Suppose that we have some collection of charged particles and a spherical shell enclosing them. How do we compute the flux of the electrical field through the shell? We just add up the charges of the particles on the inside of that shell. This is known as Gauss' Law. The divergence of the electrical (vector) field is essentially the charge on the particles which produce it.

5. Day 5: Linear algebra

Topics covered: vector spaces, inner products and norms, Cauchy-Schwartz and triangle inequalities, bases and dimension, orthonormal bases, linear transformations, rank-nullity, matrix representations, eigenvalues and eigenvectors, diagonalisation and the spectral theorem.

5.1. Vector spaces.

Right, systems of linear equations. We like solving them, don't we? Let's just cut to the chase. For the purposes of the GRE, we will always work over the fields $\mathbb{R}$ or $\mathbb{C}$.

Definition 5.1. A *vector space* over a field $F = \mathbb{R}, \mathbb{C}$ is a set V equipped with two operations: a commutative and associative addition operation $+\colon V \times V \to V$ and a scalar multiplication operation $\cdot\colon F \times V \to V$. These are required to satisfy a few axioms: we require an additive identity element $\vec{0} \in V$, that every vector $\vec{v} \in V$ have an additive inverse, that scalar multiplication behaves well with respect to multiplication in F, and that scalar multiplication and addition satisfy a distributive property. A summary of most of these axioms is:

$$(ab + c) \cdot (\vec{v} + \vec{w}) = a \cdot (b \cdot \vec{v}) + c \cdot \vec{v} + a \cdot (b \cdot \vec{w}) + c \cdot \vec{w}, \quad a, b, c \in F,\ \vec{v}, \vec{w} \in V$$

For more precise axioms, consult your local textbook.

In the case that a vector space V admits an inner product $\langle -, - \rangle\colon V \times V \to F$, we define the norm by $\|\vec{v}\|^2 = \langle \vec{v}, \vec{v} \rangle$ (where we make sure that $\langle \vec{v}, \vec{v} \rangle \in \mathbb{R}$ if $F = \mathbb{C}$). We're only ever going to be working with finite dimensional inner product spaces on the GRE, so no need to get too complicated.

Our main example is $V = \mathbb{R}^n$, which we think of as a list of n real numbers $\vec{v} = (v_1, \ldots, v_n)$ (we won't be using $\langle\ \rangle$ notation for vectors and will instead reserve that for inner products). Addition is accomplished coordinate-wise and scalar multiplication happens at each coordinate. The axioms of a vector space are satisfied by the field axioms for $\mathbb{R}$. We define our inner product (as we did the dot product before) by

$$\langle (v_1, \ldots, v_n), (w_1, \ldots, w_n) \rangle = \sum_{i=1}^{n} v_i w_i \in \mathbb{R}$$

We want to open up inner products a bit. For $V = \mathbb{C}^n$, we want to define the inner product $\langle \vec{v}, \vec{w} \rangle_{\mathbb{C}}$ to still be a sum of products of coordinates, but with a twist:

$$\langle (v_1, \ldots, v_n), (w_1, \ldots, w_n) \rangle_{\mathbb{C}} = \sum_{i=1}^{n} v_i \overline{w_i}$$

This makes the inner product not bilinear (that is, linear in each variable), but *sesquilinear*, which is fancy way of saying 'one-and-a-half linear'.

$$\alpha \cdot \langle \vec{v}, \vec{w} \rangle_{\mathbb{C}} = \langle \alpha \cdot \vec{v}, \vec{w} \rangle_{\mathbb{C}} = \langle \vec{v}, \overline{\alpha} \cdot \vec{w} \rangle_{\mathbb{C}}, \quad \alpha \in \mathbb{C}$$

The reason for this is to make sure that $\langle \vec{v}, \vec{v} \rangle_{\mathbb{C}} \in \mathbb{R}$ for any $\vec{v} \in \mathbb{C}^n$.

$$\langle \vec{v}, \vec{v} \rangle_{\mathbb{C}} = \sum_{i=1}^{n} |v_i|^2$$

Inner products and norms satisfy what's called the Cauchy-Schwartz inequality: for any $\vec{v}, \vec{w} \in V$,

$$|\langle \vec{v}, \vec{w} \rangle| \leq \|\vec{v}\| \cdot \|\vec{w}\|$$

A corollary of Cauchy-Schwratz is the triangle inequality:

$$\|\vec{v} + \vec{w}\| \leq \|\vec{v}\| + \|\vec{w}\|$$

which we all know from geometry.

Problem 5.2. When are these inequalities equalities? Note: they have different conditions.

Before tackling that problem, it might be useful to remember how the inequalities are proven. We define the projection of $\vec{w}$ onto $\vec{v}$ by

$$\operatorname{proj}_{\vec{w}} \vec{v} = \frac{\langle \vec{v}, \vec{w} \rangle}{\langle \vec{w}, \vec{w} \rangle} \cdot \vec{w}.$$

That's a good formula to remember, and does things like proves the $\cos(\theta)$ formula for inner product (in $\mathbb{R}^n$ – over $\mathbb{C}$ or in a general inner product space θ doesn't make sense):

$$\langle \vec{v}, \vec{w} \rangle = \|v\| \cdot \|w\| \cdot \cos(\theta), \quad \theta = \text{the angle between } \vec{v}, \vec{w}$$

How does this projection help prove Cauchy-Schwartz? Suppose that we are in the special case that $\vec{w} = \vec{0}$. Then the inequality holds trivially, as each side is just 0. If $\vec{w} \neq \vec{0}$, define $\vec{z} = \vec{v} - \operatorname{proj}_{\vec{w}} \vec{v}$. This is the orthogonal complement of $\vec{v}$ with respect to $\vec{w}$, as one readily checks

$$\begin{aligned}
\langle \vec{z}, \vec{w} \rangle &= \langle \vec{v} - \operatorname{proj}_{\vec{w}} \vec{v}, \vec{w} \rangle \\
&= \left\langle \vec{v} - \frac{\langle \vec{v}, \vec{w} \rangle}{\langle \vec{w}, \vec{w} \rangle} \cdot \vec{w}, \vec{w} \right\rangle \\
&= \langle \vec{v}, \vec{w} \rangle - \frac{\langle \vec{v}, \vec{w} \rangle}{\langle \vec{w}, \vec{w} \rangle} \cdot \langle \vec{w}, \vec{w} \rangle = 0
\end{aligned}$$

So by reversing the formula for $\vec{z}$, we know $\vec{v} = \vec{z} + \operatorname{proj}_{\vec{w}} \vec{v}$, then we compute

$$\begin{aligned}\|\vec{v}\|^2 = \langle \vec{v}, \vec{v}\rangle &= \langle \vec{z} + \operatorname{proj}_{\vec{w}} \vec{v}, \vec{z} + \operatorname{proj}_{\vec{w}} \vec{v}\rangle \\ &= \|\vec{z}\|^2 + \|\operatorname{proj}_{\vec{w}} \vec{v}\|^2 + \langle \vec{z}, \operatorname{proj}_{\vec{w}} \vec{v}\rangle + \langle \operatorname{proj}_{\vec{w}} \vec{v}, \vec{z}\rangle\end{aligned}$$

These inner products are zero by the above calculation, as $\operatorname{proj}_{\vec{w}} \vec{v}$ is parallel to $\vec{w}$. If we now work out what $\|\operatorname{proj}_{\vec{w}} \vec{v}\|^2$ is, we obtain

$$\begin{aligned}\|\vec{v}\|^2 &= \|\vec{z}\|^2 + \left\| \frac{\langle \vec{v}, \vec{w}\rangle}{\langle \vec{w}, \vec{w}\rangle} \cdot \vec{w} \right\|^2 \\ &= \|\vec{z}\|^2 + \frac{|\langle \vec{v}, \vec{w}\rangle|^2}{(\|\vec{w}\|^2)^2}\|\vec{w}\|^2 \\ &= \|\vec{z}\|^2 + \frac{|\langle \vec{v}, \vec{w}\rangle|^2}{\|\vec{w}\|^2} \geq \frac{|\langle \vec{v}, \vec{w}\rangle|^2}{\|\vec{w}\|^2}\end{aligned}$$

Multiplying through by the (positive) number $\|\vec{w}\|^2$ and taking square roots completes the proof.

Problem 5.3. Deduce the triangle inequality from the Cauchy-Schwartz inequality.

One fun formula that you might recall is the polarization identity. In any vector space over $\mathbb{R}$,

$$\langle \vec{v}, \vec{w}\rangle = \frac{1}{4}\left(\|\vec{v} + \vec{w}\|^2 - \|\vec{v} - \vec{w}\|^2\right)$$

which is a nice exercise.

Problem 5.4. Do that exercise.

Okay, now we need the notion of subspaces.

Definition 5.5. A subset $W \subset V$ is called a subspace if:

(1) $\vec{0} \in W$
(2) For any $\vec{w}_1, \vec{w}_2 \in W$ and $c \in F$, $\vec{w}_1 + c \cdot \vec{w}_2 \in W$

That is, W is a vector space in its own right that sits inside of V.

5.2. **Bases.** For a set $S = \{\vec{v}_1, \dots, \vec{v}_n\}$, we define the *span* of S to be all linear combinations $\sum_{i=1}^n a_i\vec{v}_i$ for any $a_i \in F$, and sometimes we write $\langle S\rangle$ for this. We say that S is *linearly independent* if whenever we have the sum $\displaystyle\sum_{i=1}^n a_i\vec{v}_i = 0$, then all coefficients $a_i = 0$. This also implies that if $\vec{w}$ is in the span of S, then there is a unique way in which to write $\vec{w}$ as a linear combination of the $\vec{v}_i$.

Problem 5.6. Prove the equivalence of the two statements above.

A maximal linearly independent set in V is called a *basis*. All bases have the same number of elements, and all (finite dimensional) vector spaces have a basis. Call that number the *dimension* of V. Note that 'infinite dimensional' vector spaces don't have a basis unless you assume the axiom of choice! It also makes sense to talk about the basis of a subspace $W \subset V$.

Let us recall a few more examples of vector spaces that are not F^n explicitly. Consider the set $P_n(F)$ of polynomials with coefficients in F of degree up to n. This set is closed under scalar multiplication by F and addition, and the zero polynomial is a zero element. The additive inverses in the field F gives us additive inverses. It is an easy exercise that $\{1, \ldots, x^n\}$ forms a basis for $P_n(F)$, which means $\dim P_n(F) = n + 1$. One can put an inner product on this space of the form $\langle f(x), g(x) \rangle = \displaystyle\int_a^b f(x)\overline{g(x)}\, dx$ (just ignore the conjugation if $F = \mathbb{R}$).

Another example is the set of matrices $M_{m\times n}(F)$, which again has a sensible addition and scalar multiplication operation. The elementary matrices $\{e_{ij} : 1 \le i \le m, 1 \le j \le n\}$ form a basis, meaning that $\dim M_{m\times n}(F) = mn$. There may be some more interesting inner products on the space of matrices, but I am not aware of them.

Remark 5.7. Of course, we will prove shortly that every vector space of the same dimension is isomorphic – though that's not a guarantee that every isomorphism preserves the inner product! We're going to keep writing V, W for arbitrary vector spaces, but on the GRE we might as well have $V = \mathbb{R}^n$ and $W = \mathbb{R}^m$ as long as $F = \mathbb{R}$, where $n = \dim V$ and $m = \dim W$.

What's the best kind of basis? An orthonormal one! Recall that $\vec{v}$ and $\vec{w}$ are *orthogonal* if $\langle \vec{v}, \vec{w} \rangle = 0$.

Definition 5.8. A set $S = \{\vec{v}_1, \ldots, \vec{v}_n\} \subset V$ is an *orthonormal basis* if

(1) $\|\vec{v}_i\| = 1$ for all $i = 1, \ldots, n$
(2) For any $i \neq j$, $\langle \vec{v}_i, \vec{v}_j \rangle = 0$

This can be phrased more succinctly by saying that $\langle \vec{v}_i, \vec{v}_j \rangle = \delta_{ij}$, where $\delta_{ij} = 1$ if $i = j$ and is zero otherwise.

The usual basis for $V = \mathbb{R}^n$ given by $\vec{e}_i$ is orthonormal. Think how much more difficult the world would be if the coordinate axes weren't perpendicular to each other!

Problem 5.9. Suppose that $S = \{\vec{v}_i\}_{i=1}^n$ is an orthonormal basis. Then we know that any $\vec{w} \in V$ has a unique expression as

$$a_1\vec{v}_1 + \cdots + a_n\vec{v}_n = \vec{w}.$$

Prove that we can compute these coefficients: $a_i = \langle \vec{v}_i, \vec{w} \rangle$.

That is very useful! But what if our basis isn't orthonormal? Luckily there's a process, called the *Gram-Schmidt process*, to transform it into an orthonormal one. The process is inductive, and goes as follows:

- Begin with the first element $\vec{v}_1$ of your basis. This may not be a unit vector, so let $\vec{u}_1 := \dfrac{\vec{v}_1}{\|\vec{v}_1\|}$. The vector $\vec{u}_1$ is orthogonal to every other element of our new basis (because we haven't added any yet).
- Now, take the element $\vec{v}_2$. This is probably not orthogonal to $\vec{u}_1$, so we force it to be so: define
$$\vec{w}_2 = \vec{v}_2 - \langle \vec{u}_1, \vec{v}_2 \rangle \cdot \vec{u}_1$$
which we can see is now orthogonal to $\vec{u}_1$. But this is probably not a unit vector, so define $\vec{u}_2 = \frac{\vec{w}_2}{\|\vec{w}_2\|}$.
- We see how to proceed from here: define
$$\vec{w}_j = \vec{v}_j - \sum_{i=1}^{j-1} \langle \vec{u}_i, \vec{v}_j \rangle \cdot \vec{u}_i, \quad \vec{u}_j = \frac{\vec{w}_j}{\|\vec{w}_j\|}$$
and eventually we'll be done!

It's important to note that at every stage we are not changing the span of our vectors. The vector $\vec{w}_2$, for instance, is a linear combination of $\vec{u}_1$ and $\vec{v}_2$, and $\vec{u}_1$ was just a multiple of $\vec{v}_1$ so was in its span. Thus the span of $\vec{w}_1, \vec{w}_2$ is the same as $\vec{v}_1, \vec{v}_2$.

Problem 5.10. Let $\{1, x, x^2\}$ be a basis for $P_2(\mathbb{R})$, the vector space of degree at most 2 polynomials with coefficients in $\mathbb{R}$, endowed with the inner product $\langle f(x), g(x) \rangle = \displaystyle\int_0^1 f(x)g(x)\,dx$. Convert this to an orthonormal basis.

5.3. **Linear transformations.** What are the functions we care about?

Definition 5.11. A linear transformation $T \colon V \to W$ is a function satisfying the axiom that $T(c \cdot \vec{v}_1 + \vec{v}_2) = c \cdot T(\vec{v}_1) + T(\vec{v}_2)$ for all $c \in \mathbb{R}$ and for all $\vec{v}_1, \vec{v}_2 \in V$.

The definition implies that $T(\vec{0}) = \vec{0}$, which is a nice feature. Many of the linear operations we already know can be recast in terms of linear transformations. If we let $P_n(\mathbb{R})$ be the space of degree at most n polynomials, then the derivative $d/dx \colon P_n(\mathbb{R}) \to P_{n-1}(\mathbb{R})$ is actually a linear transformation. Similarly, $\int dx \colon P_n(\mathbb{R}) \to P_{n+1}(\mathbb{R})$ is a linear transformation so long as we pick uniformly $C = 0$ for the '$+C$' part of an indefinite integral. (Small problem: why can't we pick any constant C that we want?)

We have already discussed linear transformations $\mathbb{R} \to \mathbb{R}$ and $\mathbb{R}^2 \to \mathbb{R}$ – these are lines and planes (respectively) that pass through the origin. Note that any line that does not have the point $(0, 0)$ on it cannot be a linear transformation. Recalling that the equation of a plane was also determined by $\vec{n}\cdot\vec{v}=d$, we can also recall that for any vector space V (over $F=\mathbb{R},\mathbb{C}$), there is a linear transformation $\langle -,\vec{v}\rangle\colon V\to F$ where we fix $\vec{v}$. Over either $\mathbb{R}$ or $\mathbb{C}$, the first argument of the inner product is linear.

Finally, given a vector space V with basis $\{\vec{v}_1,\dots,\vec{v}_n\}$, we can define a linear transformation $T\colon V\to W$ by setting $T(\vec{v}_i)$ to be whatever we like. Then 'extending by linearity', for any other $\vec{v}\in V$, we have

$$\vec{v}=\sum_{i=1}^n a_i\vec{v}_i \implies T(\vec{v})=\sum_{i=1}^n a_iT(\vec{v}_i)$$

which is not only a requirement for T to be linear, but also gives us the definition of T on all of V.

To any linear transformation, we have a couple of associated subspaces.

Problem 5.12. For a linear transformation $T\colon V\to W$, define $\operatorname{im}(T)\subset W$ to be the set $\{\vec{w}\in W:\vec{w}=T(\vec{v})\text{ for some }\vec{v}\in V\}$. Define $\ker(T)\subset V$ to be the set $\{\vec{v}\in V:T(\vec{v})=\vec{0}\}$. Prove that both of these subsets are actually subspaces.

Problem 5.13. Let $V=\mathbb{R}^n$ and let $T=\langle -,\vec{v}\rangle\colon V\to\mathbb{R}$ be the linear transformation mentioned above. Describe $\ker(T)$.

Definition 5.14. The *rank* of $T\colon V\to W$ is the dimension of its image $\operatorname{im}(T)\subset W$ as a subspace of W. The *nullity* the dimension of its kernel, $\ker(T)$.

Theorem 5.15. Rank + nullity = $\dim V$.

5.4. **Matrices.** Linear transformations are well described by matrices in the case that we identify $V=\mathbb{R}^n$ and $W=\mathbb{R}^m$. A transformation T yields a matrix $A\in M_{m\times n}(\mathbb{R})$ where the columns of A are $T(\vec{e}_i)$ for the basis vectors $\vec{e}_i$ of $\mathbb{R}^n$. But normally vector spaces don't come with automatic bases. For $T\colon V\to W$, where $\dim V=n$ and $\dim W=m$, we still get a matrix A of the same dimensions, but we have to set a basis $\beta=\{\vec{b}_i\}$ for V and $\gamma=\{\vec{c}_i\}$ for W, thus will denote it $[T]_\beta^\gamma$.

For each $\vec{b}_i\in\beta$, we can write

$$T(\vec{b}_i)=a_{1i}\vec{c}_1+\cdots+a_{mi}\vec{c}_m$$

which gives us the matrix $[T]_\beta^\gamma=(a_{ij})$ with columns these vectors $(a_{1i},\dots,a_{mi})$.

A particular case of interest is $V=W=\mathbb{R}^n$. A goal of much of linear algebra is to understand linear transformations $T\colon\mathbb{R}^n\to\mathbb{R}^n$ in the simplest terms possible, which (in one interpretation) means a basis β such that $[T]_\beta$ is as simple as possible (here we think of both the domain and codomain with basis β). In order to write the matrix $A=[T]_{\text{std}}$ in terms of a new basis, we think of converting from β to standard, doing the transformation A that we know, then back to β. The 'back to standard' matrix is P such that its columns are the $\vec{b}_i$. Therefore

$$\begin{tikzcd} \mathbb{R}^n \arrow[r, "A"] & \mathbb{R}^n \\ \mathbb{R}^n \arrow[u, "P"] \arrow[r, "{[T]_\beta^\beta}"'] & \mathbb{R}^n \arrow[u, "P"'] \end{tikzcd}$$

$$A = P[T]_\beta P^{-1} \implies [T]_\beta = P^{-1}AP$$

where the matrix P is easy to compute but P^{-1} is usually a little more unpleasant to compute. In two dimensions, there's a nice formula:

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

Astute readers will recall $ad - bc = \det A$. In general, one uses Gaussian elimination to compute the inverse of a matrix. In brief, we write the matrix next to the identify matrix as such:

$$\begin{pmatrix} a & b & | & 1 & 0 \\ c & d & | & 0 & 1 \end{pmatrix}$$

We then perform elementary row operations on the lefthand matrix to reduce it to the identity. Note that if we cannot do this, then A is not invertible, so the whole exercise is fruitless. Performing the same row operations on the righthand matrix gives us the inverse.

Problem 5.16. Verify the formula for the inverse of a 2×2 matrix using Gaussian elimination.

Problem 5.17. Using Gaussian elimination, prove that

$$\begin{bmatrix} 2 & 0 & 2 \\ 0 & 1 & 1 \\ -1 & -2 & 0 \end{bmatrix}^{-1} = \begin{bmatrix} 1/3 & -2/3 & -1/3 \\ -1/6 & 1/3 & -1/3 \\ 1/6 & 2/3 & 1/3 \end{bmatrix}$$

Solution. It would be way too hard to type out the solution here, but I would perform this process in a live lecture.

This brings up more things to say about determinants. For determinants of higher dimensional matrices, we use expansion by minors. We have covered 3×3 matrices when discussing the cross product, but for 4×4 we might want a bit of a reminder.

$$\det \begin{bmatrix} a & b & c & d \\ e & f & g & h \\ i & j & k & l \\ m & n & o & p \end{bmatrix} = a \cdot \det \begin{bmatrix} f & g & h \\ j & k & l \\ n & o & p \end{bmatrix} - b \cdot \det \begin{bmatrix} e & g & h \\ i & k & l \\ m & o & p \end{bmatrix}$$
$$+ c \cdot \det \begin{bmatrix} e & f & h \\ i & j & l \\ m & n & p \end{bmatrix} - d \cdot \det \begin{bmatrix} e & f & g \\ i & j & k \\ m & n & o \end{bmatrix}$$

In general, pick your favourite row or column of the matrix. For each element there, create an $(n-1) \times (n-1)$ submatrix by crossing out its row and column. Multiply your element by the determinant of that submatrix and then $(-1)^k$ where k is the number of steps you have taken from the top-left corner. For example, if we picked the second row, then g would come with a -1 since it is 3 steps from the top-left corner.

There are some other tricks that help us compute determinants. If you swap two rows or columns of a matrix, it multiplies the determinant by -1. If you take the transpose of the matrix (for whatever reason), it does not change the determinant. If you add a constant multiple of a row or column to another row or column, it does not change the determinant.

Problem 5.18. Prove that if the rows (or columns) of a matrix are linearly dependent, the determinant of the matrix is zero.

If the determinant of the matrix is not zero, then the matrix is invertible. That is, if $\det A \neq 0$ for an $n \times n$ matrix A, then there exists an $n \times n$ matrix B such that $AB = BA = I_n$. This is a theorem rather than a definition, but we won't need to worry about the proof for the GRE.

Let us consider all the pertinent equivalent formulations of invertibility:

- $\det A \neq 0$
- row rank of A is n
- column rank of A is n
- the linear transformation that A defines is bijective
- reduced row echelon form is the identity

Problem 5.19. Prove that if a matrix A is nilpotent, then $I - A$ is invertible. That is, if $A^k = 0$ for some $k \in \mathbb{N}$, then $I - A$ is invertible.

Solution. On the one hand, suppose $A^k = 0$. Then consider the old trick of factorisation:

$$I - A^k = (I - A)(I + A + A^2 + \cdots + A^{k-1})$$

Since $A^k = 0$, this means that $I - A$ has an inverse, which makes it invertible.

We can also use this to prove that $I + A$ is invertible, because $(-A)^k = 0$ so $-A$ is nilpotent as well.

This solves a confusing question on Practice Exam 1 that always comes up.

5.5. **Eigenvalues and eigenvectors.** The point of the above discussion is to pave the way right into (more or less) the main point of linear algebra.

Definition 5.20. Let $T\colon V \to V$ be a linear transformation of a vector space V over F. We say that a nonzero vector $\vec{v} \in V$ is an *eigenvector* for T if there exists $\lambda \in F$ such that $T(\vec{v}) = \lambda \cdot \vec{v}$. Such a λ is called an *eigenvalue* for T.

An easy first step: suppose that T has a kernel. Then for any (nonzero) $\vec{v} \in \ker(T)$, we have $T(\vec{v}) = \vec{0} = 0 \cdot \vec{v}$, so $\vec{v}$ is an eigenvector with associated eigenvalue 0. Also note that if $\vec{v}$ is an eigenvector for an eigenvalue λ, then for any $c \in F$,

$$T(c \cdot \vec{v}) = c \cdot T(\vec{v}) = c \cdot \lambda \cdot \vec{v} = \lambda \cdot (c \cdot \vec{v})$$

So any multiple of $\vec{v}$ is also an eigenvector for the same eigenvalue. A similar argument shows that the set of eigenvectors for the eigenvalue λ is actually a subspace, which we name $E_\lambda \subset V$.

Computing eigenvalues and eigenvectors for small dimensions is relatively simple.

Problem 5.21. Compute the eigenvalues for the linear transformation $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ represented in the standard basis by

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Solution. We can solve this symbolically:

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} y \\ x \end{bmatrix} = \begin{bmatrix} \lambda x \\ \lambda y \end{bmatrix}$$

which gives us the system of equations $y = \lambda x$ and $x = \lambda y$. Similar to the method of Lagrange multipliers, we can solve for λ on both sides and we conclude that

$$\frac{y}{x} = \frac{x}{y} \implies x^2 = y^2$$

This gives us only two choices: $x = y$ or $x = -y$. These correspond (respectively) to $\lambda = 1$ and $\lambda = -1$. We can therefore conclude that E_1 is the span of $(1, 1)$ and E_{-1} is the span of $(1, -1)$.

Notice that the span of the eigenspaces above is all of $\mathbb{R}^2$. That is, $\mathbb{R}^2$ admits a basis that is comprised of eigenvectors for T. We call this an *eigenbasis* and this is the best we can hope for: let $\beta = \{(1,1),(1,-1)\}$. Then

$$[T]_\beta = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

We can verify this also using the change of basis matrix $P = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$:

$$[T]_\beta = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}^{-1} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

Problem 5.22. Complete this verification.

When a matrix admits an eigenbasis, we call it *diagonalisable* because there is a change of basis to a diagonal matrix. Diagonal matrices are ideal; taking products of them is really easy. Moreover, the trace and the determinant of a diagonal matrix are super easy to compute. But this gives us an important proposition:

Proposition 5.23. Let A be a diagonalisable matrix. Then the trace of A is the sum of its eigenvalues and the determinant of A is the product of its eigenvalues.

Problem 5.24. Verify that the linear transformation $\mathbb{R}^2 \to \mathbb{R}^2$ defined by the matrix

$$B = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

does *not* have any eigenvectors.

Solution. We will use another method to make this conclusion. The *characteristic polynomial* of a linear transformation is the determinant $p_T(\lambda) = \det(A - \lambda \cdot I)$, which we think of as a polynomial in terms of λ. The roots of this polynomial are the eigenvalues of the matrix.

For our matrix above, we compute

$$\det \begin{bmatrix} -\lambda & 1 \\ -1 & -\lambda \end{bmatrix} = \lambda^2 + 1$$

This polynomial has no roots, so the linear transformation cannot have any eigenvectors (as it has no eigenvalues).

Remark 5.25. Nonetheless, we know that over $\mathbb{C}$ the eigenvalues of the above matrix are $\pm i$, which leads to the conclusion that the trace of B should be zero and the determinant should be 1. These are both true facts. In general, one can extend from $\mathbb{R}$ to $\mathbb{C}$ to see what the eigenvalues 'should' be.

How often does a linear transformation $T\colon V \to V$ have an eigenbasis?

Proposition 5.26. Let $A = [T]$ be the matrix corresponding to a linear transformation $T\colon \mathbb{R}^n \to \mathbb{R}^n$. If the characteristic polynomial $p_T(\lambda)$ has n distinct roots, then T is diagonalisable.

The short proof: we know that there are n distinct 1-dimensional eigenspaces for T. This must add up to the entire space $\mathbb{R}^n$, giving us the eigenbasis. Of course, this is not an if and only if statement. The following theorem is a restricted version, but good enough for the GRE.

Theorem 5.27 (Spectral Theorem)**.** If A is a real matrix and $A = A^T$, the transpose of A, then A is diagonalisable. If A is a complex matrix, then we require $A = A^*$, the Hermitian matrix (transpose conjugate).

The full generality only requires that $AA^T = A^TA$ (or the similar statement with the Hermitian matrix in the complex case), but we will rarely be able to verify that condition given the time constraints of the GRE.

A random note: A and A^T have the same eigenvalues. This is because a matrix has the same determinant as its transpose, which implies that A and A^T have the same characteristic polynomial.

One last fact: suppose that A and B are two matrices that admit the same eigenbasis; they don't have to have the same eigenvalues for that basis, but they need the same basis. Then $AB = BA$. This follows from the fact that diagonal matrices commute, and we can turn A and B into diagonal matrices using the same change of basis P:

$$P^{-1}ABP = P^{-1}AP \cdot P^{-1}BP = D_A \cdot D_B = D_B \cdot D_A = P^1BP \cdot P^{-1}A = P^{-1}ABP$$

If we un-conjugate by P at the end, we complete the argument. It turns out that the converse of this statement is true as well: if A and B are diagonalisable matrices that commute, then they admit a joint eigenbasis.

6. Day 6: Differential equations and complex analysis

Topics covered: separation of variables, ordinary differential equations, method of integrating factor, linear ODEs; holomorphic and analytic functions, Cauchy-Riemann functions, Cauchy integral formula.

6.1. Basic differential equations.

So we begin today by recalling the basic types of differential equations that we need to solve for the GRE. There aren't that many, but if you're like me, you've definitely forgotten about them. But before we get there, let's recall the fundamental theorem that lets us do anything:

Theorem 6.1 (Existence and uniqueness of solutions)**.** Suppose that $f(t, y)$ and $\partial_y f(t, y)$ are continuous on a compact subset $K \subset \mathbb{R}^2$. Then for any point $(t_0, y_0) \in K$, the differential equation $y' = f(t, y)$, $y(t_0) = y_0$ has a unique solution in some neighbourhood of (t_0, y_0).

So perhaps it's good to check that $\partial_y f(t, y)$ is continuous before blindly charging into a problem, but probably not.

The first type of differential equation is the basic separation of variables, like so:

Problem 6.2. Suppose that a colony of bacteria grows at a rate directly proportional to its size. Initially, the colony has 100 bacteria, and after a week it has 300 bacteria. Write a formula modelling this situation where the time t is measured in days.

Solution. The situation we have is

$$\frac{dB}{dt} = k \cdot B \implies \frac{dB}{B} = k\,dt$$

Then integrating both sides gives us

$$\log(B) = k \cdot t + C \implies B(t) = C \cdot e^{kt}.$$

We have that $B(0) = 100$, so $C = 100$. We also know that $B(7) = 300$, so

$$e^{7k} = 3 \implies k = \frac{\log(3)}{7}.$$

Putting it all together,

$$B(t) = 100e^{\frac{\log(3)}{7} \cdot t}$$

There are more sophisticated version of this problem, and they are usually in the tune of salty or sugary tanks of water.

Problem 6.3. Suppose that we have a 100L tank with 50L of water and 100g of salt in it. Suppose the tank drains at a rate of 1L/m and is filled at a rate of 2L/m with pure water. Assuming instantaneous mixing, when the tank is full, how much salt is there in the tank?

Solution. Let's set this up. We have

$$\frac{dS}{dt} = \text{in} - \text{out}$$

In this example, there's no salt coming in. For the out, we need to know what the density of salt in the tank is. The amount of salt is S, but the volume changes: we have a net +1L/m, so the volume is $50 + t$. Therefore

$$\frac{dS}{dt} = -\frac{S}{50+t} \implies \frac{dS}{S} = \frac{-dt}{50+t} \implies \log(S) = -\log(50+t) + C$$

After integration and rearranging (specifically after pulling in the -1 into the log), we get

$$S(t) = \frac{C}{t+50}$$

The amount of salt at the beginning is 100, so we have $S(0) = \dfrac{C}{50} = 100$ so $C = 5000$. The tank is full at $t = 50$, so

$$S(t) = \frac{5000}{t+50} \implies S(50) = \frac{5000}{100} = 50\text{g}.$$

Now, we turn to other types of differential equations. Let's first recall what an integrating factor is. Suppose our differential equation is of the form

$$\frac{dy}{dt} + p(t)y = q(t).$$

Then we consider the integrating factor $\mu(t) = e^{\int p(t)\,dt}$. Why does that help? Using this term,

$$\mu(t) \cdot y' + \mu(t)p(t) \cdot y = \frac{d}{dt}(\mu(t) \cdot y) = \mu(t) \cdot q(t).$$

Thus, when we integrate both sides,

$$\mu(t) \cdot y = \int \mu(t)q(t)\,dt$$

Assuming that the righthand side is integrable, we can then solve and divide out by $\mu(t)$.

Problem 6.4. Solve the linear ODE $y' - 2ty = t$.

Solution. The process implies that $\mu(t) = e^{\int -2t\,dt} = e^{-t^2}$, not something we can integrate on its own. Luckily, the whole righthand side is integrable:

$$\int te^{-t^2}\,dt = -\frac{1}{2}e^{-t^2} + C$$

Dividing through now by our integrating factor,

$$y(t) = Ce^{t^2} - \frac{1}{2}$$

Again, if we get a linear ODE of this form, this is pretty much the only way to solve it. Exact ODEs likely won't come up, but there's always that chance. Plus, it's related to multivariable calculus. Suppose that we have a differential equation of the form

$$N(x,y) \cdot y' + M(x,y) = 0 \text{ i.e. } N(x,y)dy + M(x,y)dx = 0$$

where moreover we have $\partial_x N(X,y) = \partial_y M(x,y)$. Then this implies that, at least locally, that this situation is coming from the equality is mixed partials, so we need to find a function $H(x,y)$ with $\nabla H = \langle M, N \rangle$. The general solution to the differential equation is $H(x,y) = C$.

Problem 6.5. Solve $(x^2y + 2y) \cdot y' + (xy^2 + 2x) = 0$.

Solution. This equation is exact (easily verified), so a solution looks something like

$$H(x,y) = \int xy^2 + 2x\,dx = \int x^2y + 2y\,dy$$

As before, we need to integrate but bear in mind that we might have constants that depend on one variable or the other. That is,

$$\int xy^2 + 2x\,dx = \frac{x^2y^2}{2} + x^2 + g_1(y), \quad \int x^2y + 2y\,dy = \frac{x^2y^2}{2} + y^2 + g_2(x)$$

Comparing terms, we need to use $g_2(x) = x^2$ and $g_1(y) = y^2$, so that

$$H(x,y) = \frac{x^2y^2}{2} + x^2 + y^2 = C$$

is our general solution.

6.2. **Higher order differential equations.** Now, suppose we want to solve particular differential higher order differential equations that have little interaction between the variables. For instance, examine

$$y'' - 9y = f(t)$$

The first step is to solve the corresponding homogeneous equation $y'' - 9y = 0$. We can solve this by inspection, know that $y' = ky$ is solved by e^{kt}. Hence the solutions we need are e^{3t} and e^{-3t}. The general solution to the differential equation is therefore

$$y(t) = c_1e^{3t} + c_2e^{-3t}.$$

We will look at what to do about the $f(t)$ a bit later.

Here's generally how you solve a homogeneous differential equation like this. Consider an equation

$$ay'' + by' + cy = 0$$

Then solutions to this equation are given by $e^{\lambda t}$, where λ is a root of the corresponding characteristic polynomial

$$ax^2 + bx + c = 0$$

There are three options here: the polynomial may have two distinct real roots, one double real root, or two (conjugate) complex roots.

The case of two distinct real roots is the one we examined above: the general solution is $y(t) = c_1 e^{\lambda_1 t} + c_2 e^{\lambda_2 t}$. If there is only one real root, we still need a two-dimensional solution to the system of equations, so the general solution looks like $y(t) = c_1 e^{\lambda t} + c_2 t e^{\lambda t}$. The complex roots possibility is a little more delicate, because we need to make sure that come up with a real solution.

To examine this, let $\lambda = a + bi$. Then the general solution becomes

$$y(t) = c_1 e^{(a+bi)t} + c_2 e^{(a-bi)t} = c_1 e^{at} e^{i \cdot bt} + c_2 e^{at} e^{i \cdot (-bt)}$$

Using the identity $e^{i\theta} = \cos(\theta) + i \sin(\theta)$, we change the above:

$$\begin{aligned} y(t) &= c_1 e^{at} e^{i \cdot bt} + c_2 e^{at} e^{i \cdot (-bt)} \\ &= c_1 e^{at}(\cos(bt) + i \sin(bt)) + c_2 e^{at}(\cos(-bt) + i \sin(-bt)) \end{aligned}$$

Using now that $\cos(-bt) = \cos(bt)$ and $\sin(-bt) = -\sin(bt)$,

$$y(t) = (c_1 + c_2)e^{at} \cos(bt) + (c_1 - c_2)i \cdot e^{at} \sin(bt)$$

Now we use the fact that (secretly) $c_1, c_2 \in \mathbb{C}$ now, we need that $c_1 - c_2 \in i\mathbb{R}$ and $c_1 + c_2 \in \mathbb{R}$. Luckily, it is possible to get any number we want using $c_1 = \frac{c-di}{2}$ and $c_2 = \frac{c+di}{2}$ so that $c_1 + c_2 = c$ and $c_1 - c_2 = -di$. Putting this all together, the general solution is

$$y(t) = c_1 e^{at} \cos(bt) + c_2 e^{at} \sin(bt).$$

Problem 6.6. Solve the following initial value problem: $y'' - 4y' + 9y = 0$, with $y(0) = 0$ and $y'(0) = -8$.

Solution. The characteristic equation is $x^2 - 4x + 9 = 0$, so that we have roots

$$\lambda = \frac{4 \pm \sqrt{16 - 4(9)}}{2} = 2 \pm \sqrt{5}i$$

Hence the general solution is

$$y(t) = c_1 e^{2t} \cos(\sqrt{5}t) + c_2 e^{2t} \sin(\sqrt{5}t)$$

Knowing that $y(0) = 0$ means that $c_1 = 0$, as everything else cancels out. Therefore

$$y(t) = ce^{2t} \sin(\sqrt{5}t) \text{ and } y'(t) = 2ce^{2} \sin(\sqrt{5}t) + \sqrt{5}ce^{2t} \cos(\sqrt{5}t)$$

So $y'(0) = \sqrt{5} \cdot c = -8$ thus $c = -\dfrac{8}{\sqrt{5}}$. Not the nicest solution, but

$$y(t) = -\frac{8}{\sqrt{5}} e^{2t} \sin(\sqrt{5}t)$$

6.3. **Nonhomogeneous differential equations.** How do we deal with nonhomogeneous differential equations? First, solve the homogeneous one. Then we have to guess a particular solution.

Problem 6.7. Determine a particular solution to $y'' - 4y' - 12y = 3e^{5t}$.

Solution. We need to solve $x^2 - 4x - 12 = 0$, which isn't too difficult: it factors as $(x+2)(x-6) = 0$ so we get

$$y(t) = c_1 e^{6t} + c_2 e^{-2t} + y_p(t)$$

What should $y_p(t)$ look like? Probably something of the form $y_p(t) = Ae^{5t}$, so we then need to check which A satisfies the differential equation:

$$y_p'(t) = 5Ae^{5t},\ y_p''(t) = 25Ae^{5t} \implies 25Ae^{5t} - 4 \cdot 5Ae^{5t} - 12 \cdot Ae^{5t} = 3e^{5t}$$

Solving this gives $-7Ae^{5t} = 3e^{5t}$ so $A = -\dfrac{3}{7}$. Putting this all together,

$$y(t) = c_1 e^{6t} + c_2 e^{-2t} - \frac{3}{7} e^{5t}$$

Other types of particular solutions require different guesses: sines and cosines demand sines and cosines. What if particular solutions are polynomials?

Problem 6.8. Determine a particular solution to $y'' - 4y' - 12y = t^2 + 3t + 2$.

Solution. The particular solution looks like a polynomial of the same degree, so let $y_p(t) = at^2 + bt + c$. Then

$$y_p'(t) = 2at + b, \quad y_p''(t) = 2a$$

Putting it all together,

$$2a - 4(2at + b) - 12(at^2 + bt + c) = t^2 + 3t + 2$$

We need to separate by degrees:

$$-12at^2 = t^2, \quad (-8a - 12b)t = 3t, \quad 2a - 4b - 12c = 2$$

The easiest way to solve this is left to right:

$$a = -\frac{1}{12}, \quad \left(\frac{8}{12} - 12b\right) = 3 \implies 12b = -\frac{7}{3} \implies b = -\frac{7}{36}$$

Finally, we can solve that $c = -\frac{25}{216}$. We can then put it all together as we did above.

Fortunately, this is as far as things need to go in the realm of differential equations.

6.4. **Complex analysis.** Let's recall a little the nice types of complex-valued functions.

Theorem 6.9. Let $f\colon \Omega \to \mathbb{C}$, where $\Omega \subset \mathbb{C}$ is an open subset of the complex numbers. Then the following are equivalent:

- f is differentiable in an open disc centered at $a \in \Omega$ (holomorphic)
- f has a convergent power series expansion $\sum_{n=0}^{\infty} c_n(z-a)^n$ in an open disc centered at $a \in \Omega$ (analytic)

This incredible theorem implies that differentiable functions are smooth (although we haven't defined what 'differentiable' means per se), which is one of our introductions to the wild world of complex analysis. There are some nice corollaries:

Corollary 6.10. Let $f, g\colon \Omega \to \mathbb{C}$ be two holomorphic functions on an open connected $\Omega \subset \mathbb{C}$. If $f(z) = g(z)$ on an infinite subset $S \subset \Omega$ that contains a limit point of Ω, then $f = g$ on Ω.

Corollary 6.11. A bounded holomorphic function $f\colon \mathbb{C} \to \mathbb{C}$ must be constant.

Holomorphic functions must satisfy the Cauchy-Riemann equations, and the converse is true as well.

Theorem 6.12. Let $f\colon \Omega \to \mathbb{C}$ be a function, and write $f(x+iy) = u(x+iy) + i \cdot v(x+iy)$. Then f is holomorphic if and only if

$$\partial_x u = \partial_y v \text{ and } \partial_y u = -\partial_x v.$$

This theorem is pretty useful, as it means that information about the real part of a holomorphic function can get us the whole function.

Problem 6.13. Suppose that $f(x+iy) = u(x,y) + i \cdot v(x,y)$. If $u(x,y) = x^2 - y^2$ and $v(1,1) = 2$, find $v(4,1)$.

Solution. We can use the fundamental theorem of calculus in this case:

$$v(4,1) - v(1,1) = \int_1^4 \partial_x v(x,1)\, dx$$

The equation is, what's $\partial_x v$? By the Cauchy-Riemann equations, it's $-\partial_y u = 2y$. Thus

$$v(4,1) - v(1,1) = \int_1^4 2 \cdot 1 \, dx = 6$$

This implies that $v(4,1) - 2 = 8$ so $v(4,1) = 8$.

It might be worth remembering the following:

Definition 6.14. A function $u(x,y) \colon \Omega \to \mathbb{R}$ is harmonic if $\partial_x^2 u + \partial_y^2 u = 0$.

Locally, a harmonic function is the real part of a holomorphic function, in the following way: if $u(x,y)$ is harmonic on an open and simply-connected domain Ω, there exists a holomorphic function $f(z)$ such that

$$u(x,y) = \operatorname{Re} f(x+iy)$$

We can actually construct $f(x+iy)$ pretty easily: set $g(z) = \partial_x u - i\partial_y u$. Then this function is holomorphic, as the Cauchy-Riemann equations are satisfied by the harmonic condition and the equality is mixed partials.

Then since $g(z)$ is holomorphic by a black box that relies on simply-connectedness we can integrate it to obtain $f(z)$ such that $f'(z) = g(z)$. Thus, up to a constant, we can verify that the real part of $f(z)$ is $u(x,y)$.

Remark 6.15. The main counterexample for all of these theorems that rely on a simply-connected domain is $\log(z)$ or $\log(|z|)$. For example, $\log(|x+iy|) = \log(x^2 + y^2)/2$ is harmonic on the domain $\mathbb{C} \setminus \{0\}$ but not the real part of any holomorphic function. Of course, we can restrict to simply-connected subsets of $\mathbb{C} \setminus \{0\}$ but we can't get a global statement.

6.5. **Cauchy integral formula.** At our own peril, the last thing we recall is the Cauchy integral formula.

Theorem 6.16 (Cauchy integral formula)**.** Suppose that $f \colon \Omega \to \mathbb{C}$ is a holomorphic function on an open domain, and let $D \subset \Omega$ be a closed disc in Ω. Then

$$f(a) = \frac{1}{2\pi i} \oint_{\partial D} \frac{f(z)}{z-a} \, dz$$

for every $a \in D$.

This yields the residue theorem.

Theorem 6.17 (Residue Theorem)**.** Let $U \subset \mathbb{C}$ be a simply connected open subset and $f \colon U \to \mathbb{C}$ a function holomorphic but for $a \in U$. Let γ be a closed curve in U around a, oriented counterclockwise. Then

$$\oint_\gamma f(z) \, dz = 2\pi i \operatorname{Res}(f,a)$$

where $\mathrm{Res}(f, a)$ is the coefficient of the term $\dfrac{1}{z-a}$ in the Laurent series expansion of $f(z)$ around a.

I'm not sure that this has much of a place on the GRE. Memorise the Cauchy-Riemann equations and call it a day.

7. Day 7: Algebra

Topics covered: groups, subgroups and Lagrange's theorem, abelian groups and their classification; rings, ideals, modular arithmetic; fields and structures that are almost fields.

7.1. Groups.

Definition 7.1. A group is a set G with a binary operation $\cdot\colon G \times G \to G$ that satisfies the following axioms:

- The operation is associative, so $(g \cdot h) \cdot k = g \cdot (h \cdot k)$ for any $g, h, k \in G$
- There exists an element $e \in G$ such that $e \cdot g = g \cdot e = g$ for all $g \in G$
- For every element $g \in G$, there exists an element $g^{-1} \in G$ so that $g \cdot g^{-1} = g^{-1} \cdot g = e$

Problem 7.2. Prove that the identity element is unique.

Problem 7.3. Prove that the inverse of an element g is unique.

These are two good exercises to get your hands on. Note that the group operation needn't be commutative! Consider an example you already know: let $\mathrm{GL}_n(F)$ be the subset of invertible $n \times n$ matrices with entries in a field F. Then this is a group under multiplication, with inverse and identity as one would imagine. Note that $M_n(F)$ is *not* a group under multiplication, as non-invertible matrices do not have inverses (obviously). However, $M_n(F)$ is a group under addition, and it's in fact commutative, i.e. $A + B = B + A$ for all $A, B \in M_n(F)$. Commutative groups are also called *abelian*.

Another way to get a group is to take your favourite field and start forgetting information. If F is a field, then $(F, +, 0)$ is an abelian group. If you take $F^\times = F \setminus \{0\}$, the set of nonzero elements of F, then $(F^\times, \cdot, 1)$ is an abelian group as well. We will talk about a generalisation of this below in the section on rings. We'll talk more about groups as they were developed in the next subsection.

What are the kinds of functions we're interested in?

Definition 7.4. A group homomorphism is a map of sets $\varphi\colon G \to H$ such that $\varphi(g_1 \cdot g_2) = \varphi(g_1) \cdot \varphi(g_2)$ for all $g_1, g_2 \in G$.

Problem 7.5. Prove that $\varphi(g^{-1}) = \varphi(g)^{-1}$ and $\varphi(e_G) = e_H$.

We consider an example. Let $(\mathbb{R}, +, 0)$ be the underlying (additive) abelian group of the real numbers, and let $(\mathbb{R}^{>0}, \cdot, 1)$ be the set of positive real numbers endowed with multiplication. Then we can define a group homomorphism $\exp\colon (\mathbb{R}, +) \to (\mathbb{R}^{>0}, \cdot)$ defined by $\exp(x) = e^x$. Certainly this map is a well-defined map of sets; the identity $e^{x-y} = e^x/e^y$ proves that it is a group homomorphism.

But there's a map backwards! We call this $\log\colon (\mathbb{R}^{>0}, \cdot) \to (\mathbb{R}, +)$. This is a group homomorphism because $\log(x/y) = \log(x) - \log(y)$. We can immediately verify that $\exp(\log(x)) = x$ and $\log(\exp(y)) = y$. Of course we have a special word for a group homomorphism with an inverse: it's called an *isomorphism*.

Definition 7.6. An isomorphism of groups is a group homomorphism that admits an inverse group homomorphism.

Problem 7.7. Prove an equivalent characterisation of a group isomorhpism: if $\varphi\colon G \to H$ is a bijective group homomorphism, then the set-theoretic inverse map is automatically a group homomorphism.

Definition 7.8. A subset $H \subset G$ is called a subgroup of G if $e \in H$ and for every $h_1, h_2 \in H$, $h_1 h_2^{-1} \in H$. That is, H includes the identity element and is closed under multiplication and inverses. We usually write $H < G$ in this case.

Consider the group $M_n(F)$ for a field F. Then the set of diagonal matrices is a subgroup – the sum of diagonal matrices is still diagonal, and the additive inverse of a diagonal matrix is diagonal. If we think about the multipliative group $\mathrm{GL}_n(F)$ instead, then diagonal matrices (with no nonzero entries) are still a subgroup! The product of diagonal matrices is diagonal, and to invert a diagonal matrix you just invert each element along the diagonal.

As a more abstract example of subgroups, let $g \in G$ and consider the set $\{g^n : n \in \mathbb{Z}\}$. Then it's an easy check that this satisfies the subgroup definition, and we write $\langle g \rangle < G$. Such a subgroup is called *cyclic*. Note that this set may be finite. If it is, we write $|g| = n$ for the order of g, and it's the minimal n such that $g^n = e_G$. Otherwise we say $|g| = \infty$.

There's a really important theorem on the order of subgroups (and hence of elements).

Theorem 7.9 (Lagrange's Theorem)**.** Let G be a finite group and let $H < G$ be a subgroup. Then $|H|$ divides $|G|$. In particular, $|g|$ divides $|G|$ for every $g \in G$.

Hence when we see GRE questions about the possible orders of elements and subgroups, this helps a lot.

We can talk about the subgroup generated by a number of elements $g_1, \dots, g_m$ in the obvious way. It's helpful to use the fact that the intersection of subgroups is still a subgroup (prove this if you don't believe it), and we can define

$$\langle g_1, \dots, g_m \rangle = \bigcap_{g_1, \dots, g_m \in H} H.$$

Similarly, we can talk about the subgroup generated by a family of subgroups,

$$HK = \{hk : h \in H, k \in K\} = \bigcap_{H, K \subset G'} G'$$

Since $H < HK$ and $K < HK$, this means that $|HK|$ needs to be a divisor of $|H| \cdot |K|$. In particular, $|HK| = |H| \cdot |K|$ if and only if $H \cap K = \{e_G\}$.

Problem 7.10. Prove it.

Definition 7.11. A normal subgroup is a subgroup $N < G$ such that $gNg^{-1} \subset N$ for all $g \in G$. In this case, we write $N \triangleleft G$.

Normal subgroups are very important. In particular, kernels of group homomorphisms are normal subgroups. Additionally, these are the appropriate objects in order to define quotients. For any $H < G$, we define

$$G/H = \{gH : g \in G\}$$

where $g_1H = g_2H$ if these subsets contain the same elements, i.e. there exists $h \in H$ such that $g_1 \cdot h = g_2$. In the case that $N \triangleleft G$ is a normal subgroup, G/N actually admits a group structure – $g_1N \cdot g_2N = g_1g_2N$.

Problem 7.12. Prove it.

Definition 7.13. A group G is called simple if it has no normal subgroups besides $\{e\}$ and G.

There are a variety of simple groups, but the biggest class of examples is C_p for the primes p (see below). Another choice will turn out to be A_n for $n \geq 5$, which we will define below. We elaborate on the essential types of groups below.

7.2. **Examples of groups.** It's probably about time to give some examples of (finite) groups. For every positive integer $n \in \mathbb{N}$, consider the set with n elements $X_n = \{1, \ldots, n\}$. Consider a bijection $f \colon X_n \to X_n$. We can put a group structure on this set, with the operation being composition. Identity and inverses are obvious. Call the set of these maps S_n and call it the symmetric group on n elements. Then $|S_n| = n!$, one can readily check.

We think about elements in S_n using a cycle decomposition. Let $n = 5$ for simplicity, and consider the following function:

$$f(1) = 2, \quad f(2) = 3, \quad f(3) = 5, \quad f(4) = 1, \quad f(5) = 4$$

We write this in the following format: we start by writing $(1-)$, and we then write the image of 1 to obtain $(12-)$, and so on until we get (12354). This is called a 5-cycle as it's written with 5 elements. Consider another function,

$$g(1) = 2, \quad g(2) = 3, \quad g(3) = 1, \quad f(4) = 5, \quad f(5) = 4$$

which yields up the cycle decomposition $(123)(45)$, which we call a 3-2-cycle. As a final example, consider

$$h(1) = 2, \quad h(2) = 1, \quad h(3) = 3, \quad h(4) = 4, \quad h(5) = 5$$

We could write this as (12)(3)(4)(5), but we'd rather write (12) and call it a 2-cycle or transposition. Note that in a cycle decomposition, the elements in the cycles must be disjoint. Every element of S_n has such a unique cycle decomposition up to permutation, and it's unique if we orient the cycle to begin with the lowest number left. That is

$$(123)(45) = (231)(54) = (312)(45)$$

but the first choice is canonical.

How do we multiply cycles? Consider (12)(13). This is a composition that says $1 \to 2$, $2 \to 1 \to 3$, and $3 \to 1$. This is the cycle (123). Consider now (13)(12). This says $1 \to 3$, $3 \to 1 \to 2$, and $2 \to 1$. Hence this is the cycle (132). These are different! The symmetric group S_n is not commutative. Now, we can address subgroups and orders.

Problem 7.14. The order of an m_1-m_2-$\cdots$-m_k-cycle is $\operatorname{lcm}(m_1, m_2, \ldots, m_k)$.

As such, there isn't an obvious formula for the maximal order of a cycle in S_n, but is easily computed for a given n.

There are some interesting subgroups of S_n that we can get using both group theory and geometry. The first geometric subgroup is the cyclic group C_n, which is generated by any n-cycle in S_n. All such groups are commutative. This represents rigid rotations of the regular n-gon. Of course, there's another symmetry of the regular n-gon, which is the flip along a vertical axis of symmetry. It's harder to write down a cycle decomposition, but we can do it for $n = 5$. Let the cycle be $(12345) = \sigma$, and then the flip is given by $(25)(34) = \tau$. The group generated by these two elements has $2n$ elements.

We can talk about groups in terms of generators and relations. For D_n, we write

$$D_n = \langle \sigma, \tau : \sigma^n = \tau^2 = 1, \tau\sigma\tau = \sigma^{-1} \rangle$$

There's an explicit embedding $D_n \to S_n$ as we can see above, but we can also think about D_n in abstract.

The last type of group to know is the alternating group $A_n \subset S_n$. This contains exactly half the elements, and can be described as the kernel of the map

$$\operatorname{sgn}\colon S_n \to C_2 = \{\pm 1\}$$

which sends an m_1-m_2-$\cdots$-m_k-cycle to $(-1)^{m_1+\cdots+m_k-k}$. The alternating group consists of the identity element, m-cycles for odd m, 2-2-cycles, etc.

7.3. **Abelian groups.** We now need to state the fundamental theorem on finitely generated abelian groups. Finitely generated is pretty obvious to define, but what's the theorem?

Theorem 7.15 (FTFGAG)**.** Let A be a finitely generated abelian group. Then

$$A \cong \mathbb{Z}^r \times \mathbb{Z}/n_1\mathbb{Z} \times \mathbb{Z}/n_2\mathbb{Z} \times \cdots \times \mathbb{Z}/n_k\mathbb{Z}$$

where $n_1 \mid n_2 \mid \cdots \mid n_k$. Alternatively,

$$A \cong \mathbb{Z}^r \times \mathbb{Z}/p_1^{\alpha_1}\mathbb{Z} \times \cdots \times \mathbb{Z}/p_\ell^{\alpha_\ell}\mathbb{Z}$$

for primes p_i and powers α_i.

Now, what does any of this mean? $\mathbb{Z}/n\mathbb{Z}$ is the cyclic group C_n, but we think of it additively and in terms of modular arithmetic. In particular, it's the quotient of $\mathbb{Z}$ by the normal subgroup $n\mathbb{Z} = \{n \cdot m : m \in \mathbb{Z}\}$, which we think of as generated under addition by n. In an abelian group, all subgroups are normal, so there's no problem there. The product is the same as the product of sets, and the group operation works in the obvious way: component by component.

Problem 7.16. How many abelian groups of order 4 are there (up to isomorphism)?

Solution. Expect a problem very much like this on the GRE. First of all, if $|A| = 4$, then we must have $r = 0$ in the above decomposition. Any copies of $\mathbb{Z}$ would make the group infinite.

Second, the order of $\mathbb{Z}/n_1\mathbb{Z} \times \cdots \times \mathbb{Z}/n_k\mathbb{Z}$ is the product $n_1 \cdots n_k$. Therefore we have to look at factorisations of 4, which are not plentiful. It's either 4 or $2 \cdot 2$. The fundamental theorem implies that $\mathbb{Z}/4\mathbb{Z}$ and $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ are different, so that makes 2 total groups. If you don't believe that they're different, solve the next problem.

It should be noted that $2 \cdot 2$ is a valid factorisation because $2 \mid 2$. If we were trying this same problem with 6 instead of 4, then the factorisation $2 \cdot 3$ would not be valid as $2 \nmid 3$.

Problem 7.17. Let $m, n \in \mathbb{N}$. Then $\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z} \cong \mathbb{Z}/mn\mathbb{Z}$ if and only if $\gcd(m, n) = 1$.

You can check both directions of that if you want. That means that the product of cyclic groups is still cyclic if and only if each of the terms are pairwise coprime. The two ways we think about the above decomposition depend how we group up the prime factors. We can either separate them as much as possible, or we can group them together. The specific details don't matter too much, but remember that theorem well.

7.4. **Rings.** Now we can talk about rings.

Definition 7.18. A (unital) ring is a set R such that $(R, +, 0)$ is an abelian group with another operation $\cdot$ satisfying:

- Multiplication is associative

- There exists $1 \in R$ such that $1 \cdot a = a \cdot 1 = a$ for all $a \in R$
- The distributive property holds: $a \cdot (b+c) = a \cdot b + a \cdot c$ and $(b+c) \cdot d = b \cdot d + c \cdot d$ for all $a, b, c, d \in R$

The multiplication is not necessarily commutative, as in the ring of matrices $M_n(F)$. If multiplication is a commutative operation we call R a *commutative ring* (shocking!). Here's something that ought to be true and is:

Problem 7.19. Prove that $0 \cdot a = 0$ for all $a \in R$.

There are again two types of substructures that need to consider.

Definition 7.20. A subring $S \subset R$ is an abelian subgroup that is closed under multiplication. Sometimes we demand that $1 \in S$, sometimes we don't.

Subrings aren't even that important.

Definition 7.21. A left ideal $I \subset R$ is an abelian subgroup I that is closed under left multiplication: for every $a \in R$ and $x \in I$, $a \cdot x \in I$. Similarly, we can define a right ideal and a two-sided ideal by demanding that I 'absorb' elements when multiplied on the right or on both sides (respectively).

Some examples of ideals: in the ring $\mathbb{Z}$, the ideals are all of the form $n\mathbb{Z} = \{m \cdot n : m \in \mathbb{Z}\}$ for any $n \in \mathbb{Z}$ (although $-n\mathbb{Z} = n\mathbb{Z}$). As a more complicated example, try the following:

Problem 7.22. Prove that the set of matrices

$$I = \left\{ \begin{bmatrix} a_1 & 0 & \cdots & 0 \\ a_2 & 0 & \cdots & 0 \\ \vdots & 0 & \cdots & 0 \\ a_n & 0 & \cdots & 0 \end{bmatrix} : a_i \in \mathbb{R} \right\} \subset M_n(\mathbb{R})$$

is a right ideal but not a left ideal.

The following problem is taken straight from a GRE question, and it outlines very well the ways we can construct ideals from old ideals.

Problem 7.23. Let $I, J \subset R$ be two (left/right/two-sided) ideals. Then prove that the following subsets are also (left/right/two-sided) ideals:

- $I + J = \{i + j : i \in I, j \in J\}$
- $I \cdot J = \left\{ \sum_{k=1}^{n} i_k \cdot j_k : i_k \in I, j_k \in J \right\}$
- $I \cap J$

In practice it might seem that we could define $I \cdot J = \{i \cdot j\}$, but this isn't always an ideal. The counterexample is a little off-topic, but you can find it in the solution for #50 on Practice 1. You can also prove that $I \cup J$ is *not* an ideal in general.

Ideals are important, subrings aren't (although you do need to know the definition of subring for the GRE). Here's another definition and an important consequence.

Definition 7.24. An element $a \in R$ is invertible if there exists $b \in A$ such that $a \cdot b = b \cdot a = 1$. In this case we call a a *unit* and we write the set of all units in R as $R^\times$.

As an easy exercise: prove that $R^\times$ forms a group under multiplication.

Problem 7.25. If $I \subset R$ is an ideal and $a \in I$ is a unit, then $I = R$.

This means that in a proper ideal (i.e. $I \neq R$), we can't have any invertible elements. There's one more type of element that needs defining:

Definition 7.26. A nonzero element $a \in R$ is a zero divisor if there exists a nonzero $b \in R$ such that $a \cdot b = 0$.

Problem 7.27. Prove that a zero divisor cannot be invertible and vice versa.

We can also talk about the left, right, or two-sided ideal generated by a subset of R. For an element $a \in R$, we denote by $aR = \{a \cdot r : r \in R\}$ the right ideal generated by a. The left ideal is denoted Ra and the two-sided ideal by RaR.

Finally, we can prove that if $I \subset R$ is a two-sided ideal, then R/I (the abelian group quotient) has the structure of a ring. Given two cosets $r + I$ and $s + I$, the only way to get

$$(r + I)(s + I) = rs + rI + Is + I^2 = rs + I$$

is to make sure that rI and Is are both contained in I for all $r, s \in R$, which is exactly what a two-sided ideal guarantees.

Now, what are the functions?

Definition 7.28. A (unital) ring homomorphism $\varphi\colon R \to S$ is an abelian group homomorphism such that $\varphi(1_R) = 1_S$ and $\varphi(r_1 \cdot r_2) = \varphi(r_1) \cdot \varphi(r_2)$.

As an example, consider the map $\varphi\colon \mathbb{Z} \to \mathbb{Z}$ such that $\varphi(n) = -n$. This is a perfectly good abelian group homomorphism, but it's not a ring homomorphism. In fact, since we demand that $\varphi(1) = 1$, there is only ever one map $\varphi\colon \mathbb{Z} \to R$ for any ring R, and it's defined by $\varphi(1) = 1_R$.

Kernels of ring homomorphisms are two-sided ideals, which is convenient, so that $R/\ker\varphi$ has the structure of a ring.

7.5. **Modular arithmetic.** Let $R = \mathbb{Z}$ and let $I = n\mathbb{Z}$. Then in the ring $\mathbb{Z}/n\mathbb{Z}$, we can do mathematics. The key is that we are working with 'remainders after dividing by n'. Let $n = 12$. Then for two examples,

$$8 + 7 = 15 \equiv 3, \quad 4 \cdot 5 = 20 \equiv 8$$

We can identify what elements in $\mathbb{Z}/n\mathbb{Z}$ are invertible and which are zero divisors. Supposing that d is a divisor of n, we know that $d \cdot n/d = n \equiv 0$. Even if $\gcd(d, n) = \alpha > 1$, it is still a zero divisor, because $d \cdot n/\alpha \equiv 0$. On the other hand, if $\gcd(d, n) = 1$, then we know that there's a solution to the expression $\alpha \cdot d + \beta \cdot n = 1$, so that $\alpha \cdot d \equiv 1$ and d is invertible.

There's one little theorem that we need to include that shows up on the GRE:

Theorem 7.29 (Fermat's Little Theorem)**.** Let $a \in \mathbb{Z}$ be an integer. Then for any prime p, $a^p \equiv a \pmod p$.

Put another way, in $\mathbb{Z}/p\mathbb{Z}$, we have $a^{p-1} = 1$ for any nonzero a. This can be used to prove some interesting little facts, like $n^2 - 1$ is always divisible by 3 for any n which is itself not divisible by 3.

Problem 7.30. Prove the above statement using Fermat's little theorem.

7.6. **Fields.** There's a special situation that we can see immediately. Suppose that $n = p$ is a prime. Then every $d \in \mathbb{Z}/p\mathbb{Z}$ is coprime to p, so that every element of $\mathbb{Z}/p\mathbb{Z}$ is invertible. A commutative ring in which every element is invertible is called a field.

But wait, we already know a lot of fields. We know $\mathbb{Q}, \mathbb{R}, \mathbb{C}$, and others! Well surprise, there are also finite fields. Field homomorphisms are just ring homomorphisms, but with a twist: let's examine $\varphi \colon F \to K$ for two fields F, K. We know that $\ker \varphi \subset F$ is a two-sided ideal. But since every element in F is invertible, we know that either $\ker \varphi = F$ or $\ker \varphi = \{0\}$. Since $\varphi(1_F) = 1_K$, we know that $\ker \varphi$ can't be everything. Thus $\ker \varphi = \{0\}$ and all field homomorphisms are injective.

A special case is that of field automorphisms. It's pretty hard to find field automorphisms sometimes. This is the realm of Galois theory, which isn't particularly covered on the GRE. As a special observation, there are no nontrivial automorphisms of $\mathbb{Q}$ or $\mathbb{F}_p$ (which is $\mathbb{Z}/p\mathbb{Z}$ when it has its field clothes on).

Problem 7.31. Prove that if $\varphi \colon \mathbb{Q} \to \mathbb{Q}$ with $\varphi(0) = 0$ and $\varphi(1) = 1$, then $\varphi = \mathrm{id}_{\mathbb{Q}}$.

The last thing to mention is two weaker notions than fields that are still more structured than a random unital ring.

Definition 7.32. An *integral domain* is a (unital) commutative ring that has no zero divisors.

Our favourite example is definitely $\mathbb{Z}$. If you consider the ring of polynomial functions with coefficients in a field (or an integral domain), it is still an integral domain, but not a field – you lose multiplicative inverses.

Definition 7.33. A *division ring* is a ring in which every (nonzero) element is a unit.

Again, a field is definitely a division ring, but there are non-commutative division rings (which therefore can't be fields). The *quaternions* is the key example, which are sort of like the complex numbers but a four-dimensional real vector space instead. In brief: the quaternions $\mathbb{H}$ are spanned by $\{1, i, j, k\}$ and multiplication is defined by

$$i^2 = j^2 = k^2 = ijk = -1$$

and from this you can deduce that $ij = k$, $jk = i$, $ki = j$, but we don't have commutativity: $ij = -ji$. This is intimately related to the cross product on $\mathbb{R}^3$ and the unit vectors $\hat{i}, \hat{j}, \hat{k}$. Since every element is a unit in a division ring, they don't have very many ideals (just like fields).

In general, most (unital) rings are just rings; they've got some units, some zero divisors, and some other stuff that's neither!

8. Day 8: Analysis and topology

Topics covered: suprema and infima, properties of $\mathbb{R}$, Lipschitz and uniform continuity of functions, pointwise and uniform convergence of functions, consequences of uniform convergence; abstract topological spaces and bases, compactness, connectedness and path connectedness, continuous functions, metrics and metric spaces.

8.1. The real numbers.

Let's go over some of the fancy analysis words when it comes to sequences in $\mathbb{R}$.

Definition 8.1. Let $S \subset \mathbb{R}$ be any subset. Then we define the supremum $\sup S$ (if it exists) to be the number $\alpha \in \mathbb{R}$ such that $\alpha \geq s$ for all $s \in S$ and, if $\beta < \alpha$, then there exists $s' \in S$ such that $\beta < s'$.

We can similarly define the infimum of S by $\inf S = -\sup(-S)$. The supremum is also called the least upper bound and the infimum the greatest lower bound. Another nice property of the supremum of a subset $S \subset \mathbb{R}$ is that, if $\alpha = \sup S$, then $\alpha - \varepsilon$ is not an upper bound for S for every $\varepsilon > 0$.

The following theorem/axiom is a defining property of the real numbers.

Theorem 8.2 (Axiom of Completeness)**.** Every bounded above nonempty subset of the real numbers has a supremum.

Related to this is the monotone convergence theorem.

Theorem 8.3 (Monotone Convergence)**.** Every bounded above monotone increasing sequence converges. In particular, it converges to the supremum of its set of values.

A corollary is that every bounded below monotone decreasing sequence converges.

The next theorem/axiom tells us how the real numbers interact with the natural numbers.

Theorem 8.4 (Archimedean Property)**.** Let $x \in \mathbb{R}$ be any real number. Then there exists a natural number $N \in \mathbb{N}$ such that $x < N$. Equivalently, for any $\varepsilon > 0$ in the real numbers, there exists $M \in \mathbb{N}$ such that $1/M < \varepsilon$.

Problem 8.5. Prove the equivalence of the above statements of the Archimedean Property.

The axiom of completeness (in combination with the archimedean property) is equivalent to a number of other main theorems of the first part of a real analysis course. We state them now.

Theorem 8.6 (Bolzano-Weierstrass)**.** Every bounded sequence of real numbers has a convergent subsequence.

Proof. A sketch: every (infinite) sequence of real numbers must have either an increasing or a decreasing (infinite) subsequence. (Problem: why?) In either case, the monotone convergence theorem applies so that this subsequence is bounded. □

Definition 8.7. A sequence $\{x_n\}$ is called *Cauchy* if for every $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that $|x_m - x_n| < \varepsilon$ whenever $m, n > N$.

Theorem 8.8 (Cauchy Criterion)**.** A sequence in $\mathbb{R}$ is Cauchy if and only if it converges.

Problem 8.9. Prove this using the Axiom of Completeness.

Problem 8.10. Give an example of a sequence in $\mathbb{Q}$ which is Cauchy but does not converge (in $\mathbb{Q}$).

Solution. Consider the decimal truncations of $\sqrt{2}$:

$$1, 1.4, 1.41, 1.414, \ldots$$

where a_n is accurate up to the first n decimal places (so $a_0 = 1$). This sequence is Cauchy, as $|a_n - a_m| < 10^{-\min(m,n)}$, but it converges (theoretically) to $\sqrt{2}$, which is not in $\mathbb{Q}$.

The final theorem actually relies on the Archimedean property; the preceding ones did not.

Theorem 8.11 (Nested Interval Property)**.** Suppose that $\{I_n = [a_n, b_n]\}$ are closed intervals such that $I_k \supset I_{k+1}$ for all k. Then

$$\bigcap_{n \in \mathbb{N}} I_n \neq \varnothing.$$

In particular, $\lim_{n\to\infty} a_n$ and $\lim_{n\to\infty} b_n$ are both in the intersection (though these may be equal).

If we assume that X is an ordered set such that $\mathbb{N} \subset X$ satisfies the Archimedean property, then any of the theorems {AoC, BW, CC, NIP} imply the other 3. The Archimedean property is used to make sure that the (decreasing) lengths of the intervals I_n can be turned into an enumerable list of values; without knowing that every $\varepsilon > 0$ can be described by n such that $1/n < \varepsilon$, this is not guaranteed.

8.2. **Topological interlude.** We will do some general topology later on, but we recall the various adjectives attached to subsets of the real numbers. Recall a *neighbourhood* of a point $a \in \mathbb{R}$ is the open interval $U_\varepsilon(a) = (a - \varepsilon, a + \varepsilon)$ for some $\varepsilon > 0$.

Definition 8.12. A subset $U \subset \mathbb{R}$ is called open if for every $a \in U$, there exists a neighbourhood $U_\varepsilon(a)$ that is completely contained in U.

Definition 8.13. A limit point of a subset $X \subset \mathbb{R}$ is $a \in \mathbb{R}$ (not necessarily in X) such that for every $\varepsilon > 0$, the intersection $U_\varepsilon(a) \cap X$ contains a point (of X) that is not a. A subset $X \subset \mathbb{R}$ is called closed if it contains all its limit points.

Problem 8.14. Prove that open intervals (as you understand them from Calc I) are open sets (as in the above definition). Repeat the problem but with 'closed' instead of 'open'.

Problem 8.15. Prove that $U \subset \mathbb{R}$ is open if and only if the complement U^c is closed.

The following is kind of a theorem but kind of a definition. Therefore we state it as a problem.

Problem 8.16. Let $\{U_i\}_{i\in I}$ be an arbitrary collection of open sets. Prove that the union $\bigcup_{i\in I} U_i$ is still open. Prove that any finite intersection $U_{i_1} \cap \cdots \cap U_{i_n}$ is open.

Using the previous problem and de Morgan's laws about commuting unions/intersections and complements, we obtain the corollary that closed sets are closed under arbitrary intersection and finite union.

One more definition worth stating at this point.

Definition 8.17. A subset $X \subset \mathbb{R}$ is called (sequentially) compact if even sequence in X contains a subsequence which converges in X.

But we know this is not the best description of compact subsets of $\mathbb{R}$; that comes from the Heine-Borel theorem. Two more quick things to say before stating that ...

Definition 8.18. A collection of open sets $\{U_i\}_{i\in I}$ is called an open cover of $X \in \mathbb{R}$ if $\bigcup_{i\in I} U_i = X$.

Definition 8.19. A subset $X \subset \mathbb{R}$ is called bounded if there exists $M > 0$ such that $|x| < M$ for all $x \in X$.

Theorem 8.20 (Heine-Borel)**.** Let $X \subset \mathbb{R}$ be any subset. The following are equivalent:

- Every open cover of X admits a finite subcover (usually called quasicompact).
- X is sequentially compact.
- X is closed and bounded.

Problem 8.21. Prove as many implications in the above theorem as you can.

8.3. **Continuity revisited.** Throughout let $X \subset \mathbb{R}$ denote a subset of real numbers. You can even imagine that X is an interval (bounded or unbounded; open or closed or half and half).

Definition 8.22. We say a function $f\colon X \to \mathbb{R}$ is uniformly continuous if for all $\varepsilon > 0$, there exists $\delta > 0$ such that whenever $|x_1 - x_2| < \delta$, $|f(x_1) - f(x_2)| < \varepsilon$.

This is strictly stronger than continuous: it says that the same δ can be used at any point in the domain, not just at the particular point $x = a$ you want. That's why we don't talk about '$f(x)$ is uniformly continuous at $x = a$'; this is a global condition. As such, the same function f may be uniformly continuous on some domains but not others. More on that soon.

Problem 8.23. Prove that $f(x) = mx + b$ is uniformly continuous on all of $\mathbb{R}$.

Solution. The general proof looks at $f(x_1) - f(x_2)$ and tries to transform it into something more like $x_1 - x_2$. It's very easy in this case:

$$|f(x_1) - f(x_2)| = |mx_1 + b - (mx_2 + b)| = |m| \cdot |x_1 - x_2|$$

Therefore as long as $|x_1 - x_2| < \varepsilon/|m|$, we have $|f(x_1) - f(x_2)| < \varepsilon$. That gives us our uniform choice of δ.

Being able to actually solve $K \cdot \delta = \varepsilon$, as in the preceding example, has its own name.

Definition 8.24. A function $f \colon X \to \mathbb{R}$ is called Lipschitz continuous if there exists a constant $K > 0$ such that $|f(x_1) - f(x_2)| \leq K \cdot |x_1 - x_2|$ for all $x_1, x_2 \in X$.

In particular, the choice $\delta = \varepsilon/K$ shows that Lipschitz continuous implies uniformly continuous (implies continuous). This is a piece of vocabulary that might be useful on the GRE, though I haven't seen a problem on this topic explicitly.

Problem 8.25. Prove that $f(x) = \sqrt{x}$ is uniformly continuous on $[0, \infty)$ but not Lipschitz continuous.

Solution. For uniform continuity[2],

$$|\sqrt{x_1} - \sqrt{x_2}|^2 \leq |\sqrt{x_1} - \sqrt{x_2}| \cdot |\sqrt{x_1} + \sqrt{x_2}| = |x - y|$$

Therefore the choice $\delta = \varepsilon^2$ does the job for us. This doesn't look very linear, however, which will cause the problem with Lipschitz continuity.

In particular, we would need to find $K > 0$ such that

$$|\sqrt{x} - \sqrt{0}| \leq K \cdot |x - 0|$$

for all $x > 0$. We immediately see the problem: solving this out implies that we need $K \geq 1/\sqrt{x}$ for all $x > 0$, but this is impossible once $x < K^2$. In general, a function cannot be Lipschitz if its derivative is unbounded; since $f'(x) = 1/2\sqrt{x}$, we see that it gets arbitrarily large as $x \to 0$.

Back to uniform continuity: there's a nice way to conclude a function is uniformly continuous.

[2] Lifted from here

Problem 8.26 (Heine-Cantor Theorem)**.** If X is a compact subset of $\mathbb{R}$ and $f\colon X \to \mathbb{R}$ is continuous, then it is also uniformly continuous.

Solution. Recall that one version of compact is that every open cover of X admits a finite subcover. Since X is continuous, let's define the sets C_x for all $x \in X$ as follows: fixing $\varepsilon > 0$, let

$$C_x = \{x' \in X : |f(x) - f(x')| < \varepsilon/2\}$$

In other words, it's the set around x that satisfies the conditions of uniform continuity for a slightly smaller epsilon. We can then define $U_x = U_\delta(x)$ to be the biggest open interval $U_\delta(x) \subset C_x$. For one more refinement, consider $V_x = U_{\delta/2}(x)$, the interval with half the maximal radius. The collection $\{V_x\}$ is (obviously) an open cover of X, so there's some finite collection $x_1, \ldots, x_n$ such that $V_i = U_{\delta_i/2}(x_i)$ cover X.

Consider now $\delta = \dfrac{1}{2}\min \delta_i$. This is a positive number because we are taking a minimum (instead of, say, an infimum). Moreover, take any $z_1, z_2 \in X$ with $|z_1 - z_2| < \delta$. Without loss of generality, we have $z_1 \in V_1$ (otherwise relabel). Then

$$|z_2 - x_1| \leq |z_2 - z_1| + |z_1 - x_1| < \delta + \frac{\delta_1}{2} \leq \frac{\delta_1}{2} + \frac{\delta_1}{2} = \delta_1$$

which implies that $z_1, z_2 \in U_1$. Therefore

$$|f(z_1) - f(z_2)| \leq |f(z_1) - f(x_1)| + |f(x_1) - f(z_2)| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

which proves that $f(x)$ is uniformly continuous.

This proof is a good reminder of the utility of compactness, and that on compact domains many results are upgradeable from local (i.e. continuity begin defined at points) to global (having a uniform δ for each ε).

Uniformly continuous functions have a nice property.

Theorem 8.27. Suppose that $f\colon Z \subset X \to Y$ is uniformly continuous on a subset $Z \subset X$. Then there is a unique extension $\overline{f}\colon \overline{Z} \to Y$ defined on the closure of Z that is still continuous.

This doesn't work if the function isn't uniformly continuous. For instance, let $f(x) = 1/x$ be defined on $f\colon (0,1) \to \mathbb{R}$. Then there is no way to continuously extend $f(x)$ to $[0,1]$ as we would need

$$f(0) = \lim_{x \to 0} \frac{1}{x}$$

and this limit diverges. The case of uniform continuity ensures that we don't run into this problem.

There's also absolute continuity, but I don't think we need to recall that.

8.4. **Convergence of functions.** We can now begin to talk about convergence of functions. We will restrict our attention to functions $f\colon X \subset \mathbb{R} \to \mathbb{R}$ once again.

Definition 8.28. Consider a sequence of functions $f_n\colon X \to \mathbb{R}$. Then we can define a new function $f\colon X \to \mathbb{R}$ by

$$f(x) = \lim_{n\to\infty} f_n(x)$$

for every $x \in X$, assuming all these limits exist. In this case, we say that $\{f_n\}$ converges to f pointwise.

This is a pretty good definition. For instance, let $f_n\colon [0,1] \to \mathbb{R}$ be defined by $f_n(x) = x^n$. Then it's pretty clear that the pointwise limit exists and

$$f(x) = \begin{cases} 0 & x \in [0,1) \\ 1 & x = 1 \end{cases}$$

This presents our conundrum. Each of the functions $f_n(x)$ is continuous, but their pointwise limit is not! We need to introduce more refined version of convergence that takes into account that we have an entire function, not just a series of points.

Definition 8.29. Let $\{f_n\colon X \to \mathbb{R}\}$ be a sequence of functions. We say that $\{f_n\}$ converges to f uniformly if it converges pointwise and, for every $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$|f_n(x) - f(x)| < \varepsilon \text{ for all } n > N$$

That is, the pointwise limits are all getting close to the limit function $f(x)$ simultaneously.

Note that since $\mathbb{R}$ is complete, we could also demand that the sequence $\{f_n\}$ is uniformly Cauchy rather than uniformly continuous, which is sometimes easier.

Problem 8.30. Write down the definition for a sequence of functions being uniformly Cauchy. Note that you don't need to specify a pointwise limit $f(x)$ to do so!

The point is this:

Theorem 8.31. Suppose that $\{f_n\}$ is a sequence of continuous functions that converge uniformly to f. Then f is also continuous.

This must mean that $f_n(x) = x^n$ does not converge uniformly to the limit function. To see this, fix $1 > \varepsilon > 0$ and any $N \in \mathbb{N}$. We will show that there exists $x \in [0,1]$ such that $|f_N(x) - f(x)| > \varepsilon$. Specifically, we are going to choose an $x \in (0,1)$ so we just need to prove that $x^N > \varepsilon$. But this is easy: take any $1 > \delta > \varepsilon$ and let $x = \delta^{1/N}$.

Can we upgrade this theorem in the case that we know that $\{f_n\}$ are also uniformly continuous? Yes.

Problem 8.32. Prove it.

8.5. **Integrals.** Now let's address the issue of integrals. Suppose that we have a sequence of integrable functions $f_n\colon [a,b] \to \mathbb{R}$. Suppose further that $f_n \to f$ pointwise. Does it follow that f is integrable? If so, do we have an equality

$$\lim_{n\to\infty} \int_a^b f_n(x)\,dx = \int_a^b f(x)\,dx$$

The answer is not necessarily. An easy counterexample is the following: let $f_n\colon (0,1] \to \mathbb{R}$ be defined by

$$f_n(x) = \begin{cases} n & x \in (0, 1/n] \\ 0 & \text{else} \end{cases}$$

Then each $f_n(x)$ isn't continuous, but it is bounded with finitely many discontinuities, which makes it integrable. In addition,

$$\int_0^1 f_n(x)\,dx = 1$$

for all $n \in \mathbb{N}$.

Now, $f_n \to 0$ pointwise because $f_n(x) = 0$ for all $n > 1/x$. But, as one observes

$$\int_0^1 0\,dx \neq 1$$

Thus, something has gone wrong. As something specific to notice, the functions $\{f_n\}$ are not uniformly bounded by any constant. Consequently, $\{f_n\}$ does not converge to the zero function uniformly. We have two theorems that give us the means to commute the integral and the limit.

Theorem 8.33 (Uniform convergence theorem)**.** If $f_n \to f$ uniformly as functions $[a,b] \to \mathbb{R}$ and all f_n are integrable, then f is integrable and

$$\lim_{n\to\infty} \int_a^b f_n(x)\,dx = \int_a^b f(x)\,dx$$

Theorem 8.34. Suppose $f_n \to f$ pointwise as functions $[a,b] \to \mathbb{R}$ and all f_n are integrable. Suppose further that for all $n \in \mathbb{N}$, $|f_n(x)| \leq g(x)$ for an integrable function $g\colon [a,b] \to \mathbb{R}$. Then

$$\lim_{n\to\infty} \int_a^b f_n(x)\,dx = \int_a^b f(x)\,dx$$

I'm trying to avoid measure theory in the description of this theorem, and I think we don't need it. We might not even need this theorem on the GRE, but it's easy enough to state.

8.6. **Topology.** I think we can now talk about topology in abstract.

Definition 8.35. A topological space is a set X along with a subset $\mathcal{T} \subset \mathcal{P}(X)$ such that:

- $\varnothing, X \in \mathcal{T}$
- If $\{U_i\}_{i \in I}$ is a collection of elements in $\mathcal{T}$, then so is $\bigcup_{i \in I} U_i$
- If $U_1, \ldots, U_n$ is a finite collection of elements in $\mathcal{T}$, then so is $\bigcap_{i=1}^{n} U_i$

The set $\mathcal{T}$ is called a topology on X. The sets $U \in \mathcal{T}$ are called open sets. A set V such that $V^c \in \mathcal{T}$ is called closed. Note that you could also define a topology using closed sets and a dual set of axioms.

Every subset $S \subset X$ has an interior and a closure. The interior S° is the union of all open sets $U \subset S$ and the closure $\overline{S}$ is the intersection of all closed sets $S \subset C$.

There are always two topologies on any set X, namely the maximal choice $\mathcal{T} = \mathcal{P}(X)$ called the discrete topology and the minimal choice $\{\varnothing, X\}$ called the indiscrete topology.

Something that we might care about is when two topologies are the same, i.e. when they have exactly the same open sets. Well, usually a topology is defined using a generating set, in the following sense:

Definition 8.36. A subset $B \subset \mathcal{T}$ is called a base of the topology $\mathcal{T}$ on X if:

- $\bigcup_{U \in B} U = X$
- For every $U_1, U_2 \in B$ and every $x \in U_1 \cap U_2$, there exists a $U_3 \in B$ containing x

Problem 8.37. Suppose that B_1 is a base of $\mathcal{T}_1$ and B_2 a base of $\mathcal{T}_2$ on a set X. Suppose further that for every $U_2 \in B_2$, there exists $U_1 \in B_1$ such that $U_1 \subset U_2$ and vice versa. Then $\mathcal{T}_1 = \mathcal{T}_2$.

If we have that $\mathcal{T}_1 \subset \mathcal{T}_2$, we say that $\mathcal{T}_1$ is coarser than $\mathcal{T}_2$, or that $\mathcal{T}_2$ is finer than $\mathcal{T}_1$. In linguistic terms, having more open sets makes the set X more smooth. We can also check this on bases.

We care mostly about continuous functions of topological spaces, which have the following definition.

Definition 8.38. Let $f \colon X \to Y$ be a function between two topological spaces. We say that f is continuous if for every open set $V \subset Y$, the preimage $f^{-1}(V) \subset X$ is also open.

Just like in algebra, we also want a notion of 'isomorphism', though the vocabulary differs in this case.

Definition 8.39. A function $f\colon X \to Y$ is called a homeomorphism if it is bijective, continuous, and open, where open means $f(U)$ is open in Y for every open $U \subset X$. Equivalently, we can ask that f is bijective, continuous, and f^{-1} is also continuous.

Homeomorphic spaces are essentially the same, as far as topology (though not real analysis) is concerned.

It's an important point to note that, while a bijective group homomorphism is automatically an isomorphism, a bijective continuous map of topological spaces is not necessarily homeomorphism – the condition of being open is not guaranteed!

Problem 8.40. Let X be any topological space and let X^d be the same set with the discrete topology. Prove that the map $X^d \to X$ given on underlying sets by the identity is bijective and continuous, but its inverse is not (usually) continuous. Under what conditions is the inverse continuous?

Now, what kind of sets are there besides open and closed?

Definition 8.41. A set $Z \subset X$ is called disconnected if there exist open sets $U, V \subset X$ such that $U \cup V = Z$ and $U \cap V = \varnothing$. A set that is not disconnected is called connected.

Problem 8.42. Prove that intervals in $\mathbb{R}$ are connected. Recall that an interval is defined by $I \subset \mathbb{R}$ such that for any $a, b \in I$ with $a < b$, every point $c \in (a, b)$ is also in I.

Definition 8.43. A set $Z \subset X$ is called path-connected if for every $a, b \in Z$, there exists a continuous function $\gamma\colon [0,1] \to Z$ such that $\gamma(0) = a$ and $\gamma(1) = b$.

Problem 8.44. Prove that every path-connected set is connected.

The converse is not true.

Problem 8.45. Prove that the graph $\Gamma \subset \mathbb{R}^2$ of the function

$$f(x) = \begin{cases} \sin\left(\frac{1}{x}\right) & x \neq 0 \\ 0 & x = 0 \end{cases}$$

is connected but not path connected.

As a hint: try to define a function $\gamma\colon [0,1] \to \Gamma$ with $\gamma(1) = (0,0)$.

Here is one more problem you can prove either in $\mathbb{R}$ or in the abstract topological context:

Problem 8.46. The image of a connected set under a continuous function is connected. The image of a compact set under a continuous function is compact.

The opposite of being connected is, of course, being disconnected. For instance, a space like $(0, 1) \cup (2, 3)$ is disconnected because it's in two different pieces. But each of those pieces is an interval, and intervals are connected. So this space is disconnected in basically one way.

Definition 8.47. A subset $X \subset \mathbb{R}$ is called totally disconnected if for any two points $x_1, x_2 \in \mathbb{R}$, there exist open sets $U, V \subset X$ forming a disconnection (i.e. $U \cup V = X$ and $U \cap V = \varnothing$) with $x_1 \in U$ and $x_2 \in V$.

So $(0, 1) \cup (2, 3)$ is not totally disconnected, as you can't actually separate the points within $(0, 1)$ or $(2, 3)$.

Problem 8.48. Prove that $\mathbb{Q} \subset \mathbb{R}$ is totally disconnected.

I'll leave one last general review problem.

Problem 8.49. Define the Cantor set $C \subset [0, 1]$ by the following inductive process: let $C_0 = [0, 1]$. Define C_n by removing the middle third from each of the intervals in C_{n-1}. So $C_1 = [0, 1/3] \cup [2/3, 1]$, $C_2 = [0, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup [8/9, 1]$, etc. Define

$$C = \bigcap_{n \in \mathbb{N}} C_n.$$

Prove that C is closed and totally disconnected. Then review all the other nice facts about the Cantor set using your favourite real analysis textbook.

8.7. **Metric spaces.** There's one last thing to mention, which is the concept of a general metric space.

Definition 8.50. A metric space is a set X along with a function $d \colon X \times X \to [0, \infty)$ such that for any points $x_i \in X$:

- $d(x_1, x_2) = 0$ if and only if $x_1 = x_2$ (separation).
- $d(x_1, x_2) = d(x_2, x_1)$ (symmetry).
- $d(x_1, x_2) + d(x_2, x_3) \geq d(x_1, x_3)$ (triangle inequality).

Any set always admits at least one metric, called the discrete metric:

$$\delta(x_1, x_2) = \begin{cases} 0 & x_1 = x_2 \\ 1 & x_1 \neq x_2 \end{cases}$$

You could also change 1 to any other number, but that's not really important – the distance between any two non-identical points is always the same.

Any metric space also gives rise to a topology on its underlying set: define a base of the topology on X by the open sets $U_\varepsilon(x) = \{x' \in X : d(x, x') < \varepsilon\}$ for all $\varepsilon > 0$. In $\mathbb{R}$, we obtain the usual notion of open intervals $(x - \varepsilon, x + \varepsilon)$. If X has the discrete metric, we obtain the discrete topology.

There are a number of transformations we can do to metrics and still obtain a metric. Suppose that d_1, d_2 are two metrics on a space X. Then the following are also metrics:

- $C \cdot d_1$ for any constant $C > 0$ in $\mathbb{R}$.
- $\delta = \min(d_1, d_2)$.
- $1 - \dfrac{1}{d_1 + 1} = \dfrac{d_1}{d_1 + 1}$. This scales the metric so that the distance between any two points is at most 1, but relative distances do not change.

Problem 8.51. Let X be a metric space with metric d, and consider the topological space that is defined by the metric. Does the metric $C \cdot d$ give a different (i.e. non-homeomorphic) topology? Does the metric $\dfrac{d}{d+1}$ give a different topology?

9. Day 9: Miscellaneous

Topics covered: probability and combinatorics, statistics, geometry, set theory, graph theory, algorithms. This section is purposefully shorter because a) there isn't that much to talk about and b) it leaves time for re-covering some of the more important material from previous days.

9.1. Combinatorics.

Permutations and combinations problems are some of the most straightforward, but they can be asked in tricky ways. Given n objects to be put in some order (where the order matters), we have $n!$ options. Given a choice of m objects from n objects (where the order matters), we have $n!/(n-m)!$ options – put another way, it's $n \cdot (n-1) \cdots (n-m+1)$ the product of m things counting down from n.

There's a small wrinkle that's worth discussing: if order matters only up to some cyclic permutation, then the mathematics is a little different.

Problem 9.1. How many ways are there to arrange 8 people around a circular table?

Solution. Imagine first that we put these people in a line – there are 8! options. But in a circle, there are 8 ways the table could be rotated which give the same permutation – thus there are $8!/8 = 7!$ choices. In general, arranging n people in a circle has $(n-1)!$ options.

If order doesn't matter, the situation is a little different. The choice of m objects out of a pool of n objects is, as we said above, $n!/(n-m)!$; but if order doesn't matter, we have to divide out the number of rearrangements of those m objects, thus:

$$\binom{n}{m} = \frac{n!}{m!(n-m)!}$$

where we read the lefthand side as 'n choose m'. Note that this is symmetric in the following sense:

$$\binom{n}{m} = \binom{n}{n-m}$$

because choosing m objects is the same as excluding $n-m$ objects.

These are also related to binomial expansions:

$$(a+b)^n = \sum_{m=0}^{n} \binom{n}{m} a^m b^{n-m}$$

which is sometimes useful to remember.

The last thing isn't quite combinatorics, but is related. The pigeonhole principle says that if you want to sort n objects into m bins, then if $n > m$ strictly, there must be at least two objects in one of the bins. There are a few problems on the practice exams whose reasoning comes down to this principle, e.g. one about the dartboard or one involving the distribution of primes, squares, etc.

9.2. **Probability via area.** We can imagine a probability space Ω and events A and B being subsets of Ω, such that the area/volume of Ω is 1 and hence $P(A)$ is given by the volume or area of A. Then looking at the probability of $P(A \text{ and } B)$ is just given by the intersection of these areas, and similarly for $P(A \text{ or } B)$ the union of these areas.

When you're trying to compare continuous random variables $x, y, z \in [0,1]$ (for instance), the volume approach is very useful, as we've seen.

Problem 9.2. If x, y are randomly chosen in $[0,1]$, what is the probability that $x \geq 2y$?

Solution. We can picture this as the double integral where $y \in [0,1]$ and $x \in [2y, 1]$. Except that this doesn't make total sense, because $2y > 1$ when $y > 1/2$, so we really have to integrate $y \in [0, 1/2]$.

$$\int_0^{1/2} \int_{2y}^1 1\, dx\, dy = \int_0^{1/2} 1 - 2y\, dy = \frac{1}{2} - \left(\frac{1}{2}\right)^2 = \frac{1}{4}$$

We can also do this via drawing the picture and computing the area of the triangle.

The same works in 3D.

Problem 9.3. Given three random variables x, y, z chosen in $[0,1]$, what are the odds that $x + y + z > 1$?

It will be easier to complete this problem if you find $P(x + y + z \leq 1)$, as this region is volume under the graph of a function $D \subset \mathbb{R}^2 \to \mathbb{R}$. Then $P(x + y + z > 1) = 1 - P(x + y + z \leq 1)$ since these events form a partition of the total cube $[0,1]^3$.

9.3. **General probability.** Conditional probability: $P(A|B) = \dfrac{P(A \text{ and } B)}{P(B)}$. Read 'probability of A given B'.

We say that A and B are independent if $P(A \text{ and } B) = P(A) \cdot P(B)$. Equivalently, $P(A|B) = P(A)$.

There's a nice way to swap the order of conditional probability, called Bayes' theorem.

Theorem 9.4.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

In action:

Problem 9.5. Consider drawing two cards from a deck. Compute the probability of drawing a spade first given that you drew a spade second.

Solution. To state this first for those who never ran into these problems before (due to being non-American or not knowing about cards): in a deck of cards, there are 4 suits (spades, hearts, clubs, diamonds) of 13 cards each: 1-10, Jack, Queen, and King. The 1 is called an ace and written A sometimes.

Let A be the first spade event and B the second spade event. We can work out the probability $P(B|A)$ explicitly: there are 51 cards left and 12 spades, so $P(B|A) = 12/51$. We can also compute $P(A)$ and $P(B)$: $P(A) = P(B) = 1/4$. We can see that $P(B) = 1/4$ by noting that it's just taking a random card from the deck (that happens to be second from the top), so there's even odds that it's a spade. Thus $P(A|B) = P(B|A) = 12/51$.

Now these have been discrete probabilities here, but what about continuous ones?

Definition 9.6. A probability distribution function for a random variable X is a positive integrable function $f\colon \mathbb{R} \to \mathbb{R}$ such that

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

It has a corresponding continuous distribution function defined by

$$P(X \leq a) = F(a) = \int_{-\infty}^{a} f(x)\,dx$$

An example is that of the standard normal distribution, which is given by the PDF

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

While you can't integrate $f(x)$ yourself (try it?), it can be shown that the integral is indeed 1. Since $f(x)$ is an even function, we can see that

$$F(0) = \int_{-\infty}^{0} f(x)\,dx = 0.5$$

meaning that the y-axis is the center of this distribution.

9.4. **Statistics.** The expected value of a discrete random variable X (taking values in $\mathbb{R}$) is

$$E(X) = \sum_{A \in X} P(A) \cdot A$$

Problem 9.7. Consider the roll of a 6-sided die. What is the expected value?

For a continuous random variable with pdf $f(x)$ is

$$E(X) = \int_{\infty}^{\infty} x \cdot f(x)\, dx$$

Problem 9.8. Compute the expected value for the standard normal distribution $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.

This one you can integrate (using u-substitution).

The variance can be calculated as $E(X^2) - E(X)^2$. Specifically,

$$\sum_{A \in X} P(A)(A - E(X))^2, \quad \int_{-\infty}^{\infty} (x - E(X))^2 \cdot f(x)\, dx$$

The standard deviation σ is the square root of the variance.

What do we know about standard deviation? Well, not very much in general, but on the GRE everything will be normally distributed (not necessarily using the standard normal distribution above, but it's similar).

Suppose we have a normal distribution with expected value μ (i.e. the mean) and standard deviation σ. Within ± 1 standard deviation of μ is 68% of the distribution, within $\pm 2\sigma$ is 95% and within $\pm 3\sigma$ is 99.7%.

A key example is that of binomial distributions. Suppose we have an event with probability p and we perform n trials. Then the expected value of successful trials is $n \cdot p$. If we repeat this situation a bunch of times, we can look at the number of trials that were actually successful. The variance of this distribution is $n \cdot p \cdot (1 - p)$ and so the standard deviation is the square root of this.

Problem 9.9. Suppose we roll a 20-sided die 400 times. Consider the probability of rolling a prime number. What is the expected number of successes and what is the standard deviation?

Solution. We have $p = 8/20 = 2/5$; try it from there.

9.5. **Geometry.** Triangles: let's start there. There's the law of sines and the law of cosines, which I can recall, but there's also Heron's formula for the area of a triangle: if the sides are a, b, c, then

$$A = \sqrt{s(s-a)(s-b)(s-c)}$$

where $s = \dfrac{a+b+c}{2}$ is the semiperimeter.

This is pretty nice when you know that the triangle is equilateral and then the area becomes $\dfrac{x^2\sqrt{3}}{4}$ where x is the side length.

The measure of the angle of a regular n-gon is $\dfrac{180(n-2)}{n}$. That can be useful.

You can also try to find the area of an inscribed polygon or a circumscribed polygon around a circle. This is easiest with (equilateral) triangles, squares, hexagons, and octagons.

Problem 9.10. What is the area of a regular hexagon inscribed in the unit circle? What about a hexagon circumscribed about the unit circle?

Solution. In the first case: a hexagon may be split into 6 equilateral triangles, and the side length of these triangles is the radius of the (unit) circle, which is to say 1. Since the area of an equilateral triangle is $s^2\sqrt{3}/4$, we compute the total area as

$$6 \cdot \frac{\sqrt{3}}{4} = \frac{3\sqrt{3}}{2} \approx 2.6$$

which is less than π, as we would expect.

In the second case: if the hexagon is on the outside, then the radius of the circle is now the height of the equilaterial triangles, making the side length $2 \cdot 1/\sqrt{3}$ instead. This makes the area

$$6 \cdot \frac{(2/\sqrt{3})^2\sqrt{3}}{4} = 2\sqrt{3} \approx 3.5$$

which is more than π, another good check.

9.6. **Set theory.** Let X be a set. The power set of X, denoted $P(X)$, is defined by $\{Y : Y \subset X\}$ the set of all subsets of X. If $|X|$ denotes the cardinality of the set, then necessarily $|X| < |P(X)|$. In particular, if $|X|$ is finite, then $|P(X)| = 2^{|X|}$.

Definition 9.11. We say that a set X is countable if there exists a surjective function $\mathbb{N} \to X$ or an injective function $X \to \mathbb{N}$.

A set has a cardinality greater than countable is called uncountable. Note that a countable set might be infinite or might be finite. Sometimes the phrase 'most countable' is used instead, and countable is reserved for infinite countable sets.

We denote by $\aleph_0$ the cardinality of the natural numbers. Note that any countable set must have cardinality at most $\aleph_0$, a surjective function must be from a larger (or equal) cardinality to a smaller.

A countable union of countable sets is still countable, and a finite product of countable sets is countable, but a *countable* product of countable sets is definitely not countable anymore. Consider the set $X = \{0, \ldots, 9\}$. Then taking the infinite product $\prod_{n\in\mathbb{N}} X_n$, where $X_n = X$, we can consider the map

$$[0,1] \to \prod_{n\in\mathbb{N}} X_n$$

which sends a decimal $0.a_1a_2\cdots$ to the set element $(a_1, a_2, \ldots)$. Ignoring the small detail that infinite decimal expansions aren't quite unique, this defines an injective function. Since $[0,1]$ is uncountable, so is the infinite product of these finite sets X_n.

If you can't remember why $[0, 1]$ is uncountable, I recommend the following.

Problem 9.12. Review Cantor's diagonalisation argument for the uncountability of the real numbers.

9.7. **Graph theory.** A graph is a set V of vertices and a set E of edges which connect the vertices together. There are a number of adjectives we attach to this situation. If we think of the edges as having a start and finish, we call the graph directed; otherwise it is undirected (but we usually assume graphs are undirected). If there is at most one edge between any two vertices and no loops (i.e. edges from v to itself), we call the graph simple. We say that the degree of a vertex is the number of edges attached to it.

A cycle on a graph is a traversal from a vertex v back to itself along a distinct set of edges. If an undirected graph is not simple, it will always have a cycle: either the loop $v \to v$ or the path $v \to w \to v$ along a multiple edge $v \rightrightarrows w$. A graph that has no cycles at all is called a tree; the classification or counting of trees is a common problem in graph theory. Classifying trees with 10 vertices is present in the movie Good Will Hunting, but it is not a problem that takes years to solve (as claimed).

Problem 9.13. Classify all trees with 10 vertices up to isomorphism that contain no vertices of degree 2.

The reason for asking for no vertices of degree 2 is the following: if you have a vertex of degree 2, you can split the tree in half at that vertex and obtain two other trees. Thus we are looking for trees that cannot be formed by gluing together from two other trees. If you need help with this problem, there are plenty of solutions available online.

9.8. **Algorithms.** Learn some Python? If you don't know any computer science, it's a bit tricky. I guess just try to treat the algorithm like a proof with input. In most iterations of the class there haven't been any questions about the one or two algorithm problems on the practice test.