Welcome to "Insane 2hr Cram Session of Undergraduate Mathematics," given on November 17th at MIT Splash! Let's get to it.

1 Topics we'll touch on:

1. **Calculus:** partial derivatives, multiple integrals, Stokes' theorem, differential equations
2. **Linear algebra:** an introduction to proofs, fields, vector spaces, linear transformations and functionals, matrix algebra, eigenvalues and eigenvectors
3. **Real analysis:** metric spaces, general topology, continuity, differentiation, integration, Lebesgue measure, convergence, analytic functions, Hilbert and Banach space
4. **Complex analysis:** complex numbers, holomorphic functions of a single variable, power series, contour integration, residues, Cauchy's theorem
5. **Probability theory:** random variables, measure theory, distributions (uniform, exponential, beta, gamma, normal), Poisson processes, expectation and moments, conditioning, law of large numbers, central limit theorem, martingales, stochastic processes
6. **Algebra/algebraic topology:** groups, subgroups, rings, modules, category theory, homeomorphism, homotopy, fundamental groups, covering spaces
7. **Differential topology/geometry:** manifolds, smoothness, immersions and embeddings, winding numbers, differential forms, de Rham cohomology, Lie groups
8. **A brief discussion of other applications**
9. **Tips for succeeding in mathematics at college**

2 Calculus

Let $f : \mathbb{R}^n \to \mathbb{R}, f : (x_1, ..., x_n) \mapsto y$ be a differentiable function (we'll get to what "differentiability" means later, but it's basically what you think it means). The **partial derivative of f w.r.t. x_i**, denoted by $\frac{\partial f}{\partial x_i}$ is given by $\frac{df}{dx_i}$, where x_k for $k \neq i$ is treated as a constant.

Example 2.1. Let $f : \mathbb{R}^3 \to \mathbb{R}$ be given by $f(x, y, z) := x^2y + 2ye^z + x\sin(z)$. Then

$$\frac{\partial f}{\partial x} = 2xy + \sin(z).$$

□

A **multiple integral** of f on an area $T := I_1 \times ... \times I_n$ for $I_j := [a_j, b_j]$ can be evaluated as

$$\int_T f(\mathbf{x})d\mathbf{x} = \int_T f(x_1, ..., x_n)dx_1...dx_n = \int_{a_n}^{b_n} ... \int_{a_1}^{b_1} f(x_1, ..., x_n)dx_1...dx_n.$$

This means you "evaluate the integrals one-by-one." In certain cases you can interchange the integrals, and we'll get to that later.

Example 2.2. Let $f(x, y, z) = x^2y^2 + z$, and $T = [0, 1]^3$. Then

$$\int_T f(x, y, z)dxdydz = \int_0^1 \left(\int_0^1 \left(\int_0^1 x^2y^2 + zdx\right) dy\right) dz = \int_0^1 \left(\int_0^1 \frac{y^2}{3} + zdy\right) dz$$

$$= \int_0^1 \frac{1}{9} + zdz = \frac{1}{9} + \frac{1}{2}.$$

But what if you wanted to evaluate the integral over an area that's not square (or expressible as a product of intervals)? In linear algebra, you learn how to compute the integral over a parallelogram by using linear transformations to transform the unit interval. In other cases, the theorem below may help...

What other thing is important in single-variable calculus? The fundamental theorem of course! The most general form is called **Stokes' theorem**, from which the gradient theorem, divergence theorem, and the classical Green's theorem follows (google these if you are not familiar with them; but really, all you need is Stokes'.)

Theorem 2.3. **(Stokes' theorem)** Let Δ be a "patch of area" (orientable manifold, to be exact), and ω a **differential form**. Then

$$\int_{\partial\Delta} \omega = \int_{\Delta} d\omega,$$

where ∂ denotes the boundary operator and d denotes exterior differentiation.

Proof. Relegated to the differential topology section. ■

The boundary operator ∂ is, in most cases, intuitively what you think it is. If Δ is a triangle in $\mathbb{R}^2$ made by the sides $[a_0, b_0], ..., [a_2, b_2]$, then $\partial\Delta$ is simply the union of the line segments, $\cup_i [a_i b_i]$. A differential form is basically anything that you think can be integrated, and exterior differentiation is like "differentiating a differential form."

This sounds shady, and I don't want to handwave this like usual multivariable calculus usually does (grad, div, and curl are differential forms that you can also exterior differentiate), so we'll postpone Stokes' theorem until our section of differential topology.

Our last topic of this brief introductory section is *differential equations*, equations involving derivatives. For example, the differential equation $f = \frac{df}{dx}$ is solved by $f(x) = e^x$. In general, the **n-th order linear differential equation**

$$a_n f^{(n)}(x) + ...a_1 f'(x) + a_0 f(x) + c = 0$$

is given by e^{zx}, where z in a complex number (use the Euler identity $e^{ix} = \cos x + i \sin x$ to see how you get sin and cos for degree two equations). A **partial differential equation** is a differential equation involving partial derivatives, something we'll get to later.

3 Linear algebra

Typically, linear algebra is an introduction to proof-based mathematics. You should be familiar with the following symbols: $\mathbb{R}, \mathbb{C}, \mathbb{Z}, \mathbb{N}, \mathbb{Q}, \in, \forall, \exists, !, \sim, \{\}, :, ...$

Try to decipher the following statements, and find out if they're true or not: (1) $\forall a \in \mathbb{Q}, \exists b \in \{x : |x| \neq 1\} : ab \notin \mathbb{C}$. (2) $\nexists! z \in \mathbb{C} : az \in \mathbb{N}$.

Okay, let's move on to linear algebra. The main point of linear algebra is to first generalize the notion of "vector," then define operations between vectors so that we can give the space of all vectors more structure. We can then, surprisingly, derive a lot of conclusions from the theory, and applying these conclusions back on $\mathbb{R}^n$ or something more familiar gives us nice results.

Definition 3.1. A **field** F is a set of elements satisfying the following properties:

1. Closure under addition and multiplication: $\forall a, b \in F, a + b, a \cdot b \in F$.
2. Associativity of addition and multiplication: $\forall a, b, c \in F, a+(b+c) = (a+b)+c$ and $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.
3. Commutativity of addition and multiplication: $\forall a, b \in F, a + b = b + a$ and $a \cdot b = b \cdot a$.
4. Existence of additive and multiplicative identities: $\exists 0, 1 \in F, 0 \neq 1 : \forall a \in F, a + 0 = a$ and $a \cdot 1 = a$.
5. Existence of additive inverses and multiplicative inverses: $\forall a \in F, \exists -a \in F : a + (-a) = 0$ and $\forall a \neq 0 \in F, \exists a^{-1} \in F : a \cdot a^{-1} = 1$. Let $a - b := a + (-b)$ and $a/b := a \cot b^{-1}$.

6. Distributivity of multiplication over addition: $\forall a, b, c \in F, a \cdot (b + c) = (a \cdot b) + (a \cdot c)$.

For those of you that know abstract algebra, a field is a type of commutative ring. Questions to ask yourself: is $\mathbb{R}$ a field? How about $\mathbb{Q}$? How about $\mathbb{Z}$?

Given a field F, we can define a **vector space** over it by using the field's elements for scalar multiplication.

Definition 3.2. A **vector space** V over a field F, denoted as $(V, F, +, \cdot)$ (but oftentimes just V), is a set V along with binary operations $+$ and $\cdot$ such that the following hold:

1. Commutativity for $+$ and associativity for $+, \cdot$.
2. Additive identity.
3. Additive inverse.
4. Multiplicative identity.
5. Distributive properties.

(As an exercise, write down what each criteria means.) Elements $v \in V$ are called **vectors**. A vector space over $\mathbb{R}$ is called a **real vector space**, and a vector space over $\mathbb{C}$ is called **complex vector space**.

Vector spaces have many wonderful properties. For example, we can prove uniqueness of additive identities and additive inverses:

Proposition 3.3. A vector space V has a unique additive identity, and every element in a vector space has a unique additive inverse.

Proof. For the first part, let 0 and $0'$ denote additive identities for V. Then $0' = 0' + 0 = 0$ and we have $0' = 0$. Check the second statement yourselves. ■

Definition 3.4. $U \subset V$ is called a **subspace** of V if U is also a vector space.

Since commutativity, associativity, and distributivity are inherited from V, to show that U is a subspace, we only need to verify that it has an additive identity and it's closed under $+$ and $\cdot$. For example, the subspaces of $\mathbb{R}^2$ are $\{0\}$, $\mathbb{R}^2$, and lines through the origin. What about $\mathbb{R}^3$?

Definition 3.5. The **sum** of $U_1, ..., U_m$ is given by $U_1 + ... + U_m = \{u_1 + ... + u_m : u_1 \in U_1, ..., u_m \in U_m\}$. V is a **direct sum** of $U_1, ..., U_m$, e.g. $V = U_1 \oplus ... \oplus U_m$, if each element of V can be written uniquely as a sum $u_1 + ... + u_m$ for $u_i \in U_i$.

Proposition 3.6. If $U_1, ..., U_n$ are subspaces of V, then $V = U_1 \oplus ... \oplus U_n$ iff

1. $V = U_1 + ... + U_n$
2. the only way to write 0 as a sum $u_1 + ... + u_n$, where each $u_i \in U_i$, is by taking all the u_i's equal to 0.

One direction is trivial; for the other, to show that $v = u_1 + ... + u_n$ is a unique representation, suppose we have another one and just take the difference.

Proposition 3.7. If U and W are subspaces of V, $V = U \oplus W$ iff $V = U + W$ and $U \cap W = \{0\}$.

Let's leave this subspace stuff at the back of our minds, and turn to the central object of study in linear algebra. As the name implies, linear algebra is about the study of **linear transformations** on vector spaces, but what does "linear" mean exactly?

Definition 3.8. A **linear combination** of $v_1, ..., v_m \in V$ is a vector of the form $\sum_{i=1}^m a_i v_i$, where $a_i \in F$. $\text{span}(v_1, ..., v_m) := \{\sum_{i=1}^m a_i v_i\}$. If $V = \text{span}(v_1, ..., v_m)$, then $(v_1, ..., v_m)$ **spans** V. If V can be spanned by a list of vectors, then it is **finite dimensional**; otherwise it is **infinite dimensional**.

Definition 3.9. $v_1, ..., v_m \in V$ are **linearly independent** if the only choice of $a_1, ..., a_m \in F$:

$\sum_{i=1}^m a_i v_i = 0$ is $a_1 = ... = a_m = 0$. Else it is **linearly dependent**. A **basis** of V is a list of vectors in V that is linearly independent and spans V.

Is the list $(0,1),(1,0) \in \mathbb{R}^2$ linearly dependent? What about $(0,1,0),(0,0,1),(0,1,1) \in \mathbb{R}^3$? What is a basis for $\mathbb{R}^2$? $\mathbb{R}^3$? $\mathbb{R}^n$? The set of polynomials of degree m? It should be obvious that every finite-dimensional v.s. has a basis. Below we'll state a few properties of span and basis, so that you're more confortable with them.

Lemma 3.10. If $(v_1, ..., v_m)$ is linearly dependent in V, then we can kill off one vector from it and still have the span's be equal.

Theorem 3.11. In a finite-dimensional v.s., the length of every linearly independent list of vectors is less than or equal to the length of every spanning list of vectors. We can get a basis from any spanning list by removing vectors, and we can get a basis from a linearly independent list by adding vectors. Any two bases of a finite-dimensional v.s. V have the same length; the length is called the **dimension** of V.

It should be clear that any subspace has dimension less than or equal to its ambient space. The measure of how much the difference is can be seen from the following:

Theorem 3.12. If U_1, U_2 are subspaces of a finite-dimensional v.s., then $\dim(U_1 + U_2) = \dim U_1 + \dim U_2 - \dim(U_1 \cap U_2)$. Also, if $V = U_1, ..., U_m$ and $\dim V = \dim U_1 + ... + \dim U_m$, then $V = U_1 \oplus ... \oplus U_m$.

Proof. Write down bases for all of them and it's obvious.

Having been introduced to the basic notions of basis and dimension, we proceed to define the central object:

Definition 3.13. A **linear transformation** $T : V \to W$ is a map that satisfies **linearity**, namely $T(au + bv) = aT(u) + bT(v)$ for $a, b \in F, u, v \in V$. The set of all linear maps $T : V \to W$ is denoted as $L(V, W)$.

Is the zero map linear? The identity? Differentiation? Integration?

Note that you can compose linear maps, with nice properties. Namely, if $T \in L(U, V)$ and $S \in L(V, W)$, then we can get $ST \in L(U, W)$ by doing T then S. These compositions are also associative and distributive.

To talk about linear maps and what they do intelligently, we need to know specific things about them:

Definition 3.14. The **kernel**, or **null space**, of $T \in L(V, W)$ is the set $\ker T := \{v \in V : Tv = 0\}$. The **image**, or **range**, of T is $\text{im } T = \{Tv : v \in V\}$.

It should be clear that both the kernel and the range of T are subspaces of V. One of the most important theorems of linear algebra is known as the **rank-nullity** theorem:

Theorem 3.15. If V is finite-dimensional and $T \in L(V, W)$, then $\text{im } T$ is a finite-dimensional subspace of W and

$$\dim V = \dim \ker T + \dim \text{im } T.$$

Proof. Write down bases $u_1, ..., u_m$ for $\ker T$ and extend it to a basis of V by appending $w_1, ..., w_n$. Then any $v \in V$ can be written as $v = \sum_{i=1}^m a_i u_i + \sum_{i=1}^n a_i w_i$. Apply T to get $Tv = \sum_{i=1}^n a_i T(w_i)$. To show that $(Tw_1, ..., Tw_n)$ is linearly independent, suppose $\exists b_i \in F : \sum_{i=1}^n b_i T w_i = 0$. Then by linearity of T, $T(\sum_{i=1}^n b_i w_i) = 0 \implies \sum_{i=1}^n b_i w_i \in \ker T$. Since $u_1, ..., u_m$ spans $\ker T$, we have $\sum b_i w_i = \sum c_i u_i$, which shows that all the b's are 0. This shows that $Tw_1, ..., Tw_n$ is linearly independent, so $\text{im } T$ is of dimension n as desired. ■

We can also note the general structure of linear maps. That is, are they one-to-one, onto, or both?

If they are both, then it happens that they are invertible, e.g. $\exists S : ST = TS = Id$. To make this notion more precise, we introduce the following definitions (which are, by the way, very important):

Theorem 3.16. T is **injective** if $u, v \in V$ and $Tu = Tv$ implies $u = v$. T is **surjective** if it's range is W. T is **bijective** if it's both injective and surjective.

Convince yourself that T is injective iff $\ker T = \{0\}$. This is very important outside of linear algebra, and we will use this to a large extent to show that linear maps are injective. It should also be obvious that T can only be injective when $\dim V = \dim W$, and surjective if $\dim V \geq \dim W$. A linear map is called **invertible** if there exists an inverse of T, which happens iff T is a bijection (prove this). Any two vector spaces linked by an invertible map is called **isomorphism**, and the **vector space isomorphism** is given by the bijective map.

One applicable thing about linear maps is that is allows for the **complexification** of a real vector space, which is important in differential geometry:

Theorem 3.17. A real vector space with complex structure is a complex vector space.

Proof. Let V be a real vector space with $\dim V = 2n$ and $J : V \to V$ and $J^2 = -Id$. Define multiplication $\mathbb{C} \times V \to V$ by complex scalars by $(x + iy)v = xv + yJv$ for $v \in V$ and $x, y \in \mathbb{R}$. Then, with respect to this multiplication, (V, J) defines a complex vector space of dimension n (check this). ■

Anyways, remember what matrices are? If you only thought of them as numbers in an array, you'll be surprised to see that they encode the action of a linear transformation.

Definition 3.18. Let $T \in L(V, W)$, and suppose the $v_1, ..., v_n$ is a basis of V and $w_1, ..., w_n$ is a basis of W. Then for each $k = 1, ..., n$, we can write $Tv_k = a_{1k}w_1 + ... + a_{mk}w_m$ where $a_jk \in F$ for $j = 1, ..., m$; these scalars complete determine T. The $m \times n$ matrix formed by the a's is called the **matrix of *T*** w.r.t. $(v_1, ..., v_n)$ and $(w_1, ..., w_n)$ and is denoted $M(T, (v_1, ..., v_n), (w_1, ..., w_n))$, or just $M(T)$.

With regards to this definition, everything that you know about matrices comes from the properties of linear maps (or maybe it should be the other way around). In particular, $M(T+S) = M(T) + M(S)$, $M(cT) = cM(T)$, $M(Tv) = M(T)M(v)$, etc. I will also assume that you know what diagonal matricces, upper-triangular matrices, etc. are (they are exactly what they sound like). We will return to discuss the trace and determinant of a matrix later, but first we'll divert our attention to eigenstuff. Eigenstuff help us, in a way, understand the structure of operators. We'll state what they are and a few theorems about the without proof (honestly, I get a bit bored while talking about them).

Definition 3.19. For $T \in L(V, V) = L(V)$ and U a subspace of V, U is **invariant** under T if $u \in U \implies Tu \in U$. On a separate note, $\lambda \in F$ is called an **eigenvalue** of $T \in L(V)$ if $\exists u \in V : Tu = \lambda u$, where u is the **eigenvector**.

Theorem 3.20. Let $T \in L(V)$, and suppose $\lambda_1, ..., \lambda_m$ are distinct eigenvalues of T and $v_1, ..., v_m$ are corresponding nonzero eigenvectors. Then $(v_1, ..., v_m)$ is linearly independent. Also, each operator on V has at most $\dim V$ distinct eigenvalues.

Theorem 3.21. Every operator on a finite-dimensional, nonzero, real v.s. has an invariant subspace of dimension 1 or 2. Every operator on an odd-dimensional real v.s. has an eigenvalue.

I'll assume that most of you have dealt with eigenstuff before, and I don't want to make the linear algebra section too long, so I will skip to **inner products**. In a vector space V, we can define what length and angles mean, which helps us understand the structure of vector spaces more.

Definition 3.22. An **inner product** on V is $\langle \cdot, \cdot \rangle : V \times V \to F$:

1. $\langle v, v \rangle \geq 0$ for all $v \in V$.

2. $\langle v, v\rangle = 0$ iff $v = 0$.
3. $\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$ for all $u, v, w \in V$.
4. $\langle av, w\rangle = a\langle v, w\rangle$ for all $a \in F, v, w \in V$.
5. $\langle v, w\rangle = \overline{\langle w, v\rangle}$ for all $v, w \in V$ (or $\langle v, w\rangle = \langle w, v\rangle$ if we're dealing with a real v.s. V.

Also, an **inner product space** is a v.s. V endowed with an inner product. The **norm** of $v \in V$, denoted $||v||$, is given by $||v|| = \sqrt{\langle v, v\rangle}$. Two vectors are **orthogonal** if $\langle u, v\rangle = 0$.

Theorem 3.23. (Pythagorean theorem) If $u, v \in V$ are orthogonal, then $||u + v||^2 = ||u||^2 + ||v||^2$.
Proof. Trivial. ■

Theorem 3.24. (Cauchy-Schwarz inequality) If $u, v \in V$, then $|\langle u, v\rangle| \leq ||u||||v||$, and equality holds iff one of u, v is a scalar multiple of the other.
Proof. Assume $v \neq 0$, and consider the orthogonal decomposition

$$u = \frac{\langle u, v\rangle}{||v||^2} v + w,$$

where w is orthogonal to v. Using the Pythagorean theorem, we have

$$||u||^2 = \left|\left| \frac{\langle u, v\rangle}{||v||^2} v \right|\right|^2 + ||w||^2 = \frac{|\langle u, v\rangle|^2}{||v||^2} + ||w||^2 \geq \frac{|\langle u, v\rangle|^2}{||v||^2}.$$

That equality holds iff one of u, v is a scalar multiple of the other is obvious. ■

Theorem 3.25. (Triangle inequality) If $u, v \in V$, then $||u + v|| \leq ||u|| + ||v||$, and equality holds iff one of u, v is a nonnegative multiple of the other.
Proof. Use Cauchy-Schwarz. ■

These last two inequalities are among the most important in all of mathematics, so remember them well! The notion of orthogonality is also very important, since decomposing a vector into orthogonal bases is easy.

Definition 3.26. A list of vectors is **orthonormal** if the vectors are pairwise orthogonal and each vector has norm 1. An **orthonormal basis** is a basis that's, you guessed it, orthonormal.

It should be clear that every orthonormal list of vectors is also linearly independent. Also, if $(e_1, ..., e_n)$ is an orthonormal basis of V, then the orthogonal decomposition of $v \in V$ is given by $v = \langle v, e_1\rangle e_1 + ... + \langle v, e_n\rangle$, and $||v||^2 = |\langle v, e_1\rangle|^2 + ...|\langle v, e_n\rangle|^2$. We can turn any linearly independent list into an orthonormal list with the same span as the original list using the **Gram-Schmidt procedure**: namely, let

$$e_j = \frac{v_j - \langle v_j\rangle e_1 - ... - \langle v_j, e_{j-1}\rangle e_{j-1}}{||v_j - \langle v_j, e_1\rangle e_1 - ... - \langle v_j, e_{j-1}\rangle e_{j-1}||},$$

and convince yourself that $(e_1, ..., e_j)$ is an orthonormal list. It follows that *every finite-dimensional inner product space has an orthonormal basis.*

We can also define the **orthogonal subspace** to a subspace U as $U^\perp := \{v \in V : \langle v, u\rangle = 0 \;\forall u \in U\}$. Notice that $(U^\perp)^\perp = U$, and if U is a subspace of V, then $V = U \oplus U^\perp$. Let $\pi_U : V \to U$ denote orthogonal projection w.r.t. the subspace U; then $\pi_U(v)$ is the closest point in U to v, and is given by $\pi_U(v) = \langle v, e_1\rangle e_1 + ... + \langle v, e_m\rangle e_m$. This fact has many applications in minimization problems.

Example 3.27. [Axler 114] Let's find a polynomial u with real coefficients and degree at most 5 on $[-\pi, \pi]$ that approximates $\sin x$ as well as possible, in the sense that $\int_{-\pi}^{\pi} |\sin x - u(x)|^2 dx$ is minimized. Let $C[-\pi, \pi]$ denote the real v.s. of continuous real-valued functions on $[-\pi, \pi]$ with inner product

$$\langle f, g\rangle = \int_{-\pi}^{\pi} f(x)g(x)dx,$$

let $v \in C[-\pi, \pi]$ be $v(x) = \sin x$, and let U denote the subspace of $C[-\pi, \pi]$ consisting of polynomials with real coefficients and degree ≤ 5. Then we should find $u \in U : \min ||v - u||$ is achieved. To do this, use the Gram-Schmidt procedure to get an orthonormal basis, then use the inner product given by $\langle f, g\rangle$ above to compute $\pi_U(v)$. The answer is $\pi_U(v) = 0.987862x - 0.155271x^3 + 0.00564312x^5$.

Orthogonal subspaces show up everywhere (like symplectic geometry), so keep them in mind! For the rest of this section, we deal primarily with two concepts: the notion of a functional, and properties of matrices.

Definition 3.28. A **linear functional** on V is a linear map from V to F. The space of all linear functionals is called the **dual space** of V and is denoted V^*. For $T \in L(V, W)$, the adjoint of T. denoted T^*, is the map from W to V such that for $w \in W$, T^*w is the unique vector in V such that $\langle Tv, w\rangle = \langle v, T^*w\rangle$ for all $\in V$. $T \in L(V)$ is **self-adjoint** if $T = T^*$.

For more information about the adjoint, google it. You should also try to prove properties about the adjoint on your own (I want to move on to other material), but one important theorem involving adjoints that I will mention is the **spectral theorem**, stated here without proof:

Theorem 3.29. (Real spectral theorem) Let V be a real inner-product space and $T \in L(V)$. Then V has an orthonormal basis consisting of eigenvectors of T iff T is self-adjoint.

Let's return to matrices, which will close out this section. Recall that, while defining matrices, we had to specify a basis. The following change-of-basis theorem assures us that we can switch between bases:

Theorem 3.30. Suppose $T \in L(V)$ and $(u_1, ..., u_n)$ and $(v_1, ..., v_n)$ are bases of V. If $A = M(I, (u_1, ..., u_n), (v_1, ..., v_n))$, then

$$M(T, (u_1, ..., u_n)) = A^{-1}M(T, (v_1, ..., v_n))A.$$

With this and the fact that trace(AB) = trace(BA), you should prove that the **trace** of any matrix, defined as the sum of the diagonal entries w.r.t. any choice of basis, is well-defined. We will not get into the properties of the trace here, and instead talk about determinants.

Definition 3.31. The **determinant** of a matrix is an alternating multilinear function of the columns, and for the $n \times n$ matrix $A = (a_{i,j})$ is given by

$$\det A = \sum_{(m_1,...,m_n) \in \text{perm } n} \sigma(m_1, ..., m_n) a_{m1,1} ... a_{m_n,n}.$$

For our purposes, we will define the deteminant of a linear operator to be the determinant of its matrix representation. Why we introduced the notion of determinant can be seen from the following theorem:

Theorem 3.31. A linear map is invertible iff its determinant is zero.
Proof. Left to the reader.

The determinant transforms volumes in $\mathbb{R}^n$:

Theorem 3.32. If $T \in L(\mathbb{R}^n)$, then volume $T(\Omega) = |\det T|$(volume Ω) for $\Omega \subset \mathbb{R}^n$.

This wraps up our introductory section on linear algebra. Some cool things about linear algebra that we did not get to (and you should look up): applications of linear algebra to Google's PageRank algorithm, the canonical isomorphism between a vector space and its double dual, the characteristic polynomial, Jordan form,.... Let's move on to more fun stuff!

4 Real analysis

The tools for linear algebra we developed carry nicely over to $\mathbb{R}^n$, and the study of real-analytic functions is called **real analysis**. To get started, we need to know the *topology* of $\mathbb{R}^n$. The following terms that I will use are basically what you think they mean, so in the interest of space I will just throw them at you:

The *topology of* $\mathbb{R}^n$, $\lim_{x\to a} f(x) = L$, *continuity of functions* $f : \mathbb{R}^n \to \mathbb{R}^n$, *countable* subsets of $\mathbb{R}^n$ ($A \subseteq \mathbb{R}^n$ is countable if it is finite or countably infinite.), the ball of radius r around a given by $B_r(a) := \{x \in \mathbb{R}^n | ||x-a|| < r\}$, *open* sets of $\mathbb{R}^n$ ($A \subseteq \mathbb{R}^n$ is open iff all points of A are interior points), $B \subseteq \mathbb{R}^n$ is *closed* if $\mathbb{R}^n - B$ is open or the set of *accumulation points* $acc(B) \subseteq B$.

In other words, we need to what sets are open in a space ($\mathbb{R}^n$), which defines a topology. The other definitions involve something called *epsilons* and *deltas*, which pervade real analysis. Epsilons are a quantity so small that it's identically zero.

Definition 4.1. A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is said to be **continuous** at $x = c$ if for $\delta, \epsilon > 0$ and $|x - c| < \delta$, $|f(x) - f(c)| < \epsilon$.

Definition 4.2. Given an inner product space X, a subset $A \subseteq X$ is **open** in X if it can be covered by open balls in X. $A \subseteq X$ is **closed** if its complement $X - A$ is open, or equivalently, if it contains the set of all its **accumulation points**.

Definition 4.3. A subset $A \subseteq X$ is **compact** if every sequence has a convergent subsequence. Equivalently, if we take $X = \mathbb{R}^n$, then A is compact iff it is closed and **bounded**, meaning that it can be contained in a ball of finite radius.

The equivalence of the definitions of compact subset follows from a very useful theorem in elementary real analysis, namely the **Bolzano-Weierstrass theorem:**

Theorem 4.4. (Bolzano-Weierstrass Theorem) If $A \subseteq \mathbb{R}^n$ contains infinitely many points and is bounded, then $\exists$ an *accumulation point* of A in $\mathbb{R}^n$.

Thus, TFAE criteria for compactness for $A \subseteq \mathbb{R}^n$.

1. A is compact.
2. A is closed and bounded.
3. Every infinite subset of A has an accumulation point in A.

Now consider extending these notions into generality, that is, into *metric spaces.* Metric spaces provide the backbone for doing calculus:

Definition 4.5. Let X be a set. X is a *metric space* if there exists a metric. $d : X \times X \to \mathbb{R}$ is a metric if it satisfies the following:

1. $d(p, q) > 0$ if $p \neq q$, and $d(p, p) = 0$.
2. $d(p, q) = d(q, p)$ $\forall p, q \in X$.
3. $d(p, q) \leq d(p, r) + d(r, q)$ $\forall p, q, r \in X$ (this is the *Triangle Inequality*).
4. (on $\mathbb{R}^n$) $d(x, y) = ||x - y|| = \sqrt{\langle x - y, x - y\rangle}$.

With this vocabulary of the metric, we can talk about *neighborhoods* (open balls) $N_r(p) = \{x \in X | d(x, p) < r\}$, $B_r(a)$ or $B_a(r)$, accumulation points, isolated points, interior points, open sets, closed sets, compactness, etc. Let's define more vocabulary:

Theorem 4.6. $E \subseteq X$ is *dense* if every point of X is a limit point of E or is a point of E (or both). $\mathbb{Q} \subseteq \mathbb{R}$ is dense, but $\mathbb{Z} \subseteq \mathbb{R}$ is not.

Theorem 4.7. If X is a metric space and $E \subseteq X$, then

1. $\bar{E}$ is closed.
2. $E = \bar{E}$ iff E is closed.
3. $\bar{E} \subseteq F$ for all $F \subseteq X$ closed s.t. $E \subseteq F$. $\bar{E}$ is the *smallest closed subset containing* E.

Definition 4.8. We also went over the notion of *least upper bound of* $\mathbb{R}$: for $A \subseteq \mathbb{R}$ a nonempty subset, if A is bounded then $\exists$ least upper bound called $\sup A$ (supremum). (Dedekind completeness). $\inf A$ is called the greatest lower bound of A.

For X a metric space and $Y \subseteq X$ also a metric space, we consider $E \subseteq Y \subseteq X$. We have two metric spaces potentially conflicting each other: is $E \subseteq Y$ open? Is $E \subseteq X$ open? Consider $X = \mathbb{R}^2$, $Y = [0, 1]$, $E = [0, 1/2)$. E and Y are not open in X. Convince yourself that E is open in Y by going over the definition of *neighborhood*. $E \subseteq Y$ is open in Y iff every point of E is an interior point relative to Y.

Definition 4.9. Let's introduce the general definition of *compactness*. A subset $K \subseteq X$ is compact if every open cover of K admits a finite subcover.

Theorem 4.10. Suppose $K \subseteq Y \subseteq X$. K is compact relative to X iff K is compact relative to Y.

Proof. Suppose K is a compact set relative to X. We must show that any open (relative to Y) cover of K admits a finite subcover. Take an open (relative Y) cover of K $\{V_\alpha\}$, V_α open relative to Y, $K \subseteq \cup_\alpha V_\alpha$ from the theorem above. There exists open sets relative to X, which we denote as G_α, such that $V_\alpha = Y \cap G_\alpha$. Then $K \subseteq \cup_\alpha (G_\alpha \cap Y) \subseteq \cup_\alpha G_\alpha$. K is compact relative to $X \Rightarrow \exists$ finitely many G_α's such that $K \subseteq G_{\alpha_1} \cup ... \cup G_{\alpha_n} \Rightarrow Y \cap K \subseteq (G_{\alpha_1} \cup ... \cup G_{\alpha_n}) \cap Y \Rightarrow K \cap Y \subseteq (G_{\alpha_1} \cap Y) \cup ... \cup (G_{\alpha_n} \cap Y)$. So this is finite subcover. ∎

Theorem 4.11. Compact subsets of metric spaces are closed. Closed subsets of compact sets are compact. If $\{K_\alpha\}$ is a collection of compact subsets of a metric space X such that $\cap$ of every finite subcollection is nonempty, then $\cap_\alpha K_\alpha \neq \emptyset$. This is like the *Cantor Intersection Theorem* for finite metric spaces. If E is an inf. subset of a compact set K, then E has a limit point in K. This is similar to the Bolzano-Weierstrass Theorem.

We now turn our attention to the *Cantor Set*, which we construct as follows. Consider the interval $E_0 = [0, 1]$, and remove the middle-thirds open interval $(1/3, 2/3)$ to get E_1. Repeat this process for each of the resulting intervals $E_1, E_2, ...$ ad infinitum. This is the *dyadic Cantor Set* $C \subseteq [0, 1]$. Also, $...E_2 \subseteq E_1 \subseteq E_0$, and $\cap_{k=1}^{\infty} E_k = C$. There's many interesting properties about this set: it's closed and bounded, compact, "perfect," contains no interior points, and is totally disconnected....

Anyways, to go back to the main point, we were talking about topology. Topology examines shapes as mathematical objects. For X a space and $\mathcal{T}$ consisting of all open sets of X, we can define $(X, \mathcal{T})$ to be a topological space.

To talk about shapes, we wanted to talk about notions of *bounded*, *countable*, *compact*, and *connected*. Let's move on to *connectedness*. You already have an intuition for what "connected" means, so how do we formalize it? In $\mathbb{R}$, connected subsets include (a, b), or $(-\infty, \infty)$.

Definition 4.12. $E \subseteq X$ is **connected** iff E is not disconnected. For $A, B \subseteq X$, A and B are **separated** if $A \cap \bar{B} = \emptyset$ and $\bar{A} \cap B = \emptyset$. The empty set is separated from every set. We also cannot separate anything from $\mathbb{Q}$ in $\mathbb{R}$ since $\bar{\mathbb{Q}} = \mathbb{R}$. A set $E \subseteq X$ is **disconnected** if $E = A \cup B$, where A and B are nonempty and separated.

Is the union of connected subsets connected? Not necessarily. Is the intersection of connected subsets connected? Not necessarily: consider two "banana-shaped" objects that intersect at two points. Let's move onto connected subsets of $\mathbb{R}$. The intervals should be connected:

Theorem 4.13. $E \subseteq \mathbb{R}$ is connected iff it satisfies the following property: if $x, y \in E$ and $x < z < y$, then $z \in E$. A set like $(a, c) \cup (c, b)$ violates this.

Proof. (Quick sketch) First we show that if E does not satisfy the above property, then E is disconnected. Suppose we have $x, y \in E$; we know $\exists z : x < z < y$ and $z \notin E$. Let's consider the sets $(-\infty, z) \cap E = A \cap E$ and $(z, \infty) \cap E = B \cap E$. Then $E = A_z \cup B_z$. Are these guys separated? Yes (why?). Are they nonempty? Yes (why?). So E is disconnected.

Now we want to show that E disconnected implies that the properties above are violated. By assumtion, $\exists$ nonempty, separated sets A and B such that $E = A \cup B$. How do we find x, y, z such that the properties are violated? Let $z = \sup([x, y] \cap A)$. So $x \leq z < y$. $z \in \bar{A}$ or $z \notin B$. If $z \notin A$, we have a strict inequality above and we're done. If $z \in A$, then $z \notin \bar{B}$ because A and B are separated, which implies that $\exists r > 0 : B_r(z)$ contains no points of B. Choose $z_1 : z_1 > z$, $z_1 \notin \bar{B}$. Claim: $x < z_1 < y$, $z_1 \notin E$ violates the assumption. ∎

Let's not get bogged down by metric spaces and turn to sequences and series, which as you know from AP Calc BC are very important in calculus. We want to generalize them:

Definition 4.14. A **sequence** $\{p_n\}$ is a mapping seq: $\mathbb{N} \to X : n \in \mathbb{N} \mapsto p_n \in X$.

Definition 4.15. A sequence $\{p_n\}$ in a metric space X is said to **converge** if $\exists p \in X$ with the following property: $\forall \epsilon > 0, \exists N > 0 : n \geq N \Rightarrow d(p_n, p) < \epsilon$. If a sequence does not converge, then it **diverges**.

Theorem 4.16. Let $\{p_n\}$ be a sequence in a metric space X. Then

1. $p_n \to p$ iff every neighborhood of p contains all but finitely many p_n's.
2. If $p, p' \in X$ nd $p_n \to p$, $p_n \to p'$, then $p = p'$.
3. If p_n converges, then $\{p_n\} \subseteq X$ is bounded.
4. If $E \subseteq X$ and $p \in acc(E)$, then $\exists$ a sequence $\{p_n\} \subseteq E$ s.t. $p_n \to p$.

Proof. (Sketch) (1) Suppose $p_n \to p$. Let U be a (open) neighborhood of p. $\exists \epsilon > 0 : B_\epsilon(p) = \{x \in X | d(x, p) < \epsilon\} \subseteq U$. Because $p_n \to p$, $\exists N : \forall n \geq N, d(p_n, p) < \epsilon$. Thus the neighborhood does not contain $\{p_N, p_{N+1}, ...\}$.

Now suppose every neighborhood of p contains a tail. Show $\{p_n\} \to p$. Let $\epsilon > 0$ be given. By hypothesis, $B_\epsilon(p)$ contains a tail, that is $\exists N : \forall n \geq N, p_n \in B_\epsilon(p)$, i.e. $\forall n \geq N$, $d(p_n, p) < \epsilon$.

(2) Given $\epsilon > 0$, $\exists N : \forall n \geq N, d(p, p_n) < \epsilon/2$, $\exists M : \forall n \geq M, d(p_n, p') < \epsilon$. Thus $\forall n \geq \max(N, M)$, $d(p, p') \leq d(p_n, p) + d(p_n, p')$ by the Triangle Inequality.

(3) We have $p_n \to p$. Choose $r > 0$. By part (1), $B_r(p)$ contains a tail.

(4) Consider $B_{r_n}(p)$, with $r_n := \{1/n\}$. ∎

Definition 4.17. Given a sequence $\{p_n\}$, $n \mapsto p_n \in X$, $n \in \mathbb{N}$, let $\{n_k\}$ be a sequence of positive integers s.t. $n_1 < n_2 < n_3 < ....$ Then the 'new' sequence $\{p_{n_i}\}$ is called a **subsequence** of $\{p_n\}$. If a subsequence $\{p_{n_i}\}$ converges, then its limit is called a *subsequential limit.*

Example 4.18. Consider the sequence defined by $p_n = 1/n$, $n = 1, 2, ...$ The first few terms of the sequence is $1, 1/2, 1/3, ...$ Now consider the subsequence where $n = 1, 3, 5, ...$ The first few terms of the subsequence is $1, 1/3, 1/5, ...$

Proposition 4.19. $p_n \to p$ iff every subsequence converges to p.

Proof. Suppose $p_n \to p$, and consider a subsequence $\{p_{n_i}\}$. Since $p_n \to p, \forall \epsilon > 0, \exists N : d(p_n, p) < \epsilon$. Since $\{p_{n_i}\}$ is a subsequence, $\exists M : \forall i \geq M, d(p_{n_i}, p) < \epsilon$. ∎

Note that, if $\{p_n\}$ is a sequence in a compact metric space, then $\exists$ a convergent subsequence. Also, every bounded sequence in $\mathbb{R}^n$ contains a converging subsequence.

Theorem 4.20. The set of all subsequential limits of a given $\{p_n\}$ forms a closed subset of X. Also, suppose $p_n \to p$ in X. Then $\forall \epsilon > 0, \exists N : \forall n, m \geq N$, $d(p_n, p_m) < \epsilon$.

Definition 4.21. A sequence $\{p_n\} \in X$ is **Cauchy** if $\forall \epsilon > 0, \exists N : \forall n, m \geq N, d(p_n, p_m) < \epsilon$. (Why can't we replace m with $n+1$? Consider the harmonic series.) All convergent sequences are Cauchy (why?).

Consider the sequence $1 + \frac{1}{n!}$ for $n = 1, 2, \ldots$. This should convince you that $\mathbb{Q}$ admits Cauchy sequences with no limits.

Theorem 4.22. The following are true:

1. In any metric space X, convergence implies Cauchy.
2. In any compact X, and if $\{p_n\}$ is Cauchy, then p_n converges in X.
3. In $\mathbb{R}^n$, every Cauchy sequence converges.

Definition 4.23. The metric space X is said to be *complete* if every Cauchy sequence converges. Examples are $\mathbb{R}^n$, compact metric spaces; non-examples are $\mathbb{Q}$, $\mathbb{R} - \mathbb{Q}$,....

Now let's talk about *upper and lower limits* (which we call *limit superior* and *limit inferior*). Let $\{s_n\}$ be a sequence of real numbers, and let $E \subseteq \mathbb{R} \cup \{\pm\infty\}$ be a set of all subsequential limits of $\{s_n\}$.

Definition 4.24. The upper limit of $\{s_n\}$, $s^* := \sup E$, and the lower limit $s_* := \inf E$. A sequence $\{s_n\}$ converges iff $s^* = s_* = s$.

Definition 4.25. $s_n = \sum_{k=1}^{n} a_k$ is a *series*, and $\sum a_k$ *converges* if $\{s_n\}$ converges.

Cauchy criterion for convergence of series. $\{s_n\}$ converges iff $\forall \epsilon > 0, \exists N : \forall n, m \geq N, d(S_n, S_m) = |s_n - s_m| < \epsilon$. It $\{s_n\}$ converges, it is necessary condition that, taking $m = n+1$, $|a_{n+1}| < \epsilon$ and $\lim_{n\to\infty} a_n = 0$. This is not sufficient; an example would be $\sum 1/n$.

Theorem 4.26. (Partial sums convergence) A series of nonnegative terms converges iff sequence of partial sums $\{s_n\}$ is bounded.

Theorem 4.27. (Telescoping convergence) Let $\{a_n\}, \{b_N\}$ be two sequences of real numbers s.t. $a_n = b_{n+1} - b_n$ (sequence of successive differences). Then $\sum a_n$ converges iff $\lim_{n\to\infty} b_n$ exists, in which case $\sum a_n = \lim_{n\to\infty} b_n - b_1$.

Note that we can also prove the comparison test, the power test, the root test, the ratio test, etc. from calculus. Something that we can do with series is to define some very useful numbers:

Definition 4.28. $e := \sum_{n=0}^{\infty} 1/n!$. Prove that this series converges: $s_n < 1+1+1/2+1/4+\ldots+1/2^{n-1} < 3$, so $s_n \to 3$ and $\{s_n\}$ converges.

How fast does s_n converge? $e - s_n = \frac{1}{(n+1)!} + \frac{1}{(n+2)!} + \ldots < \frac{1}{(n+1)!}(1 + \frac{1}{n+1} + \frac{1}{(n+1)^2} + \ldots) = \frac{1}{n!n}$. So it converges really fast. We also prove something that you all should know how to prove:

Theorem 4.29. e is irrational.

Proof. Suppose not; let $e = p/q$ for p and q reduced. $0 < e - s_q < \frac{1}{q!q} \Rightarrow 0 < q!(e - s_q) < 1/q$. $q!(e - s_q) = q!e - q!s_q = (q-1)!p - q!(1 + 1 + 1/2! + \ldots + 1/q!)$. This is a difference of integers, but there are no integers between 0 and $1/q$, so we have a contradiction. *Remark:* It's rather hard to prove that π is irrational, or that e is transcendental.

Consider $\sum_{n=0}^{\infty} c_n(z - z_0)^n$, $c_n, z_1, z_0 \in \mathbb{C}$. For which $z \in \mathbb{C}$ does this series converge, and what does it converge to? There exists a disk in $\mathbb{C}$, called the disk of convergence with radius R, in which the series converges.

Theorem 4.30. Given the power series $\sum_{n=0}^{\infty} a(z-z_0)^n$, set $\alpha = \lim_{n\to\infty} \sup |c_n|^{1/n}$ and set $R = \frac{1}{\alpha}$. Then series converges if $|z-z_0| < R$ and diverges if $|z-z_0| > R$.

Proof. Using the root test, we have $\lim_{n\to\infty} \sup |a_n|^{1/n} = \lim\sup(|c_n||z-z_0|^n)^{1/n} = |z-z_0| \lim_{n\to\infty} |c_n|^{1/n}$. Setting this as α, by the root test we know that $|z-z_0| < 1/\alpha = R$ implies convergence, while $>$ implies divergence.

Example 4.31.

1. $\sum_{n=1}^{\infty} nz^n > \infty$ at any point of the unit circle: if z is on the unit circle, then by definition $|z| = 1$, but then $|nz^n| = n$ and $n \to \infty$. So $\sum_{n=1}^{\infty} nz^n > \infty$ at any point of the unit circle because the terms do not approach 0 (in particular, it fails the divergence test).
2. $\sum_{n=1}^{\infty} \frac{z^n}{n^2} < \infty$ at every point of the unit circle: we show that the series converges by showing that it converges absolutely. As above, $|z| = 1$, so we can use the comparison test to write $\sum_{n=1}^{\infty} \frac{|z^n|}{n^2} \le \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty$. Thus $\sum_{n=1}^{\infty} \frac{z^n}{n^2} < \infty$, as desired.
3. $\sum_{n=1}^{\infty} \frac{z^n}{n}$ converges conditionally at every point of the unit circle, except at $z = 1$. (Try proving this using summation by parts.) This goes into *alternating series*.

Definition 4.32. $\sum a_n$ converges absolutely if $\sum |a_n|$ converges. Prove, using the comparison test, that if $\sum a_n$ converges absolutely, then $\sum a_n$ converges.

Theorem 4.33. (Alternating Series Test, not in Rudin) Let $\sum_{n=1}^{\infty}(-1)^{n+1}a_n$, $a_n \ge 0$, and $\{a_n\}$ decreasing. Then $\sum_{n=1}^{\infty}(-1)^{n+1}a_n$ converges.

Proof. Consider partial sums $s_N = \sum_{n=1}^{N} a_n$. All $s_n \ge 0$ by induction, and we want to show that $s_n \to 0$ as $n \to \infty$. Consider $s_{2n}, s_{2n+1}, \ldots$. $s_{2n+2} \ge s_{2n}$, so the even partial sums are increasing. By a similar argument, the odd partial sums are decreasing. Then the even partials are bounded above by a, and odd partials are bounded below by 0. We must show that both subsequences converge to the same number, but the even partials tend towards 0 because $a_n \ge 0$ and a_n is decreasing. ∎

Let's move on to continuity. f is continuous at a if $\lim_{x\to a} f(x) = f(a)$, and $\lim_{x\to a} f(x) = L$ means $\forall \epsilon > 0, \exists \delta > 0 : 0 < |x-a| < \delta \Rightarrow |f(x) - L| < \epsilon$. How do we generalize this to metric spaces? a should be an accumulation point of E in order to talk about limits, and absolute values should be replaced by d_x and d_y.

Definition 4.34. (Uniformly continuous) Intuitively, this means that we don't have to let δ vary depending on our point in the traditional definition of continuity. As an example, $1/x$ on $(0,1)$ is not uniformly continuous, but $\sqrt{x}$ is.

Theorem 4.35. Let $f : X \to Y$ be continuous and X compact. Then f is uniformly continuous on X. Also, if $f : X \to Y$, X and Y metric spaces, X compact, f continuous and bijective, then f is a homeomorphism.

There are many theorems about uniform convergence that we won't get to, but be assured, they are very useful in all of mathematics. We now have all the machinery to define what a **derivative** is.

Definition 4.36. f is **differentiable** if f' exists, and f' exists if the limit

Theorem 4.37. Let f be defined on $[a,b]$. If f is differentiable at $x \in [a,b]$, then f is continuous there. Also, suppose f and g are defined on $[a,b]$ and suppose they are both differentiable at $x \in (a,b)$. Then $(f+g), (f \circ g)$, and f/g are differentiable at x.

Theorem 4.38. (Chain rule) Suppose f is continuous on $[a,b]$ and $f'(x)$ exists at some $x \in [a,b]$. Also suppose g defined on some interval $I \subseteq \mathbb{R}$ and suppose I contains $f([a,b])$. Suppose that g is differentiable at $f(x) \in I$. Then $h(t) = g(f(t))$, $t \in [a,b]$ is differentiable at x and $h'(x) = g'(f(x))f'(x)$.

Definition 4.39. $f : [a,b] \to \mathbb{R}$ has a *local max* at $x \in [a,b]$ if $\exists r > 0$ s.t. for all $y \in B_r(x)$, $f(y) \leq f(x)$. Prove that, if f has a local max at a point $x \in (a,b)$, and if $f'(x)$ exists, then necessarily $f'(x) = 0$.

Theorem 4.40. Let f and g be continuous, real-valued, defined on $[a,b]$. Suppose f and g are differentiable on (a,b). Then $\exists$ a point $x \in (a,b)$ at which

$$(f(b) - f(a))g'(x) = (g(b) - g(a))f'(x).$$

The standard mean value theorem is a corollary.

Theorem 4.41. If f is differentiable on (a,b) then

1. $f'(x) \geq 0 \ \forall x \in (a,b) \Rightarrow f$ is monotonically increasing.
2. $f'(x) \leq 0 \ \forall x \in (a,b) \Rightarrow f$ is monotonically decreasing.
3. $f'(x) = 0 \ \forall x \in (a,b) \Rightarrow f$ is constant.

Proof. Use the MVT. ■

Theorem 4.42. If $f : [a,b] \to \mathbb{R}$ differentiable and $x \in (f'(a), f'(b))$, then $\exists x \in (a,b)$ s.t. $f'(x) = x$. (IVT for derivatives) Keep in mind that f' may not be continuous.

Theorem 4.43. (L'Hopital's rule) Suppose f and g are real-valued functions differentiable on (a,b) and suppose $g'(x) = 0$ for all $x \in (a,b)$, $-\infty \leq a < b \leq \infty$. Suppose $\frac{f'(x)}{g'(x)} \to A \in \mathbb{R} \cup \{\pm\infty\}$ as $x \to a$.

Case I: If $f(x) \to 0$ and $g(x) \to 0$ as $x \to a$ OR

Case II: If $g(x) \to \infty$ as $x \to a$

Then $\frac{f(x)}{g(x)} \to A$ as $x \to a$.

Proof. Suppose the hypothesis in case I holds.

Subcase I: Suppose $A \neq \infty \Rightarrow A \in \mathbb{R}$ or $A = -\infty$. Choose $q \in \mathbb{R}$ s.t. $A < q$ and find $r \in \mathbb{R}$ s.t. $A < r < q$. By hypothesis (b/c $f'(x)/g'(x) \to A$ as $x \to a$), $\exists$ a neighborhood of a (i.e. $\exists c \in (a,b)$) s.t. $\forall x \in (a,c)$, $f'(x)/g'(x) < r$. We want to show $\frac{f(x)}{g(x)} \to A$, so to do this we have to find an appropriate neighborhood of a. The MVT says $\exists t \in (x,y)$:

$$(f(x) - f(y))g'(t) = (g(x) - g(y))f'(t) \Rightarrow \frac{f(x) - f(t)}{g(x) - g(y)} = \frac{f'(t)}{g'(t)} < r$$

for $t \in (x,y)$. Taking $x \to a$ gives $\frac{f(y)}{g(y)} \leq r$.

Theorem 4.44. (Taylor's theorem) Suppose f is a real valued function of $[a,b]$: $f : [a,b] \to \mathbb{R}$. Suppose $n \in \mathbb{Z}^+$ and suppose $f^{(n-1)}$ is continuous on $[a,b]$. Suppose $f^{(n)}$ exists for all $t \in (a,b)$. Let $\alpha, \beta \in [a,b]$, $\alpha \neq \beta$, and define

$$P_{n-1}(t) = \sum_{k=0}^{n-1} \frac{f^{(k)}(\alpha)}{k!}(t - \alpha)^k$$

(of degree $n-1$). Then $\exists$ a point $x \in (\alpha, \beta) : f(\beta) = P_{n-1}(\beta) + \frac{f^{(n)}(x)}{n!}(\beta - \alpha)^n$. Check that this makes sense for the cases $n = 1$ (MVT), etc.

Let's take a detour to vector-valued functions $f : [a,b] \to \mathbb{R}^k$, $f = (f_1, ..., f_k)$. A lot of the same properties applies; also f is differentiable at $x \in (a,b)$ iff f_i is differentiable at x. There are, however, some counter-examples to L'Hopital's Rule $f : [a,b] \to \mathbb{C} \cong \mathbb{R}^2$ (can you provide one?).

Theorem 4.45. Suppose $f : [a,b] \to \mathbb{R}^k$ and f is differentiable in (a,b). Then $\exists x \in (a,b)$: $||f(b) - f(a)|| \leq (b-a)||f'(x)||$. This is not the MVT (why?).

We should also make clear what we mean by derivatives in multivalued function:

Definition 4.46. Let $v = (v_1, ..., v_n)$ be a vector; then the directional derivative of $f(x)$ is given by $D_v f(x) = \lim_{h\to 0} \frac{f(x+hv)-f(x)}{h}$. The ***i*-th partial derivative** is the directional derivative w.r.t. the i-th standard basis vector. The **Jacobian** matrix for $f : \mathbb{R}^n \to \mathbb{R}^m$ is the $n \times m$ matrix of first partials.

We will fill in the details on multivariable derivatives and Jacobians as we go. Let's change gears to integration, which we define via the **Riemann-Stieltjes integral**. In particular, let P be a partition of the domain of f, and use upper and lower **Darboux sums** to define the Riemann-Stieltjes integral. We defined the upper and lower sums by $U(P, f, \alpha) = \sum_{i=1}^n M_i \Delta\alpha_i$ and $L(P, f, \alpha) = \sum_{i=1}^n m_i \Delta\alpha_i$. You should look up the definitions of Riemann-integrable yourself, but they'r basically what you think they are.

Note that if $f \in \mathcal{R}$ (Riemann-integrable), then changing value of f at one points does not change the integral (prove this). A sufficient condition of Riemann-integrability is that f is continuous on $[a, b]$. We will skip over the formalities of integration in the interest of time (in order to make room for a wider variety of concepts later). Anyways, lets move on to FTC, in the context of Riemann integrals only.

Theorem 4.47. (Fundamental Theorem of Calculus) Let $f \in \mathcal{R}$ on $[a, b]$. For $x \in [a, b]$ define $F(x) := \int_a^x f(t)dt$. Then F is continuous on $[a, b]$, and if f is continuous at some $x_0 \in [a, b]$, then F is differentiable at x_0 and $F'(x_0) = f(x_0)$.

Proof. Since $f \in \mathcal{R}$, f must be bounded. $|f(t)| \le M \ \forall t \in [a, b]$. If $a \le x < y \le b$, then $|F(x) - F(y)| = |\int_x^y f(t)dt| \le M(y - x)$. So $\forall x, y \in [a, b], |F(x) - F(y)| \le M(y - x)$. Thus F is uniformly continuous on $[a, b]$.

Now suppose f is continuous at x_0. Let $\epsilon > 0$ be given. Then $\exists \delta > 0 : |x_0 - t| < \delta \Rightarrow |f(x_0) - f(t)| < \epsilon$. WTS: $F'(x_0)$ exists and $F'(x_0) = f(x_0)$. Observe that

$$\left|\frac{F(t) - F(s)}{t - s} - f(x_0)\right| = \left|\frac{\int_s^t f(u)du}{(t - s)} - \frac{f(x_0)(t - s)}{(t - s)}\right| = \left|\frac{1}{t - s}\int_s^t [f(u) - f(x_0)]du\right| < \epsilon$$

It follows that $F'(x_0) = f(x_0)$. ∎

Theorem 4.48. (Fundamental Theorem of Calculus) If $f \in \mathcal{R}$ on $[a, b]$ and if $\exists$ a differentiable function on $[a, b] : F' = f$, then $\int_a^b f(x)dx = F(b) - F(a)$.

Also look at integration by parts, which you should prove for yourself. Lets move on to the study of sequences of functions. Set-up: we want to study functions that are real or complex-valued, and let sequences be maps $n \mapsto (f_n : X \to \mathbb{R})$, for X a metric space. For any sequence, we have a power series: let $S_M(z) = \sum_{n=0}^M c_n z^n$. For each z_0, we get a sequence of numbers $\{S_M(z_0)\}_{M=1}$, and can ask ourselves if this converges (which is the same as asking if z_0 is in the disc of covergence). We can define a new function that sends z_0 to the number it converges to, y_0; this is just like writing $e^x = \sum_{n=0}^\infty \frac{x^n}{n!}$.

So for each $x_0 \in X$, we get a sequence of numbers $\{f_n(x_0)\}|_{n=1}^\infty$. If this sequence converges, define a function at the point x_0 to be $f(x) = \lim_{n\to 0} f_n(x_0)$. So we build a function $f : X \to \mathbb{R}$ from the sequence of functions, and do this point by point in X.

Definition 4.49. Suppose $n \mapsto (f_n : E \to \mathbb{R})$ is a sequence of functions defined on a metric space E. Suppose the sequence of numbers $\{f(x_0)\}_{n=1}$ converges for all $x_0 \in E$. Build a new function $f(x_0) := \lim_{n\to\infty} f_n(x_0)$ for all $x_0 \in E$. We say $\{f_n\}$, a sequence of functions, converges on E and $f : E \to \mathbb{R}$ is the limit, or that $\{f_n\}$ converges *pointwise* to the function f on E.

Question: If each $f_n : E \to \mathbb{R}$ has a certain property (e.g. all f_n are continuous, differentiable, integrable), to what extent does $f : E \to \mathbb{R}$ also have that property?

Answer. g is continuous at a point c means $\lim_{x\to c} g(x) = g(c)$. $f(x) := \lim_{n\to\infty} f_n(x)$. Suppose each f_n is continuous at c; this means that for all n, $\lim_{x\to x} f_n(x) = f_n(c)$. Then $\lim_{x\to c} f(x) = \lim_{x\to c}(\lim_{n\to\infty} f_n(x)) \neq \lim_{n\to\infty} f_n(c) = f(c)$ in general, since we can't just switch the limits. So f is not continuous in general. Since we can't switch limits in general, we can't conclude that f is differentiable or integrable as well.

Example 4.50. A sequence of functions for which $\lim_{n\to\infty}\int_0^1 f_n(x)dx \neq \int_0^1 \lim_{n\to\infty} f_n(x)dx$.

Note that this is really a question of whether $\lim_{x\to c}(\lim_{n\to\infty} f_n(x)) = (\lim_{n\to\infty}(\lim_{x\to c} f_n(x))$, which generally doesn't hold as limits do not commute.

So consider the function $f_n(x) = n^2x(1-x)^n$ for $X = [0,1]$. Check that one limit is 0, and the other is 1.

Example 4.51. A sequence of differentiable functions $\{f_n\}$ with pointwise limit $f(x) = 0$, but $\{f_n'\}$ diverges.

Let $f_n(x) = \frac{\sin(nx)}{\sqrt{n}}$. The pointwise limit is zero, and clearly its derivative is zero. But the derivative of the formula above is $f_n'(x) = \sqrt{n}\cos(nx)$, whose limit does not exist as $n \to \infty$.

There's a stronger notion of convergence that will help regulate f_n, called *uniform convergence* (uniform w.r.t. your domain X).

Definition 4.52. f_n converges uniformly to f if for all $\epsilon > 0$ there exists N s.t. $\forall n \geq N$, $\forall x \in X$, $|f_n(x) - f(x)| < \epsilon$; e.g. there's a kind of "epsilon tube" that you can draw around the graph so that for n large enough, all functions are "trapped" in the tube.

Theorem 4.53. The sequence of functions $\{f_n\}$ defined on E converges uniformly on E iff $\forall\epsilon > 0, \exists N : \forall n, m \geq N, |f_n(x) - f_m(x)| < \epsilon \; \forall x \in E$.

Theorem 4.54. Suppose $\lim_{n\to\infty} f_n(x) = f(x) \; \forall x \in E$. Let $M_n = \sup_{x\in E}|f_n(x) - f(x)|$. Then $f_n \to f$ uniformly on E iff $\{M_n\} \to 0$.

Theorem 4.55. (Weierstrass M-test for series) Suppose $\{f_n\}$ is a sequence of functions defined on E and suppose $|f_n(x)| \leq M_n$ for all $x \in E$. Then $\sum_{n=1}^{\infty} f_n(x)$ converges uniformly on E if $\sum_{n=1}^{\infty} M_n$ converges.

Do we know that $\sum_{n=1}^{\infty}|f_n(x)|$ converges pointwise? Yes, by the comparison test. So the content of the M-test is that we get *uniform* convergence.

Does it make sense to take a sequence of numbers $\{a_n\}_{n=1}^{\infty}$ and ask if it converges uniformly on E? Silly question– take functions equal to the constants: $f_1(x) := a_1$, which converges uniformly if it converges.

Theorem 4.56. Assume that $f_n \to f$ uniformly on E. If each f_n is continuous at a point $c \in E$, then f is also continuous at c.

Proof. Note that if c is not an accumulation point of E, then the theorem holds because functions are continuous at isolated points. So suppose c is an accumulation point of E. Let $\epsilon > 0; \exists N : \forall n \geq N, |f_n(x) - f(x)| < \epsilon/3 \; \forall x \in E$. Since f_N is continuous at c, $\exists$ a ball $B_\delta(c)$, $\delta > 0$ s.t. $x \in B_\delta(c) \Rightarrow |f_N(x) - f_N(c)| < \epsilon/3$. Apply the triangle inequality lots of times:

$$|f(x) - f(c)| \leq |f(x) - f_N(x)| + |f_N(x) - f_N(c)| + |f_N(c) - f(c)| < 3(\epsilon/3) = \epsilon.$$

So if $x \in B_\delta(c)$, then $|f(x) - f(c)| < \epsilon$. ■

Anyways, we see that uniform convergence is strong enough to allow us to interchange limits when discussing continuity. Note that compactness of the domain does not guarantee uniform convergence, but it does guarantee uniform continuity.

Theorem 4.57. Suppose K is compact and

(1) $\{f_n\}$ is a sequence of functions on K
(2) $\{f_n\}$ converges pointwise to a continuous limit function
(3) $f_n(x) \leq f_{n+1}(x) \; \forall x \in K, n \in \mathbb{N}$. Then $f_n \to f$ uniformly on K.

Proof. Set $g_n := f_n - f$, and show that $g_n \to 0$ uniformly. Note $g_n \geq g_{n+1}$. Let $\epsilon > 0$, and define a sequence of sets $K_n := \{x \in K : g_n(x) \geq \epsilon\} = g_n^{-1}([\epsilon, \mathbb{R}))$. Note each K_n is compact.

Recall the definition: we have $\mathcal{F}$ a family of functions in the set of bounded, continuous functions $\mathcal{C}(X)$ that map from $X \to \mathbb{C}$, X compact, and norm $||f|| = \sup |f(x)|$, with $d(f,g) := ||f-g||$. Prove that, w/r/t to this metric induced by the sup norm, $\mathcal{C}([a,b])$ is complete.

Definition 4.58. A normed linear space which is complete w/r/t metric induced by the norm is called a **Banach space.** So $\mathcal{C}([a,b])$ is a Banach space. Moreover, if the norm (giving the metric) comes from an inner product, then the Banach space is called a **Hilbert space**. $\mathbb{R}^n$ is a Hilbert space.

Theorem 4.59. (Stone-Weierstrass) Every continuous function on $[a,b]$ can be uniformly approximated by polynomials. (e.g. set of polynomials is dense in set of continuous functions.)

More formally: If f is a continuous complex function on $[a,b]$, then $\exists$ a sequence of polynomials P_n s.t. $\lim_{n\to\infty} P_n(x) = f(x)$; this convergence is uniform on $[a,b]$, and if f is real, then the polynomials may be taken to be real.

Proof. Use the Bernstein polynomials. First adjust the problem: let $[a,b] = [0,1]$, and let $f(0) = f(1) = 0$. Find these polynomials: let $Q_n(x) = c_n(1-x^2)^n, n \in \mathbb{N}$; these have degree $2n$ because they are even. Choose $c_n : \int_0^1 Q_n(x)dx = 1$; e.g. $c_n = \frac{1}{2\int_0^1 (1-x^2)^n dx}$ (this is like a normalization step). What can we say about c_n? We have

$$\int_{-1}^{1} (1-x^2)^n dx = 2\int_0^1 (1-x^2)^n dx \geq 2\int_0^{1/\sqrt{n}} (1-x^2)^n dx \geq 2\int_0^{1/\sqrt{n}} (1-nx^2)dx = \frac{4}{3\sqrt{n}},$$

since over the domain $x \in [0, 1/\sqrt{n}]$, we have $(1-x^2)^n \geq (1-nx^2)$. So we have

$$c_n \int_{-1}^{1} (1-x^2)^n > \frac{c_n}{\sqrt{n}} \Rightarrow c < \sqrt{n}.$$

Now choose $\delta \in (0,1]$. Then for any $x : \delta \leq |x| \leq 1$, we have $Q_n(x) = c_n(1-x^2)^n \leq \sqrt{n}(1-\delta^2)^n$. This means that $Q_n \to 0$ uniformly on $[\delta, 1]$. Then here are the polynomials that we want: $P_n(x) = \int_{-1}^{1} f(x+t)Q_n(t)dt$, for $x \in [0,1]$. Change variables: let $s = x+t$, so that

$$P_n(x) = \int_{-1}^{1} f(x+t)Q_n(t)dt = \int_{x-1}^{x+1} f(s)Q_n(s-x)ds = \int_0^1 f(t)Q(t-x)dt,$$

again for $x \in [0,1]$. f is zero outside $[0,1]$, which gives the latter equality above. So for our sequence $\{P_n(x)\}$ of polynomials, we must show that $P_n \to f$ uniformly on $[0,1]$.

Claim: $P_n \to f$ uniformly on $[0,1]$. Let $\epsilon > 0$ be given, and find

$$\delta > 0 : |y-x| < \delta \Rightarrow |f(y) - f(x)| < \epsilon/2, \quad (1)$$

which follows by continuity and compactness. Set $M := \sup_{x\in[0,1]} |f(x)| < \infty$. We know from above that $Q_n \leq \sqrt{n}(1-\delta^2)^n$ for $\delta \leq |x| \leq 1$, and $Q_n(x) \geq 0$. Now $\forall x \in [0,1]$, we have that $| \int_{-1}^{1} f(x)Q_n(t)dt = f(x)\cdot 1 = f(x)$, so

$$|P_n(x) - f(x)| = \left| \int_{-1}^{1} [f(x+t) - f(x)]Q_n(t)dt \right| \leq \int_{-1}^{1} |f(x+t) - f(x)|Q_n(t)dt,$$

where the latter inequality follows from the triangle inequality. Split this up into a sum of three integrals:

$$\int_{-1}^{1} |f(x+t)-f(x)|Q_n(t)dt = \underbrace{\int_{-1}^{-\delta} |f(x+t)-f(x)|Q_n(t)dt}_{\leq 2M\sqrt{n}(1-\delta^2)^n} + \underbrace{\int_{-\delta}^{\delta} |f(x+t)-f(x)|Q_n(t)dt}_{\leq \epsilon/2}$$

$$+\underbrace{\int_\delta^1 |f(x+t)-f(x)|Q_n(t)dt}_{\leq 2M\sqrt{n}(1-\delta^2)^n} \leq 4M\sqrt{n}(1-\delta)^n + \epsilon/2.$$

Choose n to make this as small as you want, and since this doesn't depend on x, we have uniform convergence. ■

Recall the family $Q_n(x) = c_n(1-x^2)^n$, which as $n \to \infty$ tends to the *Dirac delta* function. This is the "function"

$$\delta(x) = \begin{cases} +\infty & x = 0 \\ 0 & x \neq 0, \end{cases}$$

which also has the property that $\int \delta(x)dx = 1$.

Finally, let's jump to multivariable calculus. In particular, let's go back to derivatives on multivalued functions. We want to make sense of a derivative of the function $f : \mathbb{R}^n \to \mathbb{R}^m$, so consider how it's defined for f restricted to $\mathbb{R}^1 \to \mathbb{R}^1$. To the vector a, f' assigns a tangent vector attached to $f(a)$ in the range. *The derivative df_a is a linear transformation between tangent spaces.* To make this more rigorous, we'll need to define *tangent space* (defined using manifolds in differential topology, e.g. Math 132).

Recall from linear algebra that $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ is the nm-dimensional v.s. of all linear transformations from $\mathbb{R}^n \to \mathbb{R}^m$. Put a norm on it: for $A \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$,

$$||A|| := \sup_{|x|\leq 1} |Ax|.$$

Is this actually a norm, and does this norm give a metric we can use? If it is a norm, we can write down $||A - B||$ to be a metric and go from there. Also, if $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ is a metric space, is it complete? Compact? We'll put these questions aside for later.

Note that $|A\frac{x}{|x|}| = \frac{|Ax|}{|x|} \leq ||A||$. Show that $||A|| := \sup_{|x|\leq 1} |Ax| = \sup_{|x|=1} |Ax|$.

Theorem 4.60. The following are true:
(1) If $A \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, then $||A|| < \infty$ and $A : \mathbb{R}^n \to \mathbb{R}^m$ is uniformly continuous.
(2) If $A, B \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ and $c \in \mathbb{R}$, then $||A + B \leq ||A|| + ||B||$ and $||cA|| = |c| \cdot ||A||$.
(3) If $A \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ and $B \in \mathcal{L}(\mathbb{R}^m, \mathbb{R}^k)$, then $||BA|| \leq ||B|| \cdot ||A||$.

Let's do some calculus. The derivative is $\lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$, and we can rewrite this as $f(x+h)-f(x) = f'(x)h + r(h)$, for $\lim_{h\to 0} \frac{r(h)}{h} \to 0$. This expression is affine in h, and in particular $f'(x)h$ is linear in h. Define the derivative like this: say $f(a,b) \to \mathbb{R}^m$, $(a,b) \subseteq \mathbb{R}$. $f'(X)$ is the vector $y \in \mathbb{R}^m$ s.t. $f(x+h) - f(x) = h \cdot y + r(h)$ and $\frac{r(h)}{h} \to 0$.

Definition 4.61. Suppose $E \subseteq \mathbb{R}^n$ is an open set in $\mathbb{R}^n$ and $f : E \to \mathbb{R}^n$. Fix $x \in E$. If $\exists$ a linear transformation $A \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ s.t.

$$\lim_{h\to 0} \frac{|f(x+h) - f(x) - Ah|}{|h|} = 0,$$

then f is **differentiable** at x and $f'(x) = A$. Show that this is well-defined:

Theorem 4.62. Let $f : E \to \mathbb{R}^m$, and suppose $\exists A_1, A_2$ satisfying the limit equation for the derivative above. Then $A_1 = A_2$.

Proof. Let $B = A_1 - A_2$, and observe that

$$\lim_{h\to 0} \frac{|Bh|}{|h|} \leq \lim_{h\to 0} \frac{|f(x+h) - f(x) - A_1 h|}{|h|} + \lim_{h\to 0} \frac{|f(x+h) - f(x) - A_2 h|}{|h|},$$

so $\lim_{h\to 0}\frac{|Bh|}{h}=0$. Now take a fixed $h\neq 0$, and consider $\frac{|B(th)|}{h}=\frac{t|B(h)|}{h}\to 0$ as $t\to 0$. So B must be 0, and $A_1=A_2$ as desired. ■

For a map $f:E\to\mathbb{R}^m$, E an open subset of $\mathbb{R}^n$, $f'(x)$ exists at $x_0\in E$ if $\exists A\in\mathcal{L}(\mathbb{R}^n,\mathbb{R}^m)$:

$$\lim_{h\to 0}\frac{|f(x+h)-f(x)-Ah|}{|h|}=0,$$

and $f'(x_0)=A$. *The derivative at a point is a linear transformation between tangent spaces.*

How do we compute these guys? If $T\in\mathcal{L}(\mathbb{R}^n,\mathbb{R}^m)$, and $f(x):=Tx, x\in\mathbb{R}^n$, we claim that $f'(x)=T$. Verify this using the definition: if $f'(x)$ exists, then

$$\lim_{h\to 0}\frac{|T(x+h)-T(x)-Ah|}{|h|}=\lim_{h\to 0}\frac{|T(h)-Ah|}{|h|}=0,$$

and $A=T$.

In the above case, we relied heavily on linearity, but weird things can happen if we have maps that "mess up the domain" by squaring, folding, etc. We'll need some more tools:

Theorem 4.63. (Chain rule) Let $f:E\to\mathbb{R}^m$, $E\subseteq\mathbb{R}^n$ and $x_0\in E$. Suppose f is differentiable at x_0. Also let $g:U\to\mathbb{R}^k$, $U\subseteq\mathbb{R}^m$ and $f(E)\subseteq U$, and suppose g is differentiable at $f(x_0)$. Then $g\circ f:E\to\mathbb{R}^k$ is differentiable at x_0, and $(g\circ f)'(x_0)=g'(f(x_0))\cdot f'(x_0)$.

Definition 4.64. Let's have an aside on partial derivatives. For f a function as above, we choose standard bases $\{e_1,...,e_n\}$ of $\mathbb{R}^n$ and $\{u_1,...,u_m\}$ of $\mathbb{R}^m$. With these bases, we can work the coordinate functions. For

$$\mathbf{f}(x)=\begin{pmatrix} f_1(x)\\ \vdots\\ f_m(x)\end{pmatrix},\ f(x)=f_1(x)u_1+...+f_m(x)u_m,$$

we can define the **partial derivative of f_i in the direction** e_j as $D_jf_i(x)=\lim_{t\to 0}\frac{f_i(x+te_j)-f_i(x)}{t}$ if the limit exists. We will have the notation $\frac{\partial f_i}{\partial x_j}$ to denote this partial.

Partial derivatives help compute $f'(x)\in\mathcal{L}(\mathbb{R}^n,\mathbb{R}^m)$. Question: how intrinsic are partial derivatives? We had to choose a basis above, so they depend on coordinates, but one thing to note is that *derivatives do not depend on coordinates.*

Theorem 4.65. Let $f:E\to\mathbb{R}^m$, $E\subseteq\mathbb{R}^n$. Suppose f is differentiable at $x\in E$. Then the partial derivatives $(D_jf_i)(x)$ exist and $f'(x)e_j=\sum_{i=1}^m((D_jf_i)(x))u_i$.

Do we have enough to write down $\mathcal{M}(f'(x),\{e_1,...,e_n\},\{u_1,...,u_m\})$, the matrix representation of $f'(x)$ w.r.t. the basis $\{e_1,...,e_n\}$ in the domain and the basis $\{u_1,...,u_m\}$ in the range? **Yes.** Write this yourself.

Definition 4.66. For $f:E\to\mathbb{R}$, we define the **gradient vector**, a vector in $\mathbb{R}^n$, as $(\nabla f)(x)=\sum_{i=1}^n(D_if)(x)e_i$. This is just a vector of partial derivatives. Also note that the **directional derivative of f in the direction** u is $\lim_{t\to 0}\frac{f(x+tu)-f(x)}{t}=((\nabla f)(x)\cdot u)=D_uf(x)$. It follows from Cauchy-Schwarz that $((\nabla f)(x)\cdot u)$ is maximal for a given f and x when u is parallel to $(\nabla f)(x)$.

Definition 4.67. Let $f:E\to\mathbb{R}^m$, $E\subseteq\mathbb{R}^n$ be a differentiable mapping. f is **continuously differentiable** in E if $f':E\to\mathcal{L}(\mathbb{R}^n,\mathbb{R}^m)$ is continuous. This definition means $\forall x\in E,\forall\epsilon>0,\exists\delta>0:$ $|x-y|<\delta\Rightarrow||f'(x)-f'(y)||<\epsilon$. Notation: $f\in\mathcal{C}(E)\subseteq\mathcal{L}(\mathbb{R}^n,\mathbb{R}^m)$.

Let's move on to the *Inverse and Implicit Function Theorems*, the latter of which is perhaps the most important theorem in analysis and beyond. The inverse function theorem answers the question of when a map f is *locally* invertible (it's globally invertible if f is injective). Loosely speaking, the derivative should be invertible; e.g. a continuous differentiable map $f : E \to \mathbb{R}^m$ is invertible in a neighborhood of any point $x \in E$ at which $f'(x) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ is invertible.

Theorem 4.68. (Contraction mapping theorem) Let $G : X \to G$, X be a metric space, $\forall x, y \in X$, $d(G(x), G(y)) \leq cd(x, y)$ where $c < 1$. Then G is a contraction and there exists *at most* one fixed point. Also, if X is complete, then there exists a unique fixed point of a contraction G.

Theorem 4.69. (Inverse function theorem) Suppose f is a $\mathcal{C}'$ mapping $f : E \to \mathbb{R}^m$, $E \subseteq \mathbb{R}^n$, and suppose $f'(a)$ is invertible. Set $b = f(a)$. Then

(1) $\exists U, V := f(U)$ open s.t. $f|_U : U \to V$ is injective;

(2) $g : V \to U$, $g \in \mathcal{C}'(V)$.

To motivate the implicit function theorem, we'll give a linear algebra version.

Theorem 4.70. (IFT, linear version) Let $A \in \mathcal{L}(\mathbb{R}^{n+m}, \mathbb{R}^n)$. If $A_x \in \mathcal{L}(\mathbb{R}^n)$ is invertible, then $\forall k \in \mathbb{R}^m$, $\exists! h \in \mathbb{R}^n : A(h, k) = A_y(h) + A_y(k) = 0$.

Proof. $A(h, k) = A_y(h) + A_y(k)$. Taking $h = -(A_x)^{-1} A_y K$ finishes the proof.

Theorem. (IFT) Let f be a $\mathcal{C}'$ mapping, $f : E \to \mathbb{R}^n$ s.t. $f(a, b) = 0$ for some $(a, b) \in E$. Set $A := f'(a, b)$ and suppose A_x is invertible. Then

(1) $\exists$ open sets $U \subseteq \mathbb{R}^{n+m}$ and $W \subseteq \mathbb{R}^m$ with $(a, b) \in U$ and $b \in W$ having the following property: to every $y \in W, \exists! x : (x, y) \in U, f(x, y) = 0$;

(2) If $x := g(y)$, then g is a $\mathcal{C}'$-mapping of W into $\mathbb{R}^n$, $g(b) = a$, $f(g(y), y) = 0$ $\forall y \in W$ (the function g is implicitly defined by this equation) and $g'(b) = -(A_x)^{-1} Ay$.

We will move on to another topic without proving the inverse and implicit function theorems, but keep in mind that they are among the most important theorems in analysis, and are widely used in topology, geometry, physics, etc.

The is something obvious that remains to be done. We want to generalize the FTC to n dimensions. The tradeoff for this is understanding differential forms. Differential forms du LOVE to be integrated; e.g. $\int_D du =$ number. There's an extra issue of *orientation*, e.g. $\int_a^b F(t)dt = -\int_b^a F(t)dt$. These are both integrals of the form $\int_{[a,b]}$, with the extra information of orientation. The determinant of a matrix will help us out here. First some terminology:

Definition 4.71. A **??-field** is attaching a *??* at each point or open set in $\mathbb{R}^2$. Examples are vector fields, form fields, strawberry fields...

Definition 4.72. $\det : \mathrm{Mat}(n, n, \mathbb{R}) \to \mathbb{R}$ is the unique function s.t. it is multilinear (linear in columns of matrix), alternating (-1 comes out through row swap), and normalized ($\det(I) = 1$).

Definition 4.73. A k**-form** on $\mathbb{R}^n$ is a function φ that takes k vectors in $\mathbb{R}^n$ and returns a number denoted $\varphi(v_1, ..., v_k)$ so that:

(1) φ is k-linear in columns

(2) φ is alternating ($\varphi(v_1, v_2, v_3) = -\varphi(v_2, v_1, v_3)$)

The number k is called the **degree** of the form.

Let $i_1, ..., i_k$ be any k integers between 1 and n. Then the form $dx_{i_1} \wedge ... \wedge dx_{i_k}$ is a function of k vectors $v_1, ..., v_k$ in $\mathbb{R}$ that, for the $n \times k$ matrix

$$\begin{pmatrix} (v_1) & (v_2) & \dots & (v_k), \end{pmatrix}$$

selects k rows from this matrix and assigns them the determinant. Note that 0-forms take 0 vectors and assigns them a number; e.g. they are functions.

Geometrically, what is $dx_1 \wedge dx_2$?

$$dx_1 \wedge dx_2 \left(\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}, \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \right) = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} = a_1 b_2 - b_1 a_2.$$

Recall that the determinant of n vectors $dx_1 \wedge dx_2 \wedge ... \wedge d_{x_n}$ recalls a *signed area*. So the geometric meaning is some "signed volume."

If we project $\vec{a}$ and $\vec{b}$ onto the (x_1, x_2) plane in $\mathbb{R}^3$ gives $(a_1 \;\; a_2)^T + (b_1 \;\; b_2)^T$, the det $\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}$ gives a signed volume of the parallelogram spanned by $(a_1 \;\; a_2)^T$ and $(b_1 \;\; b_2)^T$. $dx_1 \wedge dx_2$ should be the (x_1, x_2)-component of the signed area.

Elementary/Basic Forms. Consider the set S of all k-forms on $\mathbb{R}^n$. This is a vector space. But first consider some other questions: the form $dx_1 \wedge ... \wedge dx_k$ might be redundant; for example, $dx_1 \wedge dx_3 \wedge dx_1 = 0$. We want to eliminate this redundancy, and count other things like $dx_1 \wedge dx_3$ and $dx_3 \wedge dx_1$ (which differ by a negative sign) as identical. So we must consider the most basic forms. The equivalence is as follows:

If $\{i_1, ..., i_k\}$ and $\{j_1, ..., j_k\}$ are equal on sets of positive integers, then $dx_{j_1} \wedge ... \wedge dx_{j_k} = sgn(\sigma) dx_{i_1} \wedge ... \wedge dx_{i_k}$, where σ is the permutation $j_1 = i_{\sigma(1)}, ..., j_k = i_{\sigma(k)}$.

Definition 4.74. An **elementary k-form** on $\mathbb{R}^n$ is an expression of the form $dx_{i_1} \wedge ... \wedge dx_{i_k}$ for $1 \leq i_1 < ... < i_k \leq n$ (with indices written in ascending order and all distinct). Denote the elementary 0-form as 1.

Question: can we have a k-form on $\mathbb{R}^n$ with $k > n$, elementary or otherwise?

Definition 4.75. *Addition of k-forms*: Let φ and ψ be two k-forms on $\mathbb{R}^n$. Then $(\varphi + \psi)(v_1, ..., v_k) := \varphi(v_1, ..., v_k) + \psi(v_1, ..., v_k)$. *Scalar multiplication of k-forms*: $(a\varphi)(v_1, ..., v_k) := a \cdot \varphi(v_1, ..., v_k)$.

Convince yourself that, w.r.t. addition and scalar multiplication defined above, the set $A^k(\mathbb{R}^n)$ of all k-forms on $\mathbb{R}^n$ is a vector space.

Theorem 4.76. Every k-form can be uniquely written as

$$\varphi = \sum_{1 \leq i_1 < ... < i_k \leq n} a_{i_1 ... i_k} dx_{i_1} \wedge ... \wedge dx_{i_k},$$

for $a_{i_1}, ..., a_{i_k} \in \mathbb{R}$, where $a_{i_1, ..., i_k} = \varphi(e_{i_1}, ..., e_{i_k})$, for $(e_{i_1}, ..., e_{i_k})$ being the k standard basis vectors in $\mathbb{R}^n$. The sum above means that any k-form can be written as a linear combination of basis elements of $A^k(\mathbb{R}^n)$.

Example 4.77. Fix a vector $v \in \mathbb{R}^n$, and let $\varphi_v(w) = v \cdot w$. This is a 1-form, and we should be able to write this as a linear combination of the elementary 1-forms $dx_1, ..., dx_n$. So $\varphi_v = a_1 dx_1 + a_2 dx_2 + ... + a_n dx_n$, for $a_i = \varphi_v(e_i) = v_i$. This says that φ_v is the dot product of v and $(dx_1, dx_2, ..., dx_n)$. The sum in the theorem above gets a bit more messy with 2-forms.

Remark. $\dim(A^k(\mathbb{R}^n)) = \binom{n}{k} = \frac{n!}{(n-k)!k!}$.

Let's talk more about the *wedge product.* What is $\wedge$, and how do we get a 2-form $dx_1 \wedge dx_2$ from the wedge operator on dx_1 and dx_2?

Definition 4.78. The *wedge product* of the forms $\varphi \in A^k(\mathbb{R}^n)$ and $\omega \in A^l(\mathbb{R}^n)$ is $\varphi \wedge \omega \in A^{k+1}(\mathbb{R}^n)$ defined by

$$(\varphi \wedge \omega)(v_1, ..., v_{k+l}) := \sum_{\sigma \in \text{Perm}(k,l)} \text{sgn}(\sigma) \cdot \varphi(v_{\sigma(1)}, ..., v_{\sigma(k)}) \cdot \omega(v_{\sigma(k+1)}, ..., v_{\sigma(k+l)}),$$

where the sum is over all permutations σ of the numbers $1, 2, ..., k+l$ such that $\sigma(1) < \sigma(2) < ... < \sigma(k)$ and $\sigma(k+1) < \sigma(k+2) < ... < \sigma(k+l)$. Note that σ is not required to be globally increasing on $1, 2, ..., k+l$, as then σ would be the identity.

Properties of the wedge product: Let $\varphi \in A^k(\mathbb{R}^n)$ and $\omega \in A^l(\mathbb{R}^n)$.
(1) Distributivity: $\varphi \wedge (\omega_1 + \omega_2) = \varphi \wedge \omega_1 + \varphi \wedge \omega_k$.
(2) Associativity: $(\varphi_1 \wedge \varphi_2) \wedge \varphi_3 = \varphi_1 \wedge (\varphi_2 \wedge \varphi_3)$.
(3) Skew commutativity: $\varphi \wedge \omega = (-1)^{kl} \omega \wedge \varphi$.

We need areas on which to integrate these forms:

Definition 4.79. The *support* of a real or complex function f on $\mathbb{R}^k$ ($\mathbb{R}^k \to \mathbb{R}$ or $\mathbb{C}$) is the closure of the set of all points $x \in \mathbb{R}^k$ at which $f(x) \neq 0$.

If supp(f) is bounded (and therefore compact) as a subset of $\mathbb{R}^k$, then $\int_{\mathbb{R}^k} f := \int_{I^k} f$, for some I^k containing the support. We now can integrate continuous functions on $\mathbb{R}^n$ with compact support (Lebesgue theory).

Definition 4.80. The **k-simplex** $Q^k := \{x \in \mathbb{R}^k | x_1 + ... + x_k \leq 1, x_i \geq 0\}$. This places a summing condition on the positive orthant of $\mathbb{R}^k$.

We'll need additional tools, that of primitive mappings and partitions of unity:

Theorem 4.81. (Partitions of unity) Suppose $K \subseteq \mathbb{R}^n$ is compact and $\{V_\alpha\}$ is an open cover of K. Then $\exists$ functions $\psi_1 ..., \psi_3 \in \mathcal{C}(\mathbb{R}^n)$ s.t.
(1) $0 \leq \psi_i \leq 1$
(2) Each ψ_i has support in some V_α
(3) $\phi_1(x) + ... + \phi_3(x) = 1 \; \forall x \in K$.
In particular, we can decompose f into different pieces according to each ψ_i with small support; e.g. $f(x) = \sum_{i=1}^3 f(x)\psi_i(x)$.

Theorem 4.82. (Change of variables) Suppose T is a one-to-one $\mathcal{C}'$ map $T : E \to \mathbb{R}^k$, $E \subseteq \mathbb{R}^k$. Suppose that $J_T(x) \neq 0$. If f is a continuous function on $\mathbb{R}^k$ whose support is compact and contained in $T(E)$, then

$$\int_{\mathbb{R}^k} f(y)dy = \int_{\mathbb{R}^k} f(T(x))|J_T(x)|dx.$$

We have $|\,|$ in $|J_T(x)|$ to fix orientation. Also note that $T^{-1}(\text{supp}(f))$ is indeed compact (so that we can integrate the RHS) by the inverse function theorem, since T is a closed mapping and continuous.

Note that an affine map is $T(x) = Ax + b$, where A is linear and $+b$ indicates a translation. Let's move on to more differential forms. What does the integrand in $\int_a^b f(t)dt$ mean? It's a one-form on $\mathbb{R}$, and not a function. For the ensuing discussion, assume $E \subseteq \mathbb{R}^n$ is open.

Alright, things are going to get interesting! We want to work with what we call *differential forms*:

Definition 4.83. A **k-surface** in E is a $\mathcal{C}'$-mapping $\Phi : D \to E$, for D compact in $\mathbb{R}^k$. D is called the *parameter domain*. Note that we saw something like this before, when we had $\gamma : [0,1] \to \mathbb{R}^2$.

Because of our restricted viewpoint, we restrict to the case when D is either I^k or Q^k.

Definition 4.84. A **differential form of order $k \geq 1$, or a k-form**, is a function ω s.t.

$$\omega = \sum_{1 \leq i_1 < ... < i_k \leq n} b_{i_1,...,i_k}(x) dx_{i_1} \wedge ... \wedge dx_{i_k},$$

which assigns to each k-surface Φ in E a number $\int_\Phi \omega$ according to the following rule:

$$\int_\Phi \omega = \int_D \left(\sum_i b_{i_1,\dots,i_k}(\Phi(u)) \frac{\partial(x_{i_1,\dots,i_k})}{\partial(u_1,\dots,u_k)} \right) du,$$

where D is a parameter domain of Φ.

Example 4.85. Let D be a cube in $\mathbb{R}^3$, $(r,\theta,\varphi) \in \mathbb{R}^3$, $r \in [0,1], \theta \in [0,\pi], \varphi \in [0,2\pi]$. Let $\Phi : (r,\theta,\varphi) \mapsto (r\sin\theta\cos\varphi, r\sin\theta\sin\varphi, r\cos\theta) = (x,y,z) \in \mathbb{R}^3$.

Now define ω, a 3-form, and compute $\int_\Phi \omega$. $\omega = dx \wedge dy \wedge dz$, so

$$\int_\Phi \omega = \int_D (1) \left(\begin{vmatrix} \partial x/\partial r & \partial x/\partial \theta & \partial x/\partial \varphi \\ \partial y/\partial r & \partial y/\partial \theta & \partial y/\partial \varphi \\ \partial z/\partial r & \partial z/\partial \theta & \partial z/\partial \varphi \end{vmatrix} \right) = \int_{I^3} (1)(r^2 \sin\theta) = 4\pi/3.$$

If ω is of class $\mathcal{C}'$, then we can talk about d, the exterior derivative operator. If ω is a k-form, then $d\omega$ is a $k+1$-form. If f is a 0-form, then df is a 1-form and is given by $\sum_{i=1}^n (D_i f(x))dx_i$. In general, if $\omega = \sum_I b_I(x)dx_I$, then $d\omega = \sum_I (db_I) \wedge dx_I$.

Theorem 4.86. If ω is a k-form and λ is a m-form of class $\mathcal{C}'$, in E, then
(1) $d(\omega \wedge \lambda) = (d\omega) \wedge \lambda + (-1)^k \omega \wedge d\lambda$
(2) If ω is of class $\mathcal{C}''$ in E, then $d^2\omega = 0$.

In math, there are notions of a pullback and a pushforward, both of which come from category theory. If $f : x \mapsto y$ between two spaces, then attaching a structure at y, we can pullback the local structure to x. Differential forms can be pulled back. Question: can they be pushed forward?

What does it mean to pullback a differential form? Take $x \in E \subseteq \mathbb{R}^n$, $y \in V \subseteq \mathbb{R}^m$, so that if $T : E \to V, T(x_0) = y_0$ (with the point not necessarily unique). Let ω be a k-form field on V, so we can write $\omega = \sum_I b_I(y_0)dy_I$. We want to use T to define a differential form ω_T at x in the domain. Try $\omega_T = \sum_I b_I(T(x_0))dx_I$. This should remind you of dual spaces: if $g \in W^*, T : V \to W$, then $T^t(g) = (g \circ T) \in V^*$.

What about the pushforward? Reverse the set-up: for $\omega = \sum_I b_I(x_0)dx_I$ a differential form at x in E, perhaps the best candidate is $\omega_{\text{push}} = \sum_I a_I(y_0)dy_I = \sum_I b_I(T^{-1}(y_0))dy_I$. But this isn't well-defined unless T is injective.

Definition 4.87. $\omega_T = \sum_I b_I(T(x_0))dx_I$ is called the **pullback** of $\omega = \sum_I b_I(y_0)dy_I$, a k-form, in E.

Theorem 4.88. If ω is a k-form and λ is a m-form, then
(1) $(\omega + \lambda)_T = \omega_T + \lambda_T$
(2) $(\omega \wedge \lambda) = \omega_T \wedge \lambda_T$
(3) $d(\omega_T) = (d\omega)_T$.

Now let $E \subseteq \mathbb{R}^n$, $T : E \to V \subseteq \mathbb{R}^m$, $T \in \mathcal{C}'$, $S : V \to W \subseteq \mathbb{R}^p$, $S \in \mathcal{C}'$. We can pullback twice a ω in W to get $(\omega_S)_T$ in E, and we also see that $(\omega_S)_T = \omega_{S\circ T} := \omega_{ST}$.

Theorem 4.89. Suppose T is a $\mathcal{C}'$ mapping of an open set $E \subseteq \mathbb{R}^n$ into an open set $V \subseteq \mathbb{R}^m$, and Φ is a k-surface in E and ω is a k-form in V. Then $\int_\Phi \omega_T = \int_{T\Phi} \omega$.

We'll talk briefly about a small portion of simplical homology (introduced in algebraic topology). Our previous set-up was for *De Rham cohomology* (cohomology is dual to homology). Stokes's Theorem is a duality statement as well. There are more abstract objects known as *simplices.* Recall that a k-simplex is $Q^k = \{u \in \mathbb{R}^k | u = \sum_{i=1}^k \alpha_i \epsilon_i | \alpha_i \geq 0, \sum_{i=1}^k \alpha_i \leq 1\}$. This traces out a positive orthant in $\mathbb{R}^k$. Fix

$p_0, p_1, ..., p_k \in \mathbb{R}^n$.

Definition 4.90. The **oriented affine k-simplex** $\sigma = [p_0, ..., p_k]$ is the k-surface in $\mathbb{R}^n$ with parameter domain Q^k, given by the map $\sigma(\alpha_1\epsilon_1 + ... + \alpha_k\epsilon_k) = p_0 + \sum_{i=1}^k \alpha_i(p_i - p_0)$. Also, let $\overline{\sigma} = [p_{i_0}, ..., p_{i_k}]$, where $p_{i_0}, ..., p_{i_k}$ is a permutation of $p_0, ..., p_k$. These two are related by the sign of the permutation.

Take all affine k-simplices σ's on $\mathbb{R}^n$, and put them into groups of $+$ and $-$ orientation.

Definition 4.91. An **oriented 0-simplex** takes $Q^0 \to \mathbb{R}^n$, and define this as $\sigma = +p_0$ or $\sigma = -p_0$. $\int_\sigma f := \pm 1 f(p_0)$.

Theorem 4.92. If σ is an oriented rectilinear k-simplex in an open set $E \subseteq \mathbb{R}^n$ and $\overline{\sigma} = \epsilon \cdot \sigma$ $(\epsilon = \pm 1)$, then for all k-forms ω in E,

$$\int_{\overline{\sigma}} \omega = \epsilon \int_\sigma \omega.$$

Definition 4.93. An **affine k-chain** Γ in an open set $E \subseteq \mathbb{R}^n$ is a collection of finitely many oriented affine k-simplices $\sigma_1, ..., \sigma_r$ in E (not necessarily distinct). If ω is a k-form in E, then $\int_\Gamma \omega := \sum_{i=1}^r \int_{\sigma_i} \omega$. We denote $\Gamma = \sigma_1 + ... + \sigma_r$, which is a formal sum.

Definition 4.94. For $k \geq 1$ the boundary of the **oriented affine k-simplex** $\sigma = [p_0, ..., p_k]$ is defined to be the affine $(k-1)$ chain $\partial\sigma = \sum_{j=0}^k (-1)^j [p_0, ..., p_{j-1}, p_{j+1}, ..., p_k]$, the summands of which are $k-1$ simplices.

Example 4.95. A line connecting p_0 and p_1 is a 1-simplex in $\mathbb{R}^3$; $\sigma = [p_0, p_1]$. $\partial([p_0, p_1]) = [p_1] - [p_0]$. Notice that this has a left-to-right orientation.

Example 4.96. A triangle with nodes p_0, p_1, p_2 is a 3-simplex. $\partial([p_0, p_1, p_2]) = [p_1, p_2] - [p_0, p_2] + [p_0, p_1]$, which is a chain of 2-simplices with orientation. Also check that $\partial([p_1, p_2] - [p_0, p_2] + [p_0, p_1]) = 0$ (notice that applying the boundary operator twice gives zero always).

These tools let us state Stokes' theorem more precisely, which we will state without proof:

Theorem 4.97. (Stokes theorem) If Ψ is a k-form of class $\mathcal{C}''$ in an open set $V \subseteq \mathbb{R}^m$ and if W is a $(k-1)$-form of class $\mathcal{C}'$ in V, then

$$\int_\Psi d\omega = \int_{\partial\Psi} \omega.$$

Remark 1. Note that taking $k = 1$ gives the FTC: $\int_{\text{1-chain}} df = \int_{\partial(\text{1-chain})} f$, e.g. $\int_{[a,b]} df = \int_{b-a} f = f(b) - f(a)$.
Remark 2. Note that the exterior derivative operator (with $d^2 = 0$) and the boundary operator (with $\partial^2 = 0$) are dual to each other.

5 Complex analysis

Complex analysis is the study of complex-analytic functions, which are also called **holomorphic**. Some of the most important things in complex analysis are the Cauchy-Riemann equations, contour integration, and analytic continuation. Due to time considerations, we will only go over some topics and examples en passant; please fill in the details yourself.

Definition 5.1 $f : \mathbb{C} \to \mathbb{C}$ is **holomorphic** at $z \in \mathbb{C}$ if the limit

$$\lim_{h\to 0} \frac{f(z+h) - f(z)}{h}$$

exists for $h \in \mathbb{C}$. Check that linearity and the product and quotient rules hold for holomorphic functions. To test whether or not a function is holomorphic, we have the **Cauchy-Riemann** equations:

Theorem 5.2. (Cauchy-Riemann equations) $f : \mathbb{C} \to \mathbb{C}$ is complex-analytic on Ω iff for f written as $f(x, y) = u(x, y) + iv(x, y)$, the following equations hold on Ω:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \text{ and } \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}.$$

If f satisfies the Cauchy-Riemann equations, then $f'(z) = \frac{\partial f}{\partial z}$.

Brushing aside some formality in provide provided by a complex analysis course (look up Goursat's theorem, Morera's theorem, etc.), we state Cauchy's integral formula:

Theorem 5.3. (Cauchy integral formula) Suppose $\Omega \subseteq \mathbb{C}$ is open, $f : \Omega \to \mathbb{C}$ is holomorphic and $D := \{z : |z - z_0| \leq r\}$ is contained in Ω. Let γ be the circle forming the boundary of D. Then for every a in the interior of D,

$$f(a) = \frac{1}{2\pi i} \int_\gamma \frac{f(z)}{z - a} dz$$

Cauchy's integral formula and the theory of **residues** give rise to a novel method of evaluating integrals, namely **contour integration**. We will illustrate this with an example:

Example 5.4. Evaluate

$$\int_{-\infty}^{\infty} \frac{x \sin x}{x^2 + a^2} dx$$

using the **residue theorem**.

Answer. Observe that the poles of $f(z) = \frac{ze^{iz}}{z^2+a^2}$ are $\pm ai$, and $z = ai$ in the upper half-plane has residue $\frac{iae^{-a}}{2ia} = e^{-a}/2$: to see this, observe that

$$\lim_{z\to ai} (z - ai) \frac{ze^{iz}}{z^2 + a^2} = \lim_{z\to ai} (z - ai) \frac{ze^{iz}}{(z - ai)(z + ai)} = \left.\frac{ze^{iz}}{z + ai}\right|_{z=ai} = e^{-a}/2.$$

Now we write $f(z) = \frac{z(\cos z+i \sin z)}{z^2+a^2}$, so we would like to compute the imaginary part of $\int_{-\infty}^{\infty} f(z)dz$. To do this, we will integrate $f(z)$ over a semi-disk; e.g. if we let C_R denote the counterclockwise contour of the semi-circle of radius R centered at 0, we first claim that $\int_{C_R} f(z)dz = 0$ as $R \to \infty$ because integration by parts with $u = \frac{z}{z^2+a^2}$ and $dv = e^{iz}$ gives

$$\int_{C_R} f(z)dz = \int_{C_R} \frac{ze^{iz}}{z^2 + a^2} dz = \left.\frac{ze^{iz}}{i(z^2 + a^2)}\right|_{-R}^{R} - \int_{C_R} e^{iz} \frac{-(z^2 - a^2)}{i(z^2 + a^2)^2} dz$$

Since $z \in C_R \Rightarrow |e^{iz}| \leq 1$, we impose the bound $|\frac{ze^{iz}}{i(z^2+a^2)}|_{-R}^{R}| \leq \frac{2R}{R^2-a^2}$. The **estimation lemma**, and the observation that the length of the path of integration is half the circumference of a circle with radius R, gives us the bound $|\int_{C_R} e^{iz} \frac{-(z^2-a^2)}{i(z^2+a^2)^2} dz| \leq \frac{(a^2+R^2)\pi R}{(R^2-a^2)^2}$ (for $R > a$). Thus, combining these bounds and letting $R \to \infty$, we see that

$$\frac{ze^{iz}}{i(z^2 + a^2)}|_{-R}^{R} - \int_{C_R} e^{iz} \frac{-(z^2 - a^2)}{i(z^2 + a^2)^2} dz \leq \frac{2R}{R^2 - a^2} + \frac{(a^2 + R^2)\pi R}{(R^2 - a^2)^2} \to 0.$$

Now the result follows from the residue theorem by integrating over the semi-disk of radius R at the origin and letting $R \to \infty$, that

$$\int_{-\infty}^{\infty} \frac{x \sin x}{x^2 + a^2} dx = \text{Im}(2\pi i \ \text{res}_{ai} f) = \text{Im}(2\pi i (e^{-a}/2)) = \pi e^{-a},$$

as desired. ■

Complex analysis also deals with **meromorphic functions**, **entire functions**, **gamma and zeta functions**, **conformal mappings**, and much other fun stuff. Since it will take a while to build up the background for these other concepts, we will skip to something else.... If you want to read up on complex analysis, Stein and Shakarchi's *Complex Analysis* is a good text; a more advanced text is Ahlfors's *Complex Analysis*.

6 Probability theory

Probability theory is the study of random variables and properties of their distribution. It's probably one of the funnest branches of mathematics, and certainly one of the most applicable.

The following notes are from an advanced probability theory course at Harvard, and will provide you with great context. Hopefully, the style and format of these notes will give you a sense of what college and grad school-level mathematics is like, so I will not bother to simplify them. Please look up terms that you are not familiar with. For a more basic introduction (if you don't want to fill in te gaps yourself), see DeGroot's *Probability and Statistics*.

Measure Theory. The Banach-Tarski paradox provides motivation for measure theory. We'll write (Ω, F, P) for reference, where Ω is the sample space. We'll assume that Ω is fixed.

Definition. A *σ-algebra* is defined to be $F \subseteq 2^\Omega$ (where 2^Ω is the *power set* of Ω, i.e. the set of all subsets) such that:

1. $\emptyset \in F$
2. If $A \in F$, then $A^C \in F$.
3. If $A_1, A_2, ... \in F$, then $\cup_{n=1}^{\infty} A_n \in F$.

Example. Let Ω be partitioned into disjoint pieces A, B, C, D. The σ-algebra is then $F = \{\emptyset, \Omega, A, B, C, D, A \cup B, A \cup C, A \cup B \cup D, ..., \}$, and F has 16 elements.
Remark. The number of all elements in all σ-algebras is a power of 2.
Remark. There are no countably infinite σ-algebras. *(Why?)*

There is one very important σ-algebra we will need, called the *Borel σ-algebra* with $\Omega = \mathbb{R}$. How do we construct it? Let's begin with the set of all closed intervals $[a, b] : a, b \in \mathbb{R}$. So $F_0 = \{$closed intervals$\}$. This isn't a σ-algebra, so now let's take $F_1 = \{$countable unions, intersections of sets in F_0 or their complements$\}$. F_1 is not a σ-algebra either. So let F_2 be the set in which we keep taking unions and intersections. F_2 is still not a σ-algebra (show this).

Now let $F_\infty = \cup_{j=1}^{\infty} F_j$. F_∞ is still not a σ-algebra (we can show this in analysis). Let $F_{\infty+1}$ be defined likewise ($\infty + 1$ is in the *ordinal*, but not *cardinal* sense). If we keep going, we still do not get a σ-algebra, but if we go into uncountable territory we will finally obtain a σ-algebra.

Fact. Any non-empty intersection of σ-algebras is a σ-algebra on the same Ω.
Caveat. In general, this is false for unions.

An immediate consequence of this fact is that if we have any collection F_0 (which may or may not be a σ-algebra), there exists a smallest unique σ-algebra containing F_0. This is because we can just intersect all σ-algebras that contain F_0 (at least one σ-algebra contains F_0). This is called the σ-algebra *generated* by F_0.

What about P? P is a mapping such that $P : F \to [0, 1]$, and that the following two axioms hold:

1. $P(\emptyset) = 0$, $P(\Omega) = 1$.

2. $P(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$ if $A_1, A_2, ...$ are disjoint events.

These axioms are necessary if you want to be coherent. Can we accept finite, but not countable additivity? This is an alternate philosophy.

Definition. A *random variable* is a function $X : \Omega \to \mathbb{R}$ such that X is *measurable*; i.e. $X^{-1}(B) \in \mathcal{F}$ for all Borel $B \subseteq \mathbb{R}$.

Here $X^{-1}(B) := \{\omega \in \Omega : X(\omega) \in B\} := \{x \in B\}$.

The *events* are of the following form: $X = x$, $X \leq x$.

Example. If $B = (-\infty, 3]$, then $x \in B$ is an event. So $x \leq 3$, which is $\{\omega \in \Omega : X(\omega) \leq 3\}$.
Notation. B is a Borel set, while $\mathcal{B}$ is a Borel σ-algebra. So $B \in \mathcal{B}$.

Most of probability theory is on random variables and their representations. We use measure to rigorously define r.v.'s, which are described by their distributions.

Definition. A r.v. is a mapping $X : \Omega \to \mathbb{R}$. For a σ-algebra F, $X^{-1}(B) \in F$ (the preimage) for all B Borel in $\mathbb{R}$. An inverse may not necessarily exist.

We can also write this as $X^{-1}(B) = (X \in B)$. Although this may be an abuse of notation, this allows us to write things like $X = x, (X \leq x) = X^{-1}((-\infty, x])$.

Let's spend some time on images and preimages. For $f : S \to T$, the image of $A \subseteq S$ is $f(A) = \{f(x) : x \in A\}$. $f(S)$ is the range.

The *preimage* is as follows: for $B \subseteq T$, then $f^{-1}(B) = \{x \in S : f(x) \in B\}$.

Let's practice some true-false questions.

1. $f(\cup_\alpha A_\alpha) = \cup_\alpha f(A_\alpha)$: True
2. $f(\cap_\alpha A_\alpha) = \cap_\alpha f(A_\alpha)$: False
3. $f^{-1}(\cup_\alpha A_\alpha) = \cup_\alpha f^{-1}(A_\alpha)$: True
4. $f^{-1}(\cap_\alpha A_\alpha) = \cap_\alpha f^{-1}(A_\alpha)$: True
5. $f^{-1}(A^C) = (f^{-1}(A))^C$

Why is (4) true and (2) false? Let's write a quick "definition-chasing" proof: Let $x \in f^{-1}(\cap_\alpha A_\alpha) \Leftrightarrow f(x) \in \cap_\alpha A_\alpha \Leftrightarrow f(x) \in A_\alpha \forall \alpha \Leftrightarrow x \in f^{-1}(A_\alpha) \forall \alpha \Leftrightarrow x \in \cap_\alpha f^{-1}(A_\alpha)$.

(2) can be edited to make a true statement: $f(\cap_\alpha A_\alpha) \subseteq \cap_\alpha f(A_\alpha)$.

The technical definition of a r.v. requires the assumption that $(X \in B) \in F$ for all Borel B. Why is it enough to look at intervals of the form $B = (-\infty, x]$ for all $x \in \mathbb{R}$? B's of this form actually generate all Borel sets. Now consider $\{B \text{ Borel: } X^{-1}(B) \in F\}$. Assume $(-\infty, x] \in A$ for all x, and notice that A is a σ-algebra: $\emptyset \in A$, complements are in A; now show countable unions are also in A. For $B_1, B_2, ... \in A$, $X^{-1}(\cup_{n=1}^{\infty} B_n) = \cup_{n=1}^{\infty} X^{-1}(B_n) (\in F) \in F$. So A is a σ-algebra. Since A contains the smallest algebra containing all Borel B, and actually is the smallest σ-algebra containing all such Borel B, it contains all generations of B, so we conclude that $A = B$.

Definition. The *distribution of a r.v.* X is $P(x \in B)$, as a function of $B \in \mathcal{B}$, where $\mathcal{B}$ is the set of all Borel sets over $\mathbb{R}$. This is also known as *the law* of $X, \mathcal{L}(X)$.

We don't want to find the P's for all Borel sets, so we wish to show sometime that it suffices to consider $F(x) = P(X \leq x)$. This is known as the CDF, the cumulative distributive function. It is perfectly valid to think of the CDF as the distribution.

Definition. If $\vec{X} : \Omega \to \mathbb{R}^n$ is a random vector, then its (joint) distribution is $P(\vec{X} \in B)$, where B is Borel in $\mathbb{R}^n$.

Notation. $X \sim Y$ means X and Y, which are r.v.'s, have the same distribution. This will, hopefully, not conflict with the notation $X \sim N(0, 1)$.
Caveat. A r.v. is distinct from its distribution.

Proof. (1) Assume $T \leq n$, with $n \in \mathbb{Z}^+$. $M_t = M_0 + (M_1 - M_0) + (M_2 - M_1) + ... = M_0 + \sum_{j=1}^{T}(M_j - M_{j-1})$. So we want to show $E(\sum_{j=1}^{T}(M_j - M_{j-1})) = 0$.

$$M_0 + \sum_{j=1}^{T}(M_j - M_{j-1}) = M_0 + \sum_{j=1}^{n}(M_j - M_{j-1})I_{(T\geq j)}$$

$$\Rightarrow EM_T = EM_0 + \sum_{j=1}^{n} E((M_j - M_{j-1})I_{(T\geq j)}).$$

By Adam's Law, $E((M_j - M_{j-1})I_{(T\geq j)}) = E(E((M_j - M_{j-1})(I_{(T\geq j)}|X_0, X_1, ..., X_{j-1}))$. But M is a martingale and T is a stopping time, so $I(T \geq j) = 1 - I_{(T<j)} = 1 - I_{(T\leq j-1)}$. But this is known since we have X_{j-1}. So "$= E(I(T \geq j)E((M_j - M_{j-1})|X_0, ..., X_{j-1})) = E(I(T \geq j)0) = 0$.

(2) Prove this by truncation. Let $T_n = \min(T, n)$. So this is a bounded stopping time (check that it still is a stopping time). Then $ET_n = EM_0$ by what we just proved, and taking $n \to \infty$ we observe that $T_n \to T$, $M_{T_n} \to M_T$ a.s. Since we assume that $|M_{T_n}| \leq c$, use bounded convergence and take the limit of both sides.

(3) Use a telescoping series, as before, to write $M_{T_n} - M_0 = \sum_{j=1}^{T_n}(M_j - M_{j-1}))$. Bound this using the triangle inequality: "$\leq \sum_{j=1}^{T_n} |M_j - M_{j-1}| \leq cT_n \leq cT$. But we also assumed that $E(T) \leq \infty$. Dominated convergence then says that $E(M_{T_n} - M_0) \to E(M_T - M_0)$.

Now we turn to consider the gambler's ruin problem. The problem is the same as considering a RW on the integers, which track the changes to gambler A's money. Bankruptcy for A will be at $-a$, and for B bankruptcy will be b. Let p be the probability that A wins in each round, and T the stopping time; $T = \inf\{n : M_n = -a \text{ or } M_n = b\}$.

If $p = 1/2$, then it's a SSRW again, and we consider a stopped martingale at $-a$ or b. (Note that this is also a Markov chain.) We can apply Convergence to see that M converges a.s., and we can use the Optional Stopping Theorem to show that $P(A \text{ wins}) = \frac{a}{a+b}$, and $E(T) = ab$.

Now consider the unfair case. If $p \neq 1/2$, $q = 1 - p$. This is a Markov chain, but not a martingale. But $(\frac{q}{p})^{M_n}$ is a martingale (check this; this is also related to NEFs). (If the ratio blows up, apply truncation again.)

Example. (Say "Red") We have a deck of 52 cards. Turn over cards one-by-one, but before that say "red" if you guess that the next card will be red. You win if you guess correctly, and lose if you don't. We can do this for any card, including no cards.

We'll prove a theorem that says our chances of winning can never be better or worse than 1/2. Let M_n be the fraction of red cards in the remaining cards after n cards have been revealed. $0 \leq n \leq 51$. (A pencil problem is to check that M_n is a martingale.) Thus $EM_T = EM_0 = 1/2$.

Another way to think about this result is as follows. Draw a picture of 52 cards □□□...□. Let T denote the T cards; then there are $52 - T$ cards remaning. After T cards, we can flip the $52 - T$ cards around, and therefore we can always pick the last card, which has a probability of 1/2.

7 Algebra and algebraic topology

Algebra is the study of structure, and one of the most important branches is **group theory**. With the background you've built thus far, you should have no problem grasping the fundamentals of algebra. Thus I will only mention a few points here; for a great textbook, consult Artin's *Algebra* and Hatcher's *Algebraic Topology*

A **group** is a set endowed with an operator (group multiplication) that satisfies closure, associativity, identity, inverse. For example, $\mathbb{Z}$ is a group w.r.t. integer addition: for $a, b, c \in \mathbb{Z}, a + b \in \mathbb{Z}$; $(a + b) + c = a + (b + c)$; $0 + a = a$; $-a + a = 0$. You should check the the dihedral group of order n, D_n, which has elements being the transformations of a regular polygon (that is, the identity, rotations, and flips), is also a group. Groups are important for modeling symmetry, as well as number theory etc.

Subgroup are groups in bigger groups. There are many things that can be proved about them. In the future, look up the **Sylow theorems**, **normal subgroups**, **quotient groups**....

Rings are sets with operators $+$ and $\cdot$ that satisfy closure, associativity, existence of additive/multiplication identity, existense of additive inverse, commutativity of addition, and distributive laws. Some examples of rings: $\mathbb{Z}, \mathbb{R}$.

For an elementary treatment of algebraic topology, consult Munkres's *Topology.* For a high-powered graduate introduction, see Hatcher's book. I will simply list some of the more important concepts:

Homotopy. A homotopy can be thought of as a continuous deformation of one shape to another. For example, a doughnut is homotopic to a coffee mug (they both have **genus** 1). We can define an **equivalence relation** w.r.t. homotopy. An equivalence relation is $\sim$ on a set A such that for all $a, b, c \in A$, $a \sim a, a \sim b \Longrightarrow b \sim a, a \sim b \sim c \Longrightarrow a \sim c$.

The **fundamental group** is the first homotopy group. You can associate a group to any topological space that provides a way of determining when two paths, starting and ending at a fixed base point, can be continuously deformed into each other. In $\mathbb{R}^n$, the fundamental group is the trivial group with one element. The circle, since each homotopy class consists of all loops which wind around the circle a number of times, is $\mathbb{Z}$. **van Kampen's theorem** allows you to express the structure of fundamental group of topological space in terms of the fundamental groups of two open, path-connected subspaces that cover X.

Please look up these other terms yourself: covering spaces, pushout diagrams, fibers, CW complexes, homology theory, category theory, groupoids....

8 Differential topology and geometry

Definition 8.1. A function f is said to be **smooth** if it's of class C^∞ (that is, it is infinitely continuously differentiable).

Just as linear maps were the central functions of study in linear algebra, **diffeomorphisms** are the central objects of study in differential topology.

Definition 8.2. A map f is a **diffeomorphism** if it's one-to-one and onto, and if the inverse map f^{-1} is also smooth.

Diffeomorphisms map from one space to another, and for smoothness to make sense we need something that's like Euclidean space, but we can generalize what we mean by a "space" into **manifold**.

Definition 8.3. M is a k**-dimensional manifold** if it is, intuitively, locally diffeomorphic to $\mathbb{R}^k$, meaning it has a **coordinate chart**. In particular, a topological manifold of dimension k is $(M, I, \{(U_i, \varphi_i) : i \in I\})$ where M is a paracompact, Hausdorff topological space, I is an index set, and for all $i \in I$, $U_i \subset M$ is open and $\varphi_i : U_i \to V_i \subset \mathbb{R}^n$, where φ_i is a homeomorphism and $V_i \subset \mathbb{R}^n$ is open. Also, $M = \cup_{i \in I} U_i$.

Definition 8.4. M is **smooth** if $\varphi_j \circ \varphi_i^{-1}$ is smooth; likewise for **analytic** (coordinate change is analytic), **complex** (coordinate change is holomorphic), algebraic, affine,....

Example 8.5. The circle $S^1 = \{x \in \mathbb{R}^2 : ||x|| = 1\}$ is a smooth manifold; $S^n/$ is also a smooth manifold by **stereographic projection**. Other examples include the projective space $\mathbb{CP}^n$, the torus $T^n = (S^1)^n$, Grassmannians....

Example 8.6. **Lie groups** are groups that are also manifolds. Examples of this include the general linear group $GL(n, \mathbb{R})$, the orthogonal group $O(n, \mathbb{R})$, the standard linear group $SL(n, \mathbb{R})$, the standard orthogonal group $SO(n, \mathbb{R})$, the unitary group $U(n)$, etc. Use the implicit function theorem to show that these are indeed manifolds. The **Lie algebras** of these groups can be thought of as their tangent spaces at the identity.

We will skip over things usually covered in an undergraduate differential topology course, e.g. **transversality**, **immersion** (f is one if df_x is injective at x), **embeddings** (an injective immersion, e.g. one that is diffeomorphic to its image, which is a submanifold), **Sard's theorem**, **Whitney embedding theorem**, **intersection theory**....

I'd like to skip to a part of differential topology that I find a bit more interesting, namely **de Rham cohomology**. This is the cohomology on the set of differential forms on a smooth manifold M with the exterior derivative as the link.

The background on differential forms provided in the Real Analysis should be sufficient. Differential k-forms (which are formed by tensoring more elementary differential forms; a p-tensor on V is a real-valued function T on V^p that is multilinear, such as linear functionals, dot product, determinant, etc.) are very important things here; denote the set of all k-forms on a smooth manifold M as $\Omega^k(M)$.

Remark. The **Einstein summation convention** is to write something like n_i instead of $\sum_{i=0}^{\infty} n_i$ when the sum is understood. We will adopt this convention from now on.

Anyways, let's go over the **exterior derivative** quickly. Forms can not only be integrated, but also differentiated via the exterior differentiation operator.

Definition 8.7. If $\omega = a_I dx_I$ is a k-form, then the **exterior derivative of** ω is a $k+1$ form given by $d\omega = da_I \wedge dx_I$.

Theorem 8.8. The exterior differentiation operator, defined on smooth forms on $U \subset \mathbb{R}^k$ open, has the following three properties:

1. Linearity: $d(\omega_1 + \omega_2) = d\omega_1 + d\omega_2$.
2. The Multiplication Law: $d(\omega \wedge \theta) = (d\omega) \wedge \theta + (-1)^p \omega \wedge d\theta$ if ω is a p-form.
3. The *cocycle condition*: $d(d\omega) = 0$ for $\omega \in \Omega^k(M)$.

Moreover, this is the only operator that has these properties and agrees with the previous definition of df for smooth functions f.

Proof. This is trivial; use the definition to prove these statements. ■

Definition 8.9. $\omega \in \Omega^k(M)$ is **closed** if $d\omega = 0$, and **exact** $\omega = d\theta$ for $\theta \in \Omega^{k-1}(M)$. Two closed k-forms are **cohomologous** if their difference is exact: $\omega \sim \omega' \implies \omega - \omega' = d\theta$.

De Rham cohomology, in a sense, measures the difference between closedness and exactness of differential forms, which has many applications in analysis and topology (for example, it gives the nontrivial solutions of a set of differential equations...; for more soft questions on cohomology/homology, consult google). So we have a **cochain complex**

$$0 \longrightarrow_d \Omega^0(M) \longrightarrow_d \Omega^1(M) \longrightarrow ...$$

Definition 8.10. The **k-th de Rham cohomology group** $H^k(M)$ is the set of equivalence classes of k-forms, e.g.

$$H^k(M) := \text{Ker}(d : \Omega^k(M) \to \Omega^{k+1}(M))/\text{Im}(d : \Omega^k(M) \to \Omega^{k+1}(M)).$$

I will leave you to learn more about cohomology and other parts of differential topology/geometry. In particular, you should look up **Mayer-Vietoris sequences**, **Poincaré lemma**, **vector bundles**,

Poincaré duality, **symplectic geometry**, **metrics**, **connections**, **torsion**, **geodesics**.... Due to limited time, I cannot go over these here. But some good sources are Guillemin and Pollack's *Differential Topology*, and Jeffrey Lee's *Manifolds and Differential Geometry*....

9 A brief discussion of other applications

The physics applications of all the mathematics we've discussed above should be more or less obvious, especially in the section for differential geometry (in particular, manifolds are the basic mathematical structure we use to talk about our universe and string theory; see, for example, **Calabi-Yau manifold**). Symplectic geometry and the Hamiltonian are very important concepts for mechanics.

There are some not-so-obvious applications of mathematics, namely in biology. Mathematics in the form of differential equations and stochastic processes show up a lot in biology: for example, when analyzing the effects of a drug (need differential equations), the stability of a system (need Jacobians), the processes of stochastic gene expression (need stochastic analysis/simulations). **Graph theory** helps with a lot of the computations, and is especially important in other fields of mathematics or applied mathematics, such as theoretical computer science. Evolution also uses a bunch of mathematics: some examples are the Lotka-Volterra equations, Markov chains....

This latter, more applied approach can fall under the category of dynamical systems, which involves analysis. Fractals and chaos also fall into this category. For a good introduction, see Sholomo Sternberg's *Dynamical Systems*.

10 Tips for succeeding in mathematics at college

1. You should also challenge yourself to learn at the highest level possible.
2. Make sure you really understand definitions. Have examples of every term in mind.
3. Seek out faculty and friends to talk about mathematics; the knowledge you gain from collaboration is tantamount to being a good mathematician. You will oftentimes learn more mathematics by talking about it with friends than reading.
4. Get to enjoy learning about proofs and writing them. Proofs are at the heart of mathematics, even in applied mathematics, and the sooner you come to terms with them, the better off you'll be.
5. Read more textbooks. While doing the required readings will get you through class, to develop a feel for how mathematics is done you should acquaint yourself with the styles and though processes of many mathematicians.
6. Similar to the following point, if there is a problem that you cannot get, even after consulting the teacher and many other friends, it's probably time to move on. Returning to the same question after more experience or doing other problems really helps.
7. Lastly, don't forget to enjoy your life in college! While mathematics is a very important thing, remember that you only go to college once, but your career as a mathematician (or scientist) will last a lifetime.