18.032 — Lecture Notes

Sanjana Das

Spring 2023

Notes for 18.032 (Differential Equations), taught by Professor Tristan Ozuch. All errors are my responsibility.

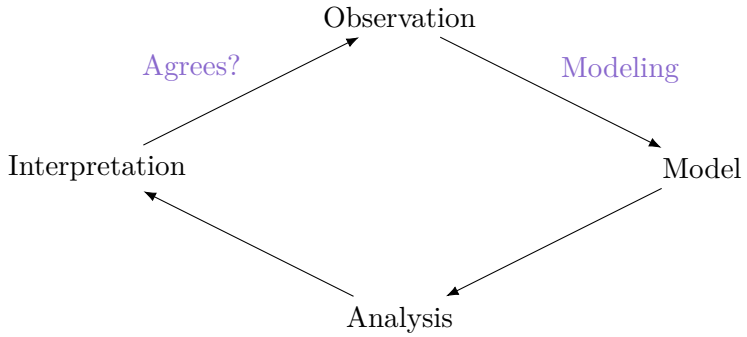
Contents

§1 February 6, 2022

§1.1 Modeling

One application of differential equations is to modeling.

To explain some observation, we can try to produce a model. Then by analyzing and interpreting the model, we may obtain and solve an equation. We then need to check whether this solution agrees with reality; if it doesn't, we may need to refine or completely change our model. We then do this over and over, and hopefully we eventually get a model that accurately matches our observations.



Today we'll see a few examples of modeling.

Modeling involves two parts:

- We need to decide which *quantities* we will follow, and which *parameters* they depend on. (For example, if we're modeling a population, our parameters may be position and time, and our quantity may be the number of people.)
- We need to decide a plausible relationship between them.

§1.1.1 Modeling an epidemic

Suppose we want to model a virus epidemic.

- Our parameter will be time (denoted t).
- Our quantities will be:
  - I, the proportion of infected people.
  - S, the proportion of people who are susceptible (to be infected in the future).
  - R, the proportion of people who are removed from the equation. In reality, this could mean many things — for example, that they've been infected and now have immunity, or that they are vaccinated, or that they are dead — but mathematically we don't need to distinguish.

Example 1.1

One simple model is the following, where there are no removed people and once you're infected you stay infected:

$$S \longrightarrow I$$

We must have $S + I = 1$ (since they're proportions).

It's plausible that the rate of infection is proportional to the number of susceptible people (if there's 10 times more susceptible people to get infected, then 10 times as many people should get infected), as well as to the number of infected people (the more infected people there are, the more people will get infected); so

$$dI = a \cdot S \cdot I \, dt$$

for some constant a.

Using $S + I = 1$, we can rewrite this as a differential equation in one variable:

$$\begin{cases} I'(t) = a(1 - I(t)) \cdot I(t) \\ S(t) = 1 - I(t). \end{cases}$$

The first equation is an ODE, the kind of equation we'll be studying through this semester.

Our class will mostly be on going from the model to the interpretation, but it still can be good to understand where the model comes from.

Now let's make this model a bit more complicated, by considering removed people.

Example 1.2

Suppose now that the infected people get removed (e.g., they recover or die).

$$S \longrightarrow I \longrightarrow R$$

As before, we have $S + I + R = 1$. We also have

$$I'(t) = a \cdot S(t) \cdot I(t) \, dt - b \cdot I(t)$$

(the bI represents people recovering, which should be proportional to the number of infected people). This also gives $R'(t) = bI(t)$, so we have the system

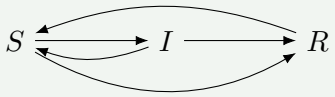
$$\begin{cases} I'(t) = a(1 - I(t) - R(t))I(t) - bI(t) \\ R'(t) = bI(t) \\ S(t) = 1 - I(t) - R(t). \end{cases}$$

This is called a *system of ODEs*.

We can also imagine more complicated models:

Example 1.3

Suppose that some people who are infected go back to being susceptible after they recover, and some removed people can become susceptible again (by mutations). Also suppose some people can go directly from susceptible to removed (by vaccinations).



We could also add a spatial variable, making this a *partial differential equation* instead (which is more complicated to study).

§1.2 Differential Equations

Definition 1.4. A **differential equation** is an equation involving functions and their derivatives.

This is a very general definition, and differential equations can be quite complicated — in particular, they may involve multiple variables.

Definition 1.5. An **ordinary differential equation** (ODE) is a differential equation with *one* parameter.

Often, the one common variable is time (denoted t).

Definition 1.6. A **partial differential equation** (PDE) is a differential equation with more than one parameter.

You can think of ODEs and PDEs as single-variable and multivariable calculus, but there's an even bigger gap between them — in research, often when you're given an ODE, you're done — we'll see this semester that there are very powerful theorems that can let you analyze the ODE. Meanwhile, starting from a PDE, you usually can't say very much. In this class, we will focus on ODEs.

Example 1.7

The differential equation

$$y'(t) \cdot y(t)^2 = \frac{1}{\sqrt{y(t)} - 1} + t$$

is an ODE.

Example 1.8

For a function $y(x,t)$, the differential equation

$$\frac{\partial}{\partial t} y(x,t) = \partial_{x^2}^2 y(x,t)$$

is a PDE.

Definition 1.9. A **general ODE** of order n is an equation of the form $F(t, y(t), y'(t), y''(t), \ldots, y^{(n)}(t)) = 0$.

Notation 1.10. Often we will exclude t, as there is only one variable, and write $F(t, y, y', \dots, y^{(n)})$ instead.

You could imagine equations involving infinitely many derivatives, but these are unlikely to come up; we focus on equations only involving finitely many derivatives.

Definition 1.11. A solution of an ODE on an interval I is a function $\varphi\colon I \to \mathbb{R}$ such that:

- φ is n times differentiable;
- $f(t, \varphi(t), \varphi'(t), \dots, \varphi^{(n)}(t)) = 0$ for all $t \in I$.

Usually our ODEs will also have some initial conditions — an order n ODE will look like

$$\begin{cases} F(t, y, y', \dots, y^{(n)}) = 0 & \text{on } I \\ y(t_0) = a_0, \dots, y^{(n-1)}(t_0) = a_{n-1} & \text{at } t_0 \in I. \end{cases}$$

(The values at t_0 are the initial conditions.)

The solution to the equation may be very different for different initial conditions (e.g., the epidemic would work very differently if everyone vs. no one was initially infected).

A well-posed problem:

- Has a solution (the existence of solutions is nontrivial — many PDEs don't have solutions); i.e., *existence*. (This is often the hardest part; sometimes we only require a weaker sense of a solution than the above definition.)
- There is only one solution, i.e., *uniqueness*. (This is important physically because if there are multiple possible solutions, we can't make physical predictions using the model because we don't know which one to use.)
- The solution is 'well-behaved' with respect to both the initial conditions and perturbations — this is called *stability*. More precisely, by 'initial conditions' we mean that if we change the initial conditions by very small numbers ε_i, we should get a very close solution (physically, we can't get completely precise measurements of the initial conditions, so if a small error in initial conditions results in very different predictions, the model is not physically useful).
- If we perturb the *equation* F by a bit — replacing it with $F_\varepsilon \simeq F$ (this notation means F_ε is 'almost equal' to F) (e.g., in the equation, we neglected a small amount of friction), the solution shouldn't change much — for the same reason, if we're not stable in this case, then we're not describing reality.

Stability is important because sometimes we'll have a mathematical model that gives completely different predictions than physical reality. For example, if we're modelling a beam being pushed to bend, it theoretically can either bend one way or the other. But physically it will always end the same way, because the beam isn't perfect (so we've perturbed the starting situation with a bit of imperfection, and the solution changes).

§2 February 8, 2023 — Integration and Separability

Today we will look at the simplest ODEs — ODEs which can be solved simply by integrating, linear ODEs, and separable ODEs. For these ODEs, we can explicitly solve the ODE — this is the best situation, but it's not that common. In the second half of the class we'll see ODEs which we're *not* able to solve; and we'll see how to find properties of the solution without solving it.

Until the first midterm, we will only look at ODEs of order 1. (Order 2 ODEs are common in physics, because the acceleration of a particle (the second derivative) depends on the force, which may depend on the position and velocity.)

An order 0 ODE is of the form $F(t, y) = 0$; to solve such an equation, we want to solve for y as a function of t — i.e., find a function $f(t)$ such that $F(t, f(t)) = 0$. (This isn't always possible, but it's something we are familiar with.)

§2.1 Integration

Question 2.1. Solve an ODE of the form

$$\begin{cases} y'(t) = f(t) & \text{on } I \subset \mathbb{R} \\ y(t_0) = y_0 & \text{for some } t_0 \in I. \end{cases}$$

Assume that f is a continuous function (in notation, $f \in C^0$).

Theorem 2.2

The unique solution of the above ODE is $y(t) = y_0 + \int_{t_0}^{t} f(s)\, ds$ for all $t \in I$.

Notation 2.3. We will sometimes drop the variable of integration and write $\int_{t_0}^{t} f$.

Proof. First we'll prove existence. We're given the solution, so it suffices to check that it works: by the fundamental theorem of calculus we have

$$h'(t) = \frac{d}{dt} \int_{t_0}^{t} f = f(t).$$

and the initial condition is satisfied as $y(t_0) = y_0 + \int_{t_0}^{t_0} f = 0$. (Usually this is not how we will prove existence, since we won't be given a solution.)

To prove uniqueness, assume we have two solutions; we will show that they have to be the same. (This is how we will always prove uniqueness.) Assume that y and z are both solutions to the equation; we then want to show that $y(t) = z(t)$ for all $t \in I$.

What's nice about our differential equation is that it is linear, so they behave nicely when we subtract. If we define $w = y - z$, then w is a solution of

$$\begin{cases} w'(t) = y'(t) - z'(t) = 0 \\ w(t_0) = 0. \end{cases}$$

(By linear we mean that the left-hand side is linear in y — $\text{LHS}(\lambda y + \mu z) = \lambda \text{LHS}(y) + \mu \text{LHS}(z)$.) Since w has derivative 0, it must be a constant. Then since $w(t_0) = 0$, this constant must be 0. So $w = 0$, meaning $y = z$. □

Exercise 2.4. If we have a nigher-order equation

$$\begin{cases} y^{(n)}(t) = f(t) \\ y(0) = y_0, \dots, y^{(n-1)}(0) = y_{n-1}, \end{cases}$$

then the unique solution is

$$y(t) = y_0 + \int_0^t \left(y_1 + \int_0^{s_1} \left(y_2 + \cdots + \int_0^{s_{n-2}} y_{n-1} + \int_0^{s_{n-1}} f(s_n)\, ds_n \right) ds_{n-1} \right) \cdots ds_1\, dt.$$

(This can be proven by iteration.)

This gives an intuitive explanation fo why we need this many initial conditions.

§2.2 Separable ODEs

A separable ODE is one that looks like $y' = f(y)g(t)$. The main point is that we can put all the terms depending on y on one side, and all terms depending on t on the other.

The method is the following:

- As long as $f(y) \neq 0$, by dividing we have

$$\frac{y'}{f(y)} = g(t) \implies \frac{dy}{f(y)} = g(t)\, dt.$$

  (We will not make this rigorous, though it can be done.)
- Integrate to get $\int \frac{dy}{f(y)} = \int g(t)$. Suppose the expression on the left is $F(y)$.
- Solve $F(y(t)) = \int g(t) + C$.
- Verify that this is a solution to the equation.
- Adjust C according to the initial condition.

Example 2.5

Suppose that $I'(t) = (1 - I(t))I(t)$. Then as long as $I(t)$ is not 0 or 1, we can divide and get

$$\frac{I'(t)}{(1-I)I} = 1.$$

We now write this as

$$I'(t) \cdot \left(\frac{1 - I(t) + I(t)}{(1 - I(t))I(t)}\right) = \frac{I'(t)}{1 - I(t)} + \frac{I'(t)}{I(t)} = 1.$$

These are $-\log(1-I)'(t)$ and $(\log I)'(t)$ respectively, so we have

$$F(y) = \int_{t_0}^{y} \left(\frac{1}{1-u} + \frac{1}{u}\right) du + c = \log\left(\frac{y}{1-y}\right) - \log\left(\frac{y_0}{1-y_0}\right) + c.$$

Now we have $F(y(t)) = t - t_0 + c$. Now this is an equation of the usual type. We have

$$\log \frac{y}{1-y} = t + c,$$

which means

$$\frac{y(t)}{1-y(t)} = e^{t+c} \implies y(t) = e^{t+c} - e^{t+c}y(t),$$

which gives us

$$y(t) = \frac{e^{t+c}}{1+e^{t+c}}.$$

Now we can verify that this satisfies the initial equation — we have

$$y'(t) = y(t)(1 - y(t))$$

(left as an exercise). Finally, we want to make sure that $y(t_0) = y_0$. Finding c (in terms of t_0 and y_0) is left as an exercise.

Next class we will look at first-order linear ODEs. These are of the form

$$\begin{cases} y'(t) = a(t)y(t) + f(t) \\ y(t_0) = y_0. \end{cases}$$

(These are actually 'affine' and not linear; truly linear ones don't have $f(t)$.) These are the most important because they're the first-order generalization of very complicated ODEs; by the end of the class we'll study more complicated ODEs with linear ones.

§3 February 10, 2023

Remark 3.1. For a separable ODE of the form $y'g(y) = f(t)$, the main point is that we want to rewrite this as $(G(y))' = f(t)$. If we let $G(y) = z$, then this is simply the equation $z' = f(t)$, which we can solve by integrating — we have

$$\begin{cases} z' = f(t) \\ z(t_0) = z_0 \end{cases} \implies z(t) = z_0 + \int_{t_0}^{t} f.$$

But in practice, you usually do this without introducing z.

§3.1 Linear Equations

Definition 3.2. A linear ODE (of order n) is of the form

$$y^{(n)} + p_{n-1}(t)y^{(n-1)} + \cdots + p_1(t)y' + p_0(t)y = f(t).$$

We will generally assume that all relevant functions $p_i(t)$ and $f(t)$ are continuous.

Remark 3.3. In some situations, we would have $p_n(t)y^{(n)}$. If it's nonzero, then we can simply divide by it; but if it sometimes vanishes, then our theory will break.

Note that this equation is linear in y, not in t. (These technically should be called *affine* because of the constant on the right-hand side, which isn't exactly linear; if $f(t) = 0$, we call the equation *homogeneous*.)

More explicitly, the operator

$$L(y) = y^{(n)} + p_{n-1}(t)y^{(n-1)} + \cdots + p_0(t)y$$

is a linear operator, because $L(\alpha y + \beta z) = \alpha L(y) + \beta L(z)$. An *operator*, or *functional*, is a function of a function; here L is a function

$$L\colon \{n+k \text{ times differentiable functions}\} \to \{k \text{ times differentiable functions}\}.$$

(In analysis, it's important to be precise about which space an operator is acting on, since its properties may be very different depending on the space.)

Linear equations are nice because of the *principle of superposition* — if $L(y) = f(t)$ and $L(z) = g(t)$, then $L(\alpha y + \beta z) = \alpha f(t) + \beta g(t)$. This lets us prove that the solution space is affine, and we'll be able to describe its dimension and sometimes do geometry to it.

§3.2 First Order Linear ODEs

A *first-order* linear ODE is a linear ODE with $n = 1$.

Question 3.4. Solve $y' + p(t) \cdot y = f(t)$.

The associated homogeneous equation is $y' + p(t) \cdot y = 0$.

Theorem 3.5

The solution to the homogeneous linear first-order ODE with initial condition $y(t_0) = y_0$ is

$$y(t) = y_0 \exp\left(-\int_{t_0}^{t} p\right) = y_0 e^{-\int_{t_0}^{t} p}.$$

Proof. It's possible to simply differentiate the formula above, but we won't do this (so that we can see how to find it). Define $P(t) = \int_{t_0}^{t} p$, so that $P'(t) = p(t)$. Then we have

$$\frac{d}{dt} e^{P(t)} \cdot y(t) = (y'(t) + p(t)y(t))e^{P(t)} = 0$$

by our equation. This means $e^{P(t)}y(t)$ is a constant; and we can use the initial condition to obtain the constant — we should have

$$e^{P(t_0)} \cdot y(t_0) = y_0$$

equal to the constant.

Note that this equation is separable, so we could solve it; the equation then becomes $\frac{y'}{y} = pp(t)$, which we know. □

Example 3.6

Suppose $p(t) = \alpha \in \mathbb{R}$, so we want to solve $y' + \alpha y = 0$ and $y(t_0) = y_0$. The solution is $y(t) = y_0 e^{-\alpha(t-t_0)}$. This extends to other settings, and it'll be important later.

We will use the same trick to solve the linear ODE iin general — we have

$$\frac{d}{dt} e^{P(t)} y(t) = (y'(t) + p(t)y(t)) e^{P(t)} = e^{P(t)}.$$

Now letting $u = e^{P(t)} y(t)$, we have $u'(t) = e^{P(t)} f(t)$. This means

$$u(t) = y(t_0) + \int_{t_0}^{t} e^{P(s)} f(s)\, ds.$$

Theorem 3.7

The unique solution to the linear ODE with $y(t_0) = y_0$ is

$$y(t) = y_0 e^{-P(t)} + e^{-P(t)} \int_{t_0}^{t} e^{P(s)} f(s)\, ds.$$

This is a separation we will see quite a lot — one part only depends on the initial conditions, and the other only depends on the left-hand. So we can write this equation as $y = y_h(t) + y_r(t)$, where

$$\begin{cases} y_h'(t) + p(t) y_h(t) = 0 \\ y(t_0) = y_0 \end{cases} \quad \text{and} \quad \begin{cases} y_r' + p(t) y_r = f(t) \\ y(t_0) = 0, \end{cases}$$

and solving these two cases is essentially separate. So you can separate the equation into these two simpler equations.

§3.3 First Big Theorem

Question 3.8. When is there a unique solution

$$\begin{cases} y'(t) = f(t, y(t)), \\ y(t_0) = y_0? \end{cases}$$

In recitation, we saw a counterexample:

Example 3.9

Consider the equation $y'(t) = y^{2/3}(t)$, with the initial condition $y(0) = 0$. Then there is *not* a unique solution.

Proving uniqueness is usually not as hard, but proving existence is — for general f, constructing a solution is really hard. One method (the Euler method) is to try approaching the solution by an approximate solution. Assume some reasonable conditions on the differentiability of f (it's differentiable and its derivative is continuous). Then for small ε, we have

$$y(t_0+\varepsilon) \approx y_0 + \varepsilon y'(t_0) + \text{negligible} = y_0 + \varepsilon f(t_0, y_0) + \text{neglibib;e.}$$

Similarly, $y(t_0+2\varepsilon) = y(t_0+\varepsilon+\varepsilon) = y(t_0+\varepsilon) + \varepsilon f(t_0+\varepsilon, y(t_0+\varepsilon)) + \cdots$. Under reasonable conditions, this scheme will converge.

§4 February 13, 2023

§4.1 Existence and Uniqueness

Since the beginning of the class, we've been considering whether solutions to our equations exist and are unique. Both are important (for being able to understand the solutions).

Question 4.1. Given a first-order ODE of the form

$$\begin{cases} y'(t) = f(t, y(t)), \\ y(t_0) = y_0, \end{cases}$$

does a solution exist, and is it unique?

The answer is not always yes — there are a lot of examples without uniqueness.

Example 4.2

Consider the ODE where

$$f(t,y) = \begin{cases} 0 & \text{if } y \le 0 \\ \sqrt{2y} & \text{if } y > 0. \end{cases}$$

This is a separable equation, so we can solve it; for all $a \ge 0$, the function

$$y_a(t) = \begin{cases} \frac{(t-a)^2}{2} & \text{if } t \ge a \\ 0 & \text{otherwise} \end{cases}$$

is a solution to the ODE

$$\begin{cases} y' = f(t,y) \\ y(0) = 0. \end{cases}$$

So in this situation, we don't have a unique solution.

When we don't have uniqueness, we often get very bad properties for our solutions.

Examples where a solution does not *exist* are harder to construct, but they exist as well:

Example 4.3

Consider the ODE where

$$f(t,y) = \begin{cases} 1 & \text{if } ty < 0 \\ -1 & \text{if } ty \ge 0, \end{cases}$$

with $y(0) = 0$. It is left as an exercise to show that this does not have a solution.

So there are weird functions f for which we don't have existence, and there are even less weird functions f where we do not have uniqueness — in the previous example, we have solutions which stay at 0 up to any point and start growing as a parabola. (The point is that the square-root function has bad behavior at 0.)

Question 4.4. Under which reasonable assumptions do we have existence and uniqueness of the solution to the ODE close to $t_0 \in \mathbb{R}$ and $y_0 \in \mathbb{R}$?

We will care about *local* solutions — we consider (t, y) inside a neighborhood of the form $[t_0 - a, t_0 + a] \times [y_0 - b, y_0 + b]$. We'll later see how local existence and uniqueness can be used to obtain global results as well (by gluing local solutions together).

By *reasonable*, we want a property that is easy to check and is satisfied by a lot of functions. The key to this, as well as to a lot of problems in analysis, is *regularity* of functions.

§4.2 Regularity of Functions

Definition 4.5. A C^k function is a function which is k times differentiable, and whose kth derivative is continuous.

Example 4.6

A C^0 function is a continuous function; a C^1 function is one which is differentiable and whose derivative is continuous.

Note that $C^k \implies C^{k-1} \implies \cdots \implies C^0$ — the larger k is, the more regular the function has to be.

Definition 4.7. A C^∞ function is one in C^k for all k.

These definitions have a bad feature, which is that they are not quantitative, in that we don't have *compactness* — a sequence of C^k functions might lose regularity in the limit (i.e., a sequence of C^k functions can converge to a function not even in C^0).

Example 4.8

Imagine a sequence of functions which become increasingly vertical in the middle (i.e., whose middle part is like rotated x^n for n odd). These converge to a white line which is not continuous (a line at $+1$ on the left and -1 on the right).

This is because we are allowing the derivative to go to ∞ in the middle — so our derivative is getting worse and worse. This is bad because one way to construct a solution is to take a limit; and here taking limits doesn't behave well.

So we will require our functions to satisfy a property called *Lipschitzness*.

Definition 4.9. A function $f\colon I \to \mathbb{R}$ (for an interval I) is *L-Lipschitz* (for some $L > 0$) if for all $x, y \in I$, we have

$$|f(x) - f(y)| \le L\,|x - y|\,.$$

This is a very important definition. The intuition is that the slope of f is at most L (the slope of the function is $\frac{f(x)-f(y)}{x-y}$ — when we let $y \to x$, we end up with the derivative). So this prevents us from approaching a vertical line.

This is in some sense a weaker assumption than being C^1; but it is also stronger in that it is *quantitative*. In particular, it is satisfied by taking limits.

Proposition 4.10

If f is C^1 on I, then f is L-Lipschitz on I if and only if $|f'(x)| \leq L$ for all $x \in I$.

(This is left as an exercise.)

Proposition 4.11

If f is L-Lipschitz, then f is continuous.

However, Lipschitz does not imply C^1. For example, the function $x \mapsto |x|$ is 1-Lipschitz on $\mathbb{R}$, but it is not differentiable. (In fact, Lipschitz does imply that f is differentiable *almost* everywhere.)

Example 4.12

The map $y \mapsto \sqrt{2y}$ is *not* L-Lipschitz for any L on the interval $(0, 1]$.

(This can be proven using the first criterion.)

Example 4.13

The map $y \mapsto \sqrt{2y}$ is L-Lipschitz on any interval $[\varepsilon, \infty)$ for some L.

So $\sqrt{2y}$ behaves differently at 0 and away from 0; this makes the difference in our example. The Lipschitz condition will be the core of our property for existence and uniqueness.

Now we will expand this definition to a function of two variables (which is the setting we are in).

Definition 4.14. For a set $D \subseteq \mathbb{R}^2$ and a function $f \colon D \to \mathbb{R}$ sending $(t, y) \mapsto f(t, y)$, we say that f is L-Lipschitz with respect to the 2nd variable if for all t_0, y_1, and y_2 (such that (t_0, y_1) and (t_0, y_2) are in D), we have

$$|f(t_0, y_1) - f(t_0, y_2)| \leq L\,|y_1 - y_2|\,.$$

This is essentially the same definition — we need $f(t_0, y)$ to be L-Lipschitz (as a function of y) for all t_0.

Often, the way we prove Lipschitz conditions is by using the first criterion from the proposition.

Proposition 4.15

If $f \colon D \to \mathbb{R}^2$ is C^1, then f is L-Lipschitz with respect to the 2nd variable on D if and only if

$$\left|\frac{\partial f}{\partial y}\right| \leq L \text{ on } D.$$

§4.3 Control of Solutions

The reason we care about Lipschitz functions is they let us control solutions to ODEs.

Take

$$\begin{cases} y' = f(t, y) \\ y(t_0) = y_0 \end{cases},$$

and suppose that f is L-Lipschitz with respect to the second variable on an interval I (where $t_0 \in I$).

Suppose we have two such solutions y_1 and y_2. We claim that they they must be equal. Here we can take their difference — define

$$z = y_1 - y_2.$$

Then z satisfies

$$z' = f(t, y_1(t)) - f(t, y_2(t))$$

and $z(t_0) = 0$.

Now f is L-Lipschitz with respect to the 2nd variable if and only if $z^1(t) \leq L\,|y_1(t) - y_2(t)|$, and similarluy $L\,|y_1(t) - y_2(t)| \leq z'(t)$. This means

$$L\,|z(t)| = z'(t) \leq L(|z|\,(t)).$$

We will show that suhc an inequialty to gether with the initial condition implis that z is 0 (whichis what we awant).dividing we have

$$-L \leq \frac{z'(t)}{|z(t)|} \leq L.$$

We will assume for contradiction that $z(t) \neq 0$, and try to obtain a contradiction.

Assume for contradiction that $z(t) > 0$ for $t > t_0$ close to t_0.

Wenow have

$$-L < \frac{z'(t)}{z(t)} < L.$$

We can integrate; suppose we integrate between s and t for some $s < t$ which we'll choose later. Then w ehave

$$-L(t-s) \leq \log(z(t)) - \log(z(s)) \leq L(t-s).$$

Now we want to take an exponential, and we get

$$e^{-L(t-s)} \leq \frac{Z(t)}{z(s)} \leq e^{L(t-s)}.$$

Now we can multiply by $z(s) > 0$, to get tjhat

$$z(s) \cdot e^{-L(t-s)} \leq z(t) \leq z(s) \cdot e^{L(t-s)}.$$

Now $s \mapsto z(s)$ is continuous at t_0, which means as $s \to t_0$, we have $z(s) \mapsto z(t_0)$. But we also have $z(t_0) = 0$, so then $0 \leq z(t) \leq 0$. This contradictis our assumption that $z(t) > 0$ for all d $t > t_0$.

In the cases $z(t) < 0$ for $t > t_0$, and all others, we get a contradictio nsimilarluy.

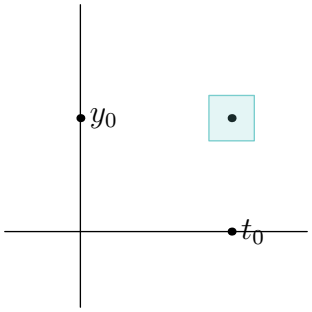
So uniqueness is not complicated if we assume Lipschitz. The rough idea is to bound our solutions and show that they have to vanish.

§5 February 15, 2023

As before, we're considering the existence and uniqueness of the solution to

$$\begin{cases} y' = f(t, y) \\ y(t_0) = y_0. \end{cases}$$

We will first focus on finding a *local* solution; later we will see how to extend it to a global one.



We will focus on a small rectangle $[t_0 - a, t_0 + a] \times [y_0 - b, y_0 + b]$; we're trying to find a solution in the rectangle (this is the key of the problem).

We will need two regularity assumptions — we assume we have $a, b, L > 0$ (where a, b are as above and L is a Lipschitz constant) such that:

- The function $t \mapsto f(t, y)$ is continuous on $[t_0 - a, t_0 + a]$ for all $y \in [y_0 - b, y_0 + b]$.
- The function $y \mapsto f(t, y)$ is L-Lipschitz on $[y_0 - b, y_0 + b]$ for all (fixed) $t \in [t_0 - a, t_0 + a]$.

(Note that the second condition is stronger — all Lipschitz functions are continuous. In fact, any function which is differentiable on a compact interval is Lipschitz (because it must have bounded derivative). (It's not true if the interval isn't compact — for example, $\sqrt{y}$ is not Lipschitz on $(0, 1]$.) The converse is false — not all Lipschitz functions are differentiable, for example $y \mapsto |y|$ — but it is true that any Lipschitz function is differentiable almost everywhere.)

§5.1 Existence–Uniqueness Theorem

In the next week or so, we'll prove the big theorem (we'll call it the Existence–Uniqueness theorem, but it may also be called the Cauchy–Lipschitz theorem or Picard–Lindelöf theorem; some of these may use slightly different assumptions, but all the proofs use the same technique, of the fixed point theorem; the method we'll use is the one that Picard used).

We'll call our ODE (IVP) (for *initial value problem*).

Theorem 5.1

If f satisfies the two assumptions above, then there exists c with $0 < c \leq a$ for which (IVP) has a unique solution on $[t_0 - c, t_0 + c] \subseteq [t_0 - a, t_0 + a]$.

So we may need to restrict our segment, but it doesn't really matter — right now we're just looking for a local solution, and we'll see how to make it global later.

The proof has two parts — existence and uniqueness. Uniqueness is much easier (and we did it last class):

Proof of Uniqueness. Assume we have two solutions $y_1(t)$ and $y_2(t)$; we'll prove that they must be equal. Define

$$z = (y_1 - y_2)^2.$$

(This avoids having to deal with cases.) Then the second assumption implies that

$$-2Lz \leq z' \leq 2Lz,$$

and we can conclude in the same way as last time.

Exercise 5.2. Show that this implies $z = 0$, using $z(t_0) = 0$.

□

Meanwhile, the existence proof uses techniques that many people use daily in their research.

§5.2 Existence — General Strategy

The technique we will use is called *Picard's iteration*; but the main idea is to recast the problem as a fixed point problem.

If we have $y' = f(t, y)$ and $y(t_0) = y_0$, then we can write down $y(t)$ by integrating the equation; by the fundamental theorem of calculus we must have

$$y(t) = y_0 + \int_{t_0}^{t} y' = y_0 + \int_{t_0}^{t} f(s, y(s))\, ds.$$

(This integral exists because $f(s, y(s))$ is continuous.) Of course, this doesn't prove that a solution exists, because the definition is recursive — we have y on both sides, and we can't define y using y. But we've rewritten the problem in a different way, where we're integrating instead of taking a derivative; and integrating is better for obtaining quantitative controls (as we will see).

The key idea is to define a functional $F\colon C^1([t_0 - a, t_0 + a]) \to C^1([t_0 - a, t_0 + a])$ by

$$F(y) = y_0 + \int_{t_0}^{t} f(s, y(s))\, ds.$$

First we need to check that this is well-defined — since y is in C^1, then $y(s)$ is C^1, so $f(s, y(s))$ is C^0 (the composition of continuous functions is continuous). So integrating gives a function which is differentiable, and whose derivative is the expression inside (which is continuous).

(Taking an integral of something raises its regularity — $F(y)'(t) = f(t, y(t))$, which is continuous.)

Now the key point is that we can restate (IVP) — y is a C^1 solution to (IVP) if and only if $y = F(y)$.

(Note that y must be in C^1 — y must be differentiable with $y' = f(t, y)$, which is continuous.)

This is because the boxed equation obtained from integrating (IVP) says precisely this.

This is a *fixed point equation* — we're looking for a value y such that $F(y) = y$. (Note that y is actually a function in C^1, so F is actually a functional.)

As a simpler example of looking for a fixed point:

Example 5.3

If $f\colon [0, 1] \to [0, 1]$ is a C^0 function, then a fixed point is an intersection between the graph of f and the graph of the identity, i.e., the line $y = x$.

There's a lot of theory for fixed points of functions; here we will also prove a fixed point theorem, but working on huge spaces of functions instead of intervals of the reals.

§5.3 Fixed Point Problem

One way to find a fixed point is by starting anywhere, and then applying f over and over — given a function $f\colon [0, 1] \to [0, 1]$, we can let $f(x_0) = x_1$, then $f(x_1) = x_2$, and so on. If we keep iterating our function over and over, then since $[0, 1]$ is a compact interval, some subsequence of these points will converge, and we will get a fixed point.

For our actual setting, we perform *Picard's iterations* in the following way:

- We start with the constant function $y_0(t) = y_0$ for all $t \in [t_0 - a, t_0 + a]$.
- We define $y_1(t) = F(y_0)(t) = y_0 + \int_{t_0}^{t} f(s, y_0)\, ds$.
- We then define $y_2(t) = F(y_1)(t)$, and so on, with $y_{n+1} = F(y_n)$ for all $n \in \mathbb{N}$.

The point is that fixed points over $[0, 1]$ exist because it's a box; but the space of C^1 functions is huge. The Lipschitz function will allow us to do something similar. (This is true in a much more general setting — if you have a function from a compact topological space to itself, it must have a fixed point for the reason we'll see here.)

The goal is to find a sublimit of our sequence of points (functions) and show that it's a fixed point. Assume that as $n \to \infty$, y_n converges in some sense we will define later to a function y_∞. In particular, $y_{n+1} \to y_\infty$ as well. Assume also that $y \mapsto F(y)$ is 'continuous' in some sense we will define later, that allows us to inverse the limit and the function. We know that $y_{n+1} = F(y_n)$. We know $y_{n+1} \to y_\infty$ (in a sense we don't understand yet); meanwhile by 'continuity' and convergence, we also have $F(y_n) \to F(y_\infty)$. So then we get

$$y_\infty = F(y_\infty),$$

which means y_∞ is a fixed point.

So what we really need to do is prove that we can extract a limit; then whenever we have a limit, we have a fixed point. (We are not going to find the limit, but we will show it exists.)

Remark 5.4. Compactness means that whenever you have a sequence of elements, you can take a subsequence that converges to something — K is compact if for all sequences $(y_n) \in K$, there exists some subsequence $(y_{\sigma(n)})$ such that $y_{\sigma(n)}$ converges to some point y_∞ as $n \to \infty$. By a subsequence, we mean that σ is a function $\mathbb{N} \to \mathbb{N}$ which is strictly increasing. (We are not really going to work with the general definition here.)

(The reason you want compactness is to take limits; but there are also other definitions.)

Exercise 5.5. Write the iterations for

$$\begin{cases} y' = y \\ y(0) = 1. \end{cases}$$

You should find that $y_n(t)$ is the order-n Taylor series of the exponential.

§6 February 17, 2023

Recall that we are trying to find a unique local solution to the ODE

$$\begin{cases} y' = f(t, y) \\ y(t_0) = y_0. \end{cases}$$

The strategy is to see this as equivalent (as long as y is C^1, which must be true if f is continuous) to saying that $y = F(y)$, where F is the *functional* $F \colon C^1 \to C^1$ defined as

$$F \colon y \mapsto \left(t \mapsto y_0 + \int_{t_0}^{t} f(s, y(s))\, ds \right).$$

(Here all functions are from some interval to $\mathbb{R}$.) This is a *fixed point equation* — almost all results about existence will be proven through some fixed point procedure.

The idea is to consider the sequence $(y_n)_{n\in\mathbb{N}}$ defined by $y_0(t) = y_0 \in \mathbb{R}$ and $y_{n+1} = F(y_n)$. The key idea is that if $y_n \to y_\infty$ (for some sense of convergence we will look at today) and we have the correct continuity properties for F (with respect to our notion of convergence that we haven't decided on), then we have $y_{n+1} = F(y_n)$, and we can take a limit of both sides (using the continuity of F) to obtain that $y_\infty = F(y_\infty)$. So y_∞ is a fixed point.

Today we'll define the notion of convergence. For numbers, we know what it means for a number to approach another. But for functions, there are *many* notions of convergence; today we will define a few and see which one we want to use to replace our unknown convergence function.

§6.1 Background on Convergence of Functions

Question 6.1. Our goal is to find a notion of convergence that is 'well-behaved' — in particular, if a sequence f_n of functions in C^0 converge to some function f_∞, then f_∞ is continuous. (This is called the *compactness* property of continuous functions.) We would also like the same to be true for C^1.

§6.1.1 Pointwise Convergence

The most natural notion of convergence is *pointwise* convergence:

Definition 6.2. Suppose we have a sequence $f_n\colon [\alpha,\beta] \to \mathbb{R}$ of functions (for all $n \in \mathbb{N}$). We say that $(f_n)_{n\in\mathbb{N}}$ *converges pointwise* to f_∞ if for all $x \in [\alpha,\beta]$, $f_n(x)$ converges to $f_\infty(x)$.

This is fairly natural, but it has a problem: it doesn't preserve continuity (or C^k for any $k \geq 0$), as seen in the following counterexample:

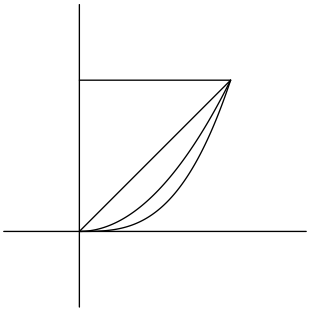
Example 6.3

Take the sequence of functions $f_n\colon [0,1] \to \mathbb{R}$ defined by $f_n(x) = x^n$. Then:

- For all $x \in [0,1)$, we have $f_n(x) \to 0$ as $n \to \infty$.
- For $x = 1$, we have $f_n(1) = 1$ for all n, which means $f_n(1) \to 1$ as $n \to \infty$.

This means the sequence (f_n) converges to the function

$$f_\infty(x) = \begin{cases} 0 & x \in [0,1) \\ 1 & x = 1. \end{cases}$$



Clearly f_∞ is not continuous. So this notion of convergence doesn't work for our purposes, since it loses all regularity (we can take a sequence of things in C^0 — even C^∞ — which converge to something not C^0).

§6.2 Uniform Convergence

So we need to improve this notion; this improved notion will be *uniform convergence*. The big difference is that it has a quantitative nature (similarly to how Lipschitz is a quantitative version of continuous).

Definition 6.4. Given a sequence of functions $f_n\colon [\alpha,\beta] \to \mathbb{R}$ for all $n \in \mathbb{N}$, the sequence $(f_n)_{n\in\mathbb{N}}$ *converges uniformly* if $\sup_{x\in[\alpha,\beta]} |f_n(x) - f_\infty(x)| \to 0$ as $n \to \infty$.

In the notion of pointwise convergence, we ask that for every x, we have $f_n(x) \to f_\infty(x)$. In uniform convergence, we ask for the same thing, but further that we approach the limit at the same rate. This requires the *whole segment* at once to get close to f_∞, rather than the individual points getting close at their own (possibly arbitrarily very slow) rates.

Example 6.5

We'll show that the functions $x \mapsto x^n$ on $[0,1]$ do not converge uniformly to f_∞. We have

$$\sup_{x\in[0,1)} |f_n(x) - f_\infty(x)| = \sup_{x\in[0,1)} |x^n - 0| = 1,$$

since for every finite n we can take $x \to 1$ (we're ignoring the point 1, since the difference there is always 0). Clearly this does not converge to 0.

The *supremum* is a sort of generalization of the maximum (which may not be defined for an infinite set):

Definition 6.6. The supremum $\sup_{x\in E} f(x) = \sup\{f(x) \mid x \in E\}$ is defined as the smallest number $a \in \mathbb{R}$ such that $f(x) \le a$ for all $x \in E$.

You can sort of think of it as a maximum that isn't necessarily attained. (In particular, you can think of it as a maximum of limits of sequences in your set.)

Exercise 6.7. Prove that uniform convergence implies pointwise convergence.

(This implies that if our sequence had a limit, then it would have to be f_∞; so by pointwise convergence we knew what the limit would be. This is how you prove uniform convergence — you first prove pointwise convergence to figure out the limit, and then you check that this supremum does indeed go to 0.)

Exercise 6.8. Show that the sequence of functions $f_n\colon [0,\frac{1}{2}] \to \mathbb{R}$ defined as $x \mapsto x^n$ converges uniformly to $f_\infty = 0$.

(This is true for any value strictly less than 1.)

Theorem 6.9

If we have a sequence of functions $f_n \in C^0([a,b])$ and $f_n \to f_\infty$ uniformly, then $f_\infty \in C^0([a,b])$ as well.

Proof. Intuitively, continuity means that $f_\infty(x) - f_\infty(y)$ should go to 0 as $x \to y$. To try to show that it goes to 0, we'll try to relate both $f_\infty(x)$ and $f_\infty(y)$ to $f_n(x)$ and $f_n(y)$ — we know $f_n(x)$ approaches $f_\infty(x)$ and $f_n(y)$ approaches $f_\infty(y)$, and we know $f_n(x)$ approaches $f_n(y)$. So we'll rewrite our difference to obtain three quantities all of which we can control. We can write

$$f_\infty(x) - f_\infty(y) = (f_\infty(x) - f_n(x)) + (f_n(x) - f_n(y)) + (f_n(y) - f_\infty(y)).$$

Because $f_n \to f_\infty$ we must have $f_\infty(x) - f_n(x) \to 0$; similarly $f_n(y) - f_\infty(y) \to 0$ as well. Meanwhile, as $x \to y$ we have $f_n(x) - f_n(y) \to 0$ because f_n is continuous. (The first two statements are as $n \to \infty$, while the last is as $x \to y$.)

This isn't rigorous because it isn't clear how the rates depend on each other (as we have different things going to 0 and their rates could depend on each other). So we'll do something else that corresponds to taking all the limits at once but not really.

Now we'll actually prove the statement. Continuity means that for all $\delta > 0$, there exists $\varepsilon > 0$ such that for all $x, y \in [a, b]$ with $|x - y| < \varepsilon$, then $|f_\infty(x) - f_\infty(y)| < \delta$.

Suppose we are given $\delta > 0$; we will show that for this particular δ we can find such an ε.

We have three terms, and we want their sum to be less than δ. So we will ask for each of these three things to be less than $\delta/3$.

- There exists N (depending on δ) such that $\sup_{z \in [a,b]} |f_N(z) - f_\infty(z)| < \delta/3$. (This is true because f_n converges uniformly to f_∞, so this expression converges to 0, which means by choosing N large enough we can make it arbitrarily small.) This takes care of the first and last term. Now fix this value of N.
- Then there exists ε (depending on N and δ) such that if $|x - y| < \varepsilon$ then $|f_N(x) - f_N(y)| < \delta/3$. (This is true because this specific function f_N is continuous.)

Now choose this value of ε. Then if $|x - y| < \varepsilon$, then

$$|f_\infty(x) - f_\infty(y)| \le |f_\infty(x) - f_N(x)| + |f_N(x) - f_N(y)| + |f_N(y) - f_\infty(y)|.$$

The first term is less than $\delta/3$ by the first property with $z = x$, and the third term is less than $\delta/3$ by the first property with $z = y$; the second is true by the second property. So then $|f_\infty(x) - f_\infty(y)| < \delta$, as desired. □

Next time we will show that in our sequence $y_{n+1} = F(y_n)$, the sequence y_n converges uniformly to some function y_∞, and y_n' also converges uniformly to y_∞'. (We will call this C^1 convergence; but it's not as useful.) We will also prove that F is continuous, meaning that if $z_n \to z_\infty$ uniformly and $z_n' \to z_\infty'$ uniformly as well, then $F(z_n) \to F(z_\infty)$ and $F(z_n)' \to F(z_\infty)'$.

§7 February 21, 2023

§7.1 Remarks on Continuity

Definition 7.1. A function f is C^0 on I if for all $y \in I$ and $\delta > 0$, there exists $\varepsilon > 0$ (depending on both y and δ) such that for all $x \in I$, with $|x - y| < \varepsilon$, we have $|f(x) - f(y)| < \delta$.

Definition 7.2. A function f is *uniformly* C^0 on I if for all $\delta > 0$, there exists ε (depending on δ) such that for all x and y in I with $|x - y| < \varepsilon$, $|f(x) - f(y)| < \delta$.

We used the definition of uniform continuity in our proof yesterday. This isn't generally the same as being continuous at every point; but if $I = [a, b]$, the two notions are equivalent (usually only the second implies the first). The main difference between them is that in the first case, ε may depend on y, while in the second the same ε must work for all y. (Our proof could also have been used to assume (1) and prove (1).)

§7.2 Convergence of Functions

Recall that we're considering the ODE

$$\begin{cases} y' = f(t, y) \\ y(t_0) = y_0 \in \mathbb{R}, \end{cases} \tag{IVP}$$

where f is continuous with respect to t and Lipschitz with respect to y. We're approaching this using *fixed points* — y being a solution to (IVP) is equivalent to $y = F(y)$, where $F\colon C^1 \to C^1$ is defined as

$$F(z) = y_0 + \int_{t_0}^{t} f(s, z(s))\, ds.$$

The method we'll use is iteration — we begin with $y_0(t) = y_0 \in \mathbb{R}$ and define $y_{n+1} = F(y_n)$. We want to show that the sequence of functions y_n converges, and that we can say certain things about its limit.

Today we'll see a bunch of technical lemmas controlling the sequence (y_n).

Let $I = [t_0 - a, t_0 + a]$ where $a > 0$ be the interval of t we're considering (to start with — we may shrink it later), and $[y_0 - b, y_0 + b]$ be the interval around y_0. (We assume a is small enough that we're always in the correct interval $[y_0 - b, y_0 + b]$, by the continuity of y.)

We'll begin by showing that y_n can't go too far away from the initial function y_0.

Lemma 7.3

For all $t \in I$ and $n \in \mathbb{N}$, we have

$$|y_n(t) - y_0| \le K\,|t - t_0|\,,$$

for $K = \max|f|$ (on the interval $[t_0 - a, t_0 + a] \times [y_0 - b, y_0 + b]$).

Proof. We know $y_n(t_0) = y_0$, and

$$|y_n'(s)| = |f(s, y_{n-1}(s))| \le K$$

for all s. Then $|y_n(t) - y_0| \le K\,|k - t_0|$ by integration. (By choosing a small enough we can assume $|t - t_0| < b$.) □

The next estimate is the main one — it tells us that (y_n) is not just bounded, but is in fact a *Cauchy sequence* (in a C^0 sense; then we'll estimate the derivatives easily).

Lemma 7.4

For all $t \in I$, and for all $n \in \mathbb{N}$, we have

$$|y_{n+1}(t) - y_n(t)| \le KL^n \cdot \frac{|t - t_0|^{n+1}}{(n+1)!},$$

if f is L-Lipschitz in y on our interval $I \times [y_0 - b, y_0 + b]$.

This is a kind of estimate you'll have to prove often in order to show that a sequence converges. In some sense, we want to say that y_n is close to y_{n+1}; but more importantly we want the sum of these errors to stay bounded. The key point is that for all $N > n$, we can (using the triangle inequality) bound

$$|y_N(t) - y_n(t)| \le |y_N(t) - y_{N-1}(t)| + \cdots + |y_{n+1}(t) - y_n(t)| \le \frac{K}{L} \sum_{k=n}^{N-1} \frac{L^{n+1}\,|t - t_0|^{n+1}}{(n+1)!}.$$

This sum looks like the expansion of the exponential; in particular, this is bounded by $\frac{K}{L}\sum_{k=n}^{\infty}$ of the same expression, which goes to 0 as $n \to \infty$ (because the Taylor series of the exponential converges everywhere on $\mathbb{R}$ — this expression comes from the Taylor series of $e^{L|t-t_0|}$).

The important thing about the lemma is that the expression we have is not just small, but in fact small enough to sum.

Definition 7.5. A Cauchy sequence is a sequence such that $|y_N(t) - y_n(t)|$ is bounded by a function of n that goes to 0 as $n \to \infty$.

(We'll come back to this later.) So we are showing that the sequence y_n is a Cauchy sequence.

Proof. We'll prove this by induction on n. We begin with the base case $n = 0$, where this statement becomes

$$y_1(t) - y_0(t) = \int_{t_0}^{t} f(s, y_0)\, ds.$$

We'll assume for convenience that $t > t_0$ (otherwise exchange the bounds of the integral); then

$$|y_1(t) - y_0(t)| \le \int_{t_0}^{t} |f(s, y_0)|\ ds \le K\,|t - t_0|\,,$$

since $f(s, y_0) \le K$ for all s. So the statement is true for $n = 0$.

Now let's try to iterate — assume that our lemma is true for all $k < n$, and we'll show that it's true for n; so we want to consider $|y_{n+1}(t) - y_n(t)|$. We can write this as

$$y_{n+1}(t) - y_n(t) = \int_{t_0}^{t} f(s, y_n(s))\, ds - \int_{t_0}^{t} f(s, y_{n-1}(s))\, ds.$$

(The two constants in the definition cancel out.) Assume again that $t > t_0$ (exchange the bounds otherwise); this gives

$$|y_{n+1}(t) - y_n(t)| \le \int_{t_0}^{t} |f(s, y_n(s)) - f(s, y_{n-1}(s))|\,.$$

Now using the fact that f is L-Lipschitz with respect to the second variable, the integrand is at most $L\,|y_n(s) - y_{n-1}(s)|$, giving the bound

$$|y_{n+1}(t) - y_n(t)| \le L \int_{t_0}^{t} |y_n(s) - y_{n-1}(s)|\,.$$

Now using the property at rank $n-1$, which gives an upper bound on the above expression, this is at most

$$L \int_{t_0}^{t} K L^{n-1} \frac{|s - t_0|^n}{n!}\, ds.$$

Now we can integrate this inequality, which gives us that this is at most

$$K L^n \frac{|t - t_0|^{n+1}}{(n+1)!},$$

as we wanted. So by induction, our lemma is true for all n. □

The next estimate we'll have is on the derivatives. (Its proof is essentially the same.)

Lemma 7.6

For all $t \in I$ and all $n \in \mathbb{N}$, we have

$$|y'_{n+1}(t) - y'_n(t)| \leq KL^n \frac{|t - t_0|^n}{n!}.$$

Proof. We have

$$y'_{n+1}(t) - y'_n(t) = f(t, y_n(t)) - f(t, y_{n-1}(t)).$$

Using the Lipschitz property, this is at most

$$L\,|y_n(t) - y_{n-1}(t)|\,,$$

and we can conclude by plugging in the main estimate. □

§7.3 Final Steps

The final steps involve finding a limit. This is a difficult question — even constructing $\mathbb{R}$ by hand is not easy, because you have to create real numbers which are limits of rational numbers without knowing they exist. What's behind all of this is what we call *completeness*.

What we want to do is show:

- There exists y_∞ such that $y_n \to y_\infty$ *uniformly* as $n \to \infty$.
- There exists z_∞ such that $y'_n \to z_\infty$ uniformly.
- $z_\infty = y'_\infty$.

Then we will be done. The *existence* is hard, because we have to construct them out of nothing. We'll use a key theorem for this.

Question 7.7. What is y_∞?

We don't yet know that the sequence converges, and we want to create a limit. Usually when we prove statements about limits, we're given a sequence of numbers and we want to find its limit; the way we do this is by finding what should be the limit, and showing that the difference goes to 0. But here we don't know what should be the limit — it's a very complicated object — so how do we say we converge?

First, what *should* y_∞ be? If y_∞ exists, then for all $t \in I$ we must have

$$y_\infty(t) = \lim_{n \to \infty} y_n(t)$$

(since uniform convergence implies pointwise convergence). This is how we'll construct y_∞ — we'll construct it one value at a time, and then we'll show that we indeed converge uniformly to y_∞; then we will do the same for z_∞.

The complicated part is proving the existence of this limit, which relies on the *completeness* of $\mathbb{R}$.

§7.4 Completeness and Cauchy Sequences

Definition 7.8. A sequence $(x_n)_{n \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ (i.e., a sequence of real numbers) is a *Cauchy sequence* if for all $\varepsilon > 0$, there exists $n \in \mathbb{N}$ (depending on ε) such that for all $N \geq n$, we have

$$|x_N - x_n| \leq \varepsilon.$$

This is the key tool when we want to construct something out of nothing — this is essentially what a sequence that *should* converge looks like, but we don't mention a limit — we don't need to assume we have a limit or know we have one. There's a deep theorem that every Cauchy sequence converges; then we're done, since we've shown essentially that our $y_n(t)$ are a Cauchy sequence for each n.

What we're essentially saying is that given ε, we can choose n such that after n, all our points x_N remain in the strip $(x_n - \varepsilon, x_n + \varepsilon)$ of size ε. The key point is that this is true for all ε — so if we now choose a much smaller ε', then the new $n(\varepsilon') = n'$ must be inside our strip, and now we must have a strip of size $2\varepsilon'$. We can keep choosing ε smaller and smaller, and this remains true; this gives us a bunch of strips included in each other where our sequence will always be after some point. What we want to say is that these strips must have a limit — the intersection of all of them.

We've never talked about any kind of limit, but this is a sequence that 'should' be converging. And if it's not converging, then we can define the limit as a new number and call that the limit — that's how you define $\mathbb{R}$ (by taking the limits of Cauchy sequences in $\mathbb{Q}$ — you take a sequence, which may not converge, and then you define a new real number for that limit).

§8 February 22, 2023

§8.1 Cauchy Sequences

Last class, we defined *Cauchy sequences*; the key point is that a Cauchy sequence is a sequence that *should* converge — its values lie in strips whose widths go to 0, so they should have *some* limit. Importantly, in defining Cauchy sequences we don't need to talk about this limit.

Definition 8.1. A *complete space* is a space where Cauchy sequences converge.

Example 8.2

Some examples of complete spaces:

- $\mathbb{N}$ is complete.
- $\mathbb{Q}$ is *not* complete.
- $\mathbb{R}$ is complete.
- Any compact set is complete (but the converse is not necessarily true, as $\mathbb{R}$ is not compact).

Remark 8.3. We won't define compactness in general. For our purposes (where we're only working in $\mathbb{R}^n$), a compact set is one of the two equivalent statements:

- Closed and bounded.
- A set K such that if (u_n) is a sequence in K, then there exists a subsequence $(u_{\sigma(n)})$ that converges.

We'll prove the first two statements.

Proof. To show that $\mathbb{N}$ is complete, the key point is that $\mathbb{N}$ is a discrete set, and so any Cauchy sequence must eventually become constant — take ε such that $2\varepsilon < 1$, and let $n = n(\varepsilon)$. Then we know $|x_N - x_n| < \varepsilon$ for all $N \geq n$, so x_N must stay within a strip that only includes one integer; this means $x_N = x_n$ for all $n \geq N$. So our sequence is eventually constant, and therefore has a limit.

To show that $\mathbb{Q}$ is *not* complete, the point is that $\mathbb{Q}$ has a lot of holes; these holes are real numbers. Rational numbers can approximate real numbers that are not rational. For example, let u_n be the first n digits of

$\sqrt{2}$ (or any other irrational number — it's not hard to prove $\sqrt{2}$ is not rational) — for example, $u_3 = 1.41$. Then $\lim_{n\to\infty} u_n = \sqrt{2}$ is not rational. (This is actually how we define the real numbers — by completing the space of rational numbers.) □

Student Question. *Is completeness equivalent to every subset having a supremum?*

Answer. If you have an order (in a way that makes sense for the notion of convergence), then yes; for real numbers, that's how you prove these things. But very often you won't have an order (e.g., on $\mathbb{R}^2$).

Now we'll see a sufficient condition for a sequence to be a Cauchy sequence (because it's what we'll use in the proof).

Proposition 8.4

If a sequence $(u_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ satisfies the property that for all n there exists $\varepsilon(n) > 0$ such that $|u_{n+1} - u_n| \le \varepsilon(n)$ and $\sum_{n\in\mathbb{N}} \varepsilon(n) < \infty$, then (u_n) is a Cauchy sequence.

Proof Sketch. For any $N \ge n$, we have

$$|u_N - u_n| \le |u_N - u_{N-1}| + \cdots + |u_{n+1} - u_n|$$

by the triangle inequality; then we can plug in the given bounds to conclude (this is left as an exercise). □

Theorem 8.5

Every Cauchy sequence of real numbers has a limit — i.e., $\mathbb{R}$ is complete.

This is a strong result because it's an *existence* result.

Proof Sketch. The proof idea will rely on the result that any monotone nondecreasing (or nonincreasing) and bounded sequence of real numbers has a limit. (This may not seem like a very big result, but it is — because it allows us to get something out of nothing.)

We'd now like to take our Cauchy sequence (u_n) and define a monotone subsequence — one way to do this is to define $v_n = \sup\{u_k \mid k \ge n\}$. This is a nonincreasing sequence, so it has a limit, called $\limsup u_n$. Similarly, we can define $w_n = \inf\{u_k \mid k \ge n\}$, which also has a limit, called $\liminf u_n$.

Then you can show that both are equal using the Cauchy property. □

§8.2 Constructing a Limit

Returning to our ODE problem, we've constructed a sequence

$$\begin{cases} y_0(t) = y_0 \\ y_{n+1}(t) = F(y_n), \end{cases}$$

where we defined

$$F(y) = y_0 + \int_{t_0}^{t} f(s, y(s))\, ds.$$

Our goal is to show that the sequences y_n converge — i.e., to construct y_∞ as the limit of our functions y_n (in the sense of uniform convergence).

We've seen that uniform convergence implies pointwise convergence, so in order to figure out what y_∞ should be, we can look at what function y_n converges to pointwise — if it exists, then y_∞ must be the pointwise limit of y_n, meaning that

$$y_\infty(t) = \lim_{n\to\infty} y_n(t)$$

for all t. (Here this limit is in the sense of $\mathbb{R}$.)

In order to prove that this limit exists, we will show that $y_n(t)$ is a Cauchy sequence; this means it must converge, and this allows us to construct y_∞ one value of t at a time.

Claim — For any fixed $t \in I$, the sequence $(u_n) = (y_n(t))$ is a Cauchy sequence.

Proof. We'll use the sufficient condition above and our main estimate, which says that

$$|u_{n+1} - u_n| \le KL^n \frac{|t-t_0|^{n+1}}{(n+1)!}.$$

Let this quantity be $\varepsilon(n)$. Then it suffices to verify that the sequence $\varepsilon(n)$ is summable. This is true — we have

$$\sum \varepsilon(n) = \frac{K}{L} e^{L|t-t_0|} < \infty.$$

So $(u_n) = (y_n(t))$ is a Cauchy sequence. □

This implies $(u_n) = (y_n(t))$ has a limit, because $\mathbb{R}$ is complete. Let this limit be $y_\infty(t)$. So now out of nothing, we've constructed the function y_∞.

§8.3 Uniform Convergence

There are a few steps remaining. First, we know that $y_n \to y_\infty$ pointwise. To show that $y_n \to y_\infty$ *uniformly*, we use the main estimate again: this tells us

$$|y_\infty(t) - y_n(t)| \le \sum_{k=n}^{\infty} KL^k \frac{|t-t_0|^{k+1}}{(k+1)!} \le \sum_{k=n}^{\infty} KL^k \frac{a^{k+1}}{(k+1)!},$$

where $I = [t_0 - a, t_0 + a]$. This expression doesn't depend on t — it's a uniform estimate, which is the key point — and it goes to 0 as $n \to \infty$ (this is equivalent to saying that the Taylor series of some exponential goes to 0 — since the entire sequence converges, its tail must go to 0). So then

$$\sup_{t\in I} |y_\infty(t) - y_n(t)| \to 0$$

as $n \to \infty$.

(The 'same rate' of uniform convergence means that we can bound the difference by something that doesn't depend on t — only depending on n.)

§8.4 Conclusions

We've shown that $y_n \to y_\infty$ uniformly. So this in particular means y_∞ is C^0 — we constructed y_∞ point-by-point by taking a bunch of limits, but in fact it's continuous because of uniform convergence.

Next, we want to show that it's C^1, and then that it satisfies the ODE.

Exercise 8.6. Show that there exists z_∞ such that $y_n' \to z_\infty$ uniformly, and consequently z_∞ is continuous.

This can be done with the same methods, but using the estimate on the derivatives instead of the main estimate. Next time we'll show that in fact $z_\infty = y_\infty'$, which implies y_∞ is C^1 and that $y_\infty'(t) = f(t, y_\infty(t))$ for all t, or equivalently that y_∞ satisfies the ODE (or $y_\infty = F(y_\infty)$).

§9 February 24, 2023

§9.1 Review

Today we'll finish our proof of the existence–uniqueness theorem. First we'll review the main steps of the proof, since they're important (and ones we'll see in many other problems).

We're trying to solve the ODE

$$\begin{cases} y' = f(t,y) \\ y(t_0) = y_0. \end{cases} \tag{IVP}$$

Our first step was to rewrite (IVP) as the *fixed point* problem $y = F(y)$, where we define

$$F(y)(t) = y_0 + \int_{t_0}^{t} f(s, y(s))\, ds.$$

In order to find a fixed point, we iterate F over and over — we define a sequence y_n with $y_0(t) = y_0$ and $y_{n+1} = F(y_n)$. (The idea is to show that y_n converges, and the point it converges to is a fixed point.)

Then we showed that (y_n) is Cauchy (in a C^1 sense — the function is Cauchy in a uniform sense, and so is its derivative). We did this using two main estimates controlling the consecutive differences $|y_{n+1} - y_n|$ — we had

$$|y_{n+1} - y_n| < \varepsilon(n)$$

for some $\varepsilon(n)$ such that $\sum_{n\in\mathbb{N}} \varepsilon(n) < \infty$. (We also showed the same statement for their derivatives.)

Then we can find a limit $y_\infty \in C^0$ such that $y_n \to y_\infty$ uniformly, and a limit $z_\infty \in C^0$ such that $y_n' \to z_\infty$ as well (this was left as an exercise).

The next step, which we'll do today, is to prove that $z_\infty = y_\infty'$. (This will imply that y_∞ is C^1.)

§9.2 Derivatives

Recall that $y_{n+1}(t) = y_0 + \int_{t_0}^{t} f(s, y_n(s))\, ds$. This implies

$$y_{n+1}'(t) = f(t, y_n(t)).$$

Now as $n \to \infty$, by pointwise convergence we have $y_{n+1}'(t) \to z_\infty(t)$ for all t. Meanwhile on the right-hand side, as $n \to \infty$ we have $f(t, y_n(t)) \to f(t, y_\infty(t))$ by the pointwise convergence of y_n and the continuity of f. This means

$$\boxed{z_\infty(t) = f(t, y_\infty(t))}.$$

Meanwhile, since we have

$$y_{n+1}(t) = y_0 + \int_{t_0}^{t} f(s, y_n(s))\, ds.$$

We have $y_{n+1}(t) \to y_\infty(t)$ by pointwise convergence as well, while

$$\int_{t_0}^{t} f(s, y_n(s))\, ds \to \int_{t_0}^{t} f(s, y_\infty(s))$$

by uniform convergence and the dominated convergence theorem. (Here we really need uniform convergence because we care about all $s \in [t, t_0]$.)

Remark 9.1. From the equation

$$y_\infty(t) = y_0 + \int_{t_0}^{t} f(s, y_\infty(s))\, ds,$$

we could have seen directly that y_∞ is C^1 — $y_\infty(s)$ is continuous, so $f(s, y_\infty(s))$ is continuous as well (as it is the composition of continuous functions), so the entire integral is C^1.

This means y_∞ is C^1, and differentiating the right-hand side gives

$$y_\infty(t) = f(t, y_\infty(t)) = z_\infty(t).$$

Now we have everything we want — in particular we have obtained $y_\infty'(t) = f(t, y_\infty(t))$, which is the equation we were trying to solve. We also have $y_\infty(t_0) = y_0$, so we're done.

Remark 9.2. The takeaway is that proving existence is complicated; so any result you have that gives you the existence of something is very important. (The fact that a monotone bounded sequence has a limit is a very powerful tool, because it gives the existence of something; we used this in going from (3) to (4), which was the hardest step.)

(You will see more of these ideas in functional analysis.)

§9.3 Examples

Now we'll see some applications of both existence and uniqueness.

Proposition 9.3

Suppose $f\colon \mathbb{R}^2 \to \mathbb{R}$ has the property that the functions $t \mapsto f(t, y)$ is continuous and $\mapsto f(t, y)$ is L-Lipschitz for some $L > 0$. Also suppose $f(-t, y) = -f(t, y)$ for all $(t, y) \in \mathbb{R}^2$. Then for all $y_0 \in \mathbb{R}$, the unique solution of the (IVP)

$$\begin{cases} y' = f(t, y) \\ y(0) = y_0 \end{cases}$$

is symmetric, meaning that $y(t) = y(-t)$ for all $t \in \mathbb{R}$.

This is important because if you act on a system in a symmetric way, you'd expect the answer to be symmetric as well. This isn't always true, but under our conditions it is.

This would generally be difficult, but there's a simple idea that solves it — we define the function on the right-hand side and show that it also satisfies (IVP), and then by uniqueness we're done.

Proof. Define the function $z(t) = y(-t)$, so our goal is to show that $y(t) = z(t)$ for all $t \in \mathbb{R}$. The main idea is to show that $y(t)$ and $z(t)$ solve the same (IVP); then by uniqueness, they must be the same function.

By assumption, we know

$$\begin{cases} y' = f(t, y) \\ y(0) = y_0. \end{cases}$$

Then we know

$$z(0) = y(-0) = y_0$$

as well, so the initial condition is satisfied. Meanwhile, by the chain rule and the symmetry condition on f we have

$$z'(t) = -y'(-t) = f(-t, y(-t)) = f(t, y(-t)) = f(t, z(t)).$$

(The first equality is by the chain rule, the second from the ODE on y, the third from the symmetry of f, and the fourth from the definition of z.)

So then z is a solution to the same ODE (IVP) — we have

$$\begin{cases} z' = f(t, z) \\ z(0) = y_0 \end{cases}$$

as well. So by uniqueness, the two solutions must be equal. □

(Note that the symmetry of f really should be called *antisymmetry* with respect to t.)

> **Remark 9.4.** This should be true for any other t_0 instead of 0, but it's harder to prove — the point is that instead of parametrizing solutions by their value at y_0, we could parametrize them by their values anywhere else, and this shouldn't affect anything. In that case, you'd want to bring the two solutions to 0 and match their values at 0.

(These arguments are really used in real life, by which we mean research in mathematics. We'll see many exercises of this form in the problem set.)

§9.4 Stability

The other important aspects of having a problem that's well-posed is that it should be *stable* with respect to perturbation. We'll look at two types of perturbation — the function f and the initial condition. (You can have both at once, but we might as well decompose into the two types.)

The first type of perturbation (Perturbation 1) — of the function f — is stated in terms of *parametrized equations*. This means we might have an ODE

$$\begin{cases} y' = f_a(t, y) \\ y(t_0) = y_0 \end{cases} \tag{PEa}$$

(where a is some parameter). Then we might look at what happens to the solutions of the ODE as a varies; we'd like them to not vary in a completely crazy way, at least if the map $a \mapsto f_a$ is nice enough.

The other type of perturbation (Perturbation 2) is of the initial conditions. Here we consider the perturbed ODE

$$\begin{cases} y' = f(t, y) \\ y(t_0) = y_0 + b \end{cases} \tag{ICb}$$

for $b \in \mathbb{R}$.

These are both perturbations of our original equation, and we'd like to see how our solutions change with respect to a and b.

Why are these important? For the first, the equations we have for modelling our physical or economic or social science phenomenon may not be the exactly true equation — it might just be right in some limit. (For example, maybe you're only looking at things at a large scale (e.g., general relativity) or small scale.) When you go from your perfect equation to real life, it might be a small perturbation of your model. You'd hope that the solution is stable when that happens.

So this is important when we have an imperfect model — if we have an imperfect model and it's not stable, then it doesn't tell us anything (because whatever we find in our perfect model won't tell us anything about real life).

For the second type, maybe you're trying to determine the weather in a week, but your measure of temperature is only accurate up to a hundredth of a degree; then we might have some error in our initial condition. So this type of perturbation is important when we have imperfect measures — we again don't want our solutions to vary too much depending on this parameter.

We'll see that often, especially for the second case, we'll have some exponential error which is small when it's small but then grows. This explains why you can only predict weather for a week, and then it becomes crazy — because you don't have any control of your ODE after large times.

First, we can reduce both problems to one problem — in the second problem, if we define $z = y - b$, then

$$z'(t) = y'(t) = f(t, y(t)) = f(t, z(t) + b).$$

This tells us z solves a new problem

$$\begin{cases} z' = g_b(t, z) = f(t, z + b) \\ z(t_0) = y_0. \end{cases}$$

This is now a problem o fthe form PEb, since we've moved the perturbation to the function rather than the initial value.

So we'll only look at the first situation, since it includes both.

§9.5 Grönwall's Lemma

We'll now see the main tool used to control ODEs when parameters vary. The intuition comes from the linear setting — suppose we have a linear equation

$$\begin{cases} y' = p(t)y \\ y(t_0) = y_0. \end{cases}$$

We want to say that somehow y is controlled. We've seen that the solutions are of the form $y(t) = y_0 exp\left(\int_{t_0}^t p\right)$, and here what we're interested in is control of this exponential.

If y_0 and $p(t)$ are both small, then the right-hand side will be small for some time.

Suppose $y \geq 0$ and $p(t) \geq 0$, and

$$\begin{cases} y'(t) \leq p(t)y(t) \\ y(t_0) = y_0 \geq 0. \end{cases}$$

Then this implies

$$y(t) \leq y_0 \exp\left(\int_{t_0}^t p\right)$$

for $t \geq t_0$.

Grönwell's lemma tells us something similar in general — it lets us have the same kind of control for a more flexible right-hand side.

Lemma 9.5 (Grönwall's Lemma)

Let $I \subset \mathbb{R}$ be an interval, and suppose that we have a function $t \mapsto p(t) \geq 0$. If $y(t) \leq C + \int_{t_0}^t p(s)y(s)\,ds$, then

$$y(t) \leq C \exp\left(\int_{t_0}^t p(s)\,ds\right).$$

This might not look much better than the original inequality, but it's much better because we only have y on one side — the original inequality doesn't tell us anything (y is controlled by a function of y). Here the key point is that there is no y — so we went from something that's not useful if we know nothing about y, to something that is. This kind of estimate is very precious, and we'll use it a lot — maybe not so much this year, but eventually.

As one place where we might see this sort of estimate, recall that we had

$$F(y) = y_0 + \int_{t_0}^t f(s, y(s))\,ds$$

in our proof of existence–uniqueness. But since f is Lipschitz, we can control $f(s, y(s)) \leq L\,|y(s)| + |f(s, 0)|$ if 0 is in the interval on whicih f is defined. Now we have something of the form in the lemma; Gronwell's solution would then give us some direct estimates on the solutions to our ODE.

Note that we allow dependence on s; this is useful because even if we're locally Lipschitz, we can allow a local Lipschitz constant $L(s)$ and that'll be enough.

§10 February 27, 2023

Today we'll mostly consider the question of stability; if we have time, we'll also start on global existence and uniqueness.

§10.1 Grönwall's Lemma

Theorem 10.1

Let $I \subseteq \mathbb{R}$ be an interval, $p\colon I \to \mathbb{R}_{>0}$ a function, and $t_0 \in I$ (and $C > 0$ a constant). If

$$0 \leq y(t) \leq C + \int_{t_0}^t p(s)y(s)\,ds,$$

then

$$y(t) \leq Ce^{\int_{t_0}^t p}.$$

Proof. Define the right-hand side of the original inequality as $z(t)$, so that we have $y(t) \leq z(t)$; we'll try to bound z from above, which also bounds y. Now

$$z'(t) = p(t) \cdot y(t) \leq p(t)z(t)$$

(since we have $y(t) \leq z(t)$ by assumption, and $p > 0$). We've already seen that this implies $z(t) \leq z(t_0)e^{\int_{t_0}^t p}$. (You can prove this by dividing by $z(t)$ on both sides, obtaining a separable inequality; we can then integrate both sides (the inequality is preserved under integration), which gives this.) This gives exactly what we want, since $z(t_0) = C$. So for all $t \in I$ we obtain

$$y(t) \leq z(t) \leq Ce^{\int_{t_0}^t p}.$$

□

The trick is really to introduce the function $z(t)$, show it satisfies one of these differential inequalities we've already seen, and conclude the upper bound on z.

This is a simple statement and proof, but it's quite powerful. One consequence is *continuous dependence on parameters*. We'll look at ODEs with parameters, as before; we'd like to say that under certain assumptions, if the parameters are close together, then so are the solutions.

Suppose we have two ODEs

$$\begin{cases} y_a' = f_a(t, y_a) \\ y_a(t_0) = y_0 \end{cases} \quad \text{and} \quad \begin{cases} y_b' = f_b(t, y_b) \\ y_b(t_0) = y_0 \end{cases}$$

(we saw last class that we can modify the equations to have the same initial condition). For a and b close enough to 0, and for all t close to t_0, we'd like to have a bound saying that $y_a(t)$ and $y_b(t)$ are close together.

Theorem 10.2 (C^0 Dependence on Parameters)

Consider the ODEs

$$\begin{cases} y_a' = f_a(t, y_a) \\ y_a(t_0) = y_0 \end{cases} \quad \text{and} \quad \begin{cases} y_b' = f_b(t, y_b) \\ y_b(t_0) = y_0 \end{cases}$$

such that:

- $t \mapsto f_a(t, y)$ is C^0 for all a and y;
- $y \mapsto f_a(t, y)$ is K-Lipschitz for all t and all a;
- $a \mapsto f_a(t, y)$ is L-Lipschitz for all t and y.

Then for a and b close together and t close to t_0, we have

$$|y_a(t) - y_b(t)| \leq \frac{L}{K} |a - b| \left(e^{K|t-t_0|} - 1 \right).$$

This is stronger than the uniqueness theorem — it implies uniqueness and also says that even if we have some perturbation, we're well-controlled for small time (where small is compared to K), although for large time we'll have exponential growth.

(There is also a C^1 statement and a C^k statement, but all the ideas are the same.)

Remark 10.3. This inequality looks like it could be bad when K is very small (since we have a K in the denominator), but it's not —

$$\lim_{K \to 0} \frac{e^{K|t-t_0|} - 1}{K} = |t - t_0|.$$

So even when K is very small, this is a good inequality.

Proof. We want to bound the difference between two functions; as before, the way we do this is by using the triangle inequality to bound the difference by some things we can actually control. More precisely, we'll bound

$$|y_a(t) - y_b(t)| \leq \left| \int_{t_0}^{t} (f_a(s, y_a(s)) - f_b(s, y_b(s))) \, ds \right|$$

(because we can integrate our equations — here we're using the fact that the initial conditions are the same, so cancel out).

Assume $t > t_0$ (otherwise we exchange the bounds of integration). Then the problem with our current expression is that we have two differences between the two terms, so we split up the difference into a

telescoping expression to only have one thing change at a time — we write this as

$$\leq \int_{t_0}^{t} |f_a(s, y_a(s)) - f_b(s, y_a(s)) + f_b(s, y_a(s)) - f_b(s, y_b(s))| \, ds.$$

What's nice is now we only have to deal with one difference at a time (while before we had two at once); and we know how to deal with both issues. For the first difference, we use Lipschitz-ness with respect to the variable a; for the second, we use Lipschitzness with respect to y; this gives

$$\leq \int_{t_0}^{t} L\,|a-b| + K\,|y_a(s) - y_b(s)| \, ds.$$

This is quite close to what we have in Grönwall's lemma — the left-hand side has $|y_a - y_b|$, and the right-hand side has an integral of it. The first term integrates to $L\,|a-b|\,|t-t_0| \leq L\,|a-b| \sup_I |t-t_0|$, which we define as our constant C.

Now define $y(t) = |y_a(t) - y_b(t)|$, define C as above, and define $p(t) = K$. Then we obtain

$$y(t) \leq C + \int_{t_0}^{t} K y(s)\, ds,$$

and so by Gronwall's inequality this is

$$\leq C e^{\int_{t_0}^{t} K} = C e^{K|t-t_0|}.$$

This is not exactly what we wanted; but this is one inequality we can get. There should be something a bit better to do, which will be in the notes. □

§10.1.1 An Example

Prof. Ozuch was solving an equation of the form

$$(1+br^2)rf'(r) + (4+3br^2)f(r) = 4 + 3cr^2,$$

which is a parametrized equation with parameters b and c. THey wanted to understand the solutions of this ODE for all b nad c, and in particular what happens around $b = c = 0$. They plugged the equation into wolfram alpha, but found that there are very tw odifferent solutions — when $b = 0$, the solutions you get are

$$f(r) = 1 + \frac{cr^2}{2} + \frac{C}{r^4}$$

for some constant C. But when $b \neq 0$, the solution is very different; it looks like

$$f(r) = \frac{c}{b} + \frac{4(b-c)(br^2+2)}{b^3 r^4} + C\frac{\sqrt{1+br^2}}{r^4}.$$

This is a somewhat nice expression — you don't expect much better — but there are a few issues when you're interested in the $b \to 0$ limit. Usually when this happens, sending $b \to 0$ in the second expressions houdl give you the first, but in this case that is not true — the b's in the denominator blow up.

But this is not what our theorem tells us about — our theorem is about

$$\begin{cases} y' = f_{b,c}(r, y) \\ y(r_0) = 0. \end{cases}$$

for instance. It took them an hour to realize that if you plug this condition into the solutions, then they become very complicated, but you do get the continuity (they are C^∞ with respect to b even at 0). But

your random solver might not see that the solutions to the ODEs have a very nice dependence; getting an explicit solution to an ODE is really not always what you want. Here the explicit solution just confused them.

So this statement should tell us that even though these two families don't look close to each other, they have to be if you match the two constants well enough.

Note that when we have two solutions and want to compare them, we need to give different names to the two constants; otherwise when you try to match the two solutions at a certain point, you need to match the two constants, which are a priori different. (In other words, you get one degree of freedom from the first equation, and another from the second; it's important to keep them both.)

§10.2 Interval of Existence

Now we'll look at how long our solutions exist for and how long our estimates hold.

We'll return to the non-parametric situation

$$\begin{cases} y' = f(t,y) \\ y(t_0) = y_0 \end{cases}. \tag{IVP}$$

Question 10.4. We've seen that close to t_0, there exists a unique solution. But how long do the solutions exist, and what can happen when they don't exist anymore?

We'll have one theorem that answers all of this. To define these things, we'll need the notion of being *locally Lipschitz*.

Definition 10.5. The function f is *locally Lipschitz* with respect to y if for all (t_1, y_1) where f is defined, there exist two constants $L, \varepsilon > 0$ such that for all $(t, y_2), (t, y_3) \in [t_1 - \varepsilon, t_1 + \varepsilon] \times [y_1 - \varepsilon, y_1 + \varepsilon]$, we have

$$|f(t,y_2) - f(t,y_3)| \le L\,|y_2 - y_3|\,.$$

This inequality is the Lipschitz one, and the ε's tell us that it's only a local statement. Note that the constants depend on where our point is — we don't necessarily have a constant L which is the same everywhere the function is defined.

Exercise 10.6. Show that C^1 implies locally Lipschitz implies C^0.

Example 10.7

$t \mapsto \sqrt{t}$ is locally Lipschitz on $(0, \infty)$.

As before, we also assume $t \mapsto f(t,y)$ is continuous (for all y).

Assume that:

- $f\colon \mathbb{R}^2 \to \mathbb{R}$ (it's possible to relax this, in the notes, but we won't do that here).
- $t \mapsto f(t,y)$ is C^0 on $\mathbb{R}$ for all $y \in \mathbb{R}$;
- $y \mapsto f(t,y)$ is locally Lipschitz on $\mathbb{R}$ for all t.

Theorem 10.8 (Maximal Interval of Existence)

Under the assumptions, for every $(t_0, y_0) \in \mathbb{R}^2$, consider the IVP

$$\begin{cases} y' = f(t, y) \\ y(t_0) = y_0. \end{cases}$$

Then the solution to (IVP) has a maximal interval of existence $I_{\max} = (\tau_-, \tau_+)$ with $\tau_\pm \in [-\infty, +\infty]$ (depending on t_0 and y_0). Moreover, if $\tau_- > -\infty$, then

$$\limsup_{t \to \tau_-} |y(t)| = +\infty.$$

Similarly if $\tau_+ < +\infty$ then

$$\limsup_{t \to \tau_+} |y(t)| = +\infty.$$

So if the solution doesn't exist everywhere, then it has to blow up. So the only thing that can go wrong is that the solution completely blows up — the solution has to exist everywhere or blow up.

Remark 10.9. If the function is only defined on a certain part of $\mathbb{R}$, then the only change to the statement is instaed of putting $-\infty$ and $+\infty$ on the τ's, you take the lower bound and upper bound of the interval I where f is defined. (If the place where the equation is defined is more complicated, then you have to be more careful.)

To prove this, we will use a connectedness argument that really tells us that a local existence result gives a global one; this is a common idea again that we'll use many times in the next few years.

§11 March 1, 2023

(The notes contain the correct application of Grönwall's lemma.)

§11.1 Connectedness

Today we'll prove the theorem on the interval of existence.

The keyword for this proof (and lots of other proofs we'll see) is *connectedness*. It's a common idea — this is the kind of proof that at first doesn't sound natural at all, but becomes natural after a while.

The idea is that to show a property is satisfied everywhere, you show that the set of points where it's satisfied is both open and closed. With connectedness, this means it must be the entire set.

Definition 11.1. A space is *connected* if the only sets which are both open and closed are the empty set and the entire space.

Here openness will come from the existence–uniqueness theorem (primarily the existence part), and closedness will come from the continuity of our solutions (this part is quite usual).

We'll not define open and closed sets in general (a general definition does exist), and only define them for intervals. Let $I \subseteq \mathbb{R}$ be an open interval (a, b) with $a \in [-\infty, +\infty)$ and $b \in (-\infty, +\infty]$.

Definition 11.2. An interval $J \subseteq I$ is *closed* in I if $J = [c, d] \cap I$ for some $c, d \in [-\infty, +\infty]$.

Definition 11.3. An interval $J \subseteq I$ is *open* in I if $J = (c, d) \cap I$ for some $c, d \in [-\infty, +\infty]$.

(We took I to be open, but all the definitions extend to closed or half-open I as well.)

(These definitions also extend to general subsets, but we won't cover that.)

Note that $\emptyset$ is both open and closed, and I is both open and closed.

Theorem 11.4 (Connectedness of Intervals)

If $J \subseteq I$ is both open and closed, then either $J = \emptyset$ or $J = I$.

This gives a way of showing a property is satisfied on the entire interval — by showing it's satisfied on a set that is both open and closed (and is nonempty).

Intuitively to see why this is true, either J has a supremum inside I or an infimum inside I (since we assumed that J is not the entire interval I). Without loss of generality assume $\sup J \in I$. Then if J is open it shouldn't include $\sup J$, and if it's closed it should. So it can't be both open and closed (unless $\sup J = \sup I$).

§11.2 Interval of Existence

Theorem 11.5 (Maximal interval of existence)

Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a function such that:

- f is C^0 on $\mathbb{R}^2$;
- f is locally Lipschitz with respect to the second variable.

Then for all $(t_0, y_0) \in \mathbb{R}^2$, there exists some interval $I = (\sigma_-, \sigma_+) \subseteq \mathbb{R}$ (with $t_0 \in I$) (called the *maximal interval of existence*) such that:

- There exists a unique solution y of our ODE
$$\begin{cases} y' = f(t, y) \\ y(t_0) = y_0 \end{cases}.$$
- If $\sigma_- > -\infty$, then
$$\limsup_{t \to \sigma_-} |y(t)| = +\infty.$$
Similarly if $\sigma_+ < +\infty$ then
$$\limsup_{t \to \sigma_+} |y(t)| = +\infty.$$

Note that I depends on t_0 and y_0. The interval I is maximal in the sense that when we reach the bounds of the interval, the solution blows up, so there's no way to continue it.

Proof. We'll divide the proof into a few steps.

The first step, called *gluing solutions*, is to show that if we have a solution on some interval I_1 and another solution on some interval I_2 (which overlaps I_1), and these solutions match on $I_1 \cap I_2$ (which contains y_0), then we can glue the two solutions to obtain a solution on $I_{12} = I_1 \cup I_2$ — so we can glue our local solutions to each other to obtain a solution that's larger and larger. We'll see that we can use connectedness to extend our solution as long as the function doesn't blow up.

Let y_1 and y_2 be solutions of (IVP) on I_1 and I_2.

Claim — We have $y_1 = y_2$ on $I_1 \cap I_2$.

Once we prove this, we'll define $y_{12}\colon I_1 \cup I_2 \to \mathbb{R}$ such that $y_{12} = y_1$ on I_1 and $y_{12} = y_2$ on I_2 (which makes sense because they both match on the intersection).

Proof. We'll define a set on which y_1 and y_2 match and show that it's both open and closed. *A priori* we only know that the two solutions match close to t_0 (by the existence–uniqueness theorem); they might be doing random things on the rest of the interval. (Sometimes when you don't have uniqueness this actually happens.)

This is where connectedness comes in — we introduce $I^* = \{t \in I_1 \cap I_2 \mid y_1(t) = y_2(t)\}$, and define I_0^* as the largest interval in I^* with t_0. We will show that I_0^* is both open and closed (in $I_1 \cap I_2$).

First we'll show that I_0^* is closed. (Usually this is a bit easier.)

For simplicity assume $I_1 \cap I_2$ is an open interval (a, b) (the other cases $(a, b]$, $[a, b)$, and $[a, b]$ can be dealt with in the same way).

Then there are two situations — either I_0^* is closed, or $I_0^* = (c, d) \cap (I_1 \cap I_2) \subsetneq I_1 \cap I_2 = [a, b]$. (In other words, either $c > a$ or $d < b$.) We want to show the latter cannot happen.

Assume for the sake of contradiction that $I_0^* = (c, d) \cap (a, b)$, with $c \in (a, b)$ and $c > a$. (The other case is similar.)

Then for all $t \in (c, d)$ we know that $y_1(t) = y_2(t)$. But we also know that y_1 and y_2 are both continuous at every point of $I_1 \cap I_2$, and in particular at c. This means $y_1(c) = y_2(c)$, which means $c \in I_0^*$, contradiction.

The next step is to prove openness, and that's where we will use the existence–uniqueness theorem. For contradiction, assume that $I_0^* = [c, d] \cap (a, b)$, and that $c \in (a, b)$. Our goal is to extend it to an interval $(c - \varepsilon, d) \cap (a, b)$ for some $\varepsilon > 0$ (which a priori depends on c).

We can do this by the existence–uniqueeness theorem — we know $y_1(c) = y_2(c)$ because $c \in I_0^*$. But there exists a unique solution y to

$$\begin{cases} y' = f(t, y) \\ y(c) = y_1(c) = y_2(c) \end{cases},$$

where this solution is from $(c - \varepsilon c, c + \varepsilon c) \to \mathbb{R}$.

Since y_1 and y_2 are solutions and they match at c, then we have that $(c - \varepsilon(c), c + \varepsilon(c)) \subseteq I_0^*$, which is a contradiction because we assumed I_0^* finished at c on the left-hand side.

Intuitively, suppose we have two solutions, and some t_0, .By definition, our two solutions are equal until c. *A priori* it's unclear they have to continue to be equal after c. But we're saying that the existence-uniqueness theorem tells us there's a unique solution on $[c - \varepsilon, c + \varepsilon]$, which means since they're equal at c that they must be equal here. So we can add this new set in, which means I_0^* is larger. So this gets a contradiction unless I_0^* is the entire thing.

So we must have $I_0^* = I_1 \cap I_2$ — we showed that it's an interval that's both closed and open, so by connectedness it must be the whole set. □

The main consequence of this is that it lets us define I as the union of all sets where a solution of (IVP) exists and is unique. Then this is our maximal interval.

This is because whenever we have two sets containing y_0 where we could have had a solution, the solution has to agree on all the intersections; so we can glue them together to come up with one common solution, and all the other solutions are solutions on subsets.

Now we've extended the solution as much as we can; we'll next class see what happens if I is not $\mathbb{R}$ (and we'll see the only thing that can happen is that everything blows up). □

§12 March 3, 2023

§12.1 Maximal Interval of Existence

Our goal is to prove the result of the maximal interval of existsence — if $f\colon \mathbb{R}^2 \to \mathbb{R}$ is such that $t \mapsto f(t,y)$ is C^0 and $y \mapsto f(t,y)$ is locally Lipschitz, then there is existence and uniqueness on a *maximal interval* $I_{\max}$, which is maximal in the sense that if its endpoints are not $\pm\infty$, then f must blow up at that point.

Last time, we saw that $I_{\max}$ is the union of all intervals where we can define a solution with the initial conditions (t_0, y_0). We showed that this makes sense because if we had two solutions on different intervals, then we could match them into one solution alone. (Imagine a picture with t_0, and I_1 on the left (ending right of t_0) and I_2 on the right (starting left of t_0), with a function on each. Then the two functions have to match on teh overlap.) The two functions have to match on their overlap, which means we can glue them together to extend the solution. Thismeans there's a maximal interval where we can define everything, and everything makes sense because we can glue together solutions on different intervals.

Now we want to understand the second point, on what happens at an endpoint of the interval — we want to show that the solutions have to blow up.

Assume that $I_{\max} = (\sigma_-, \sigma_+)$ with $\sigma_+ < +\infty$. (We'll only deal with σ_+; the same thing works for σ_-.) We want to show that the only thing that can go wrong (i.e., to cause a solution not to exist) is that the function blows up.

Assume for the sake of contradiction that $\limsup_{t\to\sigma_+, t\in I_{\max}} |y(t)| < +\infty$.

> **Claim —** $y(t)$ has a limit as $t \to \sigma_+$.

We'll first show that assuming the claim, the theorem holds. (The idea is that then we can extend the interval a little bit, contradicting maximality.)

Proof. Define $y(\sigma_+) = \lim_{t\to\sigma_+} y(t)$. Then y is a continuous function at σ_+. We can now consider the $(\mathrm{IVP}_{\sigma_+})$

$$\begin{cases} z' = f(t,z) \\ z(\sigma_+) = y(\sigma_+). \end{cases}$$

By the existence–uniqueness theorem, we know there exists $\varepsilon > 0$ such that there exists a unique solution z to $(\mathrm{IVP}_{\sigma_+})$ on $(\sigma_+ - \varepsilon, \sigma_+ + \varepsilon)$.

Now we have a unique solution for y on $I_{\max}$, with limit $y(\sigma_+)$. Now we can also take the interval $(\sigma_+ - \varepsilon, \sigma_+ + \varepsilon)$, on which we have a purple solution z. But by uniqueness, this solution has to match y on the left half of the interval. So this lets us extend the solution y by gluing together the two solutions — using $\#$ to denote gluing solutions, $y\#z$ is a solution of the initial problem on a *strictly* larger interval than $I_{\max}$, which is a contradiction to the fact that $I_{\max}$ is maximal.

(The point is that the yellow line with the inclusion of this point is a continuous point — we're reusing uniqueness at this one point.) □

Now we'll prove the claim.

Proof of Claim. We can write

$$y(t) = y_0 + \int_{t_0}^{t} f(s, y(s))\, ds.$$

We wish to show that the integral where we take $t = \sigma_+$ converges.

But we know $y(s)$ is bounded uniformly on $[t_0, \sigma_+)$ (by our assumption that y does not blow up); suppose it is bounded by Y_0. Now f is a continuous function, so $f(s, y(s))$ on $[t, \sigma_+)$ is bounded by

$$\max_{[t_0,\sigma_+]\times[-Y_0,Y_0]} |f| = M < +\infty$$

(which exists because f is continuous and this set is compact).

This tells us

$$|y(t) - y(s)| \leq |t - s| \cdot M < M\,|t - \sigma_+|$$

if $t_0 \leq t \leq s < \sigma_+$. This means $s \mapsto y(s)$ is 'Cauchy' at σ_+ (this is a continuous version of Cauchy sequences — you can call $y(t) = t_n$ and $y(s) = t_N$ if you want to return to sequences). So for any sequence we have a limit; these limits must actually match (so the continuous version is stronger, because it tells us that all the limits of the sequences $t_n \to \sigma_+$ have the same limit). So there does exist a limit of y at σ_+.

More explicitly, we first have

$$y(\sigma_+ - \frac{1}{n}) - y\left(\sigma_+ - \frac{1}{N}\right) < \frac{M}{n} \to 0$$

as $n \to \infty$, which means $y(\sigma_+ - \frac{1}{n})$ is Cauchy (as a sequence in n) and therefore has a limit.

Now we need to verify that if we take any sequence that goes to σ_+, we have the same limit. Suppose that $t_n \to \sigma_+$; then the same reasoning shows that

$$\left|y(t_n) - y(\sigma_+ - \frac{1}{N})\right| \leq M\,|\sigma_+ - t_n|\,.$$

But now we can take al imit as $N \to \infty$. The left-hand side goes to $|y(t_n) - L|$ (where L is our limit above), and so we're saying $|y(t_n) - L| < M\,|t_n|$, which goes to 0 as $n \to \infty$. □

So now we're done — we know that if our interval has an end, then the solution should blow up.

Example 12.1

Consider $y' = y^2$ with $y(0) = 1$.

If you solve this, you'll get

$$y(t) = \frac{1}{1 - t}.$$

This blows up as $t \to 1$, so the maximal interval of existence is $I_{\max} = (-\infty, 1)$ (and the solution blows up at 1).

More generally if we replace $y(0) = \alpha$, then we have

$$y(t) = \frac{1}{\frac{1}{\alpha} - t},$$

which goes to $+\infty$ at $t = \frac{1}{\alpha}$. (If $\alpha = 0$, then the only solution is 0 by uniqueness.)

Here we see that we have a finite $I_{\max}$, because $y' = y^2$ is not Lipschitz.

On the other hand, we have the following new statement:

Theorem 12.2

Suppose that either:

1. $\max_{\mathbb{R}^2} f = M < +\infty$; or
2. For all y_1, Y_2, and t, we have

$$|f(t, y_1) - f(t, y_2)| \leq L(t) |y_1 - y_2| .$$

(The key point here is that the Lipschitz constant does not depend on y.)

Then f does not blow up in finite time, and $I_{\max} = \mathbb{R}$.

So whenever we have a bounded function everything's fine; and if it is Lipschitz we can also make it work.

The first case is simple — it's the same as our proof of the claim (all we needed there was that f was bounded).

The second is a bit more painful, so we won't do it, but it's done in the notes; you can do it using Grönwall's inequality.

We will often use these. Otherwise, our maximal intervals tell us that you shouldn't worry too much about — you can talk about the solution and its property, and there will always be a maximal interval and you can work with it.

Remark 12.3. The solution can still blow up at $+\infty$ — e.g., $y' = y$ with $y(0) = 1$ has solution $y(t) = e^t$.

Remark 12.4. $y' = y^2$ doesn't satisfy (1) or (2), which is why it's possible to have a finite-time blowup.

We've spent a lot of time on existence and uniqueness for first-order ODEs of dimension 1, but we'll see later that this extends to any dimension and any order somewhat easily (replacing all the absolute values by norms in $\mathbb{R}^d$), because all of them can be written as first-order problems. Later, we might do this in random super-complicated Banach spaces of infinite dimension, but we can do the same thing (taking the norm instaed of aboslute value) — and the theorem will extend the same way. THis is why it's important to have seen it once — it extends to vastly more complicated settings, and the ideas and techniques are the same.

Remark 12.5. If we have a condition of the form $y' = f(y)$, $y(t_1) = y_0$, then if $I_{\max}(t_1, y_0) = (a, b)$, then the same problem with $y(t_2) = y_0$ instead has $I_{\max}(t_2, y_0) = (a + t_2 - t_1, b + t_2 - t_1)$ (and the solution will just be a translation). So for such things, we might as well start with 0.

§12.2 Linear Second-Order ODEs

We'll now introduce the next chapter, whose beginning will be on the exam (which is in two weeks).

We'll later work with any order ODEs in any dimension, but second-order ODEs have the special feature that you see them a lot in physics (because of acceleration being related to force), so we'll look at these specifically.

We'll mostly look at linear ones, but there's already a lot to say about these.

Definition 12.6. A second-order linear ODE is of the form

$$y''(t) + p(t)y'(t) + q(t)y(t) = f(t).$$

We want to solve this on some interval; we assume that all of $p(t)$, $q(t)$, and $f(t)$ are continuous on some interval I.

We will have a very different existence-uniqueness theorem for these equations:

Example 12.7

Consider $y'' + y = 0$. Both cos and sin are solutions!

This contradicts what would be a naive try at an existence-uniqueness theorem — existence-uniqueness tells us that two solutions to the same ODE cannot cross (i.e., they can't have teh same value at the same point), but cos and sin obviously cross (a lot). So our previous existence–uniqueness which told us that two solutions equal at one point are equal here does *not* hold here. This is why we need to be more careful with our initial value problems. Our intial value problems will require *two* conditions —

$$\begin{cases} y'' + y = 0 \\ y(0) = 0 \\ y'(0) = 1 \end{cases}$$

then gives us just sin. This is the kind of initial condition we'll need.

We'll learn how to solve these completely when they have constant coefficients, and say something nontrivial when they have variable coefficients.

§13 March 6, 2023 — Second-Order Linear ODEs

We'll now look at second-order linear ODEs, of the form

$$y'' + p(t)y' + q(t)y = f(t) \tag{LODE2}$$

on I (where p, q, and f are in $C^0(I)$).

We'll start by proving uniqueness.

Theorem 13.1

Suppose we have a differential equation

$$\begin{cases} \text{(LODE2)} \\ y(t_0) = y_0, y'(t_0) = y_1. \end{cases} \tag{IVP}$$

Then (IVP) has at most one solution on an interval I including t_0, if p and q are bounded.

So in other words, if a solution exists then it is unique. (We may not need the assumption that p and q are bounded — since they're continuous, they are bounded everywhere except possibly at the ends of the intervals — but we will use it in the proof.

Proof. Assume that y and z both solve (IVP), i.e., they solve the same ODE and satisfy the same initial conditions.

Whenever we have a linear equation, it's a good idea to take the difference; so we'll define $v = y - z$, and we want to show that $v = 0$. By linearity, we know v solves

$$\begin{cases} v'' + p(t)v' + q(t)v = 0 \\ v(t_0) = 0, v'(t_0) = 0. \end{cases}$$

Now we perform the following trick: first define $w = v^2 + (v')^2$. (In terms of physics, we can think of this as the energy of our solution; in terms of mathematics we want to think of this as some kind of norm measuring how large our function and its derivative is.) The point is that $w = 0$ if and only if $v = 0$.

Differentiating w and plugging in the equations that v satisfies, we get

$$w' = 2vv' + 2v'v'' = 2vv' + 2v'(-p(t)v' - q(t)v).$$

We can rewrite this as

$$(2 - 2q(t))vv' - 2p(t)(v')^2.$$

Now define $P = \sup_{t\in I} |p(t)|$ and $Q = \sup_{t\in I} |q(t)|$.

Our goal is to show that $w' \le cw$ for some constant c. Since we have $w(0) = 0$, this will be enough to conclude (as seen in the existence-uniqueness proof, when we proved uniqueness).

We have $|w'| \le (2+2Q)\,|vv'| + 2P(v')^2$. Our goal is to express this in terms of v^2 and $(v')^2$. To do this, we'll use the following (very commonly used — for example, to prove that the product of continuous functions is continuous) inequality:

> **Fact 13.2 —** If $a, b \in \mathbb{R}$ then $2\,|ab| \le a^2 + b^2$.

Proof. We have

$$0 \le (|a| - |b|)^2 = a^2 - 2\,|a|\,|b| + b^2.$$ □

Applying the key inequality here, we get that

$$\begin{aligned} |w'| &\le (2 + 2Q)\frac{v^2 + (v')^2}{2} + 2P(v')^2 \\ &= (1 + Q)v^2 + (2P + 1 + Q)(v')^2 \\ &\le (2P + 1 + Q)w, \end{aligned}$$

as $w = v^2 + (v')^2$. So now we have an equation $|w'| \le cw$, which means that $w = 0$. □

So it's possible to show uniqueness without complicated techniques. (The proof is kind of a trick, but the reason we saw it is the above inequality, which is frequently useful.)

§13.1 Constant Coefficients

We will first ask that p and q are constants, and f is 0. Concretely, now we're studying

$$y'' + py' + qy = 0 \tag{E}$$

where p and q are real numbers. We'll solve this equation explicitly.

The key is the characteristic polynomial.

> **Definition 13.3.** The *characteristic polynomial* $P_{(E)}(x) = X^2 + pX + q$.

This polynomial is relevant because we can 'write'

$$P_{(E)}\frac{d}{dt}(y) = \frac{d^2}{dt^2}y + p\frac{d}{dt}y + qy.$$

(This should be in quotation marks, as it doesn't exactly make sense; but later we'll see how to make sense of it.)

(For operators, squaring means repeated application — i.e., $(\frac{d}{dt})^2 y = \frac{d}{dt}\left(\frac{d}{dt}y\right)$. This defines some product in the space of operators, and in some situations you have an algebra where you can compose operators with each other — i.e., imagine if we had $\frac{d}{dx}$ as well. Then you can ask when this commutes, which is when it's C^2; there is a lot of theory here, and you'll see it in the next few years.)

Now consider the soltuion $y(t) = e^{\alpha t}$ for some α. Then we have $y' = \alpha e^{\alpha t}$ and $y'' = \alpha^2 e^{\alpha t}$. This means

$$y'' + py' + qy = (\alpha^2 + p\alpha + q)e^{\alpha t}.$$

So $t \mapsto e^{\alpha t}$ is a solution to (E) if and only if α is a root of $P_{(E)}$.

Theorem 13.4

THe set of solutions to (E), which we denote $S_{(E)}$, is 2-dimensional and generated by:

- If $P(E)$ has two distinct real roots, then

  $$S_{(E)} = \{t \mapsto Ae^{\alpha t} + Be^{\beta t}\}$$

  is the set of solutions to the ODE.
- If $P(E)$ has a double root $\alpha \in \mathbb{R}$, then

  $$S_{(E)} = \{t \mapsto Ae^{\alpha t} + Bte^{\alpha t}\}.$$

  (This is the degenerate case.)
- If $P_{(E)}$ has two complex roots (which must be complex conjugates) $\alpha_1 = a + ib$ and $\alpha_2 = a - ib$.

  If we then do the same thinga nd take real roots, we get the solution set

  $$S_{(E)} = \{Ae^{\alpha t}\cos(bt) + Be^{-\alpha t}\sin(bt)\}.$$

Remark 13.5. To solve an ODE with the conditions given, we want to solve $A+B = y_0$ and $A-B = y_1$. (There's also theory about what happens if you want to solve an ODE with two initial conditions $y(t_1)$ and $y(t_2)$; in this case, the solution might not exactly be unique — for example, $A\sin$ is not specified by two points where it's 0.)

Our proof of uniqueness would fail here because it involved considering $w = v^2 + (v')^2$; in our proof we could say that $w(t_0) = 0$, but here we can't say anything about w anywhere.

Proof. In the first case, we know that $Ae^{\alpha t} + Be^{\beta t}$ is *a* solution. But then for any initial conditions, this is the unique solution — there is a solution of this form with the given initial values. So then this is the *only* solution with $y(t_0) = A + B$ and $y'(t_0) = \alpha A + \beta B$. (This exhausts all the initial conditions, so it's the complete set of solutions.)

In the second case, we know that $e^{\alpha t}$ is a solution; we also want to check that $te^{\alpha t}$ is. We have

$$y' = (1 + \alpha t)e^{\alpha t} \text{ and } y'' = (\alpha + \alpha + \alpha^2 t)e^{\alpha t},$$

and we know $P_{(E)} = x^2 + px + q = (x - \alpha)^2$, which means $p = -2\alpha$ and $q = \alpha^2$. So now the expression we have here is

$$e^{\alpha t}((\alpha^2 t + 2\alpha) - 2\alpha(1 + \alpha t) + \alpha^2) = 0,$$

which is indeed true.

(For the third, you can solve in $\mathbb{C}$ and project onto $\mathbb{R}$, or simply verify that it's a solution.) □

§14 March 8, 2023

§14.1 Definitions

Today we will talk more about spaces of solutions to linear equations.

To prepare for this, we'll talk about what it means to be a vector subspace — we'll see some definitions and examples. (We've already seen some vector spaces — such as $\mathbb{R}^d$ — but we'll today mostly talk about a vector space of *functions*.)

We'll assume that we're working with a vector subspace of a space of functions, but you can imagine the analogs in $\mathbb{R}^d$ or in a more general setting (we're just not going to define vector spaces in general, since it's not useful for this class).

Definition 14.1. A $\mathbb{R}$-*vector subspace* E (of functions $I \to \mathbb{R}$) is a set of functions $I \to \mathbb{R}$ which is stable under taking linear combinations — i.e., for all $\lambda, \mu \in \mathbb{R}$ and for all functions y and z in E, we have $\lambda y + \mu z \in E$ as well.

Our main example (and the reason we care about the definition in this class) is the following:

Example 14.2

Let $E = \{y \in C^n(I) \mid y^{(n)} + a_{n-1}(t)y^{(n-1)} + \cdots + a_0 y_0 = 0\}$. Then E is a vector subspace.

(Note that a vector space may be infinite-dimensional — the set of C^0 functions (or C^1 functions) is a vector space since continuity (and more generally C^k) is preserved under adding and scaling functions. In particular, it may be hard to consider a basis.)

§14.2 Linear Operators

Definition 14.3. A *linear operator* $L\colon E \to E'$ (where E and E' are both vector spaces) is a functional L which preserves linear combinations, i.e.,

$$L(\lambda y + \mu z) = \lambda L(y) + \mu L(z).$$

(We may have E' be another subspace of functions, or $\mathbb{R}^d$.)

We've seen many linear operators.

Example 14.4

- $L(y) = y'$ is a linear operator, since if we sum two functions, $(f+g)' = f' + g'$.
- $L(y)(t) = \int_0^t y$ is also a linear operator.
- $L(y) = y(t_0)$ (i.e., evaluating the function at a fixed point) is a linear operator. It's a silly example, but it's quite deep in the sense that many of the other linear operators can be written as products against some well-chosen functions, but this one can't. (This is the start of the theory of distributions; it's through this example that you can define e.g. the Dirac function.)
- If L_1 and L_2 are both linear, then $L(y) = L_1(L_2(y))$ is also linear. (If L is linear and bijective, then L^{-1} is also linear.)
- If L_1 and L_2 are both linear, then $L(y) = L_1(y) + L_2(y)$ is also linear.
- $L(y)(t) = a(t)y(t)$ is also linear.
- Combining all of these, we get that

$$L(y) = y^{(n)} + a_{n-1}(t)y^{n-1} + \cdots + a_0(t)y$$

  is also a linear operation.

We care about these because they let us do linear algebra on our spaces of functions. We'll soon even be doing geometry on our space of functions (saying two functions are orthogonal or parallel or have certain heights).

Definition 14.5. The *kernel* of a linear operator L is defined as

$$\ker(L) = L^{-1}(\{0\}) = \{y \in E \mid L(y) = 0\}.$$

Note that this notation doesn't mean L is invertible. (If L is invertible then $L^{-1}(\{0\})$ is simply 0.)

Exercise 14.6. If L is a linear operator, then $\ker(L)$ is a vector subspace.

As an application, we have the following:

Example 14.7

If we define $L(y) = y^{(n)}$ on $\mathbb{R}$, then

$$\ker(L) = \{t \mapsto a_{n-1}t^{n-1} + \cdots + a_1 t + a_0 \mid a_{n-1}, \ldots, a_0 \in \mathbb{R}\}.$$

Note that we have n degrees of freedom. At some point we'll discuss dimensions and basis; we'll see that the functions $1, t, \ldots, t^{n-1}$ form a basis for our solution set, which has dimension n.

Definition 14.8. The *image* of a linear operator $L\colon E \to E'$ is defined as

$$\operatorname{im}(L) = L(E) = \{y \in E' \mid \text{exists } z \in E \text{ with } y = L(z)\}.$$

In other words, we can write $\operatorname{im}(L) = \{L(z) \mid z \in E\}$ (which is the reason for the notation $L(E)$).

Exercise 14.9. If L is a linear operator, $\operatorname{im}(L)$ is a vector subspace.

These two results are likely the most common way to show that a set is a vector subspace — it's enough to

see that it's the zero set or image of some linear operation. (Most vector spaces we care about will be either images or kernels.)

Corollary 14.10

The set of solutions of a homogeneous linear ODE is a vector subspace.

Example 14.11

The kernel of $L(y) = y'$ is constant functions; the kernel of $L(y)(t) = \int_0^t y$ is the zero function; the kernel of $L(y) = y(t_0)$ is the huge space of functions which vanish at t_0; you can check what the kernel of $L(y) = L_1(L_2(y))$ is; there's no general rule for the kernel of $L(y) = L_1(y) + L_2(y)$; the kernel of $L(y)(t) = a(t)y(t)$ is the set of functions which vanish wherever a does not vanish (but can be anything where a vanishes).

Example 14.12

The image of C^1 functions under $L(y) = y'$ is C^0 functions; the image of C^0 functions under $L(y)(t) = \int_0^t y$ is C^1 functions which vanish at 0; the image of C^0 functions under $L(y) = y(t_0)$ is $\mathbb{R}$, and so on.

§14.3 Affine Subspaces

Definition 14.13. An *affine subspace* of direction E is a set F of the form

$$F = y^* + E = \{y^* + z \mid z \in E\}$$

for some fixed function y^*.

The key point is that you want to think of y^* as a translation. For example, in $\mathbb{R}^3$ F might be some plane, and y^* a point on that plane; meanwhile E always has 0 inside it (so F is the translation of E where we translate 0 to y^*).

Note that y^* is very non-unique:

Exercise 14.14. If F is an affine subspace, then for *all* $y \in F$ we have $F = y + E$.

Our main example is the following:

Example 14.15

$L^{-1}(\{z\}) = \{y \mid L(y) = z\}$ is an affine subspace — its direction will be $\ker(L)$, and y^* can be any single function with $L(y^*) = z$.

The one that will be important for us is ODEs, but there are simpler examples:

Example 14.16

Taking $L(y) = y'$,

$$\{y \mid y' = f\} = \int_0^t f + \mathbb{R} \cdot 1$$

is an affine subspace — here $\mathbb{R} \cdot 1$ (1 denotes the constant function 1) is our direction E, and $\int_0^t f$ is y^*.

Example 14.17

$\{y \mid y(t_0) = y_0, y'(t_0) = y_1\}$ is an affine subspace; we can write it as

$$y_0 + ty_1 + \{y \mid y(t_0) = 0, y'(t_0) = 0\}.$$

Here y^* is $y_0 + ty_1$, and the direction E is the set on the right.

Our main object will be the set of solutions to an ODE:

Proposition 14.18

The set of solutions to a linear ODE $L(y) = f$ (where L involves taking derivatives and multiplying by functions) is an affine subspace whose direction is the solution set to the homogeneous problem $L(y) = 0$ — it is of the form

$$L^{-1}(\{f\}) = y^* + L^{-1}(\{0\}),$$

where y^* is any one particular solution to the original ODE.

This is useful because it means to solve a linear ODE, it suffices to find *one* solution and solve the homogeneous problem. This separates the problem in two — we're looking for one particular solution (and sometimes there's an obvious one) and trying to solve the homogeneous problem, which is often simpler.

§14.4 Bases and Dimension

Definition 14.19. A *linearly independent* family of functions $(y_1, \dots, y_n)$ is one which satisfies the following condition: if $\lambda_1 \dots, \lambda_n \in \mathbb{R}$ are such that

$$\lambda_1 y_1 + \cdots + \lambda_n y_n = 0,$$

then we must have $\lambda_1 = \cdots = \lambda_n = 0$.

In other words, our vectors are linearly independent if we can't find a nontrivial equation between them. (We've probably seen this in $\mathbb{R}^d$ before; this is just a more abstract version.)

Example 14.20

$((1,0),(0,1))$ in $\mathbb{R}^2$ is linearly independent —

$$\lambda_1(1,0) + \lambda_2(0,1) = (\lambda_1, \lambda_2),$$

so if this is 0 then we have $(\lambda_1, \lambda_2) = (0,0)$ and therefore $\lambda_1 = \lambda_2 = 0$.

We'll see that in dimension 2, there are two possibilities:

- Either there exists $(\lambda_1, \lambda_2) \neq 0$ such that $\lambda_1 y_1 + \lambda_2 y_2 = 0$, which implies that y_1 and y_2 are proportional;
- Or there exist no such $(\lambda_1, \lambda_2) \neq 0$ such that $\lambda_1 y_1 + \lambda_2 y_2 = 0$, which means they're linearly independent.

So the only way for two vectors to fail to span $\mathbb{R}^2$ is for them to be proportional.

§15 March 10, 2023

Linear algebra is about spaces where we can perform linear combinations — sums of real numbers times vectors. (Note that we're using *vectors* to refer to functions — our functions are elements of a vector space,

so we'll see them as vectors.)

Linear operators are the natural operations between vector subspaces. Any operator you've seen where you can sum (e.g., derivatives and integrals) are linear.

These two definitions help you define other things — the *kernel* (or *null space*) is the set of y for which $L(y) = 0$. Our main example is the solutions to a homogeneous linear ODE (this is why we care about linear algebra in this class, or at least the first reason).

An *affine subspace* is essentially when we allow our vector space to be translated — i.e., it's a particular vector plus all the directions in our vector subspace. For us, this is important because the set of solutions to a linear (not necessarily homogeneous) ODE form an affine subspace. If there's a solution that's simple to obtain, then you can take it to be your shift, and it remains to solve the linear *homogeneous* problem.

§15.1 Bases and Dimension

Definition 15.1. A *linearly independent family* of vectors $(y_1, \dots, y_n)$ is a collection of vectors such that whenever $\lambda_1 y_1 + \cdots + \lambda_n y_n = 0$ for some $\lambda_i \in \mathbb{R}$, we must have $\lambda_i = 0$ for all i.

Definition 15.2. A set of vectors *spans* E (the vector space) if for every $z \in E$, there exist $\lambda_i \in \mathbb{R}$ such that $\lambda_1 y_1 + \cdots + \lambda_n y_n = z$.

We should think of the first condition as uniqueness and the second as existence. More precisely, consider the map $\mathbb{R}^n \to E$ sending $(\lambda_1, \dots, \lambda_n) \mapsto \lambda_1 y_1 + \cdots + \lambda_n y_n$. Then $(y_1, \dots, y_n)$ are linearly independent if and only if this map is injective, and they span E if and only if the map is surjective.

If the map is injective, it means we can send a copy of $\mathbb{R}^n$ into E; if it's surjective, it means we can represent any vector in E with $\mathbb{R}^n$. As we'll see soon, if we have a *basis*, then both are true; this means our space is exactly like $\mathbb{R}^n$. This is what we'll do with ODEs — we'll show that the solutions to our ODE have a basis, and then working with them will be just like working in $\mathbb{R}^n$, which we're familiar with.

Definition 15.3. A *basis* of E is a linearly independent family of vectors which spans E.

These two properties tell us that the map $\mathbb{R}^n \to E$ sending $(\lambda_1, \dots, \lambda_n) \mapsto \lambda_1 y_1 + \cdots + \lambda_n y_n$ is a bijection — or in other words, that our space E is exactly like $\mathbb{R}$.

Remark 15.4. Here we won't refer to infinite-dimensional spaces, but we will later. In that case, we can still have bases — we can extend the definitions to allow infinitely many (and even uncountably many) elements of the basis, with some modifications (e.g., possibly taking an integral instead of a sum).

Having a basis also lets you define the *dimension* of a vector space.

Definition 15.5. The *dimension* of a vector space is (the following are equivalent definitions):

(1) The largest number of linearly independent vectors in our vector space.

(2) The smallest number of vectors spanning the whole space.

(3) The common number of vectors in every basis of E.

(We won't prove that these definitions are equivalent.)

The dimension can also be infinite, if our basis has infinitely many vectors; but for now we focus on finite-dimensional spaces.

Example 15.6

The space $\mathbb{R}^d$ has dimension d.

$\mathbb{R}^d$ is a space we understand very well, so we can exhibit a basis explicitly — for example, $(1, 0, \dots, 0)$, $(0, 1, 0, \dots, 0)$, $(0, \dots, 0, 1, 0)$, $(0, \dots, 0, 1)$ is a basis. So we have a basis with d elements, which means the dimension is d. This is called the *canonical basis* of $\mathbb{R}^d$.

Example 15.7

$(\sin, \cos)$ is a basis for the set of solutions to $y'' + y = 0$.

Proof. We wish to check that they span the space of solutions, and that they're linearly independent.

To see that they span the space, let y be any solution to the differential equation $y'' + y = 0$. Fix some $t_0 \in \mathbb{R}$, and let $y(t_0) = y_0$ and $y'(t_0) = y_1$.

Then y is the solution of

$$\begin{cases} y'' + y = 0 \\ y(t_0) = y_0, y'(t_0) = y_1. \end{cases}$$

But we can construct another solution to this ODE which is a linear combination of sin and cos — the equation we'd like to solve is

$$\begin{cases} a\cos t_0 + b\sin t_0 = y_0 \\ a(-\sin t_0) + b\cos t_0 = y_1. \end{cases}$$

This corresponds to an invertible matrix — we have

$$\det \begin{bmatrix} \cos t_0 & \sin t_0 \\ -\sin t_0 & \cos t_0 \end{bmatrix} = \cos^2 t_0 + \sin^2 t_0 = 1,$$

which means we can always find a solution.

Now we'll show that they're linearly independent. This can be done by noting that cos and sin are not proportional ($\sin 0 = 0$ and $\cos 0 = 1$, and $\sin \frac{\pi}{2} = 1$ while $\cos \frac{\pi}{2} = 0$). A worse method, but one that extends to more vectors, is the following — we'll assume that we have a linear combination which vanishes, and show that all the coefficients vanish. So assume that we have $\lambda_1, \lambda_2 \in \mathbb{R}$ such that

$$\lambda_1 \sin + \lambda_2 \cos = 0.$$

Plugging in 0, we get

$$0 = \lambda_1 \cos(0) + \lambda_2 \sin(0) = \lambda_1,$$

which tells us $\lambda_1 = 0$. Similarly, plugging in $\frac{\pi}{2}$, we get

$$0 = \lambda_1 \cos\frac{\pi}{2} + \lambda_2 \sin\frac{\pi}{2} = \lambda_2,$$

so $\lambda_2 = 0$. □

Remark 15.8. Two functions f and g are *proportional* if there exists $\alpha \in \mathbb{R}$ such that $f = \alpha g$ or $g = \alpha f$. For two vectors, being linearly independent is equivalent to not being proportional.

Lemma 15.9

If there exists a linear map $L\colon E \to \mathbb{R}^d$ such that $\ker L = \{0\}$, then $\dim E \le d$.

This will be useful for us because we can use it to show that we've found all the solutions to an ODE if we've found d linearly independent solutions.

Proof. We're going to find d vectors $y_1, \dots, y_d$ (or fewer) that span E; then using the definition of the dimension as the minimum number of vectors needed to span the space, we get $\dim E \le d$.

Consider $\operatorname{im} L = \{L(y) \mid y \in E\} \subseteq \mathbb{R}^d$; then $\dim \operatorname{im}(L) \le d$ (you can prove that if we have one space inside another, the dimension of the first is smaller than the dimension of the second).

Let $\dim \operatorname{im} L = k$, and let $z_1, \dots, z_k$ be a basis of $\operatorname{im} L$. Let $y_1, \dots, y_k$ be defined such that $L(y_1) = z_1, \dots, L(y_k) = z_k$ (which we can do since each of these vectors z_i is in the image).

Then we can show $y_1, \dots, y_k$ is a basis of L. □

§16 March 15, 2023

Suppose we have a polynomial

$$(x + b(t))(x + a) = x^2 + (a + b(t))x + ab(t).$$

When we have constant coefficients, we've seen that the roots of the polynomial give solutions to an ODE.

We don't when the coefficients aren't constant, but consider the ODE

$$y'' + (a + b(t))y' + ab(t)y = 0.$$

The point is that $y' + ay$ satisfies the nice ODE $z' + b(t)z = 0$, which we know how to solve; this gives

$$z(t) = z(0)e^{-\int_0^t b}.$$

Then we can rewrite our equation — we know that $y' + ay = z = z(0)e^{-\int_0^t b}$. Now we again have a first-order ODE which we know how to solve — we know $y(t) = \left(y_0 + \int_0^t e^{as - \int_0^t b}\right) e^{-at}$.

This is what was supposed to happen in the first problem of the test.

§16.1 Basis and Dimension

Definition 16.1. A *basis* is a linearly independent family that spans the whole space.

Lemma 16.2

If $L\colon E \to \mathbb{R}^d$ is injective (i.e., $\ker L = \{0\}$), then $\dim E = \operatorname{im} L \le d$.

Injectivity means that we can *embed* E into $\mathbb{R}^d$; the idea is that if we can put E into a space of dimension $\mathbb{R}^d$, then it shouldn't have dimension more than d (e.g., you can't put a box into a plane).

Remark 16.3. A linear operator L is injective if and only if $\ker L = \{0\}$ — this is because $L(y_1) = L(y_2)$ if and only if $L(y_1 - y_2) = 0$.

Proof. Last time, we constructed vectors in the following way — we took $z_1, \ldots, z_k$ to be a basis of $\operatorname{im} L$. We want to bring these vectors back to E through L — we take vectors $y_1, \ldots, y_k$ such that $L(y_i) = z_i$ (so $z_i \in \mathbb{R}^d$ and $y_i \in E$).

We will show that $y_1, \ldots, y_k$ is a basis of E. To show this, we need to verify that they are linearly independent and that they span E.

First we'll check linear independence. Assume there are $\lambda_1, \ldots, \lambda_k$ such that $\lambda_1 y_1 + \cdots + \lambda_k y_k = 0$. Then applying L to both sides gives

$$L(\lambda_1 y_1 + \cdots + \lambda_k y_k) = L(0) = 0.$$

(The fact that $L(0) = 0$ for any linear operator can be seen by the fact that $L(0 \cdot x) = 0 \cdot L(x) = 0$.)

Now expanding using linearity gives

$$\lambda_1 L(y_1) + \cdots + \lambda_k L(y_k) = 0.$$

Since by definition $L(y_i) = z_i$, then we have

$$\lambda_1 z_1 + \cdots + \lambda_k z_k = 0.$$

But since $z_1, \ldots, z_k$ are a basis, then all λ_i must be 0.

Next we'll check that they span the space. Let $y \in E$; we want to write y as a linear combination $\lambda_1 y_1 + \cdots + \lambda_k y_k = y$. We again want to pull back a property from the target space, so we'll define $z = L(y)$. Since the z_i form a basis for $\operatorname{im} L$, there exist $\lambda_1, \ldots, \lambda_k$ such that

$$z = L(y) = \lambda_1 z_1 + \cdots + \lambda_k z_k.$$

By definition and the linearity of L, this is equal to

$$\lambda_1 L(y_1) + \cdots + \lambda_k L(y_k) = L(\lambda_1 y_1 + \cdots + \lambda_k y_k).$$

Now this tells us that

$$L(y - (\lambda_1 y_1 + \cdots + \lambda_k y_k)) = 0,$$

and since $\ker L = \{0\}$ this tells us we must have $\lambda_1 y_1 + \cdots + \lambda_k y_k = y$. □

Linear algebra is at the core of a lot of things; this is one type of proof that shows up often in linear algebra. For us, it's important because it tells us something about the set of solutions of our ODEs:

Corollary 16.4

The set of solutions of a second-order ODE

$$M(y) := y'' + p(t)y' + q(t)y = 0$$

on an interval I is at most 2-dimensional.

(The set of solutions with 0 replaced by $f(t)$ is an affine subspace whose direction is given by this vector space, so the same applies.)

Proof. We want to define a linear operator from the space of these solutions to $\mathbb{R}^2$, and show that it is injective. Let $E = \ker(M)$ be the space of solutions. Then we define the map $L\colon E \to \mathbb{R}^2$ as $y \mapsto (y(t_0), y'(t_0))$, where t_0 is any element of I.

Then injectivity follows from our proof of uniqueness — $\ker L$ is the set of solutions to

$$\begin{cases} y'' + p(t)y'(t) + q(t)y(t) = 0 \\ y(t_0) = y'(t_0) = 0. \end{cases}$$

From our uniqueness proof, we know that the only solution to this equation is $y = 0$, and therefore $\ker L = \{0\}$. (We used the assumptions that p and q were bounded, but this isn't really necessary — if they're continuous, then they're bounded on every finite interval, which is enough, since it shows that y must be 0 on every finite interval and therefore everywhere.) □

Remark 16.5. The operator M itself is a map from an infinite-dimensional space to another infinite-dimensional space, but we've shown that its kernel E is actually finite (and is 2); this is not trivial, but it's a manifestation of something we'll see a lot in analysis — often you can go from infinite-dimensional to finite-dimensional problems through what's called *Fredhol theory*. We'll see some simple examples, but later we'll see that this is a very powerful tool in analysis (we'll say more about it in a few weeks).

§16.2 The Wronskian

We'll now show that the dimension cannot be 1 — i.e., it has to be either 0 or 2. We'll do this by introducing a useful tool, called the *Wronskian*.

Definition 16.6. Let $I \subseteq \mathbb{R}$. The *Wronskian* $W\colon C^1(I) \times C^1(I) \to C^0(I)$ is defined as

$$W\colon (f, g) \mapsto fg' - f'g.$$

(In general the Wronskian sends $C^{k+1}(I) \times C^{k+1}(I) \to C^k(I)$ — this is because we take one derivative.)

More explicitly, $W(f, g)$ is defined as the function where for every t we have

$$W(f,g)(t) = f(t)g'(t) - f'(t)g(t) = \det \begin{bmatrix} f(t) & g(t) \\ f'(t) & g'(t) \end{bmatrix}.$$

Note that the Wronskian is anti-symmetric, i.e., $W(f, g) = -W(g, f)$.

The key point for us is the following lemma:

Lemma 16.7

Let f and g be solutions to $y'' + p(t)y' + q(t)y = 0$. Then for some $t_0 \in I$, we have

$$W(f,g)(t) = c(f,g)e^{-\int_{t_0}^{t} p},$$

where $c(f, g) = W(f, g)(t_0)$.

We can read off the constant from the initial conditions, since they tell us what f and f' are at t_0.

This statement should make us think of a first-order ODE, and that's exactly how we prove it.

Proof. Let's take the Wronskian and differentiate it once (f and g are C^2, so their Wronskian is C^1); this gives

$$W(f,g)' = f'g' + fg'' - f''g - f'g' = fg'' - f''g.$$

Now we can use our equation to replace the second derivatives with their values; this gives

$$W(f,g)' = f(-p(t)g' - q(t)g) - (-p(t)f' - q(t)f)g = -p(t)(fg' - gf') = -p(t)W(f,g).$$

Now we can use our solution to the first-order ODE to finish. □

The main reason we care about the Wronskian here is that once you know one solution, through the Wronskian you can construct another solution that's linearly independent from it — so if our space is at least one-dimensional, then it is actually two-dimensional.

Another nice thing about this is that the exponential on the right-hand side is always nonzero. (This is one motivation for the Wronskian, but in practice this isn't usually why we use it.)

Theorem 16.8

For two solutions f and g, the following are equivalent:

- For all $t \in I$ we have $W(f,g)(t) \neq 0$.
- f and g are linearly independent.
- There exists $t_0 \in I$ such that $W(f,g)(t_0) \neq 0$.

Clearly (1) implies (3); the lemma above implies that (3) implies (1) as well.

In order to prove this theorem, we will show that (1) implies (2) implies (3) implies (1).

§17 March 17, 2023

§17.1 The Wronskian

Definition 17.1. The *Wronskian* of f and g is defined as

$$W(f,g) = fg' - f'g.$$

Theorem 17.2

Let f and g be solutions of the ODE $y'' + p(t)y' + q(t) = 0$. Then the following are equivalent:

(1) For all $t \in I$ we have $W(f,g)(t) \neq 0$.

(2) f and g are linearly independent.

(3) There exists $t_0 \in I$ for which $W(f,g)(t_0) \neq 0$.

The Wronskian can be used to track whether two solutions are linearly independent — intuitively, this is because

$$W(f,g)(t) = \det \begin{bmatrix} f(t) & g(t) \\ f'(t) & g'(t) \end{bmatrix},$$

and the two vectors $(f(t), f'(t))^\intercal$ and $(g(t), g'(t))^\intercal$ are 'initial conditions' at t. This means they determine the solution; so the solutions will be linearly independent if and only if these initial conditions are.

However, we'll now prove this naively without using this intuition.

Proof. We'll show that (1) $\implies$ (2) $\implies$ (3) $\implies$ (1).

First we'll prove (1) implies (2), by showing that if (2) is false, then (1) is false. If (2) fails, then there exists $\alpha \in \mathbb{R}$ such that either $f = \alpha g$ or $g = \alpha f$. We can check that

$$W(f, \alpha f) = W(\alpha g, g) = 0$$

for all functions f and g and scalars α (here we don't even need to use the equation).

(The only reason we need to consider both possibilities is that α could be 0 — it'd be equivalent to say that either $f = \alpha g$ or $g = 0$.)

Similarly we'll show that if we don't have (3) then we don't have (2). Since we don't have (3), for all $t \in I$ we must have $W(f, g)(t) = 0$. We want to show this implies our two functions have to be proportional. (Here it makes sense to use the contraposition because it's hard to start a proof with 'I cannot solve an equation,' whereas the contrapositive gives us a concrete equation to start with. Note that contraposition is logically the same as doing an argument by contradiction, but is more elegant according to some people who don't like arguments by contradiction.)

This means for all $t \in I$ we have

$$\det \begin{bmatrix} f(t) & g(t) \\ f'(t) & g'(t) \end{bmatrix} = 0,$$

and therefore $(f(t), f'(t))$ and $(g(t), g'(t))$ must be proportional — assume $g \neq 0$, so for all $t \in I$, there exists some $\lambda(t)$ such that $f(t) = \lambda(t)g(t)$ and $f'(t) = \lambda(t)g'(t)$. (Here we're using the fact that if g and g' are both 0 at any point, then they're identically 0.)

The first equation, by the product rule, gives

$$f(t) = \lambda(t)g'(t) + \lambda'(t)g(t).$$

(Note that at every point t, either g or g' is nonzero, so we can divide by it in the corresponding equation to obtain that $\lambda(t)$ is either $\frac{f}{g}$ or $\frac{f'}{g'}$. Since f, g, f', and g' are all continuously differentiable, the result follows.) Meanwhile the second equation says that $f'(t) = \lambda(t)g'(t)$. Taking their difference, we get that

$$\lambda'(t)g(t) = 0.$$

This requires that for all $t \in I$, we must have $\lambda'(t) = 0$ or $g(t) = 0$.

Lemma 17.3

Either the set of zeros of g is discrete, or $g = 0$.

Proof. If there exists an accumulation point of zeros of g — i.e., if there exist $t_i \to t_\infty \in I$ such that $g(t_i) = 0$ for all i — then we must have $g(t_\infty) = 0$ and $g'(t_\infty) = 0$ (the second statement is left as an exercise; it follows from the fact that our function is C^1). This means g is a solution to

$$\begin{cases} g'' + p(t)g' + q(t)g = 0 \\ g(t_\infty) = 0, g'(t_\infty) = 0. \end{cases}$$

But there's only one solution to this equation, namely 0; this implies $g = 0$ by uniqueness. □

(In particular, this tells us we cannot have an accumulation of zeros of an ODE with continuous coefficients, i.e., something like $t \sin \frac{1}{t}$.)

In particular, this means away from a discrete set of points, $\lambda'(t) = 0$ — so $\lambda'(t)$ is always 0 except at a few points, which don't accumulate. This implies (since λ' is continuous) that we must have $\lambda'(t) = 0$

everywhere. (*Discrete* means *locally finite* — i.e., the intersection of the zero set with any closed interval $[a, b]$ is finite.)

To show this, let's assume that on an interval $[a, b]$ there exists finitely many points $t_1 < \cdots < t_n$ such that $\lambda'(t) \neq 0$ if $t \neq \{t_1, \ldots, t_n\}$. Then since λ' is continuous, it immediately follows that since it's 0 on the left-hand side and right-hand side of each of these points, it must be 0 at these points as well.

Alternatively (if we didn't have this much regularity), we could do the following: let $t, s \in [a, b]$ and assume without loss of generality that $t_i < t < t_{i+1} < \cdots < t_j < s < t_{j+1}$; we want to compare λ at s and t. We can write

$$\lambda(s) = \lambda(t) + \int_t^{t_{i+1}} \lambda' + \cdots + \int_{t_{j-1}}^{t_j} \lambda' + \int_{t_j}^{s} \lambda'.$$

All of these integrals are equal to 0 (since the value of the integral on the closed and open intervals are the same — the values at the endpoints don't matter — and on all the open intervals we have $\lambda' = 0$). So then $\lambda(s) = \lambda(t)$ for all s and t in our interval; therefore λ is constant, and its derivative is 0.

Now we have that λ is constant, and so $f = \lambda g$ for a constant λ; this means f and g are not linearly independent. So we have shown that (2) implies (3).

> **Remark 17.4.** It's also possible to prove this using uniqueness — f and g satisfy the initial conditions for one to be a multiple of the other, so by uniqueness this must be the case.

Finally we'll show that (3) implies (1) — i.e., that if we don't vanish at one point, then we don't vanish anywhere. To see this, last class we showed that the Wronskian satisfies a first-order ODE, and therefore

$$W(f, g)(t) = W(f, g)(t_0) e^{-\int_{t_0}^{t} p}.$$

By assumption $W(f, g)(t_0) \neq 0$, and the exponential is nonzero as well; this means $W(f, g)(t) \neq 0$ for all t. So we are done. □

> **Remark 17.5.** You usually don't want to show that two functions are linearly independent by computing their Wronskian — you can tell just by looking whether they're proportional.

§17.2 Using the Wronskian to Find Solutions

One important use of the Wronskian is to find solutions to the ODE. Consider the homogeneous equation

$$y'' + p(t)y' + q(t)y = 0. \tag{H}$$

Assume that we have one solution (this is quite common — often one solution is obvious (e.g., the equation looks polynomial and we found a polynomial solution), or there's one solution whose behavior we understand and we want to say something about the other), and call this solution g. Assume also that $g \neq 0$ on an interval $J \subseteq I$. (It's okay if our solution vanishes at some points; we just want to stay away from these points at first, and then maybe try to match the results we've obtained on all these intervals.)

> **Question 17.6.** Can we produce a linearly independent solution f?

This is a natural question, which can be used in a number of ways (to find another solution, or to show that it's not possible for every solution to satisfy a certain property); it's extremely useful that given one solution we can explicitly get another one. (This is maybe the only reason we should care about the Wronskian.)

The idea is that we have

$$\left(\frac{f}{g}\right)' = \frac{f'g - g'f}{g^2} = -\frac{W(f, g)}{g^2}.$$

Moreover, we know $W(f,g)$ up to a constant, so we can write

$$\left(\frac{f}{g}\right)' = -\frac{ce^{-\int_{t_0}^t p}}{g^2},$$

and $c \neq 0$ if f and g are linearly independent. We may as well assume that $-c = 1$ (up to rescaling f), which gives the following.

Theorem 17.7

The function f given by

$$f(t) = g(t)\int_{t_0}^t \frac{e^{-\int_{t_0}^s p}}{g(s)^2}\,ds$$

is a solution of (H) and is linearly independent from g.

This is a remarkable result — given one solution, it lets us explicitly obtain a second one.

Proof. Showing this is indeed a solution will be on the problem set; here we'll show that they are linearly independent (essentially reusing what we wrote above, which is how we found the solution). To show that our functions are linearly independent, we just need to show that f and g are not proportional. Assume for the sake of contradiction that there exists $\alpha \in \mathbb{R}$ such that $f = \alpha g$. (We know that g does not vanish, so we don't need to consider the other case.) Then we must have

$$\int_{t_0}^t \frac{e^{-\int_{t_0}^s p}}{g(s)^2}\,ds = \alpha$$

for all t. In particular, taking the derivative gives

$$0 = \frac{e^{-\int_{t_0}^t p}}{g(s)^2}.$$

But both quantities on the right-hand side are always positive, so this is a contradiction. □

Corollary 17.8

The space of solutions is either 0-dimensional (i.e., the only solution is 0) or 2-dimensional.

We'll see a bit later that it's actually always 2-dimensional (at least, if our functions are continuous) — we'll always have one solution by an existence result (proven in the same way as our existence-uniqueness result), and we know it's two-dimensional for this reason.

Next time we'll look at the influence of the parameter p — in physics we often have equations like

$$y'' + p(t)y' + q(t)y = 0,$$

where y'' is the acceleration, $q(t)$ comes from some kind of conserved force, and $p(t)$ comes from friction or something similar. One of the key things in the Wronskian (and this formula) is that we see p but not q.

§18 March 20, 2023

§18.1 Influence of p(t)

We'll now continue our study of second-order ODEs; today we'll look at the influence of $p(t)$ on our solutions.

This term appears very often in physics, when you're looking at a problem in mechanics you have acceleration equal to some force. One of the forces depends just on position (the term $q(t)y$, which doesn't depend on the time derivative); this is typically something that's conserved, while the $-p(t)y$ term usually corresponds to energy dissipated through friction.

Mathematically, this type of operation appears very often when you try to solve some equation with symmetry. For instance, the simplest example is if you consider the Laplacian

$$\nabla = \partial_{x^2}^2 + \partial_{y^2}^2,$$

which we can write in polar coordinates. THis is a PDE, but we'll see a way to solve this PDE as an ODE, which will look like

$$y'' + \frac{y'}{t} - n^2 y/t^2 = 0$$

(if we want to understand the zeroes of our Laplacian). What's important geometrically is the $\frac{1}{r}$ lets you do some generalized versions of integration by parts. Analytically, you'd like to be able to do integration by parts with your Laplacian. On $\mathbb{R}$, the Laplacian is self-adjoint — if you have smooth functions, then (assumign functions and their derivatives vanish at 0 and 1) we have

$$\int_0^1 y''z = -\int_0^1 y'z' = \int_0^1 yz''$$

by integration by parts. So we took our two derivatives and moved them to the other side. (This works as long as there are no boundary terms at 0 and 1.)

In $\mathbb{R}^2$, we would like something similar — we would like to say

$$\int (\nabla y) z \, dv? = \int y \nabla z \, dv.$$

But to do this, we have to explain what is our volume form dv. In rectangular coordinates it's just $dx\,dy$. But in polar coordinates, this bbecomes

$$\int_{r=0}^1 \int_{\theta=0}^{2\pi} y''(r) + \frac{y'(r)}{t} z \cdot r \, dr \, d\theta = 2\pi \int_0^1 \left(y''(r) + \frac{y'(r)}{r} \right) z(r) \, dr.$$

(We're assuming y only depends on r.) We can try again to integrate by parts. We get

$$2\pi \int y''(r) z(r) r \, dr + 2\pi \int_0^1 y'(r) z(r) \, dr.$$

By the same computation from earlier this is

$$-2\pi \left(\int_0^1 (y'z'r + y'z) + \int_0^1 y'z \right).$$

Stuff cancels out. So this is equal to

$$2\pi \int_0^1 y \left(z'' + \frac{z'}{r} \right)$$

And the reason this works is because of teh r. (The thing we're using is $\partial^2 + \frac{\partial}{r} + \partial_{\theta^2}^2 = \nabla$.)

This is one of the reasons why $p(t)$ is crucial — geometrically it relates you to the volume form. Lots of operators in good coordinates need to have a term like this in order to have the right integration by parts.

This is one of the types of computations where you need to have the right p. Also p tells you a lot about your solutions through teh Wronskian — the evolution of the Wronskian only depends on p.

To understand what $p(t)$ is doing, let's try to rewrite our equation — try to plug in $y(t) = u(t)v(t)$ into (H). Our goal is to obtain another equation on u and v, with hopefully a simple enough v, to understand better the influence of p.

To do this, we then have

$$vu'' + (2v' + p(t)v)u' + (v'' + pv' + qv)u = 0.$$

The trick so to choose the right function v, namely the one that makes $\tilde{p}(t) = 2v' + p(t)v$ disappear. If we choose $\tilde{p}(t) = 2v' + p(t)v = 0$, we cknwo how to solve this ODE; it gives

$$v(t) = ce^{-\frac{1}{2}\int_{t_0}^{t} p}.$$

We choose this value of v to cancel out the first term, and now we've reduced to just an equation involving u'' and u, i.e., an equation without any $p(t)$ term.

So any solution to a second-order ODE with non-vanishing $p(t)$ can be rewritten as an equation for a new function without a term p. We may as well set our constant equal to 1, so now we want to look at solutions

$$y(t) = e^{-\frac{1}{2}\int_{t_0}^{t} p}u.$$

Then our equation becomes

$$u''(t) + (q(t) - \frac{p(t)^2}{2} - \frac{p'(t)}{2})u(t) = 0.$$

Call this new equation $\tilde{H}$.

Now we have a reduction; we went from an equation with three terms to one with two terms. So a lot of exercises will be reduced to a case with no p term, because of this trick (since we can go from one solution to the other through multiplying by this factor).

This tells us that the factor of p really just makes the function grow or decay according to the exponential. Also it explains why

$$W(t) = ce^{-\int_{t_0}^{t} p},$$

since we have

$$\tilde{W}(t) = c.$$

So the way the $p(t)$ term works is just by multiplying by this factor (up to changing the other parameter). Quite often, you do this reduction to reduce to the case where you just have something of this form $y'' + \tilde{q}(t)u = 0$. Many of our statements will therefore just be for this case.

From now on, for the next week, we'll just look at the equation $\tilde{H}$.

(For the explicit calculation,

$$v'(t) = -\frac{C}{2}p(t)e^{-\frac{1}{2}\int_{t_0}^{t} p}$$

and the same with v'', and then you divide out everything by the exponential — if you differentiate the exponential a bunch of times you always end up with it as a factor, and you can divide by it since exponentials are nonzero.)

§18.2 Solutions to TildeH ODEs

There are lots of results about these types of ODEs. Our first is to compare two solutions. A common way is to compare where they vanish.

Theorem 18.1

Assume that $q_1(t) > q_2(t)$ for all $t \in I$, and suppose we have two ODEs

$$y_1'' + q_1(t)y_1 = 0 \text{ and } y_2'' + q_2(t)y_2 = 0.$$

Then if y_2 has two consecutive zeros at $t_0 < t_0'$, then y_1 has a zero in $[t_0, t_0']$.

There will be lots of statements like this in the problem set. The idea is that we want to say something has to vanish somewhere. A common way to do that is to use the Wronskian — what happens to the Wronskian at different points, and between them?

Proof. Assume for the sake of contradiction that $y_1(t) \neq 0$ for all $t \in [t_0, t_0']$. We know the two zeros of y_2 were consecutive, so y_2 does not vanish on (t_0, t_0'). Without loss of generality we may assume that $y_1 > 0$ and $y_2 > 0$ on (t_0, t_0') (otherwise multiply them by -1).

As a picture, suppose we have t_0 and t_0'. We are saying that y_2 vanishes at our two yellow points and is 0 in between them, so we draw it as a curve up and down. This tells us in particular that the derivative at t_0 is nonnegative, and the derivative at t_0' is nonpositive — i.e.,

$$y_2'(t_0) \geq 0 \text{ and } y_2'(t_1) \leq 0$$

(you can prove this from the definition of the derivative). (In fact, by uniqueness you cannot have $y_2'(t_0) = 0$ because then y_2 is the zero function by uniqueness; but we probably won't need this.)

Now we consider the Wronskian

$$W(y_1, y_2)(t) = y_1 y_2' - y_2 y_1'.$$

We want to test this at the two only points where we have any information, namely t_0 and t_0'; these are convenient because the second term vanishes, and so

$$W(y_1, y_2)(t_0) = y_1(t_0)y_2'(t_0).$$

We know $y_2'(t_0) \geq 0$ and we assumed $y_1(t_0) > 0$, so therefore $W(y_1, y_2)(t_0) \geq 0$.

Similarly we have $W(y_1, y_2)(t_0') = y_1(t_0')y_2'(t_0') \leq 0$.

Now we'll look at the derivative of the Wronskian — we have

$$W(y_1, y_2)' = y_1'y_2' + y_1y_2'' - y_1''y_2 - y_1'y_2 = y_1y_2'' - y_1''y_2$$

(as in the single-equation case, the other terms cancel, which is very nice). This is equal to

$$y_1(-q_2y_2) - (-q_1y_1)y_2 = (q_1 - q_2)y_1y_2.$$

But by assumption $q_1 - q_2 > 0$, and $y_1 > 0$ and $y_2 > 0$ on (t_0, t_0').

Now let's draw W. At t_0 it's nonnegative and at t_0' it's nonpositive; but it's supposed to be strictly increasing. So this is a contradiction. □

This is one reason to look at the Wronskian — it gives us nontrivial information about not just one solution, but the comparison between two of them. It's often quite useful to say that we can't have two solutions with the same property (i.e., can't have two solutions vanishing at the same point that are linearly independent, or two solutions that decay to 0 at infinity). You want to think of teh Wronskian as a measure of some sort of area, and this area (once you get rid of the parameter p) becomes just a constant. If one side goes to 0, the other one can't go to 0 as welll —- if one function is very large, the other has to be very small for the area to stay the same. So we cannot have $(f, f')^\intercal$ and $(g, g')^\intercal$ both small at the same time, or else the determinant would become very small, contraditcting the fact that W is constant (e.g. 1). At the same time, both can't blow up. So you should think of teh Wronskian as saying that if one function is really large then the other is really small, and vice vrsa.

Corollary 18.2

If $q(t) < 0$ for all $t \in I$, then if $y \neq 0$ solves

$$y'' + q(t)y = 0,$$

it has at most one zero.

Proof. Apply the previous theorem with $0 = q_1$ and $q = q_2$. (This means we're comparign ourselves to $y'' = 0$, which are just affine functions; an affine function so vanish in the right places.) □

Remark 18.3. The Wronskian implies uniqueness — if two functions satisfy the same initial conditions, their Wronskian at that point is 0, and so the Wronskian is identically 0.

Very soon we will talk about some of the main arguments in analysis, such as the maximum principle, which is also oen of the tools used in uniqueness.

§18.3 Solving nonhomogeneous ODEs

So far, we've mostly worked with homogeneous ODES. NOw we'll see how to solve linear ODEs. There is a general technique for this, which we will now see.

We'll be solving

$$y'' + p(t)y' + q(t)y = f(t).$$

Again, the Wronskian will help us understand the technique. The technique is a trick, called *variation of the constant*. (This is kind of a trick, but the way to make it rigorous is using the tools we've developed.)

The trick is to replace constants by functions. What do we mean by this?

In the first-order situation (which we have already solved), consider the equation

$$y' + p(t)y = f(t)$$

(L). To solve this equation, we first solve the homogeneous one — $y' + p(t)y = 0$ (H). The solutions to (H) are $ce^{-\int_{t_0}^{t} p}$ (the c is the constant that we will vary).

The idea of variation of the constant is that we now try to look for a solution where we replace c by a function — i.e., we want to solve (L) where

$$y(t) = c(t) \int e^{-\int_{t_0}^{t} p}.$$

Let's define

$$y(t) = c(t)e^{-\int_{t_0}^{t} p}.$$

This should look familiar (it's the way we solved first-order ODES) — we can compute

$$y'(t) = (c'(t) - p(t)c(t))e^{-\int_{t_0}^{t} p},$$

so now our ODE (L) becomes

$$y' + p(t)y = c'(t)e^{-\int_{t_0}^{t} p} = f(t)$$

(the idea is that the second term cancels out with the $p(t)y$, and so we're just left with the derivative of c). This tells us

$$c(t) = \text{constant} + \int_{t_0}^{t} e^{\int_{t_0}^{s} p} f(s)\, ds.$$

Then we multiply by $e^{-\int p}$ and recover the solution to the first-order ODE.

That's what the method looks like for a first-order ODE; now we'll attempt to apply it to a second-order ODE. The idea is the same — recall that if we have

$$y'' + y = f(t)$$

(L), we know that the solutions to the associated homogeneous equation are of the form $c_1\cos(t) + c_2\sin(t)$. The variation of the ocnstant means we'll try to make these into actual functions — $c_1(t)\cos(t) + c_2(t)\sin(t)$. Now we search for a solution to (L) of this form, and hopefully we can make them solvable. The fact that the Wronskian doesn't vanish will be the key to solving these equations in the end.

§19 March 22, 2023 — Variation of Parameters/Constants

Last time we saw variation of parameters for first-order ODEs; now we'll see it for second-order ODEs (it works in any dimension). The idea is essentially that we have a basis of solutions with parameters in front of the functions, and we let these parameters become functions. (This technique is a quarter of 18.03; they'll apply it to a bunch of situations. Prof. Ozuch doesn't think it's that useful because WA will do it better than us — mathemtaically we don't care so much about computing the exact solutions — but we want to see where it comes from and how it extends to any order or dimension.)

This is one of the situations where there's a trick that becomes the way you solve this kind of problem.

Let (H) be the equation

$$y'' + p(t)y' + q(t)y = 0,$$

and let (L) be the equation

$$y'' + p(t)y' + q(t)y = f(t).$$

Suppose we know how to solve (H) — we know how to solve it when p and q are constant, and we've seen that from one solution we know how to find a second. But there are other situations where you can solve the homogeneous problem.

To start with, let's assume we can solve teh homogeneous problem, so we know a basis of solutions y_1 nad y_2. Then any solution is a linear combination of the two — the space of solutions is always 0-dimensional or 2-dimensional (we'll see soon it's always 2-dimensional). So there's always 2 solutions, and we want to ouse them as a basis to also solve the linear problem (L). If we didn't have this trick, it'd be a really hard problem, but luckily we have a way to solve this completely.

In the first-order situation we replaced one constant by a function. Here we have two constants, so we replace both by functions — the solutions of (H) are $t \mapsto c_1y_1(t) + c_2y_2(t)$ where c_1 and c_2 are constants.

Now we'll look for solutions to (L) of the form $t \mapsto y(t) = c_1(t)y_1(t) + c_2(t)y_2(t)$, where $c_1(t)$ and $c_2(t)$ are now functions.

Now we can compute

$$y' = c_1'y_1 + c_1y_1' + c_2'y_2 + c_2y_2',$$

The trick is we don't really want to differentiate c twice, so we will put together all of the things that correspond to differentiating it twice and we'll see we can assume it vanishes —

$$y'' = (c_1'y_1 + c_2'y_2)' + c_1'y_1' + c_2'y_2' + c_1y_1'' + c_2y_2''.$$

The thing $c_1'y_1 + c_2'y_2$ could be anything, but we're actually going to assume it's 0 — we'll see that we can always solve the problem assuming that this is the case. So let $c_1'y_1 + c_2'y_2 = 0$ (this is because it's an annoying term we don't want to differentiate the c's twice). Then

$$y' = c_1y_1' + c_2y_2'$$

and

$$y'' = c_1y_1'' + c_2y_2'' + c_1'y_1' + c_2'y_2'$$

under this assumption (we'll need to make sure later that we can find solutions satisfying this property, and this will be the case).

Now looking at our equation, we have

$$y'' + p(t)y' + q(t)y = c_1(y_1'' + p(t)y_1' + q(t)y_1) + c_2(y_2'' + p(t)y_2' + q(t)y_2) + c_1'y_1' + c_2'y_2'.$$

The first two expressions are 0 because y_1 and y_2 are solutions to (H). So what's left is just $c_1'y_1' + c_2'y_2'$, which means we simply want to solve

$$c_1'y_1' + c_2'y_2' = f(t).$$

This is not so bad, but there are way too many solutions to this (which makes it hard to find a solution — commonly it's easy to find something that's rare, because you have more information about it). What's missing is a second equation; but that's the equation (A) $c_1'y_1 + c_2'y_2 = 0$. We're going to ask for both of these conditions to be satisfied — (A) and (B) $c_1'y_1' + c_2'y_2' = f(t)$.

Now we have a system of ODEs. Does this look better than what we had before? (We haven't studied systems of ODEs yet; we'll see how to deal with them later. But right now we have two functions and two equations, which looks complicated.)

But the nice thing to realize is that for every t this is a linear equation, and we know how to solve linear equations — we can rewrite this as

$$\begin{bmatrix} y_1(t) & y_2(t) \\ y_1'(t) & y_2'(t) \end{bmatrix} \begin{bmatrix} c_1'(t) \\ c_2'(t) \end{bmatrix} = \begin{bmatrix} 0 \\ f(t) \end{bmatrix}.$$

We know how to solve this — it's possible to solve if and only if this matrix is invertible. But

$$\det \begin{bmatrix} y_1 & y_2 \\ y_1' & y_2' \end{bmatrix} = W(y_1, y_2)$$

is the Wronskian (so the Wronskian appears in weird places). We assumed that our two solutions y_1 and y_2 are a basis so are linearly independent, so their Wronskian doesn't vanish and this is nonzero; and therefore the system can be solved. Since we only have a 2×2 matrix, we can invert it explicitly to get an explicit solution; this gives

$$c_1'(t) = -\frac{y_2(t)f(t)}{W(y_1, y_2)(t)} \text{ and } c_2'(t) = \frac{y_1(t)f(t)}{W(y_1, y_2)(t)}$$

(using the formula for the inverse of a 2×2 matrix, where you change the diagonal terms and divide by the determinant). It is not that useful to remember this form; it's generally nicer to redo what we did here (just remember we don't want to differentiate c_1 twice, so you assume that all such things are 0 and we end up with a simpler form, and write the expression; then we know everything will vanish except the last term. So this is easier to do on your own rather than remembering the complicated formulas).

(This is one way to use the Wronskian — it tells you there will be solutions to linear equations.)

(You can also run this argument assuming $c_1'y_1 + c_2'y_2 = g(t)$; then you will end up with a $+g'(t) + p(t)g(t)$, which changes your equation slightly. You can assume anything you want as long as it lets you find one solution, and you will because the system we end up with is invertible.)

Now we can write

$$c_1(t) = -\int \frac{y_2 f}{W(y_1, y_2)} + \text{constant}_1$$

and

$$c_2(t) = \int \frac{y_1 f}{W(y_1, y_2)} + \text{cst}_2.$$

(This is the best way to solve these equations except in some specific situations, such as constant coefficients.)

When you have constant coefficients, in some situations you don't have to do all of this.

Example 19.1

Suppose that $y'' + ay' + by = a_n t^n + a_{n-1}t^{n-1} + \cdots + a_0$ where $a, b \in \mathbb{R}$ and we are assuming $f(t)$ is some polynomial.

In that case, we don't have to do all this — we can try to find a solution which itself is a polynomial. We search for $y = c_{n+2}t^{n+2} + \cdots + c_0 = \sum_{i=0}^{n+2} c_i t^i$. This will look like

$$y'(t) = \sum_{i=0}^{n+2} i c_i t^{i-1} = \sum_{j=0}^{n+1} (j+1)c_{j+1}t^j$$

and

$$y''(t) = \sum_{i=0}^{n+2} i(i-1)c_i t^{i-2} = \sum_{j=0}^{n} (j+2)(j+1)c_{j+2}t^j$$

(as long as $b \neq 0$ we could do this with just degree n instead, since the first two coefficients have to vanish; but there will be cases where we expand infinite series. The $n+2$ is just sort of the upper limit of the sum, and it won't really matter). Now comparing coefficients, we have

$$y'' + ay' + b = \sum_{j=0}^{n+2} ((j+2)(j+1)c_{j+2} + a(j+1)c_{j+1} + bc_j)t^j,$$

where we let $c_j = 0$ if $j > n+2$. Finding a solution to our ODE means we want this quantity to be equal to a_j.

Now solving the ODE becomes solving all of these — we have a sequence satisfying the requirement that all of this thing is equal to a_j. So now we can solve for c_j (we will end up with $c_{n+2} = c_{n+1} = 0$ as long as a and b are nonzero, since a_{n+2} is zero so $bc_{n+2} = 0$; then a_{n+1} is zero so $ac_{n+1} = 0$).

You solve this equation by looking first at the higher-order thing to get $c_n = \frac{a_n}{b}$, then plugging into the next coefficient, and so on — since we know everything except one number, we can always solve it iteratively. It's a little painful but it is really something we know how to do.

(Generally it's rare to have to do all these computations — you typically ask a computer to do it.)

That's one situation we can deal with without this technique. Another simpler situation (where variation of parameters isn't the best technique to solve them) is:

- If $f(t)$ is an exponential — i.e., $f(t) = e^{\alpha t}$
- or $e^{\alpha t}\cos\beta t$ or sin, or if
- $f(t) = e^{\alpha t}P(t)$ for a polynomial P.

In all these situations you search for a solution fo the same form — in the first case you search for $y(t) = ce^{at}$, which almost always works (in some situations it won't); in the second you search for $y(t) = ce^{(a+ib)t+c}$ and look at real or imaginary parts. (Solve everything in $\mathbb{C}$ and project.) In the third we search for $y(t) = e^{at}Q(t)$ for some othr polynomial Q.

The only case where you need to be careful is when $e^{\alpha t}$ is also a solution to the homogeneous thing. In that case you run into a problem, but it's not bad; you just need to add a polynomial in front of your exponential.

Now let's see an exapmle where we use variation of parameters.

Example 19.2

Solve $y'' + y = 0$ (H) — this gives $y_1 = \cos$ and $y_2 = \sin$. We want to solve

$$y'' + y = \frac{1}{\cos t}.$$

We define $y(t) = c_1(t)\cos t + c_2(t)\sin t$. Assume $c_1'\cos + c_2'\sin = 0$. Then

$$y'(t) = -c_1\sin t + c_2\cos t,$$

and

$$y'' = -c_1\cos - c_2\sin - c_1'\sin + c_2'\cos.$$

So then

$$y'' + y = -c_1'\sin + c_2'\cos = \frac{1}{\cos}.$$

This means our system is

$$\begin{bmatrix}\cos & \sin\\ -\sin & \cos\end{bmatrix}\begin{bmatrix}c_1'\\ c_2'\end{bmatrix} = \begin{bmatrix}0\\ 1/\cos\end{bmatrix}.$$

To solve this, the Wronskian is

$$\det\begin{bmatrix}\cos & \sin\\ -\sin & \cos\end{bmatrix} = \cos^2 + \sin^2 = 1$$

(recall that when $p = 0$ the Wronskian is always constant). The formula then tells us

$$c_1' = -\frac{\sin}{\cos} \text{ and } c_2' = \frac{\cos}{\cos} = 1.$$

So now integrating we obtain $c_1(t) = -\log(|\cos(t)|) + Cst_1$ (on the intervals where cos doesn't vanish) and $c_2(t) = t + Cst_2$.

Remark 19.3. In the higher-dimensional situation, it'll look very similar; we'll obtain an equation that looks like

$$\begin{bmatrix}y_1 & y_2 & \cdots & y_n\\ \vdots & \vdots & \cdots & \vdots\\ y_n^{(n-1)} & \cdots & y_n^{(n-1)} & \end{bmatrix}\begin{bmatrix}c_1'\\ c_2'\\ \vdots\end{bmatrix} = \begin{bmatrix}0\\ 0\\ \vdots\\ f(t)\end{bmatrix}$$

(The 0's come from teh same trick) The reason we can do this because the determinant ofhte matrix is $W(y_1, \ldots, y_n)$, which has the same behavior (it satisfeis a first-order ODE and never vanishes.)

This kind of trick iwll be the same as how we solve any order ODE — as a first-order one in a higher dimension.

§20 March 24, 2023

Today we don't want to start a new topic right before spring break, so we'll do some practice problems.

Example 20.1

When are the solutions to $y'' + ay' + by = 0$ bounded on $\mathbb{R}$?

We fully know how to solve this ODE — we consider the associated polynomial x^2+ax+b. The discriminant of this polynomial is $\Delta = a^2 - 4b$; the form of the solutions depends on the sign of Δ.

- If $\Delta > 0$, there are 2 real roots
$$\alpha_\pm = \frac{-a \pm \sqrt{a^2-4b}}{2}.$$
These induce the solutions to the ODE $t \mapsto Ae^{\alpha_+ t} + Be^{\alpha_- t}$, and every solution looks like this. An exponential is bounded on $\mathbb{R}$ if and only if it's constant, so here the solutions will never be bounded (meaning that for any a and b, not all the solutions will be bounded) — we know $\alpha_+ \neq \alpha_-$, so one of the two is nonzero, and for any $\beta \neq 0$, $e^{\beta t}$ is unbounded either as $t \to \infty$ or as $t \to -\infty$.
- If $\Delta = 0$, there is one real solution (with multiplicity 2) — so $\alpha = -\frac{a}{2}$ is a double root, and the solutions are $t \mapsto Ae^{\alpha t} + Bte^{\alpha t}$. Again these solutions are not bounded for any a and b — the only way for the exponential to be bounded is if $\alpha = 0$, in which case the second term is not bounded.
- If $\Delta < 0$, there are two complex conjugate roots
$$\frac{-a \pm i\sqrt{4b-a^2}}{2},$$
which lead to the solutions
$$e^{-at/2}\left(A\cos\frac{\sqrt{4b-a^2}t}{2} + B\sin\frac{\sqrt{4b-a^2}t}{2}\right).$$
In this case, if $a = 0$ then all solutions are bounded (since cos and sin are both bounded); otherwise the solutions are not generally bounded.

So all solutions are bounded on $\mathbb{R}$ if and only if $a = 0$ and $a^2 - 4b < 0$, i.e., $b > 0$. This tells us the solutions to a second-order ODE are almost always unbounded — the only case in which they are bounded is of the form $y'' + by = 0$ for $b > 0$, in which case the solutions are cos and sin.

Remark 20.2. There are many more situations in which the solutions are bounded as $t \to \infty$ or as $t \to -\infty$.

Example 20.3

Solve $y'' + 2y' + 4y = te^t$ with the conditions $y(0) = 1$ and $y(1) = 0$.

There's two ways to approach the problem:

- If we see a polynomial times an exponential on the right-hand side, we can look for a solution of the same form — so we look for $y = P(t)e^t$ for some polynomial $P(t)$.
- Use variation of parameters.

Using the first method, first consider the monomial $y(t) = t^n e^t$. Then $y'(t) = (nt^{n-1} + t^n)e^t$, and $y''(t) = (n(n-1)t^{n-2} + 2nt^{n-1} + t^n)e^t$. Then the left-hand side becomes

$$y'' + 2y' + 4y = (7t^n + 4nt^{n-1} + n(n-1)t^{n-2})e^t.$$

Here it suffices to consider $n = 1$ and 0 (because higher-degree terms will appear in the above expression). In order to take care of the t, we can take $\frac{1}{7}te^t$; this contains an extra factor of $\frac{4}{7}e^t$, which we get rid of by adding $-\frac{4}{49}e^t$.

Then we add an appropriate combination of the homogeneous solutions to match the initial conditions.

Alternatively, we can use variation of parameters — first we solve

$$y'' + 2y' + 4y = 0$$

to obtain that the solutions are $y_1 = e^{-t}\cos 3t$ and $y_2 = e^{-t}\sin 3t$. Then we attempt to solve the linear equation

$$y'' + 2y' + 4y = te^t$$

where $y(t) = c_1(t)y_1(t) + c_2(t)y_2(t)$. The equation we want to solve ends up being

$$\begin{bmatrix} y_1 & y_2 \\ y_1' & y_2' \end{bmatrix} \begin{bmatrix} c_1' \\ c_2' \end{bmatrix} = \begin{bmatrix} 0 \\ te^t. \end{bmatrix}$$

(This is slower — usually when we have a polynomial or exponential on the right-hand side, it's faster to search for solutions of the same form. But variation of parameters works as well.)

Example 20.4

Find the C^1 solutions $f\colon \mathbb{R}^{+,*} \to \mathbb{C}$ satisfying

$$f'(t) = if\left(\frac{1}{t}\right).$$

$\mathbb{R}^{+,*}$ denotes $(0, \infty)$.

First, since the right-hand side is C^1, then the left-hand side is C^1 as well, so f is C^2. Then differentiating again gives

$$f''(t) = -i \cdot \frac{1}{t^2} f'\left(\frac{1}{t}\right) = -i \cdot \frac{1}{t^2} \cdot if(t) = \frac{f(t)}{t^2},$$

and therefore

$$t^2 f'' - f = 0.$$

We don't exactly know how to solve this, but we can look for a solution of the form $g(t) = f(e^t)$. Then we get the nice ODE

$$g'' - g' - g = 0.$$

We can use this to solve for g (which we know how to do) and deduce the solution for f.

§21 April 3, 2023

We have

$$\Delta = \partial_{r^2}^2 + \frac{1}{r}\partial_r + \partial_{\theta^2}^2$$

in polar coordinates. (Δ is defined as the divergence of the gradient, or $\partial_x \partial_x f + \partial_y \partial_y f$). The divergence theorem says that

$$\int_{B(0,1)} \Delta u = \int \operatorname{div}(\nabla u)\, dx\, dy = \int_{S(0,1)} \nabla u \cdot \mathbf{n},$$

where $\mathbf{n}$ is the normal vector. (The integral of a divergence is a flux.)

You can compute all these things in polar coordinates instead. Then if you consider

$$\int_{B(0,1)} v \Delta u,$$

we have the formula

$$\operatorname{div}(u \nabla v) = \nabla u \cdot \nabla v + u \Delta v.$$

We can see the term $u\Delta v$ as an alaog to $v\Delta u$, so we also have

$$\operatorname{div}(v \nabla u) = \nabla u \cdot \nabla v + v \Delta u.$$

This then tells us that

$$\operatorname{div}(v \nabla u - u \nabla v) = v \Delta u - u \Delta v.$$

SO this tells us that

$$\int_{B(0,1)} v\Delta u - u \Delta v = \int_{B(0,1)} \operatorname{div}(v \nabla u - u \nabla v) = \int_{S(0,1)} (v \nabla u \cdot \mathbf{n} - u \nabla v \cdot \mathbf{n}).$$

If you want these formulas to work in polar coordinates, you need the second term — the $\frac{1}{r}\partial_r$ term. And if you write out these terms, you will see the Wronskian appear. We get

$$\int_{r=0}^{1} \int_{\theta=0}^{2r} v(r,\theta) \left(\partial_{r^2}^2 u + \frac{1}{r} \partial_r u + \partial_{\theta^2}^2 u \right) r\, dr\, d\theta.$$

The existence of the r in the $r\, dr\, d\theta$ means that we really need the $\frac{1}{r}\partial_r u$ term for the integration by parts to work out. (You can prove this in the case where there is no dependence on θ.) This is something you'll see often later.

§22 April 5, 2023 — Analysis of PDEs

This chapter will have ideas that are important in analysis, as well as geometry and algebraic geometry. We'll see many of these results in a much more general setting later; but we should remember the ideas.

The last part will be about systems of ODEs and will prepare us for dynamical systems, which is also quite popular.

There are a few ideas; these are the main ideas when you study PDEs. They're crucial but very simple ideas; we'll present them in a simple setting, so that when you see them in an abstract infinite-dimensional Banach or Hilbert space they'll feel easier.

- The maximum principle
- Integration by parts
- Fourier series and transforms
- Banach and hilbert geometry;
- Sobolev embeddings.

§22.1 The Maximum Principle

§22.1.1 Boundary Conditions

We've seen that to find one solution to an ODE, we need to specify initial conditions. Another way of specifying one solution is to specify *boundary conditions*.

We've been looking at ODEs of the form

$$\begin{cases} y'' + p(t)y' + q(t)y = f(t) \\ y(t_0) = y_0, y'(t_0) = y_1. \end{cases}$$

This gives *initial conditions*.

Boundary conditions are very similar, but instead of asking for the solution to have certain properties at t_0 (above, we specified the value of the function at t_0, and its slope — by adding these two constraints, we can understand teh full set of solutions). A boundary condition would be something like

$$\begin{cases} y'' + p(t)y' + q(t)y = f(t) \\ y(a) = y_a, y(b) = y_b \end{cases}.$$

To specify a solution, we need to specify two conditions (since the space of solutions is 2-dimensional); but boundary conditions may not specify them at the same point.

So for boundary conditions, we have two points t_a and t_b, and our solution is specified by the two points there.

In PDEs, it's usually more natural to give boundary conditions than initial conditions. What they look like for PDEs is — a PDE similar to our above one is for example

$$\Delta y + \langle \nabla y, \vec{p} \rangle + q \cdot y = f$$

where $y \colon \mathbb{R}^n \to \mathbb{R}$ is some function. Then the boundary conditions would really be about the boundary — suppose we look at this PDE on an open set $\Omega \in \mathbb{R}^n$, and let $\partial\Omega$ be its boundary. (We'll see a proper tdefinition later, but here's a picture (it means exactly what it sounds like))

For a pDE, fixing boundary conditions means fixing the value of our function y on the boundary — typically, we'll be looking at systmes

$$\begin{cases} (\text{PDE}) \\ y|_{\partial\Omega} = \gamma \end{cases}$$

where γ is some function.

When you have a PDE, boundary conditions are much more natural than initial conditions. (One setting where initial conditions make sense is if $\partial\Omega$ is just an interface, e.g., a line, and we want to solve it on the other side; we can ask for the function and its normal derivatives.)

We call this a *Dirichlet conditions* (the ones above with $y|_{\partial\Omega} = \gamma$). There are other conditions that make sense as well. The next simplest ones are what we call *Neumann conditions* —

for an (ODE), instead of having conditions on our function we have conditions on its derivative, e.g.,

$$\begin{cases} (\text{ODE}) \\ y'(a) = y_a, y'(b) = y_b. \end{cases}$$

These are *Neumann conditions*.

You can imagine also some linear combinations of the two; these are typically called *Robin conditions*, but they are less common, so we won't look at them.

Another condition that makes sense and is quite typical is to ask for the periodicity of a solution — sometimes asking for your solution to be periodic will fix it. That will be important in our next chapter, though we won't see it right now.

(For a PDE you automatically have infinite-dimensional solutions — so you need more initial conditions, an entire boundary — this si because the boundary is infinite instead of just two points. Physically for instance, these boundary conditions may mean something — in geometry people like minimal surfaces, so if you're looking for a surface that minimizes area, the boundary will be where the surface touches the boundary, and so if you have a soap bubble it will typically solve a PDE given a boundary condition. If you're solving the heat equation (what will be the temperature in this room in an hour), you need the outside conditions, e.g., the temperatures of the walls.)

§22.2 Maximum Principle

The maximum principle says that the boundary conditions control the solutions, under some nice enough assumptions — in particular, the maximum or minimum of the function has to be obtained at the boundary, under nice enough conditions.

Consider the Dirichlet problem

$$\begin{cases} y'' + p(t)y' + q(t)y = f(t) \\ y(0) = 0, y(1) = 0. \end{cases}$$

(The result also extends to the case when $y(0) \leq 0$ and $y(1) \leq 0$.) Suppose also that $f(t) \geq 0$ and $q(t) \leq 0$.

(This happens quite a lot if q or f is zero, or if you take the difference of two solutions of ODEs; so you find these assumptions in many situations.)

Then the maximum principle tells you something very nice:

Theorem 22.1

If $q(t) \leq 0$ for all t, and $f(t) \geq 0$ fora ll t, and $y(0) = y(1) = 0$, then $y(t) \leq 0$ for all $t \in [0, 1]$.

The reason this is claled the maximum principle is that we can restate it as that the maximum of y is attained at the boundary.

This is the maximum principle for elliptic equations; there is also one for parabolic equations, and hyperbolic equations. It's a principle that occurs a lot for very different sitations.

Proof. Assume we have a C^2 solution y. The idea is that if there exists $t_0 \in (0, 1)$ such that $y(t_0) > 0$, then there exists $t_{\max} \in (0, 1)$ such that $y(t_{\max}) = \max_{[0,1]} y$. (Draw a picture where we have O and 1 with values 0 and 0, and then if the function e.g., starts going down and then goes up and takes a positive value, there will be a maximum (since it's a continuous function on a compact interval), and this amximum will have to be positive because there's somem positive value.) So now we have our t_0 and our $t_{\max}$.

Now the idea is that we have a maximum, so what happens at the maximum? At $t_{\max}$, we have $y'(t_{\max}) = 0$ and $y''(t_{\max}) \leq 0$. (This is the important part of this proof — just what happens at the maximum of a function.)

Then plugging this into our ODE, we get that $q(t_{\max})y(t_{\max}) - f(t_{\max}) = -y''(\mathrm{t}_{\max}) \geq 0$. But we assumed that $q(t) \leq 0$, $y(t_{\max}) > 0$, and $f(t_{\max}) \leq 0$. So then $f(t_{\max}) = 0$, $y''(t_{\max}) = 0$, and $q(t_{\max}) = 0$.

Thsi is not a contradiction yet (if our inequalities on q or f were strict then we would be done). But we also know that $y'(t_{\max}) = 0$. Y

Uh.

Let's look at the case where either one of the two inequalities are strict. Then this is contradicted by the fact that everything has to vanish. It should be true with the weak inequalities, but Prof. Ozuch doesn't remember how to conclude. □

The maximum principle tells you that the boundary condition, under a quite common condition, really controls the solution inside. This is important because for PDEs, existence and uniqueness often comes just from saying that the solution is bounded — if you have a sequence of things that approach a solution, they have to be bounded, and then by some compactness theorem you can use the iteration argument we had for first-order ODES. SO controlling teh interior by the boundary is important for PDEs.

The PDE version of this is as follows:

Proposition 22.2

At a maximum of $y\colon \mathbb{R}^n \to \mathbb{R}$, you have $\nabla y = 0$, and $\Delta y \leq 0$.

So then you can use the same idea — the (PDE) at such a point becomes $q \cdot y - f \geq 0$, and so if q and f have the right signs with one of the inequalities strict, you will find that y has to be of the same sign.

Remark 22.3. It's a typical situation to not have any q or f. In higher dimensions, you will typically have the Laplace equation

$$\Delta y = 0$$

(e.g., a harmonic function) or Poisson equation

$$\Delta y = \rho$$

where ρ is some function. Here you will be able to say that if ρ has some sign, then y has some sign. For the harmonic thing $\Delta y = 0$, it says the maximum is attained at the boundary.

For the heat equation

$$(\partial_t y - \partial^2_{r^2} y) = 0$$

for a heat equation without a source, and $f(t, x)$ if there is a source, you will now have an x-direction and a t-direction (think of x as space and t as time). Then our boundary condition will be not just two points, but the whole function at the initial time, plus the two points at each time. That will be $\partial\Omega$. With a similar picture you can have the wave equation

$$(\partial^2_{t^2} y - \partial^2_{r^2} y) = 0,$$

and you'll have a maximum principle with similar boundary conditions. This is an example of Einstein's equatiosn in physics.

Every time, you have some principle that says if you have a solution defined on an open set, the maximum is always on the boundary; you have to figure out what the assumptions are, but ti's the same idea of looking at the maximum and seeing what inequalities have to be satisfied.

You might see a much more complicated maximum principle once you've seen more general functions and weak solutions of PDEs; it might not look like this statement at all, but it's the same idea.

§22.3 Weak Solutions of ODEs

Now we'll see an idea that revolutionized PDEs — maybe asking for solutions to be C^2 is not nice enough, and maybe we want something weaker.

Question 22.4. Are there 'solutions' that are not C^2, and how to make sense of them?

One reason we would do that is here when we proved existence and uniqueness, we could just take a sequence of functions and see they converge in our smooth C^1 topology. But that won't generally be true for PDEs. So we need to first show they converge in a weak sense, and then say that the weak solution is really a strong solution. So you really need this stsep of weak solutions to prove the existence of acutal solutions to most PDEs.

For ODEs, it's not extremely important, but you will see them very often in your future classes; and they will motivate Fourier series because we'll see a lot of weak solutions in Fourier series.

There are two ideas. The idea is integration by parts. The other idea we'll see later, for even weaker solutions, is Fourier series. (You can go even weaker, but this is enough for this class.)

Why do we want weak solutions? Maybe sometimes we get a PDE from some problem, but the problem isn't about functions with the same regularity.

Example 22.5 (Example/motivation)

Say we want to find the minima of some energy — e.g., what is the shortest path of a curve in $\mathbb{R}^2$? So we want to minimize the length of some curve

$$L(\gamma) = \int_0^1 |\dot{\gamma}(t)| \, dt$$

where $\gamma \colon [0,1] \to \mathbb{R}^2$ is some curve, and we've fixed the two ends of the curve. We can also write

$$L(\gamma)^2 = \int_0^1 |\dot{\gamma}(t)|^2 \, dt$$

if the speed is constant. The point is that this thing (the second) only has one derivative. But if you were to write the equation such a curve satisfies, the equation will have two derivatives — a critical point of L satisfies a second-order ODE. We will see that we're trying to minimize something that satisfies a first-order ODE, but the solution will satisfy a second-order ODE. So it can make sense to look at the C^1 solutions to the second-order ODE.

Another reason that appears a lot in physics is, say we have $f \colon [0,1] \to \mathbb{R}$ and we want to min $\int |f'|^2$ — this is really teh same problem. This is only about a first-order problem, but we will see that the solutions (the minimizers, given the boundary conditiosn e.g. $f(0) = f(1) = 0$, are exactly harmonic functions —- $f'' = 0$. You ca nsee the problem originally only had one derivative, but now you have two; that's quite typical. Th

These are the first situatiosn where you ahve a second-order ODE which is really only about something which should be differentiable once. WHen we learn how to solve ODes with Fourier series, it will often not be at all obvious that our solution is even a function; we'll have a weak solution that might not even be a function, but then we can upgrade its regularity to something that is very smooth. That is how we will solve a lot of PDEs, adn the idea is integration by parts.

§23 April 7, 2023

The main idea of the maximum principle is that at a maximum, $f' = 0$ and $f'' \leq 0$. Quite often, this can prevent the existence of a point where a function is positive (e.g., this is exactly how we proved that under some assumptions, a function has to have its extrema at the boundary). The same applies to PDEs; but

it's not just for C^2 functions, and there's a more general version that works even for functions which are not necessarily continuous.

§23.1 Weak Solutions

We want to make sense of being a solution to an ODE without being differentiable enough times to make sense of the ODE to start with.

Consider our usual ODE with constant coefficients, where

$$y'' + py' + qy = f(t).$$

We want to make sense of a solution y to this ODE which is not C^2, or maybe not even continuous. We have y'' here, so this seems questionable, but there *is* a way to make sense of this.

Assume that $y(0) = y(1) = 0$ for simplicity (if they are there, then you need to include them in your integration by parts). Call this system (ODE).

The idea is to integrate by parts against a *test function* — let's assume that we have a solution $y \in C^2$ to start with. Our goal is then to rewrite (ODE) as an equivalent statement that doesn't involve derivatives of y. For this, we will use a *test function*. This will be a function that's very smooth and has all the properties we want — C^∞ and compactly supported — and integrate by parts against it.

Suppose that y solves (ODE) on $[0, 1]$. Then for all functions $\varphi \in C^\infty([0, 1])$ such that $\varphi(0) = \varphi(1) = 0$ and $\varphi'(0) = \varphi'(1) = 0$ as well (again, for convenience so that our integration by parts doesn't contain boundary terms), we have

$$\int_0^1 (y'' + py' + qy - f)\varphi = 0$$

(since the (ODE) tells us that $y'' + py' + qy - f = 0$). But now we can integrate by parts to get rid of the derivatives on y, and put them on φ instead. Then this becomes

$$\int_0^1 -y'\varphi' - py\varphi' + qy\varphi - f\varphi = 0$$

(the boundary terms disappear because both functions vanish at the two boundary points). Integrating by parts again, this becomes

$$\int_0^1 y(\varphi'' - p\varphi' + q\varphi) - f\varphi = 0.$$

(Everything here is an equivalence, since we can also integrate by parts the other way.)

The point is that here there is no derivative on y! This gives us one sense in which we can define a weak solution.

Definition 23.1. We say that y is a *weak solution* to (ODE) if for all $\varphi \in C^\infty([0, 1])$ such that $\varphi(0) = \varphi(1) = 0$, we have that $\int_0^1 y(\varphi'' - p\varphi' + q\varphi) = \int_0^1 f\varphi$.

This is defined for functions y which are even less than continuous.

Of course, strong solutions are also weak solutions (this is what we did above — we assumed we had a C^2 strong solution, and we found a different way to express this without using the fact that our solution is C^2). Also, *a priori* a weak solution may not be continuous.

Remark 23.2. For ODEs, in fact being a weak solution is *equivalent* to being a strong solution. But there are solutions where you might have a weak solution you *don't* know has the correct amount of regularity — for example, consider the function

$$E(f) = \int_0^1 (f')^2,$$

and suppose we want to find the critical points. This function only involves one derivative. But in fact we will see that the critical points will satisfy $f'' = 0$ in a weak sense — where this weak sense will be slightly different.

Remark 23.3. There are different ways to define being a solution to an ODE which only requires weak assumptions. For example, we could even have taken $\int_0^1 -y'\varphi' - py\varphi' + qy\varphi - f\varphi = 0$ as our definition; this would be a H^1-weak solution, while our definition is L^2-weak.

Remark 23.4. In analysis, you can even define what it means to take a derivative of a function that isn't differentiable, or maybe isn't even a function. All of these ideas come from integration by parts — we want to integrate against a function that has all the nice properties we want.

§23.2 Distributions

One of the weakest senses of a weak solution is a *distribution*.

Definition 23.5. A *distribution* is a linear function on the space C_c^∞ of compactly supported functions.

Notation 23.6. C_c^∞ denotes the set of C^∞ functions which are compactly supported (i.e., φ is 0 on $[0,1] \setminus [\varepsilon, 1-\varepsilon]$ for some ε — this is a way to get rid of the boundary terms).

The first step is going from strong to weak — given a function f, we'd like to define a linear function L_f that acts on other functions, and the way we do this is by defining

$$L_f(g) = \int fg.$$

(We can take any assumptions we want — take $g \in C_c^\infty(\mathbb{R})$.)

This tells us that every function can be expressed as a linear function on the space of functions. But there are also things of this form which are *not* functions.

Example 23.7

- $L_{\delta_0} : g \mapsto g(0)$.
- $L_{\delta_0'} : g \mapsto g'(0)$.

This is a weird way to analyze functions, but it's quite powerful. (For example, the equation $\int_0^1 y(\varphi'' - py' + q\varphi) - f\varphi = 0$ is a statement about orthogonality.)

Now we'll see some different times of weak functions.

§23.3 Lebesgue and Sobolev Spaces

We'll talk about weaker solutions, but we'll also be able to talk about *geometry on functions* — by the end of next class, we'll be able to say that two functions form an angle of $\frac{\pi}{3}$, or that two functions are orthogonal,

or some function has a given length. This will be useful to solve ODEs for reasons we'll see later.

But before that, we need a bunch of definitions. In other classes, you might hear about e.g. Fourier series with weird assumptions, but the whole theory should be taught in L^2.

We'll now define $L^2([0,1])$. Intuitively, you should think of $L^2([0,1])$ as the set of functions f on $[0,1]$ such that $\int_0^1 f^2 < +\infty$ (this is not a formal definition because we're not giving regularity assumptions on f, so this statement doesn't make sense). The actual definition is the following.

Definition 23.8. $L^2([0,1])$ is the set of pointwise limits of sequences $f_m \in C^0([0,1])$ such that $\int (f_m)^2 < +\infty$.

Note that $C^0 \subseteq L^2$, but the converse is false — for example, $\int_0^1 (t^{-1/4})^2 < +\infty$. So we could shift it by taking the function $\left|t - \frac{1}{2}\right|^{-1/4}$ — this function is not even *defined* at $\frac{1}{2}$, but it's in L^2. (We can approximate it by a series of very pointy functions — we can make these functions continuous, or even C^∞, but their limits will not be continuous.)

Remark 23.9. You can come up with many different norms to define different spaces of functions — all the stories here will come from the L^2 norm, but you can use others.

Definition 23.10. The L^2-norm of f, denoted $\|\bullet\|_{L^2}$, is defined as $\|f\|_{L^2} = \left(\int_0^1 f^2\right)^{1/2}$.

The previous discussion of weak solutions is in some sense 'abstract nonsense' (putting this into context), but this is something we'll use this semester.

First, why is this a norm? Norms need to satisfy a few axioms:

- For all $\lambda \in \mathbb{R}$, we have $\|\lambda f\|_{L^2} = |\lambda| \cdot \|f\|_{L^2}$.
- $\|f + g\|_{L^2} \leq \|f\|_{L^2} + \|g\|_{L^2}$ (i.e., the *triangle inequality*).

The former is easy to check. In order to prove the latter, we'll use the Cauchy–Schwarz inequality.

Fact 23.11 (Cauchy–Schwarz Inequality) — We have $\int_0^1 fg \leq \left(\int_0^1 f^2\right)^{1/2} \left(\int_0^1 g^2\right)^2$.

Very soon, we will write this as $\langle f, g\rangle_{L^2} \leq \|f\|_{L^2} \cdot \|g\|_{L^2}$.

Proof of Triangle Inequality. Squaring the left-hand side to get rid of the square root, we have

$$\|f+g\|_{L^2}^2 = \int_0^1 (f+g)^2 = \int_0^1 f^2 + \int_0^1 g^2 + 2fg.$$

Applying the Cauchy–Schwarz identity, this is at most $\left(\left(\int_0^1 f^2\right)^{1/2} + \left(\int_0^1 g^2\right)^{1/2}\right)^2$ (if we expand out, the first two terms match up and the last is Cauchy–Schwarz), which gives the desired inequality. □

Cauchy–Schwarz is nice because it allows us to separate a product into two separate integrals (where we only look at one function at a time).

Remark 23.12. There's also a L^p space and norm (where we replace the exponent 2 with p), but we won't use it here.

Another useful space is the *Sobolev spaces* (which we'll link back to our previous notions of C^k via *Sobolev embedding*).

We'll first define $H^1([0,1])$. (Sobolev spaces are defined for every positive integer k.) Informally, $H^1([0,1])$ is the set of functions f in $[0,1]$ such that $f \in L^2$ and $f' \in L^2$. The formal definition is as follows:

Definition 23.13. $H^1([0,1])$ is the set of limits of functions $f_m \in C^0([0,1])$ such that $\int (f_m)^2 + (f_m')^2 < +\infty$.

Definition 23.14. The norm $\|f\|_{H^1}$ is defined as $\sqrt{\int_0^1 f^2} + \sqrt{\int_0^1 (f')^2} = \|f\|_{L^2} + \|f'\|_{L^2}$.

(You can also define it as $\sqrt{\int_0^1 f^2 + (f')^2}$.)

We saw that $C^0 \subseteq L^2$, but the converse is false. Similarly, we have $C^k \subseteq H^k$, but the converse is false. (H^k is defined in the same way, but with k derivatives instead of just one.)

Definition 23.15. For any k, $\|f\|_{H^k} = \|f\|_{L^2} + \cdots + \left\|f^{(k)}\right\|_{L^2}$.

Soon we will see the *Sobolev embeddinng theorem*, which tells us that H^{k+1} is embedded in C^k — this is very strong.

In fact, you can say $H^{k+1} \subseteq C^{k,\frac{1}{2}}$. What does a non-integer regularity mean? We saw what it means to be Lipschitz — we define $C^{0,1}$ to mean that $|f(x) - f(y)| \leq L\,|x-y|$ for some L. We can similarly define $C^{0,\alpha}$ (for any $\alpha \in (0,1]$) to mean that $|f(x) - f(y)| \leq L\,|x-y|^\alpha$; then $C^{k,\alpha}$ means that the kth derivative is $C^{0,\alpha}$.

A function which satisfies such an inequality is more than continuous; and these are good for many reasons. For example, the function $C^{k+2} \to C^k$ sending $y \mapsto y''$ is actually quite a bad operator, but if we add the condition $C^{k+2,\alpha} \to C^{k,\alpha}$ for any α, then things behave much better. (We saw that a function being Lipschitz means that our slope cannot be too vertical — we can draw this by bounding our function by the absolute value function at our point. The picture is similar for different α, but how we have curves instead of straight lines bounding our region — for example, the function $\sqrt{t}$ is $C^{0,1/2}$ but not $C^{0,1}$.)

So the Sobolev spaces are great for a lot of tasks, and they're also quite directly related to the C^k spaces that we know and like.

Similarly, the map $H^{k+2} \to H^k$ sending $y \mapsto y''$ is also well-behaved, though this comes from more advanced functional analysis that is mostly beyond the scope of this class.

(The C^k spaces are likely the first ones you'll see, but they're not suitable for a lot of analysis.)

(A Sobolev space is a *Hilbert space* because it comes from a dot product — we'll later define the dot product $\langle f, g\rangle_{L^2} = \int_0^1 fg$, and say that two functions are orthogonal if their dot product is 0, and $\|f\|_{L^2}^2 = \langle f, f\rangle_{L^2}$. This Hilbert structure means that there's many things you can do in L^2 but not L^p for general p — we can extend Pythagorean theorems (for example) to infinite-dimensional spaces very easily. All the H_k are Hilbert spaces as well — we can define $\langle f, g\rangle_{H^k} = \int fg + f'g' + \cdots + f^{(k)}g^{(k)}$.)

§24 April 10, 2023

§24.1 Minimizing Functionals

We'll now see an example which motivates the use of weak solutions, namely *Dirichlet energy*.

We define the functional $E\colon C^2([0,1]) \to \mathbb{R}$ according to

$$f \mapsto E(f) = \int_0^1 (f')^2.$$

Many equations in physics (e.g., electromagnetism and relativity) can be written as the critical points of some functional.

Suppose we'd like to find which functionals minimize this energy given some additional information. Suppose we require that $f(0) = f(1) = 0$.

Our goal is to show that we can go from such a functional to an ODE (and in higher dimensions, a PDE).

The first step is to linearize this equation — since we've seen that for *functions*, to look for critical points, we compute the derivative (which is a sort of linearization) and set it to 0.

Question 24.1. How can we compute a 'derivative' of E at f?

We'll want to define two things:

- The *differential* of E at f, denoted as $D_f E$;
- The *gradient* of E at f.

If f is a minimizer, then for all $h \in C^2([0,1])$ with $h(0) = h'(0) = h(1) = h'(1) = 0$ (we again use these assumptions to avoid boundary terms when we integrate by parts), we must have

$$E(f + \varepsilon h) \geq E(f)$$

for all $\varepsilon > 0$ (since every perturbation of our minimizer should be larger).

We have

$$E(f+\varepsilon h) = \int_0^1 ((f+\varepsilon h)')^2 = \int_0^1 (f' + \varepsilon h')^2 = \int_0^1 (f')^2 + \varepsilon \int_0^1 2f'h' + \varepsilon^2 \int_0^1 h''^2.$$

We'll call the second term — $\int_0^1 2f'h'$ — the *differential* of E at f in the direction of h, which we denote by $D_f E(h)$.

You should think of this as your usual linearization formula — if E were just a functional, we'd have $E(f+\varepsilon h) = E(f) + \varepsilon D_f E(h) + o(\varepsilon)$ — perhaps more familiarly, we would define

$$D_f E(h) = \lim_{\varepsilon \to 0} \frac{E(f+\varepsilon h) - E(f)}{\varepsilon}.$$

(When we do functional analysis, we'll see there are several such notions of differentials, but they all have the same idea; it's just a matter of how bad the space you're linearizing in is. Here since everything is linear, we can take this limit, so we can make sense of this differential.)

In our situation, this means

$$D_f E(h) = 2\int_0^1 f'h'.$$

This is a linearization of our functional, and just like for functions, a minimizer f must be a critical point of this differential — if f is a minimizer, then we must have $D_f E(h) = 0$ for all h (such that $h(0) = h(1) = h'(0) = h'(1) = 0$). This is for the same reason as with functions — imagine graphing $E(f+\varepsilon h)$ vs. ε. If the linearization in the direction of h is not zero, then we can move in one of the two directions and decrease our value of $E(f+\varepsilon h)$, contradicting the fact that f is a minimizer.

Definition 24.2. If $D_f E(h) = 0$ for all h, we say that f is a *critical point* of E.

This is how many PDEs appear; the name of the field that studies the variation properties of functionals is *calculus of variations*. As some examples:

Example 24.3

(1) The shortest path question in $\mathbb{R}^m$ at constant speeds — then for a curve $\gamma\colon [0,1] \to \mathbb{R}^m$, we consider $E(\gamma) = \int_0^1 |\dot\gamma|^2_{\mathbb{R}^m}$ (i.e., the length-squared).

(2) The shortest path question on a surface $\Sigma \subseteq \mathbb{R}^m$ — in this case, we consider curves $\gamma\colon [0,1] \to \Sigma$, and define the energy $E(\gamma) = \int |\dot\gamma|^2_{\mathbb{R}^m}$ in the same way. (The critical points in both this question and the one above are called *geodesics*.)

(3) We could even define the energy of a *surface* as $E(\Sigma) = \text{Area}(\Sigma)$. The critical points are called *minimal surfaces*.

(4) The integral of the scalar curvature gives Einstein's equations.

In our above example, if we consider

$$E(f) = \int_{\mathbb{R}^m} |\nabla f|^2,$$

then the critical points would be harmonic functions.

So far, we haven't really obtained an equation; we've just written the differential. In order to obtain an equation, we want to find the *gradient* associated to a linearization. There's a very general theorem that says these things always exist, but first we'll just consider this particular situation.

Question 24.4. What is the gradient of E at f?

For motivation, in $\mathbb{R}^n$ we define the gradient based on the *dot product* — given a function $u\colon \mathbb{R}^n \to \mathbb{R}$, we can consider its differential $du\colon \mathbb{R}^n \to \mathbb{R}$, where the differential of u at t in the direction s is

$$d_t u(s) = \lim_{\varepsilon \to 0} \frac{u(t+\varepsilon s) - u(t)}{\varepsilon}.$$

(If u is a function of one variable, then this is simply $s \cdot u'$.)

This *differential* is the most natural object in this setting; the *gradient* is much more dependent on the geometry of the space. But the way we would define the gradient $\nabla_t u$ is through the equation

$$d_t u(s) = (\nabla_t u) \cdot s.$$

In more complicated spaces than $\mathbb{R}^n$, this is the only way you can define the gradient — the differential will always come first, and from the differential you can define the gradient. The gradient will define on the geometry through the scalar product. (For example, in $\mathbb{R}^n$, the gradient will depend on an orthonormal basis — we choose the dot product so that $(1,0,\dots,0)$, $\dots$, $(0,1,\dots)$ is orthonormal. But if we had some bad vectors, we might not want to take the canonical basis. (This won't even make sense in some situations, like L^2.))

The geometry we'll use is the one given by L^2. We've seen that for our functional $E\colon C^2([0,1]) \to \mathbb{R}$, we can write

$$D_f E(h) = 2\int_0^1 f'h'.$$

We want to write this as a dot product of *something* with h — i.e., the integral of some function against h (using the definition of the dot product in L^2). The problem is we currently have a derivative of h, and we'll get rid of it using integration by parts — we have

$$D_f E(h) = 2\int_0^1 f'h' = -2\int_0^1 f''h = \langle -2f'', h\rangle_{L^2}.$$

So then we have $\nabla_f E = -2f''$.

Now our critical point equation becomes that $\nabla_f E$ has to vanish at the point we're considering — this gives us an ODE

$$\nabla_f E = -2f'' = 0.$$

In this simple example, we can say that any C^2 minimizer must satisfy the equation $-2f'' = 0$. This is quite a strong equation — the only solutions are affine functions, and with the boundary conditions $f(0) = f(1) = 0$ the only solution is 0. But there are many more interesting equations that can be written in this way — where we introduce a functional, calculate its differential $D_f E(h)$ by taking a limit of $E(f + \varepsilon h)$, and then using integration by parts to write it in terms of just h.

(There's an example in the problem set where we do something like this.)

This is one of the reasons we want to do some geometry in our spaces of functions — we want to be able to say that two functions are L^2-orthogonal and talk about a L^2-dot product in order to define all these objects.

§24.2 Some More Linear Algebra

To make this more concrete or formal, we'll do some more linear algebra, this time *Hilbert* linear algebra.

Definition 24.5. A *scalar product* (or *inner product*) on a $\mathbb{R}$-vector space E (or $\mathbb{C}$-vector space), is a map $\langle -, - \rangle \colon E \times E \to \mathbb{R}$ (or $\mathbb{C}$) satisfying the properties that for all $u, v, w \in E$:

- Symmetry — $\langle u, v \rangle = \langle v, u \rangle$ if E is a $\mathbb{R}$-vector space (and $\langle u, v \rangle = \overline{\langle v, u \rangle}$ if E is a $\mathbb{C}$-vector space).
- Bilinearity — $\langle \alpha u + \beta v, w \rangle = \alpha \langle u, w \rangle + \beta \langle v, w \rangle$ for all α and β in $\mathbb{R}$ (or $\mathbb{C}$).
- Positivity — $\langle u, u \rangle > 0$ if $u \neq 0$.

Such objects are generalization of the dot product. This is what will let us talk about orthogonality or the angle between two functions or the length of a function, and having this geometric intuition will be useful.

Example 24.6

(1) On $\mathbb{R}^n$, wecan take $\langle x, y \rangle = \sum_{i=1}^n x_i y_i$ (the standard dot product).

(2) On $\mathbb{C}^n$, we can take $\langle x, y \rangle = \sum_{i=1}^m x_i \overline{y_i}$.

(3) The L^2-product (which will be our main object of study) is

$$\langle f, g \rangle_{L^2} = \int_0^1 f \overline{g}.$$

(4) The H^k-product is defined as

$$\langle f, g \rangle_{H^k} = \int_0^1 f\overline{g} + f'\overline{g'} + \cdots + f^{(k)}\overline{g^{(k)}}.$$

Definition 24.7. We call a space E together with an inner product $\langle -, - \rangle$ a *pre-Hilbert space* or *inner product space*.

Definition 24.8. Let $(E, \langle -, - \rangle)$ be an inner product space. We say $f, g \in E$ are *orthogonal* if $\langle f, g \rangle = 0$.

This will be very helpful — again, what we want to do is extend our geometric intuition and results from two dimensions to any dimension.

We'll first see a bunch of examples — we'll see a bunch of functions that are orthogonal to each other. We'll change the definition of the dot product a bit so that we can use cos and sin.

Example 24.9

Define $\langle f, g\rangle = \int_0^{2\pi} f\overline{g}$

- $\cos \perp \sin$.
- $\cos \perp 1$ and $\sin \perp 1$.
- More generally, the function $t \mapsto \cos(mt)$ is orthogonal to the function $t \mapsto \cos(nt)$ for integers $m \neq n$. (The same is true for sin.)

Soon we will see that 1, cos, sin, and their higher frequencies form an *orthonormal basis* of L^2; this will be the key result of the second part of the semester. (This is one of the most important results in analysis, and in lots of applications to e.g., computer vision, radio frequencies, and so on.)

Definition 24.10. Let $L\colon E \to F$ be a linear map, where E and F are vector spaces which can both be embedded into a third space G (which has an inner product). We say that L is *symmetric*, or *self-adjoint*, if for all $f, g \in E$ we have $\langle L(f), g\rangle = \langle f, L(g)\rangle$.

Example 24.11

The map $L(y) = y''$ on $E = \{u \in C^2 \mid u(0) = u(1) = 0\}$ is self-adjoint (taking $G = C^0$) — we have

$$\int_0^1 f''g = -\int_0^1 f'g' = \int_0^1 fg''$$

by integration by parts, which gives $\langle L(f), g\rangle = \langle f, L(g)\rangle$.

The whole theory only works for such operators, but most operators we'll study will be of this form.

This may seem like abstract nonsense, but we'll see that there's a lot of stuff we can do with it — where very theoretical and abstract theory leads to a lot of concrete applications, and the best way to see what's happening is to go through the L^2 geometry.

§25 April 12, 2023

Definition 25.1. A *pre-Hilbert space* is a vector space E with an inner product $\langle -, - \rangle$, which satisfies the following conditions:

- $\langle u, v \rangle = \overline{\langle v, u \rangle}$.
- $\langle \alpha u + \beta v, w \rangle = \alpha \langle u, w \rangle + \beta \langle v, w \rangle$.
- $\langle u, u \rangle > 0$ for $u \neq 0$.

You should think of this as a generalization of $\mathbb{R}^n$ with the dot product; our goal is to extend the results we've seen on $\mathbb{R}^n$ with the dot product to the space of L^2 functions $(L^2([0,1]), \langle -, - \rangle_{L^2})$, where the dot product is defined as $\langle f, g \rangle = \int_0^1 f\bar{g}$. (Depending on context, we might scale the integral or change the boundary conditions.)

Definition 25.2. A function $L \colon E \to F$ (for E and F both subspaces of some pre-Hilbert space) is *symmetric* (or *self-adjoint*) if for all $f, g \in E$ we have $\langle L(f), g \rangle = \langle f, L(g) \rangle$.

§25.1 Eigenvectors

We'll need some linear algebra (e.g., eigenvectors and eigenbases); we'll do this in infinite dimensions.

Definition 25.3. We say that $y \in E$ is an *eigenvector* of L with eigenvalue λ if $L(y) = \lambda y$ and $y \neq 0$.

(We may call f an *eigenfunction*, since it is a function.)

Note that in some instances, it may not be obvious that the function we obtain as our eigenfunction is nonzero, so we'd need to verify it. (This doesn't seem important right now, but it's an easy mistake that leads to problems, since the theory collapses if you include 0.)

Example 25.4

cos and sin are eigenfunctions of the functional $L \colon y \mapsto y''$ with eigenvalue -1. (The same is true for $e^{\pm it}$.)

More generally, $t \mapsto \cos(\alpha t)$ and $t \mapsto \sin(\alpha t)$, or $t \mapsto e^{\pm i\alpha t}$, are eigenvectors of L with eigenvalue $-\alpha^2$ for any $\alpha \in \mathbb{R}$.

These will be our key examples for Fourier series.

Understanding an operator and its eigenvectors and eigenvalues can help us understand the space itself, as we will soon see — one of the key ways to prove that a family of functions is linearly independent is to prove that they are eigenfunctions of the same eigenvector.

Theorem 25.5

If $f_1, \ldots, f_n$ are eigenfunctions (i.e., eigenvectors) of a linear operator L associated to eigenvalues $\lambda_1, \ldots, \lambda_n$ which are pairwise distinct (i.e., $\lambda_i \neq \lambda_j$ if $i \neq j$), then $(f_1, \ldots, f_n)$ is linearly independent.

This gives a powerful way to construct linearly independent families. There are many situations where without this simple theorem, it's hard to prove that functions are linearly independent.

Before we prove this theorem, here is a corollary to show why it's important.

Corollary 25.6

Families of the form $(e^{\lambda_i t})_i$, $(\cos(\lambda_i t))_i$, and $(t^{\lambda_i})_i$ are linearly independent for $\lambda_i > 0$ distinct.

Without this theorem, these statements would be hard to prove — in fact, probably the simplest proof is to reprove the theorem for these examples.

Proof. For (1), we take the operator $y \mapsto y'$; for (2), we take the operator $y \mapsto y''$; and for (3), we take the operator $y \mapsto ty'$. In all these cases, we can verify that our functions are all eigenfunctions of these operators, with distinct eigenvalues; so we automatically obtain linearly independent families. □

For us, the two first ones will be the most important. But this theorem is quite general (and we will use it in finite dimensions as well), and we'll see that when L is symmetric we have even more.

Proof of Theorem. We use iteration on the number of vectors, which we denote n. In the case $n = 1$, this is obvious because eigenfunctions are by definition nonzero.

Now assume the property at n; we will prove it at $n+1$. Suppose we have $(\lambda_1, \dots, \lambda_{n+1})$ and $(f_1, \dots, f_{n+1})$ such that for all $i \in \{1, \dots, n\}$, we have $L(f_i) = \lambda_i f_i$.

Assume for contradiction that there exists $(\alpha_1, \dots, \alpha_{n+1})$, not all zero, such that $\sum_i \alpha_i f_i = 0$.

In linear algebra, the only things we can really do are take linear combinations and apply linear operators, so they're always worth trying. Here, let's apply our operator to this equality; we have $\sum_i \alpha_i f_i = 0$, and applying L to both sides, by linearity we get

$$\sum_i \alpha_i L(f_i) = 0.$$

Now using the fact that the f_i are eigenfunctions, this means

$$\sum_i \alpha_i \lambda_i f_i = 0.$$

Now we have *both* the equations $\sum \alpha_i f_i = 0$ and $\sum \alpha_i \lambda_i f_i = 0$. Both of these sums have $n+1$ terms, and the only thing we know is that the property is true with n terms, so our goal is to get rid of one of these $n+1$ terms. The key idea is to take a linear combination of these two quantities — we have

$$\gamma \sum_{i=1}^{n+1} \alpha_i f_i - \sum_{i=1}^{n+1} \alpha_i \lambda_i f_i = \sum_{i=1}^{n+1} \alpha_i (\gamma - \lambda_i) f_i = 0.$$

Conveniently, if we choose γ to equal one of the λ_i, then we can make one of these terms disappear — so we want to choose $\gamma = \lambda_{i_0}$ for some appropriately chosen i_0.

To do so, by assumption there exists i_0 such that $\alpha_{i_0} \neq 0$ (from the assumption we made towards a contradiction). So then we have

$$0 = \sum_{i=1}^{n+1} \alpha_i (\lambda_{i_0} - \lambda_1) f_i.$$

But we may as well remove i_0 from this sum, since the term for i_0 is simply 0.

(In this case we could have just taken $i_0 = n+1$ instead, but there exist situations where you need to be more careful.)

Now we have a sum of n eigenfunctions with distinct eigenvalues — define $\tilde{\alpha}_i = \alpha_i(\lambda_{i_0} - \lambda_i)$. Then the property at n applied to the set of n functions $(f_i)_{i \neq i_0}$ tells us that these functions are linearly independent, so we must have $\tilde{\alpha}_i = 0$ for all $i \neq i_0$. This means $\alpha_i(\lambda_{i_0} - \lambda_i) = 0$ for all $i \neq i_0$. But since $\lambda_{i_0} \neq \lambda_i$ (by the assumption that the eigenvalues are distinct), so for all $i \neq i_0$, we must have $\alpha_i = 0$. This implies α_{i_0} is also 0 (because $\alpha_{i_0} f_{i_0} = -\sum_{i \neq i_0} \alpha_i f_i$, and $f_{i_0} \neq 0$). □

The first idea of this proof was to apply the linear operator, since it's what we had information about. Then the second was that if we have two linear combinations, then we can combine them to get rid of one of the variables. This is often useful in iteration because it lets us go down to one less element in our sum.

This is something we might have seen with linear algebra on matrices.

Corollary 25.7

If $M \in \mathbb{R}^{n\times n}$ has n distinct eigenvalues, then it has a basis of eigenvectors.

What's really cool is that this extends to infinite dimensions as well. But first we'll see something else — if our operator is *symmetric* (this is the equivalent of a symmetric matrix in $\mathbb{R}^n$), then there is more. The eigenvectors associated to different eigenvalues will be not only linearly independent, but even *orthogonal*.

Theorem 25.8

If L is symmetric with respect to $\langle -, - \rangle$, and f_1 and f_2 are eigenvectors associated to the eigenvalues $\lambda_1 \neq \lambda_2$, then $\langle f_1, f_2 \rangle = 0$.

As a corollary in $\mathbb{R}^n$ (which we may have seen before):

Corollary 25.9

If $M \in \mathbb{R}^{n\times n}$ is symmetric (i.e., $M^\intercal = M$), then M admits an *orthonormal* basis of eigenvectors.

(In the complex case, you'd want the matrix to be *Hermitian* rather than symmetric, meaning $M = \overline{M}^\intercal$.)

This is really useful, because symmetric matrices appear a lot in many problems. One instance where it will be important for us is that if we have a C^2 function in $\mathbb{R}^n$, then its *Hessian* will be a symmetric matrix, so it will always be diagonalizable with an orthonormal eigenbasis. But this works not just in $\mathbb{R}^n$, but any pre-Hilbert space. The proof is even simpler than the previous one.

Proof. The only things we can do are take scalar products and apply L, so let's do that — we have

$$\langle L(f_1), f_2 \rangle = \lambda_1 \langle f_1, f_2 \rangle.$$

But since L is symmetric, this is also equal to $\langle f_1, L(f_2) \rangle = \lambda_2 \langle f_1, f_2 \rangle$. This implies $(\lambda_1 - \lambda_2)\langle f_1, f_2 \rangle = 0$. But by assumption $\lambda_1 \neq \lambda_2$; this means $\langle f_1, f_2 \rangle = 0$. □

§25.2 Decomposition of Vectors in an Orthonormal Basis

Now we've gone from having random function spaces to being able to do some geometry on them. Next time, we will prove the infinite-dimensional Pythagorean theorem (which is basically trivial given what we've done so far).

Definition 25.10. For some set S (possibly infinite), a collection of vectors $(e_i)_{i\in S}$ in $(E, \langle -, - \rangle)$ is an *orthonormal basis* if

$$\langle e_i, e_j \rangle = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise,} \end{cases}$$

and for every $f \in E$, there exist $(f_i)_{i\in S}$ (with $f_i \in \mathbb{R}$ or $\mathbb{C}$) such that $f = \sum_{i\in S} f_i e_i$.

The first condition is just called being *orthonormal*, and the second then corresponds to being a basis.

Everything we know about geometry on the plane will extend to any infinite-dimensional space of functions, with the same intuition.

Proposition 25.11
If (e_i) are orthonormal, then $f_i = \langle f, e_i \rangle$ for all i.

So then we can rewrite the formula in the definition of an orthonormal basis as

$$f = \sum_{i \in S} \langle f, e_i \rangle e_i.$$

As another example of a theorem that extends:

Theorem 25.12 (Pythagorean Theorem)
We have $\|f\|^2 = \sum_{i \in S} \langle f, e_i \rangle^2$, where $\|f\|^2$ is defined as $\langle f, f \rangle$.

§26 April 14, 2023

§26.1 Eigenvalues and Eigenvectors

Eigenvalues and eigenvectors are very important — we'll see them everywhere.

Definition 26.1. A *nonzero* vector y is an eigenvector of L with eigenvalue $\lambda \in \mathbb{R}$ (or $\mathbb{C}$) if $L(y) = \lambda y$.

(Note that in the midterm we'll have to name the eigenvalues of certain operators; every time, we should verify that our solutions are not zero, as this is not always obvious — this is important because if we allowed $y = 0$ then every λ would be an eigenvalue, so the notion would not be useful.)

Example 26.2
$t \mapsto \sin(\alpha t)$, $\cos(\alpha t)$, and $e^{i\alpha t}$ are eigenfunctions of $L(y) = y''$ with eigenvalue $-\alpha^2$.

Last time, we mentioned and proved the following key theorem.

Theorem 26.3
If $y_1, \ldots, y_n$ are eigenvectors of a symmetric operator L on $(E, \langle -, - \rangle)$ with *distinct* eigenvalues $\lambda_1, \ldots, \lambda_n$, then $\langle y_i, y_j \rangle = 0$ for all $i \neq j$.

This is a very powerful way to construct families of orthogonal (and then orthonormal) vectors — by finding a good operator and then the basis of eigenfunctions. This is the most common way to construct bases of functions; then once we have such a basis, we can write down every function in this basis. This is roughly the basic idea of Fourier series, Fourier transforms, Laplace transforms, Taylor expansion, and so on — in all of these situations, the idea is to choose a certain basis of functions and decompose functions in this basis.

Fourier's idea is that 'any function is a sum of cosine and sine' (with some caveats). In this week and next week, we'll work towards being able to understand what this means. Even nicer, we'll be able to say that if $f \in L^2([0, \pi])$, then

$$f(t) = \frac{a_0}{2} + \sum_{n \in \mathbb{N}} a_m \cos(mt) + b_m \sin(mt).$$

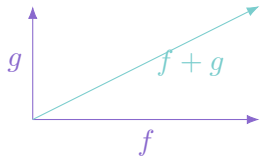
This will give an *orthogonal decomposition* of f. The statement of Fourier transforms will let us do this on infinite segments as well.

Similarly, the Laplace transform uses decompositions with $e^{\lambda\rho}$, and the Taylor expansion uses decompositions with t^n. But in each of these cases, the point to decompose with respect to a basis of eigenvectors.

§26.2 The Pythagorean Theorem

The nice thing about an *orthonormal* basis is that the Pythagorean theorem applies as usual (this is a very simple result to prove once we have the right setup).

Theorem 26.4 (Pythagorean Theorem)
Let $\|y\|^2 = \langle y, y\rangle$ (as in Euclidean space). If $\langle f, g\rangle = 0$, then $\|f+g\|^2 = \|f\|^2 + \|g\|^2$.



This is the usual Pythagorean theorem, but for functions. This is very useful because it lets us compute integrals — for example, we can say that

$$\int_0^{2\pi} (\cos t + \sin t)^2 = \int_0^{2\pi} \cos^2 t + \int_0^{2\pi} \sin^2 t.$$

This extends directly to any *finite* collection of orthogonal vectors — given a collection of functions $f_1, \ldots, f_n$ such that for all $i \neq j$ we have $\langle f_i, f_j\rangle = 0$, we have

$$\|f_1 + \cdots + f_n\|^2 = \|f_1\|^2 + \cdots + \|f_n\|^2.$$

But in fact, there's also an *infinite* version.

Theorem 26.5
If $(e_i)_{i\in S}$ is an orthonormal basis of E, then:

(a) For all $f \in E$, there exist $f_i \in \mathbb{R}$ (or $\mathbb{C}$) such that $f = \sum_{i\in S} f_i e_i$.

(b) $f_i = \langle f, e_i\rangle$.

(c) $\|f\|^2 = \sum_{i\in S} |f_i|^2 = \sum_{i\in S} \langle f, e_i\rangle^2$.

Note that we have to be a bit careful about the equality $f = \sum_{i\in S} f_i e_i$ — here we mean equality in the sense that $\|f - \sum_{i\in S} f_i e_i\| = 0$. Being zero in an L^2 sense is not quite the same as being zero in a continuous sense — for example, the function that is zero everywhere except at two isolated points has norm zero in the L^2 sense, but f is not quite zero. (This will only be a problem when the functions are continuous.)

(The correct way to deal with this is that in a L^2 sense we say that $f = g$ if $\|f-g\|_{L^2} = 0$ — this creates an equivalence relation, and you can quotient the usual space by this equivalence relation. This gives a slight subtlety about functions with low regularity which might not be pointwise equal to their Fourier series everywhere; but when f is C^1 we won't have this problem. But this is one of the reasons we need weak solutions — it's hard to get these formulas without going through L^2.)

(Right now we're assuming that S is countable. It's possible to make sense of it even if S is not countable — and we'll do this for Fourier transforms — but there we'd integrate instead of summing.)

Remark 26.6. Note that the e_i need to be *orthonormal*, not just orthogonal (otherwise we have to deal with inverses of certain expressions).

Our main application of this will be the following:

Example 26.7

Suppose $f \in L^2$ and e_n is the map $t \mapsto e^{int}$ for $n \in \mathbb{Z}$, with $\langle -, - \rangle = \frac{1}{2\pi} \int_0^{2\pi} fg$. Then from (a), we have $f = \sum_{n \in \mathbb{Z}} c_n e^n$ for some $c_n \in \mathbb{C}$; then (b) tells us that

$$c_n = \frac{1}{2\pi} \int_0^{2\pi} f e^{-int},$$

and (c) tells us that

$$\|f\|_{L^2} = \sum_{n \in \mathbb{Z}} |c_n|^2 .$$

First let's prove all these aspects of the Pythagorean theorem.

Proof. For the first statement $\|f+g\|^2 = \|f\|^2 + \|g\|^2$, we have

$$\langle f+g, f+g \rangle = \langle f, f \rangle + \langle f, g \rangle + \langle g, f \rangle + \langle g, g \rangle$$

by linearity. But $\langle f, g \rangle = \langle g, f \rangle = 0$, so this is simply equal to

$$\|f+g\|^2 = \|f\|^2 + \|g\|^2 .$$

The second statement is the same — we'll have a bunch of terms instead of just 2, but all will vanish except the terms $\langle f_i, f_i \rangle$.

For the infinite version, (a) is the definition of being a basis. For (b), we have $\langle f, e_i \rangle = \langle \sum_j f_j e_j, e_i \rangle = \sum_j f_j \langle e_j, e_i \rangle$ by linearity. But we know $\langle e_j, e_i \rangle$ is 0 if $j \neq i$, and 1 if $j = i$. So all the terms in the sum disappear except the one with $j = i$, and this becomes exactly f_i. □

In general, we have the extension of the Pythagorean theorem that

$$\|f+g\|^2 = \langle f+g, f+g \rangle = \|f\|^2 + 2\langle f, g \rangle + \|g\|^2 .$$

We also have the Cauchy–Schwarz formula

$$|\langle f, g \rangle| \leq \|f\| \cdot \|g\| .$$

This will let us define the *angle* between two functions — we'll say

$$\cos(\angle(f, g)) = \frac{\langle f, g \rangle}{\|f\| \cdot \|g\|}.$$

§26.3 Midterm Review

This is the last class before the next midterm, so we'll spend a bit of time on a few practice problems.

Consider the dot product

$$\langle f, g \rangle_w = \int_0^1 f(t) g(t) w(t)\, dt.$$

This dot product is used quite a lot to come up with bases of vectors — specific choices of w will give certain well-known families of polynomials (the Legendre polynomials, Laguerre polynomials, Hermitte polynomials, and others) which are orthogonal with respect to this given dot product. So in lots of settings, we may want to use this dot product.

There are two ways to look at this — one is that we might have an equation and want to find a nice dot product for it (i.e., one that makes it symmetric). Another is that we might want to choose a good weight to match some situation we're studying.

Here we'll think of the problem in the first way: suppose we *want* to have

$$\langle L(f), g\rangle_w = \langle f, L(g)\rangle_w.$$

Depending on what our problem is (which determines L), we'll want to take a different weight.

Exercise 26.8. Show that if $w > 0$ is C^0, then $\langle -, -\rangle_w$ is an inner product on functions $C^0 \colon [0,1] \to \mathbb{R}$.

Proof. We need to verify all our axioms; but they come from the fact that the integral is linear, the product is commutative, and the integral of a positive C^0 function is positive. □

We'll now look into a specific situation — in analysis, there are lots of equations that come from different problems, and one common equation is the *Laplace equation*, where $L(y) = y''$. This operator is 'adapted' to $\langle -, -\rangle_1$, meaning that it's symmetric and therefore leads to an orthonormal basis of solutions (with appropriate boundary conditions).

Question 26.9. What is symmetric with respect to $\langle -, -\rangle_w$?

In other words, we'd like an operator which satisfies

$$\langle L(f), g\rangle_w = \langle f, L(g)\rangle_w.$$

The first operator in terms of *diffusion* leads to the *heat equation* $(\partial_t - \partial_{x^2}^2)y = 0$; this appears in physics, but also probability and statistics (because of Brownian statistics). But this is only the equation we get if our weight is 1 (i.e., our space is homogeneous); we might want to take into account factors like some places having higher population and so more propagation, and we handle this by inserting a weight.

A classi way to do this is by taking

$$L(y) = y'' + p(t)y'$$

(the constant term is always symmetric, so we might as well ignore it). In order to choose p for which this is symmetric, we integrate by parts — we have

$$\begin{aligned}\int_0^1 (f'' + pf')gw &= [f'gw]_0^1 + \int_0^1 (-f'g'w - f'gw' + pf'gw) \\ &= [fgw - fg'w]_0^1 + \int_0^1 (fg''w + fg'w' - f'gw + pf'gw).\end{aligned}$$

(we want to get rid of the derivatives on f, and put both derivatives on g, so we integrate by parts repeatedly until this happens). The first two terms are good. To simplify the remaining terms, we have

$$\int_0^1 -f'gw + pf'gw = [-fgw + pgw]_0^1 + \int_0^1 f(g'w + gw' + p'gw + pg'w + pgw').$$

Now we want this to equal $fg'A(t)$. So we'd like to get rid of the terms involving g. This is where we use p — we can use our degree of freedom to cancel out all these terms, and then we should be able to rearrange all the other terms to look the way we want — we want to take p such that

$$w' + p'w + pw' = 0.$$

This should then give us the right solution.

§27 April 21, 2023 — Fourier Series

Fourier series are a really important tool in analysis, especially in a lot of PDEs. The idea is to decompose any function into a basis of functions we understand, namely cosines and sines — the idea is very simple but extremely powerful, and it'll hopefully help us understand what we saw previously about weak solutions and integration by parts more concretely (when we'll write solutions as sums which make sense but represent functions which might not make sense).

We'll work with real functions on $L^2([-\pi,\pi])$ with the inner product $\langle -,-\rangle$ defined as

$$\langle f,g\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} f\overline{g}.$$

(Using complex Fourier series, it is more convenient to use $\frac{1}{2\pi}$; the point is that we want a certain basis to be orthonormal. But we will spend time on the real version because we've mostly been working with reals up to now.)

The fact that sines and cosines form an orthonormal *family* is quite easy, and we've already seen it. But in fact they form an orthonormal *basis* — we won't prove it, as it's pretty difficult.

Theorem 27.1

The families $(\frac{1}{\sqrt{2}}, \cos mt, \sin mt)_{m\in\mathbb{N}}$ and $(\frac{1}{\sqrt{2}}e^{im})$ are orthonormal bases under $\langle -,-\rangle$.

For convenience, we'll use the following notation.

Notation 27.2. We use $c_m(t)$ to denote $\cos(mt)$, $s_m(t)$ to denote $\sin(mt)$, and $e_m(t)$ to denote e^{imt}.

It's easy to check that these are orthonormal (we can simply integrate them against each other). The fact that it's a basis is equivalent to stating that for all $f\in L^2([-\pi,\pi])$, there exist unique constants $a_m(f)$ and $b_m(f)$ in $\mathbb{R}$ such that

$$f = \frac{a_0(f)}{2} + \sum_{m\in\mathbb{N}} a_m(f)c_m + b_m(f)s_m.$$

(So we can decompose any function into sines and cosines.) However, this equality is only in a L^2 sense — importantly, this does *not* mean that

$$f(t) = \frac{a_0}{2} + \sum_{m\in\mathbb{N}} a_m c_m(t) + b_m s_m(t)$$

for any given t. So we get L^2 equality, but *not* pointwise equality. What this more precisely means is that

$$\lim_{M\to+\infty}\left\| f - \left(\frac{a_0}{2} + \sum_{m=1}^{M} a_m c_m + b_m s_m\right)\right\|_{L^2} = 0,$$

where $\|f\|_{L^2} = \sqrt{\langle f,f\rangle}$. (In particular, the functions may differ at a few points, possibly infinitely many — they're basically equal, but there is a small set at which they may be different.)

Example 27.3

There exist discontinuous functions f in L^2 (e.g., f consisting of a few separate parts). Then the Fourier series will equal f wherever f is smooth enough, but at the points of discontinuity, the Fourier series will equal the average value of the two limits.

Definition 27.4. The real Fourier series of f, denoted $\hat{f}$, is defined as

$$\hat{f} = \frac{a_0(f)}{2} + \sum a_m(f)c_m + b_m(f)s_m.$$

The beautiful thing is that we can compute exactly what a_0, a_m, and b_m are — we have

$$a_0(f) = \frac{1}{\pi}\int_{-\pi}^{\pi} f \cdot 1$$
$$a_m(f) = \frac{1}{\pi}\int_{-\pi}^{\pi} f \cdot c_m$$
$$b_m(f) = \frac{1}{\pi}\int_{-\pi}^{\pi} f \cdot s_m.$$

We refer to these quantities as the *real Fourier coefficients*.

Definition 27.5. The *complex Fourier series* of f, denoted $\hat{f}$, is defined as

$$\hat{f} = \sum_{n\in\mathbb{Z}} c_n(f)e_n.$$

Here $c_n(f)$, called the *complex Fourier coefficient*, is given by

$$c_n(f) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f \cdot \overline{e_n} = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)e^{-int}\,dt.$$

(For a real function, the complex and real Fourier series will give the same sum.)

Remark 27.6. The reason we use $\frac{1}{\pi}$ for the real case and $\frac{1}{2\pi}$ for the complex case is because we want our relevant functions to have length 1 — otherwise the formula $f = \sum_i \langle f, e_i\rangle e_i$ breaks (we have to divide by $\langle e_i, e_i\rangle$).

Remark 27.7. The reason we have a sum over all of $\mathbb{Z}$ in the complex case and only $\mathbb{N}$ in the real case is because they're related through the formula $e^{imx} + e^{-imx} = 2\cos(mx)$ and $e^{imx} - e^{-imx} = 2i\sin(mx)$.

So then computing the Fourier series then just amounts to computing a few integrals (actually infinitely many, but they're all of the same form). Last recitation, we computed some Fourier coefficients of some simpler functions. We'll start by seeing one which is even simpler, and will show the behaviour mentioned (with discontinuities).

Example 27.8

Compute the Fourier coefficients of

$$f(t) = \begin{cases} 0 & \text{if } t < 0 \\ 1 & \text{if } t \geq 0. \end{cases}$$

If we do this, we'll see that

$$\hat{f}(t) = \begin{cases} 0 & \text{if } t \leq 0 \\ \frac{1}{2} & \text{if } t = 0 \\ 1 & \text{if } t > 0. \end{cases}$$

One thing that'll be important for us is how fast the a_m and b_m decay as $m \to \infty$ — when m is large, then cos and sin oscillate very fast (e.g., at high frequency). Typically, a function with high frequency is not very

regular. We'll be able to say later that for a C^k function, they should decay as $\frac{1}{m^k}$; so just from the decay of these coefficients, we'll be able to say how regular the function is. (If the function is C^∞, then the terms will have to decay faster than polynomial — e.g., if $a_m = O(e^{-m\alpha t})$.) We'll see that it's easier to work with H_k than with C_k, but we'll have a way to go from one to the other.

Remark 27.9. Note that both formulas make sense as formal series whatever a_m, b_m, and a_0 are, so we'll be able to talk about functions that aren't actually functions and do some analysis with them. For example, the Dirac operator (which is not even L^2, and is a terrible operator) still has Fourier coefficients — it will have $a_m = 1$ for all m, and $b_m = 0$ for all m. This is a very weak function — it's not a function at all, and not even a measure — but we can still represent it as such a Fourier series. This series is zero everywhere and goes to ∞ at 0.

§27.1 Geometry of Fourier Series

We'll now make use of some of the facts we saw earlier about orthonormal vectors, which hopefully have given us a good idea of where they'll come from.

Theorem 27.10 (Parseval's Identity)
If $f \in L^2(-\pi, \pi)$, then

$$\|f\|_{L^2}^2 = \frac{a_0^2}{2} + \sum_{m\in\mathbb{N}} a_m^2 + b_m^2 = 2\sum_{n\in\mathbb{Z}} \|c_n\|^2 .$$

(This essentially is the Pythagorean theorem.)

This tells us we can compute the L^2 norm of a function just from its coefficients — we have a beautiful identity between an integral and some sum of numbers. There are lots of sums which can be computed from this — sometimes you'll be able to compute the integral but not the sum, and this lets you get the value of the sum as well. (We'll see some exercises on this, where we compute some values of the zeta function using this and integration by parts.)

In particular, this tells us that every coefficient is bounded by the L^2-norm of f (which is not trivial). This has a simple but useful corollary:

Corollary 27.11 (Bessel's Inequality)
For every M,

$$\frac{a_0^2}{2} + \sum_{m=1}^{M} a_m^2 + b_m^2 \le \|f\|_{L^2}^2 = \frac{1}{\pi}\int_{-\pi}^{\pi} |f|^2 .$$

(The reason the 2 isn't squared is because 1 doesn't have norm 1 — we have $\frac{1}{\pi}\int_{-\pi}^{\pi} a_0^2 = 2a_0^2$.)

This clearly follows because the sum of fewer positive terms is less than the sum of more positive terms. But it's an important consequence — it also tells us that all of the sub-sums are bounded. So no matter what the function is, if it's in L^2, then none of these coefficients can be too large — and moreover, their *sum* cannot be too large either.

§27.2 Fourier Series and Differentiation

We're going to compute a lot of Fourier series, adn eventually we'll do more of this in class. But before that, we'll see the main reason we like Fourier series. It might be complicated to differentiate a function — and

it might not be continuous or differentiable. But the Fourier series is always differentiable, so we'll be able to differentiate *any* function this way, even ones which are not continuous.

Definition 27.12. We define $\hat{f}' = \sum_{m\in\mathbb{N}} a_m(f)c_m' + b_m(f)s_m' = \sum_{m\in\mathbb{N}} -ma_m(f)s_m(f) + mb_m(f)c_m$.

By integration by parts, if f is C^2, then $\hat{f}' = \hat{f}' = f'$. If f is H^1, then the first equality is still true. So this actually makes sense (we're not just randomly differentiating things, this does give us the correct value).

Proof. We use integration by parts. We want to consider

$$\frac{1}{\pi}\int_{-\pi}^{\pi} fc_m = \frac{1}{\pi}[\cdots]_{-\pi}^{\pi} - \frac{1}{n}f'\frac{1}{m}s_m.$$

□

Example 27.13

In our discontinuous function described earlier, $\hat{f}'$ will be reasonable, then go up to $+\infty$ and come back at the point of discontinuity, then be reasonable again for the next stretch, then go down to $-\infty$ and come back, and so on.

Similarly, we can use this to obtain the derivative — and integral — of Dirac.

In particular, this means

$$a_m(f') = mb_m(f) \text{ and } b_m(f') = -ma_m(f)$$

if f is differentWe can do the same with integrals —- we'll have $a_m(\int f) = -\frac{1}{m}b_m(f)$ and $b_m(f) - \frac{1}{m}a_m(f)$.

Note that taking integrals makes our functions more regular, and taking derivatives makes them less regular (since it pulls out m's, making the coefficients bigger).

§27.3 Poincare Inequalities

Theorem 27.14 (Poincaré–Wirtinger Inequality)

If $f \in H^1(-\pi,\pi)$, then

$$\frac{1}{\pi}\int_{-\pi}^{\pi}\left(f - \frac{a_0(f)}{2}\right)^2 \le \frac{1}{\pi}\int_{-\pi}^{\pi}|f'|^2.$$

This says that a function is not too far from its average if its derivative is small — note that by definition $\frac{a_0(f)}{2} = \frac{1}{2\pi}\int_{-\pi}^{\pi} f$. (The constant is the square root of the first nonzero eigenvalue of $y \mapsto y''$.)

Exercise 27.15. Show that $\hat{c_n} = c_n$, $\hat{s_n} = s_n$, and $\hat{1} = 1$ (by computing the Fourier coefficients — this follows from the definition of an orthonormal basis, but you can also obtain it by computing the coefficients directly).

§28 April 24, 2023

§28.1 Realizations of L^2

Last time we saw that any function in $L^2(-\pi,\pi)$ — functions which are square-integrable — can be decomposed as a Fourier series. We also saw Parseval's identity that

$$\|f\|_{L^2}^2 = \frac{a_0^2}{2} + \sum_{n\geq 1}(a_n^2 + b_n^2).$$

This gives a natural representation of what L^2 is — we can think of L^2 as the space

$$L^2(-\pi,\pi) = \left\{ \frac{a_0}{2} + \sum_{n\in\mathbb{N}} a_m c_m + b_m s_m \mid \frac{a_0^2}{2} + \sum a_m^2 + b_m^2 < +\infty \right\}.$$

This is one of the main reasons why we care about L^2 — a lot of identities about Fourier series can be expressed in terms of geometry in L^2 (this space has much better properties than e.g., C^0).

We can rewrite the definition of H^k in a similar way: we can define

$$H^k([-\pi,\pi]) = \left\{ \frac{a_0}{2} + \sum_{m\geq 1} a_m c_m + b_m s_m \mid \frac{0^k a_0^2}{2} + \sum_{m\geq 1} m^{2k}(a_m^2 + b_m^2) < \infty \right\}.$$

So the faster our coefficients decay at ∞, the better the regularity of our function will be (in particular $L^2 = H^0$).

Example 28.1

If $a_m^2 + b_m^2 < m^{-(2k+2)}$, then our function is in H^k. (This is true if we replace $+2$ with $+1+\varepsilon$ for any $\varepsilon > 0$.)

Example 28.2

If our sum is finite (i.e., only finitely many coefficients are nonzero), then our function is in H^k for all k. Less trivially, if $a_m^2 + b_m^2 < e^{-\alpha m}$ for some $\alpha > 0$, then our function is again in H^k for all k (which we will see soon implies that our function is in C^∞).

There's lots of equations we'll solve by finding the Fourier equations. Once we have the coefficients, *a priori* our function might not actually be a function, but by looking at how fast our coefficients decay we can see that our solution really is a smooth function. This is a general principle in analysis — first solving over things that don't necessarily make sense as functions, and then improving the regularity in some way.

Remark 28.3. This definition also makes sense for $k < 0$, and for any $k \in \mathbb{R}$ — when we do analysis, we will talk about being H^s for some s which may not be a positive integer. For example, it is useful to say that $\Delta\colon H^{s+2} \to H^s$ for any *real* s (not just positive integers).

This is why we've introduced L^2 in this class, and why it's so important for Fourier series.

Remark 28.4. Given a function in L^2, we can integrate or differentiate it many times at the level of Fourier coefficients in order to make it more or less regular.

Remark 28.5. It's possible to construct functions in H^s for s not a positive integer by their Fourier coefficients. In fact, the function t^s should probably be in H^s — you can prove this by calculating a bound on $\int_{-\pi}^{\pi} t^s \cos(mt)\,dt < m^{f(s)}$.

Now we have an idea of what it means to be L^2 and H^k at the level of Fourier series.

Remark 28.6. The main reason we care about negative k is that we can naturally pair functions which are H^k with functions that are H^{-k}, though we won't discuss this more.

§28.2 Differentiation

Last time, we defined the function

$$\hat{f}' = \sum_{m\geq 1} -ma_m(f)s_m + mb_m(f)c_m.$$

We can verify that if $f \in H^k$, then $\hat{f}' \in H^{k-1}$. Then we can also make sense of $\int f$, where instead of multiplying by m, we divide by m (and fix signs).

If f is smooth enough, then

$$\begin{cases} a_m(f') = mb_m(f) \\ b_m(f') = -ma_m(f). \end{cases}$$

We'll use this to solve ODEs:

Example 28.7

Suppose that we want to solve

$$y'' + ay' + by = f,$$

where $f \in L^2$ and we want y to be 2π-periodic.

To solve this, we'll decompose both y and f into Fourier series — assume that

$$y = \sum_{n\in\mathbb{Z}} c_n(y)e_n \text{ and } f = \sum_{n\in\mathbb{Z}} c_n(f)e_n.$$

Then we can rewrite our equation (E) in Fourier series — differentiating twice, we get that

$$\sum_{n\in\mathbb{Z}} (-n^2c_n(y) + ainc_n(y) + bc_n(y))e_n = \sum_{n\in\mathbb{Z}} c_n(f)e_n.$$

This is an equality between infinite sums, so *a priori* we can't say anything; but because e_m form an orthonormal basis, the decomposition into this basis is unique. This means for every $n \in \mathbb{Z}$ we must have $(-n^2 + ain + b)c_n(y) = c_n(f)$. Now solving the equation is easy — we can just divide by this expression to get that

$$c_n(y) = \frac{c_n(f)}{-n^2 + ain + b},$$

as long as $-n^2 + ain + b \neq 0$.

Remark 28.8. We may have heard of solving ODEs with Taylor coefficients — the principle is the same, of differentiating coefficient-by-coefficient.

In particular, this tells us that for large n, $|c_n(y)| \le 2\frac{|c_n(f)|}{n^2}$. So if $f \in H^k$, then $y \in H^{k+2}$. This is important because it shows that we gain regularity on the left-hand side. This is how Fourier coefficients are used — we assume we have a solution $y = \sum c_n(y)e_n$ (which may be very nonregular), solve the equation explicitly to get a unique solution (which is very easy), and then find that these coefficients decay fast and our function is actually quite regular.

§28.3 Sobolev Embeddings

Now we'll talk about a topic that bridges the gap between these Sobolev spaces and the C^k spaces we are more familiar with. We know that $C^k \subseteq H^k$ (we've already proven this). Meanwhile, we will see that $H^k \subseteq C^{k-1}$ if $k \ge 1$. (In fact, we will be able to get a $\frac{1}{2}$, but we won't do this.)

Theorem 28.9 (Sobolev Embedding)
We have $H_k \subseteq C^{k-1}$.

(This will have some nice applications, because we'll then be able to say that a Fourier series is often equal to the function itself, and we even have uniform convergence.)

Proof. Take $f \in H^1$; we will show that f must be continuous. Our goal is to show that for all $s, t \in [-\pi, \pi]$, we have $|f(s) - f(t)| \le \varepsilon(t-s)$, where ε is a function of $t-s$ such that $\varepsilon > 0$ and $\lim_{x\to 0}\varepsilon(x) = 0$.

In order to do this, we use the Cauchy–Schwarz inequality — assume without loss of generality that $s > t$. Then

$$|f(s) - f(t)| = \left|\int_t^s f'\right|.$$

The only information we have is that $\int |f'|^2$ is bounded (we don't know the same is true for $\int f'$). To bound this, we can use Cauchy–Schwarz — we have

$$\left|\int_t^s f' \cdot 1\right| \le \left|\int_t^s f'^2\right|^{1/2} \left|\int_t^s 1^2\right|^{1/2} \le \sqrt{s-t} \cdot \|f\|_{H^1}.$$

We can define this last expression as $\varepsilon(s-t)$, and we are done. (The crucial part is that this last expression only depends on the norm in H^1 — we did not use any more regularity assumptions on f.) □

Remark 28.10. The reason we can obtain a $\frac{1}{2}$ is because $\sqrt{s-t} = (s-t)^{1/2}$, so we're actually not just continuous but even Hölder continuous.

This shows any function in H^1 must be continuous; this should give a good idea of what it means to have a Fourier series that converges, and hopefully it makes Sobolev spaces feel more familiar (since they're not so far from the C^k spaces we're used to).

As a corollary:

Theorem 28.11
If $f \in H^1$, then:

- $f(t) = \hat{f}(t)$ for all $t \in [-\pi, \pi]$.
- Letting f_M be the partial Fourier decomposition $f_M = \frac{a_0(f)}{2} + \sum_{m=1}^{M} a_m(f)c_m + b_m(f)s_m$, we have $f_M \to \hat{f}$ *uniformly.*

Note that (2) is another way to see that $\hat{f}$ is continuous — our f_M are continuous and converge uniformly to a function, so the limit has to be continuous.

Corollary 28.12
If $f \in C^1$, then $f = \hat{f}$.

So the complicated sum we're producing here is actually *pointwise* equal to our function, and the convergence is not just pointwise but even uniform. We have to be careful when our function is just continuous — the Fourier series might be different from teh function — but whenever our function is regular enough, then the function is exactly realized by its Fourier series.

Corollary 28.13
If $a_m(f)^2 + b_m(f)^2 < m^{-4}$, then $f = \hat{f}$ pointwise (i.e., $f(t) = \hat{f}(t)$).

This is something we can measure by computing integrals to get a_m and b_m, and seeing that they decay fast enough.

(Note that saying $f = \hat{f}$ pointwise is a stronger statement than saying that they're equal in the L^2 sense, which is always true.)

Corollary 28.14
If f is continuous on $[-\pi, \pi]$and *piecewise* C^1 (i.e., f is C^1 on $[-\pi, \pi] \setminus \{p_1, \dots, p_i\}$ for some finitely many points p_i), then $f = \hat{f}$ pointwise (i.e., $f(t) = \hat{f}(t)$).

Example 28.15
A 'mountainous' function may fit this definition; so does $|x|$, so $|x| \in H^1$.

We'll see that Fourier coefficients often behave very nicely, and they can be used to prove a lot of regularity statements.

Remark 28.16. In higher dimensions this becomes more complicated because the Sobolev embedding is not as good, and you need higher regularity to show that your function equals its Fourier series.

§28.4 Poincaré–Wirtinger Inequality

Theorem 28.17
If $f \in H^1$, then

$$\frac{1}{\pi}\int_{-\pi}^{\pi}\left(f - \frac{a_0(f)}{2}\right)^2 \le \frac{1}{\pi}\int_{-\pi}^{\pi}(f')^2.$$

This gives an inequality between a function's distance from its average and the norm of its derivative — the L^2-norm of the dertivative controls how far we can get from the average value on the whole interval.

Proof. We'll compute both sides by looking at their Fourier coefficients and using Parseval's identity, and we'll see that there is an obvious inequality. First, Parseval's identity tells us that

$$\frac{1}{\pi}\int_{-\pi}^{\pi}\left(f - \frac{a_0}{2}\right)^2 = \sum_{m\ge 1} a_m^2 + b_m^2.$$

It also tells us (using the Fourier coefficients of the derivative) that

$$\frac{1}{\pi}\int_{-\pi}^{\pi}(f')^2 = \sum_{m\geq 1} m^2(a_m^2 + b_m^2).$$

Of course $m^2 \geq 1$ for all m, so we're done. □

This proof also lets us characterize the equality cases — equality means that

$$\sum (m^2-1)(a_m^2+b_m^2) = 0.$$

The coefficient $m^2 - 1$ is 0 if $m = 1$ and strictly positive otherwise. So for equality to hold, we must have $(m^2-1)(a_m^2+b_m^2) = 0$, so one of the two terms must be zero; this means for $m > 1$, we must have $a_m^2 + b_m^2 = 0$, and therefore $a_m = b_m = 0$. So this means $f = \frac{a_0}{2} + a_1\cos + b_1 \sin$.

As an application of this:

Exercise 28.18. The first nonzero eigenvalue of the operator $L(y) = y''$ is 1 on 2π-periodic functions.

This can be proven by assuming we have an eigenvalue and using integration by parts to get the right-hand side.

§29 April 26, 2023

Last time, we saw that we can write

$$\hat{f} = \frac{a_0}{2} + \sum a_m c_m + b_m s_m.$$

We can look at how fast a_m and b_m decay; and if $\sum m^{2k}(a_m^2+b_m^2) < +\infty$ then our function is in H^k (by Parseval's identity). We also saw that $H^{k+1} \subseteq C^k$. As a consequence, if $f \in H^1$ (i.e., $\sum m^2(a_m^2+b_m^2) < +\infty$ — in fact, an exponent of 1 would be enough since $H^{1/2} \subseteq C^0$), then $f = \hat{f}$ pointwise (rather than just in L^2). We saw an easy criterion to verify that something is in H^1 — if it is continuous, and C^1 everywhere except at finitely many points.

Example 29.1
$x \mapsto |x|$ is H^1.

Remark 29.2. Quite often in analysis, you want to go from one space to another. It's crucial that if we have an operator $C^{k+2} \to C^k$ we can't say much, but if it's $H^{k+2} \to H^k$ then we can use a lot of stuff regarding Banach spaces. This is one reason to use Sobolev spaces; and here they are much better adapted to Fourier series.

Remark 29.3. This is a good way to compute sums — if you know your function has a certain value and $f = \hat{f}$, then this tells you the value of the series corresponding to the a_n and b_n. But you can only do this when your coefficients decay fast enough (otherwise it'll be true at most points, but you won't be able to say where it is or isn't true).

§29.1 Solving PDEs Using Fourier Series

Quite often you classify equations by their properties, and there are different kinds of equations.

Last time we talked a little about how to solve some ODEs; this was similar to using e.g., Taylor series. But for PDEs it's essentially something you can only do with Fourier series, and it's kind of remarkable.

Definition 29.4. The Laplace equation is $\Delta u = 0$; in dimension 2 this means $(\partial_{x^2}^2 + \partial_{y^2}^2)u = 0$.

Definition 29.5. The *Poisson equation* is $\Delta u = f$ for some function $f : \mathbb{R}^2 \to \mathbb{R}$.

Definition 29.6. The *heat equation* is

$$\partial_t u = \Delta u,$$

so in dimension 1 this means $\partial_t u = \delta_{x^2}^2 u$.

Definition 29.7. The *Wave equation* is $\partial_{t^2}^2 u = \Delta u$, or in one dimension $\partial_{t^2}^2 u = \partial_{x^2}^2 u$.

Definition 29.8. *Scrhodinger's equation* is $\partial_t u = i\Delta u$, or $\partial_t u = i\partial_{x^2}^2 u$.

These equations fall into different categories. THe Laplace and Poisson equation are elliptic equations; the heat equation is parabolic; the wave equation is hyperbolic (the Einstein equations in physics are also hyperbolic, but they are elliptic in geometry). (These are just names used to classify equations whose solutions all have similar properties — elliptic equation comes from the fact that the Laplacian is an elliptic operator, while the parabolic type have derivative of time equal to some elliptic operator. Last year on the midterm they looked at non-integer powers of the Laplacian, and these stay elliptic. The point is that all these things might look close to each other, but depending on how many e.g., time derivatives we have, the solutions will look extremely different. For example, there's an elliptic regularity theorem that says all the solutions to an elliptic problem are very regular; for parabolic it won't necessarily be true at the initial time but will be true furthher; and hyperbolic equations will propagate the non-regularity. So the name sort of keeps track of what happens to the equations; there is a criterion that keeps track of whether you are elliptic. We will mostlly look at $\partial_{x^2}^2$, which is an elliptic operator.)

We will see some way to solve all t hese equations. (You could add any function f to the right-hand side and we'll still be able to solve them.) These are the most classical equations in PDEs and we'll see how to solve all of them; this is actually quite easy once you know Fourier series.

§29.2 Heat Equation

We'll limit ourselves to functions which are 2π-periodic — this means we're trying to solve the equation on some circle. So imagine we have some circle where $x = 0$ is identified with $x = 2\pi$, and we have time going up. Then the question is, we have such a circle. (Next year we'll learn about Fourier transform and be able to solve these things not just on a circle but on a line.)

Let's suppose we have an initial function at $t = 0$ on this circle, which we can call $u_0(x)$; this is our initial conditions. The question is, as time goes on, what happens to this. The answer is that it's going to sort of homogenize — the heat equation homogenizes your solutions, so then you'll get a little bump and then at time ∞ it'll converge to some constant.

Remarkably, if you start with no regularity (e.g., the dirac), then at any positive time you automatically end up with something C^∞; this is a very powerful way to use parabolic equations. (Sometimes you're given

some very rough data, and if you just run some flow for an arbitrarily short time you get something very smooth.) We'll hopefully understand where this comes from from the evolution fo a solution.

We want to solve

$$\begin{cases} \partial_t u = \partial_{x^2}^2 u \\ u(0,x) = u_0(x) \end{cases} \tag{H}$$

where $u = u(t,x)$ and $u(0,x)$ can be some very weak function.

For now, assume that $u_0(x) \in L^2$ so that we can tak about its Fourier decomposition.

A uniqueness statement we'll see later will imply there is nonly one solution. For now we'll assume it; so assume that $u(t,\cdot)$ can be realized by some Fourier series

$$\hat{u}(t,\cdot) = \frac{a_0(t)}{2} + \sum a_m(t) c_m.$$

Assume that $u_0(x) = u_0(-x)$ for convenience (to avoid having sines); if it were not the case we would also have sines.

Uniqueness will tell us this is the one solution so we are done; we won't see this now, since we will solve some equations first.

Now (H) tells us that

$$\begin{cases} \partial_t \left(\frac{a_0(t)}{2} + \sum_{m \geq 1} a_m(t) c_m \right) = \partial_{x^2}^2 \left(\frac{a_0(t)}{2} + \sum a_m(t) c_m \right) \\ \frac{a_0}{2} + \sum a_m(0) c_m = \frac{a_0(u_0)}{2} + \sum_{m \geq 1} a_m(u_0) c_m. \end{cases}$$

($\hat{u}(t,\cdot)$ denotes teh map $x \mapsto \hat{u}(t,x)$; we're not writing it this way because we don't yet know it is a pointiwse equality.)

By uniqueness of the Fourier decomposition, we know that for all m we have $a_m(0) = a_m(u_0)$ (the Fourier coefficicents of u at time 0 must exactly be those of our initial condition). Meanwhile, we can rewrite the first equation – what we'll do is differentiate whatever's in the sum,, and assume we can differentiate inside the sum. THis tlels us that

$$\frac{a_0'(t)}{2} + \sum a_m'(t) c_m = \sum_m (-m^2) a_m(t) c_m(t).$$

(In particular, note that a_0 doesn't appear.) This is because $c_m'' = -m^2 c_m$ with respect to x. (Note that the c_m are functions of x.)

Again by uniqueness of the Fourier decomposition, taking $n = 0$ gives $a_0'(t) = 0$, and for allwe have

$$a_m'(t) = -m^2 a_m(t).$$

Now we are done — we have a lot of first-order ODEs with constant coefficients, which we know hwo to solve, and we also know the initial conditions $a_m(0) = a_m(u_0)$. So we can exactly tell what the a_m should be.

(The decomposition is always unique when we have L^2 functions.)

The last step is verifying that our functions are somewhat regular.

Now we have reduced our big equation (H) as,

$$\begin{cases} a_0(t) = a_0(u_0) = \frac{1}{\pi} \int_{-\pi}^{\pi} u_0 \\ a_m'(t) = -m^2 a_m(t) \\ a_m(0) = \hat{a_m}(u_0) \end{cases}$$

where the latter two equations imply that $a_m(t) = \hat{a_m}(u_0) e^{-m^2 t}$. So the average value is constant, and the higher frequencies go to 0 very fast. This is why we gain lots of regularity, and also why we converge to the average value. (The heat equation essentially averages everything — if you heat one region of the room then it spreads everywhere and eventually you've reached constant temperature.)

Remark 29.9. Note that we have *parabolic smoothing* — for every $t > 0$, $u(t, \bullet)$ is C^∞. We started with a function that might be very bad (not even continuous), and we run the flow for any positive time (e.g., a tenth of a second) and everything becomes C^∞. This is because we know that to verify we are H^k (or C^{k-1}) it's enough to verify the Fourier coefficients decay fast enough. But here they decay exponentially fast, so for every $k \in \mathbb{N}$ we have $\sum m^{2k} a_m^2 = \sum m^{2k} \hat{a_m}(u_0) e^{-m^2 t}$. We have $\sum \hat{a_m}^2(u_0) < +\infty$ because we assumed our thing was L^2. Then the thing decays exponentially fast so is summable. This tells us that for every k we are H^k, and so since we are H^k for every k, we are also C^{k-1} for all k, and therefore C^∞.

So you can start with something very irregular and apply the heat equation to obtain something very smooth.

Here we assumed u_0 is L^2 but it's enough to take it bounded; you can even take u_0 to be the Dirac at 0 (i.e., δ_0). This is an element of L^1 (but not L^2); its Fourier series is $\hat{u_0} = \frac{1}{2} + \sum c_m$ (there is no decay in the coefficients so there is no regularity, but it is still possible to make sense of the solutions). You can still make sense of

$$\begin{cases} \partial_t u = \Delta u \\ u(0, x) = \delta_0. \end{cases}$$

The solution to this is

$$u(t, x) = \frac{1}{2} + \sum e^{-m^2 t} c_m.$$

So we start with something with zero regularity and obtain something that is C^∞; this is what we call the *Heat kernel*. (If you were to solve the equation on Euclidean space, you would obtain a Gaussian.)

These are the most important solutions to the heat equation because from them you can recover all the other ones. (This will be in the notes.)

There are lots of PDEs we can try in this way.

§29.3

Consider the equation

$$\begin{cases} \partial_t u = \Delta u + f(t, x) \\ u(0, \cdot) = u_0. \end{cases}$$

The way we solve this is by decomposing the problem into two. We have two difficulties to deal with; one is the initial condition adn the other is a source term. It turns out that we can search for solutiosn to two rpoblems: we can take

$$u = u_H + u_S$$

where u_H is the solution to a homogeneous problem with an initial condition (H for homogeneous and S for source), just using the linearity of the equation which means we can superpose solutions, where one has the initial condition adn the other has no initial condition but the source term. This means

$$\begin{cases} (\partial_t - \Delta) u_H = 0 \\ u_H(0, \bullet) = u_0 \end{cases}$$

and

$$\begin{cases} (\partial_t - \Delta) u_S = f(t, x) \\ u_H(0, \bullet) = 0. \end{cases}$$

This superposition is what we call the *Duhamel principle*. In general, you want to write

$$u(t, x) = \frac{u_0(t)}{2} + \sum_{m \geq 1} u_m(t) c_m(x),$$

where

$$u_m(t) = u_m(0)e^{-m^2t} + e^{-m^2t}\int_0^t e^{m^2s} f_m(s)\,ds,$$

The first term comes from the homogeneous part, and the second term comes from the other part (the source part), where $f = \frac{f_0(t)}{2} + \sum_{m\geq 1} f_m(t)c_m$.

Really you want to solve the two equations separately, find each of these solutions, and then sum them; you recognize the solutions to the first order ODE where the initial condition adn right-hand side have impact. (But it's really a fact in general for PDEs that you can decompose into something depending on the source and something depending on the initial data.)

This is the same proof as what we did, but you end up with a slightly more complicated first-order ODE where the f's will appear.

Now you can see what happens if we take ∂_{t^2} instead. It'll turn out that we don't have exponential decay of the coefficients, so there is a big difference. But this is the general way you solve PDEs, adn you can recover all of the good properties of solutions to elliptic, parabolic, and so on equations.

§30 April 28, 2023

Today we'll see two more examples of PDEs which we can solve using Fourier analysis; we'll see some extensions of this on the problem set due next week.

§30.1 The Wave Equation

The principle is always the same — consider an equation

$$\begin{cases} \partial_{t^2}^2 u = \partial_{x^2}^2 u \\ u(0,x) = u_0(x). \end{cases}$$

We want to assume that both our initial condition and our entire solution have Fourier decompositions (where the x-dependence only comes from the cosines and sines, and the t-dependence is only in the Fourier coefficients).

Remark 30.1. Being L^2 — or even L^1 — is enough to have a Fourier decomposition, so almost everything does; if our functions don't, then we usually can't solve the equation. For example, $u_0(x) = \frac{1}{|x|}$ shouldn't have a Fourier decomposition, because it doesn't even have an average value — $\frac{1}{\pi}\int_{-\pi}^{\pi} u_0 = +\infty$, and every integral against cos should be infinite as well. If we don't have Fourier decompositions, it might even be that there is no solution, and a solution would definitely have to be weird, since the average value is constant. But we'll focus on cases where we do have Fourier decompositions; in that case we'll be able to solve the equation by the previous method.

Suppose that

$$u_0 = \frac{u_0}{2} + \sum u_m c_m$$

and that

$$u(t,x) = \frac{a_0(t)}{2} + \sum_{m\geq 1} a_m(t)c_m(x).$$

Our goal is to rewrite our PDE in terms of ODEs on the $a_m(t)$. We can rewrite (H) as

$$\begin{cases} a_0''(t) = 0 \\ a_0(0) = u_0 \end{cases}$$

and for all $m \geq 1$,

$$\begin{cases} a_m''(t) = -m^2 a_m(t) \\ a_m(0) = u_m. \end{cases}$$

This comes by taking the time-derivative and spatial derivative of both sides (applying the equation to each term of the Fourier decomposition — differentiating $\cos mx$ twice provides the factor of $-m^2$).

Remark 30.2. For the heat equation, there isn't a closed form solution at all; but we can still write it in terms of Fourier series.

We can try to solve this ODE, but there's a problem — the solution isn't unique, because we should have *two* initial conditions (since we have second-order ODEs). So we need to be a bit careful — what we wrote looked like a good initial value problem, but it isn't one, and we can see this at the level of ODEs. So we need a second initial condition, which we will obtain by taking a derivative (i.e., fixing the initial position and the initial speed) — add the condition that

$$\partial_t u(0,x) = \widetilde{u_0}(x) = \frac{\widetilde{u_0}}{2} + \sum_{m \geq 1} \widetilde{u_m} c_m.$$

Now we can rewrite our systems. The first equation, for the average value, says that

$$\begin{cases} a_0''(t) = 0 \\ a_0(0) = u_0, a_0'(0) = \widetilde{u_0}, \end{cases}$$

and the second equation, for a_m, says that

$$\begin{cases} a_m''(t) = -m^2 a_m(t) \\ a_m(0) = u_m, a_m'(0) = \widetilde{u_m}. \end{cases}$$

Luckily, we know how to solve both these ODEs. The solution to the first is

$$a_0(t) = u_0 + t\widetilde{u_0}.$$

The solutions to the second are

$$a_m(t) = \alpha c_m(t) + \beta s_m(t)$$

(note that here we have t, and not x), and we can determine α and β from the initial conditions; we can check that $\alpha = u_m$ and $\beta = \frac{\widetilde{u_m}}{m}$, so

$$a_m(t) = u_m c_m(t) + \frac{\widetilde{u_m}}{m} s_m(t).$$

So overall, our solution can be written as

$$u(t,x) = \frac{u_0 + t\widetilde{u_0}}{2} + \sum_{m \geq 1} \left(u_m c_m(t) + \frac{\widetilde{u_m}}{m} s_m(t) \right) c_m(x).$$

Interestingly, we have a product of a cosine in time and a cosine in space. There is an explanation of this, which is easier to see if we write this in terms of exponentials — the second sum is a sum of terms of the form $e^{in(t+x)}$ and $e^{in(t-x)}$ for $n \in \mathbb{Z}$. So we really have two possibilities — this is saying that we create a wave in both directions (one going forwards and one going backwards). This is what happens when you drop something in water (the waves don't just go in one direction, but fall in a circle). Pictorially, this means our initial condition will split and travel in two directions (a wave is propagated in both directions).

Remark 30.3. Note that the coefficients $a_m(t)$ do not decay in time — we don't gain (or lose) regularity, and regularity is instead preserved. (This is unlike the heat equation.)

§30.2 Solving PDEs with Boundary Conditions

We'll now see how to solve PDEs with boundary conditions, which will be a bit different. Again the principle is to decompose our problem into a homogeneous problem and one with zero boundary condition; we will only look at the homogeneous problem.

We consider Laplace's equation on a disk — consider the equation

$$\begin{cases} \Delta u = 0 & \text{on } B(0,1) \\ u|_S = u_0 \text{ on } S(0,1) = \{x^2 + y^2 = 1\}. \end{cases}$$

We'd like to say how big a solution can be inside, depending on the values on the boundary. (The maximum principle already says something.)

We'll ues polar coordinates — write $(x, y) = (r\cos\theta, r\sin\theta)$. We'll naturally get functions which are 2π-periodic in θ, which means we can use Fourier series (more naturally than in the previous case).

The Laplacian in polar coordinates becomes

$$\Delta u = \left(\partial_{r^2}^2 + \frac{1}{r}\partial_r + \frac{1}{r^2}\partial_{\theta^2}^2\right) u.$$

We know u is 2π-periodic in θ; let's assume that u is smooth enough that we can discuss pointwise values, and write that

$$u(r,\theta) = \sum_{n\in\mathbb{Z}} c_m(r) e^{im\theta}.$$

We will assume that

$$u_0(\theta) = u(1,\theta) = \sum_{n\in\mathbb{Z}} u_m e^{im\theta}.$$

We'll first rewrite our equation — we need to have

$$\begin{cases} \left(\partial_{r^2}^2 + \frac{1}{r}\partial_r + \frac{1}{r^2}\partial_{\theta^2}^2\right) u = 0 \\ u(1,\theta) = u_0(\theta). \end{cases}$$

In Fourier series, this becomes

$$\begin{cases} c_m''(r) + \frac{1}{r}c_m'(r) + \frac{1}{r^2}(-m^2) = 0 \\ c_m(1) = u_m. \end{cases}$$

This is an ODE, but we again have the issue that we only have one initial condition instead of two. But we'll see that there is actually a natural assumption to make.

The solutions to these ODEs are not so complicated (though they're actually more complicated in dimension 2 than in higher dimension). Let (E) be the above problem; then if $n = 0$, we get the solution $c_n(r) = \alpha_0 + \beta_0 \log r$, and for $m \neq 0$ we get

$$c_m(r) = \alpha_m r^{|m|} + \beta_m r^{-|m|}.$$

Now we want to fix α_m and β_m using the initial conditions. Note that both the terms $\beta_0 \log r$ and $\beta_m r^{-|m|}$ blow up at 0. So just asking for our solutions to be bounded automatically kills the last two terms; and then we have enough information to solve the ODEs. (We'll see that these parts do also make sense, but if we were trying to solve the equation *outside* the circle instead of inside the circle; if we want the solutions bounded outside the circle instead, then we'd need to kill the r^m terms instead.)

This gives us

$$u(r,\theta) = u_0 + \sum_{|n|\geq 1} u_n r^{|n|} e^{in\theta}.$$

If we were instead trying to solve *outside*, then we'd get the same thing, but with $-|n|$ instead — so we'd get

$$u(r,\theta) = u_0 + \sum_{|n|\geq 1} u_n r^{-|n|} e^{in\theta}.$$

Remark 30.4. It's also a good exercise to try to solve the same problem with f instead of 0; the ODEs are a bit more complicated, and you'll need to use variation of constants, but it's doable and we have all the tools needed to do it.

Remark 30.5. Sometimes you keep the log and get rid of the other terms which blow up at 0 — then you're formally solving

$$\begin{cases} \Delta u = \delta_0 \\ u(1,\theta) = u_0(\theta) \end{cases}$$

(where δ_0 denotes the Dirac equation at 0), and you solve this by keeping the $\log r$ term. (You want this because then convolution lets you find all solutions.)

Remark 30.6. This is the last class on Fourier series, but we will have a problem set to practice them (involving reducing things to ODEs and solving the ODEs we understand). Fourier series are an essential tool everywhere — number theory and algebraic geometry, analysis. It lets you find some identities that you can't find in other ways, lets you control even discrete things (because you're relating a sequence of numbers to any function — here we've been reducing numbers to a sequence of numbers we understand one by one, but in number theory people will often do the opposite, where they have a sequence of numbers (a_n) and want to say something about their sum or something similar, and they define the function $f(x) = \sum_{n\in\mathbb{Z}} a_n e^{inx}$. Quite often, by looking at information about $f(x)$, you can get information about the sequence itself. We won't see this in this class, but e.g., Larry Guth uses Fourier series to solve problems that are completely outside analysis and more close to geometry and number theory. It also applies to combinatorics. This is another reason to care about Fourier series. You can also obtain by hand all elliptic and parabolic regularity theorems in this way — in the future, you'll look at equations on complicated domains with complicated operators, and you will be trying to solve something in this region, and it's generally really hard to get controls on the series. But it turns out that if you take a Fourier series (this time in two directions — functions which are 2π-periodic in both the x-direction and the y-direction, by drawing a big grid), all these theorems which are extremely complicated without Fourier series become almost trivial, and you will be solving equations just for 2π-periodic functions, which is even easier than what we did here, and you will be able to get regularity on your solutions. This is a powerful tool that simplifies lots of things, and you don't always see it in analysis books.

Remark 30.7. Next we'll see systems of ODEs (which are actualy not more complicated than in dimension 1).

§31 May 1, 2023 — Systems of ODEs

A system of ODEs is essentially an ODE but in higher dimensions — so instead of solving one equation, we'll solve a system of equations.

In dimension 1, we had ODEs of the form

$$y' = f(t,y).$$

In 3 dimensions, we might have

$$\begin{aligned} x' &= f_1(t, x, y, z) \\ y' &= f_2(t, x, y, z) \\ z' &= f_3(t, x, y, z). \end{aligned}$$

Each of x, y, z is a function of time — for example, the first equation really states that

$$x'(t) = f_1(t, x(t), y(t), z(t)).$$

Note that everything is coupled — so for example, x appears in the equations for y' and z' — which is what makes systems of ODEs more difficult.

We can interpret this geometrically — for example, let $X = (x, y, z)$, and write $(f_1, f_2, f_3) = F(t, X)$. This lets us rewrite the equation as

$$X'(t) = F(t, X(t)).$$

(So we've essentially just replaced our function x in dimension 1 with a time-dependent vector X.)

A general 1st order ODE in dimension d is of the following form: consider a function $Y \colon \mathbb{R} \to \mathbb{R}^d$ sending $t \mapsto (y_1(t), \dots, y_d(t))$, and a function $F \colon I \times \mathbb{R}^d \to \mathbb{R}^d$ sending

$$(t, y_1, \dots, y_d) \mapsto F(t, y_1, \dots, y_d) = (F_1(t, y_1, \dots, y_d), \dots, F_d(t, y_1, \dots, y_d))).$$

Definition 31.1. A general first-order ODE in dimension d is an equation

$$Y'(t) = F(t, Y(t)).$$

As in the 3-variable case, we can write this as a system, which will have d lines.

Remark 31.2. We're only looking at first-order ODEs here. But we'll see later that an ODE of any order can be reduced to a first-order ODE in higher dimensions. This is useful because then we can use the same existence–uniqueness theorem for *every* situation.

Theorem 31.3 (Existence–Uniqueness)

Assume that $t \mapsto F(t, Y)$ is C^0 for all fixed Y and $Y \mapsto F(t, Y)$ is L-Lipschitz for all t. Then there exists $c > 0$ (depending on the initial conditions) such that

$$\begin{cases} Y' = F(t, Y) \\ Y(t_0) = Y_0 \end{cases}$$

has a unique solution on $[t_0 - c, t_0 + c] \times B_d(Y_0, c)$.

(Everything we said about global existence–uniqueness and blowup in finite time also works — the blowup condition means that the *norm* goes to $\pm\infty$.)

Notation 31.4. The notation $B_d(Y_0, c)$ denotes the ball centered at Y_0 in $\mathbb{R}^d$ of radius c.

This may sound like a much more general statement than what we did in dimension 1, but the proof is exactly the same. (You can try to do it, replacing all $|\bullet|$ by norms. You can still integrate, and you'll have the same estimates (the estimates come from integrating absolute values). It is also true that $\mathbb{R}^d$ is complete (this is a consequence of the 1-dimensional statement), so the same proof with Cauchy sequences then works.)

Remark 31.5. This works in infinite dimensions as well. You could even take any norm on $\mathbb{R}^d$ and all would work (in infinite dimensions you'd need to be more careful to choose a good norm, but the same proof would still work).

§31.1 Higher Order ODEs

We'll now see how any ODE of order d can be seen as a first-order ODE. (It's a simple trick, but it lets us use the theorem without having to reprove it for every order.)

We'll see how to go from a dth order ODE in dimension 1 to a first order ODE in dimension d. More generally, you can go from a dth order ODE in dimension n to a first order ODE in dimension dn; we won't do this, to avoid having too many coordinates, but the extension is clear.

We saw earlier that a dth order ODE is one of the form

$$y^{(d)} = f(t, y, y^{(1)}, \dots, y^{(d-1)})$$

(if we can't write our ODE in this form, then we have other problems). We want to rewrite this equation so that we only take one derivative at a time. The simple realization here is that

$$(y^{(k)})' = y^{(k+1)}.$$

Using this, we define $y_k = y^{(k)}$. Then we can write the system

$$\begin{aligned} y_0' &= y_1 \\ y_1' &= y_2 \\ &\vdots \\ y_{d-2}' &= y_{d-1} \\ y_{d-1}' &= f(t, y_0, \dots, y_{d-1}). \end{aligned}$$

This gives us a system of ODEs equivalent to our original one, since all we're saying is that differentiating a function gives its derivative.

Remark 31.6. The converse is false — not every first-order ODE in dimension d corresponds to an ODE in dimenion 1.

This trick is rather simple, but it lets us replace an ODE of any order with a first-order ODE. This also explains the initial conditions we've been taking since the beginning — why we take our initial conditions to consider all the derivatives at the time t_0. This is because if we write $Y = (y_0, \dots, y_{d-1})$ and

$$F(t, Y) = (y_1, y_2, \dots, y_{d-1}, f(t, y_0, \dots, y_{d-1}))$$

(we use capital letters to denote vectors, and lowercase letters for things which are not vectors), then our equation becomes

$$Y' = F(t, Y).$$

Then we can apply existence–uniqueness to this function — we just need F to be continuous in t and Lipschitz in y, which is equivalent to the same conditions on f.

So given any e.g., second-order ODE, we have existence and uniqueness as long as we fix an initial condition $Y(t_0) = Y_0 = (Y_0^0, \dots, Y_0^{d-1})$, which in coordinate scan be written as $y(t_0) = Y_0^0$, …, $y^{(d-1)} = Y_0^{d-1}$. Thsi explains why these were the initial conditions we were always taking — it's the natural extension of our initial conditions for first-order ODEs.

(You can easily magine what happens for dimension n; it's the same thing.)

§31.2 Linear ODEs

In dynamics, the main issue is to understand what happens at critical points (or fixed points), and in order to do this, we look at the first derivative.

Example 31.7

Consider the ODE

$$y^{(d)}(t) = \sum_{i=0}^{d-1} a_i(t) y^{(i)}(t) + f(t).$$

(Note that $f(t)$ is not inside the summation.)

Then this corresponds to a matrix

$$\begin{bmatrix} y_0' \\ y_1' \\ \vdots \\ y_{d-2}' \\ y_n' \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & & & & & & \\ \vdots & \vdots & \vdots & \vdots & & & \\ a_0(t) & a_1(t) & \cdots & a_{d-1}(t) & & & \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ f(t) \end{bmatrix}.$$

So we can rewrite this equation as $Y'(t) = A(t)Y(t) + F(t)$ (where $Y(t) \in \mathbb{R}^d$ and $A(t) \in \mathbb{R}^{d \times d}$). So every dth order ODE in dimension 1 can be written in this form.

Definition 31.8. A *linear* first order ODE in dimension d is of the form

$$Y'(t) = A(t)Y(t) + F(t),$$

for any fixed matrix $A(t) = ((a_{ij}(t))) \in \mathbb{R}^{d \times d}$.

Very often, we'll be able to reduce these equations to a bunch of first-order ODEs; this will hopefully give us some explanation of everything we've done for second-order ODEs.

Example 31.9

If $A(t)$ is diagonal (i.e., $a_{ij}(t) = 0$ for $i \neq j$), then the system simplifies to

$$\begin{aligned} y_1' &= a_{11}(t) y_{11} + f_1(t) \\ &\vdots \\ y_d'(t) &= a_{dd}(t) y_d + f_d(t). \end{aligned}$$

This is nice because we know how to solve first-order 1-dimensional ODEs — this gives

$$y_k(t) = y_k(t_0) e^{\int_{t_0}^t a_{kk}(s)\,ds} + e^{\int_{t_0}^t a A_{kk}(s)\,ds} \int_{t_0}^t e^{-\int_{t_0}^s a_{kk}} f_k(s)\,ds.$$

When A doesn't depend on time, we can essentially always reduce to the diagonal case, and therefore solve our ODE. When A does depend on time, its regularity will start mattering.

Definition 31.10 (Fundamental Matrix)**.** Consider $Y' = AY$ where A has constant coefficients. A *fundamental solution* (or *fundamental matrix*) a is $d \times d$ matrix $U(t) \in \mathbb{R}^{d\times d}$such that

$$U = \begin{bmatrix} U_1(t) & \cdots & U_d(t) \end{bmatrix}$$

where each U_i is a solution to the ODE; for every $k \in \{1, \dots, d\}$, we have $U_k'(t) =$ for all t in our interval; and for every t, the $U_k(t)$ are linearly independent, or equivalently $\det U(t) \neq 0$.

Intuitively, a fundamental solution is a solution from which we can recover any other solution. The result is that if we apply this matrix to any constant vector, then we get all the solutions.

Remark 31.11. $\det U$ is the Wronskian of the column vectors $(U_1(t), \dots, U_d(t))$.

This explains the 2-dimensional situation — we had

$$W(y_1, y_2) = \det \begin{bmatrix} y_1 & y_2 \\ y_1' & y_2' \end{bmatrix}.$$

§32 May 3, 2023

Last time we began discussing ODEs in dimension d; we saw that we can extend our notion of a first-order linear ODE to a linear ODE

$$Y'(t) = AY(t)$$

for a matrix A. These are the simplest ODEs, but they're in some sense the most important — e.g., if we know the speed at every point. At critical points, to understand what happens at that point, you linearize your equation. So these are the most important equations, and luckily they're also the simplest.

In a lot of instances (especially today), you could let A vary with time. Many of the definitions still work (e.g., the Wronskian makes sense, we just can't solve the equation explicitly). But we're always going to assume A is constant; most of what we'll say will be true in the A variable case as well.

One way to obtain all solutions at once is by writing down the fundamental matrix. You should think of it as having d solutions and making sure they're all linearly independent. Some people like putting these solutions into a matrix (because then you can apply this matrix to a vector).

Definition 32.1. We say U is a *fundamental matrix* of our ODE (E) if:

(1) $U \in \mathbb{R}^{d\times d}$ is invertible (i.e., $\det U(t) \neq 0$ for all t).

(2) $U' = AU \in \mathbb{R}^{d\times d}$.

(It'll turn out that $\det U(t) \neq 0$ for all t is equivalent to $\det U(t) \neq 0$ for some t.)

Definition 32.2. The *Wronskian* of U (or equivalently, the Wronskian of the d columns of U) is defined as

$$W(U_1(t), \dots, U_d(t)) = \det U(t).$$

(Here $U_i(t)$ denote the columns of U.)

This corresponds to our old definition of the Wronskian:

Example 32.3

For a second-order ODE $y'' + py' + qy = 0$ (ODE2), we make it a first-order ODE in dimension 2 by creating the system

$$\begin{cases} (y)' = y' \\ (y')' = -py' - qy \end{cases}.$$

(ODE1) This tells us that the vector $Y = (y, y')^\intercal$ evolves according to $Y' = AY$ where

$$A = \begin{bmatrix} 0 & 1 \\ -q & -p \end{bmatrix}.$$

This is now a first-order linear ODE in dimension 2. If y_1 and y_2 are solutions of (ODE2), then we can define

$$U(t) = \begin{bmatrix} y_1(t) & y_2(t) \\ y_1'(t) & y_2'(t) \end{bmatrix}.$$

(y_1 and y_2 are a basis of solutions, both in the one-dimensional second-order case and in this case.) Then the two columns Y_1 and Y_2 are solutions to (ODE1).

U satisfies the equation $U' = AU$. If y_1 and y_2 are linearly independent, then

$$detU(t) = y_1 y_2' - y_1' y_2.$$

We've seen that this is nonzero (at one point, or equivalently at every point) if and only if y_1 and y_2 are linearly independent.

We'll see very soon that the derivative of the determinant will also satisfy a first-order ODE, which will lead to similar results.

Another thing that was useful in the second-order case was the characteristic polynomial; that will again be useful here.

Definition 32.4. The *characteristic polynomial* of (E) (the equation $Y' = AY$ in $\mathbb{R}^d$) is defined as the characteristic polynomial of A, which is

$$P_A(s) = \det(A - sI_d).$$

We use I_d to denote the identity matrix in dimension d (with 1's on the diagonal and 0's everywhere else — it's the unique matrix such that $AI = IA = A$ for all matrices A).

We'll see that this corresponds to what we saw for second-order ODEs.

The reason this is useful is the following proposition:

Proposition 32.5

$s \in \mathbb{C}$ is a root of P_A if and only if s is an eigenvalue of A, i.e., there exists a nonzero vector $X \neq 0$ in $\mathbb{R}^d$ such that $AX = sX$.

This comes from linear algebra in finite dimensions — if $A - sI_d$ has determinant zero, then it is not invertible, so it must have nonzero kernel; then taking any nonzero element of the kernel gives us such a vector.

Example 32.6

For our matrix

$$A = \begin{bmatrix} 0 & 1 \\ -q & -p \end{bmatrix},$$

we have

$$A - sI_2 = \begin{bmatrix} -s & 1 \\ -q & -p-s \end{bmatrix},$$

and so

$$P_A(s) = \det(A - sI_2) = s(p+s) + q = s^2 + ps + q.$$

This is what we defined as being the characteristic polynomial of the second-order ODE.

Similarly to before, finding the roots of the characteristic polynomial gives us eigenvalues, and the eigenvalues will give us simple solutions.

As another example (in dimension 2 specifically — this extends to greater dimensions, but becomes more complicated):

Exercise 32.7. We have $P_A(s) = s^2 - \operatorname{tr}(A)s + \det(A)$.

In any dimension, $-\operatorname{tr} A$ will be the first coefficient and $\det(A)$ will be the last, but there will be more terms in the middle. The determinant is obvious, since when $s = 0$ we simply get $\det(A)$, but the trace will be the key term in what we'll see when we prove the Wronskian doesn't vanish.

§32.1 First-order ODEs in dimension 2 to Second-Order ODEs

Take A to be the matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

and consider the ODE $Y' = AY$. Say that $Y = (x, y)^\intercal$ (where x and y are functions of time). Then the equation becomes a system of equations

$$\begin{cases} x' = ax + by \\ y' = cx + dy \end{cases}.$$

This is a coupled equation, so we can't just solve one equation and then solve the other. But there is a way to separate them — let's say we only want to find an equation in x. The way we do this is by differentiating once again, and replacing all y''s we see by this equation and trying to organize the terms to simplify them.

Letting our equations be (1) and (2), differentiating (1) gives

$$x'' = ax' + by'.$$

We want to get rid of all the y's that we see — we're happy with all x's and differentiated x's, but not y's. Here we see y', so we replace it by our equation to get

$$x'' = ax' + b(cx + dy).$$

Now we can get rid of this y using the first equation — we can write this as

$$x'' = ax' + bcx + d(x' - ax),$$

using the equation $x' = ax + by$ to replace by.

This gives a second-order ODE in only x, which is

$$x'' = (a+d)x' - (ad-bc)x.$$

(So if we have a solution to this system, then x should be a system to this other equation — then we can look at hte solutions to this equation and use them to solve.) Note that here we have $a+d = \operatorname{tr} A$ and $ad-bc = \det A$. So the characteristic polynomial kills the matrix somehow; that's a general fact.

If we try to write down an equation for only y, we'll end up with the exact same equation — we similarly have $y'' = (a+d)y' - (ad-bc)y$.

Remark 32.8. This may or may not be a general fact (that we can go from a dimension-n ODE to an nth order ODE) — it might be true in general dimensions, because of the fact that the characteristic polynomial kills the matrix.

§32.2 The Wronskian

Lemma 32.9

$\det(I_d + \varepsilon A) = 1 + \varepsilon \operatorname{tr}(A) + \varepsilon^2(\cdots)$ for any matrix A.

In other words, the linearization of the determinant at the identity is equal to the trace.

We aren't going to prove it (since it requires knowing a formula for the determinant), but intuitively, the determinant measures a volume, and the trace captures the variation of the volume.

Exercise 32.10. Prove it in dimension 2.

This comes from the same reason as why we see the trace in P_A as the first term.

In particular, using the fact that $\det(AB) = \det(A)\det(B)$, we can also linearize the determinant at any other matrix, just by multiplying the above formula by any matrix — the application that we want is

$$\det(U + \varepsilon AU) = \det(U)\det(I + \varepsilon A) = \det(U) + \varepsilon \det(U)\operatorname{tr}(A) + \varepsilon^2(\cdots).$$

This is useful because this lets us compute the derivative of the determinant of U, from the derivative of U:

Corollary 32.11

For a matrix U depending on t which is a fundamental matrix (i.e., $U' = AU$) and nonzero determinant), we have $(\det U)' = \operatorname{tr}(A)\det(U)$.

This is a first-order ODE which is linear. So either $\det(U)$ vanishes at every point, or at no points —- in toher words, there are two exclusive possibilities:

- For all t, $\det U(t) = 0$.
- For all t, $\det U(t) \neq 0$.

This is because we have

$$W'(t) = \operatorname{tr} A \cdot W(t).$$

Since thsi satisfies a first-order ODE, by uniqueness (since 0 is a solution) if we touch 0 at one point then we are identically 0.

We can also write down exactly the evolution of the Wronskian — it's given by the solution to this simple ODE.

So the trace measures how the solutions are growing or becoming more or less linearly independent. Note that $e^{\operatorname{tr} A}$ comes up; we'll soon start talking about exponentials of matrices, which will be the key to solving these systems.

Remark 32.12. This can be used to show a system is *not* stable (if the trace has the wrong sign), but not the other way — a system is stable if *all* the eigenvalues have the correct sign.

§32.3 Forecast

In dimension 1, the solutions to $y' = ay$ are of the form $y(t) = y(t_0)e^{a(t-t_0)}$. Very soon, we will prove that in dimension d, we formally have the same behavior — $Y' = AY$ tells us that

$$Y(t) = e^{(t-t_0)A}Y(t_0).$$

Here $e^{(t-t_0)A}$ will be a fundamental matrix $U(t)$, and $Y(t_0)$ a constant vector. The key difficulty will be what $e^{(t-t_0)A}$ means; we'll define it in two ways (one as a solution to this, and a bit more concrete definition).

§33 May 5, 2023

Last time, we started discussing finding a *fundamental matrix* U – U will be our tool to obtain solutions to any ODE in dimension d. Last time, we saw that there is always a unique solution; uniqueness and existence generally come from our existence-uniqueness theorem, but it's even simpler here.

We are trying to solve $U' = AU$ with $U(t_0) = U_0$, and with $\det(U_0) \neq 0$ (or equivalently $\det(U(t)) \neq 0$ for all t).

Definition 33.1. The *exponential* of $A \in \mathbb{R}^{d\times d}$ is the unique solution of $U' = AU$, $U(0) = I_d$. We denote it as $t \mapsto e^{tA} = \exp(tA)$.

This is not the most explicit way to define the exponential — we will have other ways to define it. But this is basically teh only fundamental solution we care about — the one starting at the identity. (Prof. Ozuch doesn't like the definition of a fundamental matrix because you'll always have to calculate the exponential, and it's in some sense the best fundamental matrix.)

Exercise 33.2. Any fundamental matrix is of the form $e^{tA}U_0$ with $\det(U_0) \neq 0$.

(This can be seen by showing that this is a solution to the equation above.)

We will later see more concrete definitions of the matrix exponential as well.

Remark 33.3. Recall that in dimension 1,

$$\begin{cases} u' = au \\ u(0) = 1 \end{cases} \implies u(t) = e^{ta}.$$

This is where the name comes from — we extend the exponential in dimension 1 to any dimension, and one possible definition is by this ODE. The other definitions also extend to the matrix setting.

§33.1 Computing the Exponential of a Matrix

The generic situation is when $A \in \mathbb{R}^{d\times d}$ has d distinct eigenvalues. (This is generic in the sense that if you take a matrix at random, there is a 100% chance it is of this type.) Call these eigenvalues $\lambda_1, \ldots, \lambda_d$ (so $\lambda_i \neq \lambda_j$ for all $i \neq j$).

What will be important for us is that these matrices are diagonalizable — in this case, we know that

$$P_A(s) = (\lambda_1 - s)\cdots(\lambda_d - s).$$

As we've seen, the roots of the characteristic polynomial lead to eigenvalues, so for every i, there exists $y_i \neq 0 \in \mathbb{R}^d$ such that $Ay_i = \lambda_i y_i$.

Then the $(y_i)_{i\in\{1,\dots,d\}}$ form a basis of $\mathbb{R}^d$ — we've seen a few weeks ago that if we have a family of vectors associated to distinct eigenvalues, then they are linearly independent. Also, any d linearly independent vectors in dimension d must form a basis.

Definition 33.4. A is *diagonalizable* if it admits a basis of eigenvectors.

We can rephrase what we just proved in this terms:

Corollary 33.5

If $A \in \mathbb{R}^{d\times d}$ has d distinct eigenvalues, then it is diagonalizable.

(We can compute whether A has distinct eigenvalues from the characteristic polynomial.)

So almost every matrix is diagonalizable; we will see that this is the easiest situation for us.

Remark 33.6. An equivalent definition of diagonalizability is that A is diagonalizable if there exists an invertibel matrix P such that $P^{-1}AP$ is diagonal (i.e., of the form

$$\begin{bmatrix}\lambda_1 & & \\ & \lambda_2 & \\ & & \end{bmatrix}.$$

) This definition is equivalent, and you can go from one to the other in the following way — suppose we have a basis of eigenvectors $y_1, \dots, y_d$. Then we can define P to be the matrix with columns y_1, y_2, $\dots$, y_d. Since the y_i rform a basis, this matrix is automatically invertible.This is also the matrix such that $P((0,0,1,9,\cdots) = y_i$.)

This is how you go from the first definition to the second. To go from the second to the first, define y_i as the ith column of P, i.e., as $P((0,\cdots,1))$. So both derfinitions are equivalent.

Definition 33.7. Two matrices A and B are similar if there exists an invertible matrix P such that $M = P^{-1}AP$.

So being diaonlia]able is the same as being similar to a diagonal matrix.

Diagonalizability is something we proved in any (even infinite) dimension.

Corollary 33.8

If $A \in \mathbb{R}^{d\times d}$ is a *symmetric* matrix (i.e., $a_{ij} = a_{ji}$), then it is diagonalizable in an orthonormal basis.

Equivalently, there exists P such that $P^{-1} = P^\intercal$ and $P^{-1}AP = P^\intercal AP$ is diagonal. (This is true for complex as wel,)

Now we'll find the exponential of a diagonal matrix. Assume that $P^{-1}AP$ is the diagonal matrix $\lambda_1, \dots, \lambda_d$, which we denote by D. Then the exponential of D is the diaognal matrix

$$e^{tD} = \begin{bmatrix} e^{t\lambda_1} & & \\ & e^{t\lambda_2} & \\ & & \end{bmatrix}.$$

Meanwhile, if A is diagonalizable then we define its exponentail as $Pe^{tD}p^{-1}$.

Idea of Proof. For the first statement we can jjust do it. For the second, we want to solve$U' = AU$ with $U(0) = I$. This is equivalent to $U' = PDP^{-1}U$, and $U(0) = I$. Since P is invertible, we can diide by both sides to get $P^{-1}U = DP^{-1}U$ with the initial condtiion $P^{-1}U(0) = P^{-1}$. The only solution to this can be written thanks to the exponential of D. LETTING $V = P^{-1}U$ we then get $V = e^{tD}P^{-1}$ by the definition of $\exp tD$. SO $U = Pe^{tD}e^{-1}$. (Note that P is a constnant matrix so we can say that $V' = P^{-1}U'$ — doing this in coordinates, we have $(P^{-1}U)_{ij} = \sum$)

(Solving

$$\begin{cases} U' = AU \\ U(0) = U_0 \end{cases}$$

is — the solution of this is $e^{tA} \cdot U_0$, because you can verify that this is a solution, and you have uniqueness. □

Now we've dealt with the diagonalizable case. This gives the following:

Proposition 33.9

The solution in $\mathbb{R}^d$ of $Y' = AY$ and $Y(t_0) = Y_0$ is $Y(t) = e^{(t-t_0)A}Y_0$.

So once you know the exponentail of a matrix, youknow exactly what the solution is.

This is true in general (not just in the diagonalizable case).

§33.2 Non-diagonalizable case

The diagonalizable case is the generic one, but what happens if you're not diagonalizable? (In math your matrices will come from specific situations so they won't be generic; in CS if you want stable algorithms then you can't always assume diagonalizability, and if you want to compute the exponential of a matrix it is a bad idea to diagonalize it first, you should instead use the formula we will see soon.)

First, what matrices aren't diagonalizable? A typical not diagonalizable matrix is nilpotent — meaning that some power of it is zero, meaning that $A^k = 0$ for some k.

Example 33.10

Any matrix with 0's on and below the diagonal (and anything above it) is nilpotent, e.g.,

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

(If you are nilpotent, you will always satisfy $A^d = 0$.)

These matrices are not diagonalizable, but they come up in a lot of situations. Some exercises (though we won't need all of them; some are much harder than others).

Exercise 33.11. (1) The characteristic matrix of a nilpotent matrix has only 0 as a root, i.e., it is $(-s)^d$. (Equivalently, the only eigenvalue is 0.)

(2) (Harder) Any nilpotent matrix is similar to a matrix with all 0's on and below the diagonal, some 0's and then 1's immediately above it, and all 0's above that. The number of 1's should be the smallest k such that $A^k = 0$, minus 1.

One way to prove (2) is to define k_0 to be the smallest such that $A^k = 0$, and then consider $v \neq 0$ such that $A^{k-1}v \neq 0$. The $A^\ell v$ for $0 \leq \ell \leq k_0 - 1$ will give you the beginning of a basis.

Nilpotent matrices will be the main problem — we'll see soon that any matrix is a diagonalizable plus nilpotent matrix, and they commute. One thing we will conclude with is the next definition:

Definition 33.12. The *exponential* of A is

$$e^{tA} = \sum_{k=0}^{\infty} \frac{(tA)^k}{k!}.$$

(This isan equivalent definition.) This matches the definition of the exponential in the usual case. (There are many functions you can define on matrices just through their Taylor series, e.g., the logarithm or square root of a matrix. THis will hae all the right properties, e.g., exponential of the log is A itself, and so on.,) We will show next time that these are equivalent and how to use it. One key point is that if you are nilpotent, then this sum is always finite (and only goes up to d), so it's just a fiini

($k_0 \leq d$ because of e.g. the harder question).

§34 May 8, 2023

There are two types of matrices we'll care about. The first type are *diagonalizable*, meaning that A is similar to a diagonal matrix

$$D = \begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \ddots \end{bmatrix}$$

(where *similar* means that $A = PDP^{-1}$). The second type are *nilpotent*, meaning that A is symmetric to a matrix

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

We'll see how to decompose matrices into a diagonalizable and nilpotent part, which will be useful to compute matrix exponentials.

Definition 34.1. We have the following two equivalent definitions of the matrix exponential:

- $t \mapsto e^{tA}$ is the unique solution of
$$\begin{cases} U = AU \in \mathbb{R}^{d\times d} \\ U(0) = I. \end{cases}$$
- $e^{tA} = \sum_{k=0}^{\infty} \frac{(tA)^k}{k!}$.

(The fact that these are equivalent in general follows from the fact that you can solve the ODE in terms of Taylor coefficients.)

Both equations give the same exponential for diagonal matrices. Meanwhile if $A = PDP^{-1}$ for D diagonal, then $e^{sA} = Pe^{sD}P^{-1}$ (in either definition). So in this case, we can see explicitly that the two definitions lead to the same formula.

§34.1 Exponential of a Nilpotent Matrix

The second formula is particularly nice for nilpotent matrices, because this sum will be finite (since large enough powers of our matrix must vanish). From the second definition, if N is nilpotent — which in particular implies $N^d = 0$ — then we have

$$e^{sN} = \sum_{k=0}^{d-1} \frac{(sN)^k}{k!}.$$

In particular, the exponential of a nilpotent matrix is just a polynomial.

The first definition is a bit more annoying, but we'll see that something happens that also explains why we only see polynomial coefficients. We'll first solve the general ODE

$$\begin{cases} Y' = NY \in \mathbb{R}^d \\ Y(t_0) = Y_0 \end{cases}$$

(we're solving the equation for vectors first; then the same applies to every column of our matrix). Then we can write

$$N = PDP^{-1},$$

where P has zeroes everywhere except the diagonal immediately above the main one, where it has k_0 1's and the remaining 0's. Call this matrix $\widetilde{N}$; then solving the equation for $\widetilde{N}$ will give us a solution for N as well.

We want to solve

$$\begin{cases} Y' = \widetilde{N}Y \\ Y(t_0) = Y_0. \end{cases}$$

This is equivalent to a system of ODEs

$$\begin{cases} y_1' = y_2 \\ y_2' = y_3 \\ \vdots \\ y_{k_0}' = y_{k_0+1} \\ y_{k_0+1}' = 0 \\ \vdots \\ y_d' = 0. \end{cases}$$

We also have initial conditions $y_1(t_0) = y_1^0, \ldots, y_d(t_0) = y_d^0$. Some of these equations can be solved instantly — for example, we get $y_{k_0+1}(t) = y_{k_0+1}^0, \ldots, y_d(t) = y_d^0$. For the remaining equations, we can solve all of them by repeated integration — first we get

$$y_{k_0}(t) = y_{k_0}^0 + (t - t_0)y_{t_0+1}^0$$

from integrating the equation $y_{k_0}'(t) = y_{k_0+1}(t) = y_{k_0+1}^0$. We can solve all the other equations in the same way, by repeatedly integrating — then we know

$$\begin{aligned} y_{k_0-1}(t) &= y_{k_0-1}^0 + \int_{t_0}^t y_{k_0}(s)\,ds \\ &= y_{k_0-1}^0 + \int_{t_0}^t \left(y_{k_0}^0 + (s - t_0)y_{k_0+1}^0\right) ds \\ &= y_{k_0-1}^0 + y_{k_0}^0(t - t_0) + y_{k_0+1}^0 \frac{(t - t_0)^2}{2}. \end{aligned}$$

We can keep on doing this, and we'll eventually get

$$y_1(t) = y_1^0 + \int_{t_0}^t y_2(s)\,ds,$$

which will be a polynomial in $t - t_0$ of degree k_0.

So all the coefficients we'll get for the exponential of $\widetilde{N}$ are polynomials; when we multiply back by P and P^{-1} to obtain $e^{sN} = Pe^{s\widetilde{N}}P^{-1}$, this remains true.

Remark 34.2. The fact that $e^{sN} = Pe^{s\widetilde{N}}P^{-1}$ is true in general, and useful whenever we have a matrix similar to a simpler one. We already saw this argument in the diagonalizable case — solving $Y' = NY$ is the same as solving $P^{-1}Y' = \widetilde{N}(P^{-1}Y)$. Letting $V = P^{-1}Y'$, we get a new ODE $V' = P^{-1}V$; then to recover Y we just multiply by P.

Alternatively, from the second definition, note that

$$N^2 = (P\widetilde{N}P^{-1})(P\widetilde{N}P^{-1}) = P\widetilde{N}^2P^{-1},$$

and the same occurs for any power — we have $N^k = P\widetilde{N}^kP^{-1}$, and so we can get the same formula from the sum.

Exercise 34.3. Verify that this recovers the original formula $e^{s\widetilde{N}} = \sum_{k=0}^{\infty} \frac{(s\widetilde{N})^k}{k!}$.

The reason we care about diagonalizable and nilpotent matrices is that every matrix can be written as a sum of two such matrices, which *commute* (this will be important).

Proposition 34.4

If A and B commute (i.e., $AB = BA$), then $e^{t(A+B)} = e^{tA}e^{tB}$.

Note that this is *not* true in general (i.e., when A and B don't commute); be careful not to apply it to matrices that don't commute. (A lot of formulas which hold for usual numbers break for matrices in general because of commutativity — for example, without commutativity you can only say $(A+B)^2 = A^2 + AB + BA + B^2$, and not $A^2 + 2AB + B^2$. So you can imagine that without commutation, the infinite sum of powers of $A+B$ is really bad.)

Remark 34.5. It is also true that $e^{s(tA)} = e^{stA}$.

Proof. The proof is essentially the same as in the usual exponential of a number. We'll use the formula

$$e^{t(A+B)} = \sum_{k\geq 0} \frac{t^k}{k!}(A+B)^k = \sum_{k\geq 0}\sum_{\ell=0}^{k} \frac{t^k}{k!}\binom{k}{\ell}A^\ell B^{k-\ell}.$$

(Note that in order to get this, we used the fact that A and B commute, so we can rearrange any term with ℓ A's and $k-\ell$ B's into $A^\ell B^{k-\ell}$.) We can rewrite this as

$$\sum_{k\geq 0}\sum_{\ell=0}^{k} \frac{t^\ell A^\ell}{\ell!}\frac{t^{k-\ell}B^{k-\ell}}{(k-\ell)!} = \sum_{m\geq 0}\frac{t^m}{m!}A^m \cdot \sum_{n\geq 0}\frac{t^n}{n!}B^m = e^{tA}e^{tB}.$$

□

Proof 2. We want to solve

$$\begin{cases} U' = (A+B)U \\ U(0) = I. \end{cases}$$

We know there's a unique solution; we'll verify that $e^{tA}e^{tB}$ also satisfies the above equation, and therefore by uniqueness it'll be $e^{t(A+B)}$ (since $e^{t(A+B)}$ is defined as the unique solution).

Define $V(t) = e^{tA}e^{tB}$; we will prove that V solves

$$\begin{cases} V' = (A+B)V \\ V(0) = I. \end{cases}$$

To do this, we can just differentiate — we have

$$V'(t) = Ae^{tA}e^{tB} + e^{tA}Be^{tB}.$$

(Here we used the ODE definition of the exponential, which implies that $\partial_t e^{tA} = Ae^{tA}$ — this is the definition we're using, though it can also be checked term by term in the Taylor expansion).

We'll again use the fact that $AB = BA$ to conclude that this equals

$$(A+B)e^{tA}e^{tB} = (A+B)V(t).$$

Of course $V(0) = I \cdot I = I$. So the two definitions lead to different proofs, but what's at the core of the proof is the commutation of A and B. □

Exercise 34.6. The formula is actually *equivalent* to commutativity — we have $e^{t(A+B)} = e^{tA}e^{tB}$.

This can be proven just by looking at the second derivative at 0 of the two sides.

Finally, we'll state an important theorem (which is commonly used in math, computer science, and probably other fields).

Theorem 34.7 (Jordan–Chevalley, Dunford)

Let M be a $d \times d$ matrix. Then there exist A and N such that:

(a) $M = A + N$;

(b) A is diagonalizable;

(c) N is nilpotent;

(d) A and N commute.

(These are essentially unique, though with some subtleties.)

Remark 34.8. We won't prove this; you'll see the proof in a theoretical version of a linear algebra class.

Remark 34.9. You always want to decompose your matrices into nice pieces, since they're big; here our nice pieces will be diagonalizable and nilpotent (which have much fewer terms).

Then you can use this to compute the exponential of any matrix M — we'll have $e^{tM} = e^{tA}e^{tN}$.

Remark 34.10. The matrices A and N are constructed in the following way: we first attempt to find an invertible matrix P such that

$$P^{-1}MP = \begin{bmatrix} M_1 & & \\ & M_2 & \\ & & \ddots \\ M_\ell & & \end{bmatrix}$$

is 'diagonal by block,' and where each of the M_i is of a very specific form (they correspond to the eigenvalues) — we have

$$M_i = \begin{bmatrix} \lambda_i & 1 & \\ 0 & \lambda_i & 1 \end{bmatrix}.$$

Then the commutativity will be obvious, since it's true on each of these blocks. The size of M_i will be the multiplicity of λ_i as a root of P_M (the characteristic polynomial), and the number of 1's will be the order of nilpotence of $M - \lambda_i I$ (or maybe the order ± 1).

The first step of the proof uses the Cayley–Hamilton theorem, a very nice and simple theorem which states the following:

Theorem 34.11

$P_M(M) = 0$.

This tells you that all the $M - \lambda_i$ to some power vanish; then once you know that they vanish, it tells you that $M - \lambda_i$ is a nilpotent matrix. We know that nilpotent matrices look like these, so you can do this on each block. It's not an easy proof, but it's uses very nice linear algebra ideas.

The next step will be to apply this to matrices; we'll then be able to compute the exponential of any matrix thanks to this theorem.

We'll later talk about what happens dynamically in an equation $Y' = AY$ (does Y blow up? Go to 0?) We'll see how this relates to the eigenvalues of A.

§35 May 10, 2023

In the last few lectures, we'll talk about the more dynamical aspect of ODEs. What we've done lately in L^2, H^k spaces, and Fourier series is extremely important in analysis; to learn more about them, you want to take classes in functional analysis, and analysis of PDEs.

But there is another aspect to ODEs (ODEs are related to lots of aspects of math, including number theory and algebra) — *dynamical systems*. We'll relate this to our exponential matrices and linear ODEs.

When you have a dynamical system, everything can be described in terms of what happens to the fixed points of the dynamic — you may have a few points where the dynamic is trivial. For example, suppose you have a ball and a mountain, and you want to put the ball at a point and say where it goes. If you put it at any of the orange points (the critical points), then it will not move — these are fixed points of the dynamic in the sense that if you put a ball there, it will not move. The idea is that by only analyzin these fixed points of the dynamic (meaning the points that don't move), we should be able to understand the dynamic from every other point.

The picture is that by looking at any of these points, we should be able to say what happens if we're close to these points. For the first point, a bit on the left we'll fall to the left, and a bit on the right we'll fall towards the critical point. If we start at the second, then on either side we'll fall; if we start at the third, then on either side we go back to the critical point; from the fourth we fall in either direction. So at each of these points, we can write whether we go left or right in each direction. This tells us everything about the dynamic at a qualitative level.

All of these can be read from the behavior of the dynamical system close to one of the fixed points; our analysis of some linear ODEs will exactly tell us what happens close to these points, and we'll be able to understand what happens to the global dynamic as well (if you let it evolve, you'll be able to say given where the ball starts, where it will end up).

For this, what is crucial is the notion of *stability*. We'll say that the first critical point is unstable, as are the second and the fourth; and the third is stable. Understanding the stability helps us understand the dynamic, so that's what we're going to try to define and understand today, and we'll relate it to the dynamic we've seen for ODEs in dimension d — just by knowing the matrix of a linear ODE, we'll be able to say directly from computing its eigenvalues, whether it is stable or not.

The notion of stability is one where a lot of people use different definitions, but we will see two definitions that have the same idea — for a fixed point to be stable means that if we perturb around the point, we should flow back to the fixed point.

Definition 35.1. Consider a first-order ODE

$$\Big\{Y' = F(Y) \in \mathbb{R}^d. \tag{E}$$

Then $Y_* \in \mathbb{R}^d$ is a *fixed point* of (E) if (the following two statements hold; they are equivalent):

(1) The constant vector Y_* is a solution to

$$\begin{cases} Y' = F(Y) \\ Y(t_0) = Y_*. \end{cases}$$

(2) Equivalently, $F(Y_*) = 0$.

(Here we restrict ourselves to the case where the right-hand side doesn't depend on time; the notes consider the general case.)

This is a point that doesn't move — if we start the dynamic at that point, then it won't move (note that there's only one solution through Y_* by uniqueness).

Definition 35.2. A fixed point Y_* is *Lyapunov–stable* if for all $\varepsilon > 0$, there exists $\delta > 0$ such that if Z solves

$$\begin{cases} Z' = F(Z) \\ Z(0) = Z_0 \end{cases}$$

and $\|Z_0 - Y_*\| \le \delta$, then for all $t > 0$ we have $\|Z(t) - Y_*(t)\| < \varepsilon$.

In other words, this definition of stability says that if we start close to a fixed point, then we should stay close to it. Take our fixed point Y_* (so if we start there, we stay there). We'll have two radii δ and ε (where δ is a smaller radius). If we start at a point Z_0 in the smaller circle of radius δ, then if we run the dynamic, we should always stay in the larger circle of radius ε.

Remark 35.3. There are ways to write dynamical systems while keeping track of momentum as well; this is similar to how we went from second-order ODEs to higher-dimensional first-order ODEs.

We'll also have a second type of stability:

Definition 35.4. We say that Y_* is *asymptotically stable* if there exists $\delta > 0$ such that if $\|Z_0 - Y^*\| < \delta$, then $\lim_{t\to\infty} \|Z(t) - Y^*(t)\| = 0$.

This time we're not saying that if we start close, we stay close; we're more strongly saying that if we start close, then we converge to the point. So if we start in the circle, then we might initially go far away, but we eventually have to approach our point as $t \to \infty$.

Example 35.5

If we have a flat plane, so that every point is a fixed point, then if we start at Z_0 we will always stay at Z_0 (we don't flow to Y_*). This is Lyapunov stable (you can take $\delta = \varepsilon$), but not asymptotically stable.

Remark 35.6. We will see that in some situations, you can never be asymptotically stable for specific reasons (e.g., Hamiltonian systems will never be asymptotically stable, but always Lyapunov stable; gradient flows will be the opposite). There are examples the other way, but typically Lyapunov stable is weaker than asymptotically stable (you can imagine something goes far away and then flows back to the fixed point, but that doesn't happen very often).

Definition 35.7. We say that Y_* is *unstable* otherwise (i.e., it doesn't satisfy either definition of stability).

Exercise 35.8. Consider the following situations:

(1) $y' = \lambda y$, with the fixed point $y_* = 0$. This is:
   - Lyapunov stable if $\lambda \leq 0$.
   - Asymptotically stable if $\lambda < 0$.
   - Unstable otherwise.

(2) Consider $y'' + y = 0$, with $Y_* = (0, 0)$ (meaning that $y(0) = 0$ and $y'(0) = 0$). This is Lyapunov stable, but not asymptotically stable.

(3) Consider $y' = y^2 - 1$. This has fixed points $Y_* = -1$ and $Y_* = 1$. Check their stability.

(The first example will be the most important for us.)

§35.1 Linear Stability

There's another notion of stability, which is *linear* stability. This is often the only thing you can easily compute (while Lyapunov stability and asymptotic stability may be very hard to verify by hand, linear stability is quite simple).

Consider the ODE

$$Y' = F(Y).$$

If Y is close to Y_* and F is C^1, then you can expand $F(Y)$ close to Y_* as

$$F(Y) = d_{Y_*}F(Y - Y_*) + \text{negligible}$$

(where $d_{Y_*}F(Y)$ is the differential of F at Y_* in the direction of $Y - Y^*$, and negligible means small compared to $|Y - Y_*|$). (Here what we did is linearize the function at Y_*.)

(Note that there is no $F(Y_*)$ because the fact that Y_* is a fixed point means $F(Y_*) = 0$. At any other point, the leading term would be $F(Y_*)$ and you could just look at that term to understand the dynamic. If this term vanished as well, then you'd have to go to the next one, and so on.)

The first term is the leading-order one, so the dynamic is really sort of determined by what happens to this linear operation on the vector. And this linear operation can be written as a matrix.

We can write this as

$$F(Y) = M(Y - Y_*) + \text{negligible},$$

which lets us rewrite the dynamic in a new way.

Define $Z = Y - Y_*$. Then the leading order dynamic is

$$Z' = MZ, \tag{$\text{lin}_{Y_*}E$}$$

where M is the Jacobian matrix of F at Y_* (which is exactly the matrix of the linear operation $d_{Y_*}F$). You can compute this matrix, and then we will analyze this dynamic next — this equation is solved by the exponential of the matrix M, and knowing the eigenvalues of M will let us completely understand the dynamic.

In other words, the exponential of M determines the local dynamic. This is what we'll try to prove.

Definition 35.9. Y_* is *linearly stable* if 0 is stable for the dynamic $Z' = MZ$.

This is much easier to compute — it just needs us to compute one matrix (which comes frmo differentiating once), and then its eigenvalues (we'll very soon see how to deduce stability from the eigenvalues).

Theorem 35.10

Let M be a real $d \times d$ matrix. Then $e^{tM} \in \mathbb{R}^{d\times d}$ is a matrix whose components are linear combinations of $t \mapsto t^k e^{at}\cos(bt)$ and $t \mapsto t^k e^{at}\sin(bt)$, where $\lambda = a + ib$ is an eigenvalue of M, and $k \leq d-1$.

Corollary 35.11

If for every eigenvalue λ of M we have $\Re(\lambda) < 0$, then 0 (i.e., the zero vector) is Lyapunov and asymptotically stable.

Remark 35.12. The criterion $\Re(\lambda) < 0$ is quite common.

Proof of Corollary. For all k, all $a < 0$, and $b \in \mathbb{R}$, the map $t \mapsto t^k e^{at}\cos(bt)$ (or sin) is bounded for $t > 0$ (which gives Lyapunov stability) and goes to 0 as $t \to \infty$ (which gives asymptotic stability).

More explicitly, the solution to

$$\begin{cases} Z' = MZ \\ Z(0) = Z_0 \end{cases}$$

is $e^{tM}Z_0$. All entries of e^{tM} will be sums of terms of this form, and all these terms are bounded and go to 0 as $t \to \infty$. This means $e^{tM} \to 0$ as $t \to \infty$, so $e^{tM}Z_0 \to 0$; and the fact that they're bounded also lets you prove Lyapunov stability. □

This is similar to saying you are stable if your Jacobian is negative definite. It's not the same thing because if it's not a symmetric matrix, you can't talk about being negative definite. But the symmetric case is fairly similar, so as another corollary:

Corollary 35.13

If M is symmetric and negative definite, then 0 is stable (both Lyapunov and asymptotic).

Remark 35.14. If M is symmetric, then it only has real eigenvalues. Asking for all the real eigenvalues to be negative is equivalent to asking for the entire matrix to be negative definite. You can diagonalize symmetric real matrices in an orthonormal basis, and they have real eigenvalues.

Proof Sketch of Theorem. We know that M can be written as $M = A + N$ such that A is diagonalizable and N is nilpotent, and A and N commute. This tells us $e^{tM} = e^{tA}e^{tN}$.

First we have

$$e^{tA} = P \begin{bmatrix} e^{t\lambda_1} & \\ & e^{t\lambda_2} \end{bmatrix} P^{-1}.$$

This is a linear combination of the $e^{t\lambda_i}$, which gives you what you want. So the coefficients of this matrix are all exponentials of the eigenvalues.

(The fact that we only have cosines and sines is because we have a real matrix, so its exponential must stay real.)

Then we need to understand the exponential of N. We've seen that the exponential of a nilpotent matrix is simply a polynomial — $e^{tN} = \sum_{k=0}^{d-1} \frac{(tN)^k}{k!}$ is a polynomial of degree $d-1$ (or rather, a matrix with polynomial entries of degree $d-1$). Then taking products of these two matrices, all our terms are of the desired form. □

We'll see soon that the local dynamic close to a fixed point can be read at the linear level (where we look at the linear dynamic given by the Jacobian instead), and this is directly seen from the eigenvalues of M.

§36 May 12, 2023

We're going to skip the proof of the fact that stability can be read off the eigenvalues of the matrix, and just give the statement:

Theorem 36.1

Consider the equaiton $Y' = MY$ with eigenvalues of M $\lambda_1, \ldots, \lambda_n$. Then

- 0 is asymptotically stable if and only if for all i we have $\Re(\lambda_i) < 0$.
- 0 is Lyapunov stable if and only if for all i we have $\Re(\lambda_i) \leq 0$, and if $\Re(\lambda_i) = 0$ then λ_i corresponds to a diagonal block (i.e., it only has zeroes) in the Jordan decomposition.

We won't prove (2) — it requires more about the Jordan decomposition — but (1) is because $\exp(tM)$ is a linear combniation fo $t^k e^{\lambda_i t}$, which goes to 0 as $t \to +\infty$ if and only if $\Re(\lambda_i) < 0$.

For (2), if we have

$$\begin{bmatrix} i\alpha & & \\ & i\alpha & \\ & & i\alpha \end{bmatrix},$$

the solutions will be linear combinations of $\cos(\alpha t)$ and $\sin(\alpha t)$. A priori it should be multiplied by a power of tk, but this power of t won't appear if the matrix is diagonal; so asking for only 0's off the diagonal rules out powers of t (which would make us blow up at ∞). (But this case is very nongeneric, so you probably won't see it often.) The point is that t^k times an exponential will still decay, but if we don't have the exponential then we need to eliminate the t^k's.

We'll see a few more definitions, and then classify the behavior of a 2-dimensional system.

§36.1 Directions of Stability and Instability

We've seen that being stable essentially meant that our eigenvalues have negative real part. For us, these correspond to the stable directions, and the ones with positive real part will be unstable directions (and the ones with real part 0 will be unclear, where we need to verify these properties).

The definition is a bit more complicated because it involves *generalized* eigenvalues as well.

Let $M \in \mathbb{R}^{d\times d}$, and decompose its characteristic polynomial as

$$P_M(s) = (\lambda_1 - s)^{m_1} \cdots (\lambda_\ell - s)^{m_\ell}.$$

(Here $\lambda_1, \ldots, \lambda_\ell$ are the eigenvalues of M, and $m_1, \ldots, m_\ell$ are their multiplicities.)

Definition 36.2. The *generalized eigenspace* is $E_{\lambda_i} = \ker((M - \lambda_i I)^{m_i})$.

One crucial result (which we will not prove — we almost have the tools to prove it, but we don't have time) is:

Theorem 36.3

$\mathbb{R}^d = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_\ell}$.

(This is the next-best thing we can do — we might not be diagonalizable so we might not be able to write the whole space as a direct sum of eigenspaces, but we can write it as a direct sum of generalized eigenspaces.)

The operation $\oplus$ denotes 'direct sum' — it means that if $v_1 + v_2 + \cdots + v_\ell = 0$ with $v_i \in E_i$, then we must have $v_i = 0$ for all i, and every vector $v \in \mathbb{R}^d$ can be written as $v_1 + \cdots + v_\ell$ for $v_i \in E_{\lambda_i}$. (This is basically being a basis, except that instead of lines we're taking subspaces.)

The key point is that this lets us define stable and unstable subspaces:

- The *stable* subspace for M is $E^S = \bigoplus_{\Re(\lambda_i)<0} E_{\lambda_i}$.
- The *unstable* subspace for M is $E^U = \bigoplus_{\Re(\lambda_j)>0} E_{\lambda_j}$.
- The *center* subspace (which is more complicated to deal with, so we won't) is $E^C = \bigoplus_{\Re(\lambda_k)=0} E_{\lambda_k}$.

We have $\mathbb{R}^d = E^S \oplus E^U \oplus E^C$. (So any vector can be written as a sum of a stable vector, an unstable vector, and an unclear one.)

If we start with a vector in E^S, then 0 on this subspace will be asymptotically and Lyapunov stable. If we start with one in E^U it will be unstable in either direction. If we start with one in E^C, it won't be asymptotically stable and may or may not be Lyapunov stable.

Example 36.4

Suppose that

$$M = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Then the x-axis E^U is unstable, and the y-axis E^S is stable. What the dynamic will look like on these subspaces is very simple —- on the orange one the dynamic will always make you flow back to 0 (orange line with arrows towards origin), while in the blue direction we will flow away from 0 (arrows away from origin). In general we say this dynamic is unstable; if you start anywhere, you will start following a line (shaped as like $1/x$ with arrows) where you flow vertically towards 0 but then horizontally away.

This is sometimes crucial to understanding in which directions you will be stable or unstable; this lets you completely read what the dynamic will be close to a fixed point (i.e., 0). This says wherever we start, if it's not on the orange line then we will flow away exponentially fast; but if we start on the orange line then we will stay in it and will converge exponentially fast to 0.

So understanding the eigenvalues and generalized eigenspaces lets you understand the dynamic — it tells you not only there are some direcctions we are unstable, but also which directions are stable or unstable.

§36.2 Local Behavior of 2-dimensional Systems

In dimension 2 we'll look at a few possible matrices, and see how the dynamic behaves through a drawing like this. We'll see that even in dimension 2, we'll have a lot of different situations to consider, separated by stability.

Case 1 (Asymptotically stable). There are a few situations in which we will be asymptotically stable.

- Both eigenvalues are real and $\lambda_1, \lambda_2 < 0$ — then our eigenspaces will look like two lines, and $\mathbb{R}^2 = E^S$. But it might be stable at different rates depending on whether we are on λ_2 or λ_1. The dynamic will be that we flow back to 0 (drawn with arrows), but we need to be a bit careful — it won't be straight lines, because one of the two eigenvalues might be larger than the other. So our dynamic will look like curves that go through 0.

  Exercise 36.5. Solve $Y' = \begin{bmatrix} -1 & 0 \\ 0 & -2 \end{bmatrix} Y$ with initial condition $Y(0) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

  This should give

  $$Y(t) = \begin{bmatrix} e^{-t} \\ e^{-2t} \end{bmatrix},$$

  and the picture will look like one of our yellow lines.

- The eigenvalues are complex conjugates; call them $(\lambda_1, \lambda_2) = (a + ib, a - ib)$, where $\Re(\lambda_1) = \Re(\lambda_2) = a < 0$.

  Then it's hard to represent the two eigenspaces (which are complex). The whole space will again be stable, and the picture will look like a spiral going towards 0 (exponentially fast). If you solve

  $$Y' = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$

  you will find that

  $$\exp(tM) = e^{at} \begin{bmatrix} \cos(bt) & -\sin(bt) \\ \sin(bt) & \cos(bt) \end{bmatrix}.$$

  (The fact that the eigenvalues are in complex conjugate pairs is because if we have any polynomial with real coefficients, the complex roots come in conjugate pairs. So for a quadratic polynomial, we have two real roots, two complex conjugates, or a double real root (if the discriminant is 0). The eigenvalues of a matrix are the roots of a degree 2 polynomial, so one of these situations must occur.)

- The last situation is when we have a double real root and the matrix is not diagonalizable. This corresponds to

  $$Y' = \begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix}.$$

  This will still be asymptotically stable, but more difficult to draw. You'll see that the worst case is that you will have some directions with te^{-t}, which does go to 0 as $t \to \infty$ but is not as small as e^{-t}.

  (A diagonalizable thing fits into the first case.)

Case 2 (Lyapunov stable but not asymptotically).

Again we can look at the three situations:

- Two real roots (or eigenvalues) $\lambda_1 = 0$ and $\lambda_2 < 0$ — so one is asymptotically stable and the other is not. In this case we have two directions E_{λ_1} and E_{λ_2}. The first direction does not move; and E_{λ_2} is stable so we flow inwards. (So if we start on E_{λ_1}, we don't move.) The dynamic we get is where we flow exponentially fast to not 0, but our projection on E_{λ_1} parallel to E_{λ_2}.

  This is non-generic, but it can happen — for example, imagine you have half a cylinder. Then you will have a purple line at the bottom, where if you put a ball on this prruple line it won't move; meanwhile if you take an orange point it flows towards the purple line.

  In this case $E_{\lambda_2} = E^S$ and $E_{\lambda_1} = E^C$.
- Two complex conjugate roots. Since they have 0 real part, this will just correspond to rotation —- suppose we have $\lambda_1 = ib$ and $\lambda_2 = -ib$. Then the dynamic will still be rotating, but not going inwards — it'll just look like a bunch of circles (depending on th direction). Note that in this case we will be Lyapunov stable but not asympotitcally stable. In this case $E^C = \mathbb{R}$.
- The final case is when $M = 0$, in which case our dynamic does not move; if we start anywhere we stay there.

(The moment we have something like $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ we become unstable because of the powers of t.)

Case 3 (The Unstable Situation). This is basically the opposite of the stable situation, though it also contains the situations where the eigenvalues have different signs.

- Two real roots:
  - $\lambda_1 > 0$ and $\lambda_2 < 0$. In this case E_{λ_2} is stable, and E_{λ_1} i s unstalbe; the dynamic will ook like $1/x$ lines. In this case $E^S = E_{\lambda_2}$ and $E^U = E_{\lambda_1}$.
  - $\lambda_1, \lambda_2 > 0$. Then both directions are unstable, s owe'll have outwards arrows on both. We'll have behavior kind of like the other one, but flowing away (lines through 0, but the arrows go outwards).
  - Two complex conjugate roots $a \pm ib$ with $a > 0$ — then we will be spiralling away from 0.
  - A double real root 0, which means our matrix looks like

$$Y' = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} Y.$$

    Then $e^{tM} = \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}$.

You can imagine that in higher dimensions there are more cases, but they'll essentially be combinations of the things we saw here.

Remark 36.6. Usually when you're unstable you flow away exponentially fast, but in th e0 case you flow away much slower, at polynomial rates.

Theorem 36.7

If we have $Y' = F(Y)$ for not necessarily linear F, and Y_* is fixed (i.e., $F(Y_*) = 0$), then let $M = d_{Y_*}F$ be the Jacobian matrix (we'll see some examples next time) of F at Y_*. Then:

- If *all* of the eigenvalues of M have negative real part, then Y^* is Lyapunov and asymptotically stable.
- If even one eigenvalue has positive real part, then Y_* is unstable.

To prove this, you use the fact that you know how to solve the linearized equation, adn you approximate the actual dynamic by the linearized one. This is done in the notes.

NOw given a dynamic, we can just look at the critical points, linearize, and be able to say its stable and unstable directions.

Example 36.8

COnsider

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} x^2 - y^2 \\ xy \end{bmatrix} = F(\begin{bmatrix} x \\ y \end{bmatrix}).$$

First you look at the fixed points, meaning when is $F(\begin{bmatrix} x_* \\ y_* \end{bmatrix}) = 0$, and then you linearize F at Y_* to obtain M, and find its eigenvalues and stability.

§37 May 15, 2023

For the last class, we will talk about the method for solving a system of ODEs. If we have time, we'll talk about two types of ODEs that are very common — gradient flow and Hamiltonian systems. (They will be in the notes if we are interested — gradient flow appears in computer science, and Hamiltonian systems in physics and chemistry.)

We'll look into systems of two equations (since our focus isn't on multivariable calculus). When we have a system of ODEs

$$Y' = F(Y)$$

for $Y \in \mathbb{R}^2$, it's sometimes useful to write it in this way, and sometimes easier to write it in coordinates as

$$\begin{cases} x' = f_1(x, y) \\ y' = f_2(x, y). \end{cases}$$

One example where it's easier to think of the function as a vector is in gradient flows:

Example 37.1 (Gradient Flow)

For $\psi \colon \mathbb{R}^2 \to \mathbb{R}$, define $F \colon \mathbb{R}^2 \to \mathbb{R}^2$ as $Y \mapsto \nabla\psi(Y)$ (the gradient of ψ at Y, also denoted $\Delta_Y \psi$).

Here the vector interpretation is better, because the coordinates lose the geometric interpretation. But often if we aren't given such a motivation, we'll mostly look at the system in coordinates.

The first thing to do is find the fixed points — where is the dynamic trivial? It's trivial when solving the ODE with initial coordinates (x_i, y_i) is constant; this means we should have $F(Y_i) = 0$, or equivalently $f_1(x_i, y_i) = f_2(x_i, y_i) = 0$.

This tells us where we have points that won't move along the dynamic. This already gives you strong information — for example, whether your flow goes to infinity or converges to a point can be seen from where and how many fixed points you have. (If you have an even number of fixed points, there are conclusions you can make directly.)

The next thing is to determine whether it's stable at the fixed points, and what the dynamic does. This means we linearize — so we compute the matrix

$$M_* = d_{Y_*} F = \begin{bmatrix} \partial_x f_1(x_i, y_i) & \partial_y f_1(x_i, y_i) \\ \partial_x f_2(x_i, y_i) & \partial_y f_2(x_i, y_i) \end{bmatrix}$$

(which doesn't depend on x and y).

Then once we have this matrix, we should try to compute its eigenvalues, by computing the roots of the characteristic polynomial $\det(M - sI) = 0$.

Just by knowing the eigenvalues (we don't even need to know the eigenvectors), we can say whether the fixed point is stable or not — we know that if $\Re(\lambda_1^i) < 0$ and $\Re(\lambda_2^i) < 0$ then we're (both Lyapunov and asymptotically) stable, while if $\Re(\lambda_1^i)$ or $\Re(\lambda_2^i) > 0$, then we're unstable. So we just need to compute the roots of a degree 2 polynomial, and look at the signs of their real parts.

(The i's represent the index for our fixed point.)

Then we might want to understand in *which* directions our points are stable or unstable — this can be important because sometimes maybe the system has some symmetry and as long as we preserve the symmetry we will be stable, but when we break the symmetry we lose stability. So it can be important to understand this, and it can have physical meaning (if you run a gradient flow then you're trying to minimize a function, and the best thing you can hope for is to find an actual minimum that's minimal in all directions. BUt sometimes it's also important to understand saddle points, which are stable in one direction but unstable in others).

To do this, we want to find the corresponding eigenvectors — i.e., find Y_1^i such that $MY_1^i = \lambda Y_1^i$ and likewise for λ_2^i. Understanding these vectors then lets you understand E_i^U and E_i^S and E_i^C. (If $\Re(\lambda_1^i) < 0$ then Y_1^i is a stable direction, if it's positive then it's an unstable direction.)

THis already tells you a lot — e.g. which points will attract the flow, which will repulse it, and so on.

If your fixed points are in some region, you can compute the vector field $f(x, y)$ at some other (i.e., non-fixed) points; for example, if you want to understand what's happening in the region between them (e.g. understand if the flow is going from one fixed point to another). This can also help you verify your calculations of stability.

(Here we are using i to index our fixed points.)

(Note that everything is a function of t, and all our derivatives are with respect to t.)

§37.1 Some Examples

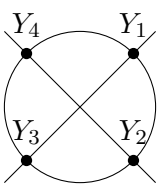
Example 37.2

Consider

$$\begin{cases} x' = x^2 + y^2 - 1 \\ y' = x^2 - y^2. \end{cases}$$

First we want to find the fixed points. We have $f_1(x, y) = 0$ if and only if $x^2 + y^2 = 1$, which tells us that we're somewhere on a circle of radius 1; the second condition $f_2(x, y) = 0$ tells us that $x^2 = y^2$, so $x = \pm y$. Plugging this into the first equation gives us four fixed points —

$$(x, y) \in \left\{ \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right), \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} \right), \left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} \right), \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right) \right\}.$$



The next step is to linearize at these points. First we'll do $Y_1 = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$. First linearizing the first equation with respect to x, the derivative of the right-hand side is $2x = \sqrt{2}$. This gives us

$$\begin{bmatrix} \sqrt{2} & \sqrt{2} \\ \sqrt{2} & -\sqrt{2} \end{bmatrix}$$

(the first equation is symmetric, and the second is antisymmetric in x and y). Similarly, at $Y_2 = (\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})$, the matrix becomes

$$\begin{bmatrix} \sqrt{2} & -\sqrt{2} \\ \sqrt{2} & \sqrt{2} \end{bmatrix}.$$

Now the next step is to compute the eigenvalues; this means we compute

$$\det \begin{bmatrix} \sqrt{2} - s & \sqrt{2} \\ \sqrt{2} & -\sqrt{2} - s \end{bmatrix} = s^2 - 2 - 2 = s^2 - 4.$$

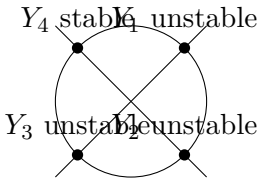
(You should always verify that you get the correct determinant when $s = 0$, that we have leading term s^2 (in general the coefficient is $(-1)^{\dim}$), and that the trace is the linear part.) The roots of this polynomial are ± 2. This lets us directly answer step 4 — this point is unstable, because one of the two roots has positive real part.

For the second one, we have

$$\det \begin{bmatrix} \sqrt{2} - s & -\sqrt{2} \\ \sqrt{2} & \sqrt{2} - s \end{bmatrix} = s^2 - 2\sqrt{2}s + 4 = (s - \sqrt{2})^2 + 2.$$

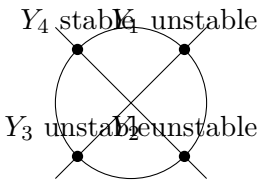
The roots of this polynomial are $\sqrt{2} \pm \sqrt{2}i$. We can see that the real part $\sqrt{2}$ is positive, so this will be unstable as well.

In this case, we can't really compute the stable and unstable directions; it'll just be unstable in every direction. But in the first case, we can compute this.

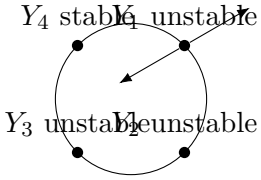


(You can do the same computations for Y_3 and Y_4.)

We also know that both Y_2 and Y_4 have roots which are complex conjugates; this means that we will have an outwards spiral at Y_2, and an inwards spiral at Y_4.



Now the question is — we've seen Y_1 has one stable and unstable direction associated to $+2$ and -2 (and the same is true for Y_3). To obtain this, you compute the eigenvectors. We are not going to do this in class, but you want to solve $MY = \pm 2Y$ with M the above matrix and Y a vector; you want to understand which Y's are associated to $+2$ and -2, and that tells you which directions will be stable and unstable.



What's really cool is that then we can understand which points are sent to which points and how. For example, some points on the Y_2-spiral get sent to the curve on Y_3, and then to Y_4. Typically you go from unstable directions to stable directions. For example, fron the right place at Y_1 we'll go to infinity, but on the left-hand side all curves come from ∞ and go to Y_4, and all the Y_3 and Y_2 curves are part of this.

So —- you can try to understand in which part of the plane, e.g. x and y are increasing or decreasing (i.e., look at the sign of x^2+y^2-1 and of x^2-y^2). The first tells you that inside the circle, x will typically be decreasing (so you typically go left), while outside the circle you typically go to the right. Meanwhile for x^2-y^2, you have two diagonal lines, and on the left and right you go upwards, while on the top and bottom you go downwards.

If you want to say *how* stable or unstable a point is, the more negative the eigenvalue the more stable you are, and the more positive the more unstable you are. But this only tells you what's happening locally around your point, it doesn't tell you everything.

Example 37.3

Consider the system

$$\begin{cases} x' = \sin(x+y) \\ y' = \sin(x-y). \end{cases}$$

Here we will get a bunch of fixed points — we want to solve $x+y=n\pi$ and $x-y=m\pi$ for $n,m\in\mathbb{Z}$, which gives $x=\frac{n+m}{2}\pi$ and $y=\frac{n-m}{2}\pi$. Then for each of these points, we can compute their derivatives and say whether they are stable or unstable.

The two examples we'll see in practice later are the gradient flow and Hamiltonian flow. We won't define them here, but one important fact is that for a gradient flow, the Jacobian matrix M that we get is actually the Hessian of the function, meaning the matrix

$$\begin{bmatrix} \partial^2_{x^2}\psi & \partial^2_{yx}\psi \\ \partial^2_{xy}\psi & \partial^2_{y^2}\psi \end{bmatrix}.$$

This is a real matrix and is symmetric if $\psi\in C^2$, which means it has real eigenvalues and orthogonal eigenvectors. (All the matrices we saw with complex conjugate roots will never happen when we have a gradient flow, because the eigenvalues are real.)