MATH 27300 (Basic Theory of Ordinary Differential Equations) Notes

Steven Labalme

December 6, 2022

Weeks

List of Figures

Week 1

Introduction to ODEs

1.1 Definitions and Scope

9/28:

- Questions:
  - When will the PDFs be made available?
- Office: Eckhart 309.
  - Office hours: MWF 3:00-4:00.
- Reader: Walker Lewis. His contact info is in the syllabus.
- Final grade is based on...
  - 2 midterms (15 pts. each; weeks 4 and 8).
  - Final exam (35 pts.).
  - HW (35 pts.).
  - Bonus problems (15 pts).
- Total points for the quarter is 115. The bonus problems usually arise from advanced math and incorporate more advanced knowledge, and we are encouraged to seek out all relevant resources as long as we write up our own solutions.
- **Ordinary differential equation**: An equation that involves an unknown function of a single variable; an equation that takes the form $F(t, y, y', \dots, y^{(n)}) = 0$. *Also known as* **ODE**.
  - F is a known function.
  - t is an argument (time). x is also used (when space is involved).
  - $y = y(t)$ is an unknown function.
- **Order n** (ODE): An ODE for which the n^{th} derivative of y is the highest-order derivative involved (and is involved).
- ODEs are of the form $y' = f(t, y)$ or, more generally, $y^{(n)} = F(t, y, y', \dots, y^{(n-1)})$.
  - We can transform this second form into the first form via

$$Y = \begin{pmatrix} y \\ y' \\ \vdots \\ y^{(n-1)} \end{pmatrix} \qquad\qquad f(t, y) = \begin{pmatrix} Y^2 \\ Y^3 \\ \vdots \\ F(t, Y^1, Y^2, \dots, Y^{n-1}) \end{pmatrix}$$

This makes $Y' = f(t, Y)$ equal to the system of equations

$$\begin{aligned}(Y^1)' &= Y^2\\ (Y^2)' &= Y^3\\ &\vdots\\ (Y^{n-1})' &= F(t, Y^1, Y^2, \dots, Y^{n-1})\end{aligned}$$

- Think about this conversion more.

– Thus, we mainly focus on equations of the form $y' = f(t, y)$ (where y may be a scalar or vector function), because that's general enough.

- **Linear** (ODE): Any ODE that can be written in the form

$$y' = A(t)y + f(t)$$

- Because of the above, this naturally includes equations of the form

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

– Indeed, if we define $Y = (y, y', \dots, y^{(n-1)})$, then we may express this equation in the form

$$\begin{aligned}
\underbrace{\begin{pmatrix} Y^1\\ Y^2\\ \vdots\\ Y^n \end{pmatrix}'}_{Y'} &= \underbrace{\begin{pmatrix} Y^2\\ Y^3\\ \vdots\\ b(t) - a_0(t)Y^1 - \cdots - a_{n-1}(t)Y^{n-1} \end{pmatrix}}_{g(t,y)}\\
&= \begin{pmatrix} Y^2\\ Y^3\\ \vdots\\ -a_0(t)Y^1 - \cdots - a_{n-1}(t)Y^{n-1} \end{pmatrix} + \underbrace{\begin{pmatrix} 0\\ 0\\ \vdots\\ b(t) \end{pmatrix}}_{f(t)}\\
&= \underbrace{\begin{pmatrix} 0 & 1 & 0 & \cdots & 0\\ 0 & 0 & 1 & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \cdots & 1\\ -a_0(t) & -a_1(t) & -a_2(t) & \cdots & -a_{n-1}(t) \end{pmatrix}}_{A(t)} \underbrace{\begin{pmatrix} Y^1\\ Y^2\\ \vdots\\ Y^n \end{pmatrix}}_{Y} + \underbrace{\begin{pmatrix} 0\\ 0\\ \vdots\\ b(t) \end{pmatrix}}_{f(t)}
\end{aligned}$$

– This conversion and its implications is covered in more depth in Lecture 4.1.

- **Nonlinear** (ODE): An ODE that is not linear.

- **Autonomous** (ODE): An ODE that can be written in the form

$$y' = f(y)$$

– Remember that y can be a scalar or a vector function.

– Solutions to autonomous ODEs can start at *any* time t and still be valid.

- For example, take the scalar ODE $y' = y$. It's general solution is $y(t) = ae^{t-t_0}$ for some $a \in \mathbb{R}$ and t_0 being the start time. Importantly, notice that we can make t_0 take any value we want and $y(t)$ will still solve $y' = y$.

- **Nonautonomous** (ODE): An ODE that is not autonomous.
  - We will not investigate these in this course.
- **Initial value problem**: A problem of the form, "find $y(t)$ such that the following holds."

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

  Also known as **IVP**, **Cauchy problem**.

- Locally well-posed (LWP) conditions:
  1. Existence (local in time).
  2. Uniqueness (you cannot have multiple solutions).
  3. Local stability (if you perturb your initial value or equation a little bit, you do not expect your solution to vary crazily [esp. locally]).
- Example of a nonunique ODE:
  - $y' = \sqrt{y}$, $y(0) = 0$ has solutions $y_1(t) = 0$ ($t \geq 0$) and $y_2(t) = t^2/4$ ($t \geq 0$).
  - We will investigate the reason later.
- Preview of the reason: **Cauchy-Lipschitz Theorem** or **Picard-Lindelöf Theorem**.
  - As long as the ODE is **Lipschitz continuous**, it's locally stable.
- **Lipschitz continuous** (function): A function f such that

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

  - But in the counterexample above, the slope of the chord from 0 to $y(t)$ approaches infinity as $t \to 0$.
- **Peano Existence Theorem**: Under certain conditions, there exists a solution to a given IVP.
- **Dynamical system**: A law under which a particle evolves over time. $y' = f(t,y)$, IVP is LWP.
- If the IVP $y' = f(t,y)$, $y(t_0) = y_0$ is locally well-posed, then the map $\Phi(t,x)$ which solves

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t}\Phi(t,x) = f(t,\Phi(t,x)) \\ \Phi(0,x) = x \end{cases}$$

  is well-defined and satisfies the property

$$\Phi(t_2, \Phi(t_1, x)) = \Phi(t_1 + t_2, x)$$

  - Φ is very related to y, though how exactly is still a bit of a mystery?? Perhaps it's

$$\Phi(t,x) = y(t)$$

    where y is the solution to the IVP $y' = f(t,y)$, $y(0) = x$.
  - It appears that $\Phi(t,x)$ is related to $f_t(x)$ from *Differential Forms*, i.e., we are picking a point x and traveling along its integral curve for time t.
  - Think about $y(t) = ae^{t-t_0}$ as an integral curve of the one-dimensional vector field $X(x) = x$.
  - The final property appears to express the notion that if you have a system and evolve it by time t_1 and then time t_2, that's equivalent to evolving it by time $t_1 + t_2$.

- **Steady flow**: A vector field on a manifold contained in $\mathbb{R}^2$ or $\mathbb{R}^3$ that does not vary with time.
- Let X be a vector field.
  - Trajectory of a particle: At $x \in \Omega$, the velocity of the particle should coincide with $X(x)$.
  - The differential equation $\dot{x} = X(x)$ is what we're interested in.
  - A solid shape gets shifted and deformed (imagine a chunk of water falling out of the end of a pipe). This is the **local group of transformation**.
  - Differential geometry is the purview of such things.
- Newton's law of motion $F = m \cdot a$ applied to n particles is nothing but the system of equations

  $$m_i x_i'' = F_i(x_1, \dots, x_n)$$

  for $i = 1, \dots, n$.
  - Many well-known examples.
  - The best known one perhaps is that of uniform acceleration of a single particle. In this case,

    $$m_0 x'' = f_0$$

    - The solution is

      $$x(t) = \frac{f_0}{2m_0} t^2 + v_0 t + x_0$$

      where $x_0 = x(0)$ and $v_0 = x'(0)$ are the initial conditions.
  - A simple example is downwards motion due to gravity. Then

    $$x(t) = \frac{1}{2} \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix} t^2 + v_0 t + x_0$$

    - The trajectory in general is a parabola.
  - Another example: The mathematical pendulum.
    - The radial directions balance ($mg\cos\theta$).
    - The tangential directions do not ($mg\sin\theta$). Thus, our ODE is

      $$l \frac{\mathrm{d}^2\theta}{\mathrm{d}t^2} = g \sin\theta$$

  - One last set of examples from ecology:
    - Imagine an petri dish of infinite nutrition. The population growth of the bacteria will obey the exponential growth law

      $$\frac{\mathrm{d}y}{\mathrm{d}t} = ky$$

    - Suppose we have a system capacity M. Then we obey the logistic growth law

      $$\frac{\mathrm{d}y}{\mathrm{d}t} = k(M - y)$$

    - Lotka-Volterra prey-predator model: Wolf population (W) and rabbit population (R). We have

      $$\begin{aligned} R' &= k_1 R - aWR \\ W' &= -k_2 W + bWR \end{aligned}$$

    - We can also introduce more species and capacities and et cetera, et cetera.
- Conclusion: Dynamical systems are everywhere, especially in physics, chemistry, and ecology.
- We can also consider long-term behavior.
  - We can have chaos, but chaos can be reasoned with using oscillation, systems that converge to oscillation, etc. We will mostly be focusing on the regular aspect of the long-term behavior.

1.2 Origin of ODEs: Boundary Value Problems

9/30:

- Textbook PDFs will be posted today.
- Note: Equations of order n generally require n parameters to solve.
- Today, we will consider boundary value problems, which are separate from dynamical systems but not entirely unrelated.
- **Boundary Value Problem**: A problem in which we are solving for a y that has fixed values at the boundaries $x = a, b$. *Also known as* **BVP**.
- The **Brachistochrone problem** is an example of a BVP.
- **Brachistochrone problem**: Suppose you have a frictionless track from $(0,0)$ to (a, y_0) and release a particle from $(0,0)$. Which path allows the particle to get to (a, y_0) in the shortest amount of time? *Etymology* brákhistos "shortest" + khrónos "time."

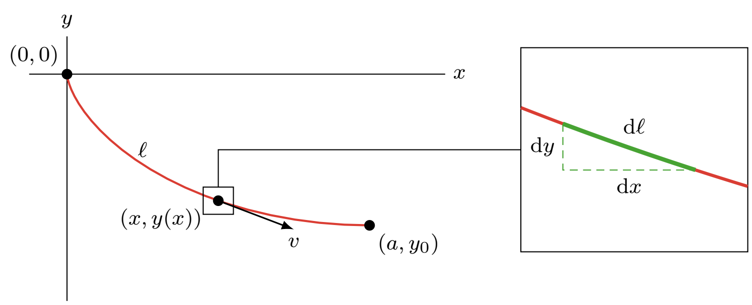


Figure 1.1: Brachistochrone problem.

  - Throughout this derivation, we will make several assumptions. We will do our best to note these assumptions as we go in footnotes. Note that while all of these assumptions are justified in the case of solving this problem, they may not be justified in every related variational problem. Let's begin.
  - Since the track is frictionless, the mechanical energy should be conserved.
  - At a given point along the curve, the particle has a velocity v and is vertical distance y from where it started. We know from physics that

$$\begin{aligned}\frac{1}{2}mv^2 &= mgy\\ v &= \sqrt{2gy}\end{aligned}$$

  - Since $v = \mathrm{d}\ell/\mathrm{d}t$, the time $\mathrm{d}t$ it takes for the particle to traverse an infinitesimal section of track of arc length $\mathrm{d}\ell$ is $\mathrm{d}t = \mathrm{d}\ell/v$.
  - The track should be given by $y = y(x)$[1].
  - Let ℓ denote the arc length of the whole track. Then

$$\mathrm{d}\ell = \sqrt{1 + (y'(x))^2}\,\mathrm{d}x$$

  - Thus, the total time for the particle to traverse the curve is

$$t(y) = \int_0^t \mathrm{d}\tau = \int_0^a \frac{\mathrm{d}\ell}{v} = \int_0^a \frac{\sqrt{1 + (y'(x))^2}\,\mathrm{d}x}{\sqrt{2gy(x)}}$$

[1]There are paths that connect $(0,0)$ and (a, y_0) that are not functions of x. We are taking those out of consideration.

- We also have $y(0)=0$ and $y(a)=y_0$.
- We want to find y such that the above integral is minimized. Thus, we define the following **functional**, which is used to solve general fixed-endpoint variational problems (the Brachistochrone problem is a problem of this type).
- Let $J[y]=\int_a^b F(x,y(x),y'(x))\,\mathrm{d}x$.
- The space of functions we're considering is C^1 (the set of all continuously differentiable functions)[2].
- Take a function h, vanishing at a,b.
- Let $f(t)=J[y+th]$. Then

$$f(t)=\int_a^b F(x,\underbrace{y(x)+th(x)}_{z(x,t)},\underbrace{y'(x)+th'(x)}_{w(x,t)})\,\mathrm{d}x$$

- We know that[3]

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\int_a^b F\,\mathrm{d}x &= \int_a^b \frac{\mathrm{d}F}{\mathrm{d}t}\,\mathrm{d}x\\ &= \int_a^b \left(\frac{\partial F}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t}+\frac{\partial F}{\partial z}\frac{\mathrm{d}z}{\mathrm{d}t}+\frac{\partial F}{\partial w}\frac{\mathrm{d}w}{\mathrm{d}t}\right)\mathrm{d}x\\ &= \int_a^b \left(\frac{\partial F}{\partial x}\cdot 0+\frac{\partial F}{\partial z}\cdot h(x)+\frac{\partial F}{\partial w}\cdot h'(x)\right)\mathrm{d}x\\ &= \int_a^b \left(\frac{\partial F}{\partial z}\cdot h(x)+\frac{\partial F}{\partial w}\cdot h'(x)\right)\mathrm{d}x\end{aligned}$$

  - The last term in the above equation may be integrated by parts as follows. Note that we make use of the hypothesis $h(a)=h(b)=0$ in eliminating the $[uv]_a^b$ term.

$$\begin{aligned}\int_a^b \frac{\partial F}{\partial w}h'(x)\,\mathrm{d}x &= \left[\frac{\partial F}{\partial w}h(x)\right]_{x=a}^b-\int_a^b h(x)\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial w}\right)\mathrm{d}x\\ &= \left[\left.\frac{\partial F}{\partial w}\right|_b\cdot 0-\left.\frac{\partial F}{\partial w}\right|_a\cdot 0\right]-\int_a^b \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial w}\right)h(x)\,\mathrm{d}x\\ &= -\int_a^b \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial w}\right)h(x)\,\mathrm{d}x\end{aligned}$$

  - Substituting back into the original equation, we obtain

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\int_a^b F\,\mathrm{d}x &= \int_a^b \left[\frac{\partial F}{\partial z}\cdot h(x)-\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial w}\right)h(x)\right]\mathrm{d}x\\ &= \int_a^b \left[\frac{\partial F}{\partial z}-\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial w}\right)\right]h(x)\,\mathrm{d}x\end{aligned}$$

  - Therefore,

$$f'(t)=\frac{\mathrm{d}}{\mathrm{d}t}\int_a^b F\,\mathrm{d}x=\int_a^b \left\{\frac{\partial F}{\partial z}-\frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{\partial F}{\partial w}\right]\right\}h(x)\,\mathrm{d}x$$

- Thus,

$$f'(0)=\int_a^b \left\{\frac{\partial F}{\partial z}(x,y(x),y'(x))-\frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{\partial F}{\partial w}(x,y(x),y'(x))\right]\right\}h(x)\,\mathrm{d}x=0$$

  for all h.

[2] This also eliminates some possible paths from consideration.

[3] We must assume sufficient regularity of F here. In particular, we must assume that the derivative of the integral of F is equal to the integral of the derivative of F.

    - Now suppose y is the solution. Then y minimizes $J[y]$. But if this is true, then any variation th will cause $J[y+th] > J[y]$. It follows that for every h, $f(t)$ has a minium at $t=0$. But if f has a minimum at 0 for all h, then $f'(0)=0$ for all h.
- Lemma: Let ϕ be continuous on (a,b). If for every $h \in C^1([a,b])$ vanishing on a,b we have that

$$\int_a^b \phi(x)h(x)\,\mathrm{d}x = 0$$

then $\phi(x)=0$.

Proof. Suppose for the sake of contradiction that (WLOG) $\phi(x_0) > 0$. Then within some neighborhood $N_\delta(x)$ of x_0, $\phi(x) > 0$ for all $x \in N_\delta(x)$. Now choose h to be a bump function on that interval. Then $\int_a^b \phi(x)h(x)\,\mathrm{d}x > 0$, a contradiction. □

- It follows that

$$\frac{\partial F}{\partial z}(x,y(x),y'(x)) - \frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{\partial F}{\partial w}(x,y(x),y'(x))\right] = 0$$

    - This is a second-order differential equation, specifically the **Euler-Lagrange equation**.
    - It is a necessary condition for y to be an extrema.
    - Euler-Lagrange equations are not easy to solve in general. However, we're lucky here.
- In our example,

$$F(x,z,w) = \sqrt{\frac{1+w^2}{2gz}}$$

    - What's nice here is that $F(x,z,w) = F(z,w)$, i.e., there is no dependence on x. This is crucial.
- With this observation in mind, notice that

$$\begin{aligned}\frac{\mathrm{d}F}{\mathrm{d}x} &= \frac{\partial F}{\partial z}\frac{\mathrm{d}z}{\mathrm{d}x} + \frac{\partial F}{\partial w}\frac{\mathrm{d}w}{\mathrm{d}x}\\ &= \frac{\partial F}{\partial z}\frac{\mathrm{d}y}{\mathrm{d}x} + \frac{\partial F}{\partial w}\frac{\mathrm{d}y'}{\mathrm{d}x}\\ &= \frac{\partial F}{\partial z}\frac{\mathrm{d}y}{\mathrm{d}x} + \frac{\partial F}{\partial w}\frac{\mathrm{d}^2y}{\mathrm{d}x^2}\end{aligned}$$

- We now rearrange the E-L equation and multiply through by $\mathrm{d}y/\mathrm{d}x$.

$$\begin{aligned}\frac{\partial F}{\partial z} - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial w}\right) &= 0\\ \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial w}\right)\frac{\mathrm{d}y}{\mathrm{d}x} &= \frac{\partial F}{\partial z}\frac{\mathrm{d}y}{\mathrm{d}x}\end{aligned}$$

- Subtracting the last two results yields

$$\begin{aligned}\frac{\mathrm{d}F}{\mathrm{d}x} - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\mathrm{d}F}{\mathrm{d}w}\right)\frac{\mathrm{d}y}{\mathrm{d}x} &= \frac{\partial F}{\partial w}\frac{\mathrm{d}^2y}{\mathrm{d}x^2}\\ \frac{\mathrm{d}F}{\mathrm{d}x} &= \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\mathrm{d}F}{\mathrm{d}w}\right)\frac{\mathrm{d}y}{\mathrm{d}x} + \frac{\partial F}{\partial w}\frac{\mathrm{d}^2y}{\mathrm{d}x^2}\\ &= \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\mathrm{d}F}{\mathrm{d}w}\frac{\mathrm{d}y}{\mathrm{d}x}\right)\\ \frac{\mathrm{d}}{\mathrm{d}x}\left(F - \frac{\mathrm{d}F}{\mathrm{d}w}\frac{\mathrm{d}y}{\mathrm{d}x}\right) &= 0\\ F - \frac{\mathrm{d}F}{\mathrm{d}w}\frac{\mathrm{d}y}{\mathrm{d}x} &= A\end{aligned}$$

where $A \in \mathbb{R}$ depends on the initial conditions.

- From the definition of F, we can calculate

$$\frac{\partial F}{\partial w} = \frac{w}{\sqrt{1+w^2}} \cdot \frac{1}{\sqrt{2gz}} = \frac{y'}{\sqrt{1+(y')^2}} \cdot \frac{1}{\sqrt{2gy}}$$

- It follows that our solution function y satisfies the separable differential equation

$$\begin{aligned}
\sqrt{\frac{1+(y')^2}{2gy}} - \frac{y'}{\sqrt{1+(y')^2}} \cdot \frac{1}{\sqrt{2gy}} \cdot y' &= A \\
\frac{1+(y')^2}{\sqrt{1+(y')^2}\sqrt{2gy}} - \frac{(y')^2}{\sqrt{1+(y')^2}\sqrt{2gy}} &= A \\
\frac{1}{\sqrt{2gy(1+(y')^2)}} &= A \\
(y')^2 &= \frac{1/2A^2g - y}{y}
\end{aligned}$$

- The solution, as we can determine using methods from Calculus I-II, is the **cycloid**

$$\begin{cases} x = a(\theta - \sin\theta) \\ y = a(1 - \cos\theta) \end{cases}$$

  where the specific parameters come from the boundary values.

- **Functional**: A map from a function space to a set of numbers.
- **Sturm-Liouville problems**: Boundary value problems concerning the integral

$$\int_a^b \left[p(x)(y'(x))^2 + q(x)(y(x))^2\right] \mathrm{d}x$$

  - The most basic BVP is a vibrating string. In finding the eigenmode of the vibration, you need to solve the above differential equation.
  - Very important in physics.
  - If time permits at the end of the course, Shao will return to the following topic in detail.
- Next several weeks: *Solvable* differential equations.

Week 2

Solving Simple ODEs

2.1 Separable ODEs

10/3:

- Do not sit on the left side of the classroom: The sun sucks!
- **Separable** (ODE): An ODE of the form

$$\frac{\mathrm{d}y}{\mathrm{d}t} = f(t)g(y)$$

where y is a real[1], unknown, scalar function of t.

- Solving separable ODEs: Formally, evaluate

$$\int \frac{\mathrm{d}y}{g(y)} = \int f(t)\,\mathrm{d}t$$

- Rearrange the initial separable ODE to $\mathrm{d}y/\mathrm{d}t \cdot 1/g = f$ and invoke the law of composite differentiation to get

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{y_0}^{y(t)} \frac{\mathrm{d}w}{g(w)} - \int_{t_0}^{t} f(\tau)\,\mathrm{d}\tau\right] = 0$$

- It follows that

$$\int_{y_0}^{y(t)} \frac{\mathrm{d}w}{g(w)} = \int_{t_0}^{t} f(\tau)\,\mathrm{d}\tau$$

- Examples:
  1. Exponential growth.
     - We have that

       $$\frac{\mathrm{d}y}{\mathrm{d}t} = ky$$

       for $k > 0$ and $y(0) = y_0 > 0$.
     - The solution is

       $$\begin{aligned} \frac{1}{y}\cdot\frac{\mathrm{d}y}{\mathrm{d}t} &= k \\ \log y(t) - \log y_0 &= kt \\ y(t) &= y_0\mathrm{e}^{kt} \end{aligned}$$

[1] We'll deal with complex functions later.

2. Logistic growth.
   - We have that
   $$\frac{\mathrm{d}y}{\mathrm{d}t} = ky\left(1 - \frac{y}{M}\right)$$
   for $k, M > 0$ and $y(0) = y_0 > 0$.
   - The solution is
   $$\begin{aligned} \frac{M\,\mathrm{d}y}{y(M-y)} &= k\,\mathrm{d}t \\ \log\frac{y}{M-y} - \log\frac{y_0}{M-y_0} &= kt \\ \frac{y(M-y_0)}{y_0(M-y)} &= \mathrm{e}^{kt} \\ y\cdot\frac{M-y_0}{y_0} &= (M-y)\mathrm{e}^{kt} \\ y\cdot\frac{M-y_0}{y_0} + y\mathrm{e}^{kt} &= M\mathrm{e}^{kt} \\ y\left(\frac{M-y_0}{y_0} + \mathrm{e}^{kt}\right) &= M\mathrm{e}^{kt} \\ y\left(\frac{M-y_0+y_0\mathrm{e}^{kt}}{y_0}\right) &= M\mathrm{e}^{kt} \\ y\left(\frac{M+y_0(\mathrm{e}^{kt}-1)}{y_0}\right) &= M\mathrm{e}^{kt} \\ y(t) &= \frac{My_0\mathrm{e}^{kt}}{M+y_0(\mathrm{e}^{kt}-1)} \end{aligned}$$
   - Sketches the graph of logistic growth and discusses the turning point (for which there is a formula; zero of the second derivative) as well as general trends.
   - If $y_0 < 0$, the solution is not physically meaningful, but it is mathematically insightful.
     - When we integrate, the arguments of our logarithms now have absolute values.
     $$\log\left|\frac{y}{M-y}\right| - \log\left|\frac{y_0}{M-y_0}\right| = kt$$
     - We need to make sure that the denominator of the final logistic form is never equal to zero, but now that y_0 is negative, as t increases, the denominator will approach zero exponentially. It reaches zero when
     $$\begin{aligned} M + y_0(\mathrm{e}^{kt}-1) &= 0 \\ \mathrm{e}^{kt} &= -\frac{M}{y_0} + 1 \end{aligned}$$
     In other words, $t_{\max} = (1/k)\log(1 - M/y_0)$ because when $t = t_{\max}$, the equation blows up.
     - This is an example of **finite lifespan**.
   - If $y_0 > M$, then you will exponentially decrease to M.
3. Lotka-Volterra predator-prey model.
   - We have that
   $$r' = k_1 r - awr \qquad\qquad w' = -k_2 w + bwr$$
   where r is rabbits and w is wolves.

  - We can rename the variables to

$$\begin{cases} x' = Ax - Bxy \\ y' = -Cy + Dxy \end{cases}$$

  - Dividing, we get

$$\frac{x'}{y'} = \frac{Ax - Bxy}{-Cy + Dxy}$$
$$\frac{By - A}{y} y' + \frac{Dx - C}{x} x' = 0$$

  - Use the fact that x, y are independent variables, so both terms in the above equation are equal to zero??
  - Invoke the law of composite differentiation twice and, from the above, know that $0 + 0 = 0$, so we can add the two solutions:

$$\frac{\mathrm{d}}{\mathrm{d}t}(By(t) - A\log y(t)) + \frac{\mathrm{d}}{\mathrm{d}t}(Dx(t) - C\log x(t)) = 0$$
$$By(t) - A\log y(t) + Dx(t) - C\log x(t) = E$$

  - Sketches some of the trajectories (they're all closed curves in the xy-plane).

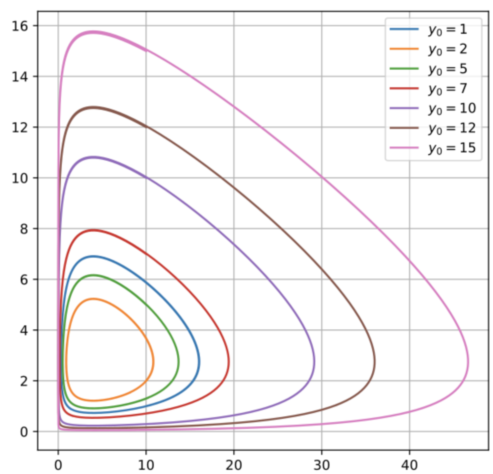


Figure 2.1: Lotka-Volterra solution curves.

  - Properties of the curves:
    - The implicit relation which determines them: By the implicit function theorem, the y derivative of the LHS is $B - A/y$ and the x-derivative of the LHS is $D - C/x$. When the partial derivatives are equal to zero, $(C/D, A/B)$ becomes interesting. Turning points happen when the y-coordinate is A/B or the x-coordinate is C/D.

- **Finite lifespan**: Even if the RHS of $\mathrm{d}y/\mathrm{d}t = f(t, y)$ is very regular, the solution can still blow up at some finite time.

- Consider the final ODE from the Brachistochrone problem.

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \sqrt{\frac{B - y}{y}}$$

  - Finding the **primitives**.

- What are these "primitives" Shao keeps talking about??
  - We should have
$$\int \sqrt{\frac{y}{B-y}}\,\mathrm{d}y = x$$
  - Change of variables: $y = B\sin^2\phi$ and $\mathrm{d}y = 2B\cos\phi\sin\phi\,\mathrm{d}\phi$. Thus,
$$\int \sqrt{\frac{y}{B-y}}\,\mathrm{d}y = \int \frac{\sin\phi}{\cos\phi}\cdot 2B\cos\phi\sin\phi\,\mathrm{d}\phi = 2B\int \sin^2\phi\,\mathrm{d}\phi$$
  - The solution is
$$\begin{cases} x = B\phi - \frac{B}{2}\sin(2\phi) + C \\ y = B\sin^2\phi \end{cases}$$
    - This is a parameterization of a cycloid.
- Later in the week, we will do the SHM, the pendulum, the Kepler 2-body problem, and the Michaelis-Menten equation.
- Separable ODEs are a subset of ODEs of **exact form**.
- ODEs of exact form are of the form
$$g(x,y)\frac{\mathrm{d}y}{\mathrm{d}x} + f(x,y) = 0$$
  where for some $F(x,y)$, $g = \partial F/\partial y$, $f = \partial F/\partial x$, and partials commute. Equivalently,
$$\frac{\partial g}{\partial x} = \frac{\partial f}{\partial y}$$
  is our necessary and sufficient condition.
- By the law of composite differentiation,
$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}x}\left[F(x,y(x))\right] &= \frac{\partial F}{\partial x} + \frac{\partial F}{\partial y}\cdot y'(x) \\ &= f(x,y(x)) + g(x,y(x))y'(x) \\ &= 0 \end{aligned}$$
  - We solve these with an integrating factor $\mu \neq 0$ such that $(\mu g, \mu f)$ satisfy the constraint.

2.2 Office Hours (Shao)

- **Primitive**: An antiderivative.
- **Law of composite differentiation**: The chain rule.
- Went over how Shao has been applying the law of composite differentiation with respect to separable ODEs:
  - Rearrange the initial separable ODE as follows.
$$\frac{1}{g(y)}\cdot\frac{\mathrm{d}y}{\mathrm{d}t} = f(t)$$

- Define $\mathrm{d}H/\mathrm{d}y = 1/g(y)$. Then, continuing from the above, we have by the law of composite differentiation that
$$\begin{aligned}\frac{\mathrm{d}H}{\mathrm{d}y}\cdot\frac{\mathrm{d}y}{\mathrm{d}t} &= f(t)\\ \frac{\mathrm{d}H}{\mathrm{d}t} &= f(t)\end{aligned}$$
- From the definition of H, we know that $H(y) = \int_{y_0}^{y} \mathrm{d}w\,/g(w)$. We also have from the FTC that $f(t) = \frac{\mathrm{d}}{\mathrm{d}t}\int_{t_0}^{t} f(\tau)\,\mathrm{d}\tau$. Thus, continuing from the above, we have that
$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}(H) &= f(t)\\ \frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{y_0}^{y}\frac{\mathrm{d}w}{g(w)}\right] &= \frac{\mathrm{d}}{\mathrm{d}t}\int_{t_0}^{t} f(\tau)\,\mathrm{d}\tau\\ \frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{y_0}^{y(t)}\frac{\mathrm{d}w}{g(w)} - \int_{t_0}^{t} f(\tau)\,\mathrm{d}\tau\right] &= 0\end{aligned}$$
as desired.
- It follows since $y(t_0) = y_0$ that $C = H(y_0) = 0$, so we can take the above to
$$\int_{y_0}^{y(t)}\frac{\mathrm{d}w}{g(w)} = \int_{t_0}^{t} f(\tau)\,\mathrm{d}\tau$$
knowing that our constant of integration is zero.

- Take away from Brachistochrone problem: Just an example of a BDE; we won't have to answer questions on it.

2.3 ODEs of Exact Form

10/5:

- Last time, we discussed separable ODEs.
- Today, we will study **exact form** equations, as discussed last class.
- **Exact form** (ODE): An ODE of the form
$$g(x,y)\frac{\mathrm{d}y}{\mathrm{d}x} + f(x,y) = 0$$
where
$$\frac{\partial g}{\partial x} = \frac{\partial f}{\partial y}$$
- For equations of this form, there exists $F(x,y)$ such that
$$\frac{\partial F}{\partial x} = f \qquad \frac{\partial F}{\partial y} = g \qquad F(x,y(x)) = C$$
for some $C \in \mathbb{R}$.
- To solve equations of this form, we need an **integrating factor**.
- **Integrating factor**: A number or function μ such that
$$\mu g\frac{\mathrm{d}y}{\mathrm{d}x} + \mu f = 0 \qquad \frac{\partial}{\partial x}(\mu g) = \frac{\partial}{\partial y}(\mu f)$$

- The solution to linear homogeneous equations of the form $\mathrm{d}y/\mathrm{d}t = p(t)y$ is

$$y(t) = y_0 \exp\left[\int_{t_0}^{t} p(\tau)\mathrm{d}\tau\right]$$

- Recall that $\mathrm{e}^{a+ib} = \mathrm{e}^a(\cos b + i\sin b)$, so

$$\mathrm{e}^{ix} = \cos x + i\sin x \qquad \cos x = \frac{1}{2}(\mathrm{e}^{ix} + \mathrm{e}^{-ix}) \qquad \sin x = \frac{1}{2i}(\mathrm{e}^{ix} - \mathrm{e}^{-ix})$$

- Example: If $y' = ky$, then $y' = -\lambda y$.
- If we have an inhomogeneous linear equation $\mathrm{d}y/\mathrm{d}t = p(t)y + f(t)$, then

$$\frac{\mathrm{d}y}{\mathrm{d}t} - py - f = 0$$

but

$$0 = \frac{\mathrm{d}}{\mathrm{d}t}(1) \neq \frac{\mathrm{d}}{\mathrm{d}y}(-p(t)y - f(t))$$

- We wish to find an integrating factor $\mu(t, y)$ such that

$$\mu(t,y)\frac{\mathrm{d}y}{\mathrm{d}t} - \mu(t,y)p(t)y - \mu(t,y)f(t) = 0$$

and

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mu) = \frac{\mathrm{d}}{\mathrm{d}y}(-\mu p y - \mu f)$$

- Solution: Take μ to be a function of t, alone. Then

$$\mu'(t) = \frac{\mathrm{d}}{\mathrm{d}y}(-\mu p y - \mu f) = -\mu(t)p(t)$$

and we now have a homogeneous linear equation with solution

$$\mu(t) = \exp\left[-\int_{t_0}^{t} p(\tau)\mathrm{d}\tau\right]$$

  - If we let $P(t) = \int_{t_0}^{t} p(\tau)\mathrm{d}\tau$, then

$$\begin{aligned} \mathrm{e}^{-P(t)}y'(t) - p(t)y(t)\mathrm{e}^{-P(t)} &= \mathrm{e}^{-P(t)}f(t) \\ \frac{\mathrm{d}}{\mathrm{d}t}\left(\mathrm{e}^{-P(t)}y(t)\right) &= \mathrm{e}^{-P(t)}f(t) \\ \mathrm{e}^{-P(t)}y(t) &= \int_{t_0}^{t} \mathrm{e}^{-P(\tau)}f(\tau)\mathrm{d}\tau \end{aligned}$$

  - Thus, we finally have the solution to the inhomogeneous problem as follows: The IVP $y' = py + f$, $y(t_0) = y_0$ has solution

$$y(t) = y_0\mathrm{e}^{P(t)-P(t_0)} + \mathrm{e}^{P(t)}\int_0^t \mathrm{e}^{-P(\tau)}f(\tau)\mathrm{d}\tau$$

  where P is any anti-derivative of p.

- In particular, when $p(t) = a$, we get the **Duhamel formula** (which we should memorize).

- **Duhamel formula**: The following equation, which is the solution to an inhomogeneous linear equation with $p(t)=a$.

$$y(t)=y_0\mathrm{e}^{a(t-t_0)}+\int_{t_0}^{t}\mathrm{e}^{a(t-\tau)}f(\tau)\mathrm{d}\tau$$

  - Important for computing forced oscillation.
- Inspecting the inhomogeneous solution.
  - The first term is the solution to the homogeneous form. The second term deals with the initial value.
- Given an inhomogeneous equation, you can always write its solution as the combination of the solution to the homogeneous problem plus a particular solution, i.e.,

$$y=y_h+y_p$$

  - "The general solution equals the homogeneous solution plus a particular solution."
  - This is related to linear algebra, where the solution to $Ax=b$ is a particular solution x_p plus any vector $x\in\ker A$.
  - Thus, this idea will reappear in the theory of systems of linear ODEs.
- We now look at systems of linear ODEs.
- Consider the harmonic oscillator: A particle of mass m connected to an ideal spring (obeys Hooke's law) with no friction or gravity.
  - Newton's second law: The acceleration is proportional to the restoring force.
  - Hooke's law: The restoring force is of magnitude kx in the opposite direction to the displacement.
  - Thus, the ODE is of the form

$$x''=-\frac{k}{m}x$$

  - However, if there is damping (which will be proportional to the velocity), then the ODE is of the form

$$x''+\frac{b}{m}x'+\frac{k}{m}x=0$$

- Consider an ODE of the form

$$y''+ay'+by=0$$

  for $a,b\in\mathbb{C}$.
  - Aim: Find $\mu,\lambda\in\mathbb{C}$ such that

$$(y'-\mu y)'-\lambda(y'-\mu y)=0$$

  - To find the parameters, we expand the above to

$$y''-(\mu+\lambda)y'+\mu\lambda y=0$$

  - Comparing with the original form, we have that $a=-(\mu+\lambda)$ and $b=\mu\lambda$.
  - It follows that μ,λ are the roots of $x^2+ax+b=0$, which we will call the **characteristic polynomial** of the ODE.
  - Substitute $x=y'-\mu y$. Then we can solve

$$x'-\lambda x=0$$

    to determine that $x=A\mathrm{e}^{\lambda t}$.

- Returning the substitution, we have that

$$y' - \mu y = A\mathrm{e}^{\lambda t}$$

- Since the above is of the form $y' = ay + f$, we can apply the Duhamel formula. It follows that a particular solution is

$$A\int_0^t \mathrm{e}^{\mu(t-\tau)}\mathrm{e}^{\lambda\tau}\mathrm{d}\tau$$

- Thus, general solutions are of the form

$$y(t) = B\mathrm{e}^{\mu t} + C\mathrm{e}^{\mu t}\int_0^t \mathrm{e}^{(\lambda-\mu)\tau}\mathrm{d}\tau$$

- Evaluating the integral, we get

$$y(t) = B\mathrm{e}^{\mu t} + C\mathrm{e}^{\mu t}\frac{\mathrm{e}^{(\lambda-\mu)t}-1}{\lambda-\mu}$$

  which simplifies (by incorporating constants, etc.) to

$$y(t) = A_1\mathrm{e}^{\mu t} + B_1\mathrm{e}^{\lambda t}$$

  for $\mu \neq \lambda$, or

$$y(t) = A_1\mathrm{e}^{\mu t} + B_1 t\mathrm{e}^{\mu t}$$

  for $\mu = \lambda$.
- These linearly independent solutions form a basis of the space of solutions; all solutions can be expressed as a linear combination of these two functions.

- If our equation is of the form $y'' + ay' + by = f(t)$, then we just need to apply the Duhamel formula twice.
- Returning to the simple harmonic oscillator problem, we substitute $\omega = \sqrt{k/m}$ to get

$$x'' = \omega^2 x$$

  - The characteristic polynomial is

$$0 = x^2 + \omega^2 = (x+i\omega)(x-i\omega)$$

  - Thus, solutions are of the form

$$x = A_1\mathrm{e}^{i\omega t} + B_1\mathrm{e}^{-i\omega t}$$

  - It follows that the period is $T = 2\pi/\omega$.
  - To get a real (usable) solution, apply Euler's formula to get

$$\begin{aligned} x(t) &= A_1(\cos\omega t + i\sin\omega t) + B_1(\cos\omega t - i\sin\omega t) \\ &= A\cos\omega t + B\sin\omega t \end{aligned}$$

    where $A = A_1 + B_1$, $B = iA_1 - iB_1$.
  - To match the initial condition $x(0) = x_0$, $x'(0) = v_0$, we use

$$x(t) = x_0\cos\omega t + \frac{v_0}{\omega}\sin\omega t$$

  - In other words,

$$\begin{cases} A = x_0 \\ \omega B = v_0 \end{cases} \qquad \begin{cases} A_1 + B_1 = x_0 \\ i\omega A_1 - i\omega B_1 = v_0 \end{cases}$$

    so

$$\begin{cases} A = x_0 \\ B = \frac{v_0}{\omega} \end{cases} \qquad \begin{cases} A_1 = \frac{1}{2}\left[x_0 - \frac{iv_0}{\omega}\right] \\ B_1 = \frac{1}{2}\left[x_0 + \frac{iv_0}{\omega}\right] \end{cases}$$

2.4 ODE Examples

10/7:

- Today, we will investigate a variety of examples of ODEs arising in real life.
- Michaelis-Menten kinetics: If E is an enzyme, S is its substrate, and P is the product, then the mechanism is

$$\mathrm{E} + \mathrm{S} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{ES} \xrightarrow{k_2} \mathrm{E} + \mathrm{P}$$

- The concentrations that we are concerned with are $[\mathrm{E}], [\mathrm{S}], [\mathrm{ES}], [\mathrm{P}]$.
- From the above mechanism, we can write the four rate laws

$$\frac{\mathrm{d}}{\mathrm{d}t}[\mathrm{S}] = -k_1[\mathrm{E}][\mathrm{S}] + k_{-1}[\mathrm{ES}] \tag{1}$$
$$\frac{\mathrm{d}}{\mathrm{d}t}[\mathrm{E}] = -k_1[\mathrm{E}][\mathrm{S}] + (k_{-1} + k_2)[\mathrm{ES}] \tag{2}$$
$$\frac{\mathrm{d}}{\mathrm{d}t}[\mathrm{ES}] = k_1[\mathrm{E}][\mathrm{S}] - (k_{-1} + k_2)[\mathrm{ES}] \tag{3}$$
$$\frac{\mathrm{d}}{\mathrm{d}t}[\mathrm{P}] = k_2[\mathrm{ES}] \tag{4}$$

- The initial conditions are $[\mathrm{S}] = [\mathrm{S}]_0$ and $[\mathrm{E}] = [\mathrm{E}]_0$.
- We can reduce these rate laws to the 2D system

$$\frac{\mathrm{d}}{\mathrm{d}t}[\mathrm{S}] = -k_1([\mathrm{E}]_0 - [\mathrm{ES}])[\mathrm{S}] + k_{-1}[\mathrm{ES}] \tag{5}$$
$$\frac{\mathrm{d}}{\mathrm{d}t}[\mathrm{ES}] = k_1([\mathrm{E}]_0 - [\mathrm{ES}])[\mathrm{S}] - (k_{-1} + k_2)[\mathrm{ES}] \tag{6}$$

  - Note that to do so, we have used two conservation laws: The conservation of the enzyme plus enzyme-substrate complex, and the conservation of the substrate plus enzyme-substrate complex plus products.
- QSSA: Quasi steady-state assumption.
  - Assume that $[\mathrm{E}]_0/[\mathrm{S}]_0 \lll 1$.
  - It follows that $\mathrm{d}[\mathrm{ES}]/\mathrm{d}t \approx 0$ due to saturation of the enzyme and $[\mathrm{S}] \approx [\mathrm{S}]_0$ due to ever-more substrate being available.
- Then

$$[\mathrm{ES}] = \frac{[\mathrm{E}]_0[\mathrm{S}]}{K_M + [\mathrm{S}]}$$

  where $k_M = (k_{-1} + k_2)/k_1$ is the **Michaelis-Menten constant**, a usual indication of enzyme activity.
- Substitute the above into Equation 5:

$$\frac{\mathrm{d}}{\mathrm{d}t}[\mathrm{S}] = -\frac{v_{\max}[\mathrm{S}]}{k_M + [\mathrm{S}]}$$

  - Note that $v_{\max} = k_2[\mathrm{E}]_0$.
- The above is now a differential equation of separable form; it's solution is

$$\int_{[\mathrm{S}]_0}^{[\mathrm{S}]} -\frac{(k_M + z)\,\mathrm{d}z}{zv_{\max}} = \int_0^t \mathrm{d}t$$
$$-\frac{k_M}{v_{\max}} \log\frac{[\mathrm{S}]}{[\mathrm{S}]_0} - \frac{1}{v_{\max}}([\mathrm{S}] - [\mathrm{S}]_0) = t$$

$$
\begin{aligned}
\log\frac{[\mathrm{S}]}{[\mathrm{S}]_0}+\frac{[\mathrm{S}]}{k_M} &= \frac{[\mathrm{S}]_0-v_{\max}t}{k_M}\\
\frac{[\mathrm{S}]}{[\mathrm{S}]_0}\mathrm{e}^{[\mathrm{S}]/k_M} &= \exp\left(\frac{[\mathrm{S}]_0-v_{\max}t}{k_M}\right)\\
\frac{[\mathrm{S}]}{k_M}\mathrm{e}^{[\mathrm{S}]/k_M} &= \frac{[\mathrm{S}]_0}{k_M}\exp\left(\frac{[\mathrm{S}]_0-v_{\max}t}{k_M}\right)\\
\frac{[\mathrm{S}]}{k_M} &= W\left[\frac{[\mathrm{S}]_0}{k_M}\exp\left(\frac{[\mathrm{S}]_0-v_{\max}t}{k_M}\right)\right]\\
[\mathrm{S}] &= k_M W\left[\frac{[\mathrm{S}]_0}{k_M}\exp\left(\frac{[\mathrm{S}]_0-v_{\max}t}{k_M}\right)\right]
\end{aligned}
$$

  - Getting from line 5-6 (i.e., the introduction of W): Suppose we have an equation of the form $y\mathrm{e}^y = x$. We cannot express x in terms of y using elementary functions, so we must define W such that $y = W(x)$. Explicitly, W is the unique function of x that satisfies $W(x)\mathrm{e}^{W(x)} = x$.

- Harmonic oscillator.
- Recall that

$$x'' + \frac{k}{m}x = 0$$

- Substituting $\omega = \sqrt{k/m}$, we can solve the above for

$$x(t) = x(0)\cos(\omega t) + \frac{x'(0)}{\omega}\sin(\omega t)$$

- This is an integrable system with n degrees of freedom and $n-1$ scalar conservation laws??
- Conservation of mechanical energy:

$$E = \frac{1}{2}m|x'|^2 + \frac{1}{2}kx^2$$

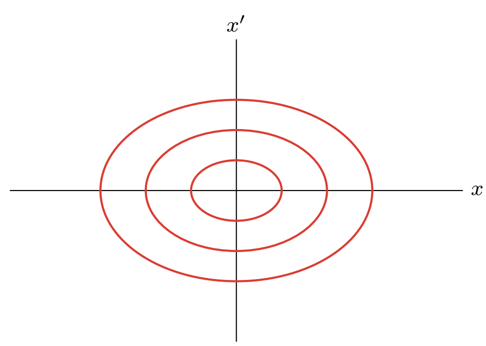


Figure 2.2: Conservation of mechanical energy in the harmonic oscillator.

  - Differentiating wrt. x yields

$$
\begin{aligned}
0 &= mx'x'' + kxx'\\
&= \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{2}m(x')^2\right) + \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{2}kx^2\right)
\end{aligned}
$$

  - This means that the solution is an ellipse in the xx'-plane, where each ellipse corresponds to an initial displacement and velocity.

- Mathematical pendulum.
- Equation of motion:

$$\begin{aligned}0 &= l\theta'' + g\sin\theta\\ &= \ell\theta''\theta' + g\sin\theta\cdot\theta'\\ &= \frac{\mathrm{d}}{\mathrm{d}t}\bigg(\underbrace{\frac{\ell}{2}|\theta'|^2 - g\cos\theta}_{E}\bigg)\end{aligned}$$

- Initial values:

$$\theta(0) = \theta_0 \qquad\qquad \theta'(0) = 0$$

- It follows from the above that

$$\begin{aligned}\frac{\ell}{2}|\theta'|^2 - g\cos\theta_0 &= -g\cos\theta\\ \frac{\mathrm{d}\theta}{\mathrm{d}t} &= \sqrt{\frac{2g}{\ell}(\cos\theta_0 - \cos\theta)}\\ \int_{\theta_0}^{\theta}\sqrt{\frac{\ell}{2g(\cos\theta_0 - \cos\phi)}}\mathrm{d}\phi &= t\end{aligned}$$

  - This is an elliptical integral (and thus cannot be expressed in terms of the elementary functions).
- Suppose θ_0 is small. Then θ is small, and we can invoke the small angle approximation $\sin\theta \approx \theta$.
  - This yields an approximate equation of motion:

$$\ell\theta'' + g\theta = 0$$

  - From here, we can determine that $\theta(t) \approx \theta_0\cos\sqrt{g/\ell}\cdot t$ and $T = 2\pi\sqrt{\ell/g}$.
- Kepler problem.
- Two bodies of mass m_1, m_2 are located at positions x_1, x_2 pulling on each other gravitationally.
  - The force of attraction is a conservative central force.
  - The potential between the two masses is a function of their distance, i.e, $U(|x_1 - x_2|)$.
- From Newton's second and third law, we get

$$m_1x_1'' = U'(|x_1 - x_2|)\frac{x_2 - x_1}{|x_2 - x_1|} \qquad\qquad m_2x_2'' = U'(|x_1 - x_2|)\frac{x_1 - x_2}{|x_1 - x_2|}$$

  - The derivative of potential is force.
  - The vector term provides direction.
- Conservation of momentum:

$$\begin{aligned}(m_1x_1 + m_2x_2)'' &= 0\\ m_1x_1' + m_2x_2' &= C\end{aligned}$$

  - Let $M = m_1 + m_2$. Then the center of mass

$$\frac{m_1}{M}x_1 + \frac{m_2}{M}x_2$$

    moves inertially (i.e., does not accelerate or decelerate; is a stable reference frame) — we'll define it to be the origin.

- Conservation of angular momentum:

$$[m(x_1 - x_2)' \times (x_1 - x_2)]' = 0$$

  - $m = m_1 m_2/(m_1 + m_2)$.
  - $\times$ indicates the cross product.
  - $L = m(x_1 - x_2)' \times (x_1 - x_2)$.
- It follows that $x_1 - x_2$ is always in a fixed plane, which we may call the **horizon plane**.
- Conservation of mechanical energy:

$$mq'' + U'(|q|)\frac{q}{|q|} = 0$$
$$\frac{m}{2}|q'|^2 + U(|q|) = E$$

  - $q = x_1 - x_2$.
- Introduce polar coordinates (r, ϕ).
  - Then $r^2\phi' = \ell_0$, $r = r(\phi)$, and $\mathrm{d}r/\mathrm{d}\phi = r'(t)/\phi'(t)$.
  - It follows that
$$\frac{m}{2}(|r'|^2 + |\phi'|^2) + U(r) = E$$
  - Since $U(r) = -Gm_1m_2/r$ for Newtonian gravity,
$$\left(\frac{\mathrm{d}r}{\mathrm{d}\phi}\right)^2 + r^2 = \frac{2GMr^3}{\ell_0^2} + \frac{2Er^4}{m\ell_0^2}$$
  - The substitution $\mu = 1/r$ yields
$$\left(\frac{\mathrm{d}\mu}{\mathrm{d}\phi}\right)^2 + \mu^2 = \frac{2GM}{\ell_0^2}\mu + \frac{2E}{m\ell_0^2}$$
  - Differentiating again gives
$$2\frac{\mathrm{d}\mu}{\mathrm{d}\phi}\frac{\mathrm{d}^2\mu}{\mathrm{d}\phi^2} + 2\mu\frac{\mathrm{d}\mu}{\mathrm{d}\phi} = \frac{2GM}{\ell_0^2}\frac{\mathrm{d}\mu}{\mathrm{d}\phi}$$
  - Substituting $\mu = \cos(t)$ gives
$$\frac{\mathrm{d}^2\mu}{\mathrm{d}\phi^2} + \mu - \frac{GM}{\ell_0^2} = 0$$
    or
$$r = \frac{1}{GM/\ell_0^2 + \varepsilon\cos(\phi - \phi_0)}$$
    - This is a conic section!
- Thus, for example, we can calculate the precession of Mercury.
- Note that while we have determined the trajectory of our 2 bodies in terms of elementary functions, the n-body problem cannot be solved analytically.

Week 3

Linear Algebra Review

3.1 Elements of Linear Algebra

10/10:
- Today: Review of linear algebra.
- Start with a **vector space** over $\mathbb{R}$ or $\mathbb{C}$ or, more generally, any field K.
- **Vector space** (over K): A set equipped with addition and scalar multiplication such that the following axioms are satisfied.
  1. Commutativity and associativity of addition.
  2. Additive identity and inverse.
  3. Compatibility of scalar multiplication and addition (distributive laws).
  4. The additive identity times any vector is zero.
- In $\mathbb{R}^n, \mathbb{C}^n$, addition is component-wise and scalar multiplication is scaling of the element.
- For a homogeneous equation

$$y' = A(t)y = \begin{pmatrix} a_{11}(t)y^1 + a_{12}(t)y^2 + \cdots \\ \vdots \end{pmatrix} \begin{pmatrix} y^1 \\ y^2 \\ \vdots \\ y^n \end{pmatrix}$$

  - If y_1, y_2 are solutions, any linear combination of them is a solution. This is called the **solution space** of the equation.
- **Linearly independent** (set of vectors): A set of vectors $x_1, \dots, x_m \in V$ for which the only coefficients $\lambda_1, \dots, \lambda_m$ such that

$$\lambda_1 x_1 + \cdots + \lambda_m x_m = 0$$

  is $\lambda_1 = \cdots = \lambda_m = 0$.
  - $\lambda_m \neq 0$ implies

$$x_m = -\frac{1}{\lambda_m}(\lambda_1 x_1 + \cdots + \lambda_{m-1} x_{m-1})$$

- **Maximal linear independence group**: A subset $X \subset V$ such that for any $y \in V$, $\{y\} \cup X$ is not linearly independent. *Also known as* **basis**.
- Theorem: Any basis in V has the same cardinality.
- **Dimension** (of V): The cardinality given by the above theorem. *Denoted by* $\mathbf{dim}\,\boldsymbol{V}$.

- We usually denoted a basis as an ordered n-tuple since the order often matters (for orientation?).
- Notational conventions.
  - For $\mathbb{R}^n, \mathbb{C}^n$, we will always use column vectors.
  - $x_1, x_2, \dots$ denotes vectors.
  - $x^1, x^2, \dots$ denotes the components of a column vector.
  - A vector component squared may be denoted $(x^1)^2$.
- **Standard basis** (for $\mathbb{R}^n$): The set of n vectors of length n which have a 1 as one entry and a zero in all the others and are all distinct.
- **Linear transformation** (of V to V): A mapping $\phi : V \to V$ satisfying

$$\phi(\lambda x + \mu y) = \lambda\phi(x) + \mu\phi(y)$$

- A mapping is completely determined by its action on the basis vectors:

$$\phi\left(\sum_{k=1}^n x^k e_k\right) = \sum_{k=1}^n x^k \phi(e_k)$$

- **Matrix** (of a linear transformation wrt. the standard basis): The $n \times n$ array

$$\begin{pmatrix} \phi(e_1) & \cdots & \phi(e_n) \end{pmatrix}$$

- If $\phi, \psi : V \to V$ are linear, $\phi \circ \psi$ is also linear.
  - Composition of linear transformations corresponds to matrix multiplication.
- Matrix multiplication: If

$$B = \begin{pmatrix} b_1 & \cdots & b_n \end{pmatrix}$$

  then

$$AB = \begin{pmatrix} Ab_1 & \cdots & Ab_n \end{pmatrix}$$

  where

$$Ax = \begin{pmatrix} a_{11}x^1 + \cdots + a_{1n}x^n \\ \vdots \\ a_{n1}x^1 + \cdots + a_{nn}x^n \end{pmatrix}$$

- We can talk about matrix inverses: If it exists, it is unique, and

$$AA^{-1} = A^{-1}A = I_n$$

- Matrix multiplication is not commutative in general. Shao gives a counterexample.
- A is invertible iff the columns of A are a basis for $\mathbb{R}^n$ (resp. $\mathbb{C}^n$).
- **Determinant** (of A): Not explicitly defined.
- Properties of the determinant.
  - Multilinear.

$$\det\begin{pmatrix} a_1 & \cdots & \lambda a_k + \mu\tilde{a}_k & \cdots & a_n \end{pmatrix} = \lambda\det\begin{pmatrix} a_1 & \cdots & a_k & \cdots & a_n \end{pmatrix} + \mu\det\begin{pmatrix} a_1 & \cdots & \tilde{a}_k & \cdots & a_n \end{pmatrix}$$

  - Skew-symmetric.

$$\det\begin{pmatrix} a_1 & \cdots & a_i & \cdots & a_j & \cdots & a_n \end{pmatrix} = -\det\begin{pmatrix} a_1 & \cdots & a_j & \cdots & a_i & \cdots & a_n \end{pmatrix}$$

- Theorem: The determinant is uniquely characterized by these two axioms.
- $\det I_n = 1$.
- Shao goes over computing the determinant via minors.
- Special cases:
  - If the matrix is upper- or lower-triangular, the determinant is equal to the product of the diagonal entries.
  - If the matrix is blocked upper- or lower-triangular, e.g.,

$$A = \begin{pmatrix} A_1 & * \\ 0 & A_2 \end{pmatrix}$$

    then $\det A = \det A_1 \cdot \det A_2$.
- $\det(AB) = \det(A)\det(B)$.
- $\det A \neq 0$ iff A is invertible.
- Direct formula to compute the inverse.

$$A^{-1} = \frac{1}{\det A}\left(a_{\ell k}(-1)^{k+\ell}\det A_{k\ell}\right)$$

  - Tedious for higher-dimensional cases, but quite sufficient for $n = 2, 3$.
- Let A be $n \times n$, and let $Ax = b$.
  - If A is invertible, then $x = A^{-1}b$.
  - If A is not invertible and $b \in \operatorname{span}(a_1, \dots, a_n)$, then $x = x_h + x_p$ where $Ax_h = 0$ and $Ax_p = b$.
- **Kernel** (of A): The set of all vectors $y \in \mathbb{R}^n$ (resp. $\mathbb{C}^n$) such that $Ay = 0$.
- **Range** (of A): The set of all linear combinations of $a_1, \dots, a_n$.
- Suppose $\phi : \mathbb{R}^n \to \mathbb{R}^n$ has matrix A under $(e_1, \dots, e_n)$. Let $(q_1, \dots, q_n)$ be another basis.
  - There exists a matrix Q such that $q_k = Qe_k$. Q is called the **connecting matrix** between $(e_1, \dots, e_n)$ and $(q_1, \dots, q_n)$.
  - Claim: Let $x \in \mathbb{R}^n$ have representation $x = (x^1, \dots, x^n)$ under the standard basis. Then under the Q basis, x has representation $x' = Q^{-1}(x^1, \dots, x^n)$. Similarly, $x = Qx'$.
  - Claim: ϕ has matrix $B = Q^{-1}AQ$ with respect to the Q basis.
- Matrix similarity: $A \sim B$ iff there exists Q invertible such that $B = Q^{-1}AQ$.
  - Implies that A and B describe the same matrix under different bases.
  - Matrix product under the old and new bases are related.

$$Q^{-1}ABQ = (Q^{-1}AQ)(Q^{-1}BQ)$$

  - Similarity preserves the determinant:

$$\det\left(Q^{-1}AQ\right) = \det\left(Q^{-1}\right)\det(A)\det(Q) = \det(A)\det\left(Q^{-1}\right)\det(Q) = \det(A)$$

- There is an extra example in Shao's notes (of a linear transformation in two bases).

3.2 Diagonalization and Jordan Normal Form

10/12:

- Similar matrices and Jordan Normal Form (JNF).
- Suppose $A : \mathbb{C}^n \to \mathbb{C}^n$ is linear. We can express A in a different basis with the help of the connecting matrix Q.
- In this lecture, we seek to find the most convenient basis in which to discuss our linear transformation.
- Today we will work in $\mathbb{C}^n$ (but all results hold for $\mathbb{R}^n$, too).
- **Invariant subspace** (of A): A subspace $K \subset \mathbb{C}^n$ such that $A(K) = K$.
- Suppose you have m invariant subspaces $K_1, \dots, K_m \subset \mathbb{C}^n$ whose pairwise intersection is $\{0\}$.
- **Direct sum** (of $K_1, \dots, K_m$): The collection of all vectors which can be represented as sums from each of the subspaces. *Denoted by* $\boldsymbol{K_1 \oplus \cdots \oplus K_m}$. *Given by*

$$K_1 \oplus \cdots \oplus K_m = \left\{ x \in \mathbb{C}^n \mid x = \sum_{j=1}^{m} x_j,\ x_j \in K_j \right\}$$

- Suppose $K_1, K_2 \in \mathbb{C}^n$ are invariant subspaces of A of dimension n_1, n_2, respectively, such that $K_1 \oplus K_2 = \mathbb{C}^n$. Then choosing a basis for K_1 and K_2, the matrix A takes the form

$$\begin{pmatrix} B_1 & 0 \\ 0 & B_2 \end{pmatrix}$$

  where B_1 is an $n_1 \times n_1$ block and B_2 is an $n_2 \times n_2$ block.
- **Eigenvalue** (of A): A complex number $\lambda \in \mathbb{C}$ such that $A - \lambda I$ is not invertible. *Denoted by* $\boldsymbol{\lambda}$.
  - Equivalently, $\det(A - \lambda I) = 0$.
- **Characteristic polynomial**: The polynomial in z defined as follows. *Denoted by* $\boldsymbol{\chi_A(z)}$. *Given by*

$$\chi_A(z) = \det(A - zI)$$

  - Similar matrices have the same characteristic polynomials.
- **Spectrum** (of A): The set of all eigenvalues of A.
- **Eigenvector** (of A): A vector $v \in \mathbb{C}^n$ corresponding to an eigenvalue λ via

$$Av = \lambda v$$

- Claim: The set of all eigenvectors corresponding to λ form an invariant subspace.

  Proof.

$$A(v_1 + v_2) = \lambda v_1 + \lambda v_2 = \lambda(v_1 + v_2)$$

  □
- **Eigenspace** (of λ): The vector subspace of $\mathbb{C}^n$ equal to the span of the eigenvectors of λ. *Denoted by* $\boldsymbol{V_\lambda}$.
- **Algebraic multiplicity** (of λ): The degree of the $(z-\lambda)$ term in the factorization of the characteristic polynomial. *Denoted by* $\boldsymbol{\alpha_\lambda}$.
- **Geometric multiplicity** (of λ): The dimension of the eigenspace of λ. *Denoted by* $\boldsymbol{\gamma_\lambda}$.

- $\gamma_\lambda \leq \alpha_\lambda$.
- If $\alpha_\lambda = \gamma_\lambda$ for each λ, then each eigenspace V_λ has a basis such that $\oplus_\lambda V_\lambda = \mathbb{C}^n$.
  - Under this basis, the matrix of A is diagonal with all λ's (along the diagonal) repeated according to their algebraic multiplicity.
- **Superdiagonal**: The set of entries in a matrix which are directly above a diagonal entry.
- **Jordan block**: A $d \times d$ matrix corresponding to an eigenvalue λ that has λ as every diagonal entry, 1 as every superdiagonal entry, and zeroes everywhere else. *Denoted by* $\boldsymbol{J_d(\lambda)}$.
  - A Jordan block is an example of a matrix with algebraic multiplicity d and geometric multiplicity 1.
  - The geometric multiplicity γ_j is the number of Jordan blocks with eigenvalue λ_j. Of course, when $\gamma_j = \alpha_j$ (in particular, if $\alpha_j = 1$), there is no Jordan block corresponding to λ_j at all.
- For any linear transformation, we can find a basis such that the matrix is the diagonalized Jordon blocks.
- Theorem: Let A be an $n \times n$ complex matrix. Then there is a **Jordan basis** Q under which

$$Q^{-1}AQ = \begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_k \end{pmatrix}$$

  i.e., under which the matrix of $Q^{-1}AQ$ is block-diagonalized Jordan blocks.
  - The proof will not be tested — it is very hard. Shao will sketch it, though.
  - The proof is constructive: It will tell you how to convert a matrix into the Jordan normal form.
- Proof procedure:
  1. Determine the eigenvalues as well as their algebraic and geometric multiplicities.
     (a) Compute $\chi_A(z)$.
     (b) Find $\lambda_1, \dots, \lambda_m$ (factor $\chi_A(z)$).
     (c) Find $\alpha_1, \dots, \alpha_m$ (combine like terms in the factorization of $\chi_A(z)$).
     (d) Find $\gamma_1, \dots, \gamma_m$ ($\gamma_i = n - \operatorname{rank}(A - \lambda_i I)$).
  2. Find the **generalized eigenspaces** of each λ_i. This will allow us to block-diagonalize A.
     (a) For each λ_i, compute the $\ker(A - \lambda_i I) \subset \ker(A - \lambda_i I)^2 \subset \ker(A - \lambda_i I)^3 \subset \cdots$.
     (b) The sequence will stop at some $d_i \in \mathbb{N}$. In particular, it will stop when $\dim \ker(A - \lambda I)^{d_j} = \alpha_i$.
        - Claim: $\mathbb{C}^n = K_1 \oplus \cdots \oplus K_m$.
     (c) Since each K_i is an invariant subspace of A, we know that there is a matrix of the linear transformation corresponding to A of the form

$$\begin{pmatrix} B_1 & & 0 \\ & \ddots & \\ 0 & & B_m \end{pmatrix}$$

     We now just need to choose the *best* basis of each K_i, i.e., the one that makes each B_i into a (direct sum of) Jordan block(s).
  3. Find the best basis for each K_i.

(a) Recall that each λ_i corresponds to $\gamma = \gamma_i$ linearly independent eigenvectors, which we will denote $v_{i,1}, \dots, v_{i,\gamma}$. We will block-diagonalize B_i into γ Jordan blocks, each of which corresponds to a $v_{i,j}$ as follows.
Every Jordan block is of the form

$$\begin{pmatrix} \lambda_i & 1 & & 0 \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_i \end{pmatrix}$$

Let the above block be of dimension $k_{i,j} = d$. It follows that this block will be responsible for linearly transforming d vectors in the Jordan basis. Let $v_{i,j,1} = v_{i,j}$ be the first of these d vectors. Then the submatrix of $v_{i,j,1}$ in the Jordan basis corresponding to this Jordan block is

$$\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

which should make sense since we want $Av_{i,j} = \lambda_i v_{i,j}$ and under this definition,

$$\begin{pmatrix} \lambda_i & 1 & & 0 \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_i \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \lambda_i \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Now let $v_{i,j,2}$ be the second of the d vectors. Naturally, its submatrix in the Jordan basis should be

$$\begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}$$

But this implies that

$$\begin{aligned} \begin{pmatrix} \lambda_i & 1 & & 0 \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_i \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} &= \begin{pmatrix} 1 \\ \lambda_i \\ \vdots \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \lambda_i \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} \\ Jv_{i,j,2} &= v_{i,j,1} + \lambda_i v_{i,j,2} \\ (J - \lambda_i I)v_{i,j,2} &= v_{i,j,1} \end{aligned}$$

Naturally, this process will generalize to show that $(J - \lambda_i I)v_{i,j,k} = v_{i,j,k-1}$, i.e., we can recursively determine the $v_{i,j,1}, \dots, v_{i,j,k_{i,j}}$.

(b) However, there is slightly more subtlety than we might guess at first glance. Indeed, of our γ eigenvectors corresponding to λ_i, pick the first γ' to be elements of $\ker(A - \lambda_i I) \cap \operatorname{im}(A - \lambda_i I)$. This is a necessary condition for the existence of $v_{i,j,2}$ such that $(A - \lambda_i I)v_{i,j,2} = v_{i,j,1}$ for $j = 1, \dots, \gamma'$.

(c) Thus, using the above process, we will find $k_{i,j}$ elements of the Jordan basis for each $v_{i,j}$. The full, ordered set of these vectors, listed as follows, constitutes the Jordan basis.

$$\begin{array}{llll} v_{i,1,1}, & v_{i,1,2}, & \cdots, & v_{i,1,k_{i,1}} \\ v_{i,2,1}, & v_{i,2,2}, & \cdots, & v_{i,2,k_{i,2}} \\ \cdots & & & \\ v_{i,\gamma',1}, & v_{i,\gamma',2}, & \cdots, & v_{i,\gamma',k_{i,\gamma'}} \\ v_{i,\gamma'+1}, & v_{i,\gamma'+2}, & \cdots, & v_\gamma \end{array}$$

(d) Note that each of these vectors is naturally an element of the generalized eigenspace K_i since for each $k = 1, \dots, k_{i,j}$, the formula $(A - \lambda_i I)v_{i,j,k} = v_{i,j,k-1}$ implies that

$$(A - \lambda_i I)^k v_{i,j,k} = 0$$

Also note that each $k_{i,j} \leq d_i$ and $k_{i,1} + \cdots + k_{i,\gamma'} + \gamma - \gamma' = \alpha_i$.

(e) Under this basis, the Jordan normal form of A on the generalized eigenspace K_i will be

$$\begin{pmatrix} J_{k_{i,1}}(\lambda) & & & & \\ & J_{k_{i,2}}(\lambda) & & & \\ & & \ddots & & \\ & & & J_{k_{i,\gamma'}}(\lambda) & \\ & & & & \lambda I_{\gamma-\gamma'} \end{pmatrix}$$

- **Generalized eigenspace** (of λ): The kernel of $(A - \lambda I)^{d_\lambda}$. *Denoted by* $\boldsymbol{K_\lambda}$. *Given by*

$$K_\lambda = \ker(A - \lambda I)^{d_\lambda}$$

- $\boldsymbol{d_\lambda}$: The power of $A - \lambda I$ for which the kernel stabilizes.
- The JNF computation can be really heavy; we'll only ever compute 2×2 or 3×3 versions.
- Example[1]:
  - Consider

$$A = \begin{pmatrix} -2 & 2 & 1 \\ -7 & 4 & 2 \\ 5 & 0 & 0 \end{pmatrix}$$

  - Then

$$\chi_A(z) = z(z-1)^2$$

  - (1) It follows that

$$\lambda_1 = 0 \qquad\qquad \lambda_2 = 1$$

  - (2) We have that

$$\ker(A - 0I) = \operatorname{span}\left\{\begin{pmatrix} 0 \\ -1 \\ 2 \end{pmatrix}\right\} \qquad\qquad \ker(A - 1I) = \operatorname{span}\left\{\begin{pmatrix} 1 \\ -1 \\ 5 \end{pmatrix}\right\}$$

    - We call the left vector above q_1 and the right vector above q_2.
  - Thus,

$$A \sim \left(\begin{array}{c|cc} 0 & & \\ \hline & 1 & x \\ & & 1 \end{array}\right)$$

[1] Largely ignore this misguided relic of class that day.

- We find that
$$(A - 1I)^2 = \begin{pmatrix} 0 & 0 & 0 \\ 10 & -5 & -3 \\ -20 & 10 & 6 \end{pmatrix}$$
so
$$\ker(A - I)^2 = \text{span}\left\{ \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 \\ 0 \\ 10 \end{pmatrix} \right\}$$
- Clearly,
$$\ker(A - I) \subsetneq \ker(A - I)^2$$
so we can stop here because the dimension of the kernel has reached the algebraic multiplicity.
- Since $q_2 \in K_1$, q_3 solves the equation $(A - I)q_3 = q_2$.
- We know that
$$\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} e_1 = \lambda e_1 \qquad \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} e_2 = e_1 + \lambda e_2$$
- It follows that
$$q_3 = \begin{pmatrix} 0 \\ 3 \\ -5 \end{pmatrix}$$
and hence
$$Q = \begin{pmatrix} q_1 & q_2 & q_3 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & -1 & 3 \\ 2 & 5 & -5 \end{pmatrix}$$
and
$$Q^{-1}AQ = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

- Simple cases.
- The 2×2 case.
  - $A \in \mathcal{M}^2(\mathbb{C})$ can only have nontrivial Jordan form if it has a single eigenvalue λ with $\alpha_\lambda = 2$ and $\gamma_\lambda = 1$. If both equal 2, then $A = \lambda I_2$. If it has two eigenvalues, then it is regularly diagonalizable.
  - In this particular case, calculate λ from $\chi_Z(z) = (z - \lambda)^2$, find one eigenvector v, and find the other generalized eigenvector u; u will satisfy $(A - \lambda I)u = v$. The connecting matrix will be $Q = (v|u)$[2] and the JNF is
$$Q^{-1}AQ = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$$
- The 3×3 case.
  - We divide into three nontrivial cases: $\chi_A(z) = (z - \lambda)^3$ with $\gamma_\lambda = 2$, $\chi_A(z) = (z - \lambda)^3$ with $\gamma_\lambda = 1$, and $\chi_A(z) = (z - \lambda)^2(z - \mu)$ with $\gamma_\lambda = 1$.
  - In the first case, we have two eigenvectors v_1, v_2 (make sure to pick v_1 such that it is also in the column space of $A - \lambda I$!). We can find the third (generalized) eigenvector by solving $(A - \lambda I)u = v_1$. Then $Q = (v_1|u|v_2)$ and the JNF is
$$Q^{-1}AQ = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix}$$

[2] Order matters! We need the eigenvector, specifically, to get scaled by λ only.

- In the second case, we have one eigenvector v. We can find the second and third generalized eigenvectors by solving $(A-\lambda I)u_1 = v$ and $(A-\lambda I)u_2 = u_1$. Then $Q = (v|u_1|u_2)$ and

$$Q^{-1}AQ = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}$$

- In the third case, we have two eigenvectors v_λ, v_μ. We can find the third (generalized) eigenvector by solving $(A-\lambda I)u = v_\lambda$. Then $Q = (v_\lambda|u|v_\mu)$ and

$$Q^{-1}AQ = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \mu \end{pmatrix}$$

3.3 Matrix Calculus

10/14: • Today: Matrix calculus.

- We introduced the Jordan normal form because it is an easy form on which to do matrix calculus.
- **Matrix norm**: A function for $n \times n$ complex matrices such that
  1. $\|A\| \geq 0$, $\|A\| = 0$ iff $A = 0$.
  2. $\|A+B\| \leq \|A\| + \|B\|$.
  3. $\|\lambda A\| = |\lambda|\|A\|$.
  4. $\|AB\| \leq \|A\|\|B\|$.

  Denoted by $||\cdot||$.

  - The first three axioms above are the normal norm axioms; the last one is unique to matrix norms.
- **Operator norm**: The norm defined by

$$\|Ax\| = \sup_{|x|=1} |Ax|$$

- **??**: The norm defined by

$$\|A\| = \sum_{i,j=1}^{n} |a_{i,j}|$$

- Theorem: Any two matrix norms are equivalent.
- **Convergent** (sequence of matrices): A sequence of matrices A_n for which there exists A such that $\|A_n - A\| \to 0$ as $n \to \infty$. *Denoted by* $\boldsymbol{A_n \to A}$.
  - Note that $\|A_n - A\| \to 0$ iff the entries of A_n converge to the entries of A.
- Suppose $A(t) = (a_{ij}(t))_{i,j=1}^n$ is a matrix function. Then

$$A'(t) = (a'_{ij}(t))_{i,j=1}^n \qquad \int_{t_0}^{t} A(t)\,\mathrm{d}t = \left(\int_{t_0}^{t} a_{ij}(\tau)\mathrm{d}\tau\right)_{i,j=1}^{n}$$

- The product rule holds:

$$\frac{\mathrm{d}}{\mathrm{d}t}[A(t)B(t)] = A'(t)B(t) + A(t)B'(t)$$

- However, matrix multiplication is not commutative. This can get us into trouble in the following situation: We might think that
$$\frac{\mathrm{d}}{\mathrm{d}t}[A(t)^2] = 2A'(t)A(t)$$
but, in fact,
$$\frac{\mathrm{d}}{\mathrm{d}t}[A(t)^2] = A'(t)A(t) + A(t)A'(t)$$
  - For example, let
$$A(t) = \begin{pmatrix} 0 & 1 \\ t & 0 \end{pmatrix}$$
  - Then
$$A'(t) = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$
  - It follows that
$$\frac{\mathrm{d}}{\mathrm{d}t}[A'(t)^2] = \underbrace{\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}}_{A'(t)A(t)} + \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}}_{A(t)A'(t)}$$
  - Notice that $A'(t)A(t) \neq A(t)A'(t)$.
- Suppose we have a matrix A and we want to compute A^{100}.
- If A is diagonalizable, then $A^n = Q\Lambda^n Q^{-1}$.
- What if A is not diagonalizable?
  - Then we convert to A to Jordan normal form $A = QBQ^{-1}$. Thus, we just need to compute the powers of the Jordan blocks.
  - Suppose
$$J_d(\lambda) = \begin{pmatrix} \lambda & 1 & & 0 \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda \end{pmatrix}$$
    - In a given Jordan block, all entries above the diagonal are 1.
  - Decompose
$$J_d(\lambda) = \lambda I_d + N_d$$
  - Note that N_d is nilpotent — every successive power to which you raise it shifts the 1s up one row until it becomes the zero matrix.
  - In computing $[J_d(\lambda)]^m$, invoke the binomial expansion. When $m < d$ invoke the full expansion. When $m \geq d$, neglect all zero terms (terms with N_d^i for $i \geq m$):
$$[J_d(\lambda)]^m = \binom{m}{0}\lambda^m I_d + \binom{m}{1}\lambda^{m-1}N_d + \cdots + \binom{m}{d-1}\lambda^{m-d+1}N_d^{d-1}$$
  - Example: When $d = 3$, then
$$\begin{pmatrix} \lambda & 1 & 0 \\ & \lambda & 1 \\ & & \lambda \end{pmatrix}^m = \begin{pmatrix} \lambda^m & m\lambda^{m-1} & m(m-1)\lambda^{m-2} \\ & \lambda^m & m\lambda^{m-1} \\ & & \lambda^m \end{pmatrix}$$
- We will only compute JNF for 2×2 and 3×3; Shao reviews these cases from last class.

- We now have a formula to compute the powers of matrices with ease, so we can move onto more complicated functions of matrices now.
- Consider the power series
$$f(z) = c_0 + c_1 z + c_2 z^2 + \cdots$$
  - The c_i are complex coefficients.
- **Analytic** (function): A function whose Taylor series (locally) converges and converges to the function in question.
- We can consider an analytic function of matrices:
$$f(A) = c_0 I + c_1 A + c_2 A^2 + \cdots$$
- **Radius of convergence**: The number R such that the series converges absolutely for $\|A\| < R$.
  - We do not talk about the radius of convergence any more in this course.
- **von Neumann series**: The series $I + A + A^2 + \cdots$ converging to $(I_n - A)^{-1}$ for any $\|A\| < 1$.
  - Example: We can check that the von Neumann series for N_d converges.
- Suppose $A = QBQ^{-1}$. Then
$$\begin{aligned} f(A) &= f(QBQ^{-1}) \\ &= c_0 I + c_1(QBQ^{-1}) + c_2(QBQ^{-1})^2 + \cdots \\ &= Q(c_0 I + c_1 B + c_2 B^2 + \cdots)Q^{-1} \\ &= Qf(B)Q^{-1} \end{aligned}$$
  - Going even further,
$$B = \begin{pmatrix} B_1 & 0 \\ 0 & B_2 \end{pmatrix} \quad \Longrightarrow \quad f(B) = \begin{pmatrix} f(B_1) & 0 \\ 0 & f(B_2) \end{pmatrix}$$
  - In particular, if A is diagonalizable, then
$$f(A) = Q \begin{pmatrix} f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_n) \end{pmatrix} Q^{-1}$$
- Suppose A is not diagonalizable, and f is some analytic function.
  - Then in the viscinity of a, f can be approximated by the Taylor series
$$f(z) = f(a) + f'(a)(z-a) + \frac{1}{2!} f^{(2)}(a)(z-a)^2 + \cdots$$
  - Similarly, we can approximate $f[J_d(\lambda)]$ in the viscinity of λI_d with the Taylor series
$$\begin{aligned} f[J_d(\lambda)] &= f(\lambda I_d + N_d) \\ &= f(\lambda I_d) + f'(\lambda I_d)[(\lambda I_d + N_d) - \lambda I_d] + \frac{1}{2!} f^{(2)}(\lambda I_d)[(\lambda I_d + N_d) - \lambda I_d]^2 + \cdots \\ &= f(\lambda) I_d + f'(\lambda) N_d + \frac{1}{2!} f^{(2)}(\lambda) N_d^2 + \cdots \\ &= \begin{pmatrix} f(\lambda) & f'(\lambda) & \cdots & \frac{f^{(d-1)}(\lambda)}{(d-1)!} \\ & f(\lambda) & \ddots & \vdots \\ & & \ddots & f'(\lambda) \\ & & & f(\lambda) \end{pmatrix} \end{aligned}$$

- **Matrix exponential** (of A): The matrix with identical dimensions to A defined by the following power series. *Denoted by* $\mathbf{e}^{\boldsymbol{A}}$. *Given by*

$$\mathrm{e}^A = I_n + A + \frac{1}{2!}A^2 + \cdots$$

  - This power series is convergent for matrices with $\|A\| < 1$ since $\|A^m\| \leq \|A\|^m \to 0$.
  - Usual rules that you might expect the matrix exponential to obey based on the notation are obeyed.

$$\mathrm{e}^{(t+\tau)A} = \mathrm{e}^{tA}\mathrm{e}^{\tau A} \qquad\qquad \mathrm{e}^{A+B} = \mathrm{e}^A\mathrm{e}^B$$

- An explicit formula for the e^{tA}.
  - We know that $tA = tQBQ^{-1}$, where we may take B be in JNF.
  - Consider $\mathrm{e}^{tJ_3(\lambda)}$, for example.
  - Then from the above, we have that

$$\mathrm{e}^{tJ_3(\lambda)} = \begin{pmatrix} \mathrm{e}^{t\lambda} & t\mathrm{e}^{t\lambda} & \frac{t^2}{2}\mathrm{e}^{t\lambda} \\ & \mathrm{e}^{t\lambda} & t\mathrm{e}^{t\lambda} \\ & & \mathrm{e}^{t\lambda} \end{pmatrix}$$

  - Thus,

$$\mathrm{e}^{tA} = Q\mathrm{e}^{tB}Q^{-1}$$

- Next time: First order linear systems with constant coefficients; will make use of e^{tA}.
- Next Wednesday: Review; next Friday: Midterm.

Week 4

Linear Systems

4.1 Autonomous Linear Systems

10/17:

- Today: General theory for autonomous linear systems.
- Review session Wednesday (no new material).
- First midterm Friday.
  - Test problems will be slight variations of homework problems or examples given in class.
- **Linear autonomous system**: A system of n linear equations written in the following form. *Denoted by* $\boldsymbol{y' = Ay}$. *Given by*

$$\begin{pmatrix} y^1 \\ \vdots \\ y^n \end{pmatrix}' = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} y^1 \\ \vdots \\ y^n \end{pmatrix} \qquad y(0) = 0$$

  - Note that the a_{ij}'s are complex or real.
- The explicit solution is given by $y(t) = \mathrm{e}^{tA} y_0$.
  - Recall that $\mathrm{d}/\mathrm{d}t \left(\mathrm{e}^{tA}\right) = A\mathrm{e}^{tA}$, as we can show via the power series expansion.
- **Picard iteration**: We take

$$\begin{aligned}
y'(t) &= Ay(t) \\
\int_0^t y'(\tau)\,\mathrm{d}\tau &= \int_0^t Ay(\tau)\mathrm{d}\tau \\
y(t) &= y_0 + \int_0^t Ay(\tau_1)\mathrm{d}\tau_1 \\
&= y_0 + \int_0^t A\left[y_0 + \int_0^{\tau_1} Ay(\tau_2)\mathrm{d}\tau_2\right]\mathrm{d}\tau_1 \\
&= y_0 + tAy_0 + \int_0^t \int_0^{\tau_1} A^2 y(\tau_2)\mathrm{d}\tau_2\mathrm{d}\tau_1 \\
&= y_0 + tAy_0 + \int_0^t \int_0^{\tau_1} A^2\left[y_0 + \int_0^{\tau_2} Ay(\tau_3)\mathrm{d}\tau_3\right]\mathrm{d}\tau_2\mathrm{d}\tau_1 \\
&= y_0 + tAy_0 + \frac{t^2A^2}{2} + \int_0^t \int_0^{\tau_1} \int_0^{\tau_2} A^3 y(\tau_3)\mathrm{d}\tau_3\mathrm{d}\tau_2\mathrm{d}\tau_1
\end{aligned}$$

$$= \sum_{k=0}^{m} \frac{t^k A^k}{k!} y_0 + A^{m+1} \underbrace{\int_0^t \cdots \int_0^{\tau_m}}_{m+1} y(\tau_{m+1}) \mathrm{d}\tau_{m+1} \cdots \mathrm{d}\tau_1$$

- We get from the second to the third line by substituting $y(t)$, as defined into the second line, into where it appears in the integral.

- We want to show that the integral converges to zero.
  - The magnitude of the remainder is less than or equal to

$$\|A\|^{m+1} \left(\sup_{\tau \in [0,t]} |y(\tau)| \right) \frac{t^{m+1}}{(m+1)!}$$

    - Justification of this term: Look at the rightmost term in the last line of the Picard iteration above. Imagine taking the norm of it. Splitting the "scalar" integral from the matrix allows us to take a matrix norm, and the property $\|AB\| \leq \|A\| \|B\|$ tells us that $\left\|A^{m+1}\right\| \leq \|A\|^{m+1}$. Then with respect to the integral, if we evaluate it, we will get the next polynomial term in the sequence — $t^{m+1}/(m+1)!$ — times at most the maximum value of y at every infinitesimal.
  - We can visualize lower-dimensional integrals as the volume of the corresponding unit **simplex**.
    - For example, in $\mathbb{R}^2$,

$$\int_0^1 \int_0^{\tau_1} 1 \mathrm{d}\tau_2 \, \mathrm{d}\tau_1$$

      can be visualized as the area of the unit triangle. This rationalizes why it evaluates to $1/2$, the area of said triangle.
    - In $\mathbb{R}^3$,

$$\int_0^1 \int_0^{\tau_1} \int_0^{\tau_2} 1 \mathrm{d}\tau_3 \mathrm{d}\tau_2 \, \mathrm{d}\tau_1$$

      can be visualized as the area of the unit simplex. This rationalizes why it evaluates to $1/3! = 1/6$, the volume of said simplex.
  - Since $(m+1)! \to \infty$ faster than any other term, the whole thing goes to zero.
  - Thus, since the remainder goes to zero as we add more terms, we eventually reach the limit

$$\begin{aligned} y(t) &= \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k y_0 \\ &= \mathrm{e}^{tA} y_0 \end{aligned}$$

- **Simplex**: A higher-dimensional generalization of a triangle.

- We now consider the following inhomogeneous equation. An appropriate integrating factor still helps.

$$\begin{aligned} y' &= Ay + f(t) \\ y' - Ay &= f(t) \\ \mathrm{e}^{-tA} y' - A\mathrm{e}^{-tA} y &= \mathrm{e}^{-tA} f(t) \\ \frac{\mathrm{d}}{\mathrm{d}t} \left(\mathrm{e}^{-tA} y(t) \right) &= \mathrm{e}^{-tA} f(t) \\ \mathrm{e}^{-tA} y(t) - y_0 &= \int_0^t \mathrm{e}^{-\tau A} f(\tau) \mathrm{d}\tau \\ y(t) &= \mathrm{e}^{tA} y_0 + \int_0^t \mathrm{e}^{(t-\tau)A} f(\tau) \mathrm{d}\tau \end{aligned}$$

  - We also call this the Duhamel formula.

- Note that if your time scale starts from t_0, then
$$y(t) = \mathrm{e}^{(t-t_0)A}y(t_0) + \int_{t_0}^{t} \mathrm{e}^{(t-\tau)A}f(\tau)\mathrm{d}\tau$$
- The utility of JNF: If we want to understand $\mathrm{e}^{tA}y_0$, we convert $A = QBQ^{-1}$, allowing us to evaluate e^{tA}.
  - Shao reviews some facts of JNF from previous lectures.
- From last lecture, we have that
$$\mathrm{e}^{tA}y_0 = Q\mathrm{e}^{tB}Q^{-1}y_0$$
- Example: Let
$$A = \begin{pmatrix} -2 & 2 & 1 \\ -7 & 4 & 2 \\ 5 & 0 & 0 \end{pmatrix}$$
  - This is the same matrix from a previous lecture. As before, we have that
$$Q = \begin{pmatrix} 0 & 1 & 0 \\ -1 & -1 & 3 \\ 2 & 5 & -5 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$
    - Recall that the left two vectors are normal eigenvectors (the leftmost one corresponds to $\lambda_1 = 0$ and the middle one corresponds to $\lambda_2 = 1$) and the rightmost one is a generalized eigenvector.
  - We can compute that
$$\mathrm{e}^{tB} = \begin{pmatrix} \mathrm{e}^{0t} & 0 & 0 \\ 0 & \mathrm{e}^{1t} & 1t\mathrm{e}^{1t} \\ 0 & 0 & \mathrm{e}^{1t} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \mathrm{e}^{t} & t\mathrm{e}^{t} \\ 0 & 0 & \mathrm{e}^{t} \end{pmatrix}$$
  - It follows that
$$\begin{aligned} \mathrm{e}^{tA}y_0 &= Q \begin{pmatrix} 1 & 0 & 0 \\ 0 & \mathrm{e}^{t} & t\mathrm{e}^{t} \\ 0 & 0 & \mathrm{e}^{t} \end{pmatrix} Q^{-1}y_0 \\ &= \begin{pmatrix} -3t\mathrm{e}^t + \mathrm{e}^t & 2t\mathrm{e}^t & t\mathrm{e}^t \\ 3t\mathrm{e}^t - 10\mathrm{e}^t + 10 & -2t\mathrm{e}^t + 6\mathrm{e}^t - 5 & -t\mathrm{e}^t + 3\mathrm{e}^t - 3 \\ -15t\mathrm{e}^t + 20\mathrm{e}^t - 20 & 10t\mathrm{e}^t - 10\mathrm{e}^t + 10 & 5t\mathrm{e}^t - 5\mathrm{e}^t + 6 \end{pmatrix} \begin{pmatrix} y_0^1 \\ y_0^2 \\ y_0^3 \end{pmatrix} \end{aligned}$$
- **Stable** (eigenvalue): An eigenvalue $\lambda_j = \sigma_j + i\beta_j$ for which $\sigma_j < 0$.
- **Unstable** (eigenvalue): An eigenvalue $\lambda_j = \sigma_j + i\beta_j$ for which $\sigma_j > 0$.
- **Stable** (subspace of the system): A generalized eigenspace corresponding a stable eigenvalue.
- **Unstable** (subspace of the system): A generalized eigenspace corresponding an unstable eigenvalue.
  - If λ_j is unstable, then the corresponding entries in e^{tB_j} are exponentially growing functions.
  - If λ_j is stable, then the corresponding entries in e^{tB_j} are exponentially decreasing functions.
  - If $\sigma_j = 0$, then the "stability" depends on the geometric multiplicity??
  - Along the stable subspaces, your points will be attracted to zero.
  - Along the unstable subspaces, your points will be repelled from zero.
  - If $\sigma_h = 0$, then we have rotation around a point, oscillation about zero, or oscillation whose magnitude grows to infinity. We do not talk about its stability.

        - We do not include the eigenvector corresponding to $\lambda_1 = 0$ in the above basis of the stable subspace because the solution oscillates about y_1??
    - The stable subspace of our example is

$$\operatorname{span}\left\{\begin{pmatrix}1\\-1\\5\end{pmatrix},\begin{pmatrix}0\\3\\-5\end{pmatrix}\right\}$$

- Recall that B_j acts on K_j.
    - ... in picture??
    - Recall that $\mathbb{C}^n = K_1 \oplus \cdots \oplus K_m$.
    - P_j is not an *orthogonal* projection, but it is a projection of y_0 onto K_j. It's also a polynomial??
- Consider the order n linear differential equation

$$x^{(n)} + a_{n-1}x^{(n-1)} + \cdots + a_1x' + a_0x = 0$$

    - Then we can make a system out of it:

$$\begin{pmatrix}y^1\\\vdots\\y^n\end{pmatrix}' = \underbrace{\begin{pmatrix}0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1\\ -a_0 & -a_1 & \cdots & -a_{n-1}\end{pmatrix}}_{F[p]}\begin{pmatrix}y^1\\\vdots\\y^n\end{pmatrix}$$

        - Recall how to do the transformation from Lecture 1.
    - $F[p]$ is the **Frobenius matrix**.
    - The transpose of this matrix is a very special matrix called the **companion matrix** $C[p] = F[p]^T$.
    - Claim: Let $p(z) = z^n + a_{n-1}z^{n-1} + \cdots + a_1z + a_0$. Then $\chi_{C[p]} = \chi_{F[p]} = p(z)$.

      Proof. Do the Laplace expansion with respect to the last column of $A - zI$ (companion) or last row (Frobenius). □

    - Roots of $p(z)$ are the eigenvalues of $F[p]$ and $C[p]$.
    - Claim: $C[p]$ has **minimal polynomial** $p(z)$.

      Proof. We have that $C[p]e_i = e_{i+1}$ for $i = 1, \dots, n-1$ and

$$C[p]e_n = -a_0e_1 - \cdots - a_{n-1}e_n$$

      which implies that if $r(z)/\deg r < n$ nullifies $C[p]$, then necessarily $r(z) = p(z)$ since $(z-\lambda_j)^{<\alpha_j}$?? □

    - Claim: $C[p], F[p]$ have the same Jordan normal form.
        - More generally, transpose matrices are similar so they have the same JNF.
- **Monic polynomial**: A polynomial whose highest-degree coefficient equals 1.
- **Minimal polynomial** (of A): The unique monic polynomial p of smallest degree such that $p(A) = 0$.
- Theorem: In the Jordan normal form $F[p]$, each λ_j corresponds to only one Jordan block.

- Thus,
$$F[p] \sim \begin{pmatrix} J_{\alpha_1}(\lambda_1) & & \\ & \ddots & \\ & & J_{\alpha_m}(\lambda_m) \end{pmatrix}$$
The implication is that
$$J_d(\lambda) \neq \begin{pmatrix} \lambda & & \\ & \lambda & 1 \\ & & \lambda \end{pmatrix}$$
ever??

- Corollary: The solution $y(t)$ is of the form
$$(\cdots) + a_1 \mathrm{e}^{t\lambda_j} + \cdots + c_{\alpha_j - 1} t^{\alpha_j - 1} \mathrm{e}^{t\lambda_j} + \cdots$$

- Example: Solving a second-order ODE.
$$x'' + ax' + bx = 0 \quad \Longleftrightarrow \quad \begin{pmatrix} y^1 \\ y^2 \end{pmatrix}' = \begin{pmatrix} 0 & 1 \\ -b & -a \end{pmatrix} \begin{pmatrix} y^1 \\ y^2 \end{pmatrix}$$
  - The characteristic polynomial of the equation (and this matrix) is $z^2 + az + b = 0$.
  - If $\lambda_1 \neq \lambda_2$, then $x(t) = A\mathrm{e}^{t\lambda_1} + B\mathrm{e}^{t\lambda_2}$. If $\lambda_1 = \lambda_2 = \lambda$, then $x(t) = A\mathrm{e}^{t\lambda} + Bt\mathrm{e}^{t\lambda}$.

4.2 Midterm 1 Review

10/19:
- Notes on Friday's exam.
  - Three problems. All will be calculations for specific equations. They will all be standard examples that appeared in the lectures or homeworks.
  - The materials that you can bring to the exam are the notes on JNF (printed). You will be dealing with the JNF of 2×2 or 3×3 matrices.
- Review session today, no new content.
- Remind Shao to post teaching notes from more recent weeks.
- **Ordinary differential equation**: An equation that involves an unknown function together with its derivatives. *Given by*
$$F(t, y, y', y'', \dots, y^{(n)}) = 0$$
- **Order** (of an ODE): The highest order derivative present in the ODE.
- Two types of ODE problems: IVPs and BVPs.
  - IVPs arise in dynamical systems.
  - BVPs arise in variational problems in physics.
- We are primarily interested in ODEs which can be explicitly solved for $y \in C^1(\mathbb{R}^n)$ (resp. $C^1(\mathbb{C}^n)$).
- Two types of equations:
  - A higher-order scalar equation.
  - The more general form of vector-valued systems of the form $y' = f(t, y)$.
- In order to determine y, the initial value $y(t_0) = y_0$ is needed.
  - If a vector-valued system, you need $y_0^1, \dots, y_0^n$ (all components).

- If a scalar system, you need $y(t_0), y'(t_0), \dots, y^{(n-1)}(t_0)$.

- The idea of well-posedness is not yet well-defined in the course; we will cover it after the midterm.
- **Well-posed** (IVP): For every initial value, there is only one unique solution, and for a small change in the initial value, there is only a small change in the solution (continuous dependence on initial values).
- The theorem that we've been relying on but haven't proven yet: **Cauchy-Lipschitz / Picard-Lindelöf theorem**.
- **Cauchy-Lipschitz theorem**: If $f(t, y)$ is Lipschitz continuous with respect to y, then the IVP is locally well-posed. *Also known as* **Picard-Lindelöf theorem**.
  - The term **locally well-posed** has not been rigorously defined either.
- Given any ODE, it is usually very easy to verify the Lipschitz condition for the RHS.
- Example of an IVP that is not locally well-posed.
  - $y = \sqrt{y}$, $y(0) = 0$.
  - Note that if we start at any $t_0 > 0$, then this IVP *is* locally well-posed.
- No Cauchy-Lipschitz in the first midterm; just calculations. We will need the precise statement in the second midterm, though.
- We are not going to talk about solutions that require power series because that inevitably involves complex analysis.
- Explicitly solvable equations: Equations of separable form, i.e., the IVP $y'(t) = f(y)g(t)$, $y(t_0) = y_0$.
- From C-L theorem: If $f(y)$ is continuously differentiable in some neighborhood of y_0, then the solution is unique.
- If $f(y_0) = 0$, then $y(t) = y_0$.
  - Because then $y'(t) = f(y_0)g(t) = 0$, so y is a constant function.
- If $f(y) \neq 0$ in some neighborhood of y_0, then the solution should satisfy the implicit equation

$$\int_{y_0}^{y} \frac{\mathrm{d}w}{f(w)} = \int_{t_0}^{t} g(\tau)\mathrm{d}\tau$$

  - We use the chain rule to make separation of variables rigorous: We can differentiate the LHS above wrt. t and get $y'(t)/f(y(t))$.
  - Relating the $f(y_0) = 0$ and $f(y) \neq 0$ cases and not making them overlap: We start integrating from the nonzero value.
- Examples: $y'(t) = p(t)y(t)$ is homogeneous linear. It follows that

$$y(t) = \exp\left[\int_{t_0}^{t} p(\tau)\mathrm{d}\tau\right] y_0$$

- If $p(t) = r \neq 0$, then the solution is exponential growth or decay:

$$y(t) = y_0 \mathrm{e}^{r(t-t_0)}$$

- Logistic growth:

$$y'(t) = ry\left(1 - \frac{y}{M}\right) \quad\Longleftrightarrow\quad y(t) = \frac{My_0\mathrm{e}^{rt}}{M + y_0(\mathrm{e}^{rt} - 1)}$$

- Shao gives the related implicit integral equation and logarithmic equation as well.

- There exist equations which cannot be solved by separation of variables. One case is equations of the form
$$g(x,y)\frac{\mathrm{d}y}{\mathrm{d}x}+f(x,y)=0$$
where $\partial_x g(x,y)=\partial_y f(x,y)$.
  - In this case, there exists $F(x,y)$ such that $\partial_x F=f$, $\partial_y F=g$, and $F(x,y)=C$ is the relation satisfied by the solution.
  - These are **exact form** equations.
- Not all equations satisfy this relation. However, it is often possible (though potentially quite hard) to find an **integrating factor** by which you can multiply your equation to put it in exact form.
- Special case where it is easy to find the integrating factor: Consider the inhomogeneous linear equation $y'(t)=p(t)y(t)+f(t)$. Then the integrating factor is
$$\mu=\exp\left[-\int_{t_0}^{t}p(\tau)\mathrm{d}\tau\right]$$
- Multiplying through, we get
$$\begin{aligned}\exp\left[-\int_{t_0}^{t}p(\tau)\mathrm{d}\tau\right]f(t)&=\exp\left[-\int_{t_0}^{t}p(\tau)\mathrm{d}\tau\right]y'(t)-\exp\left[-\int_{t_0}^{t}p(\tau)\mathrm{d}\tau\right]p(t)y(t)\\&=\frac{\mathrm{d}}{\mathrm{d}t}\left\{\exp\left[-\int_{t_0}^{t}p(\tau)\mathrm{d}\tau\right]y(t)\right\}\\y(t)&=\exp\left[\int_{t_0}^{t}p(\tau)\mathrm{d}\tau\right]y_0+\exp\left[\int_{t_0}^{t}p(\tau)\mathrm{d}\tau\right]\cdot\int_{t_0}^{t}\exp\left[-\int_{t_0}^{\tau}p(\tau')\mathrm{d}\tau'\right]f(t)\mathrm{d}\tau\end{aligned}$$
- The above formula is complicated, though, so it is probably better to remember the method than to memorize the above.
- When $p(t)=a$ for all t, $y'(t)=ay+f(t)$. The solution is given by the **Duhamel formula**.
- **Duhamel formula**: The following equation, which solves ODEs of the form $y'(t)=ay+f(t)$. *Given by*
$$y(t)=\mathrm{e}^{a(t-t_0)}y_0+\int_{t_0}^{t}\mathrm{e}^{a(t-\tau)}f(\tau)\mathrm{d}\tau$$
  - We should understand the derivation, but we can apply the Duhamel formula on PSets and exams without further justification.
- Other things (??) are related to this form by some smart transformation.
- Final example of explicitly solvable ODEs: Linear autonomous systems.
- **Linear autonomous system**: A system of equations of the form $y'=Ay$ where A is a constant $n\times n$ matrix and y takes its value in $\mathbb{R}^n$ (resp. $\mathbb{C}^n$).
- The homogeneous solution is
$$y(t)=\mathrm{e}^{tA}y_0$$
where $\mathrm{e}^{tA}=1+\frac{tA}{1!}+\frac{t^2A^2}{2!}+\cdots$.
- In the inhomogeneous case $y'=Ay+f(t)$, our solution is
$$y(t)=\mathrm{e}^{tA}y_0+\int_0^t\mathrm{e}^{(t-\tau)A}f(\tau)\mathrm{d}\tau$$

- We don't want to compute e^{tA} using an infinite power series. Thus, we introduce similarity.
- Let Q be the connecting matrix from the standard basis to the new basis. Then the matrix of Q is the set of new basis vectors q_1, q_2, q_3, i.e., $Q = \begin{pmatrix} q_1 & q_2 & q_3 \end{pmatrix}$. Then $B = Q^{-1}AQ$ or $A = QBQ^{-1}$.
- We want B to be in the most convenient basis possible. Thus, we take the basis to be the Jordan basis.
- We fortunately have $\mathrm{e}^{tA} = Q\mathrm{e}^{tB}Q^{-1}$.
- Consider $\chi_A(z) = \det(zI_n - A)$ where $n = 2, 3$. If χ_A has distinct roots, then the eigenvalues of A are distinct. At this point, we can find an eigenvector corresponding to each eigenvalue and diagonalize our matrix.
- Alternatively, if χ_A has multiple roots...
  - 2×2 case, A is not diagonal. Then there is only one eigenvector v_λ. In this case, solve $(A - \lambda)u = v_\lambda$. Here, we say that the algebraic multiplicity is 2 and the geometric multiplicity is 1. Then

$$Q = \begin{pmatrix} v_\lambda & u \end{pmatrix} \qquad B = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} \qquad \mathrm{e}^{tA} = Q \begin{pmatrix} \mathrm{e}^{t\lambda} & t\mathrm{e}^{t\lambda} \\ 0 & \mathrm{e}^{t\lambda} \end{pmatrix} Q^{-1}$$

  - 3×3 case: If we have λ of $\alpha_\lambda = 2$ and μ of $\alpha_\mu = 1$, or if we have λ with $\alpha_\lambda = 3$. First case: Check geometric multiplicity of λ, i.e., how many linearly independent v give $(A - \lambda I)v = 0$. If there is one, solve $(A - \lambda I)u = v_\lambda$. If there are more than one, A is diagonalizable. Second case: Check geometric multiplicity of λ. Divide into two subcases. If $\gamma_\lambda = 1$, then we need to solve $(A - \lambda I)u_1 = v_\lambda$ and $(A - \lambda I)u_2 = u_1$, and we get

$$Q = \begin{pmatrix} v_\lambda & u_1 & u_2 \end{pmatrix} \qquad B = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}$$

    If $\gamma_\lambda = 2$, then cleverly choose v_1 such that v_1 is in the column space of $A - \lambda I$. This will allow us to solve $(A - \lambda I)u = v_1$. Then

$$Q = \begin{pmatrix} v_1 & u & v_2 \end{pmatrix} \qquad B = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix}$$

- For our linear autonomous system $y' = Ay$, λ is an eigenvector of A. Write $\lambda = \sigma + i\beta$. If $\lambda > 0$, then λ is **unstable** and the corresponding generalized eigenspace is said to be an **unstable eigenspace**.
- For example, if the JNF is

$$A = \left(\begin{array}{cc|c} 1 & 1 & \\ & 1 & \\ \hline & & -2 \end{array}\right)$$

  then the eigenspace corresponding to the upper block is said to be unstable, and the other one is said to be stable.
- Consider the vector $\mathrm{e}^{tA}v$. The entries consist of linear combinations of functions of the form $t^k\mathrm{e}^{t\lambda}$. If the real part is greater than zero, the solution grows exponentially fast in the t direction (notice how $t \to \infty$ implies $t^k\mathrm{e}^{t\lambda} \to \infty$). Otherwise, the solution decays exponentially fast (notice how $t \to \infty$ implies $t^k\mathrm{e}^{t\lambda} \to 0$).

Week 5

End Quantitative and Intro to Qualitative

5.1 Planar Autonomous Linear Systems

10/24:

- Review of vector fields.
- **Phase diagram**: A diagram that shows the qualitative behavior of an autonomous ordinary differential equation. *Also known as* **phase portrait**.

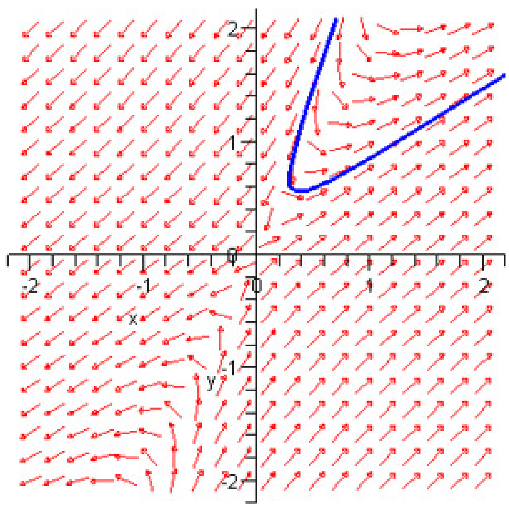
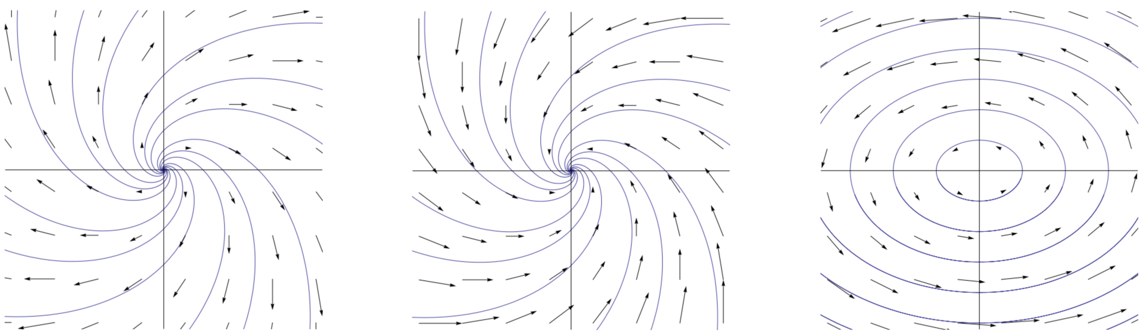


Figure 5.1: Phase diagram example.

  - Consists of a selection of arrows describing, to some extent, a vector field and is often paired with integral curves.
- Suppose $\Omega \subset \mathbb{R}^n$ is open.
- **Vector field** (on Ω): A mapping from $\Omega \to \mathbb{R}^n$. *Denoted by* $\boldsymbol{X}$.
  - Essentially, a vector field assigns to every point of some region a vector; the definition just formalizes this notion.
- **Flow**: A formalization of the idea of the motion of particles in a fluid.
  - The solution to the IVP $\frac{\mathrm{d}y}{\mathrm{d}t} = X(y)$, $y(0) = x$.

- If X is C^1, then for all $x \in \Omega$, there exists a unique solution y to the above IVP.
- **Orbit** (of x under X): The trajectory $y(t, x)$.
  - Recall that the tangent vector to any trajectory at any point coincides with the vector to which X maps that point.
- **Fixed point**: A point $x_0 \in \Omega$ such that $X(x_0) = \bar{0}$.
  - If x_0 is a fixed point, then the trajectory is $y(t) = x_0$.
- Today: We will consider flows on vector fields where the dimension is two and our vector field is linear. In particular...
- Let A be a 2×2 real matrix, and let $X(x) = Ax$.
  - In this case, $x_0 = 0$ is the only fixed point.
  - The flow is given by the linear differential equation $y' = Ay$, $y(0) = x$. The solution is $y(t) = \mathrm{e}^{tA}x$.
- Case 1: A has no real eigenvalues.



(a) $\sigma > 0$. (b) $\sigma < 0$. (c) $\sigma = 0$.

Figure 5.2: Phase diagrams for a planar system with no real eigenvalues.

  - We know that $\chi_A(z)$ is a real polynomial: $\chi_A(z) = z^2 + (\operatorname{trace} A)z + \det A$, and since A is real, both $\operatorname{trace} A$ and $\det A$ are real.
  - Thus, the eigenvalues appear as conjugate pair, i.e., we may write $\lambda = \sigma + i\beta$ and $\bar{\lambda} = \sigma - i\beta$.
    - $\alpha = \gamma = 1$ for both eigenvalues.
    - The eigenvectors must also be complex conjugates.
  - Distinct eigenvalues imply that A is diagonalizable.
  - However, this is not what we want because if we use the complex form, then

$$\mathrm{e}^{tA} = Q \begin{pmatrix} \mathrm{e}^{t\lambda} & 0 \\ 0 & \mathrm{e}^{t\bar{\lambda}} \end{pmatrix} Q^{-1}$$

  - Indeed, we want to get a real matrix out of $Q, \mathrm{e}^{t\Lambda}, Q^{-1}$ all complex. We have

$$\begin{aligned}
\mathrm{e}^{tA}x &= Q \begin{pmatrix} \mathrm{e}^{t(\sigma+i\beta)} & 0 \\ 0 & \mathrm{e}^{t(\sigma-i\beta)} \end{pmatrix} \underbrace{Q^{-1}x}_{z} \\
&= Q \begin{pmatrix} \mathrm{e}^{t(\sigma+i\beta)} z^1 \\ \mathrm{e}^{t(\sigma-i\beta)} z^2 \end{pmatrix} \\
&= z^1 \mathrm{e}^{t(\sigma+i\beta)} v + z^2 \mathrm{e}^{t(\sigma-i\beta)} \bar{v}
\end{aligned}$$

- Since $y(0) = x = z^1 v + z^2 \bar{v} \in \mathbb{R}^2$ (i.e., $z^1 v + z^2 v$ is *real*), we know that it is equal to its complex conjugate. This tells us that

$$
\begin{aligned}
z^1 v + z^2 \bar{v} &= \bar{z}^1 \bar{v} + \bar{z}^2 v \\
z^1 &= \bar{z}^2
\end{aligned}
$$

- It follows that

$$
\begin{aligned}
y(t) &= \mathrm{e}^{tA} x \\
&= z^1 \mathrm{e}^{t(\sigma + i\beta)} v + \bar{z}^1 \mathrm{e}^{t(\sigma - i\beta)} \bar{v} \\
&= z^1 \mathrm{e}^{t(\sigma + i\beta)} v + \overline{z^1 \mathrm{e}^{t(\sigma + i\beta)} v} \\
&= 2 \operatorname{Re}(z^1 \mathrm{e}^{t(\sigma + i\beta)} v) \\
&= 2 \operatorname{Re}(z^1 \mathrm{e}^{\sigma t} (\cos(\beta t) + i \sin(\beta t))(v_1 + i v_2)) \\
&= 2 \operatorname{Re}(z^1 \mathrm{e}^{\sigma t} (\cos(\beta t) v_1 + i \cos(\beta t) v_2 + i \sin(\beta t) v_1 - \sin(\beta t) v_2)) \\
&= 2 \mathrm{e}^{\sigma t} \cos(\beta t) \cdot \operatorname{Re}(z^1 v) - 2 \mathrm{e}^{\sigma t} \sin(\beta t) \cdot \operatorname{Im}(z^1 v)
\end{aligned}
$$

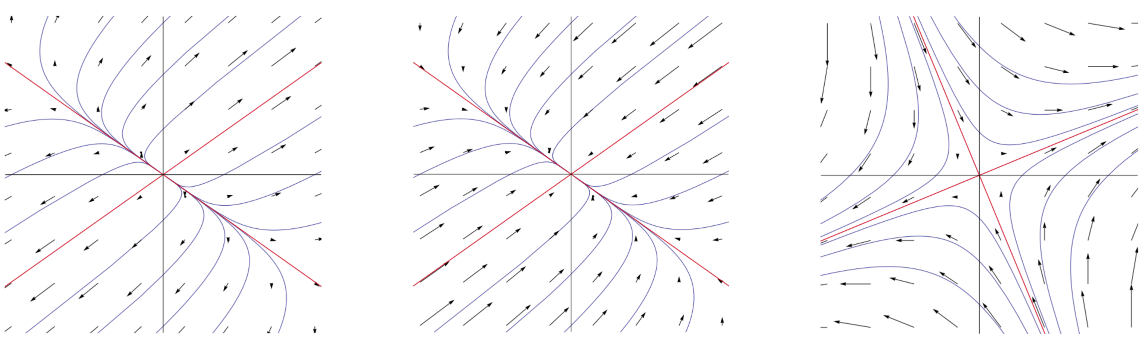
- Suppose $\sigma \neq 0$. Then

$$
x \mapsto \begin{pmatrix} \operatorname{Re}(z^1 v) \\ \operatorname{Im}(z^1 v) \end{pmatrix}
$$

  is a real linear transformation on $\mathbb{R}^2$.
- It follows that the trajectories are just spirals in the complex plane.
- If $\sigma > 0$, then the spiral repels from the origin. If $\sigma < 0$, then the spiral attracts to the origin. If $\sigma = 0$, we get an ellipse.
- Therefore, we have completely classified equations of the form

$$
\begin{pmatrix} y^1 \\ y^2 \end{pmatrix}' = \begin{pmatrix} y^2 \\ -\omega^2 y^1 \end{pmatrix}
$$

- Case 2: A has real eigenvalues and *is* diagonalizable.



(a) $\lambda_1, \lambda_2 > 0$. (b) $\lambda_1, \lambda_2 < 0$. (c) $\lambda_1 < 0 < \lambda_2$ (WLOG).

Figure 5.3: Phase diagrams for a diagonalizable planar system with real eigenvalues.

- Suppose $\lambda_1, \lambda_2 \in \mathbb{R}$ have corresponding linearly independent eigenvectors v_1, v_2.
- If we choose v_1, v_2 to be our basis, then

$$
\mathrm{e}^{tA} = Q \begin{pmatrix} \mathrm{e}^{t\lambda_1} & 0 \\ 0 & \mathrm{e}^{t\lambda_2} \end{pmatrix} Q^{-1}
$$

  where $Q = \begin{pmatrix} v_1 & v_2 \end{pmatrix}$.

- Thus, as before, the solution may be expressed in the following form, where $z = Q^{-1}x$.

$$y(t) = \mathrm{e}^{tA}x = \mathrm{e}^{\lambda_1 t}z^1 v_1 + \mathrm{e}^{\lambda_2 t}z^2 v_2$$

- Moving forward, it will be convenient to work in the v_1, v_2 basis. We divide into three subcases ($\lambda_1, \lambda_2 > 0$ [Figure 5.3a], $\lambda_1, \lambda_2 < 0$ [Figure 5.3b], and WLOG $\lambda_1 < 0 < \lambda_2$ [Figure 5.3c]).
  1. Notice that
  $$\mathrm{e}^{\lambda_2 t} = \mathrm{e}^{(\lambda_2/\lambda_1)(\lambda_1 t)}$$
  i.e., $\mathrm{e}^{\lambda_2 t}$ is a power of $\mathrm{e}^{\lambda_1 t}$. Thus, when the signs are the same, we get a power function $v_2 = v_1^{\lambda_2/\lambda_1}$.
     - Both subspaces v_1, v_2 are unstable here.
  2. If $\lambda_1, \lambda_2 < 0$, then we have the same trajectories, but they're all attracted to the origin instead of repelled.
     - Both subspaces v_1, v_2 are stable here.
  3. When both eigenvalues have different signs, we are considering power functions of a negative power.
     - The stable subspace is v_2 and the unstable subspace is v_1 here.

- Case 3: A has real eigenvalues and *is not* diagonalizable.

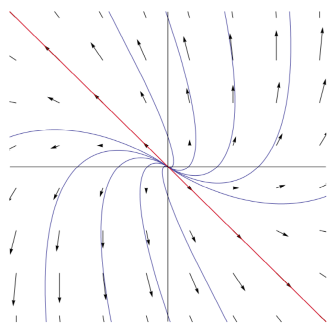


Figure 5.4: Phase diagrams for a nondiagonalizable planar system with real eigenvalues.

- In this case, the matrix exponential is given by

$$\mathrm{e}^{tA} = Q \begin{pmatrix} \mathrm{e}^{t\lambda} & t\mathrm{e}^{t\lambda} \\ 0 & \mathrm{e}^{t\lambda} \end{pmatrix} Q^{-1}$$

- The solution is given by

$$\mathrm{e}^{tA}x = (z^1\mathrm{e}^{t\lambda} + z^2 t\mathrm{e}^{t\lambda})v + z^2\mathrm{e}^{t\lambda}u$$

  where $Q^{-1}x = z$ again.
- In graphing, note that here we have (a distorted version of) the form $y = x \pm x \log x$:

$$y = (z^1\mathrm{e}^{t\lambda} + z^2 t\mathrm{e}^{t\lambda})\hat{\imath} + z^2\mathrm{e}^{t\lambda}\hat{\jmath}$$

  Define $x := \mathrm{e}^{t\lambda}$. Then $t = \lambda^{-1}\ln x$. Substituting, we have

$$\begin{aligned} &= (z^1 x + z^2(\lambda^{-1}\ln x)x)\hat{\imath} + z^2 x\hat{\jmath} \\ &= (z^1 x + z^2\lambda^{-1}x\ln x)\hat{\imath} + z^2 x\hat{\jmath} \end{aligned}$$

- When $\lambda > 0$, the whole space is unstable. Thus, the phase diagram is tangent to the origin.
- When $\lambda < 0$, the trajectories take the same form but this time are attracted to zero. In this case, the whole space is stable.

- We can take x_1 to x_2 iff they are in the same orbit. Conclusion: Orbits never cross.
- Takeaway: You should be able to compute the eigenvalues and eigenvectors and sketch these graphs.
- Shao will post lecture notes after today's lecture!
- Next lecture: The final explicitly solveable case, which is the driven harmonic oscillator.

5.2 Driven Harmonic Oscillator and Resonance

10/26: • We are interested in the 2nd order constant coefficient equation

$$x'' + \mu x' + \omega_0^2 x = H_0 \mathrm{e}^{i\omega t}$$

where $\mu \geq 0$ and $\omega_0, \omega > 0$.

- Two cases where this ODE arises:

(a) A driven harmonic oscillator. (b) An RLC circuit.

Figure 5.5: Origins of the driven harmonic oscillator equation.

1. The driven harmonic oscillator.
   - Consider a mass on a spring.
   - The extent of friction between the mass point and the surface is described by μ.
   - The oscillation is periodically driven by a force of magnitude $H_0 \cos \omega t$.
2. RLC circuit.
   - R is resistance, L is inductance, C is capacitance.
   - We have the laws

$$LI_L'(t) = V_L \qquad CV_C'(t) = I_C \qquad I_R(t) = V_R(t)/R$$

     - Left: Self-inductance.
     - Right: Ohm's law.
   - Combining them with Kirchhoff's laws

$$I(t) = I_R = I_C = I_L \qquad V(t) = V_R + V_L + V_C$$

     we get the RLC circuit equation

$$LI'' + RI' + \frac{I}{C} = V'(t)$$

   - The most interesting cases is when we have a source of alternating current of frequency ω. In this case, $V(t) = V_0 \cos \omega t$ or, in the complex case, $V(t) = V_0 \mathrm{e}^{i\omega t}$. This yields the complex equation

$$I'' + \frac{R}{L} I' + \frac{1}{LC} I = \frac{i\omega V_0}{L} \mathrm{e}^{i\omega t}$$

   - Here, the friction coefficient $\mu = R/L$ and the frequency is $\omega_0 = \sqrt{1/LC}$.

- Recall that we want to solve the following ODE.

$$x'' + \mu x' + \omega_0^2 x = H_0 \mathrm{e}^{i\omega t}$$

  - The homogeneous linear equation $x'' + \mu x' + \omega_0^2 x = 0$ is well-understood, i.e., we can find all of the *homogeneous* solutions to the above equation.
  - Thus, to solve the above inhomogeneous equation, we just have to find a particular solution.
- WLOG let $\omega > 0$.
- From the homework, a particular solution $x_p(t)$ with initial condition $x_p(0) = x_p'(0) = \mu = 0$ can be obtained from the Duhamel formula as follows.

$$x_p(t) = H_0 \int_0^t \frac{\sin \omega_0 (t - \tau)}{\omega_0} \mathrm{e}^{i\omega\tau} \mathrm{d}\tau$$

  - Substituting

$$\sin\theta = \frac{\mathrm{e}^{i\theta} - \mathrm{e}^{-i\theta}}{2i}$$

    into the above allows us to evaluate it.
  - In particular, it follows that

$$x_p(t) = \begin{cases} \dfrac{H_0}{\omega_0^2 - \omega^2}\left(\mathrm{e}^{i\omega t} - \cos\omega_0 t - \dfrac{i\omega}{\omega_0}\sin\omega_0 t\right) & \omega \neq \omega_0 \\ -\dfrac{iH_0}{2\omega_0}\left(t\mathrm{e}^{i\omega_0 t} - \dfrac{\sin\omega_0 t}{\omega_0}\right) & \omega = \omega_0 \end{cases}$$

    - We compute the $\omega = \omega_0$ case using L'Hôpital's rule to analyze the $\omega \neq \omega_0$ case as $\omega \to \omega_0$.
  - If we pump in energy at the same point that we have deviation ($\omega = \omega_0$), then the amplitude of oscillation goes to ∞.
    - Practically, when $\omega \approx \omega_0$, the long-time behavior of the driven oscillator will be very much like a growing oscillator.
    - Eventually, the amplitude will be approximately $(\omega - \omega_0)^{-1}$.
- **Resonance catastrophe**: Inputing energy into a system at its natural frequency, causing the total energy to grow until a mechanical failure occurs.
  - This is what happened at the Millenium Bridge in London; synchronized footsteps caused the bridge to shake really wildly.
- If $\mu > 0$, there will be a particular solution of the form

$$x_p(t) = A(\omega) H_0 \mathrm{e}^{i\omega t}$$

  - From HW1, we have three cases when $\mu > 0$: $0 < \mu < 2\omega_0$, $\mu = 2\omega_0$, and $\mu > 2\omega_0$. These are just the three cases of the characteristic polynomial??
  - Substituting the proposed form of the particular solution into the differential equation, we get

$$\begin{aligned} x_p'' + \mu x_p' + \omega_0^2 x_p &= H_0 \mathrm{e}^{i\omega t} \\ (-\omega^2 + i\omega\mu + \omega_0^2) H_0 A(\omega) \mathrm{e}^{i\omega t} &= H_0 \mathrm{e}^{i\omega t} \\ (-\omega^2 + i\omega\mu + \omega_0^2) A(\omega) &= 1 \\ A(\omega) &= \frac{1}{\omega_0^2 - \omega^2 + i\mu\omega} \end{aligned}$$

  - In theory, we avoid the resonance catastrophe in this case. In practice, however, when $\omega \to 0$, we still run into issues.

- For mass point:

$$|H_0A(\omega)| = \frac{|H_0|}{\sqrt{(\omega^2-\omega_0^2)^2+\mu^2\omega^2}}$$

  - The norm $|H_0A(\omega)|$ is maximized when $\omega_r = \sqrt{\omega_0^2+\mu^2/2}$.
  - $\omega_r \to \omega_0$ implies $\mu \to 0$??
- As for the argument/angle,

$$\arg(H_0A(\omega)) = \arg H_0 + \arg A(\omega)$$

- We consider $\omega : 0 \to \omega_0 \to +\infty$.
  - When $\omega = 0$, the complex amplitude is $1/\omega_0^2$ so it's a real number in the complex plane.
  - If ω is increased a bit, we get the reciprocal of a complex number. Norm is reciprocal, argument is negative.
  - For $\omega = \omega_0$, we have a purely imaginary number.
  - As $\omega \to \infty$, the argument approaches $-\pi$??
  - Showing the shape of the norm and the argument with respect to ω. This allows us to completely describe the resonance phenomena.

- For the RLC circuit, the discussion is a bit different.
  - The external voltage $V(t) = V_0 e^{i\omega t}$. Thus, $V'(t) = iV_0\omega e^{i\omega t}$.
  - Here,

$$x_p(t) = \frac{iV_0\omega e^{i\omega t}}{\omega_0^2-\omega^2+iR\omega/L}$$

  - Look at the complex amplitude.
    - Multiply the numerator and denominator by the inductance L to get

$$x_p(t) = \frac{iV_0\omega L e^{i\omega t}}{L\omega_0^2 - L\omega^2 + iR\omega}$$

    - Then,

$$\text{Norm} = \frac{V_0L}{\sqrt{R^2+\left(\frac{1}{C\omega}-\frac{\omega}{L}\right)^2}}$$

    - For an RLC circuit, the resistance does not affect the resonance frequency.

$$\omega_r = \sqrt{\frac{1}{LC}} = \omega_0$$

    - If you have an external source of voltage, then you can vary the capacity of your circuit to ensure that the voltage will be maximized at a given frequency. We can tune our circuit to a very specific resonance frequency (this is used to filter our radio stations). The RLC circuit is only observable when the resonance coincides with the external resonance.

- There will be a bonus problem which is a PDE describing the vibration of a string.
  - Suppose we have a string with fixed endpoints, and suppose it is undergoing a small vibration.
  - Deviation from the equilibrium is described by a function $u(x,t)$.
  - The simplest equation we can derive is the 1D linear wave equation

$$\frac{\partial^2 u}{\partial t^2} - \frac{1}{c^2}\frac{\partial^2 u}{\partial x^2} = f(x,t)$$

    - c is the speed of the wave.

    - $f(x,t)$ is the given external force.
  - We can show that when $f(x,t)=0$, then the vibration of the string is the linear supposition of infinitely many standing waves.
$$u(x,t)=\sum_{k=1}^{\infty} a_k \mathrm{e}^{\frac{\pi k t}{\ell}} \sin \frac{k\pi}{\ell} x$$
    - There are $k-1$ nodes in the string. These are called standing waves.
  - If you drive it with frequency
$$f(x,t)=\cos \omega t \sin \frac{k\pi}{\ell} x$$
    you encounter the resonance catastrophe.

- We are interested in the driven harmonic oscillator because it describes the vibrations, even of PDEs.
- This concludes our discussion of explicitly solvable differential equations.
- Those that are solvable by power series require complex analysis.
- Starting this Friday, we will talk about the qualitative theory of differential equations.
- Cauchy-Lipschitz this Friday.
- Next week: Continuous dependence on initial values and differentiation with respect to the parameter of this equation.
- After this, we will be able to compute classical examples in the theory of perturbations.
- We will be able to solve the procession of Mercury problem (which was the first experimental verification of general relativity).

5.3 Qualitative Theory of ODEs

10/28:

- First issue: Uniqueness — we want to be able to talk about *the* solution to the IVP.
- We will be considering the IVP $y'(t)=f(t,y)$, $y(t_0)=y_0$ for $y(t)$ an $\mathbb{R}^n$-valued function.
- To embed our rough outline of the Cauchy-Lipschitz theorem into analysis, we start with metric spaces.
- **Metric space**: A set and a metric. *Denoted by* $\boldsymbol{(X,d)}$.
- **Metric**: A function from $X \times X \to [0,+\infty)$ that satisfies the following three axioms. *Denoted by* $\boldsymbol{d}$.
  1. $d(x,y)=d(y,x)$.
  2. $d(x,y)\geq 0$ and $d(x,y)=0$ iff $x=y$.
  3. Triangle inequality: $d(x,z)\leq d(x,y)+d(y,z)$.
- Examples:
  1. $\mathbb{R}^n$ with $d(x,y)=\|x-y\|=\sqrt{\sum_{j=1}^{n}|x^j-y^j|^2}$.
  2. Continuous functions $y:[a,b]\to\mathbb{R}^n$ with $d(y_1,y_2)=\|y_1-y_2\|=\sup_{t\in[a,b]}|y_1(t)-y_2(t)|$.
- In Euclidean spaces, we have **completeness**.
- **Cauchy** (sequence): A sequence $\{x_n\}\subset X$ such that for all $\varepsilon>0$, there exists $N\geq 0$ such that $d(x_m,x_n)<\varepsilon$ for all $m,n>N$.

- **Convergent** (sequence): A sequence $\{x_n\} \subset X$ for which there exists $x \in X$ such that
$$\lim_{n\to\infty} d(x, x_n) = 0$$
- **Complete** (metric space): A metric space (X, d) such that every Cauchy sequence is convergent.
- Theorem (Banach fixed point theorem): Let (X, d) be a complete metric space and let $\Phi : X \to X$ be a function for which there exists $q \in (0, 1)$ such that for all $x, y \in X$,
$$d(\Phi(x), \Phi(y)) \leq q \cdot d(x, y)$$
Then there exists a unique $x \in X$ such that $x = \Phi(x)$.

Proof. We first construct the desired fixed point x.

Pick any $x_0 \in X$. Inductively define $\{x_n\}$ by $x_{n+1} = \Phi(x_n)$, starting from $n = 0$. We will now show that $\{x_n\}$ is a Cauchy sequence. As a lemma, we will prove by induction that
$$d(x_j, x_{j+1}) \leq q^j \cdot d(x_0, x_1)$$
for all $j \in \mathbb{N}_0$. For the base case $j = 0$, equality evidently holds. Now suppose inductively that we have proven that $d(x_j, x_{j+1}) \leq q^j \cdot d(x_0, x_1)$; we want to prove the claim for $j + 1$. But we have that
$$\begin{aligned} d(x_{j+1}, x_{j+2}) &= d(\Phi(x_j), \Phi(x_{j+1})) \\ &\leq q \cdot d(x_j, x_{j+1}) \\ &\leq q \cdot q^j \cdot d(x_0, x_1) \\ &= q^{j+1} \cdot d(x_0, x_1) \end{aligned}$$
as desired.

It follows that
$$\begin{aligned} d(x_n, x_{n+m}) &\leq \sum_{k=0}^{m-1} d(x_{n+k}, x_{n+k+1}) && \text{Triangle inequality} \\ &\leq \sum_{k=0}^{m-1} q^{n+k} \cdot d(x_0, x_1) && \text{Lemma} \\ &= q^n(1 + q + \cdots + q^{m-1}) \cdot d(x_0, x_1) \\ &< q^n(1 + q + \cdots + q^{m-1} + \cdots) \cdot d(x_0, x_1) \\ &= \frac{q^n}{1-q} \cdot d(x_0, x_1) \end{aligned}$$
It follows that the above term will converge to zero as $n \to \infty$, so $\{x_n\}$ is a Cauchy sequence and there exists an x such that $x_n \xrightarrow{d} x$.

We now prove that x is a fixed point of Φ, i.e., that $\Phi(x) = x$. We have that
$$\begin{aligned} d(x, \Phi(x)) &\leq d(x, x_n) + d(x_n, \Phi(x_n)) + d(\Phi(x_n), \Phi(x)) \\ &\leq d(x, x_n) + d(x_n, x_{n+1}) + q \cdot d(x_n, x) \\ &= (1 + q) \cdot d(x, x_n) + d(x_n, x_{n+1}) \end{aligned}$$
where the first term converges since $\{x_n\}$ is convergent and the second term converges since $\{x_n\}$ is Cauchy. Thus, $d(x, \Phi(x)) \to 0$ as $n \to \infty$, so $x = \Phi(x)$, as desired.

Lastly, we prove that x is unique. Suppose that there exists $y \in X$ such that $y = \Phi(y)$. Then
$$d(x, y) = d(\Phi(x), \Phi(y)) \leq q \cdot d(x, y)$$
It follows that $d(x, y) \leq q^n \cdot d(x, y)$, i.e., that $d(x, y) \to 0$ as $n \to \infty$. Therefore, we must have that $d(x, y) = 0$, from which it follows that $x = y$, as desired. ∎

- Notes on the Banach fixed point theorem.
  - Φ is a **contraction**.
  - Shao gives the example of crumpling a sheet of paper (more specifically, dropping a map of a park in that park; a point coincides).
- Example: Fixed point of the cosine function.
  - Define $\{x_n\}$ by $x_{n+1} = \cos x_n$. If $x_0 \in \mathbb{R}$, then $x_1 \in [-1, 1]$ and $x_2 \in [\cos 1, 1]$.
  - Thus, while cosine is not a contraction on the real numbers ($\cos'(-\pi/2) = 1$, for example), we can show that $\cos : [\cos 1, 1] \to [\cos 1, 1]$ is a contraction: If $x, y \in [\cos 1, 1]$, then

$$\begin{aligned} |\cos x - \cos y| &= \left| \int_y^x -\sin t \, \mathrm{d}t \right| \\ &\leq |x - y| \sup_{t \in [\cos 1, 1]} |\sin t| \\ &\leq (\sin 1)|x - y| \end{aligned}$$

  - Thus, cosine has a fixed point at the intersection of $y = \cos x$ and $y = x$ of approximate value $0.739\ldots$
  - Overall, this is a pretty bad example, though.
- Theorem: Let $y_k : [a, b] \to \mathbb{R}^n$ be a Cauchy sequence of continuous functions under the sup norm. Then the limit exists and is continuous.
  - The proof is based on uniform convergence, which we've encountered before in analysis.
  - It follows that $C[a, b]$ (the metric space of all continuous functions on $[a, b]$) is complete.
  - If $\{y_k\} \subset \bar{B}(0, M)$[1], then the limit y is in $\bar{B}(0, M)$.
- Let's return to our ODE $y'(t) = f(t, y(t))$, $y(t_0) = y_0 \in \mathbb{R}^n$.
- We now have the tools to prove the Cauchy-Lipschitz theorem, and we will presently build up to that.
- Although we do not typically think of it this way, f is still a function with a domain and range. In particular, its domain is the set of ordered pairs where the first entry is a real number and the second entry is an element of the range of y, i.e., an element of $\mathbb{R}^n$. Thus, to begin, we are allowed to impose the following conditions on f.
  - Let $f(t, z)$ be defined on $[t_0, t_0 + a] \times \bar{B}(y_0, b)$ for some $a, b \in \mathbb{R}_+$ (we will put further constraints on the values of a, b later).
  - On this domain, suppose $|f|$ is bounded by some $M \in \mathbb{R}$, i.e., $|f(t, z)| \leq M$ for all t, z in the above set.
  - Let f be Lipschitz continuous in the second argument. In particular, there exists $L > 0$ such that $|f(t, z_1) - f(t, z_2)| \leq L|z_1 - z_2|$ for any $z_1, z_2 \in \mathbb{R}^n$.
- We usually consider a given ODE in differential form. However, there's no reason we can't consider the equivalent integral equation

$$y(t) = y_0 + \int_{t_0}^t f(\tau, y(\tau)) \mathrm{d}\tau$$

  - The reason for this change of perspective will become apparent shortly.

[1] $\bar{B}(0, M)$ denotes the set of all functions $y : [a, b] \to \mathbb{R}^n$ with sup norm at most M; topologically, it is the closed ball of radius M centered at the origin in $C[a, b]$.

- Let $\Phi : C[t_0, t_0 + a] \to C[t_0, t_0 + a]$ map functions to functions. Specifically, let it send
$$y(t) \mapsto y_0 + \int_{t_0}^{t} f(\tau, y(\tau))\mathrm{d}\tau$$
  - We denote this by writing $\Phi[y](t) = y_0 + \int_{t_0}^{t} f(\tau, y(\tau))\mathrm{d}\tau$.
  - Notice that the solution of our IVP is exactly the point of $C[t_0, t_0 + a]$ fixed by Φ because
$$y(t) = y_0 + \int_{t_0}^{t} f(\tau, y(\tau))\mathrm{d}\tau \quad\iff\quad y = \Phi[y]$$
  - *This* motivates all steps taken thus far.
- All that remains is to show that Φ is a contraction on some subset of $C[t_0, t_0 + a]$. Then we can apply the Banach fixed point theorem.
- We first identify this subset. Let
$$X_b = \{y : [t_0, t_0 + a] \to \bar{B}(y_0, b)\}$$
  - By the previous theorem, this is a complete metric space.
  - We now want to relate a and b so that $\Phi(X_b) \subset X_b$ and Φ is a contraction.
  - For $\Phi(X_b) \subset X_b$, we need
$$\|\Phi[y] - y_0\| \leq \int_{t_0}^{t_0+a} |f(\tau, y(\tau))|\mathrm{d}\tau \leq a \cdot M \leq b$$
    so we want $a < b/M$.
  - Moreover, if Φ is to be a contraction, then since
$$\begin{aligned}\|\Phi[y_1] - \Phi[y_2]\| &\leq \int_{t_0}^{t+a} |f(\tau, y_1(\tau)) - f(\tau, y_2(\tau))|\mathrm{d}\tau \\ &\leq La \cdot \|y_1 - y_2\|\end{aligned}$$
    we want $La \in (0, 1)$. We can achieve this by requiring $a < 1/2L$.
  - Thus, choosing
$$a < \min\left(\frac{1}{2L}, \frac{b}{M}\right)$$
    accomplishes all of our goals.
- Therefore, by the Banach fixed point theorem, there exists a unique y such that $y = \Phi[y]$.
  - As we have already remarked, this fixed point is exactly the aforementioned solution to the IVP.
- Conclusion:
- Theorem (Cauchy-Lipschitz theorem): Let $f(t, z)$ be defined on an open subset $\Omega \subset \mathbb{R}_t \times \mathbb{R}_z^n$, $(t_0, y_0) \in \Omega$, such that f is Lipschitz continuous wrt. t, z in some neighborhood of (t_0, y_0). Then the IVP $y'(t) = f(t, y(t))$, $y(t_0) = y_0$ has a unique solution for some $T > 0$ on $[t_0, t_0 + T]$ such that $y(t)$ does not escape that neighborhood.
- If $f \in C_1$, then
$$|f(t, z_1) - f(t, z_2)| \leq \sup_{z \in \bar{B}(y_0, r)} \left\|\frac{\partial f}{\partial z}(t, z)\right\| \cdot |z_1 - z_2|$$

- We use the finite increment theorem of differential calculus to prove that f is Lipschitz continuous if it's continuously differentiable.
- The norm on the RHS above is the matrix norm.
- We have $y(t) = \int_{t_0}^t f(\tau, y(\tau))\mathrm{d}\tau + y_0$ and we use the Banach fixed point theorem (which is proved constructively).
- We have $y_{n+1}(t) = y_0 + \int_{t_0}^t f(\tau, y_n(\tau))\mathrm{d}\tau$.
  - Thus, the Picard iteration is justified by the Banach fixed point theorem.
- We do not use the algorithm from the proof (the Picard iteration) computationally; we use the polygon algorithm. This algorithm is only of theoretical significance.
- **Interval of existence**: The union of intervals containing the interval $[t_0, t_0 + T]$ on which the IVP has a solution.
  - The interval of existence is always open. If $t_0 \in I$ such that $y(t_0) = y$, then $y'(t) = f(t, y(t))$, $y(t_0) = y$.
  - Note that the theorem does not predict when singularity can occur.
  - Example: The interval of existence will always be $x' = 1 + x^2$, $x(t_0) = x_0$. Then $x(t) = \tan(t - t_0 + \arctan(x_0))$. The length of existence is always π.
- Interval of existence: If you consider the IVP $y'(t) = f(t, y(t))$, $y(t_0) = y_0$, then $[t_0, t_0 + T_1], [t_0 + T_1, t_0 + T_2]$. The first is of length T_1, and the second of length $T_2 - T_1$. Continuing on, we get $T_n - T_1$ so that $T_n \to \infty$ or T_n is bounded. This gives us the maximal solution/interval of existence.
- The motherfucker (Shao) made us stay 10 minutes late.

Week 6

Qualitative Theory of ODEs

6.1 More Cauchy-Lipschitz and Intro to Continuous Dependence

10/31: • Last time, we built up a proof to the Cauchy-Lipschitz theorem intuitively.

  – We begin today with a direct proof that is very similar, but slightly different.

• Theorem (Cauchy-Lipschitz theorem): Let $f(t,z)$ be defined on an open subset $\Omega \subset \mathbb{R} \times \mathbb{R}^n$, let $(t_0, y_0) \in \Omega$, let $|f|$ be bounded on Ω, and let f be Lipschitz continuous in z and continuous wrt. t in some neighborhood of (t_0, y_0). Then the IVP $y'(t) = f(t, y(t))$, $y(t_0) = y_0$ has a unique solution on $[t_0, t_0 + T]$ for some $T > 0$ such that $y(t)$ does not escape Ω.

Proof. Let $f(t,z)$ be defined for $(t,z) \in [t_0, t_0 + a] \times \bar{B}(y_0, b) \subset \Omega$. Let $|f(t,z)| \leq M$. Let $|f(t,z_1) - f(t,z_2)| \leq L|z_1 - z_2|$ for all $z_1, z_2 \in \bar{B}(y_0, b)$.

Define $\{y_n\}$ recursively, starting from $y_0(t) = y_0$, by

$$y_{k+1}(t) = y_0 + \int_{t_0}^{t} f(\tau, y_k(\tau))\mathrm{d}\tau$$

Since f is continuous with respect to t, it is integrable with respect to t, so the above sequence is well-defined on $[t_0, t_0 + T]$. Choose $T = \min(a, b/M, 1/2L)$. Then

$$\|y_k - y_0\| \leq T \cdot M \leq \frac{b}{M} \cdot M = b$$

so no y_k escapes $\bar{B}(y_0, b)$. Additionally,

$$\begin{aligned}
\|y_{k+1} - y_k\| &\leq \int_{t_0}^{t} \|f(\tau, y_k(\tau)) - f(\tau, y_{k-1}(\tau))\|\mathrm{d}\tau \\
&\leq TL\|y_k - y_{k-1}\| \\
&\leq \frac{1}{2}\|y_k - y_{k-1}\| \\
&\leq \left(\frac{1}{2}\right)^k \|y_1 - y_0\|
\end{aligned}$$

Thus, the difference between successive terms in the sequence is controlled by a geometric progression, so $\{y_n\}$ is a Cauchy sequence in the function space. It follows that $\{y_k\}$ is uniformly convergent to some continuous $y : [t_0, t_0 + T] \to \mathbb{R}^n$. $\square$

• This completes the proof. Although it's more concrete than the contraction mapping one, they are virtually the same: In both cases, we obtain an approximate sequence controlled by a geometric progression.

- Examples of the Picard iteration:
  1. Consider an linear autonomous systems $y' = Ay$, A an $n \times n$ matrix, and $y(0) = y_0$.
     - We know that the solution is $y(t) = \mathrm{e}^{tA} y_0$. However, we can derive this using the Picard iteration.
     - Indeed, via this procedure, let's determine the first couple of Picard iterates.

$$y_0(t) = y_0 \qquad \begin{aligned} y_1(t) &= y_0 + \int_0^t A y_0(\tau) \mathrm{d}\tau \\ &= y_0 + tAy_0 \end{aligned} \qquad \begin{aligned} y_2(t) &= y_0 + \int_0^t A y_1(\tau) \mathrm{d}\tau \\ &= y_0 + tAy_0 + \frac{1}{2} t^2 A^2 y_0 \end{aligned}$$

     - It follows inductively that

$$y_k(t) = \sum_{j=0}^{k} \frac{t^j A^j}{j!} y_0$$

     - Since the term above is exactly the power series definition of e^{tA}, we have that $y_k(t) \to \mathrm{e}^{tA} y_0$ with local uniformity in t, as desired.
  2. Consider the ODE $y' = y^2$, $y(0) = 1$.
     - We know that the solution is $y(t) = 1/(1-t)$. We will now also derive this via the Picard iteration.
     - Choose $b = 1$, so that

$$\bar{B}(y_0, b) = \{y \mid |y - y(0)| \leq 1\} = \{y \mid |y - 1| \leq 1\} = [0, 2]$$

     - On this interval, $f(t, y) = y^2$ has maximum slope $L = 4$. Thus, we should take $T \leq 1/2L = 1/8$.
     - It follows that $|y_1^2 - y_2^2| \leq 4|y_1 - y_2|$ for all $y_1, y_2 \in \bar{B}(y_0, b)$.
     - Calculate the first few Picard iterates.

$$\begin{aligned} y_1(t) &= 1 + \int_0^t (y_0(\tau))^2 \mathrm{d}\tau = 1 + t \\ y_2(t) &= 1 + \int_0^t (1+\tau)^2 \mathrm{d}\tau = 1 + t + t^2 + \frac{t^3}{3} \\ y_3(t) &= 1 + \int_0^t \left(1 + \tau + \tau^2 + \frac{\tau^3}{3}\right)^2 \mathrm{d}\tau = 1 + t + t^2 + t^3 + \frac{2t^4}{3} + \frac{t^5}{3} + \frac{t^6}{9} + \frac{t^7}{63} \end{aligned}$$

     - It follows by induction that

$$\begin{aligned} |y_k(t) - (1 + t + \cdots + t^k)| &\leq t^{k+1} \\ \left| y_k(t) - \frac{1 - t^{k+1}}{1 - t} \right| &\leq t^{k+1} \end{aligned}$$

       It follows that $|t| < 1/8$.
     - For $|t| < 1/8$, $y(t) = 1/(1-t)$. Blows up as $t \to 1$.
     - Some more details on the bounding of the error term are presented in the lecture notes document.
- Lemma (Grönwall's inequality): Let $\varphi(t)$ be a real function defined for $t \in [t_0, t_0 + T]$ such that

$$\varphi(t) \leq f(t) + a \int_{t_0}^t \varphi(\tau) \mathrm{d}\tau$$

  Then

$$\varphi(t) \leq f(t) + a \int_{t_0}^t \mathrm{e}^{a(t-\tau)} f(\tau) \mathrm{d}\tau$$

Proof. Multiply both sides by e^{-at}:

$$\begin{aligned}
\mathrm{e}^{-at}\varphi(t) - a\mathrm{e}^{-at}\int_{t_0}^{t}\varphi(\tau)\mathrm{d}\tau &\leq \mathrm{e}^{-at}f(t) \\
\frac{\mathrm{d}}{\mathrm{d}t}\left(\mathrm{e}^{-at}\int_{t_0}^{t}\varphi(\tau)\mathrm{d}\tau\right) &\leq \mathrm{e}^{-at}f(t) \\
\mathrm{e}^{-at}\int_{t_0}^{t}\varphi(\tau)\mathrm{d}\tau &\leq \int_{t_0}^{t}\mathrm{e}^{-a\tau}f(\tau)\mathrm{d}\tau \\
\int_{t_0}^{t}\varphi(\tau)\mathrm{d}\tau &\leq \int_{t_0}^{t}\mathrm{e}^{a(t-\tau)}f(\tau)\mathrm{d}\tau
\end{aligned}$$

Substituting back into the original equality yields the result at this point. □

- Note that there is no sign condition on $f(t)$ or a.
- Grönwall's inequality is very important and we should remember it.
- It is also exactly what we need to prove continuous dependence.
- Theorem: Let $f(t,z), g(t,z)$ be defined on $\Omega \subset \mathbb{R}_t^1 \times \mathbb{R}_z^n$, an open and bounded a region containing (t_0, y_0) and (t_0, w_0). Let the functions be L - Lipschitz wrt. z. Consider two initial value problems $y' = f(t,y)$, $y(t_0) = y_0$ and $w' = g(t,w)$, $w(t_0) = w_0$. If $|f(t,z) - g(t,z)| < M$, then for $t \in [t_0, t_0 + T]$,

$$|y(t) - w(t)| \leq \mathrm{e}^{LT}|y_0 - w_0| + \frac{M}{L}(\mathrm{e}^{LT} - 1)$$

Proof. We have that

$$\begin{aligned}
|y(t) - w(t)| &= \left|\left[y_0 + \int_{t_0}^{t} f(\tau, y(\tau))\mathrm{d}\tau\right] - \left[w_0 + \int_{t_0}^{t} g(\tau, y(\tau))\mathrm{d}\tau\right]\right| \\
&= \left|[y_0 - w_0] + \int_{t_0}^{t}[f(\tau, y(\tau)) - g(\tau, y(\tau))]\mathrm{d}\tau\right| \\
&\leq |y_0 - w_0| + \left|\int_{t_0}^{t}[f(\tau, y(\tau)) - g(\tau, w(\tau))]\mathrm{d}\tau\right| \\
&\leq |y_0 - w_0| + \int_{t_0}^{t}|f(\tau, y(\tau)) - g(\tau, w(\tau))|\mathrm{d}\tau
\end{aligned}$$

where we get from the second to the third line using the triangle inequality, and the third to the fourth line using Theorem 13.26 of Honors Calculus IBL. We also know that

$$\begin{aligned}
|f(\tau, y(\tau)) - g(\tau, w(\tau))| &\leq |f(\tau, y(\tau)) - f(\tau, w(\tau))| + |f(\tau, w(\tau)) - g(\tau, w(\tau))| \\
&\leq L|y(\tau) - w(\tau)| + M
\end{aligned}$$

Combining what we've obtained, we have

$$\begin{aligned}
\underbrace{|y(t) - w(t)|}_{\psi(t)} &\leq \underbrace{|y_0 - w_0| + M(t - t_0)}_{f(t)} + \underbrace{L}_{a}\int_{t_0}^{t}\underbrace{|y(\tau) - w(\tau)|}_{\psi(t)}\mathrm{d}\tau \\
&\leq MT + |y_0 - w_0| + L\int_{t_0}^{t}\mathrm{e}^{L(t-\tau)}[|y_0 - w_0| + M(t - \tau)]\mathrm{d}\tau && \text{Grönwall} \\
&\leq \mathrm{e}^{LT}|y_0 - w_0| + \frac{M}{L}(\mathrm{e}^{TL} - 1)
\end{aligned}$$

as desired. □

- Note: Getting from directly from Grönwall's inequality in the second line above to the last line above is quite messy. A consequence of Grönwall's inequality explored in the book makes this much easier. *Prove Equation 2.38 via Problem 2.12.*
- Implication: The IVP is not just solvable itself, but is solvable wrt. perturbation of the initial conditions and RHS within a small, finite interval in time.
- Suppose $y' = 0$, $y(0) = 1$ and $w' = \varepsilon w$, $w(0) = 1$. Then $y(t) = 1$ and $w(t) = \mathrm{e}^{\varepsilon t}$ and solutions are only close when t is small.
  - $t \leq 1/\varepsilon$??
- This is important in physics. In most physical scenarios, the RHS is C^1. This is called determinism.

6.2 Differentiability With Respect To Parameters

11/2:

- Review: Implicit Function Theorem.
  - Gives you a sufficient condition for which an implicit relation defines a function.
  - Does not give you the function, but tells you that it must exist and that it is unique.
- Theorem (Implicit Function Theorem): Let $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ be C^k in some neighborhood of $(x_0, y_0) \in \mathbb{R}^n \times \mathbb{R}^m$ a point satisfying $F(x_0, y_0) = 0$. If the truncated Jacobian matrix $\frac{\partial F}{\partial y}(x_0, y_0)$, which is $m \times m$, is invertible, then there is a neighborhood U of x_0 such that there is a unique function $f : U \to \mathbb{R}^m$ with $y_0 = f(x_0)$ and $F(x, f(x)) = 0$ and

$$f'(x) = -\left(\frac{\partial F}{\partial y}(x, y)\right)^{-1} \cdot \frac{\partial F}{\partial x}(x, f(x))$$

  - The proof is based on the Banach fixed point theorem (this may be false?? I think Shao is confusing the proof of this theorem with the proof of the Inverse Function Theorem).
  - The motivation for the last equality (the line above) is that if $F(x, f(x)) = 0$, then by the chain rule for partial derivatives,

$$\begin{aligned} 0 &= \frac{\mathrm{d}}{\mathrm{d}x}(F(x, f(x))) \\ &= \frac{\partial F}{\partial x}(x, f(x)) \cdot \frac{\mathrm{d}x}{\mathrm{d}x} + \left[\frac{\partial F}{\partial y}(x, y)\right] \cdot \frac{\mathrm{d}f}{\mathrm{d}x} \\ &= \frac{\partial F}{\partial x}(x, f(x)) + \left[\frac{\partial F}{\partial y}(x, y)\right] \cdot f'(x) \\ f'(x) &= -\left(\frac{\partial F}{\partial y}(x, y)\right)^{-1} \cdot \frac{\partial F}{\partial x}(x, f(x)) \end{aligned}$$

  - Recall that we know that the matrix bracketed in line 2 is invertible by hypothesis.
  - Additionally, since $\partial F/\partial x = A$ is $n \times m$ and $\partial F/\partial y = B$ is $m \times m$, $f' = -A^{-1}B$ is $n \times m$, as it should be for a function $f : \mathbb{R}^n \to \mathbb{R}^m$.
- Consider the IVP

$$y' = f(t, y; \mu), \quad y(t_0) = x(\mu)$$

  - This ODE and its initial condition both depend on a parameter $\mu \in B(0, r) \subset \mathbb{R}^m$ (usually we take $m = 1$ so μ is just real).
  - We denote the solution by $y(t; \mu)$.

  - Suppose $|x(\mu)| < C$ for $\mu \in B(0,r)$ and $x(\mu) \in C^1$. Suppose the RHS $f(t,z;\mu)$ of the ODE is defined on $[t_0, t_0+a] \times \bar{B}(x(0), b+C) \times B(0,r)$, is C^1 in all variables, is bounded by M on its domain, and is L-Lipschitz in z.

- By Cauchy-Lipschitz, for small
$$T \leq \min\left(a, \frac{b}{M}, \frac{1}{2L}\right)$$
and $\mu \in B(0,r)$ (r small), the solution *exists* on $[t_0, t_0+T]$ and its value does not escape $\bar{B}(x(0), b+C)$.
  - We now aim to show that the solution is *differentiable* wrt. μ on this interval.
- If $y(t;\mu)$ satisfies $y'(t;\mu) = f(t, y(t;\mu); \mu)$ and if the Jacobian matrix $J = \partial y/\partial\mu$ exists, then J satisfies the **first variation equation**.
- **First variation equation**: The following linear differential equation. *Given by*
$$\frac{\mathrm{d}}{\mathrm{d}t}\underbrace{\frac{\partial y}{\partial\mu}(t;\mu)}_{J(t;\mu)} = \underbrace{\frac{\partial f}{\partial z}(t, y(t;\mu);\mu)}_{A(t;\mu)} \cdot \underbrace{\frac{\partial y}{\partial\mu}(t;\mu)}_{J(t;\mu)} + \frac{\partial f}{\partial\mu}(t, y(t;\mu);\mu), \quad \frac{\partial y}{\partial\mu}(t_0,\mu) = \frac{\partial x}{\partial\mu}(\mu)$$
- The first variation equation has a unique solution, but we do not yet know that $y(t;\mu)$ is even differentiable with respect to μ. We presently verify this claim.
- Theorem[1]: $y(t;\mu)$ is C^1 in μ and $\partial y/\partial\mu\,(t;\mu)$ satisfies the first variation equation.

  Proof. Let $\Theta(t;\mu) = y(t;\mu+h) - y(t;\mu) - J(t;\mu)h$ for h small. Aim, show that $\Theta(t;\mu) = o(h)$ as $h \to 0$.

  We compute
$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\Theta(t;\mu) &= y'(t;\mu+h) - y'(t;\mu) - J'(t;\mu)h \\ &= \underbrace{f(t, y(t;\mu+h);\mu+h) - f(t, y(t;\mu);\mu)}_{I} - \underbrace{\frac{\partial f}{\partial z}(t, y(t;\mu);\mu)J(t;\mu) + \frac{\partial f}{\partial\mu}(t, y(t;\mu);\mu)}_{II}\end{aligned}$$

  I denotes the first term; II denotes the second term.

  We have that
$$I = \frac{\partial f}{\partial z}(t, y(t;\mu);\mu)[y(t;\mu+h) - y(t;\mu)] + \frac{\partial f}{\partial\mu}(t, y(t;\mu);\mu)h + \underbrace{R(t;\mu,h)}_{o(h)}$$

  color coding
$$\begin{aligned}I - II &= \underbrace{\text{green} - \text{blue}}_{\Theta(t;\mu)} + R(t;\mu,h) \\ &= \frac{\mathrm{d}}{\mathrm{d}t}\Theta(t;\mu) = \Theta(t;\mu) + \underbrace{R(t;\mu,h)}_{o(h)} \\ \Theta(t_0;\mu) &= o(h) \\ |\Theta(t;\mu)| &\leq C\int_{t_0}^{t} |R(\tau;\mu,h)|\mathrm{d}\tau && \text{Grönwall} \\ &= o(h)\end{aligned}$$

  circle terms cancel. □

[1]See the proof from the book, transcribed below.

- Example: First order derivatives must satisfy the first variational equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial y}{\partial \mu}(t;\mu) = \frac{\partial f}{\partial z}(t, y(t;\mu);\mu) \cdot \frac{\partial y}{\partial \mu}(t;\mu)$$

and the second order derivative must satisfy the second variational equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial^2 y}{\partial \mu^2} = \frac{\partial^2 f}{\partial z^2}\left(\frac{\partial y}{\partial \mu}\frac{\partial^2 y}{\partial \mu^2}\right) + \frac{\partial^2 f}{\partial z \partial \mu}\frac{\partial y}{\partial \mu} + \frac{\partial^2 f}{\partial \mu \partial z}(-)\frac{\partial y}{\partial \mu} + \frac{\partial^2 f}{\partial \mu^2}(-)$$

- Corollary: If $f(t,z;\mu)$ is C^k in (t,z,μ), $y(t_0) = x(\mu)$ is C^k, then $y(t;\mu)$ is C^k in μ.
- The Taylor expansion

$$y(t;\mu) = y(t;0) + y_1\mu + y_2\mu^2 + \cdots + y_k\mu^k + O(\mu^{k+1})$$

of $y(t;\mu)$ about 0 gives an approximation of said function up to order k in μ.

  - Misc notes: but you can cut off the expansion at k?? $y(t;0)$ being solvable implies inductively that the rest are solvable??
  - We can take this Taylor expansion because we assume that y is continuously differentiable k times with respect to μ.
  - The coefficients y_j are given as follows.

$$y_j = \frac{1}{j!}\frac{\partial^j y}{\partial \mu^j}\ (t;0)$$

- Application of the Taylor expansion: It can be substituted into the ODE as follows.

$$\begin{aligned} \dot{y} &= f(t,y;\mu) \\ \frac{\mathrm{d}}{\mathrm{d}t}(y(t;\mu)) &= f(t,y(t;\mu);\mu) \\ \frac{\mathrm{d}}{\mathrm{d}t}(y(t;0)) + \frac{\mathrm{d}y_1}{\mathrm{d}t}\mu + \cdots + \frac{\mathrm{d}y_k}{\mathrm{d}t}\mu^k + O(\mu^{k+1}) &= f(t, y(t;0) + y_1\mu + \cdots + y_k\mu^k + O(\mu^{k+1});\mu) \end{aligned}$$

  - Then you can match coefficients of the various μ terms on the LHS and RHS and solve for $y_0, \dots, y_k$.
  - When to use this method: Sometimes, you can view equations that aren't explicitly solvable as perturbations of an easily solvable system.

- Simple example (more complex ones next lecture):

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \mu y, \quad y(0) = 1$$

  - First off, we know that there is an explicit solution ($y(t) = \mathrm{e}^{\mu t}$). Thus, we will be able to check our final answer.
  - Suppose $y \in C^2$ with respect to μ. Then

$$y(t;\mu) = y_0 + y_1\mu + y_2\mu^2 + O(\mu^3)$$

  - It follows by substituting into the above differential equation that

$$\begin{aligned} \frac{\mathrm{d}y}{\mathrm{d}t} &= \mu y \\ \frac{\mathrm{d}}{\mathrm{d}t}\left(y_0 + y_1\mu + y_2\mu^2\right) &= \mu(y_0 + y_1\mu + y_2\mu^2) \\ \frac{\mathrm{d}y_0}{\mathrm{d}t} + \frac{\mathrm{d}y_1}{\mathrm{d}t}\mu + \frac{\mathrm{d}y_2}{\mathrm{d}t}\mu^2 &= 0 + y_0\mu + y_1\mu^2 + y_2\mu^3 \end{aligned}$$

  - By comparing coefficients, this yields the sequentially solvable differential equations

$$\frac{\mathrm{d}y_0}{\mathrm{d}t} = 0 \qquad \frac{\mathrm{d}y_1}{\mathrm{d}t} = y_0 \qquad \frac{\mathrm{d}y_2}{\mathrm{d}t} = y_1$$

    where we apply the initial condition $y_0(0) = 1$ to solve the left ODE above.
  - Solving, we get

$$y_0(t) = 1 \qquad y_1(t) = t \qquad y_2(t) = \frac{t^2}{2}$$

    - Where do the other initial conditions (all zero) come from??
  - Therefore, our approximate solution is

$$y(t) = 1 + t\mu + \frac{1}{2}t^2\mu^2 + O(\mu^3)$$

    which does indeed give the first three terms in the Taylor series expansion of the solution $\mathrm{e}^{\mu t}$.

- The perturbative solution fails in large time intervals — polynomials inevitably grow slower than exponential functions.
- Next time: Several examples applying what we've learned today.
- This week's homework: Some basic Lipschitz definitions and also computations with the perturbative series.

6.3 Variational Examples

11/4: - We begin today with a more direct and less involved proof of the variation of parameters theorem.

Proof. Let $y'(t;\mu) = f(t, y(t;\mu);\mu)$ with $y(t_0;\mu) = x(\mu)$. Assume Lipschitz continuity and C^1-ness of the ODE and the initial condition on μ. Then differentiation with respect to μ must satisfy the first variational equation. In particular, let $J(t;\mu)$ be the solution of

$$J'(t;\mu) = \underbrace{\frac{\partial f}{\partial z}(t, y(t;\mu);\mu)}_{A(t;\mu)} J(t,\mu) + \underbrace{\frac{\partial f}{\partial \mu}(t, y(t;\mu);\mu)}_{F(t;\mu)}, \quad J(t_0;\mu) = \frac{\partial x}{\partial \mu}$$

Consider the Picard iteration sequence defined by

$$y_{n+1}(t;\mu) = \underbrace{f(t; y_n(t;\mu);\mu)}_{A_n(t;\mu)}, \quad y_n(t_0,\mu) = x(\mu)$$

Differentiating we get

$$\frac{\partial y_n}{\partial \mu}(t;\mu)$$

which we may call $J_n(t;\mu)$. We want to prove that the sequence of functions J_n converges uniformly to J. This makes sense since A and F uniformly converge. Moreover, under this definition of J_n, we have that

$$J_{n+1}'(t;\mu) = \underbrace{\frac{\partial f}{\partial z}(t, y_n(t;\mu);\mu)}_{A_n(t;\mu)} J_n(t;\mu) + \underbrace{\frac{\partial f}{\partial \mu}(t, y_n(t;\mu);\mu)}_{F_n(t;\mu)}, \quad J_n(t_0;\mu) = \frac{\partial x}{\partial \mu}(\mu)$$

Thus, Step 1 is to show that $\{\|J_n\|\}$ is bounded on $[t_0, t_0 + T]$. To do so, we note that

$$\|J_{n+1}\| \leq \frac{1}{2}\|J_n\| + \sup\left|\frac{\partial f}{\partial \mu}\right|$$

so that $\|J_n\|$ forms a bounded sequence. By induction,

$$\|J_n\| \leq 2C$$

We now embark on Step 2: Proving $J_n \to J$ uniformly. First off, we have that

$$\begin{aligned}
(J - J_{n+1})'(t;\mu) &= \frac{\mathrm{d}}{\mathrm{d}t}(J(t;\mu) - J_{n+1}(t;\mu)) \\
&= A(t;\mu)J(t;\mu) + F(t;\mu) - A_n(t;\mu)J_n(t;\mu) - F_n(t;\mu) \\
&= A(t;\mu)J(t;\mu) + A_n(t;\mu)J(t;\mu) - A_n(t;\mu)J(t;\mu) \\
&\quad - A_n(t;\mu)J_n(t;\mu) + F(t;\mu) - F_n(t;\mu) \\
&= A_n(t;\mu)(J - J_n)(t;\mu) + (A - A_n)(t;\mu)J(t;\mu) + (F - F_n)(t;\mu)
\end{aligned}$$

and

$$J(t_0;\mu) - J_{n+1}(t_0;\mu) = 0$$

Integrating once again on $[t_0, t_0 + T]$, we get

$$\|J - J_{n+1}\| \leq \frac{1}{2}\|J - J_n\| + \delta_n$$

where $\delta_n \to 0$ since we "obviously" have that $A_n \to A$ and $F_n \to F$ uniformly.

We now proceed via a standard analysis argument. Fix $\delta > 0$, choose N such that $\delta_n < \delta$ for $n \geq N$. Then we can control it by $\frac{1}{2}\|J - J_n\| + \delta$ for $n \geq N$. Then

$$\|J - J_{n+1}\| - 2\delta \leq \frac{1}{2}\|J - J_n\| - 2\delta$$

for all $n \geq N$, so we have by iteration that $\|J - J_{n+1}\| \leq 2\delta + \frac{1}{2^{n-N}}\|J - J_N\|$, so $\lim_{n\to\infty}\|J - J_n\| < 2\delta$ for arbitrary $\delta > 0$. Therefore, $\|J - J_n\| \to 0$, so $J_n \to J$ uniformly.

So in conclusion, $J_n \to J$ uniformly and we recall that $J_n = \partial y_n/\partial\mu$ where $y_n \to y$ uniformly. $\square$

- We now look at examples. The ones in the HW will be no more difficult than these.
- Example (same one as last time):
  - Consider $y' = \mu y$ with $y(0) = 1$.
  - In order to find asymptotic expansion wrt. μ, we use the **ansatz** $y(t;\mu) = y_0 + y_1\mu + y_2\mu^2 + \cdots + y_n\mu^n + O(\mu^{n+1})$.
    - The differentiation theorem asserts that $y(t;\mu)$ can be differentiated wrt. μ so many times.
  - We can compute
    $$\mu y(t;\mu) = 0 + y_0\mu + y_1\mu^2 + \cdots + y_{n-1}\mu^n + O(\mu^{n+1})$$
    and
    $$y'(t;\mu) = y_0' + y_1'\mu + y_2'\mu^2 + \cdots + y_n'\mu^n + O(\mu^{n+1})$$
    and set them equal to yield a system of differential equations.
  - The initial conditions are $y_0(0) = 1$ and then $y_1(0) = \cdots = y_n(0) = 0$.
  - $y_0' = 0$ with $y_0(0) = 1$ implies that $y_0(t) = 1$.
  - Then the first order approximation is $y_1' = y_0 = 1$, so solving and applying the initial conditions, we get $y_1(t) = t$.
  - Continuing on, the second order approximation is $y_2(t) = t^2/2$.
  - Inductively, $y_m(t) = t^m/m!$.
  - In conclusion, we obtain the desired approximate solution.

- **Ansatz**: The form of the solution that you guess.
- In general, this shows the technique well: Use a polynomial ansatz and compare terms to yield an inductive sequence of explicitly solvable equations up to a certain point.
- Example: Mathematical pendulum.
  - Suppose that the length of the rope is ℓ and the gravitational acceleration is g. Then
  $$\theta''(t;\mu) = -\frac{g}{\ell}\sin[\theta(t;\mu)]$$
  - Assume a small angle, $\theta(0) = \mu$ and $\theta'(0) = 0$.
  - Substitute $\omega_0^2 = g/\ell$.
  - In HS, we learned that the harmonic oscillator approximation of the mathematical pendulum is justified for small θ. We now justify this.
  - Ansatz: $\theta_0 + \theta_1\mu + \theta_2\mu^2 + \theta_3\mu^3 + O(\mu^4)$.
  - Recall that
  $$\sin\theta = \theta - \frac{\theta^3}{6} + O(\theta^5)$$
  - First step, solve to determine $\theta_0 = 0$.
  - Then we only have a term of order $O(\mu)$ and $O(\mu^3)$ to worry about.
  - Substitute the expansion in:
  $$\begin{aligned}\sin\theta &= \theta - \frac{\theta^3}{6} + O(\theta^5)\\ &= \left(\theta_0 + \theta_1\mu + \theta_2\mu^2 + \theta_3\mu^3\right) - \frac{1}{6}\left(\theta_0 + \theta_1\mu + \theta_2\mu^2 + \theta_3\mu^3\right)^3\\ &= 0 + \theta_1\mu + \theta_2\mu^2 + \left(\theta_3 - \frac{\theta_1^3}{6}\right)\mu^3 + O(\mu^4)\end{aligned}$$
  - We also have that
  $$\theta''(t;\mu) = \theta_1''\mu + \theta_2''\mu^2 + \theta_3''\mu^3 + O(\theta^4)$$
  and
  $$-\omega_0^2\sin\left(\theta_1\mu + \theta_2\mu^2 + \theta_3\mu^3 + O(\mu^4)\right) = -\omega_0^2\theta_1\mu - \omega_0^2\theta_2\mu^2 - \omega_0^2\left(\theta_3 - \frac{\theta_1^3}{6}\right)\mu^3 + O(\mu^4)$$
  - Initial conditions: $\theta_0 = 0$, $\theta_1(0) = 1$, and $\theta_2(0) = \theta_3(0) = \theta_1'(0) = \cdots = \theta_3'(0) = 0$.
  - First order: $\theta_1'' = -\omega_0^2\theta_1$, $\theta_1(0) = 1$, $\theta_1'(0) = 0$. Implies $\theta_1(t) = \cos\omega_0 t$. This is why we can use the harmonic oscillator approximation.
  - Second order: $\theta_2 = -\omega_0^2\theta_2$. Initial conditions imply $\theta_2(t) = 0$.
  - Third order: $\theta_3'' = -\omega_0^2\theta_3 + \frac{\omega_0^2\theta_1^3}{6}$. Implies that
  $$\theta_3(t) = \frac{\omega_0 t}{16}\sin\omega_0 t + \frac{1}{192}(\cos\omega_0 t - \cos 3\omega_0 t)$$
    - We have to apply some trigonometric identities to verify this??
  - In conclusion, we have the approximation of our solution up to order $O(\mu^3)$ as
  $$\theta(t;\mu) = \mu\cos\omega_0 t + \mu^3\left[\frac{\omega_0 t}{16}\sin\omega_0 t + \frac{1}{192}(\cos\omega_0 t - \cos 3\omega_0 t)\right] + O(\mu^4)$$
    - This approximation is only good for T in a fixed, small time interval because the second term is not periodic.

- We now investigate the period of the mathematical pendulum.
  - The first order approximation (harmonic oscillator) gives the period as $T \approx 2\pi/\omega_0 = 2\pi\sqrt{\ell/g}$.
  - Let $T(\mu)$ denote the period of the mathematical pendulum as a function of the starting angle μ.
  - $T(\mu)$ should be approximately equal to the period of $\theta(t;\mu)$. Additionally, thinking about the mathematical pendulum intuitively, the period $T(\mu)$ should be about four times the first positive zero of $\theta(t;\mu)$.
    - Indeed, in a full cycle, the pendulum must go from the positive extreme, to zero, to the negative extreme, back to zero, and back to the original position, so there are our four parts.
    - Example: In the harmonic oscillator approximation, the first zero is at $\pi/2\omega_0$, and the period is $2\pi/\omega_0 = 4 \cdot \pi/2\omega_0$.
  - Thus, determining the period $T(\mu)$ becomes a problem of finding t such that $\theta(t;\mu) = 0$.
  - The zeroes of $\theta(t;\mu)$ will be equal to the zeroes of $\theta(t;\mu)/\mu$, so we seek t such that the implicit function

$$F(t;\mu) = \frac{\theta(t;\mu)}{\mu} = \cos\omega_0 t + \mu^2\left[\frac{\omega_0 t}{16}\sin\omega_0 t + \frac{1}{192}(\cos\omega_0 t - \cos 3\omega_0 t)\right] = 0$$

  - When $\mu = 0$, the mathematical pendulum is stationary, but this does technically mean that it has a zero at $(\pi/2\omega_0;0)$. This point is important because for μ small enough that the harmonic oscillator approximation is good, the first zero should be very close to $\pi/2\omega_0$. Thus, we choose to solve $F(t;\mu) = 0$ around $(t_0;\mu_0) = (\pi/2\omega_0;0)$.
  - The requirement for the Implicit Function Theorem is met since

$$\begin{aligned}
\frac{\partial F}{\partial t}(t_0;\mu_0) &= -\omega_0\sin\omega_0 t_0 + \mu_0^2\left(\frac{\omega_0}{16}\sin\omega_0 t_0 + \frac{\omega_0^2 t_0}{16}\cos\omega_0 t_0 + \frac{1}{192}(-\omega_0\sin\omega_0 t_0 + 3\omega_0\sin 3\omega_0 t_0)\right)\\
&= -\omega_0\sin\frac{\pi}{2} + 0^2(\dots)\\
&= -\omega_0\\
&\neq 0
\end{aligned}$$

  - Thus, there exists $t_1(\mu)$ smooth defined on some neighborhood of $\mu_0 = 0$ satisfying $t_1(0) = \pi/2\omega_0$ and $F(t_1(\mu);\mu) = 0$.
  - We cannot (easily??) obtain $t_1(\mu)$ directly, so we will look for its second-order Taylor expansion

$$t_1(\mu) = \frac{\pi}{2\omega_0} + b_1\mu + b_2\mu^2 + O(\mu^3)$$

  - We need not compute a bunch of derivatives to find b_1, b_2, though. Indeed, we can just substitute into $F(t_1(\mu);\mu) = 0$ and compare different powers of μ. Doing so, we obtain

$$\begin{aligned}
0 &= F(t_1(\mu);\mu)\\
&= \cos\left(\frac{\pi}{2} + \omega_0 b_1\mu + \omega_0 b_2\mu^2 + O(\mu^3)\right)\\
&\quad + \mu^2\left[\frac{1}{16}\left(\frac{\pi}{2} + \omega_0 b_1\mu + \omega_0 b_2\mu^2 + O(\mu^3)\right)\sin\left(\frac{\pi}{2} + \omega_0 b_1\mu + \omega_0 b_2\mu^2 + O(\mu^3)\right)\right.\\
&\quad \left. + \frac{1}{192}\left(\cos\left(\frac{\pi}{2} + \omega_0 b_1\mu + \omega_0 b_2\mu^2 + O(\mu^3)\right) - \cos 3\left(\frac{\pi}{2} + \omega_0 b_1\mu + \omega_0 b_2\mu^2 + O(\mu^3)\right)\right)\right]\\
&= -\omega_0 b_1\mu + \left(\frac{\pi}{32} - \omega_0 b_2\right)\mu^2 + O(\mu^3)
\end{aligned}$$

  from which we can determine that

$$\begin{aligned}
0 &= -\omega_0 b_1 & \qquad 0 &= \frac{\pi}{32} - \omega_0 b_2\\
b_1 &= 0 & b_2 &= \frac{\pi}{32\omega_0}
\end{aligned}$$

- Thus,

$$
\begin{aligned}
T(\mu) &= 4 \cdot t_1(\mu) \\
&= \frac{2\pi}{\omega_0} + \frac{\pi}{8\omega_0}\mu^2 + O(\mu^3) \\
&= 2\pi\sqrt{\frac{\ell}{g}}\left(1 + \frac{1}{16}\mu^2 + O(\mu^3)\right)
\end{aligned}
$$

- We calculate an accumulation that is a perturbation of an ODE in the bonus this week, reproducing Einstein's work.