18.03 LECTURE NOTES, SPRING 2018

BJORN POONEN

These are an approximation of what was covered in lecture. (Please clear your browser's cache before reloading this file to make sure you are getting the current version.) This PDF file is divided into sections; the instructions for viewing the table of contents depend on which PDF viewer you are using.

If your PDF viewer is linked to a browser, you should be able to click on URLs, such as the one below to go to the online mathlets:

`http://mathlets.org/mathlets/`

Small text contains a technical explanation that you might want to ignore when reading for the first time.

February 7

1. Solutions to differential equations

1.1. Introduction. A differential equation (DE) is an equation relating an unknown function and some of its derivatives. DEs arise in engineering, chemistry, biology, physics, economics, etc., because many laws of nature describe the instantaneous rate of change of a quantity in terms of current conditions.

Overall goals: to learn how to

- model real-world problems with DEs.
- solve DEs exactly when possible, or else solve numerically (get an approximate solution)
- extract qualitative information from a DE, whether or not it can be solved.

(There is ongoing research on these questions. For instance, there is a $1,000,000 prize for understanding the solutions to the Navier–Stokes equations modeling fluid flow.)

1.2. Notation and units. (To be done in recitation on Feb. 6.)

Notation for higher derivatives of a function $y(t)$:

$$\begin{aligned}
\text{first derivative:}\quad & \dot{y},\ y',\ \frac{dy}{dt} \\
\text{second derivative:}\quad & \ddot{y},\ y'',\ \frac{d^2y}{dt^2} \\
\text{third derivative:}\quad & \dddot{y},\ y^{(3)},\ \frac{d^3y}{dt^3} \\
& \vdots \\
n^{\text{th}}\text{ derivative:}\quad & y^{(n)}
\end{aligned}$$

Warning: The notation $\dot{y}$ is a standard abbreviation for $\frac{dy}{dt}$; use it only for the derivative with respect to *time*. If y is a function of x, write y' or $\frac{dy}{dx}$ instead.

If y has units m (meters) and t has units s (seconds), then $\frac{dy}{dt}$ has the same units as $\frac{y}{t}$ would, namely m/s (meters per second). Similarly, $\ddot{y}$ would have units m/s^2.

1.3. **A secret function.**

Example: Can you guess my secret function $y(t)$?

Clue: It satisfies the DE

$$\dot{y} = 3y. \tag{1}$$

(This might model population growth in some biological system.)

Maybe you guessed $y = e^{3t}$. This is *a* solution to the differential equation (1), because substituting it into the DE gives $3e^{3t} = 3e^{3t}$. But it's not the function I was thinking of! Some other solutions are $y = 7e^{3t}$, $y = -5e^{3t}$, $y = 0$, etc. Later we'll explain why the general solution to (1) is

$$y = ce^{3t}, \qquad \text{where } c \text{ is a parameter;}$$

saying this means that

- for each number c, the function $y = ce^{3t}$ is a solution, and
- there are no other solutions besides these.

So there is a 1-parameter family of solutions to (1).

You still haven't guessed my secret function.

Clue 2: My function satisfies the initial condition $y(0) = 6$.

Solution: There is a number c such that $y(t) = ce^{3t}$ holds for all t; we need to find c. Plugging in $t = 0$ shows that $6 = ce^0$, so $c = 6$. Thus, among the infinitely many solutions to the DE, the particular solution satisfying the initial condition is $y(t) = 6e^{3t}$. $\square$

Important: Checking a solution to a DE is usually easier than finding the solution in the first place, so it is often worth doing. Just plug in the function to both sides, and also check that it satisfies the initial condition.

In (1) only one initial condition was needed, since only one parameter c needed to be recovered.

1.4. **Classification of differential equations.** There are two kinds:

- Ordinary differential equation (ODE): involves derivatives of a function of *only one* variable.
$$\ddot{y} = -9y \qquad \text{(solve for } y(t)\text{)}$$
- Partial differential equation (PDE): involves *partial derivatives* of a *multivariable* function.
$$\frac{\partial u}{\partial t} = 9\,\frac{\partial^2 u}{\partial x^2} \qquad \text{(solve for } u(x,t)\text{)}$$

Order of an ODE: the highest n such that the n^{th} derivative of the function appears.

(The definition of order for a PDE is similar; just know that a term like $\frac{\partial^2 u}{\partial x^2}$ or $\frac{\partial^2 u}{\partial x\,\partial t}$ counts as a 2^{nd} derivative since it is a partial derivative of a partial derivative.)

Example 1.1. Is

$$\begin{aligned} &707099375\cos(t^5)\,\ddot{y}^4 + 3487980982(y+t^3)^7\,\dot{y} \\ &- 389750387y^{(3)}y^{(4)} + 2ty^{(5)} + 845372305472398573 0987 \\ &= 80970874y^6 - 2809754087\sin(t/y) + 8957092\ln(1-t^7) \\ &+ 64893745723786e^{y^8-t^3} + 987t^6 + 543y^2 + 18.03\ ? \end{aligned}$$

an ODE or a PDE? ODE.

Flashcard question: *What is its order?*

The order is 5, because the highest derivative that appears is the 5^{th} derivative, $y^{(5)}$.

2. Modeling

I sometimes tell people that I have a career in modeling. We're going to talk about mathematical modeling, which is converting a real-world problem into mathematical equations.

Guidelines:

1. Identify relevant quantities, both known and unknown, and give them symbols. Find the units for each.
2. Identify the independent variable(s). The other quantities will be functions of them, or constants. Often *time* is the only independent variable.

3. Write down equations expressing how the functions change in response to small changes in the independent variable(s). Also write down any "laws of nature" relating the variables. As a check, make sure that each summand in an equation has the same units.

Often simplifying assumptions need to be made; the challenge is to simplify the equations so that they can be solved but so that they still describe the real-world system well.

2.1. **Example: savings account.**

Problem 2.1. I have a savings account earning interest compounded daily, and I make frequent deposits or withdrawals into the account. Find an ODE with initial condition to model the balance.

Simplifying assumptions: Daily compounding is almost the same as continuous compounding, so let's assume that interest is paid continuously instead of at the end of each day. Similarly, let's assume that my deposits/withdrawals are frequent enough that they can approximated by a continuous money flow at a certain rate, the net deposit rate (which is negative when I am withdrawing). Finally, let's assume that the interest rate and net deposit rate vary continuously with time, but do not depend on the balance.

Variables and functions (with units): Define the following:

P : the principal, the initial amount that the account starts with (dollars)

t : time from the start (years)

x : balance (dollars)

I : the interest rate (year^{-1}) (e.g., 4%/year = 0.04 year^{-1})

q : the net deposit rate (dollars/year).

Here t is the independent variable, P is a constant, and x, I, q are functions of t.

Equations: During a time interval $[t, t+dt]$ for an "infinitesimally small" increment dt, the following hold (technically speaking, dt is a differential; if dt were replaced by a positive number Δt, then the equations below would be only approximations, but when we divide by Δt and take a limit, the end result is the same):

$$\text{interest earned } \textit{per dollar} = I(t)\,dt$$

$$\text{interest earned} = I(t)x(t)\,dt \quad \text{(asked as flashcard question)}$$

$$\text{amount deposited into the account} = q(t)\,dt$$

so

$$dx = \text{change in balance} = I(t)x(t)\,dt + q(t)\,dt$$

$$\frac{dx}{dt} = I(t)x(t) + q(t).$$

(Check: the units in each of the three terms are dollars/year.) Also, there is the initial condition $x(0) = P$. Thus we have an ODE with initial condition:

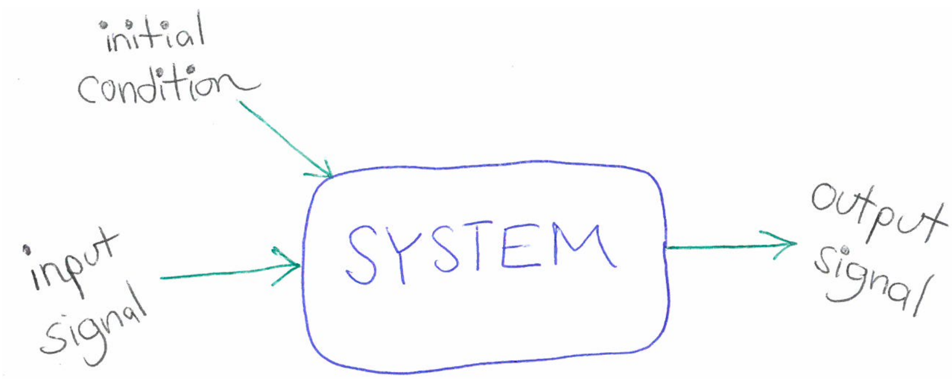
$$\dot{x} = I(t)x + q(t), \qquad x(0) = P. \quad \square \tag{2}$$

Now that the modeling is done, the next step might be to *solve* (2) for the function $x(t)$, but we won't do that yet.

2.2. **Systems and signals.** Maybe for financial planning I am interested in testing different saving strategies (different functions q) to see what balances x they result in. To help with this, rewrite the ODE as

$$\underset{\text{controlled by bank}}{\dot{x} - I(t)x} = \underset{\text{controlled by me}}{q(t)}.$$

In the "systems and signals" language of engineering, q is called the *input signal*, the bank is the *system*, and x is the *output signal*. These terms do not have a mathematical meaning dictated by the DE alone; their interpretation is guided by what is being modeled. But the general picture is this:



- The input signal is a function of the independent variable alone, a function that enters into the DE somehow (usually the right side of the DE, or part of the right side).
- The system processes the input signal by solving the DE with the given initial condition.
- The output signal (also called system response) is the solution to the DE.

3. Separation of variables for first-order ODEs

(Done in recitation.)

Separation of variables is a technique that quickly solves some simple first-order ODEs. Here is how it works:

1. Check that the DE is a *first-order ODE*. (If not, give up and try another method.) Suppose that the function to be solved for is $y = y(t)$.
2. Rewrite $\dot{y}$ as $\dfrac{dy}{dt}$.

3. Add and/or subtract to move terms to the other side of the DE so that the term with $\frac{dy}{dt}$ is on the left and all other terms are on the right.
4. Try to separate the y's and t's. Specifically, try to multiply and/or divide (and in particular move the dt to the right side) so that it ends up as an equality of differentials of the form
$$f(y)\,dy = g(t)\,dt.$$
Note: If there are factors involving both variables, such as $y+t$, then it is impossible to separate variables; in this case, give up and try a different method.
Warning: Dividing the equation by an expression invalidates the calculation if that expression is 0, so at the end, check what happens if the expression is 0; this may add to the list of solutions.
5. Integrate both sides to get an equation of the form
$$F(y) = G(t) + C.$$
These are implicit equations for the solutions, in terms of a parameter C.
6. If possible (and if desired), solve for y in terms of t.
7. Check for extra solutions coming from the warning in Step 4. The solutions in the previous step and this step comprise the general solution.
8. (Optional, but recommended) Check your work by verifying that the general solution actually satisfies the original DE.

Problem 3.1. Solve $\dot{y} - 2ty = 0$.

Solution:

Step 1. This involves only the *first* derivative of a *one-variable* function $y(t)$, so it is a first-order ODE. Thus we can *attempt* separation of variables.

Step 2. Rewrite as $\frac{dy}{dt} - 2ty = 0$.

Step 3. Isolate the $\frac{dy}{dt}$ term: $\frac{dy}{dt} = 2ty$.

Step 4. We can separate variables! Namely, $\frac{1}{y}\,dy = 2t\,dt$. (Warning: We divided by y, so at some point we will have to check $y = 0$ as a potential solution.)

Step 5. Integrate: $\ln|y| = t^2 + C$.

Step 6. Solve for y:
$$|y| = e^{t^2+C}$$
$$y = \pm e^C e^{t^2}.$$
As C runs over all real numbers, and as the $\pm$ sign varies, the coefficient $\pm e^C$ runs over all nonzero real numbers. Thus these solutions are $y = ce^{t^2}$ for all nonzero c.

Step 7. Because of Step 4, we need to check also the constant function $y = 0$; it turns out that it is a solution too. It can be considered as the function ce^{t^2} for $c = 0$.

Conclusion: The general solution to $\dot{y} - 2ty = 0$ is

$$y = ce^{t^2}, \quad \text{where } c \text{ is an } \textit{arbitrary} \text{ real number.} \quad \square$$

Step 8. Plugging in $y = ce^{t^2}$ to $\dot{y} - 2ty = 0$ gives $ce^{t^2}(2t) - 2tce^{t^2} = 0$, which is true, as it should be.

4. Linear ODEs vs. nonlinear ODEs

4.1. **Linear ODEs.**

4.1.1. *Building a homogeneous linear ODE.* One way to build a DE is as follows:

1. Start with a list like

   $$\ddot{y} \qquad \dot{y} \qquad y$$

   in which each term is one of y, $\dot{y}$, $\ddot{y}$, ... (it's OK to skip some).
2. Multiply each term by a function of t (possibly a *constant* function):

   $$e^t\,\ddot{y} \qquad 5\,\dot{y} \qquad t^9\,y.$$

3. Add them up and set the result equal to 0:

   $$e^t\,\ddot{y} + 5\,\dot{y} + t^9\,y = 0.$$

Any DE that arises in this way is called a homogeneous linear ODE.
("Homogeneous" has an e after the n, and the e is pronounced!)
The functions e^t, 5, and t^9 used in Step 2 are called the coefficients.

Most general n^{th} *order homogeneous linear ODE:*

$$p_n(t)\,y^{(n)} + \cdots + p_1(t)\,\dot{y} + p_0(t)\,y = 0$$

for some functions $p_n(t), \ldots, p_0(t)$.

4.1.2. *Building an inhomogeneous linear ODE.* If you start with a homogeneous linear ODE, and replace the 0 on the right by a function of t only, the result is called an inhomogeneous linear ODE. The function of t could be a constant function, but it is not allowed to involve y. For example,

$$e^t\,\ddot{y} + 5\,\dot{y} + t^9\,y = 7\sin t + 2$$

is an inhomogeneous linear ODE. So is

$$e^t\,\ddot{y} + 5\,\dot{y} + t^9\,y = 2.$$

Most general n^{th} *order inhomogeneous linear ODE:*

$$p_n(t)\, y^{(n)} + \cdots + p_1(t)\, \dot{y} + p_0(t)\, y = q(t)$$

for some functions $p_n(t)$, …, $p_0(t)$, $q(t)$.

Imagine feeding different "input signals" $q(t)$ into the right hand side of an inhomogeneous linear ODE to see what "output signals" $y(t)$ the system responds with.

4.1.3. *Both kinds together.* In testing whether an ODE is a homogeneous linear ODE or inhomogeneous linear ODE, you are allowed to rearrange the terms. A linear ODE is an ODE that can be rearranged into one of these two types.

Remark 4.1. If you already know that an ODE is linear, there is an easy test to decide if it is homogeneous or not: plug in the constant function $y = 0$.

- If $y = 0$ is a solution, the ODE is homogeneous.
- If $y = 0$ is not a solution, the ODE is inhomogeneous.

4.1.4. *Standard linear form.* Actually done on Feb. 9.

Dividing the DE

$$e^t\, \ddot{y} + 5\, \dot{y} + t^9\, y = 0.$$

by the leading coefficient e^t gives an equivalent DE

$$\ddot{y} + \frac{5}{e^t}\, \dot{y} + \frac{t^9}{e^t}\, y = 0.$$

The same can be done for any linear ODE, to put it in standard linear form

$$\boxed{y^{(n)} + p_{n-1}(t)\, y^{(n-1)} + \cdots + p_1(t)\, \dot{y} + p_0(t)\, y = q(t)}$$

for some functions $p_{n-1}(t)$, …, $p_0(t)$, $q(t)$ (not the same ones as before the division).

For now, we assume that we are looking for a solution $y(t)$ defined on an open interval I, and that the functions $p_{n-1}(t)$, …, $p_0(t)$, $q(t)$ are continuous (or at least piecewise continuous) on I. Open interval means a connected set of real numbers without endpoints, i.e., one of the following: (a, b), $(-\infty, b)$, (a, ∞), or $(-\infty, \infty) = \mathbb{R}$.

4.2. **Nonlinear ODEs.** For an ODE to be nonlinear, the functions y, $\dot{y}$, … must enter the equation in a more complicated way: raised to powers, multiplied by each other, or with nonlinear functions applied to them.

Flashcard question: Which of the following ODEs is linear?

$$\begin{aligned}
\ddot{y} - 7t\, y\dot{y} &= 0 \\
\ddot{y} &= e^t(y + t^2) \\
\dot{y} - y^2 &= 0 \\
\dot{y}^2 - ty &= \sin t \\
\dot{y} &= \cos(y + t).
\end{aligned}$$

Answer: The second one is linear since it can be rearranged into

$$\ddot{y} + (-e^t)y = t^2 e^t.$$

The others are nonlinear (the nonlinear portion is highlighted in red).

February 9

5. Solving a first-order linear ODE

Every *first-order* linear ODE in standard linear form is as follows:

$$\begin{aligned}
\text{Homogeneous:} &\quad \dot{y} + p(t)\, y = 0 \\
\text{Inhomogeneous:} &\quad \dot{y} + p(t)\, y = q(t).
\end{aligned}$$

5.1. **Homogeneous equations: separation of variables.** Homogeneous first-order linear ODEs can always be solved by separation of variables:

$$\begin{aligned}
\dot{y} + p(t)\, y &= 0 \\
\frac{dy}{dt} + p(t)\, y &= 0 \\
\frac{dy}{dt} &= -p(t)\, y \\
\frac{dy}{y} &= -p(t)\, dt \quad \text{(assume for now that } y \text{ is not 0).}
\end{aligned}$$

Choose *any* antiderivative $P(t)$ of $p(t)$. Integrating gives

$$\begin{aligned}
\ln|y| &= -P(t) + C \\
|y| &= e^{-P(t)+C} \\
y &= \pm e^C\, e^{-P(t)} \\
y &= ce^{-P(t)},
\end{aligned}$$

where c is any number (we brought back the solution $y = 0$ corresponding to $c = 0$).

If you choose a different antiderivative, it will have the form $P(t) + d$ for some constant d, and then the new $e^{-P(t)}$ is just a constant e^{-d} times the old one, so the set of *all* scalar multiples of the function $e^{-P(t)}$ is the same as before.

Conclusion:

Theorem 5.1 (General solution to first-order homogeneous linear ODE)**.** *Let $p(t)$ be a continuous function on an open interval I* (this ensures that $p(t)$ has an antiderivative)*. Let $P(t)$ be any antiderivative of $p(t)$. Then the general solution to $\dot{y} + p(t)\,y = 0$ is $\boxed{y = ce^{-P(t)}}$, where c is a parameter.*

5.2. **Inhomogeneous equations: variation of parameters.** Variation of parameters is a method for solving inhomogeneous linear ODEs. Given a first-order inhomogeneous linear ODE

$$\dot{y} + p(t)\,y = q(t), \tag{3}$$

follow these steps:

1. Find a nonzero solution, say y_h, of the associated homogeneous ODE

$$\dot{y} + p(t)\,y = 0.$$

   (You need just *one* nonzero solution. If instead you found the *general* solution to the homogeneous ODE, set the parameter c equal to 1, say, to get one solution.)
2. For an undetermined function $u(t)$, substitute

$$y = u(t)\,y_h(t) \tag{4}$$

   into the *inhomogeneous* equation (3) and solve for $u(t)$, to find all choices of $u(t)$ that make this y a solution to the inhomogeneous equation.
3. Now that the general $u(t)$ has been found, plug it back into $y = u(t)y_h(t)$ to get the general solution to the inhomogeneous equation. $\square$

The reason for considering uy_h in Step 2 is this: we know that if we try cy_h for a *constant* c in the inhomogeneous equation

$$\dot{y} + p(t)\,y = q(t),$$

it won't work since the left side will evaluate to 0 (the function cy_h is a solution to the *homogeneous* equation). Therefore instead we try uy_h for a *function* $u(t)$, and try to figure out *which* functions u will make the left side evaluate to $q(t)$. (That's why it's called variation of parameters: the parameter c has been replaced by something varying.)

Problem 5.2. Solve $t\dot{y} + 2y = t^5$ on the interval $(0, \infty)$.

Solution:

Step 1. The associated homogeneous equation is $t\dot{y} + 2y = 0$, or equivalently, $\dot{y} + \frac{2}{t}y = 0$. Solve by separation of variables:

$$\begin{aligned}
\frac{dy}{dt} &= -\frac{2}{t}y \\
\frac{dy}{y} &= -\frac{2}{t}\,dt \\
\ln|y| &= -2\ln t + C \qquad \text{(since } t > 0\text{)} \\
y &= ce^{-2\ln t} \\
y &= ct^{-2}.
\end{aligned}$$

Choose *one* nonzero solution, say $y_h = t^{-2}$.

Step 2. Substitute $y = ut^{-2}$ into the inhomogeneous equation: the left side is

$$t\dot{y} + 2y = t(\dot{u}t^{-2} + u\,(-2t^{-3})) + 2ut^{-2} = t^{-1}\dot{u},$$

so the inhomogeneous equation becomes

$$\begin{aligned}
t^{-1}\dot{u} &= t^5 \\
\dot{u} &= t^6 \\
u &= \frac{t^7}{7} + c.
\end{aligned}$$

Step 3. The general solution to the inhomogeneous equation is

$$y = ut^{-2} = \left(\frac{t^7}{7} + c\right)t^{-2} = \frac{t^5}{7} + ct^{-2}.$$

(If you want, check by direct substitution that this really is a solution.) □

5.3. **Inhomogeneous equations: integrating factor.** (Done in recitation.)

Another approach to solving

$$\dot{y} + p(t)\,y = q(t) \tag{5}$$

is to use an integrating factor:

1. Find an antiderivative $P(t)$ of $p(t)$.

2. Multiply both sides of the ODE by the integrating factor $e^{P(t)}$ in order to make the left side the derivative of something:

$$\begin{aligned} e^{P(t)}\dot{y} + e^{P(t)}p(t)\,y &= q(t)e^{P(t)} \\ \frac{d}{dt}\left(e^{P(t)}y\right) &= q(t)e^{P(t)} \\ e^{P(t)}y &= \int q(t)e^{P(t)}\,dt \\ y &= e^{-P(t)}\int q(t)e^{P(t)}\,dt. \end{aligned}$$

Here $\int q(t)e^{P(t)}\,dt$ represents all possible antiderivatives of $q(t)e^{P(t)}$, so there are infinitely many solutions.

If you fix one antiderivative, say $R(t)$, then the others are $R(t)+c$ for a constant c, so the general solution is

$$y = R(t)e^{-P(t)} + ce^{-P(t)}. \quad \square$$

5.4. **Linear combinations.** A linear combination of a list of functions is any function that can built from them by scalar multiplication and addition.

linear combinations of $f(t)$:	the functions $cf(t)$, where c is any number
linear combinations of $f_1(t)$ and $f_2(t)$:	the functions of the form $c_1f_1(t)+c_2f_2(t)$, where c_1 and c_2 are any numbers.

$$\vdots$$

Examples:

- $2\cos t + 3\sin t$ is a linear combination of the functions $\cos t$ and $\sin t$.
- $9t^5+3$ is a linear combination of the functions t^5 and 1.

Flashcard question: One of the functions below is not a linear combination of $\cos^2 t$ and 1. *Which one?*

1. $3\cos^2 t - 4$
2. $\sin^2 t$
3. $\sin(2t)$
4. $\cos(2t)$
5. 5
6. 0

Answer: 3.

All the others are linear combinations:

$$\begin{aligned} 3\cos^2 t - 4 &= 3\cos^2 t + (-4)\cdot 1 \\ \sin^2 t &= (-1)\cos^2 t + 1\cdot 1 \\ \sin(2t) &= ??? \\ \cos(2t) &= 2\cos^2 t + (-1)\cdot 1 \\ 5 &= 0\cos^2 t + 5\cdot 1 \\ 0 &= 0\cos^2 t + 0\cdot 1. \end{aligned}$$

Could there be some fancy identity that expresses $\sin(2t)$ as a linear combination of $\cos^2 t$ and 1? No; here's one way to see this: Every linear combination of $\cos^2 t$ and 1 has the form

$$c_1 \cos^2 t + c_2$$

for some numbers c_1 and c_2. All such functions are *even* functions, but $\sin(2t)$ is an *odd* function. (Warning: This trick might not work in other situations.)

5.5. **Superposition.**

Problem 5.3 (Multiplying an input signal by 9)**.** Fill in the blank:

$$\text{Given that } t^5/7 \text{ is one solution to } t\dot{y} + 2y = t^5,$$
$$\text{it follows that } \underline{\qquad} \text{ is one solution to } t\dot{y} + 2y = 9t^5.$$

You probably guessed that to get a right hand side that is 9 times as large, the solution needs to be 9 times as large: $9t^5/7$. This is a correct possible answer!

Why does this work? Imagine plugging in $y = 9t^5/7$ instead of $y = t^5/7$ into the left side of the DE; then each of $\dot{y}$ and y will be 9 times larger, so $t\dot{y} + 2y$ will be 9 times larger too; that is, it will equal $9t^5$ instead of t^5.

What special property of $t\dot{y} + 2y$ ensured that it would be 9 times larger? It is that each summand ($t\dot{y}$ and $2y$) is a function of t times one of y, $\dot{y}$, ..., so that when y is multiplied by 9, each summand gets multiplied by 9. In other words, this worked precisely because the DE was *linear*!

In the language of systems and signals, if the input signal (right hand side) is multiplied by 9, the output signal (the solution) is multiplied by 9.

Problem 5.4 (Adding two input signals)**.** Fill in the blank:

$$\text{Given that } t^5/7 \text{ is one solution to } t\dot{y} + 2y = t^5,$$
$$\text{and } 1/2 \text{ is one solution to } t\dot{y} + 2y = 1,$$
$$\text{it follows that } \underline{\qquad\quad} \text{ is one solution to } t\dot{y} + 2y = t^5 + 1.$$

As you probably guessed, $t^5/7 + 1/2$ is a possible answer. Again, this works because the DE is *linear*.

Adding input signals is also called superimposing them; this explains the name of the general principle:

Superposition principle (works for linear DEs only)**.**

1.

$$\begin{aligned} \text{Multiplying a solution to} \quad & p_n(t)\, y^{(n)} + \cdots + p_0(t)\, y = \ \ q(t) \quad \text{by a number } a \\ \text{gives a solution to} \quad & p_n(t)\, y^{(n)} + \cdots + p_0(t)\, y = aq(t). \end{aligned}$$

2.

$$\begin{aligned} \text{Adding a solution of} \quad & p_n(t)\, y^{(n)} + \cdots + p_0(t)\, y = q_1(t) \\ \text{to a solution of} \quad & p_n(t)\, y^{(n)} + \cdots + p_0(t)\, y = \qquad\qquad q_2(t) \\ \text{gives a solution of} \quad & p_n(t)\, y^{(n)} + \cdots + p_0(t)\, y = q_1(t) + q_2(t). \end{aligned}$$

Using both parts gives a principle for linear combinations, as in the following example.

Problem 5.5 (Linear combination of input signals)**.** Fill in the blank:

$$\begin{aligned} \text{Given that } t^5/7 &\text{ is one solution to } t\dot{y} + 2y = t^5, \\ \text{and } 1/2 &\text{ is one solution to } t\dot{y} + 2y = 1, \\ \text{it follows that } \underline{\qquad\quad} &\text{ is one solution to } t\dot{y} + 2y = 9t^5 + 3. \end{aligned}$$

One possible answer: $9t^5/7 + 3/2$.

5.6. **Consequence of superposition for an inhomogeneous linear DE.**

Problem 5.6. Fill in the blank:

$$\begin{aligned} \text{Given that } ct^{-2} &\text{ is the } \textit{general} \text{ solution to the homogeneous DE } t\dot{y} + 2y = 0, \\ \text{and } t^5/7 &\text{ is a particular solution to the inhomogeneous DE } t\dot{y} + 2y = t^5, \\ \text{it follows that } \underline{\qquad\qquad} &\text{ is the } \textit{general} \text{ solution to the inhomogeneous DE } t\dot{y} + 2y = t^5. \end{aligned}$$

Answer: $t^5/7 + ct^{-2}$. This is great news: if you already solved the homogeneous DE, you just have to find *one* solution to the inhomogeneous DE to build *all* solutions to the inhomogeneous DE!

Why does this work? For any number c, superposition says that adding ct^{-2} and $t^5/7$ will give *a* solution to the inhomogeneous DE with right side $0 + t^5 = t^5$, but why do *all* solutions of that inhomogeneous DE arise this way? It is because the process can be reversed: given any solution to the inhomogeneous DE $t\dot{y} + 2y = t^5$ we can *subtract* the particular solution $y_p = t^5/7$ to the same DE to get a solution to the homogeneous DE $t\dot{y} + 2y = 0$.

The strategy suggested by Problem 5.6 can help you find the general solution y_i to any

$$\text{inhomogeneous linear DE:} \quad p_n(t)\, y^{(n)} + \cdots + p_0(t)\, y = q(t) :$$

1. List all solutions to the associated

$$\text{homogeneous linear DE:} \quad p_n(t)\, y^{(n)} + \cdots + p_0(t)\, y = 0;$$

   i.e., write down *its* general solution y_h.
2. Find (in some way) any *one* particular solution y_p to the *inhomogeneous* DE.
3. Add y_p to all the solutions of the homogeneous DE to get all the solutions to the inhomogeneous DE.

Summary:

$$\underset{\textit{general}\text{ inhomogeneous solution}}{y_i} = \underset{\text{particular inhomogeneous solution}}{y_p} + \underset{\textit{general}\text{ homogeneous solution}}{y_h}.$$

5.7. **Newton's law of cooling.**

Problem 5.7. My minestrone soup is in an insulating thermos. Model its temperature as a function of time.

Simplifying assumptions:

- The insulating ability of the thermos does not change with time.
- The rate of cooling depends only on the *difference* between the soup temperature and the external temperature.

Variables and functions (with units): Define the following:

$$t: \ \text{time (minutes)}$$

$$x: \ \text{external temperature (°C)}$$

$$y: \ \text{soup temperature (°C)}$$

Here t is the independent variable, and x and y are functions of t.

Equation:

$$\dot{y} = f(y - x)$$

for some function f. Another simplifying assumption: $f(z) = -kz + \ell$ for some constants k and ℓ (any reasonable function be approximated on small inputs by its linearization at 0); this leads to

$$\dot{y} = -k(y - x) + \ell.$$

Common sense says

- If $y = x$, then $\dot{y} = 0$. Thus ℓ should be 0.

- If $y > x$, then y is decreasing. This is why we wrote $-k$ instead of just k.

So the equation becomes

$$\boxed{\dot{y} = -k(y - x)}$$

This is Newton's law of cooling: the rate of cooling of an object is proportional to the difference between its temperature and the external temperature. The (positive) constant k is called the coupling constant, in units of minutes^{-1}; smaller k means better insulation, and $k = 0$ is perfect insulation. This ODE can be rearranged into standard form:

$$\dot{y} + ky = kx.$$

It's a first-order inhomogeneous linear ODE! The input signal is x, the system is the thermos, and the output signal is y.

If the external temperature x is constant, then

- One particular solution to the inhomogeneous ODE: $y_p = x$
  (the solution in which the soup is already in equilibrium with the exterior)
- General solution to the *homogeneous* ODE: $y_h = ce^{-kt}$.
- General solution to the inhomogeneous ODE: $y = x + ce^{-kt}$.
  (As $t \to \infty$, the soup temperature approaches x; this makes sense.)

February 12

5.8. **Things to remember when solving first-order ODEs.**

1. Find the general solution first!
   - Try separation of variables, since this is often easier than the other methods, if it works.

   If the ODE is linear:
   - **Homogeneous**: Use separation of variables. The general solution will be $cf(t)$ for some function f.
   - **Inhomogeneous**: Choose one of the following methods.
     (i) Feeling lucky? Guess a particular solution y_p and write down

     $$y_p + \text{(general homogeneous solution)}.$$

     (ii) Variation of parameters: Find *one* homogeneous solution y_h and plug in $u\, y_h$.
     (iii) Integrating factor (to be discussed in recitation).
2. Use the initial conditions. This should always be the very last step.

6. Existence and uniqueness of solutions

Using separation of variables (in the homogeneous case) and variation of parameters (in the inhomogeneous case), we showed that every first-order linear ODE has a 1-parameter family of solutions. To nail down a specific solution in this family, we need one initial condition, such as $y(0)$.

It will turn out that every second-order linear ODE has a 2-parameter family of solutions. To nail down a specific solution, we need two initial conditions at the same starting time, $y(0)$ and $\dot{y}(0)$. The starting time could also be some number a other than 0.

Here is the general result:

Existence and uniqueness theorem for a linear ODE. *Let $p_{n-1}(t)$, …, $p_0(t)$, $q(t)$ be continuous functions on an open interval I. Let $a \in I$, and let $b_0, \ldots, b_{n-1}$ be given numbers. Then there exists a unique solution to the n^{th} order linear ODE*

$$y^{(n)} + p_{n-1}(t)\, y^{(n-1)} + \cdots + p_1(t)\, \dot{y} + p_0(t)\, y = q(t)$$

satisfying the n initial conditions

$$y(a) = b_0, \quad \dot{y}(a) = b_1, \quad \ldots, \quad y^{(n-1)}(a) = b_{n-1}.$$

Existence means that there is *at least* one solution.

Uniqueness means that there is *only* one solution.

Remark 6.1. For a linear ODE as above, the solution $y(t)$ is defined on the whole interval I where the functions $p_{n-1}(t)$, …, $p_0(t)$, $q(t)$ are continuous. In particular, if $p_{n-1}(t)$, …, $p_0(t)$, $q(t)$ are continuous on all of $\mathbb{R}$, then the solution $y(t)$ will be defined on all of $\mathbb{R}$.

7. Complex numbers

Complex numbers are expressions of the form $a + bi$, where a and b are real numbers, and i is a new symbol. Multiplication of complex numbers will eventually be defined so that $i^2 = -1$. (Electrical engineers sometimes write j instead of i, because they want to reserve i for current, but everybody else thinks that's weird.)

Just as the set of all real numbers is denoted $\mathbb{R}$, the set of all complex numbers is denoted $\mathbb{C}$. The notation "$\alpha \in \mathbb{C}$" means literally that α is an element of the set of complex numbers, so it is a short way of saying "α is a complex number".

Flashcard question: Is 9 a real number or a complex number?

Possible answers:

1. real number
2. complex number
3. both

4. neither

Answer: Both, because 9 can be identified with $9 + 0i$.

7.1. Operations on complex numbers.

real part	$\operatorname{Re}(a+bi) := a$	
imaginary part	$\operatorname{Im}(a+bi) := b$	(Note: It is b, not bi, so $\operatorname{Im}(a+bi)$ is real!)
complex conjugate	$\overline{a+bi} := a - bi$	(negate the imaginary component)

Flashcard question: *What is* $\operatorname{Im}(17 - 83i)$*?*

Possible answers:

1. 17
2. $17i$
3. 83
4. -83
5. $83i$
6. $-83i$

Answer: 4. The imaginary part is -83, without the i.

(In lecture there was a joke about the Greek letter Ξ; you had to be there.)

One can add, subtract, multiply, and divide complex numbers (except for division by 0). Addition, subtraction, and multiplication are defined as for polynomials, except that after multiplication one simplifies by using $i^2 = -1$; for example,

$$\begin{aligned}(2+3i)(1-5i) &= 2 - 7i - 15i^2 \\ &= 17 - 7i.\end{aligned}$$

To divide z by w, multiply z/w by $\overline{w}/\overline{w}$ so that the denominator becomes real; for example,

$$\frac{2+3i}{1-5i} = \frac{2+3i}{1-5i} \cdot \frac{1+5i}{1+5i} = \frac{2+13i+15i^2}{1-25i^2} = \frac{-13+13i}{26} = -\frac{1}{2} + \frac{1}{2}i.$$

The arithmetic operations on complex numbers satisfy the same properties as for real numbers ($zw = wz$ and so on). The mathematical jargon for this is that $\mathbb{C}$, like $\mathbb{R}$, is a field. In particular, for any complex number z and integer n, the n^{th} power z^n can be defined in the usual way (need $z \neq 0$ if $n < 0$); e.g., $z^3 := zzz$, $z^0 := 1$, $z^{-3} := 1/z^3$. (Warning: Although there is a way to define z^n also for a *complex* number n, when $z \neq 0$, it turns out that z^n has more than one possible value for non-integral n, so it is ambiguous notation. Anyway, the most important cases are e^z, and z^n for integers n; the other cases won't even come up in this class.)

If you change every i in the universe to $-i$ (that is, take the complex conjugate everywhere), then all true statements remain true. For example, $i^2 = -1$ becomes $(-i)^2 = -1$. Another example: If $z = v + w$, then $\overline{z} = \overline{v} + \overline{w}$; in other words,

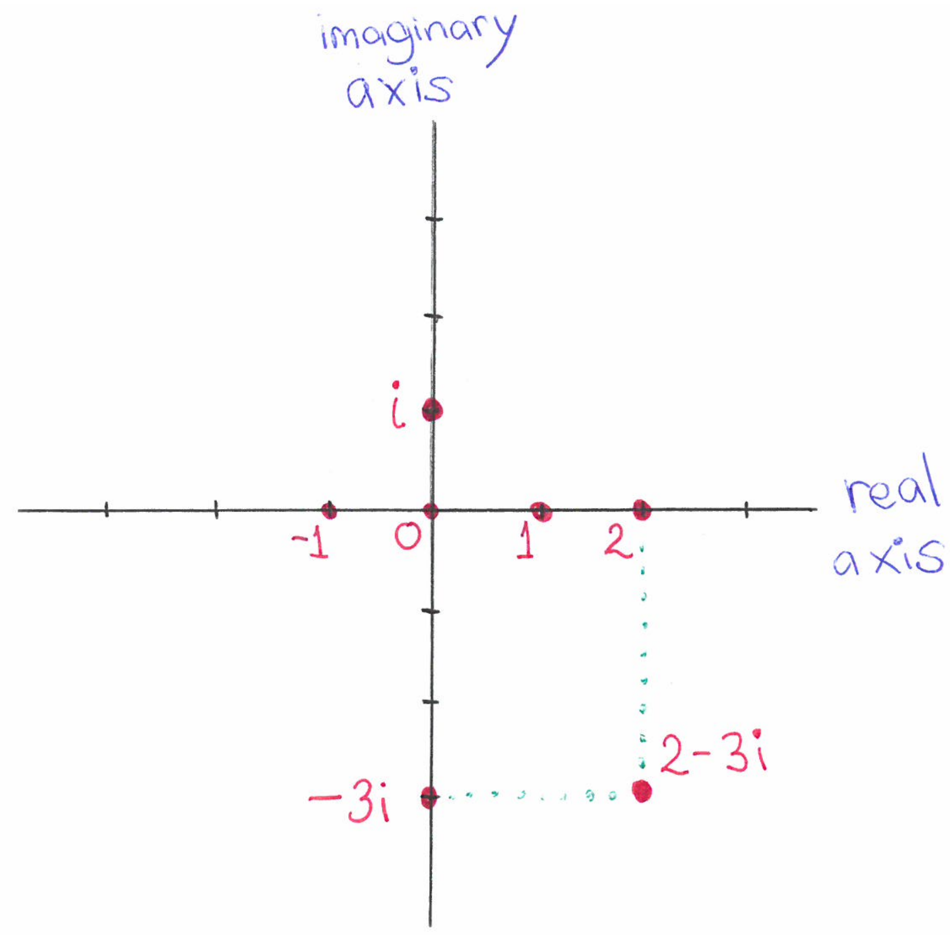
$$\overline{v+w} = \overline{v} + \overline{w}$$

for any complex numbers v and w. Similarly,

$$\overline{v\,w} = \overline{v}\,\overline{w}.$$

(These two identities say that complex conjugation respects addition and multiplication.)

7.2. **The complex plane.** Just as real numbers can be plotted on a line, complex numbers can be plotted on a plane: plot $a + bi$ at the point (a, b).

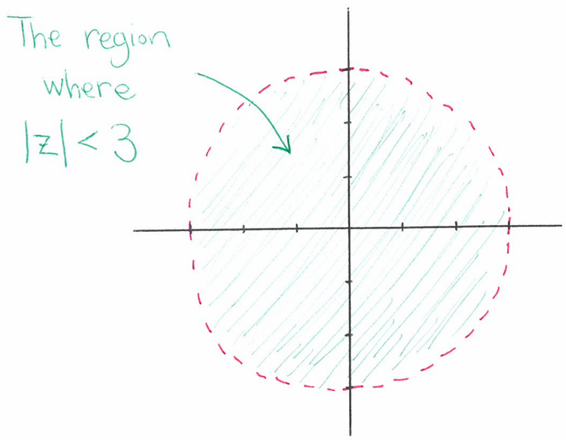


Addition and subtraction of complex numbers has the same geometric interpretation as for vectors. The same holds for scalar multiplication by a *real* number. (The geometric interpretation of multiplication by a *complex* number is different; we'll explain it soon.) Complex conjugation reflects a complex number in the real axis.

The absolute value (also called magnitude or modulus) $|z|$ of a complex number $z = a + bi$ is its distance to the origin:

$$|a + bi| := \sqrt{a^2 + b^2} \qquad \text{(this is a real number).}$$

For a complex number z, inequalities like $z < 3$ do not make sense, but inequalities like $|z| < 3$ do, because $|z|$ is a real number. The complex numbers satisfying $|z| < 3$ are those in the open disk of radius 3 centered at 0 in the complex plane. (Open disk means the disk without its boundary.)



7.3. **Some useful identities.** The following are true for all complex numbers z:

$$\operatorname{Re} z = \frac{z + \overline{z}}{2}, \qquad \operatorname{Im} z = \frac{z - \overline{z}}{2i}, \qquad \overline{\overline{z}} = z, \qquad z\overline{z} = |z|^2.$$

Also, for any *real* number c and complex number z,

$$\operatorname{Re}(cz) = c \operatorname{Re} z, \qquad \operatorname{Im}(cz) = c \operatorname{Im} z.$$

(These can fail if c is not real.)

Proof of the first identity: Write z as $a + bi$. Then

$$\operatorname{Re} z = a,$$
$$\frac{z + \overline{z}}{2} = \frac{(a + bi) + (a - bi)}{2} = a,$$

so $\operatorname{Re} z = \dfrac{z + \overline{z}}{2}$. $\square$

The proofs of the others are similar.

Many identities have a geometric interpretation too. For example, $\operatorname{Re} z = \dfrac{z + \overline{z}}{2}$ says that $\operatorname{Re} z$ is the midpoint between z and its reflection $\overline{z}$.

7.4. **Complex roots of polynomials.**

real polynomial: polynomial with real coefficients

complex polynomial: polynomial with complex coefficients

Example 7.1. How many roots does the polynomial $z^3 - 3z^2 + 4$ have? It factors as $(z-2)(z-2)(z+1)$, so it has only two distinct roots (2 and -1). But if we count 2 twice, then the number of roots *counted with multiplicity* is 3, equal to the degree of the polynomial.

Some real polynomials, like $z^2 + 9$, cannot be factored completely into degree 1 real polynomials, but do factor into degree 1 complex polynomials: $(z+3i)(z-3i)$. In fact, *every* complex polynomial factors completely into degree 1 complex polynomials — this is proved in advanced courses in complex analysis. This implies the following:

Fundamental theorem of algebra. *Every degree n complex polynomial $f(z)$ has exactly n complex roots, if counted with multiplicity.*

Since real polynomials are special cases of complex polynomials, the fundamental theorem of algebra applies to them too. For real polynomials, the non-real roots can be paired off with their complex conjugates.

Example 7.2. The degree 3 polynomial z^3+z^2-z+15 factors as $(z+3)(z-1-2i)(z-1+2i)$, so it has three distinct roots: -3, $1+2i$, and $1-2i$. Of these roots, -3 is real, and $1+2i$ and $1-2i$ form a complex conjugate pair.

Example 7.3. Want a fourth root of i? The fundamental theorem of algebra guarantees that $z^4 - i = 0$ has a complex solution (in fact, four of them). We'll soon learn how to find them.

The fundamental theorem of algebra will be useful for constructing solutions to higher order linear ODEs with constant coefficients, and for discussing eigenvalues.

7.5. **Real and imaginary parts of complex-valued functions.**

Suppose that $y(t)$ is a complex-valued function of a real variable t. Then

$$y(t) = f(t) + i\,g(t)$$

for some *real-valued* functions of t. Here $f(t) := \operatorname{Re} y(t)$ and $g(t) := \operatorname{Im} y(t)$. Differentiation and integration can be done component-wise:

$$\begin{aligned} y'(t) &= f'(t) + i\,g'(t) \\ \int y(t)\,dt &= \int f(t)\,dt + i \int g(t)\,dt. \end{aligned}$$

Example 7.4. Suppose that $y(t) = \dfrac{2+3i}{1+it}$. Then

$$y(t) = \frac{2+3i}{1+it} = \frac{2+3i}{1+it} \cdot \frac{1-it}{1-it} = \frac{(2+3t)+i(3-2t)}{1+t^2} = \underbrace{\left(\frac{2+3t}{1+t^2}\right)}_{f(t)} + i \underbrace{\left(\frac{3-2t}{1+t^2}\right)}_{g(t)}.$$

The functions in parentheses labelled $f(t)$ and $g(t)$ are real-valued, so these are the real and imaginary parts of the function $y(t)$. $\square$

7.6. **Defining the complex exponential function: introduction.** Raising e to a complex number has no *a priori* meaning; it needs to be defined. People long ago tried to define it so that the key properties of the function e^t for real numbers t would be true for complex numbers too. They succeeded, and we will too!

The most important property of e^t is that it satisfies

$$\dot{y} = y, \qquad y(0) = 1,$$

so we use this to guide the definition for complex numbers.

7.7. **The function e^{it}.**

Definition 7.5. The function e^{it} of a real number t is defined to be the solution to

$$\dot{y} = iy, \qquad y(0) = 1.$$

(The existence and uniqueness theorem for linear ODEs guarantees that there is exactly one solution.)

Remark 7.6. Strictly speaking, we need a *vector-valued* variant of the existence and uniqueness theorem, since a complex-valued function of t is equivalent to a pair of real-valued functions of t. Anyway, it is true.

Euler's formula. *We have* $\boxed{e^{it} = \cos t + i \sin t}$ *for every real number* t.

Proof. By definition, the function e^{it} is a solution to

$$\dot{y} = iy, \qquad y(0) = 1.$$

The calculation

$$\begin{aligned} \frac{d}{dt}(\cos t + i \sin t) &= -\sin t + i \cos t \\ &= i(\cos t + i \sin t) \end{aligned}$$

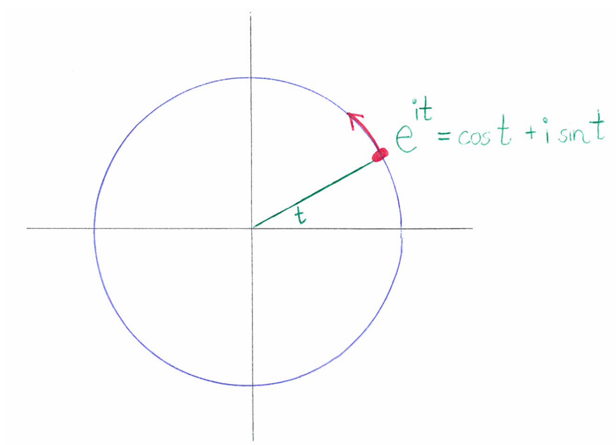
shows that the function $\cos t + i \sin t$ is *another* solution to

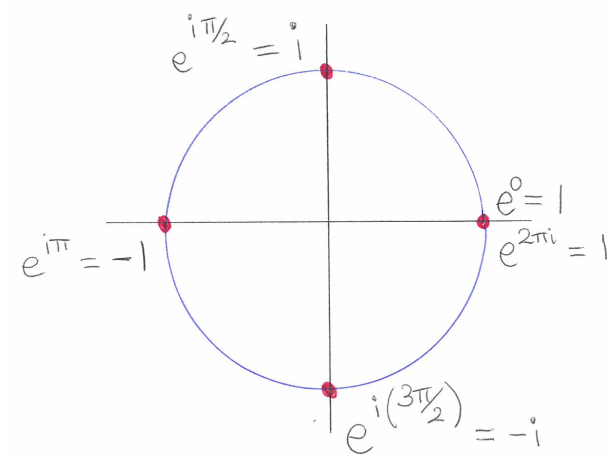
$$\dot{y} = iy, \qquad y(0) = 1.$$

But the uniqueness part of the existence and uniqueness theorem says that there is only one solution! Therefore e^{it} and $\cos t + i \sin t$ must be the same function. $\square$

Remark 7.7. Some older books use the awful abbreviation $\operatorname{cis} t := \cos t + i \sin t$, but this belongs in a cispool [*sic*], since e^{it} is a more useful expression for the same thing.

As t increases, the complex number $e^{it} = \cos t + i \sin t$ travels counterclockwise around the unit circle.





7.8. **The complex exponential function.**

Definition 7.8. For each complex constant α, the function $e^{\alpha t}$ is defined to be the unique solution to

$$\dot{y} = \alpha y, \qquad y(0) = 1.$$

For each nonzero α, the value of αt traces out the line through 0 and α as t ranges over real numbers, so the function $e^{\alpha t}$ specifies the value of e^z for every z on this line. These lines for varying α cover the whole complex plane, so defining $e^{\alpha t}$ for every α assigns a value to e^z for every complex number z.

Remark 7.9. Each value of e^z has been defined multiple times: for example, if $\beta = 2\alpha$, then e^{β} has been defined as both the value of $e^{\beta t}$ at $t = 1$ and the value of $e^{\alpha t}$ at $t = 2$. Fortunately these multiple definitions are consistent: in the example, $e^{\alpha(2t)}$ satisfies the same DE and initial condition that specified $e^{\beta t}$, so $e^{\alpha(2t)} = e^{\beta t}$; now set $t = 1$.

February 14

7.9. **Basic properties of the complex exponential function.** Let's show how the definition implies that the function e^z for complex z satisfies many of the same properties as e^t for real t.

Lemma 7.10. *For any complex numbers α and β, the functions $e^{\alpha t} e^{\beta t}$ and $e^{(\alpha+\beta)t}$ are equal.*

Proof. We will show that both $e^{\alpha t}e^{\beta t}$ and $e^{(\alpha+\beta)t}$ are solutions to

$$\dot{y} = (\alpha + \beta)y, \qquad y(0) = 1.$$

This implies that they are the same function, since the uniqueness part of the existence and uniqueness theorem says that there is only one solution!

The second function, $e^{(\alpha+\beta)t}$, satisfies that DE and initial condition by definition. The first function $y(t) := e^{\alpha t}e^{\beta t}$, also satisfies the DE and initial condition:

$$\begin{aligned}\dot{y} &= (\alpha e^{\alpha t})e^{\beta t} + e^{\alpha t}(\beta e^{\beta t}) \qquad \text{(product rule)}\\ &= (\alpha+\beta)e^{\alpha t}e^{\beta t}\\ &= (\alpha+\beta)y\end{aligned}$$

and $y(0) = e^{\alpha 0}e^{\beta 0} = 1 \cdot 1 = 1$. □

Theorem 7.11.

(a) $e^0 = 1$.
(b) $e^z e^w = e^{z+w}$ *for all complex numbers* z *and* w.
(c) $\dfrac{1}{e^z} = e^{-z}$ *for every complex number* z.
(d) $(e^z)^n = e^{nz}$ *for every complex number* z *and integer* n.

Proof.

(a) True by definition.
(b) The previous lemma said $e^{\alpha t}e^{\beta t} = e^{(\alpha+\beta)t}$. Evaluating at $t = 1$ gives $e^{\alpha}e^{\beta} = e^{\alpha+\beta}$. Renaming variables gives $e^z e^w = e^{z+w}$.
(c) (The proofs of (c) and (d) were skipped in lecture.) We have

$$e^z e^{-z} \overset{\text{(b)}}{=} e^{z+(-z)} = e^0 \overset{\text{(a)}}{=} 1,$$

so e^{-z} is the inverse of e^z.
(d) If $n = 0$, then this is $1 = 1$ by definition. If $n = 3$,

$$(e^z)^3 = e^z e^z e^z \overset{\text{(b) repeatedly}}{=} e^{z+z+z} = e^{3z};$$

the same argument works for any positive integer n. If $n = -3$, then

$$(e^z)^{-3} = \frac{1}{(e^z)^3} \overset{\text{(just shown)}}{=} \frac{1}{e^{3z}} \overset{\text{(c)}}{=} e^{-3z};$$

the same argument works for any negative integer n. □

7.10. **Further properties of the complex exponential function.**

Theorem 7.12.

(a) $e^{a+bi} = e^a(\cos b + i \sin b)$ *for all real numbers* a *and* b.
(b) $e^{-it} = \cos t - i \sin t = \overline{e^{it}}$ *for every real number* t.
(c) $|e^{it}| = 1$ *for every real number* t.

Proof.

(a) We have $e^{a+bi} = e^a e^{ib} = e^a(\cos b + i \sin b)$.
(b) Changing every i in the universe to $-i$ transforms $e^{it} = \cos t + i \sin t$ into $e^{-it} = \cos t - i \sin t$. (Substituting $-t$ for t would do it too.) On the other hand, applying complex conjugation to both sides of $e^{it} = \cos t + i \sin t$ gives $\overline{e^{it}} = \cos t - i \sin t$.
(c) By Euler's formula, $|e^{it}| = |\cos t + i \sin t| = \sqrt{\cos^2 t + \sin^2 t} = \sqrt{1} = 1$. □

Of lesser importance is the power series representation

$$e^z = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots . \tag{6}$$

This formula can be deduced by using Taylor's theorem with remainder, or by showing that the right hand side satisfies the DE and initial condition. Some books use $e^{a+bi} = e^a(\cos b + i \sin b)$ or the power series $e^z = 1 + z + \frac{z^2}{2!} + \cdots$ as the *definition* of the complex exponential function, but the DE definition we gave is less contrived and focuses on what makes the function useful.

7.11. **Polar forms of a complex number.** Given a nonzero complex number $z = x + yi$, we can express the point (x, y) in polar coordinates r and θ:

$$x = r \cos \theta, \qquad y = r \sin \theta.$$

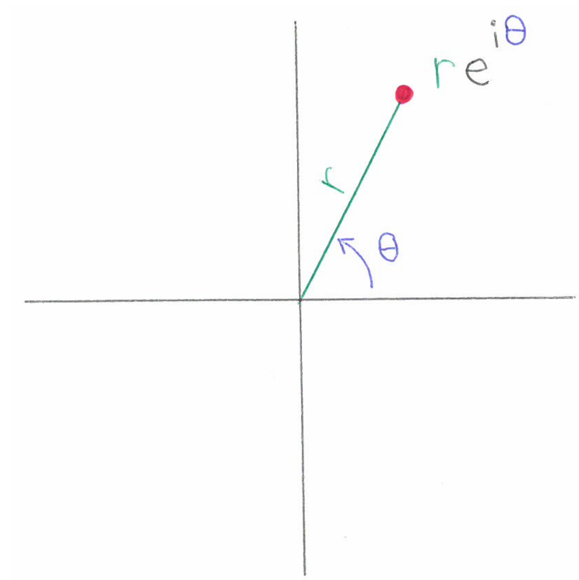
Then

$$x + yi = (r \cos \theta) + (r \sin \theta)i = r(\cos \theta + i \sin \theta).$$

In other words,

$$\boxed{z = re^{i\theta}}.$$

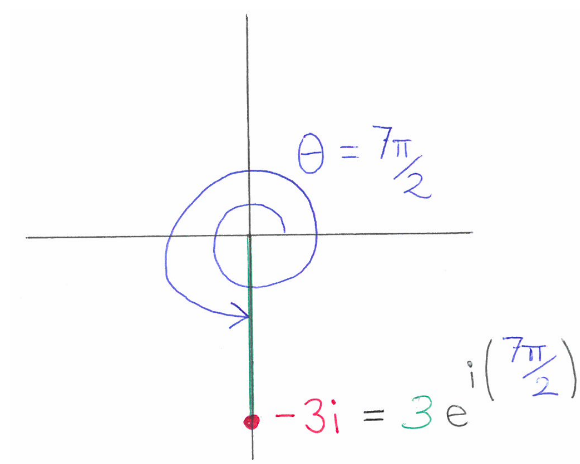
The expression $re^{i\theta}$ is called a polar form of the complex number z. Here r is required to be a *positive* real number (assuming $z \neq 0$), so $r = |z|$.



Any possible θ for z (a possible value for the angle or argument of z) may be called $\arg z$, but this is dangerously ambiguous notation since there are many values of θ for the same z: this means that $\arg z$ is not a function.

Example 7.13. Suppose that $z = -3i$. So z corresponds to the point $(0, -3)$. Then $r = |z| = 3$, but there are infinitely many possibilities for the angle θ. One possibility is $-\pi/2$; all the others are obtained by adding integer multiples of 2π:

$$\arg z \quad = \quad \ldots,\ -5\pi/2,\ -\pi/2,\ 3\pi/2,\ 7\pi/2,\ \ldots.$$



So z has many polar forms:

$$\cdots = 3e^{i(-5\pi/2)} = 3e^{-i\pi/2} = 3e^{i(3\pi/2)} = 3e^{i(7\pi/2)} = \cdots. \quad \square$$

To specify a *unique* polar form, we would have to restrict the range for θ to some interval of width 2π. The most common choice is to require $-\pi < \theta \leq \pi$. This special θ is called the principal value of the argument, and is denoted in various ways:

$$\theta = \operatorname{Arg} z = \underbrace{\texttt{Arg[z]}}_{\text{Mathematica}} = \underbrace{\texttt{ArcTan[x,y]}}_{\text{Mathematica}} = \underbrace{\texttt{atan2(y,x)}}_{\text{MATLAB}}.$$

Warning: Some people require $0 \le \theta < 2\pi$ instead.

Warning: In MATLAB, be careful to use (y, x) and not (x, y).

Test for equality of two nonzero complex numbers in polar form:

$$r_1 e^{i\theta_1} = r_2 e^{i\theta_2} \qquad \Longleftrightarrow \qquad r_1 = r_2 \text{ and } \theta_1 = \theta_2 + 2\pi k \text{ for some integer } k.$$

This assumes that r_1 and r_2 are positive real numbers, and that θ_1 and θ_2 are real numbers, as you would expect for polar coordinates.

7.12. **Converting from $x + yi$ form to a polar form.** This is the same as converting from rectangular coordinates to polar coordinates, so you are supposed to know this already. This section was not covered in lecture.

Problem 7.14. Convert a nonzero complex number $z = x + yi$ to polar form. In other words, given real numbers x and y, find r and a possible θ.

Finding r is easy: $r = |z| = \sqrt{x^2 + y^2}$.

Finding θ is trickier. If $x = 0$ or $y = 0$, then $x + yi$ is on one of the axes and θ will be an appropriate integer multiple of $\pi/2$. So assume that x and y are nonzero. The correct θ satisfies $\tan\theta = y/x$, but there are also other angles that satisfy this equation, namely $\theta + k\pi$ for any integer k. Some of these other angles point in the opposite direction. In particular, $\tan^{-1}(y/x)$ might be in the opposite direction. By definition, the angle $\tan^{-1}(y/x)$ always lies in $(-\pi/2, \pi/2)$, pointing into the right half plane, so it will be wrong when $x + yi$ lies in the left half plane; in that case, adjust $\tan^{-1}(y/x)$ by adding or subtracting π to get a possible θ. Finally, if desired, add an integer multiple of 2π to get the *principal* value of the argument, which is the θ satisfying $-\pi < \theta \le \pi$.

The "2-variable arctangent function" in Mathematica and MATLAB mentioned above looks not only at y/x, but also at the point (x, y), to calculate a correct θ.

Example 7.15. Suppose that $z = -1 - i$. Evaulating $\tan^{-1}(y/x)$ at $(-1, -1)$ gives $\tan^{-1}(1) = \pi/4$, pointing in the direction *opposite* to $(-1, -1)$. Subtracting π gives $-3\pi/4$ as a possible θ. (The other possible angles θ are the numbers $-3\pi/4 + 2\pi k$, where k can be any integer.)

7.13. **Operations in polar form.** Some arithmetic operations on complex numbers are easy in polar form:

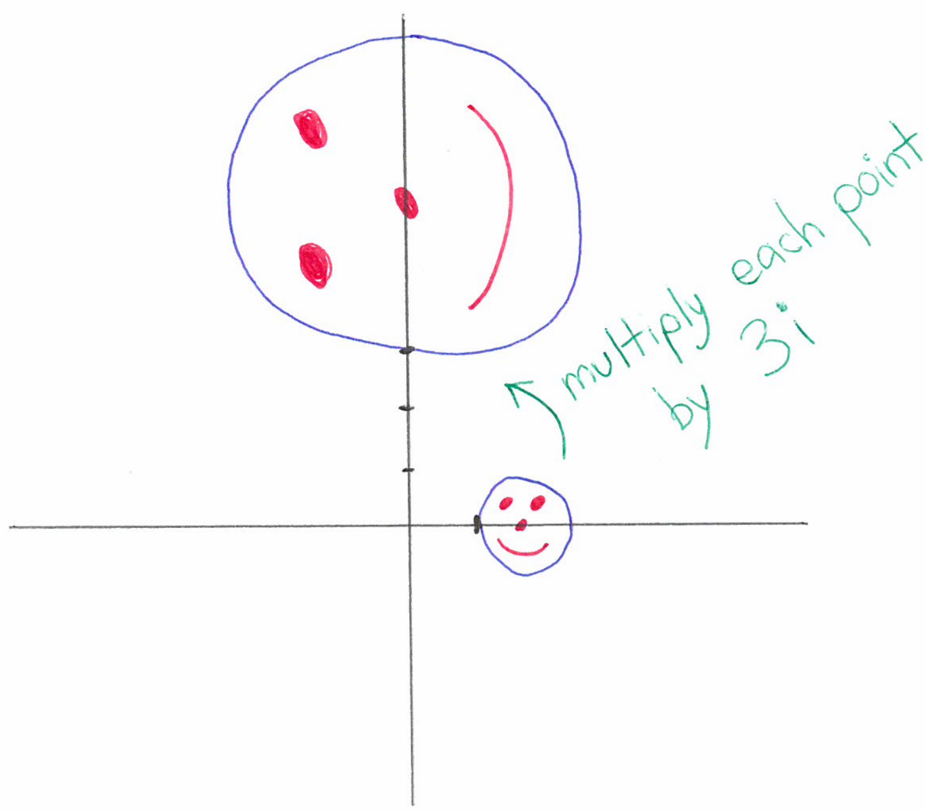
$$\begin{aligned}
&\text{multiplication: } (r_1 e^{i\theta_1})(r_2 e^{i\theta_2}) = r_1 r_2 e^{i(\theta_1+\theta_2)} \quad \text{(multiply absolute values, add angles)}\\
&\text{reciprocal: } \quad \frac{1}{re^{i\theta}} = \frac{1}{r} e^{-i\theta}\\
&\text{division: } \quad \frac{r_1 e^{i\theta_1}}{r_2 e^{i\theta_2}} = \frac{r_1}{r_2} e^{i(\theta_1-\theta_2)} \quad \text{(divide absolute values, subtract angles)}\\
&n^{\text{th}} \text{ power: } \quad (re^{i\theta})^n = r^n e^{in\theta} \qquad \text{for any integer } n\\
&\text{complex conjugation: } \quad \overline{re^{i\theta}} = re^{-i\theta}.
\end{aligned}$$

Taking absolute values gives identities:

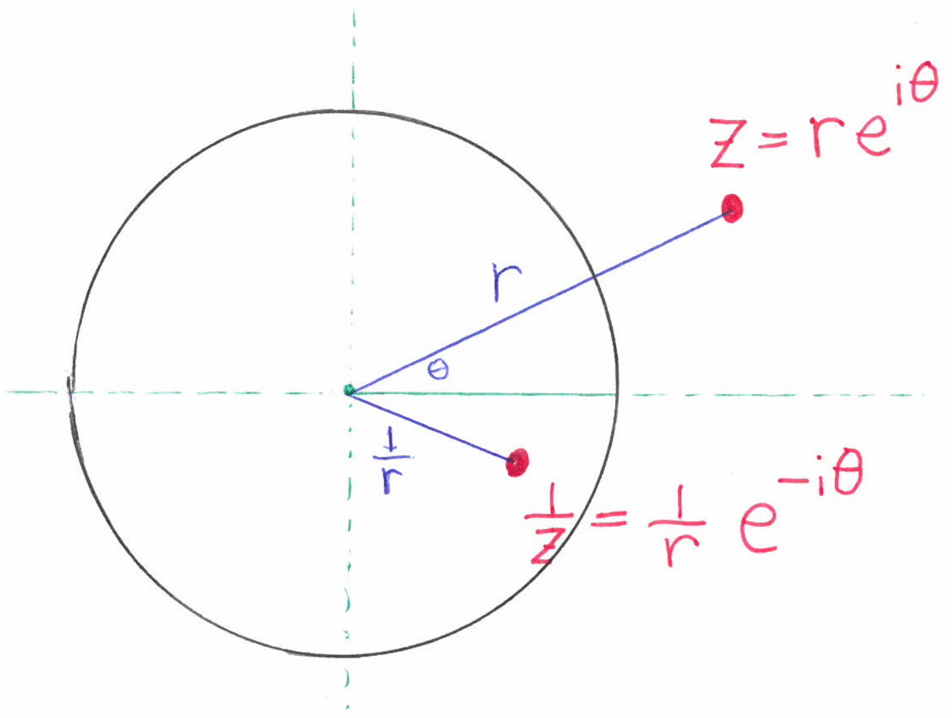
$$|z_1 z_2| = |z_1|\,|z_2|, \qquad \left|\frac{1}{z}\right| = \frac{1}{|z|}, \qquad \left|\frac{z_1}{z_2}\right| = \frac{|z_1|}{|z_2|}, \qquad |z^n| = |z|^n, \qquad |\overline{z}| = |z|.$$

Question 7.16. What happens if you take a smiley in the complex plane and multiply each of its points by $3i$?

Solution: Since $i = e^{i\pi/2}$, multiplying by i adds $\pi/2$ to the angle of each point; that is, it rotates counterclockwise by 90° (around the origin). Next, multiplying by 3 does what you would expect: dilate by a factor of 3. Doing both leads to...



For example, the nose was originally on the real line, a little less than 2, so multiplying it by $3i$ produces a big nose close to $(3i)2 = 6i$. $\square$



Question 7.17. How do you trap a lion?

Answer: Build a cage in the shape of the unit circle $|z| = 1$. Get inside the cage. Make sure that the lion is outside the cage. Apply the function $1/z$ to the whole plane. Voilà! The lion is now inside the cage, and you are outside it. (Only problem: There's a lot of other stuff inside the cage too. Also, don't stand too close to $z = 0$ when you apply $1/z$.)

Question 7.18. Why not always write complex numbers in polar form?

Answer: Because addition and subtraction are difficult in polar form!

7.14. **The function $e^{(a+bi)t}$.** Fix a nonzero complex number $a + bi$. As the real number t increases, the complex number $(a + bi)t$ moves along a line through 0, and $e^{(a+bi)t}$ moves along part of a line, a circle, or a spiral, depending on the value of $a + bi$. Try the "Complex Exponential" mathlet

`http://mathlets.org/mathlets/complex-exponential/`

to see examples of this.

Example 7.19. Consider $e^{(-5-2i)t} = e^{-5t}e^{i(-2t)}$ as $t \to \infty$. Its absolute value is e^{-5t}, which tends to 0, so the point is moving inward. Its angle is $-2t$, which is decreasing, so the point is moving clockwise. It's spiraling inwards clockwise.

7.15. **Finding n^{th} roots.**

7.15.1. *An example.*

Problem 7.20. What are the complex solutions to $z^5 = -32$?

Solution: Rewrite the equation in polar form, using $z = re^{i\theta}$:

$$(re^{i\theta})^5 = 32e^{i\pi}$$

$$r^5 e^{i(5\theta)} = 32e^{i\pi}$$

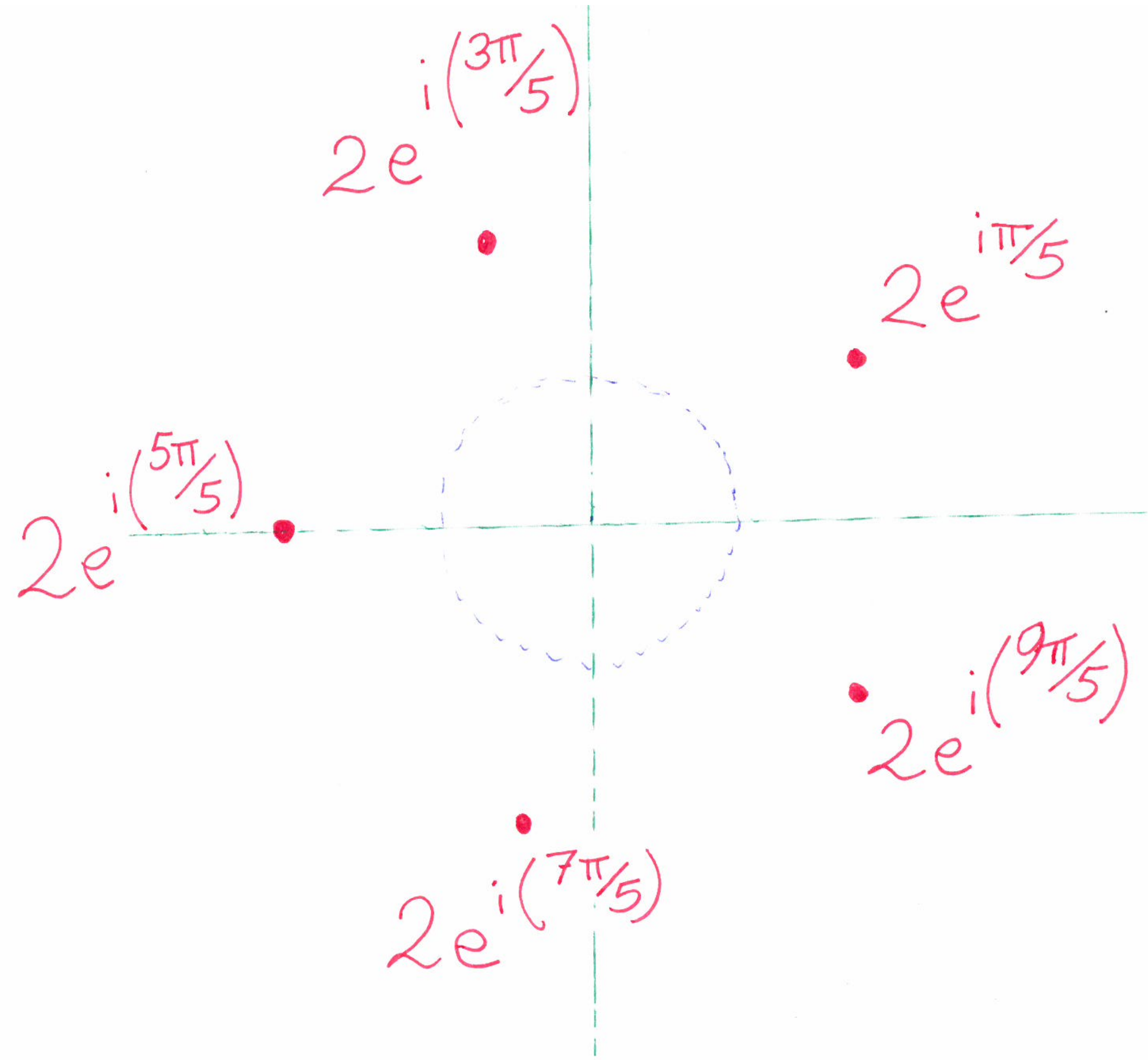
$$\underbrace{r^5 = 32}_{\text{absolute values}} \qquad \underbrace{5\theta = \pi + 2\pi k}_{\text{angles}} \text{ for some integer } k$$

$$r = 2 \qquad \text{and} \qquad \theta = \frac{\pi}{5} + \frac{2\pi k}{5} \text{ for some integer } k$$

$$z = 2e^{i\left(\frac{\pi}{5}+\frac{2\pi k}{5}\right)} \text{ for some integer } k.$$

These are numbers on a circle of radius 2; to get from one to the next (increasing k by 1), rotate by $2\pi/5$. Increasing k five times brings the number back to its original position. So it's enough to take $k = 0, 1, 2, 3, 4$. Answer:

$$2e^{i(\pi/5)},\ 2e^{i(3\pi/5)},\ 2e^{i(5\pi/5)},\ 2e^{i(7\pi/5)},\ 2e^{i(9\pi/5)}. \quad \square$$



Remark 7.21. The fundamental theorem of algebra predicts that the polynomial $z^5 + 32$ has 5 roots when counted with multiplicity. We found 5 roots, so each must have multiplicity 1.

Remark 7.22. The same approach (write in polar form, solve for absolute value and angle) finds the solutions to $z^n = \alpha$ for any positive integer n and nonzero complex number α.

7.15.2. *Roots of unity.*

Problem 7.23. Let n be a positive integer. The n^{th} roots of unity are the complex solutions to $z^n = 1$. Find them all.

Solution: Rewrite the equation in polar form, using $z = re^{i\theta}$:

$$(re^{i\theta})^n = 1$$

$$r^n e^{i(n\theta)} = 1$$

$$\underset{\text{absolute values}}{r^n = 1} \qquad \underset{\text{angles}}{n\theta = 2\pi k} \text{ for some integer } k$$

$$r = 1 \qquad \text{and} \qquad \theta = \frac{2\pi k}{n} \text{ for some integer } k$$

$$z = e^{i\left(\frac{2\pi k}{n}\right)} \text{ for some integer } k.$$

As before, it's enough to take $k = 0, 1, 2, \ldots, n-1$. One of the solutions, obtained by taking $k = 1$, is the number $e^{2\pi i/n}$; call this ζ. In terms of ζ, the complete list of n^{th} roots of unity is

$$\boxed{1,\ \zeta,\ \zeta^2,\ \ldots,\ \zeta^{n-1}}.$$

February 16

7.15.3. *Another approach to describing all n^{th} roots of a complex number.* Here is another approach to describing all complex solutions to $z^n = \alpha$, analogous to the $y_i = y_p + y_h$ approach to inhomogeneous linear DEs:

Problem 7.24. Fill in the blank:

If z_0 is one solution to $z^n = \alpha$,

and $1, \zeta, \zeta^2, \ldots, \zeta^{n-1}$ are all the solutions to $z^n = 1$,

then ________________ are all the solutions to $z^n = \alpha$.

Answer: $\boxed{z_0,\ \zeta z_0,\ \zeta^2 z_0,\ \ldots,\ \zeta^{n-1} z_0}$. Why? These are solutions, since for any integer k,

$$(\zeta^k z_0)^n = (\zeta^k)^n z_0^n = 1 \cdot \alpha = \alpha;$$

there can't be any more, since a degree n polynomial equation has at most n solutions. □

Remark 7.25. To use this approach to find all solutions to $z^n = \alpha$, you need to know one solution z_0 in advance. If there is no obvious z_0, you'll still need polar form to solve $z^n = \alpha$.

The answer above says that once you know one solution z_0, you get the others by repeatedly multiplying by $\zeta = e^{2\pi i/n}$, which rotates by $2\pi/n$; after n rotations, you return to z_0. Thus the n solutions to $z^n = \alpha$ form the vertices of a regular n-gon (at least if $n \geq 3$).

Try the "Complex Roots" mathlet

`http://mathlets.org/mathlets/complex-roots/`

7.16. e^{it} and e^{-it} as linear combinations of $\cos t$ and $\sin t$, and vice versa.

Example 7.26. The functions e^{it} and e^{-it} are linear combinations of the functions $\cos t$ and $\sin t$:

$$\begin{aligned} e^{it} &= \cos t + i \sin t \\ e^{-it} &= \cos t - i \sin t. \end{aligned}$$

If we view e^{it} and e^{-it} as known, and $\cos t$ and $\sin t$ as unknown, then this is a system of two linear equations in two unknowns, and can be solved for $\cos t$ and $\sin t$. This gives

$$\boxed{\cos t = \frac{e^{it} + e^{-it}}{2}}, \qquad \boxed{\sin t = \frac{e^{it} - e^{-it}}{2i}}.$$

Thus $\cos t$ and $\sin t$ are linear combinations of e^{it} and e^{-it}. (Explicitly, $\sin t = \frac{1}{2i} e^{it} + \frac{-1}{2i} e^{-it}$.)

Important: The function e^z has nicer properties than $\cos t$ and $\sin t$, so it is often a good idea to use these formulas to replace $\cos t$ and $\sin t$ by these combinations of e^{it} and e^{-it}, or to view $\cos t$ and $\sin t$ as the real and imaginary parts of e^{it}.

Replacing t by ωt in the identities above leads to

$$\begin{aligned} e^{i\omega t} &= \cos \omega t + i \sin \omega t \\ e^{-i\omega t} &= \cos \omega t - i \sin \omega t. \end{aligned}$$

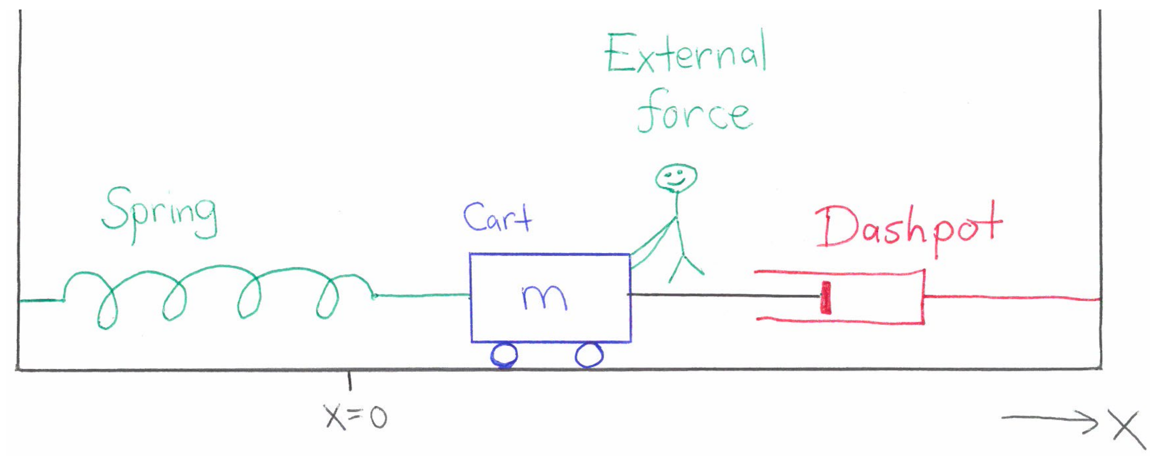
and

$$\boxed{\cos \omega t = \frac{e^{i\omega t} + e^{-i\omega t}}{2}}, \qquad \boxed{\sin \omega t = \frac{e^{i\omega t} - e^{-i\omega t}}{2i}}.$$

8. Introduction to second-order linear ODEs with constant coefficients

8.1. Modeling: a spring-mass-dashpot system.

Problem 8.1. A cart is attached to a spring attached to a wall. The cart is attached also to a dashpot, a damping device. (A dashpot could be a cylinder filled with oil that a piston moves through. Door dampers and car shock absorbers often actually work this way.) Also, there is an external force acting on the cart. Model the motion of the cart.



Solution: Define variables

$$
\begin{aligned}
t &: \text{ time (s)} \\
x &: \text{ position of the cart (m), with } x = 0 \text{ being where the spring exerts no force} \\
m &: \text{ mass of the cart (kg)} \\
F_{\text{spring}} &: \text{ force exerted by the spring on the cart (N)} \\
F_{\text{dashpot}} &: \text{ force exerted by the dashpot on the cart (N)} \\
F_{\text{external}} &: \text{ external force on the cart (N)} \\
F &: \text{ total force on the cart (N).}
\end{aligned}
$$

The independent variable is t; everything else is a function of t (well, maybe m is constant).

Physics tells us that

- F_{spring} is a function of x (of opposite sign), and
- F_{dashpot} is a function of $\dot{x}$ (again of opposite sign).

To simplify, approximate these by *linear* functions (probably OK if x and $\dot{x}$ are small):

$$\underbrace{F_{\text{spring}} = -kx}_{\text{Hooke's law}}, \qquad F_{\text{dashpot}} = -b\dot{x},$$

where k is the spring constant (in units N/m) and b is the damping constant (in units N s/m); here $k, b > 0$. Substituting these and Newton's second law $F = m\ddot{x}$ into

$$F = F_{\text{spring}} + F_{\text{dashpot}} + F_{\text{external}}$$

gives

$$m\ddot{x} = -kx - b\dot{x} + F_{\text{external}},$$

a second order linear ODE, which we would usually write as

$$\boxed{m\ddot{x} + b\dot{x} + kx = F_{\text{external}}(t)}. \quad \square$$

All this works even if m, b, k are functions of time, but we'll assume from now on that they are constants.

$$\begin{aligned} \text{input signal:} &\quad F_{\text{external}}(t) \\ \text{system:} &\quad \text{spring, mass, and dashpot} \\ \text{output signal:} &\quad x(t). \end{aligned}$$

Carts attached to springs are not necessarily what interest us. But oscillatory systems arising in all the sciences are governed by the same math, and this physical system lets us visualize their behavior.

8.2. **The differential equation $\ddot{x}+x=0$.** Suppose that $m=k=1$ and there is no dashpot and no external force, only a mass and spring. Then the DE is simply

$$\boxed{\ddot{x}+x=0}.$$

Each solution $x(t)$ gives rise to a pair of numbers $(x(0), \dot{x}(0))$. Conversely, the existence and uniqueness theorem says that for each pair of numbers (a, b), there is exactly one solution to $\ddot{x}+x=0$ satisfying $(x(0), \dot{x}(0)) = (a, b)$. What are these solutions?

$$\begin{aligned} \text{Well, } \cos t &\text{ is the solution to } \ddot{x}+x=0 \text{ such that } x(0)=1 \text{ and } \dot{x}(0)=0, \\ \text{and } \sin t &\text{ is the solution to } \ddot{x}+x=0 \text{ such that } x(0)=0 \text{ and } \dot{x}(0)=1, \\ \text{so } \underline{\qquad\qquad} &\text{ is the solution to } \ddot{x}+x=0 \text{ such that } x(0)=a \text{ and } \dot{x}(0)=b. \end{aligned}$$

The answer is the function $a\cos t + b\sin t$. In other words, the 2-parameter family of solutions

$$\boxed{a\cos t + b\sin t}$$

is the *general* solution to $\ddot{x}+x=0$.

There are other ways to construct solutions. For example, for any constant ϕ, the time-shifted function $\cos(t-\phi)$ is a solution, and if A is another constant, then

$$A\cos(t-\phi)$$

is a solution too. It turns out that these functions are the same as the functions $a\cos t + b\sin t$, just written in a different form, so the family of such functions $A\cos(t-\phi)$ is the general solution again!

To explain this and to solve other DEs, we'll need to understand functions like these and learn how to convert between the different forms.

9. Sinusoidal functions

9.1. **Construction.** Start with the curve $y = \cos x$. Then

1. Shift the graph ϕ units to the right (ϕ is phase lag, measured in radians). (For example, shifting by $\phi = \pi/2$ gives the graph of $\sin x$, which reaches its maximum $\pi/2$ radians after $\cos x$ does.)
2. Compress the result horizontally by *dividing* by a scale factor ω (angular frequency, measured in radians/s).
3. Amplify (stretch vertically) by a factor of A (amplitude).

(Here $A, \omega > 0$, but ϕ can be any real number.)

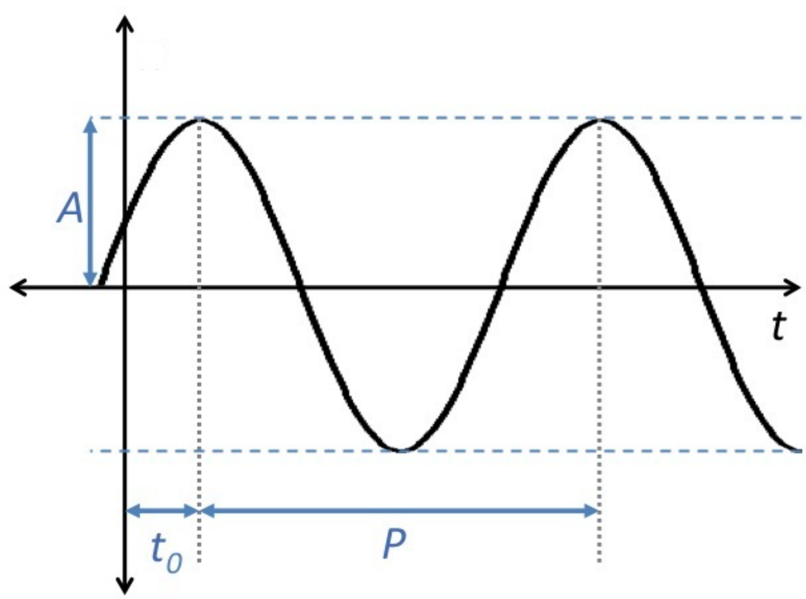
Result? The graph of a new function $f(t)$, called a sinusoidal function (or just sinusoid).

9.2. **Formula.** What is the formula for $f(t)$? According to the instructions, each point (x, y) on $y = \cos x$ is related to a point $(t, f(t))$ on the graph of f by

$$t = \frac{x + \phi}{\omega}, \quad f = Ay.$$

Solving for x gives $x = \omega t - \phi$; substituting into $f = Ay = A \cos x$ gives

$$\boxed{f(t) = A\cos(\omega t - \phi)}.$$



9.3. **Alternative geometric description.** Alternatively, the graph of $f(t)$ can be described geometrically in terms of

A : its amplitude, as above, how high the graph rises above the t-axis at its maximum

t_0 : its time lag, also sometimes called τ, a t-value at which a maximum is attained (s)

P : its period, the time for one complete oscillation (= width between successive maxima) (s or s/cycle)

How do t_0 and P relate to ω and ϕ?

- $t_0 = \phi/\omega$, since this is the t-value for which the angle $\omega t - \phi$ becomes 0.
- $P = 2\pi/\omega$, since adding $2\pi/\omega$ to t increases the angle $\omega t - \phi$ by 2π.

There is also frequency $\nu := 1/P$, measured in Hz = cycles/s. It is the number of complete oscillations per second. To convert from frequency ν to angular frequency ω, multiply by $\dfrac{2\pi \text{ radians}}{1 \text{ cycle}}$; thus $\omega = 2\pi\nu = 2\pi/P$, which is consistent with the formula $P = 2\pi/\omega$ above.

Question 9.1. What is the difference between phase lag and time lag?

Answer: Phase lag ϕ and time lag t_0 both measure how much a sinusoid $A\cos(\omega t - \phi)$ is shifted relative to the standard sinusoid $\cos(\omega t)$ of the same frequency, but ϕ is measured as a fraction of a cycle (expressed in radians), and t_0 is expressed in time units.

For example, if ϕ is π radians, that is half a cycle, so it means that the sinusoid is completely out of phase, attaining a maximum where $\cos(\omega t)$ has a minimum, and vice versa. The time lag of this same sinusoid, however, will depend on the duration of a cycle: if the angular frequency ω is very high, then there will be many cycles per second, so each cycle represents a very short time period, so the time for half a cycle will also be very short.

To convert between phase lag and time lag, multiply or divide by the angular frequency ω. To remember whether to multiply or divide, compare units. Since ω is measured in radians/s, the conversion is $t_0 = \phi/\omega$.

Another way to think about this: in terms of the construction of a sinusoid, ϕ represents the shift in angle, and then compressing horizontally by dividing by ω gives the shift in time. $\square$

Try the "Sinusoids" mathlet

`http://web.mit.edu/jorloff/www/jmoapplets/svg-master/sinusoids.html`.

9.4. **Three forms.** There are three ways to write a sinusoidal function of angular frequency ω:

- amplitude-phase form: $A\cos(\omega t - \phi)$, where A and ϕ are real numbers with $A \geq 0$;
- complex form: $\operatorname{Re}\left(ce^{i\omega t}\right)$, where c is a complex number;
- linear combination: $a\cos\omega t + b\sin\omega t$, where a and b are real numbers.

Different forms are useful in different contexts, so we'll need to know how to convert between them. The following proposition explains how.

Proposition 9.2. *If constants A, ω, ϕ, a, b, c are set so that the key equations*

$$\begin{tikzcd} & Ae^{i\phi} \arrow[dl, equal] \arrow[dr, equal] & \\ \overline{c} \arrow[rr, equal] & & a+bi \end{tikzcd}$$

hold, then

$$\begin{tikzcd} & A\cos(\omega t - \phi) \arrow[dl, equal, "1"'] \arrow[dr, equal, "3{,}4"] & \\ \operatorname{Re}\left(ce^{i\omega t}\right) \arrow[rr, equal, "2"'] & & a\cos\omega t + b\sin\omega t. \end{tikzcd}$$

(The little numbers refer to the proofs below.)

February 20

Warning: Don't forget that it is $\overline{c}$ and not c itself that appears in the key equations. An equivalent form of the key equations (obtained by taking complex conjugates) is

$$\begin{tikzcd} & Ae^{-i\phi} \arrow[dl, equal] \arrow[dr, equal] & \\ c \arrow[rr, equal] & & a-bi \end{tikzcd}$$

If you ever forget the key equations above, you can do the conversion manually by going through the steps in the proof below.

Proof of Proposition 9.2.

1.

$$\begin{aligned} \operatorname{Re}\left(ce^{i\omega t}\right) &= \operatorname{Re}\left(Ae^{-i\phi}\, e^{i\omega t}\right) \\ &= \operatorname{Re}\left(Ae^{i(\omega t-\phi)}\right) \\ &= A\operatorname{Re}\left(e^{i(\omega t-\phi)}\right) \\ &= A\cos(\omega t-\phi). \end{aligned}$$

2.

$$\begin{aligned} \operatorname{Re}\left(ce^{i\omega t}\right) &= \operatorname{Re}\left((a-bi)(\cos\omega t + i\sin\omega t)\right) \\ &= \operatorname{Re}\left((a\cos\omega t + b\sin\omega t) + i(\cdots)\right) \\ &= a\cos\omega t + b\sin\omega t. \end{aligned}$$

3. Using $\cos(x-y) = \cos x\cos y + \sin x \sin y$ shows that

$$\begin{aligned} A\cos(\omega t-\phi) &= A\cos\phi\cos\omega t + A\sin\phi\sin\omega t \\ &= a\cos\omega t + b\sin\omega t \end{aligned}$$

since $a = A\cos\phi$ and $b = A\sin\phi$ when (A,ϕ) are polar coordinates of (a,b).

4.

$$
\begin{aligned}
a\cos\omega t + b\sin\omega t &= (a,b)\cdot(\cos\omega t, \sin\omega t) \\
&= |(a,b)|\;|(\cos\omega t,\sin\omega t)|\;\cos(\text{angle between the vectors}), \\
&\qquad \text{(by the geometric interpretation of the dot product)} \\
&= A\cos(\omega t - \phi)
\end{aligned}
$$

(Actually, it would have been enough to prove equality on two sides of the triangle.) □

Here are two sample problems showing how to use the key equations.

Problem 9.3. (We actually did this in lecture on Fri Feb 16.) Convert $7\cos(2t - \pi/2)$ into complex form $\operatorname{Re}(ce^{i\omega t})$.

Solution: Given: $A = 7$, $\omega = 2$, $\phi = \pi/2$. Needed: c, ω (actually, we know ω already). The c we need is the one that satisfies the key equation $\overline{c} = Ae^{i\phi}$. This says $\overline{c} = 7e^{i\pi/2} = 7i$, so $c = -7i$. The answer is

$$\operatorname{Re}\left(ce^{i\omega t}\right) = \operatorname{Re}\left(-7ie^{i(2t)}\right). \quad \square$$

Problem 9.4. (Skipped.) Convert $-\cos 5t - \sqrt{3}\sin 5t$ to amplitude-phase form $A\cos(\omega t - \phi)$.

Solution: Given: $a = -1$, $b = -\sqrt{3}$, $\omega = 5$. Wanted: A, ω, ϕ. So we use $Ae^{i\phi} = a + bi$, which says that $Ae^{i\phi} = -1 - i\sqrt{3}$. First, $A = \sqrt{(-1)^2 + (-\sqrt{3})^2} = 2$. The real-part equation $A\cos\phi = -1$ says that $\cos\phi = -1/2$, so the angle of $(-1, -\sqrt{3})$ is $\phi = -2\pi/3$ (or this plus $2\pi k$ for any integer k). Thus the answer is

$$A\cos(\omega t - \phi) = 2\cos(5t + 2\pi/3). \quad \square$$

Remark 9.5. Sinusoids are always real-valued functions. It would be wrong to write one as $a\cos\omega t + bi\sin\theta t$; the i should not be in there. Even the complex form $\operatorname{Re}(ce^{i\omega t})$ of a sinusoid is a real-valued function, because of the Re on the outside.

Try the "Trigonometric Identity" mathlet

`http://mathlets.org/mathlets/trigonometric-id/`

9.5. **Complex gain, gain, and phase lag.** Later in the class, we'll talk about LTI systems (LTI stands for linear and time-invariant). These include all systems built of springs, masses, and dashpots, and also all RLC circuits (circuits built of resistors, inductors, and capacitors).

It turns out that such a system, when fed a sinusoidal input signal, produces a sinusoidal output signal of the same frequency. How can we compare the input and output sinusoids?

Write each sinusoid in complex form (convert if necessary):

$$\text{input signal: } \operatorname{Re}\left(ce^{i\omega t}\right)$$
$$\text{output signal: } \operatorname{Re}\left(Ce^{i\omega t}\right).$$

Imagine feeding a corresponding "complex replacement" signal into the system and getting a complex output signal (this probably makes no physical sense, but do it anyway):

$$\text{complex input: } ce^{i\omega t}$$
$$\text{complex output: } Ce^{i\omega t}.$$

Define complex gain as the factor by which the complex input signal has gotten "bigger":

$$G := \frac{\text{complex output}}{\text{complex input}} = \frac{Ce^{i\omega t}}{ce^{i\omega t}} = \frac{C}{c}.$$

Complex gain is a complex number.

Question 9.6. What is the physical interpretation of the complex gain G, in terms of amplitudes and phases of the real signals?

The answers are in the two boxes below. The conversion $\operatorname{Re}\left(ce^{i\omega t}\right) = A\cos(\omega t - \phi)$ uses the key equation

$$c = Ae^{-i\phi},$$

so multiplying c by the complex scale factor G (to get C) amounts to

- multiplying the amplitude A by $|G|$, and
- increasing ϕ by $-\arg G$.

The amplitude scale factor is called gain:

$$\boxed{\text{gain} := \frac{\text{output amplitude}}{\text{input amplitude}} = |G|}.$$

Gain is a nonnegative real number.

The increase in ϕ is called the phase lag (of the output *relative to the input*):

$$\boxed{\text{phase lag} := \phi_{\text{output}} - \phi_{\text{input}} = -\arg G}.$$

Phase lag is a real number, measured in radians. Warning: This is a *relative* phase lag, different from the absolute phase lag defined earlier comparing a sinusoid to the standard sinusoid $\cos x$.

Example 9.7. If the phase lag is $\pi/2$, that means that the maximum of the output sinusoid occurs $\pi/2$ radians after the maximum of the input signal. (To get instead the relative *time* lag, divide the phase lag by ω.)

Remark 9.8. For an LTI system, it turns out that the complex gain depends only on ω. In other words, the complex gain is the same for all sinusoidal input signals having a fixed angular frequency ω. Mathematically, this is because when the complex input signal $ce^{i\omega t}$ is multiplied by a nonzero complex number, linearity implies that the complex output signal is multiplied by the *same* complex number, so the complex gain ($= \frac{\text{complex output}}{\text{complex input}}$) stays the same.

9.6. **Beats.** Bonus section! Not covered in lecture.

Try the "Beats" mathlet

`http://mathlets.org/mathlets/beats/`

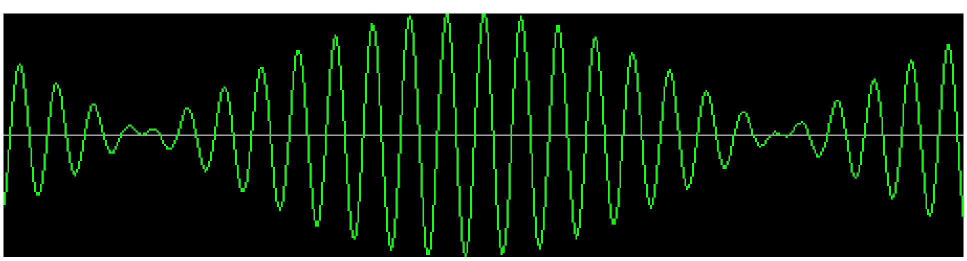
Beats occur when two very nearby pitches are sounded simultaneously.

Problem 9.9. Consider two sinusoid sound waves of angular frequencies $\omega + \epsilon$ and $\omega - \epsilon$, say $\cos((\omega+\epsilon)t)$ and $\cos((\omega-\epsilon)t)$, where ϵ is much smaller than ω. What happens when they are superimposed?

Solution: The sum is

$$\begin{aligned}
\cos((\omega+\epsilon)t) + \cos((\omega-\epsilon)t) &= \operatorname{Re}\left(e^{i(\omega+\epsilon)t}\right) + \operatorname{Re}\left(e^{i(\omega-\epsilon)t}\right) \\
&= \operatorname{Re}\left(e^{i(\omega+\epsilon)t} + e^{i(\omega-\epsilon)t}\right) \\
&= \operatorname{Re}\left(e^{i\omega t}(e^{i\epsilon t} + e^{-i\epsilon t})\right) \\
&= \operatorname{Re}\left(e^{i\omega t}(2\cos\epsilon t)\right) \\
&= (2\cos\epsilon t)\operatorname{Re}\left(e^{i\omega t}\right) \\
&= 2(\cos\epsilon t)(\cos\omega t).
\end{aligned}$$

The function $\cos\omega t$ oscillates rapidly between ± 1. Multiplying it by the slowly varying function $2\cos\epsilon t$ produces a rapid oscillation between $\pm 2\cos\epsilon t$, so one hears a sound wave of angular frequency ω whose amplitude is the slowly varying function $|2\cos\epsilon t|$. □



10. Second-order linear ODEs with constant coefficients

10.1. **Consequence of superposition for a homogeneous linear ODE; vector spaces.**

Problem 10.1 (Multiplying a solution by 9)**.** Fill in the blank:

$$\text{Given that } \cos t \text{ is a solution to } \ddot{x} + x = 0,$$
$$\text{it follows that } 9\cos t \text{ is a solution to } \ddot{x} + x = \underline{\quad\quad}.$$

Answer: $9 \cdot 0$, which is 0 again. Thus $9\cos t$ is a solution to the *same* DE. □

Similarly, adding two solutions to $\ddot{x} + x = 0$ gives another solution to $\ddot{x} + x = 0$.

Conclusion: The set S of solutions to $\ddot{x} + x = 0$ has the following properties:

0. The zero function 0 is in S.
1. Multiplying any one function in S by any scalar gives another function in S.
2. Adding any two functions in S gives another function in S.

A set S of functions satisfying all three properties is called a vector space of functions, since these properties say that you can scalar-multiply and add such functions, as you can with vectors. (One can also talk about vector spaces of vectors, or vector spaces of matrices. There is a more abstract notion of vector space that includes all these as special cases.)

The three properties above hold not just for the set of all solutions to $\ddot{x} + x = 0$, but also for the set of all solutions of any homogeneous linear DE. In other words:

Theorem 10.2. *For any homogeneous linear DE, the set of all solutions is a vector space.*

Theorem 10.2 is why homogeneous linear DEs are so nice. It says that if you know some solutions, you can form linear combinations to build new solutions, with no extra work! **This is the key point of linearity in the homogeneous case.** We will use it over and over again in applications throughout the course.

10.2. **Span.** Last week we used the existence and uniqueness theorem to show that the solutions to $\ddot{x} + x = 0$ are exactly the linear combinations of $\cos t$ and $\sin t$:

$$\{\text{all solutions to } \ddot{x} + x = 0\} \quad = \quad \{a\cos t + b\sin t, \text{ where } a \text{ and } b \text{ range over all numbers}\}.$$

Abbreviation for the set on the right:

$$\operatorname{Span}(\cos t, \sin t).$$

Here is what span means in general:

Definition 10.3. Suppose that $f_1, \ldots, f_n$ is a list of functions.
The span of $f_1, \ldots, f_n$ is the set of *all* linear combinations of $f_1, \ldots, f_n$:

$$\operatorname{Span}(f_1, \ldots, f_n) := \{\text{functions } c_1 f_1 + \cdots + c_n f_n, \text{ where } c_1, \ldots, c_n \text{ range over all numbers}\}.$$

(To be completely precise, we should say whether $c_1, \ldots, c_n$ are ranging over real numbers only, or over all complex numbers too. The answer depends on the context.)

Flashcard question: How many functions are in the set $\operatorname{Span}(\cos t, \sin t)$?

Answer: Infinitely many! Here are some functions in this set:

$$2\cos t + 3\sin t, \quad -5\sin t, \quad 0, \quad \pi\cos t, \quad \ldots.$$

Problem 10.4. Let S be the set of polynomials $p(t)$ whose degree is ≤ 2. Express S as a span.

Solution: The set S equals the set of polynomials of the form

$$at^2 + bt + c,$$

where a, b, c range over all numbers. Thus one possible answer is that $S = \operatorname{Span}(t^2, t, 1)$. □

Problem 10.5. Let T be the set of all solutions to $\dot{y} = 7y$. Express T as a span.

Solution: The set T is the set of functions of the form ce^{7t}, where c ranges over all numbers. Thus one possible answer is that $T = \operatorname{Span}(e^{7t})$. (The linear combinations of a single function are just the scalar multiples of that function.) □

Last week we showed that the set of all solutions to $\ddot{x} + x = 0$ equals the set of the linear combinations of two solutions. The same reasoning applies to *any* 2^{nd}-order homogeneous linear ODE.

Conclusion: For *any* 2^{nd}-order homogeneous linear ODE, the set of solutions can be expressed as the span of 2 solutions.

10.3. **Linearly dependent functions.** How do you know *which* two solutions to use? Will the span of *any* two solutions give the set of all solutions? No! It turns out that most pairs of solutions will work, but not every pair. To determine which pairs work, we need the notion of linear dependence.

Example 10.6 (Why the span of two solutions might not equal the set of all solutions). The functions $\cos t$ and $2\cos t$ are two solutions to $\ddot{x} + x = 0$, but $\operatorname{Span}(\cos t, 2\cos t)$ consists only of the functions

$$a\cos t + b(2\cos t) = (a + 2b)\cos t,$$

which vary only over the scalar multiples of $\cos t$. Thus

$$\operatorname{Span}(\cos t, 2\cos t) = \operatorname{Span}(\cos t).$$

The solution $\sin t$ is missing from this set, so this is not the set of all solutions to $\ddot{x} + x = 0$.

Definition 10.7. Call two functions f and g linearly dependent when either f is a scalar multiple of g or g is a scalar multiple of f. Call f and g linearly independent otherwise.

Warning: The definition of linear dependence for three or more functions is more complicated. We'll discuss it later.

Flashcard question: Which of the following is a pair of linearly independent functions?

- $\cos t$, $2\cos t$
- $\cos t$, $\cos(t+\pi)$
- $\cos t$, $\cos(t-\pi/2)$
- $\cos t$, 0.

Answer: The third pair is linearly independent since $\cos(t-\pi/2)=\sin t$, and neither $\cos t$ nor $\sin t$ is a scalar multiple of the other function (they have zeros in different places, for instance). The second pair is linearly dependent since $\cos(t+\pi)=(-1)(\cos t)$. The fourth pair is linearly dependent since $0=0(\cos t)$. $\square$

It turns out that in solving a 2^{nd}-order homogeneous linear ODE, any two solutions can be used to generate the others, *provided that they are linearly independent*:

Strategy for solving a 2^{nd}-order homogeneous linear ODE:

1. Find two solutions, say f and g, by any method (guessing is OK, as long as you plug f and g in to verify that they are solutions!)
2. Check that f and g are linearly independent (that is, check that f is not a scalar multiple of g and that g is not a scalar multiple of f).
3. If so, then the set of all solutions to the DE is $\operatorname{Span}(f,g)$.

10.4. **Example: a second-order homogeneous linear ODE with constant coefficients.**

Problem 10.8. What are the solutions to

$$\ddot{y}+\dot{y}-6y=0\ ? \tag{7}$$

Solution: Try $y=e^{rt}$, where r is a constant to be determined. Let's find out for which constants r this is really a solution. We get $\dot{y}=re^{rt}$ and $\ddot{y}=r^2e^{rt}$, so (7) becomes

$$\begin{aligned} r^2e^{rt}+re^{rt}-6e^{rt}&=0\\ (r^2+r-6)e^{rt}&=0. \end{aligned}$$

This holds as an equality of functions if and only if

$$\begin{aligned} r^2+r-6&=0\\ (r-2)(r+3)&=0\\ r=2 \quad &\text{or} \quad r=-3. \end{aligned}$$

So e^{2t} and e^{-3t} are solutions.

Luckily, neither is a constant times the other (the ratio is $e^{2t}/e^{-3t} = e^{5t}$, a nonconstant function); in other words, e^{2t} and e^{-3t} are linearly independent. Conclusion:

$$\boxed{\text{The set of all solutions to } \ddot{y} + \dot{y} - 6y = 0 \quad \text{is} \quad \operatorname{Span}(e^{2t}, e^{-3t})}.$$

An equivalent way to express this answer:

$$\boxed{\text{The general solution to } \ddot{y} + \dot{y} - 6y = 0 \text{ is } c_1 e^{2t} + c_2 e^{-3t}, \text{ where } c_1 \text{ and } c_2 \text{ are parameters.}}$$

Remark 10.9. In the future, we'll jump directly from the ODE to solving $r^2 - r - 6 = 0$, now that we know how this works.

10.5. **Basis.** Here is a third way to express the answer to the previous problem:

$$\boxed{\text{The functions } e^{2t} \text{ and } e^{-3t} \text{ form a basis of the space of solutions to } \ddot{y} + \dot{y} - 6y = 0.}$$

This terminology means that

- $\operatorname{Span}(e^{2t}, e^{-3t})$ is the space of solutions to $\ddot{y} + \dot{y} - 6y = 0$, and
- e^{2t} and e^{-3t} are linearly independent.

Think of the functions e^{2t} and e^{-3t} in the basis as the "basic building blocks":

- the first condition says that every solution can be built from e^{2t} and e^{-3t} by taking linear combinations, and
- the second condition says that there is no redundancy in the list (neither building block could have been built from the other one).

The plural of basis is *bases*, pronounced BAY-sees.

February 21

10.6. **How to solve any second-order homogeneous linear ODE with constant coefficients.** To solve

$$a_2 \ddot{y} + a_1 \dot{y} + a_0 y = 0,$$

where a_2, a_1, a_0 are constants (with $a_2 \neq 0$), do the following:

1. Write down the characteristic equation

$$a_2 r^2 + a_1 r + a_0 = 0,$$

in which the coefficient of r^i is the coefficient of $y^{(i)}$ from the ODE. The polynomial $a_2 r^2 + a_1 r + a_0$ is called the characteristic polynomial. (For example, $\ddot{y} + 5y = 0$ has characteristic polynomial $r^2 + 5$.)

2. Solve the characteristic equation to list the complex roots with multiplicity. (To do this, factor the characteristic polynomial, or complete the square, or use the quadratic formula.)

3a. If the roots are *distinct* numbers $r_1 \neq r_2$ then the functions

$$\boxed{e^{r_1 t}, \quad e^{r_2 t}}$$

form a basis of the space of solutions to the ODE. In other words, the general solution is

$$c_1\, e^{r_1 t} + c_2\, e^{r_2 t}.$$

3b. If the roots are equal, r and r, then the functions

$$\boxed{e^{rt}, \quad te^{rt}}$$

form a basis of the space of solutions to the ODE. In other words, the general solution is

$$c_1\, e^{rt} + c_2\, te^{rt}.$$

(We'll explain later on why this works.) □

In both cases, there is a basis consisting of 2 functions.

10.7. **Complex roots.** The general method in the previous section works even if some of the roots are not real.

Problem 10.10. What is a basis for the space of all solutions to $\ddot{x} + x = 0$?

Answer 1 (given last week): $\boxed{\cos t, \sin t}$.
(In other words, the general solution is $c_1 \cos t + c_2 \sin t$.) □

Answer 2 (using the new general method): Characteristic polynomial: $r^2 + 1 = (r-i)(r+i)$. Roots: i and $-i$. Basis: $\boxed{e^{it}, e^{-it}}$. (In other words, the general solution is $c_1 e^{it} + c_2 e^{-it}$.) □

The answers agree since any (complex) linear combination of e^{it} and e^{-it} is also a (complex) linear combination of $\cos t$ and $\sin t$, and vice versa. For example,

$$\begin{aligned} 5e^{it} + 3e^{-it} &= 5(\cos t + i \sin t) + 3(\cos t - i \sin t) \\ &= 8 \cos t + 2i \sin t. \end{aligned}$$

In other words, $\operatorname{Span}(\cos t, \sin t)$ and $\operatorname{Span}(e^{it}, e^{-it})$ (both defined using complex coefficients) are the same set of functions; each is the set of all solutions to $\ddot{x} + x = 0$.

Question 10.11. Which basis should be used, $\boxed{e^{it}, e^{-it}}$ or $\boxed{\cos t, \sin t}$?

Answer: It depends:

- The basis e^{it}, e^{-it} is easier to calculate with, but it's not immediately obvious which linear combinations of these functions are real-valued.
- The basis $\cos t$, $\sin t$ consisting of *real-valued* functions is useful for interpreting solutions in a physical system. The general real-valued solution is $c_1 \cos t + c_2 \sin t$ where c_1, c_2 are real numbers.

So we will be converting back and forth.

The same principles apply more generally:

Complex basis vs. real-valued basis. *Let $y(t)$ be a complex-valued function of a real-valued variable t. If $\boxed{y, \overline{y}}$ is a basis for a space of functions, then $\boxed{\operatorname{Re}(y), \operatorname{Im}(y)}$ is another basis for the same space of functions, but having the advantage that it consists of real-valued functions.*

Proof. (Not done in detail in lecture.) Any linear combination of y and $\overline{y}$ can be re-expressed as a linear combination of $\operatorname{Re}(y)$ and $\operatorname{Im}(y)$ by substituting

$$y = \operatorname{Re}(y) + i\ \operatorname{Im}(y), \qquad \overline{y} = \operatorname{Re}(y) - i\ \operatorname{Im}(y).$$

Conversely, any linear combination of $\operatorname{Re}(y)$ and $\operatorname{Im}(y)$ can be re-expressed as a linear combination of y and $\overline{y}$ by substituting

$$\operatorname{Re}(y) = \frac{y + \overline{y}}{2}, \qquad \operatorname{Im}(y) = \frac{y - \overline{y}}{2i}.$$

Thus $\operatorname{Span}(\operatorname{Re}(y), \operatorname{Im}(y)) = \operatorname{Span}(y, \overline{y})$.

To finish checking the $\operatorname{Re}(y), \operatorname{Im}(y)$ is a basis for the space, we need to check that they are linearly independent. If $\operatorname{Im}(y)$ were a scalar multiple of $\operatorname{Re}(y)$, say $\operatorname{Re}(y) = f(t)$ and $\operatorname{Im}(y) = af(t)$, then y and $\overline{y}$ would be multiples of $f(t)$ too, so one of y and $\overline{y}$ would be a multiple of the other, so y and $\overline{y}$ would be linearly dependent, which is nonsense since by assumption they form a basis. Thus $\operatorname{Im}(y)$ cannot be a scalar multiple of $\operatorname{Re}(y)$. Similarly, $\operatorname{Re}(y)$ cannot be a scalar multiple of $\operatorname{Im}(y)$. Thus $\operatorname{Re}(y), \operatorname{Im}(y)$ are linearly independent. □

Question 10.12. Would it be OK to replace $\boxed{y,\ \overline{y}}$ instead by $\boxed{\operatorname{Re}(y),\ i \operatorname{Im}(y)}$?

Answer: Yes, but this would be less useful, because the whole point was to obtain a basis consisting of *real-valued* functions.

Question 10.13. Would it be OK to replace $\boxed{y,\ \overline{y}}$ instead by $\boxed{\operatorname{Re}(y),\ \operatorname{Re}(\overline{y})}$.

Answer: No, because if $y = f + ig$, then $\overline{y} = f - ig$, so $\operatorname{Re}(y)$ and $\operatorname{Re}(\overline{y})$ are both f! They are linearly dependent, so they can't form a basis.

Here is an example of how this is used in practice:

Problem 10.14. Find a basis of solutions to

$$\ddot{y} + 4\dot{y} + 13y = 0$$

consisting of *real-valued* functions.

Solution:

Characteristic equation: $r^2+4r+13=0$. The quadratic formula, or completing the square (rewriting the polynomial as $(r+2)^2+9$), shows that the roots are $-2+3i$, $-2-3i$. Basis: $e^{(-2+3i)t}$, $e^{(-2-3i)t}$. But these functions are not real-valued!

So replace $y=e^{(-2+3i)t}$ and $\overline{y}=e^{(-2-3i)t}$ by $\operatorname{Re}(y)$ and $\operatorname{Im}(y)$, found by expanding

$$\begin{aligned} e^{(-2+3i)t} &= e^{-2t}e^{i(3t)} \\ &= e^{-2t}\left(\cos(3t)+i\sin(3t)\right) \\ &= e^{-2t}\cos(3t)+i\,e^{-2t}\sin(3t). \end{aligned}$$

Thus

$$\boxed{e^{-2t}\cos(3t), \quad e^{-2t}\sin(3t)}$$

is another basis, this time consisting of real-valued functions. □

10.8. **Harmonic oscillators and damped frequency.** Let's apply all this to the spring-mass-dashpot system, assuming no external force.

10.8.1. *Undamped case.* If there is no damping, the DE is

$$m\ddot{x}+kx=0.$$

Characteristic polynomial: $p(r)=mr^2+k$.
Roots: $\pm\sqrt{-k/m}=\pm i\omega$, where $\boxed{\omega := \sqrt{k/m}}$.
Basis of solution space: $e^{i\omega t}$, $e^{-i\omega t}$.
Real-valued basis: $\cos\omega t$, $\sin\omega t$.
General real solution: $\boxed{c_1\cos\omega t+c_2\sin\omega t}$, where c_1, c_2 are real constants.

In other words, the real-valued solutions are all the sinusoid functions of angular frequency ω. They could also be written as $A\cos(\omega t-\phi)$, where A and ϕ are real constants.

This system, or any other system governed by the same DE, is also called a simple harmonic oscillator. The angular frequency ω is also called the natural frequency (or resonant frequency) of the oscillator.

10.8.2. *Damped case.* If there is damping, the DE is

$$m\ddot{x}+b\dot{x}+kx=0.$$

Characteristic polynomial: $p(r)=mr^2+br+k$.
Roots: $\dfrac{-b\pm\sqrt{b^2-4mk}}{2m}$ (by the quadratic formula)

There are three cases, depending on the sign of b^2-4mk.

Case 1: $b^2 < 4mk$ (underdamped).

Then there are two complex roots $-s \pm i\omega_d$, where

$$s := \frac{b}{2m} \qquad \text{(positive)}$$

$$\text{damped frequency} \quad \omega_d := \frac{\sqrt{4mk - b^2}}{2m} \qquad \text{(positive)}$$

Basis of solution space: $e^{(-s+i\omega_d)t}$, $e^{(-s-i\omega_d)t}$

Real-valued basis: $e^{-st}\cos(\omega_d t)$, $e^{-st}\sin(\omega_d t)$.

General real solution: $\boxed{e^{-st}(c_1\cos(\omega_d t) + c_2\sin(\omega_d t))}$, where c_1, c_2 are real constants.

This is a sinusoid multiplied by a decaying exponential. It can also be written as $e^{-st}(A\cos(\omega_d t - \phi))$ for some A and ϕ. Each nonzero solution tends to 0, but changes sign infinitely many times along the way. The system is called underdamped, because there was not enough damping to eliminate the oscillation completely.

The damping not only causes the solution to decay exponentially, but also *changes the frequency of the sinusoid.* The new angular frequency, ω_d, is called damped frequency. It is less than the undamped frequency we computed earlier (same formula, but with $b = 0$):

$$\omega_d = \frac{\sqrt{4mk - b^2}}{2m} \qquad < \qquad \frac{\sqrt{4mk}}{2m} = \sqrt{\frac{k}{m}} = \omega.$$

Warning: The damped solutions are not actually periodic (they don't repeat exactly, because of the decay). Sometimes $2\pi/\omega_d$ is called the pseudo-period.

Case 2: $b^2 = 4mk$ (critically damped).

There there is a repeated real root: $-\frac{b}{2m}, -\frac{b}{2m}$. Call it $-s$.

Basis of solution space: e^{-st}, te^{-st}.

General real solution: $\boxed{e^{-st}(c_1 + c_2 t)}$, where c_1, c_2 are real constants.

What happens to the solutions as $t \to +\infty$? The solution e^{-st} tends to 0. So does $te^{-st} = \frac{t}{e^{st}}$: even though the numerator t is tending to $+\infty$, the denominator e^{st} is tending to $+\infty$ faster (in a contest between exponentials and polynomials, exponentials always win). Thus all solutions eventually decay.

This case is when there is just enough damping to eliminate oscillation. The system is called critically damped.

Case 3: $b^2 > 4mk$ (overdamped).

In this case, the roots $\frac{-b \pm \sqrt{b^2 - 4mk}}{2m}$ are real and distinct. Both roots are negative, since $\sqrt{b^2 - 4mk} < b$. Call them $-s_1$ and $-s_2$.

General real solution: $\boxed{c_1 e^{-s_1 t} + c_2 e^{-s_2 t}}$, where c_1, c_2 are real constants.

As in all the other damped cases, all solutions tend to 0 as $t \to +\infty$. The term corresponding to the *less negative* root eventually controls the rate of return to equilibrium. The system is called overdamped; there is so much damping that it is slowing the return to equilibrium.

Summary:

Case	Roots	Situation
$b = 0$	two complex roots $\pm i\omega$	undamped (simple harmonic oscillator)
$b^2 < 4mk$	two complex roots $-s \pm i\omega_d$	underdamped (damped oscillator)
$b^2 = 4mk$	repeated real root $-s, -s$	critically damped
$b^2 > 4mk$	distinct real roots $-s_1, -s_2$	overdamped

Midterm 1 covers everything up to here.

February 23

Problem 10.15. Analyze the spring-mass-dashpot system with $m = 1$, $b = 2$, $k = 4$.

Solution: The ODE is

$$\ddot{x} + 2\dot{x} + 4x = 0.$$

Characteristic polynomial: $p(r) = r^2 + 2r + 4 = (r+1)^2 + 3$.
Roots: $-1 \pm i\sqrt{3}$. These are complex, so the system is *underdamped*.
Basis of the solution space: $e^{(-1+i\sqrt{3})t}$, $e^{(-1-i\sqrt{3})t}$.
Real-valued basis: $e^{-t}\cos(\sqrt{3}t)$, $e^{-t}\sin(\sqrt{3}t)$.
General real solution: $\boxed{e^{-t}(a\cos(\sqrt{3}t) + b\sin(\sqrt{3}t))}$, where a, b are real constants.
The damped frequency is $\sqrt{3}$.

Flashcard question: Any nonzero solution $x = e^{-t}(a\cos(\sqrt{3}t) + b\sin(\sqrt{3}t))$ crosses the equilibrium position $x = 0$ infinitely many times.
How much time elapses between two consecutive crossings?

Possible answers:

1. $\pi\sqrt{3}$
2. $\pi/\sqrt{3}$
3. $2\pi\sqrt{3}$
4. $2\pi/\sqrt{3}$
5. $\sqrt{3}/\pi$
6. $\sqrt{3}/(2\pi)$
7. None of the above

Answer: $\pi/\sqrt{3}$. Why? The solution has the same zeros as the sinusoid $a\cos(\sqrt{3}t) + b\sin(\sqrt{3}t)$ of angular frequency $\sqrt{3}$, period $2\pi/\sqrt{3}$. But a sinusoid crosses 0 twice within each period, so the answer is half a period, $\pi/\sqrt{3}$.

Try the "Damped Vibrations" mathlet

`http://mathlets.org/mathlets/damped-vibrations/`

11. Higher-order homogeneous linear ODEs with constant coefficients

Before talking about ODEs of order higher than 2, let's continue the discussion of linear algebra topics such as span, linear dependence, and basis.

11.1. **Vector spaces and span.**

Question 11.1. Is $\operatorname{Span}(f_1, \ldots, f_n)$ always a vector space?

Answer: Yes, because

0. The zero function 0 is a linear combination of $f_1, \ldots, f_n$, namely $0f_1 + \cdots + 0f_n$.
1. Multiplying a linear combination of $f_1, \ldots, f_n$ by a scalar gives another linear combination of $f_1, \ldots, f_n$ (with different coefficients).
2. Adding two linear combinations of $f_1, \ldots, f_n$ gives another linear combination of $f_1, \ldots, f_n$. $\square$

11.2. **Linearly dependent functions.** Recall that *two* functions are called *linearly dependent* when one of them is a scalar multiple of the other one. To motivate the definition for *more than two* functions, first consider the following:

Flashcard question: True or false? The set of all solutions to $\ddot{x} + x = 0$ is

$$\operatorname{Span}(\cos t,\ \sin t,\ 3\cos t + 4\sin t).$$

Answer: TRUE. Let's explain why. The span of these three functions is the set of linear combinations

$$a\cos t + b\sin t + c(3\cos t + 4\sin t).$$

But each such linear combination is also just a linear combination of $\cos t$ and $\sin t$ alone: for example,

$$100\cos t + 10\sin t + 2(3\cos t + 4\sin t) = 106\cos t + 18\sin t.$$

Thus

$$\operatorname{Span}(\cos t,\ \sin t,\ 3\cos t + 4\sin t) = \operatorname{Span}(\cos t, \sin t),$$

which we already know is the set of all solutions to $\ddot{x} + x = 0$. $\square$

Even though the statement was true, including $3\cos t + 4\sin t$ in the list was redundant: it gave no new linear combinations. The general definition of linearly dependent functions captures this notion of redundancy:

Definition 11.2. Functions $f_1, \ldots, f_n$ are linearly dependent (think redundant) if at least one of them is a linear combination of the others. Otherwise, call them linearly independent.

Example 11.3. The three functions $\cos t$, $\sin t$, and $3\cos t + 4\sin t$ are linearly dependent since the third function is a linear combination of the first two.

Remark 11.4. When there are only two functions, f_1 and f_2, then to say that one of them is a linear combination of the others is the same as saying that one of them is a scalar multiple of the other one. So the new definition is compatible with the earlier definition for two functions.

Equivalent definition: Functions $f_1, \ldots, f_n$ are linearly dependent if there exist numbers $c_1, \ldots, c_n$ *not all zero* such that $c_1 f_1 + \cdots + c_n f_n = 0$.

Why is this definition equivalent to the other one just given?

- If there exists such a nontrivial linear combination summing to 0, say
$$3f + 5g + 7h = 0,$$
then one can solve for one of the functions to express it as a linear combination of the others:
$$f = \left(-\frac{5}{3}\right) g + \left(-\frac{7}{3}\right) h.$$
- Conversely, if one of the functions is a linear combination of the others, say,
$$g = 6f + 8h,$$
then moving all the terms to the left side gives a nontrivial linear combination summing to 0:
$$(-6)f + g + (-8)h = 0.$$

Moral: When describing a vector space of functions (such as the set of solutions to a homogeneous linear ODE) as a span, it is most efficient to give it as the span of *linearly independent* functions.

11.3. **Basis.**

Definition 11.5. A basis of a vector space S is a list of functions $f_1, f_2, \ldots$ such that

1. $\operatorname{Span}(f_1, f_2, \ldots) = S$, and
2. The functions $f_1, f_2, \ldots$ are linearly independent.

Flashcard question: What is a basis for the space of solutions to $\dot{y} = 3y$?
Possible answers:

1. This is not a homogeneous linear ODE, so the solutions don't form a vector space. It's a trick question.
2. The function e^{3t} by itself is a basis.
3. The function $2e^{3t}$ by itself is a basis.
4. The basis is the set of all functions of the form ce^{3t}.

Answer: 2 or 3! First, answer 1 is wrong: each term *is* a (constant) function of t times either y or $\dot{y}$, so this *is* a homogeneous linear ODE, and the solutions *do* form a vector space. The basis is supposed to consist of linearly independent functions such that all the solutions can be built from them. Answer 4 is wrong since the functions in the basis are supposed to be linearly independent; if e^{3t} is in a basis, then $5e^{3t}$ should not be in the same basis since it is a linear combination of e^{3t} by itself. Answer 2 is correct since the solutions are exactly the functions ce^{3t} for all numbers c. Answer 3 is correct too since the functions $c(2e^{3t})$ also run through all solutions as c ranges over all numbers.

Key point: A vector space usually has infinitely many functions. To describe it compactly, give a *basis* of the vector space.

11.4. **Dimension.** It turns out that, although a vector space can have different bases, each basis has *the same number of functions* in it.

Definition 11.6. The dimension of a vector space is the number of functions in any basis.

Example 11.7. The space of solutions to $\ddot{x} + x = 0$ is 2-dimensional since the basis $\boxed{\cos t, \sin t}$ has 2 functions. (The basis $\boxed{e^{it}, e^{-it}}$ also has 2 functions.)

Example 11.8. The space of solutions to $\dot{y} = 3y$ is 1-dimensional.

In these two examples, the dimension equals the order of the homogeneous linear ODE. It turns out that this holds in general:

Dimension theorem for a homogeneous linear ODE. *The dimension of the space of solutions to an n^{th} order homogeneous linear ODE is n.*

In other words, the number of parameters needed in the general solution to an n^{th} order homogeneous linear ODE is n.

Remember when we proved that all solutions of $\ddot{x} + x = 0$ were linear combinations of $\cos t$ and $\sin t$, by showing that no matter what the values of $x(0)$ and $\dot{x}(0)$, we could find a linear combination of $\cos t$ and $\sin t$ that solved the DE with the same initial conditions? The same idea proves the dimension theorem for any homogeneous linear ODE.

Let's explain the idea for a 3rd-order ODE, in which we use 0 as starting time in the existence and uniqueness theorem. Define

$$\begin{aligned} f &:= \text{the solution such that } y(0) = 1,\ \dot{y}(0) = 0,\ \ddot{y}(0) = 0 \\ g &:= \text{the solution such that } y(0) = 0,\ \dot{y}(0) = 1,\ \ddot{y}(0) = 0 \\ h &:= \text{the solution such that } y(0) = 0,\ \dot{y}(0) = 0,\ \ddot{y}(0) = 1. \\ \text{Then } af + bg + ch &\text{ is the solution such that } y(0) = a,\ \dot{y}(0) = b,\ \ddot{y}(0) = c, \end{aligned}$$

so every solution $y(t)$, no matter what its values of $y(0)$, $\dot{y}(0)$, $\ddot{y}(0)$ are, is some linear combination $af + bg + ch$. Thus $\boxed{\text{the set of all solutions to the DE is } \operatorname{Span}(f, g, h)}$.

If a, b, c are numbers such that $y := af + bg + ch$ is the zero function, then its values of $y(0)$, $\dot{y}(0)$, $\ddot{y}(0)$ must all be 0, which means that a, b, c are all 0. Thus $\boxed{f, g, h \text{ are linearly independent}}$.

The previous two paragraphs imply that $\boxed{f, g, h}$ is a basis for the space of solutions to the DE. There are 3 functions in this basis, so the dimension of the space of solutions is 3.

11.5. **Solving a homogeneous linear ODE with constant coefficients.** Earlier we gave a method to solve any *second-order* homogeneous linear ODEs with constant coefficients. Now we do the same for n^{th} order for *any* n.

Given

$$a_n y^{(n)} + \cdots + a_1 \dot{y} + a_0 y = 0, \tag{8}$$

where $a_n, \ldots, a_0$ are constants, do the following:

1. Write down the characteristic equation
$$a_n r^n + \cdots + a_1 r + a_0 = 0,$$
in which the coefficient of r^i is the coefficient of $y^{(i)}$ from the ODE. The left hand side is called the characteristic polynomial $p(r)$. (For example, $\ddot{y} + 5y = 0$ has characteristic polynomial $r^2 + 5$.)
2. Factor $p(r)$ as
$$a_n (r - r_1)(r - r_2) \cdots (r - r_n)$$
where $r_1, \ldots, r_n$ are (possibly complex) numbers.
3. If $r_1, \ldots, r_n$ are *distinct*, then the functions $e^{r_1 t}, \ldots, e^{r_n t}$ form a basis for the space of solutions to the ODE (8). In other words, the general solution is
$$c_1 e^{r_1 t} + \cdots + c_n e^{r_n t}.$$
4. If $r_1, \ldots, r_n$ are *not* distinct, then $e^{r_1 t}, \ldots, e^{r_n t}$ cannot be a basis since some of these functions are redundant (definitely not linearly independent!) If a particular root r is

repeated m times, then

$$\begin{array}{rccccc} \text{replace} & \overbrace{e^{rt}, & e^{rt}, & e^{rt}, & \ldots, & e^{rt}}^{m \text{ copies}} \\ \text{by} & e^{rt}, & te^{rt}, & t^2e^{rt}, & \ldots, & t^{m-1}e^{rt}. \end{array}$$

(We'll explain later on why this works.) □

In all cases,

$$\boxed{\#\text{ functions in basis} \quad = \quad \#\text{ roots of } p(r) \text{ counted with multiplicity} \quad = \quad \text{order of ODE}},$$

as predicted by the dimension theorem.

Problem 11.9. Find the general solution to

$$y^{(6)} + 6y^{(5)} + 9y^{(4)} = 0.$$

Solution: The characteristic polynomial is

$$r^6 + 6r^5 + 9r^4 = r^4(r+3)^2,$$

whose roots listed with multiplicity are

$$0, 0, 0, 0, -3, -3.$$

Since the roots are not distinct, the basis is *not*

$$e^{0t}, \quad e^{0t}, \quad e^{0t}, \quad e^{0t}, \quad e^{-3t}, \quad e^{-3t}.$$

We need to replace the first block of four functions, and also the last block of two functions. So the correct basis is

$$\underbrace{e^{0t}, \quad te^{0t}, \quad t^2e^{0t}, \quad t^3e^{0t}}, \quad \underbrace{e^{-3t}, \quad te^{-3t}},$$

which simplifies to

$$1, \quad t, \quad t^2, \quad t^3, \quad e^{-3t}, \quad te^{-3t}.$$

Thus the general solution is

$$c_1 + c_2t + c_3t^2 + c_4t^3 + c_5e^{-3t} + c_6te^{-3t}.$$

(As expected, there is a 6-dimensional space of solutions to this 6^{th} order ODE.) □

Problem 11.10. Find the simplest constant-coefficient homogeneous linear ODE having $(5t+7)e^{-t} - 9e^{2t}$ as one of its solutions.

Solution: The given function is a linear combination of

$$e^{-t}, \quad te^{-t}, \quad e^{2t}$$

so the roots of the characteristic polynomial (with multiplicity) should include -1, -1, 2. So the simplest characteristic polynomial is

$$(r+1)(r+1)(r-2) = r^3 - 3r - 2$$

and the corresponding ODE is

$$y^{(3)} - 3\dot{y} - 2y = 0. \quad \square$$

Remember when for a second-order ODE whose characteristic polynomial had complex roots we got a basis like

$$e^{(-2+3i)t}, \quad e^{(-2-3i)t},$$

consisting of a complex-valued function y and its complex conjugate $\overline{y}$? We explained that it was OK to replace $\boxed{y, \overline{y}}$ by a new basis $\boxed{\operatorname{Re} y, \operatorname{Im} y}$ consisting of real-valued functions.

We can do the same replacement even if $\boxed{y, \overline{y}}$ is just *part* of a basis.

February 26

Midterm 1

February 28

Here is an example of how this is used in practice:

Problem 11.11. Find a basis of solutions to

$$y^{(3)} + 3\ddot{y} + 9\dot{y} - 13y = 0$$

consisting of *real-valued* functions.

Solution: The characteristic polynomial is $p(r) := r^3 + 3r^2 + 9r - 13$. Checking the divisors of -13 (as instructed by the rational root test), we find that 1 is a root, so $r - 1$ is a factor. Long division (or solving for unknown coefficients) produces the other factor:

$$p(r) = (r-1)(r^2 + 4r + 13).$$

Roots: 1, $-2+3i$, $-2-3i$.
Basis: e^t, $e^{(-2+3i)t}$, $e^{(-2-3i)t}$.
Leave e^t as is but replace $y := e^{(-2+3i)t}$ and $\overline{y} = e^{(-2-3i)t}$ to get a new basis

$$\boxed{e^t, \quad e^{-2t}\cos(3t), \quad e^{-2t}\sin(3t)}$$

consisting of real-valued functions. $\square$

12. Inhomogeneous linear ODEs with constant coefficients

12.1. **Operator notation.**

12.1.1. *The operator D.*

- A function, like $f(t) = t^2$, takes an input number and returns another number.
- An operator takes an input *function* and returns another function.

For example, the differential operator $\dfrac{d}{dt}$ takes an input function $y(t)$ and returns $\dfrac{dy}{dt}$. This operator is also called D. So $De^{3t} = 3e^{3t}$, for instance (chain rule).

Any number can be viewed as the "multiply-by-the-number" operator: for instance, the operator 5 transforms the function $\sin t$ into the function $5 \sin t$. (Similarly, a function $f(t)$ can be viewed as the "multiply-by-$f(t)$" operator: the operator t^2 transforms $\cos t$ into $t^2 \cos t$.)

12.1.2. *Multiplying and adding operators.* To apply a product $L_1 L_2$ of operators to a function, first apply L_2 and then apply L_1 to the result: $(L_1 L_2)(f) = L_1(L_2(f))$. For instance, DDy means take the derivative of y, and then take the derivative of the result; therefore we write $D^2 y = \ddot{y}$.

To apply a sum of two operators, apply each operator to the function and add the results. For instance, $(D^2 + D)y = D^2 y + Dy = \ddot{y} + \dot{y}$.

Problem 12.1. Rewrite the ODE

$$2\ddot{y} + 3\dot{y} + 5y = 0,$$

in operator form.

Answer:

$$(2D^2 + 3D + 5)y = 0. \quad \square$$

The same argument shows that every constant-coefficient homogeneous linear ODE

$$a_n y^{(n)} + \cdots + a_0 y = 0$$

can be written simply as

$$p(D)\, y = 0,$$

where p is the characteristic polynomial.

12.2. **Linear operators.** An operator L is linear if

$$L(f + g) = Lf + Lg, \qquad L(af) = a\, Lf$$

for any functions f and g, and any number a. Any linear operator L respects linear combinations, meaning that

$$L(c_1 f_1 + \cdots + c_n f_n) = c_1\, Lf_1 + \cdots + c_n\, Lf_n$$

for any numbers $c_1, \dots, c_n$ and functions $f_1, \dots, f_n$.

The operator D is linear. Also, any "multiply-by-a-function" operator is linear. Multiplying and adding these operators shows that

$$2D^2 + 3D + 5$$

is linear, as is any polynomial $p(D)$.

Remark 12.2. Any homogeneous linear ODE can be written as $Ly = 0$ for some linear operator L, even if the coefficients are *nonconstant* functions of t.

12.3. **Time invariance.**

Remark 12.3. Delaying an input signal $F(t)$ in time by a seconds gives a new input signal $F(t-a)$ (the minus sign is correct: the new input signal has the value at $t = a$ that the old input signal has at $t = 0$).

When p is a polynomial with constant coefficients, then $p(D)$ is time-invariant, which means that

$$\text{if } f(t) \text{ is a solution to } p(D)\, x = F(t)$$

and a is a number,

$$\text{then } f(t-a) \text{ is a solution to } p(D)\, x = F(t-a).$$

In words: if an input signal $F(t)$ is delayed in time by a seconds, then the output signal is delayed by a seconds.

This can simplify the solution to some DEs:

Problem 12.4. Fill in the blank:

$$\text{Given that } x(t) := \frac{1}{2}\cos t + \frac{1}{2}\sin t \text{ is a solution to } \dot{x} + x = \cos t,$$
$$\text{it follows that } \underline{\qquad\qquad\qquad\qquad} \text{ is a solution to } \dot{x} + x = \sin t.$$

(*Hint:* $\sin t = \cos(t - \pi/2)$.)

One answer is

$$\begin{aligned} x(t-\pi/2) &= \frac{1}{2}\cos(t-\pi/2) + \frac{1}{2}\sin(t-\pi/2) \\ &= \frac{1}{2}\sin t - \frac{1}{2}\cos t. \end{aligned}$$

A system that is defined by a linear time-invariant operator is called an LTI system.

12.4. **Shortcut for applying an operator to an exponential function.**

Warm-up problem: If r is a number, what is $(2D^2 + 3D + 5)e^{rt}$?
Solution: First, $De^{rt} = re^{rt}$ and $D^2e^{rt} = r^2e^{rt}$ (keep applying the chain rule). Thus

$$\begin{aligned}(2D^2 + 3D + 5)e^{rt} &= 2r^2e^{rt} + 3re^{rt} + 5e^{rt}\\ &= (2r^2 + 3r + 5)e^{rt}. \quad \square\end{aligned}$$

The same calculation, but with an arbitrary polynomial, proves the general rule:

Theorem 12.5. *For any polynomial p and any number r,*

$$\boxed{p(D)\,e^{rt} = p(r)e^{rt}}.$$

12.5. **Basis of solutions when there are repeated roots.**

Problem 12.6. Find a basis of solutions to $\ddot{y} - 10\dot{y} + 25y = 0$.

Solution:
Characteristic polynomial: $p(r) = r^2 - 10r + 25 = (r-5)^2$.
Roots: 5, 5.
Basis: $\boxed{e^{5t},\ te^{5t}}$. $\square$

But why does this work? Using operators, we can now explain!
In operator form, the DE is

$$(D-5)^2y = 0.$$

The calculation

$$(D-5)^2e^{5t} = p(D)\,e^{5t} \overset{\text{shortcut}}{=} p(5)e^{5t} = 0e^{5t} = 0$$

shows that e^{5t} is one solution. But the DE is second-order, so the basis should have two functions. Taking the second function to be ce^{5t} for a *constant* c does not give a basis, since e^{5t}, ce^{5t} are linearly dependent.

Let's try variation of parameters! Plug in $y = ue^{5t}$, where u is a function to be determined. To calculate $(D-5)^2\,ue^{5t}$, let's apply $D-5$ twice:

$$\begin{aligned}(D-5)\,ue^{5t} &= \left(\dot{u}e^{5t} + u(5e^{5t})\right) - 5ue^{5t}\\ &= \dot{u}e^{5t}.\end{aligned}$$

Similarly,

$$(D-5)^2\,ue^{5t} = \ddot{u}e^{5t}.$$

Thus in order for ue^{5t} to be a solution to $(D-5)^2y = 0$ we must have

$$\begin{aligned}
\ddot{u} &= 0 \\
\dot{u} &= c_1 \\
u &= c_1 t + c_2 \\
y &= (c_1 t + c_2)e^{5t} \\
y &= c_1 t e^{5t} + c_2 e^{5t}.
\end{aligned}$$

In other words, the set of all solutions is $\operatorname{Span}(te^{5t}, e^{5t})$. Neither te^{5t} nor e^{5t} is a *constant* multiple of the other, so they are linearly independent. Thus they form a basis.

A similar approach handles more complicated characteristic polynomials involving many repeated roots.

Remark 12.7. You don't have to go through the discussion of this section each time you want to solve $p(D)y = 0$; we are just explaining why the method given earlier actually works.

12.6. **Exponential response.** For any polynomial p and number r,

$$p(D)\, e^{rt} = p(r)e^{rt},$$

so

$$\underbrace{e^{rt}}_{\text{output signal}} \quad \text{is a particular solution to} \quad p(D)\, y = \underbrace{p(r)e^{rt}}_{\text{input signal}}.$$

New problem: What if the input signal is just e^{rt}?
Answer (superposition): Multiply by $\dfrac{1}{p(r)}$ to get...

Exponential response formula (ERF).
For any polynomial p and any number r such that $p(r) \neq 0$,

$$\boxed{\underbrace{\frac{1}{p(r)}e^{rt}}_{\text{output signal}} \quad \textit{is a particular solution to} \quad p(D)\, y = \underbrace{e^{rt}}_{\text{input signal}}.}$$

In other words, multiply the input signal by the number $\dfrac{1}{p(r)}$ to get an output signal.

Problem 12.8. Find the general solution to $\ddot{y} + 7\dot{y} + 12y = -5e^{2t}$.

Solution:
Characteristic polynomial: $p(r) = r^2 + 7r + 12 = (r+3)(r+4)$.
Roots: -3, -4.
General solution to *homogeneous* equation: $y_h := c_1 e^{-3t} + c_2 e^{-4t}$.

ERF says:

$$\frac{1}{p(2)}e^{2t} \quad \text{is a particular solution to} \quad p(D)\,y = e^{2t};$$

i.e.,

$$\frac{1}{30}e^{2t} \quad \text{is a particular solution to} \quad \ddot{y} + 7\dot{y} + 12y = e^{2t},$$

so

$$\underbrace{-\frac{1}{6}e^{2t}}_{\text{call this } y_p} \quad \text{is a particular solution to} \quad \ddot{y} + 7\dot{y} + 12y = -5e^{2t}.$$

General solution to inhomogeneous equation:

$$\begin{aligned} y &= y_p + y_h \\ &= -\frac{1}{6}e^{2t} + c_1 e^{-3t} + c_2 e^{-4t}. \quad \square \end{aligned}$$

March 2

The existence and uniqueness theorem says that

$$p(D)\,y = e^{rt}$$

should have a solution even if $p(r) = 0$ (when ERF does not apply). Here is how to find a particular solution in this bad case:

Generalized exponential response formula.

If p is a polynomial having r as a root with multiplicity m, then

$$\boxed{\underbrace{z_p = \frac{1}{p^{(m)}(r)}t^m e^{rt}}_{\text{output signal}} \quad \textit{is a particular solution to} \quad p(D)\,z = \underbrace{e^{rt}}_{\text{input signal}}.}$$

In other words, multiply the input signal by t^m, and then multiply by the number $\dfrac{1}{p^{(m)}(r)}$, where $p^{(m)}$ is the m^{th} derivative of p.

(The proof that this works involves a generalized shortcut, obtained by applying $\frac{\partial}{\partial R}$ to the shortcut equation

$$p(D)\,e^{Rt} = p(R)e^{Rt}$$

m times and then setting the variable R equal to the number r. We'll skip it.)

Generalized ERF comes up less often than regular ERF, since in most applications, $p(r) \neq 0$.

12.7. **Sinusoidal response (complex replacement).** Suppose that $p(t)$ is a real polynomial, and ω is a real number. Let z_p denote a complex-valued function.

Problem 12.9. Fill in the blank:

$$\text{If } p(D)\, z_p = e^{i\omega t}$$
$$\text{then } p(D)\,\_\_\_\_ = \cos \omega t.$$

Answer: Taking real parts of both sides shows that $\mathrm{Re}(z_p)$ works.

The observation above leads to the complex replacement method, which we now explain.

Given: The inhomogeneous linear ODE

$$p(D)\, x = \cos \omega t,$$

where p is a real polynomial, and ω is a real number.
Goal: To find *one particular solution* x_p.
Method:

1. Replace $\cos \omega t$ by $e^{i\omega t}$ (whose real part is $\cos \omega t$) and use a different letter z for the unknown function in this new "complex replacement DE":
$$p(D)\, z = \underbrace{e^{i\omega t}}_{\text{complex replacement}}.$$
2. Find a particular solution z_p to the complex replacement DE. (Use ERF for this, provided that $p(i\omega) \neq 0$.)
3. Take the real part: $x_p := \mathrm{Re}(z_p)$. Then x_p is a particular solution to the *original* DE.

Problem 12.10. Find a particular solution x_p to

$$\ddot{x} + \dot{x} + 2x = \cos 2t.$$

Solution: The characteristic polynomial is $p(r) := r^2 + r + 2$.

Step 1. Since $\cos 2t$ is the real part of e^{2it}, replace $\cos 2t$ by e^{2it}:

$$\ddot{z} + \dot{z} + 2z = e^{2it}.$$

Step 2. ERF says that one particular solution to this new ODE is

$$z_p := \frac{1}{p(2i)} e^{2it} = \frac{1}{-2+2i} e^{2it}.$$

Step 3. A particular solution to the original ODE is

$$\boxed{x_p := \mathrm{Re}(z_p) = \mathrm{Re}\left(\frac{1}{-2+2i} e^{2it}\right)}.$$

This is a sinusoid expressed in complex form.

It might be more useful to have the answer in amplitude-phase form or as a linear combination of cos and sin, but we are given $x_p = \operatorname{Re}\left(ce^{2it}\right)$ with $c := \dfrac{1}{-2+2i}$. To convert, we need to rewrite $\overline{c}$ as $Ae^{i\phi}$ or $a + bi$, using

$$\begin{tikzcd} & Ae^{i\phi} \arrow[dl, equal] \arrow[dr, equal] & \\ \overline{c} \arrow[rr, equal] & & a+bi. \end{tikzcd}$$

Converting to amplitude-phase form. The number $-2+2i$ has absolute value $2\sqrt{2}$ and angle $3\pi/4$, so

$$\begin{aligned} -2+2i &= 2\sqrt{2}e^{i(3\pi/4)} \\ c = \frac{1}{-2+2i} &= \frac{1}{2\sqrt{2}}e^{i(-3\pi/4)} \\ \overline{c} &= \frac{1}{2\sqrt{2}}e^{i(3\pi/4)}, \end{aligned}$$

which is supposed to be $Ae^{i\phi}$. Thus the amplitude is $A = \frac{1}{2\sqrt{2}} = \frac{\sqrt{2}}{4}$ and the phase lag is $\phi = 3\pi/4$. Conclusion: In amplitude-phase form,

$$\boxed{x_p = \frac{\sqrt{2}}{4}\cos(2t - 3\pi/4)}.$$

Converting to a linear combination of cos *and* sin. We have

$$\begin{aligned} c &= \frac{1}{-2+2i} = \frac{1}{-2+2i}\left(\frac{-2-2i}{-2-2i}\right) = \frac{-2-2i}{8} = -\frac{1}{4} - \frac{1}{4}i \\ \overline{c} &= -\frac{1}{4} + \frac{1}{4}i, \end{aligned}$$

which is supposed to be $a + bi$, so $a = -1/4$ and $b = 1/4$. Conclusion:

$$\boxed{x_p = -\frac{1}{4}\cos 2t + \frac{1}{4}\sin 2t}.$$

Try the "Amplitude and Phase: Second Order IV" mathlet

`http://mathlets.org/mathlets/amplitude-and-phase-second-order-iv/`

with $m = 1$, $b = 1$, $k = 2$, $\omega = 2$ to see the input signal $\cos 2t$, and output signal (in yellow). Can you see the amplitude and phase lag of the output signal? The red segment indicates the time lag $t_0 = \phi/\omega = (3\pi/4)/2 = 3\pi/8 \approx 1.18$. $\square$

12.8. **Complex gain, gain, and phase lag for an ODE.** The operator $p(D)$ defines an LTI system: when solving

$$p(D)\, x = \cos \omega t,$$

the input signal is $\cos \omega t$ and the output signal is the steady-state solution x_p. Here is what happens in general, assuming $p(i\omega) \neq 0$:

- The complex replacement ODE

$$p(D)\, z = e^{i\omega t}$$

  has complex input signal $e^{i\omega t}$ and complex output signal $\dfrac{1}{p(i\omega)} e^{i\omega t}$ by ERF, so

$$\boxed{\text{complex gain } G = \frac{1}{p(i\omega)}}$$

  (a complex number). We can write $z_p = Ge^{i\omega t}$.
- The original ODE

$$p(D)\, x = \cos \omega t$$

  has sinusoid output signal $x_p := \operatorname{Re}(Ge^{i\omega t})$.
- The angular frequency of the output signal is the same as the angular frequency of the input signal: ω.
- For this system,

$$\boxed{\text{gain} = |G| = \frac{1}{|p(i\omega)|}}$$

  and

$$\boxed{\text{phase lag} = -\arg G = \arg p(i\omega)}.$$

Remark 12.11. What happens if we change the input signal $\cos \omega t$ to a different sinusoidal function of angular frequency ω? As mentioned when we introduced complex gain, this multiplies the complex input signal and the complex output signal by the same complex number, so the complex gain is the same. Thus the complex gain, gain, and phase lag are given by the same formulas as above: they depend only on the system and on ω.

Complex replacement is helpful also with other real input signals, with any real-valued function that can be written as the real part of a reasonably simple complex input signal. Here are some examples:

Real input signal	Complex replacement
$\cos \omega t$	$e^{i\omega t}$
$A\cos(\omega t - \phi)$	$Ae^{-i\phi}e^{i\omega t}$
$a\cos \omega t + b\sin \omega t$	$(a - bi)e^{i\omega t}$
$e^{at}\cos \omega t$	$e^{(a+i\omega)t}$

Each function in the first column is the real part of the corresponding function in the second column. The nice thing about these examples is that the complex replacement is a constant times a complex exponential, so ERF (or generalized ERF) applies.

12.9. **Stability.**

12.9.1. *Steady-state solution, transient.*

Problem 12.12. What is the general solution to $\ddot{x} + 7\dot{x} + 12x = \cos 2t$?

Solution: The characteristic polynomial is

$$p(r) = r^2 + 7r + 12 = (r + 3)(r + 4).$$

The complex gain is

$$G = \frac{1}{p(2i)} = \frac{1}{(2i)^2 + 7(2i) + 12} = \frac{1}{8 + 14i}.$$

Complex replacement and ERF show that

$$x_p = \operatorname{Re}\left(\frac{1}{8 + 14i}e^{2it}\right)$$

is a particular solution.

On the other hand, the general solution to the associated homogeneous ODE is

$$x_h = c_1 e^{-3t} + c_2 e^{-4t}.$$

Therefore the general solution to the original inhomogeneous ODE is

$$\begin{aligned} x &= x_p + x_h \\ &= \underbrace{\operatorname{Re}\left(\frac{1}{8 + 14i}e^{2it}\right)}_{\text{steady-state solution}} + \underbrace{c_1 e^{-3t} + c_2 e^{-4t}}_{\text{transient}}. \quad \square \end{aligned}$$

In general, for a forced damped oscillator, complex replacement and ERF will produce a periodic output signal, and that particular solution is called the steady-state solution. Every other solution is the steady-state solution plus a transient, where the transient is a function that decays to 0 as $t \to +\infty$. As time progresses, the solution $x(t)$ will approximate the steady-state solution more and more closely.

Changing the initial conditions gives a new solution $x_{\text{new}}(t)$. But this changes only the c_1, c_2 above, so $x_{\text{new}}(t)$ will approximate the same steady-state solution that $x(t)$ approximates, and $x_{\text{new}}(t) - x(t)$ will tend to 0 as $t \to +\infty$. A system like this, in which changes in the initial conditions have vanishing effect on the long-term behavior of the solution (i.e., $x_{\text{new}}(t) - x(t)$ tends to 0 as $t \to +\infty$), is called stable.

Try the "Forced Damped Vibration" mathlet

`http://mathlets.org/mathlets/forced-damped-vibration/`

Notice that changing the initial conditions (dragging the yellow square on the left) does not change the long-term behavior of the output signal (yellow curve) much. So this is a stable system.

12.9.2. *Testing a second-order system for stability in terms of roots.* Stability depends on the shape of the solution to the associated homogeneous solution. In the problem above, this was $c_1 e^{-3t} + c_2 e^{-4t}$, which decays as $t \to +\infty$ no matter what c_1 and c_2 are, because -3 and -4 are negative. For a general 2nd-order constant-coefficient linear ODE, stability depends on the roots of the characteristic polynomial, as shown in the following table:

Roots	General solution x_h	Condition for stability	Characteristic poly.
complex $a \pm bi$	$e^{at}(c_1 \cos(bt) + c_2 \sin(bt))$	$a < 0$	$r^2 - 2ar + (a^2 + b^2)$
repeated real s, s	$e^{st}(c_1 + c_2 t)$	$s < 0$	$r^2 - 2sr + s^2$
distinct real r_1, r_2	$c_1 e^{r_1 t} + c_2 e^{r_2 t}$	$r_1, r_2 < 0$	$r^2 - (r_1 + r_2)r + r_1 r_2$

More generally:

Theorem 12.13 (Stability test in terms of roots)**.** *A constant-coefficient linear ODE of any order is stable if and only if every root of the characteristic polynomial has negative real part.*

12.9.3. *Testing a second-order system for stability in terms of coefficients.* In the 2nd-order case, there is also a simple test directly in terms of the coefficients:

Theorem 12.14 (Stability test in terms of coefficients, 2nd-order case)**.** *Assume that a_0, a_1, a_2 are real numbers with $a_0 > 0$. The ODE*

$$(a_0 D^2 + a_1 D + a_2)x = F(t)$$

is stable if and only if $a_1 > 0$ and $a_2 > 0$.

Proof. By dividing by a_0, we can assume that $a_0 = 1$. Break into cases according to the table above.

- When the roots are $a \pm bi$, we have $a < 0$ if and only if the coefficients $-2a$ and $a^2 + b^2$ are both positive.

- When the roots are s, s, we have $s < 0$ if and only if the coefficients $-2s$ and s^2 are both positive.
- When the roots are r_1, r_2, we have $r_1, r_2 < 0$ if and only if the coefficients $-(r_1 + r_2)$ and $r_1 r_2$ are both positive. (Knowing that $-(r_1 + r_2)$ is positive means that at least one of r_1, r_2 is negative; if moreover the product $r_1 r_2$ is positive, then the other root must be negative too.)

Summary: In all three cases, the roots have negative real part if and only if the coefficients of the characteristic polynomial are positive. So to test for stability, instead of checking whether the roots have negative real part, we can check whether the coefficients of the characteristic polynomial are positive. $\square$

There is a generalization of the coefficient test to higher-order ODEs, called the Routh–Hurwitz conditions for stability, but the conditions are much more complicated.

March 5

12.10. **Resonance.** Recall that a harmonic oscillator has a natural frequency. Resonance is a phenomenon that occurs when a harmonic oscillator is driven with an input sinusoid whose frequency is close to or equal to the natural frequency.

- "Near resonance" (frequency is *close to* the natural frequency): The oscillations of the output signal will be much larger than the oscillations of the input signal — the gain will be large. In fact, the closer that the input frequency gets to the natural frequency, the larger the gain becomes.
- "Pure resonance" (frequency is *equal to* the natural frequency): The oscillations grow with time, and there is no steady-state solution. But in a realistic physical situation, there is at least a tiny amount of damping, which prevents the runaway growth, so that the oscillations are bounded (but still large).

We now explain all of this by solving ODEs explicitly.

12.10.1. *Warm-up: harmonic oscillator with no input signal.* A typical ODE modeling a harmonic oscillator is

$$\ddot{x} + 9x = 0.$$

Characteristic polynomial: $r^2 + 9$.
Roots: $\pm 3i$.
Basis of solutions: e^{3it}, e^{-3it}.
Real-valued basis: $\cos 3t$, $\sin 3t$.
General real-valued solution: $a \cos 3t + b \sin 3t$, for real numbers a, b. These are all the sinusoids with angular frequency 3. The natural frequency is 3.

12.10.2. *Near resonance.* Now let's drive the harmonic oscillator with an input sinusoid. A typical ODE modeling this situation is

$$\ddot{x} + 9x = \underbrace{\cos \omega t}_{\text{input signal}}.$$

The complex replacement ODE is

$$\ddot{z} + 9z = e^{i\omega t}.$$

Characteristic polynomial: $p(r) = r^2 + 9$.
Assume that $\omega \neq 3$, so that $i\omega$ is not a root of $p(r)$. Then ERF gives the particular solution

$$z_p := \frac{1}{p(i\omega)} e^{i\omega t} = \frac{1}{9 - \omega^2} e^{i\omega t}.$$

Then a particular solution to the original ODE is

$$x_p := \frac{1}{9 - \omega^2} \cos \omega t.$$

Complex gain: $G = \dfrac{1}{p(i\omega)} = \dfrac{1}{9 - \omega^2}$.
Gain: $|G| = \dfrac{1}{|9 - \omega^2|}$. This becomes very large as ω approaches 3.
Phase lag: $-\arg G$, which is 0 or π depending on whether $\omega < 3$ or $\omega > 3$.

Try the "Harmonic Frequency Response: Variable Input Frequency" mathlet

`http://mathlets.org/mathlets/harmonic-frequency-response-i/`

to see this. (In this mathlet, the natural frequency is 1, and the frequency of the input signal is adjustable. RMS stands for root mean square, which for a sinusoid is amplitude/$\sqrt{2}$.)

In engineering, the graph of gain as a function of ω is called a Bode plot (Bode is pronounced Boh-dee). (Actually, engineers usually instead use a log-log plot: they plot log(gain) as a function of $\log \omega$). On the other hand, a Nyquist plot shows the trajectory of the complex gain G as ω varies.

Also try the "Harmonic Frequency Response: Variable Natural Frequency" mathlet

`http://mathlets.org/mathlets/harmonic-frequency-response-variable-natural-frequency/`

(In this one, the input signal is fixed to be $\sin t$, and the natural frequency is adjustable.)

12.10.3. *Pure resonance.*

Question 12.15. What happens if $\omega = 3$ exactly?

This time, the complex replacement ODE

$$\ddot{z} + 9z = e^{3it}$$

cannot be solved by ERF, since $3i$ is a root of $p(r) = r^2 + 9$. This one requires generalized ERF. First, $p(r)$ has distinct roots $3i$ and $-3i$, so $m = 1$, and $p^{(m)}(r) = p'(r) = 2r = 6i$ at $r = 3i$. Generalized ERF gives

$$\begin{aligned} z_p &:= \frac{1}{6i} t e^{3it} \\ &= -\frac{i}{6} t \left(\cos(3t) + i \sin(3t)\right) \\ &= \frac{1}{6} t \left(-i \cos(3t) + \sin(3t)\right), \end{aligned}$$

so

$$x_p := \frac{1}{6} t \sin(3t)$$

is a particular solution to the original ODE. This is not a sinusoid, but an oscillating function whose oscillations grow without bound as time progresses.

12.10.4. *Resonance with damping.* In a realistic physical situation, there is at least a tiny amount of damping, and this prevents the runaway growth of the previous section.

Question 12.16. What happens if $\omega = 3$ exactly, but there is a tiny amount of damping, so that the ODE is

$$\ddot{x} + \underbrace{b\dot{x}}_{\text{damping term}} + 9x = \underbrace{\cos \omega t}_{\text{input signal}}$$

for some small positive constant b?

New characteristic polynomial: $p(r) = r^2 + br + 9$. Since $3i$ is no longer a root, ERF applies.

Complex gain: $G = \dfrac{1}{p(3i)} = \dfrac{1}{3bi}$.

Gain: $|G| = \dfrac{1}{3b}$. This is large, but the oscillations are bounded; there is a steady-state solution.

Try the "Amplitude and Phase: First Order" mathlet

`http://mathlets.org/mathlets/amplitude-and-phase-1st-order/`

Try the "Amplitude and Phase: Second Order I" mathlet

`http://mathlets.org/mathlets/amplitude-and-phase-2nd-order/`

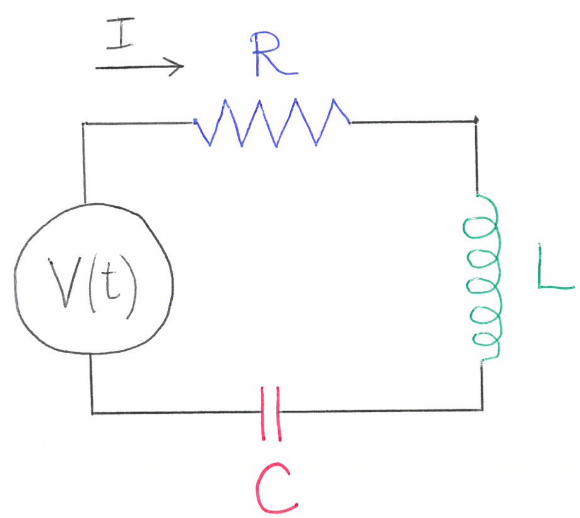
Try the "Amplitude and Phase: Second Order II" mathlet

`http://mathlets.org/mathlets/amplitude-and-phase-2nd-order-ii/`

Try the "Amplitude and Phase: Second Order III" mathlet

`http://mathlets.org/mathlets/amplitude-and-phase-2nd-order-iii/`

12.11. **RLC circuits.** Let's model a circuit with a voltage source, resistor, inductor, and capacitor attached in series: an RLC circuit.



Variables and functions (with SI units):

$$
\begin{aligned}
t &: \text{ time (s)} \\
R &: \text{ resistance of the resistor (ohms)} \\
L &: \text{ inductance of the inductor (henries)} \\
C &: \text{ capacitance of the capacitor (farads)} \\
Q &: \text{ charge on the capacitor (coulombs)} \\
I &: \text{ current (amperes)} \\
V &: \text{ voltage source (volts)} \\
V_R &: \text{ voltage drop across the resistor (volts)} \\
V_L &: \text{ voltage drop across the inductor (volts)} \\
V_C &: \text{ voltage drop across the capacitor (volts).}
\end{aligned}
$$

The independent variable is t. The quantities R, L, C are constants. Everything else is a function of t.

Equations: Physics says

$$\begin{aligned}
I &= \dot{Q} \\
V_R &= RI \qquad \text{Ohm's law} \\
V_L &= L\dot{I} \qquad \text{Faraday's law} \\
V_C &= \frac{1}{C}Q \\
V &= V_R + V_L + V_C \qquad \text{Kirchhoff's voltage law.}
\end{aligned}$$

The last equation can be rearranged into

$$V_L + V_R + V_C = V,$$

which becomes

$$\boxed{L\ddot{Q} + R\dot{Q} + \frac{1}{C}Q = V(t)},$$

a second-order inhomogeneous linear ODE with unknown function $Q(t)$. Mathematically, this is equivalent to the spring-mass-dashpot ODE

$$\boxed{m\ddot{x} + b\dot{x} + kx = F_{\text{external}}(t)},$$

with the following table of analogies:

Spring-mass-dashpot system		RLC circuit	
displacement	x	Q	charge
velocity	$\dot{x}$	I	current
mass	m	L	inductance
damping constant	b	R	resistance
spring constant	k	$1/C$	$1/$capacitance
external force	$F_{\text{external}}(t)$	$V(t)$	voltage source

Similarly, a harmonic oscillator (undamped) is analogous to an LC circuit (no resistor).

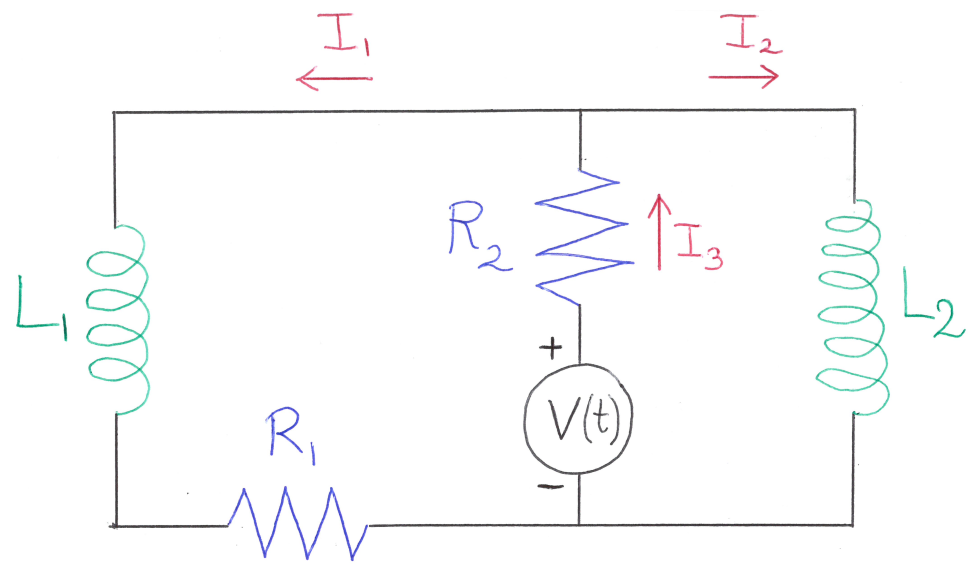
Remark 12.17. Differentiating the DE involving Q gives a DE involving I:

$$L\ddot{I} + R\dot{I} + \frac{1}{C}I = \dot{V}.$$

Try the "Series RLC Circuit" mathlet

`http://mathlets.org/mathlets/series-rlc-circuit/`

13.1. Motivation: modeling a two-loop circuit. Consider a two-loop circuit



in which $R_1 = 8$ ohms, $R_2 = 4$ ohms, $L_1 = 2$ henries, $L_2 = 1$ henry.

- Kirchhoff's current law says that at each junction, the current flowing in equals the current flowing out; applying this at the junction at the top gives

$$I_3 = I_1 + I_2$$

(and the bottom junction gives the same).

- Kirchhoff's voltage law says that around each loop in the circuit, the sum of the electric potential differences (voltages) is 0:

$$\begin{aligned} V - R_2 I_3 - L_1 \dot{I}_1 - R_1 I_1 &= 0 \\ V - R_2 I_3 - L_2 \dot{I}_2 &= 0. \end{aligned}$$

(As one goes around the left loop counterclockwise, the electric potential increases by V as one crosses the voltage source, and then *decreases* as one crosses the resistor R_2 since the current I_3 flows from high potential to low potential, and so on.)

Substitute $I_3 = I_1 + I_2$, and substitute the given values:

$$\begin{aligned} V - 4(I_1 + I_2) - 2\dot{I}_1 - 8I_1 &= 0 \\ V - 4(I_1 + I_2) - \dot{I}_2 &= 0. \end{aligned}$$

Isolate $\dot{I}_1$ and $\dot{I}_2$:

$$\begin{aligned}\dot{I}_1 &= -6I_1 - 2I_2 + V/2 \\ \dot{I}_2 &= -4I_1 - 4I_2 + V,\end{aligned}$$

where V is a function of t alone.

If the function $V(t)$ is given, this is a system in *two* unknown functions $I_1(t)$ and $I_2(t)$. We will develop methods for solving for $I_1(t)$ and $I_2(t)$.

13.2. **Definitions.**

Flashcard question: Consider the system

$$\begin{aligned}\dot{x} &= 2t^2x + 3y \\ \dot{y} &= 5x - 7e^ty\end{aligned}$$

involving *two* unknown functions, $x(t)$ and $y(t)$. Which of the following describes this system?

Possible answers:

- first-order homogeneous linear system of ODEs
- second-order homogeneous linear system of ODEs
- first-order inhomogeneous linear system of ODEs
- second-order inhomogeneous linear system of ODEs
- first-order homogeneous linear system of PDEs
- second-order homogeneous linear system of PDEs
- first-order inhomogeneous linear system of PDEs
- second-order inhomogeneous linear system of PDEs

Answer: It's a first-order homogeneous linear system of ODEs. The system is first-order since it involves only the first derivatives of the unknown functions. This is a homogeneous linear system since every summand is a function of t times one of $x, \dot{x}, \ldots, y, \dot{y}, \ldots$. (If there were also terms that were functions of t, then it would be an inhomogeneous linear system.) The equations are ODEs since the functions are still functions of only one variable, t. $\square$

13.3. **Rewriting a linear system of ODEs in matrix form.**

- The homogeneous system in the flashcard question can be written in matrix form,

$$\begin{array}{ccccc}\begin{pmatrix}\dot{x} \\ \dot{y}\end{pmatrix} & = & \begin{pmatrix}2t^2 & 3 \\ 5 & -7e^t\end{pmatrix} & \begin{pmatrix}x \\ y\end{pmatrix} \\ \dot{\mathbf{x}} & = & A(t) & \mathbf{x},\end{array}$$

by defining

$$\underset{\text{vector-valued function}}{\mathbf{x}} := \begin{pmatrix} x \\ y \end{pmatrix} \qquad \text{and} \qquad \underset{\text{matrix-valued function}}{A(t)} := \begin{pmatrix} 2t^2 & 3 \\ 5 & -7e^t \end{pmatrix}.$$

- Similarly, in the two-loop circuit example, if we define

$$\underset{\text{vector-valued function}}{\mathbf{I}} := \begin{pmatrix} I_1 \\ I_2 \end{pmatrix}, \qquad \text{and} \qquad \underset{\text{vector-valued function}}{\mathbf{q}(t)} := \begin{pmatrix} V(t)/2 \\ V(t) \end{pmatrix}$$

then the inhomogeneous system can be written compactly as

$$\dot{\mathbf{I}} = \begin{pmatrix} -6 & -2 \\ -4 & -4 \end{pmatrix} \mathbf{I} + \mathbf{q}(t).$$

In this example, the matrix-valued function is constant.

March 7

13.4. **Theory.** Before trying to solve such systems, we might want to have some assurance that solutions exist. Fortunately, as in the case of a single linear ODE, they do:

Existence and uniqueness theorem for a linear system of ODEs. *Let $A(t)$ be a matrix-valued function and let $\mathbf{q}(t)$ be a vector-valued function, both continuous on an open interval I. Let $a \in I$, and let $\mathbf{b}$ be a vector. Then there exists a unique solution $\mathbf{x}(t)$ to the system*

$$\dot{\mathbf{x}} = A(t)\,\mathbf{x} + \mathbf{q}(t)$$

satisfying the initial condition $\mathbf{x}(a) = \mathbf{b}$.

(Of course, the sizes of these matrices and vectors should match in order for this to make sense.)

Remark 13.1. Write $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ and $\mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$. Then the vector initial condition $\mathbf{x}(a) = \mathbf{b}$ amounts to n scalar initial conditions: $x_1(a) = b_1, \ldots, x_n(a) = b_n$.

Once the system $\dot{\mathbf{x}} = A(t)\,\mathbf{x}$ and the starting time a are fixed, there is a one-to-one correspondence

$$\{\text{solutions to } \dot{\mathbf{x}} = A(t)\,\mathbf{x}\} \longleftrightarrow \{\text{possibilities for } \mathbf{b}\}$$

under which each solution $\mathbf{x}(t)$ corresponds to its initial condition vector $\mathbf{b} := \mathbf{x}(a)$ (the existence and uniqueness theorem says that for each $\mathbf{b}$, there is one solution $\mathbf{x}(t)$). Adding solutions corresponds to adding their $\mathbf{b}$ vectors, and scalar multiplication of solutions corresponds to scalar multiplication of their $\mathbf{b}$ vectors too. Therefore the concepts of span, linear independence, basis, and dimension on the left side correspond to the same concepts on the right side. In particular,

$$\text{dimension of } \{\text{solutions to } \dot{\mathbf{x}} = A(t)\,\mathbf{x}\} \qquad = \qquad \text{dimension of } \{\text{possibilities for } \mathbf{b}\}.$$

The latter dimension is n, since $\mathbf{b}$ ranges over all vectors in $\mathbb{R}^n$. Conclusion:

Dimension theorem for a homogeneous linear system of ODEs. *For any first-order homogeneous linear system of n ODEs in n unknown functions*

$$\dot{\mathbf{x}} = A(t)\,\mathbf{x},$$

the set of solutions is an n-dimensional vector space.

13.5. **Converting a second-order ODE to a system of two first-order ODEs.**

Problem 13.2. Convert $\ddot{x} + 5\dot{x} + 6x = 0$ to a first-order system of ODEs.

Solution: Introduce a new function variable $y := \dot{x}$. Now try to express the derivatives $\dot{x}$ and $\dot{y}$ in terms of x and y:

$$\begin{aligned} \dot{x} &= y \\ \dot{y} &= \ddot{x} = -5\dot{x} - 6x = -6x - 5y. \end{aligned}$$

In matrix form, this is $\dot{\mathbf{x}} = A\mathbf{x}$ with $A = \begin{pmatrix} 0 & 1 \\ -6 & -5 \end{pmatrix}$. $\square$

(The matrix $\begin{pmatrix} 0 & 1 \\ -6 & -5 \end{pmatrix}$ arising this way is called the companion matrix of the polynomial $r^2 + 5r + 6$.)

Remark 13.3. For constant-coefficient ODEs, the characteristic polynomial of the second-order ODE (scaled, if necessary, to have leading coefficient 1) equals the characteristic polynomial (to be defined soon) of the matrix of the first-order system.

Remark 13.4. Given a *3rd-order* ODE with unknown function x, we can convert it to a system of first-order ODEs by introducing $y := \dot{x}$ and $z := \ddot{x}$. In general, we can convert an n^{th}-order ODE to a system of n first-order ODEs.

Remark 13.5. One can also convert *systems* of higher-order ODEs to systems of first-order ODEs. For example, a system of 4 fifth-order ODEs can be converted to a system of 20 first-order ODEs. That's why it's enough to study *first-order* systems.

13.6. **Converting a system of two first-order ODEs to a second-order ODE.** Conversely, given a system of two first-order ODEs, one can eliminate function variables to find a second-order ODE satisfied by one of the functions.

Problem 13.6. Given that

$$\dot{x} = 2x - y$$
$$\dot{y} = 5x + 7y,$$

find a second-order ODE involving only x.

Solution: Solve for y in the first equation ($y = 2x - \dot{x}$) and substitute into the second:

$$2\dot{x} - \ddot{x} = 5x + 7(2x - \dot{x}).$$

This simplifies to

$$\ddot{x} - 9\dot{x} + 19x = 0. \quad \square$$

Remark 13.7. First-order systems with more than two equations can be converted too, but the conversion is not so easy.

Remark 13.8. In principle, we could solve a first-order linear system of ODEs by first converting it in this way. But usually it is better just to leave it as a system.

14. Homogeneous linear systems of ODEs

14.1. **Guessing solutions.** Consider a first-order 2×2 homogeneous linear system of ODEs with *constant* coefficients:

$$\dot{\mathbf{x}} = A\mathbf{x},$$

where A is an 2×2 matrix with constant entries.

The dimension theorem predicts that the space of solutions is 2-dimensional, so if we are clever enough to guess 2 solutions and they turn out to be linearly independent, then we know the general solution!

The solutions to the ODE $\dot{x} = ax$ were the functions ce^{at}. So let's try

$$\mathbf{x} = e^{\lambda t}\mathbf{v},$$

where λ is a number and $\mathbf{v}$ is a nonzero constant vector. (Note: In contrast with ce^{at}, we put the $e^{\lambda t}$ first when writing $e^{\lambda t}\mathbf{v}$ in order to follow the convention of writing the scalar first in a scalar-vector multiplication. Some people nevertheless write $\mathbf{v}e^{\lambda t}$ instead of $e^{\lambda t}\mathbf{v}$; it means the same thing.)

Question 14.1. For which pairs $(\lambda, \mathbf{v})$ consisting of a scalar and a nonzero vector is the vector-valued function $\mathbf{x} = e^{\lambda t}\mathbf{v}$ a solution to the system $\dot{\mathbf{x}} = A\mathbf{x}$?

Solution: Plug it in, to see what has to happen in order for it to be a solution:

$$\lambda e^{\lambda t}\mathbf{v} = Ae^{\lambda t}\mathbf{v} \qquad \text{(for all } t\text{)}.$$

Interchanging sides and dividing by $e^{\lambda t}$ (also a reversible operation) shows that this is equivalent to

$$\boxed{A\mathbf{v} = \lambda\mathbf{v}}.$$

14.2. **Eigenvalues and eigenvectors.** Given A, we face the problem of

- finding all possible λ, and
- for each λ, finding all vectors $\mathbf{v}$ such that $A\mathbf{v} = \lambda\mathbf{v}$.

This is called the eigenvalue-eigenvector problem.

(Eigen is the German word for "own", as in "the matrix's own vectors" — the eigenvectors belong to the matrix.)

Definition 14.2. Suppose that A is an $n \times n$ matrix.

- An eigenvalue of A is a scalar λ such that $A\mathbf{v} = \lambda\mathbf{v}$ for some nonzero vector $\mathbf{v}$.
- An eigenvector of A associated to a given λ is a vector $\mathbf{v}$ such that $A\mathbf{v} = \lambda\mathbf{v}$.
  (Warning: Some authors require an eigenvector to be nonzero.)

Try the "Matrix Vector" mathlet

`http://mathlets.org/mathlets/matrix-vector/`

Problem 14.3. Let $A = \begin{pmatrix} 1 & -2 \\ -1 & 0 \end{pmatrix}$ and let $\mathbf{v} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$. Is $\mathbf{v}$ an eigenvector of A?

Solution: The calculation

$$A\mathbf{v} = \begin{pmatrix} 1 & -2 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} 2 \\ -1 \end{pmatrix} = \begin{pmatrix} 4 \\ -2 \end{pmatrix} = 2\mathbf{v},$$

shows that $\mathbf{v}$ is an eigenvector, and that the associated eigenvalue is 2. $\square$

In order to find eigenvalues and eigenvectors of a matrix, we need a few concepts from linear algebra.

14.3. **Identity matrix.** Covered in recitation.

The diagonal of a square matrix consists of the entries along the straight line from the upper left to the lower right:

$$\begin{pmatrix} 4 & 6 & 9 \\ 1 & 7 & 8 \\ 2 & 3 & 5 \end{pmatrix}.$$

The $n \times n$ identity matrix is the matrix with ones along the diagonal and zeros elsewhere. For example, the 2×2 identity matrix is

$$I := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Key property: $\boxed{I\mathbf{v} = \mathbf{v} \text{ for any vector } \mathbf{v}.}$ (Check this yourself in the 2×2 case!)

14.4. **Trace.** Covered in recitation.

Definition 14.4. The trace of a square matrix A is the sum of the entries along the diagonal. It is denoted $\operatorname{tr} A$.

Example 14.5. If $A = \begin{pmatrix} 4 & 6 & 9 \\ 1 & 7 & 8 \\ 2 & 3 & 5 \end{pmatrix}$, then $\operatorname{tr} A = 4 + 7 + 5 = 16$.

14.5. **Determinant.** Covered in recitation.

To each square matrix A is associated a number $\det A$ called the determinant.

$$\boxed{\text{Key property: } A\mathbf{v} = \mathbf{0} \text{ has a nonzero solution } \mathbf{v} \text{ if and only if } \det A = 0.}$$

(Thus the determinant "determines" whether a system of linear equations has a nonzero solution.) In the 2×2 case, the determinant is given by the formula

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.$$

Warning: Trace and determinant make sense only for *square* matrices.

14.6. **Characteristic polynomial of a square matrix.** Covered in recitation.

Use λ to denote a scalar-valued variable.

Definition 14.6. The characteristic polynomial of an $n \times n$ matrix A is $\det(\lambda I - A)$. This is a degree n polynomial in the variable λ and its leading coefficient is 1, so the polynomial looks like $\lambda^n + \dots$).

The reason for this definition will be clear in the next section when we show how to compute eigenvalues.

(Warning: This is not the same concept as the characteristic polynomial of a constant-coefficient linear ODE, but there is a connection, arising when such a DE is converted to a first-order system of linear ODEs.)

Remark 14.7. We often calculate the characteristic polynomial using $\det(A - \lambda I)$ instead. This turns out to be the same as $\det(\lambda I - A)$, except negated when n is odd. (The reason is that changing the signs of all n rows of the matrix $A - \lambda I$ flips the sign of the determinant n times.) Usually we care only about the *roots* of the polynomial, so negating the whole polynomial doesn't make a difference. In any case, $\det(A - \lambda I) = \det(\lambda I - A)$ for 2×2 matrices (since 2 is even).

Problem 14.8. What is the characteristic polynomial of $A := \begin{pmatrix} 7 & 2 \\ 3 & 5 \end{pmatrix}$?

Solution: We have

$$\begin{aligned} A - \lambda I &= \begin{pmatrix} 7 & 2 \\ 3 & 5 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} = \begin{pmatrix} 7-\lambda & 2 \\ 3 & 5-\lambda \end{pmatrix} \\ \det(A - \lambda I) &= (7-\lambda)(5-\lambda) - 2(3) = \boxed{\lambda^2 - 12\lambda + 29}. \end{aligned}$$

Here is a shortcut for 2×2 matrices:

Theorem 14.9. *If A is a 2×2 matrix, then the characteristic polynomial of A is*

$$\lambda^2 - (\operatorname{tr} A)\lambda + (\det A).$$

Proof. Write $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Then the characteristic polynomial is

$$\begin{aligned} \det(A - \lambda I) &= \det \begin{pmatrix} a-\lambda & b \\ c & d-\lambda \end{pmatrix} \\ &= (a-\lambda)(d-\lambda) - bc \\ &= \lambda^2 - (a+d)\lambda + (ad - bc) \\ &= \lambda^2 - (\operatorname{tr} A)\lambda + (\det A). \end{aligned}$$

□

We can solve Problem 14.8 again, using this shortcut: the matrix $A := \begin{pmatrix} 7 & 2 \\ 3 & 5 \end{pmatrix}$ has $\operatorname{tr} A = 12$ and $\det A = 29$, so the characteristic polynomial of A is $\lambda^2 - 12\lambda + 29$.

Remark 14.10. Suppose that $n > 2$. Then, for an $n \times n$ matrix A, the characteristic polynomial has the form

$$\lambda^n - (\operatorname{tr} A)\lambda^{n-1} + \cdots \pm \det A$$

where the $\pm$ is $+$ if n is even, and $-$ if n is odd. So knowing $\operatorname{tr} A$ and $\det A$ determines some of the coefficients of the characteristic polynomial, but not all of them.

14.7. Computing all the eigenvalues.

Warm-up problem: Given a square matrix A, how can we test if 5 is an eigenvalue?

Solution: The following are equivalent:

- 5 is an eigenvalue.

- There exists a nonzero solution to

$$\begin{aligned} A\mathbf{v} &= 5\mathbf{v} \\ 5\mathbf{v} - A\mathbf{v} &= \mathbf{0} \\ 5I\mathbf{v} - A\mathbf{v} &= \mathbf{0} \\ (5I - A)\mathbf{v} &= \mathbf{0}. \end{aligned}$$

- $\det(5I - A) = 0$.
- Evaluating the characteristic polynomial $\det(\lambda I - A)$ at 5 gives 0.
- 5 is a root of the characteristic polynomial. □

The same test works for any number in place of 5. (Now that we know how this works, we never again have to go through the argument above.) Conclusion:

$$\boxed{\text{eigenvalues} = \text{roots of the characteristic polynomial}}.$$

Steps to find all the eigenvalues of a square matrix A:

1. Calculate the characteristic polynomial $\det(\lambda I - A)$ or $\det(A - \lambda I)$.
2. The roots of this polynomial are all the eigenvalues of A.

Problem 14.11. Find all the eigenvalues of $A := \begin{pmatrix} 1 & -2 \\ -1 & 0 \end{pmatrix}$.

Solution: We have $\operatorname{tr} A = 1 + 0 = 1$ and $\det A = 0 - 2 = -2$, so the characteristic polynomial is

$$p(\lambda) = \lambda^2 - (\operatorname{tr} A)\lambda + (\det A) = \lambda^2 - \lambda - 2 = (\lambda - 2)(\lambda + 1).$$

Its roots are 2 and -1; these are the eigenvalues. □

The multiplicity of an eigenvalue is just its multiplicity as a root of the characteristic polynomial.

14.8. **Computing eigenvectors.**

Problem 14.12. Find all the eigenvectors of $A := \begin{pmatrix} 1 & -2 \\ -1 & 0 \end{pmatrix}$ associated with the eigenvalue 2.

Solution: By definition, an eigenvector associated to the eigenvalue 2 is a vector $\mathbf{v} = \begin{pmatrix} v \\ w \end{pmatrix}$ satisfying

$$\begin{aligned} A\mathbf{v} &= 2\mathbf{v} \\ (A - 2I)\mathbf{v} &= \mathbf{0} \\ \left(\begin{pmatrix} 1 & -2 \\ -1 & 0 \end{pmatrix} - \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \right) \mathbf{v} &= \mathbf{0} \\ \begin{pmatrix} -1 & -2 \\ -1 & -2 \end{pmatrix} \begin{pmatrix} v \\ w \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \end{aligned}$$

which is equivalent to

$$-v - 2w = 0.$$

We set w to be any number c, and solve for v to get the general solution $\begin{pmatrix} -2c \\ c \end{pmatrix} = c \begin{pmatrix} -2 \\ 1 \end{pmatrix}$.

In other words, the eigenvectors with eigenvalue 2 are all the scalar multiples of $\begin{pmatrix} -2 \\ 1 \end{pmatrix}$. □

Remark 14.13. In this example, the matrix equation became two copies of the same equation $-v - 2w = 0$. More generally, for any 2×2 matrix A and eigenvalue λ, one of the two equations will be a scalar multiple of the other, so again we need to consider only one of them. In particular, the system of two equations will always have a nonzero solution (as there must be, by definition of eigenvalue).

A similar calculation shows that the eigenvectors of A associated with the eigenvalue -1 are the scalar multiples of $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

To summarize:

Steps to find all the eigenvectors associated to a given eigenvalue λ of a 2×2 matrix A:

1. Calculate $A - \lambda I$.
2. Expand $(A - \lambda I)\mathbf{v} = \mathbf{0}$ using $\mathbf{v} = \begin{pmatrix} v \\ w \end{pmatrix}$; this gives a system of two equations in x and y.
3. Solve the system; one of the equations will be redundant, so nonzero solutions will exist.
4. The solution vectors $\begin{pmatrix} v \\ w \end{pmatrix}$ are the eigenvectors associated to λ.

Remark 14.14. Let A be a 2×2 matrix.

- If A is $aI = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$ for some number a, then the only eigenvalue is a (with multiplicity 2), and every vector is an eigenvector with eigenvalue a.
- Otherwise, for each eigenvalue λ, the system $(A-\lambda I)\mathbf{v} = 0$ amounts to one nontrivial equation (the other is redundant), so the eigenvectors associated to λ will be the scalar multiples of a single nonzero vector. In this case, if λ is real, then the set of all real eigenvectors forms a line through the origin, called the eigenline of λ.

14.9. **Solving a 2×2 homogeneous linear system of ODEs with constant coefficients.** Steps to find a basis of solutions to $\dot{\mathbf{x}} = A\mathbf{x}$, given a 2×2 constant matrix A with distinct eigenvalues:

1. Compute the characteristic polynomial $\det(\lambda I - A)$ or $\det(A - \lambda I)$ or $\lambda^2 - (\operatorname{tr} A)\lambda + (\det A)$.
2. Find the roots λ_1 and λ_2 of the characteristic polynomial; these are the eigenvalues.
3. Solve $(A - \lambda_1 I)\mathbf{v} = 0$ to find a nonzero eigenvector $\mathbf{v}_1$ associated with λ_1. (Assuming that λ_1 is not a repeated root, the eigenvectors associated to λ_1 will be just the scalar multiples of $\mathbf{v}_1$.)
4. Solve $(A - \lambda_2 I)\mathbf{v} = 0$ to find a nonzero eigenvector $\mathbf{v}_2$ associated to λ_2.
5. Then $e^{\lambda_1 t}\mathbf{v}_1$ and $e^{\lambda_2 t}\mathbf{v}_2$ form a basis for the space of solutions. (Under our assumption $\lambda_1 \neq \lambda_2$, these two vector-valued functions are linearly independent.)

The "simple" solutions forming a basis, here of the shape $e^{\lambda t}\mathbf{v}$, are sometimes called normal modes. There is not a precise mathematical definition of normal mode, however, since what counts as simple is subjective.

(Lecture actually ended here.)

Problem 14.15. Find the solution to

$$\begin{aligned} \dot{x} &= x - 2y \\ \dot{y} &= -x \\ x(0) &= -1 \\ y(0) &= 8. \end{aligned}$$

Solution: This is $\dot{\mathbf{x}} = A\mathbf{x}$ with $A := \begin{pmatrix} 1 & -2 \\ -1 & 0 \end{pmatrix}$.

We already found the eigenvalues and eigenvectors of A:

- Eigenvalues: 2, -1.

- Eigenvector associated to the eigenvalue 2: $\begin{pmatrix} -2 \\ 1 \end{pmatrix}$.
- Eigenvector associated to the eigenvalue -1: $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

Basis of the space of solutions: $e^{2t}\begin{pmatrix} -2 \\ 1 \end{pmatrix}, \quad e^{-t}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

General solution: $\mathbf{x}(t) = c_1 e^{2t}\begin{pmatrix} -2 \\ 1 \end{pmatrix} + c_2 e^{-t}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

Finally, set $t = 0$ and plug in the initial conditions:

$$\begin{pmatrix} -1 \\ 8 \end{pmatrix} = c_1 \begin{pmatrix} -2 \\ 1 \end{pmatrix} + c_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

This is a system of linear equations

$$\begin{aligned} -2c_1 + c_2 &= -1 \\ c_1 + c_2 &= 8. \end{aligned}$$

Solving it gives $c_1 = 3$ and $c_2 = 5$. Putting these values back into the general solution gives

$$\boxed{\mathbf{x}(t) = 3e^{2t}\begin{pmatrix} -2 \\ 1 \end{pmatrix} + 5e^{-t}\begin{pmatrix} 1 \\ 1 \end{pmatrix}}.$$

In other words,

$$\begin{aligned} x(t) &= -6e^{2t} + 5e^{-t}, \\ y(t) &= 3e^{2t} + 5e^{-t}. \end{aligned}$$

Since there were many opportunities to make errors, it would be wise to check the answer by verifying that these functions satisfy the original DEs and initial condition:

$$\begin{aligned} \dot{x} &= -12e^{2t} - 5e^{-t} = x - 2y \\ \dot{y} &= 6e^{2t} - 5e^{-t} = -x \\ x(0) &= -6 + 5 = -1 \\ y(0) &= 3 + 5 = 8. \quad ☺ \end{aligned}$$

Remark 14.16. The system of linear equations involving c_1 and c_2 could have been written in matrix form:

$$\begin{pmatrix} -2 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} -1 \\ 8 \end{pmatrix}.$$

This point of view will be helpful for more complicated systems.

14.10. **Complex eigenvalues.**

Suppose that A is a real 2×2 matrix whose eigenvalues are not real.

- If λ is one of the eigenvalues, the other is $\overline{\lambda}$.
- If $\mathbf{v}$ is a nonzero eigenvector associated to λ, then $\overline{\mathbf{v}}$ is a nonzero eigenvector associated to $\overline{\lambda}$.
- $\boxed{e^{\lambda t}\mathbf{v},\ e^{\overline{\lambda}t}\,\overline{\mathbf{v}}}$ is a basis for the space of solutions.
- $\boxed{\operatorname{Re}\left(e^{\lambda t}\mathbf{v}\right),\ \operatorname{Im}\left(e^{\lambda t}\mathbf{v}\right)}$ is a basis consisting of *real* vector-valued functions.

14.11. **Phase plane.**

Question 14.17. Let A be a constant 2×2 real matrix. How can you visualize a real-valued solution $\begin{pmatrix} x(t) \\ y(t) \end{pmatrix}$ to $\dot{\mathbf{x}} = A\mathbf{x}$?

Answer: Make two plots, the first showing $x(t)$ as a function of t, and the second showing $y(t)$ as a function of t.

Better answer: Draw the solution as a parametrized curve $\begin{pmatrix} x(t) \\ y(t) \end{pmatrix}$ in the phase plane with axes x and y. In other words, plot the point $\begin{pmatrix} x(t) \\ y(t) \end{pmatrix}$ for every real number t (including negative t). The ODE specifies, in terms of the current position, which direction the phase plane point will move next (and how fast).

Question 14.18. Suppose that λ is a real eigenvalue of A, and $\mathbf{v}$ is a nonzero real eigenvector associated to λ. Then $e^{\lambda t}\mathbf{v}$ is a solution to $\dot{\mathbf{x}} = A\mathbf{x}$. (It is *the* solution satisfying $\mathbf{x}(0) = \mathbf{v}$.) Evaluating $e^{\lambda t}\mathbf{v}$ at any time t gives a positive scalar multiple of $\mathbf{v}$, so the trajectory is contained in the ray through $\mathbf{v}$. What is the direction of the trajectory?

Two approaches:

1. Consider the length and direction of $e^{\lambda t}\mathbf{v}$ as t changes.
2. Use the ODE itself to get the velocity vector at each point.

Answers:

- If $\lambda > 0$, the phase point tends to infinity (repelled from $(0, 0)$).
- If $\lambda < 0$, the phase point tends to $(0, 0)$ (attracted to $(0, 0)$).
- If $\lambda = 0$, the phase point is stationary at $\mathbf{v}$! The point $\mathbf{v}$ is called an critical point since $\dot{\mathbf{x}} = \mathbf{0}$ there. $\square$

Flashcard question: One of the solutions to $\dot{\mathbf{x}} = \begin{pmatrix} -1 & 0 \\ 0 & -2 \end{pmatrix} \mathbf{x}$ is

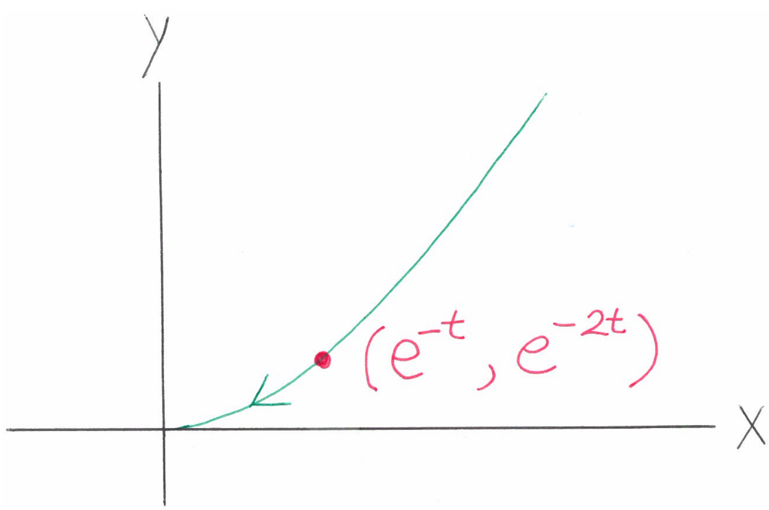
$$\begin{pmatrix} x \\ y \end{pmatrix} = e^{-t} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + e^{-2t} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} e^{-t} \\ e^{-2t} \end{pmatrix}.$$

Which of the following describes the motion in the phase plane with axes x and y as $t \to +\infty$?

Possible answers:

- approaching infinity, along a curve asymptotic to the x-axis
- approaching infinity, along a curve asymptotic to the y-axis
- approaching infinity, along a straight line
- approaching the origin, along a curve tangent to the x-axis
- approaching the origin, along a curve tangent to the y-axis
- approaching the origin, along a straight line
- spiraling
- none of the above

Answer: Approaching the origin, along a curve tangent to the x-axis. As $t \to +\infty$, both $x = e^{-t}$ and $y = e^{-2t}$ tend to 0, but the y-coordinate tends to 0 faster than the x-coordinate, so the trajectory is tangent to the x-axis. (In fact, the y-coordinate is always the square of the x-coordinate, so the trajectory is part of the parabola $y = x^2$.) $\square$



The phase plane trajectory by itself does not describe a solution fully, since it does not show at what time each point is reached. The trajectory contains no information about speed, though one can specify the direction by drawing an arrow on the trajectory.

The phase portrait (or phase diagram) is the diagram showing all the trajectories in the phase plane. We are now ready to classify all possibilities for the phase portrait in terms of the eigenvalue behavior. The most common ones are indicated in green; the others are degenerate cases.

Try the "Linear Phase Portraits: Matrix Entry" mathlet

`http://mathlets.org/mathlets/linear-phase-portraits-matrix-entry/`

See if you can get every possibility listed below.

14.11.1. *Distinct real eigenvalues.* $\boxed{\text{Suppose that the eigenvalues } \lambda_1, \lambda_2 \text{ are real and distinct.}}$ Let $\mathbf{v}_1$, $\mathbf{v}_2$ be corresponding eigenvectors. The set of all eigenvectors associated to λ_1 consists of all scalar multiples of $\mathbf{v}_1$; these form the eigenline of λ_1. General solution:

$$c_1 e^{\lambda_1 t}\mathbf{v}_1 + c_2 e^{\lambda_2 t}\mathbf{v}_2.$$

Opposite sign: $\lambda_1 > 0$, $\lambda_2 < 0$. This is called a saddle. Trajectories flow outward along the positive eigenline (the eigenline of λ_1) and inward along the negative eigenline (the eigenline of λ_2). Other trajectories are asymptotic to both eigenlines, tending to infinity towards the positive eigenline. (Typical solution: $\mathbf{x} = e^{2t}\begin{pmatrix}1\\1\end{pmatrix} + e^{-3t}\begin{pmatrix}-1\\1\end{pmatrix}$. When $t = +1000$, this is approximately a large positive multiple of $\begin{pmatrix}1\\1\end{pmatrix}$. When $t = -1000$, this is approximately a large positive multiple of $\begin{pmatrix}-1\\1\end{pmatrix}$.)

In the next two cases, in which the eigenvalues have the same sign, we'll want to know which is bigger. If $|\lambda_1| > |\lambda_2|$, call λ_1 the fast eigenvalue and λ_2 the slow eigenvalue; use the same adjectives for the eigenlines.

Both positive: $\lambda_1, \lambda_2 > 0$. This is called a repelling node (or node source). All nonzero trajectories flow from $(0,0)$ towards infinity. Trajectories not contained in the eigenlines are tangent to the slow eigenline at $(0,0)$, and far from $(0,0)$ have direction approximately parallel to the fast eigenline.

Both negative: $\lambda_1, \lambda_2 < 0$. This is called an attracting node (or node sink). All nonzero trajectories flow from infinity towards $(0,0)$. Trajectories not contained in the eigenlines are tangent to the slow eigenline at $(0,0)$, and far from $(0,0)$ have direction approximately parallel to the fast eigenline.

One eigenvalue is zero: $\lambda_1 \neq 0$, $\lambda_2 = 0$. General solution: $c_1 e^{\lambda_1 t}\mathbf{v}_1 + c_2\mathbf{v}_2$. This is a degenerate case called a comb. It could also be described by the words nonisolated critical points, since every point on the 0 eigenline is stationary. Other trajectories are along lines parallel to the other eigenline, tending to infinity if $\lambda_1 > 0$, and approaching the 0 eigenline if $\lambda_1 < 0$.

14.11.2. *Complex eigenvalues.* $\boxed{\text{Suppose that the eigenvalues } \lambda_1, \lambda_2 \text{ are not real.}}$ Then they are $a \pm bi$ for some real numbers a, b. In $e^{(a+bi)t}$, the number a controls repulsion/attraction, while b controls rotation (angular frequency).

Zero real part: $a = 0$. This is called a center. The nonzero trajectories are concentric ellipses. Solutions are periodic with period $2\pi/b$.
(Typical solution: $\mathbf{x} = \operatorname{Re}\left[e^{it}\begin{pmatrix} 2 \\ -i \end{pmatrix}\right] = \begin{pmatrix} 2\cos t \\ \sin t \end{pmatrix}$, a parametrization of a fat ellipse.)

Positive real part: $a > 0$. This is called a repelling spiral (or spiral source). All nonzero trajectories spiral outward.

Negative real part: $a < 0$. This is called an attracting spiral (or spiral sink). All nonzero trajectories spiral inward.

In these spiraling or rotating cases, how can one determine whether trajectories go clockwise or counterclockwise? It's complicated to see this in terms of eigenvalues and eigenvectors, but easy to see by testing a single velocity vector. The velocity vector at $\mathbf{x} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ is $\dot{\mathbf{x}} = A\mathbf{x} = A\begin{pmatrix} 1 \\ 0 \end{pmatrix}$; trajectories go counterclockwise if and only if this velocity vector has positive y-coordinate.

14.11.3. *Repeated real eigenvalue.* $\boxed{\text{Suppose that there is a repeated real eigenvalue, say } \lambda, \lambda.}$ The eigenspace of λ (the set of all eigenvectors associated to λ) could be either 1-dimensional or 2-dimensional.

$\lambda \neq 0$ and $A \neq \lambda I$ (1-dimensional eigenspace): This is a called a degenerate node (or improper node or defective node). There is just one eigenline. It serves as both the slow eigenline and the fast eigenline: every trajectory not contained in it is tangent to it at $(0, 0)$, and approximately parallel to it when far from $(0, 0)$. Such trajectories are repelled from $(0, 0)$ if $\lambda > 0$, and attracted to $(0, 0)$ if $\lambda < 0$. (This is a borderline case between a node and a spiral.)

$\lambda \neq 0$ and $A = \lambda I$ (2-dimensional eigenspace): This is a called a star node. Every vector is an eigenvector. Nonzero trajectories are along rays, repelled from $(0, 0)$ if $\lambda > 0$, and attracted to $(0, 0)$ if $\lambda < 0$.

$\lambda = 0$ and $A \neq 0$ (1-dimensional eigenspace): This could be called parallel lines. Points on the eigenline are stationary. All other trajectories are lines parallel to the eigenline.

$\lambda = 0$ and $A = 0$ (2-dimensional eigenspace): This could be called stationary. Every point is stationary.

14.11.4. *Summary.* Although all of the above may be needed for homework problems, for exams you are expected to know only the main cases listed in green above and also the case of a center, not the other "borderline" cases.

Steps to sketch a phase portrait of $\dot{\mathbf{x}} = A\mathbf{x}$ (when A has distinct nonzero eigenvalues):

1. Find the eigenvalues of A.
2. If the eigenvalues are distinct real numbers ($(\operatorname{tr} A)^2 - 4 \det A > 0$) and are nonzero, find and draw the two eigenlines, and indicate the direction of motion along each (repelling/attracting according to eigenvalue being $+/-$).
   - If opposite sign, saddle. Other trajectories are asymptotic to both eigenlines, in the direction matching that of the nearby eigenline.
   - If same sign, then repelling/attracting node. Other trajectories are tangent to the slow eigenline at $(0, 0)$.
3. If the eigenvalues are complex, say $a \pm bi$, check the sign of a:
   - If $+$, repelling spiral.
   - If $-$, attracting spiral.
   - If 0, center.

   To determine whether it is clockwise or counterclockwise, choose a starting vector $\mathbf{x}(0)$ and compute the velocity vector $\dot{\mathbf{x}}(0)$ there as $A\,\mathbf{x}(0)$ to see which direction the particle will move next.

March 12

Problem 14.19. Sketch the phase portrait of the system $\dot{\mathbf{x}} = \begin{pmatrix} -5 & -2 \\ -1 & -4 \end{pmatrix} \mathbf{x}$.

Solution: Call the matrix A. Then $\operatorname{tr} A = -9$ and $\det A = 20 - 2 = 18$, so the characteristic polynomial is $\lambda^2 + 9\lambda + 18 = (\lambda + 6)(\lambda + 3)$. The eigenvalues are the roots, which are -6 and -3.

Eigenvectors of -6: These are the solutions to

$$(A - (-6)I)\mathbf{v} = \mathbf{0}$$
$$\begin{pmatrix} 1 & -2 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} v \\ w \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

which is the linear system

$$v - 2w = 0$$
$$-v + 2w = 0.$$

Here w can be any number c, and then $v = 2c$, so

$$\begin{pmatrix} v \\ w \end{pmatrix} = \begin{pmatrix} 2c \\ c \end{pmatrix} = c \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

Thus the eigenline of -6 is the line through the origin in the direction of $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$.

Eigenvectors of -3: These are the solutions to

$$(A - (-3)I)\mathbf{v} = \mathbf{0}$$
$$\begin{pmatrix} -2 & -2 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} v \\ w \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

which is the linear system

$$\begin{aligned} -2v - 2w &= 0 \\ -v - w &= 0. \end{aligned}$$

Here w can be any number c, and then $v = -c$, so

$$\begin{pmatrix} v \\ w \end{pmatrix} = \begin{pmatrix} -c \\ c \end{pmatrix} = c \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

Thus the eigenline of -3 is the line through the origin in the direction of $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$. This is the slow eigenline, since $|-3| < |-6|$.

The eigenvalues are distinct real numbers, and they are negative, so the phase portrait is an $\boxed{\text{attracting node}}$. The trajectories along the eigenlines tend to $(0, 0)$ as $t \to +\infty$ because the eigenvalues are negative. All other trajectories tend to $(0, 0)$ too, and are tangent at $(0, 0)$ to the slow eigenline (the line in the direction of $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$).

Problem 14.20. Sketch the phase portrait of the system $\dot{\mathbf{x}} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \mathbf{x}$.

Solution: Call the matrix A, so $A = 2I$. Every vector $\mathbf{v}$ satisfies $A\mathbf{v} = 2I\mathbf{v} = 2\mathbf{v}$, so every vector is an eigenvector associated with the eigenvalue 2. At every position in the phase plane, the system $\dot{\mathbf{x}} = A\mathbf{x}$ says that the velocity vector is 2 times the position vector, so every trajectory moves out radially along a ray. This phase portrait is called a $\boxed{\text{star node}}$ (a degenerate case).

(If instead A were $-2I$, then every trajectory would tend to $(0, 0)$ along a ray.)

Problem 14.21. Bonus problem: not done in lecture. Sketch the phase portrait of the system $\dot{\mathbf{x}} = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix} \mathbf{x}$.

Solution: Call the matrix A. Then $\operatorname{tr} A = 4$ and $\det A = 4$. Characteristic polynomial: $\lambda^2 - 4\lambda + 4$. Eigenvalues: 2, 2.

Eigenvectors of 2: These are the solutions to

$$(A - 2I)\mathbf{v} = \mathbf{0}$$
$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} v \\ w \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

which is the linear system

$$w = 0$$
$$0 = 0.$$

Here v can be any number c, but $w = 0$, so

$$\begin{pmatrix} v \\ w \end{pmatrix} = \begin{pmatrix} c \\ 0 \end{pmatrix} = c \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Thus there is only one eigenline, and it is horizontal. Trajectories inside this line (other than the one that sits at $\mathbf{0}$) tend to infinity along the line, since the eigenvalue is positive. All other trajectories are tangent to the eigenline at $(0, 0)$, and tend to infinity as $t \to +\infty$ while becoming approximately parallel to the eigenline. This phase portrait is called a $\boxed{\text{degenerate node}}$.

14.12. **Trace-determinant plane.** The type of phase portrait is determined by the eigenvalues λ_1, λ_2 (except in the case of a repeated eigenvalue, when one needs to know whether A is a scalar times I). And the eigenvalues are determined by the characteristic polynomial

$$\det(\lambda I - A) = \lambda^2 - (\operatorname{tr} A)\lambda + (\det A) = (\lambda - \lambda_1)(\lambda - \lambda_2).$$

(Comparing coefficients shows that $\operatorname{tr} A = \lambda_1 + \lambda_2$ and $\det A = \lambda_1 \lambda_2$.)

Therefore the classification of phase portraits can be re-expressed in terms of $\operatorname{tr} A$ and $\det A$. First, by the quadratic formula, the number of real eigenvalues is determined by the sign of the discriminant $(\operatorname{tr} A)^2 - 4 \det A$.

Only some of the cases below were discussed in lecture.

14.12.1. *Distinct real eigenvalues.* $\boxed{\text{Suppose that } (\operatorname{tr} A)^2 - 4 \det A > 0.}$ Then the eigenvalues are real and distinct.

- If $\det A < 0$, then $\lambda_1 \lambda_2 < 0$, so the eigenvalues have opposite sign: saddle.
- If $\det A > 0$, then the eigenvalues have the same sign.

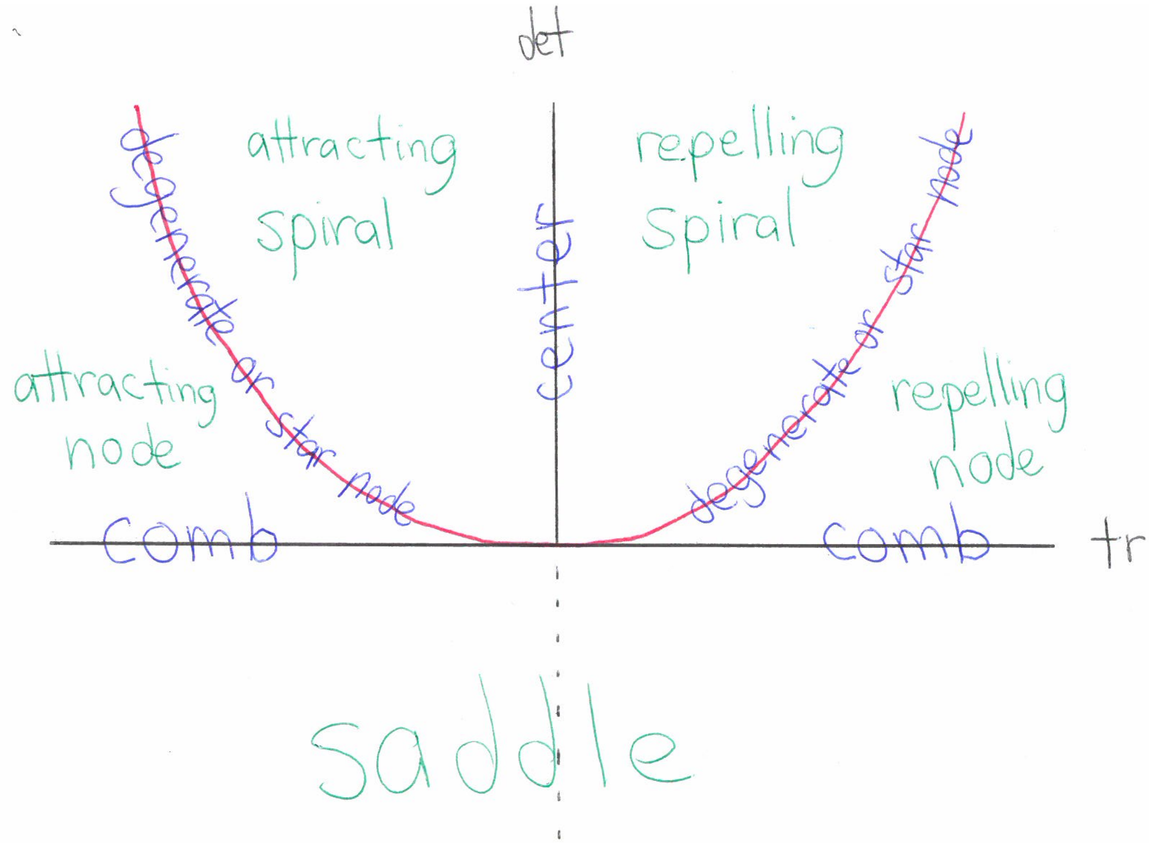
  - If tr $A > 0$, repelling node.
  - If tr $A < 0$, attracting node.
- If $\det A = 0$, then one eigenvalue is 0; comb.

14.12.2. *Complex eigenvalues.* $\boxed{\text{Suppose that } (\operatorname{tr} A)^2 - 4\det A < 0.}$ Then the eigenvalues are $a \pm bi$, and their sum is $\operatorname{tr} A = 2a$.

- If tr $A = 0$, center.
- If tr $A > 0$, repelling spiral.
- If tr $A < 0$, attracting spiral.

14.12.3. *Repeated real eigenvalues.* $\boxed{\text{Suppose that } (\operatorname{tr} A)^2 - 4\det A = 0.}$ Then we get a repeated real eigenvalue λ, λ, and $\operatorname{tr} A = 2\lambda$.

- If tr $A \neq 0$, degenerate node or star node.
- If tr $A = 0$, parallel lines or the stationary case.



The trace-determinant plane is the plane with axes tr and det. This is completely different from the phase plane (because the axes are different).

Whereas the phase portrait shows all possible trajectories for a system $\dot{\mathbf{x}} = A\mathbf{x}$, the trace-determinant plane has just one point for the system. The position of that point contains information about the kind of phase portrait.

Above the parabola $\det = \frac{1}{4}\operatorname{tr}^2$, the eigenvalues are complex. Below the parabola, the eigenvalues are real and distinct.

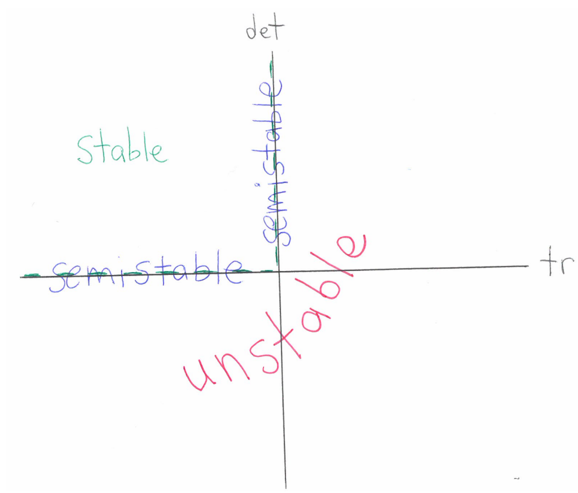
14.13. **Stability.** Consider a system $\dot{\mathbf{x}} = A\mathbf{x}$.

- If all trajectories tend to $\mathbf{0}$ as $t \to +\infty$, the system is called stable.
- If some trajectories are unbounded as $t \to +\infty$, then the system is called unstable.
- In the borderline case in which all solutions are bounded, but do not all tend to $\mathbf{0}$, the system is called semistable or neutrally stable. Example: a center.

The tests for stability are the same as for a single higher-order ODE, in terms of the roots or coefficients of the characteristic polynomial:

$$\begin{aligned} \text{stable} \iff & \text{ all eigenvalues have negative real part} \\ & \text{ (that makes each } e^{\lambda t} \text{ in the general solution tend to 0)} \\ \iff & \text{ the characteristic polynomial has positive coefficients} \\ & \text{ (equivalently, } \operatorname{tr} A < 0 \text{ and } \det A > 0). \end{aligned}$$

(The green test is for the 2×2 case only.)



14.14. **Structural stability.** Stability is a question of what happens to solutions of a fixed system of ODEs. What happens if the system of ODEs itself is changed, by changing the matrix A? There is a new definition to describe this:

Definition 14.22. If the phase portrait type is robust in the sense that small perturbations in the entries of A cannot change the type of the phase portrait, then the system is called structurally stable.

Warning: A system $\dot{\mathbf{x}} = A\mathbf{x}$ can be structurally stable without being stable, and can be stable without being structurally stable. It is unfortunate that the two concepts have similar names, since they are independent of each other.

The structurally stable cases are those corresponding to the large regions in the trace-determinant plane, not the borderline cases. For a 2×2 matrix A, the system $\dot{\mathbf{x}} = A\mathbf{x}$ is structurally stable if and only if A has either

- distinct nonzero real eigenvalues (saddle, repelling node, or attracting node), or
- complex eigenvalues with nonzero real part (spiral).

14.15. **Energy conservation and energy loss.**

14.15.1. *Conservation of energy in the harmonic oscillator.* Consider the harmonic oscillator described by $m\ddot{x} + kx = 0$. Let's check conservation of energy.

Kinetic energy: $\mathrm{KE} = \dfrac{m\dot{x}^2}{2}$.

Potential energy PE is a function of x, and

$$\underbrace{\mathrm{PE}(x) - \mathrm{PE}(0)}_{\text{change in PE}} = \underbrace{-\int_0^x F_{\text{spring}}(X)\,dX}_{\text{work done by } F_{\text{spring}}} = -\int_0^x -kX\,dX = \frac{kx^2}{2}.$$

If we declare $\mathrm{PE} = 0$ at position 0, then $\mathrm{PE}(x) = \dfrac{kx^2}{2}$.

Total energy:

$$E = \mathrm{KE} + \mathrm{PE} = \frac{m\dot{x}^2}{2} + \frac{kx^2}{2}.$$

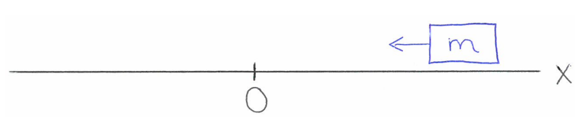
How does total energy change with time?

$$\dot{E} = m\dot{x}\ddot{x} + kx\dot{x} = \dot{x}(m\ddot{x} + kx) = 0.$$

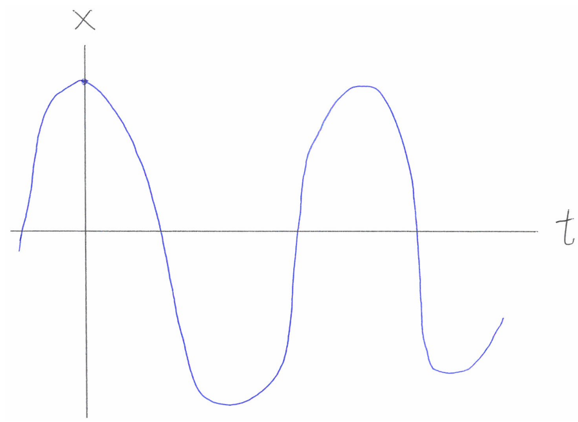
So energy is conserved.

14.15.2. *Phase plane depiction of the harmonic oscillator.* Three ways to depict the harmonic oscillator:

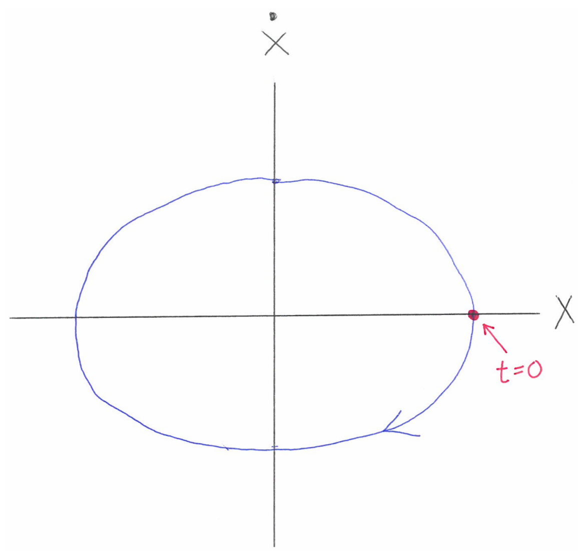
- Movie showing the motion of the mass directly:



- Graph of x as a function of t:



- Trajectory of $\begin{pmatrix} x(t) \\ \dot{x}(t) \end{pmatrix}$ (a parametrized curve) in the phase plane whose horizontal axis shows x and whose vertical axis shows $\dot{x}$ (the phase plane after converting the second-order ODE to a 2×2 system):



Let's start with the mass to the right of equilibrium, and then let go. At $t = 0$, we have $x > 0$ and $\dot{x} = 0$. At the first time the mass crosses equilibrium, $x = 0$ and $\dot{x} < 0$. When the mass reaches its leftmost point, $x < 0$ and $\dot{x} = 0$ again. These give three points on the phase plane trajectory.

March 14

Here are two ways to see that the whole trajectory is an ellipse:

1. We have $x = A\cos\omega t$ for some A and ω. Thus $\dot{x} = -A\omega\sin\omega t$. The parametrized curve

$$\begin{pmatrix} A\cos\omega t \\ -A\omega\sin\omega t \end{pmatrix}$$

is an ellipse. (This is like the parametrization of a circle, but with axes stretched by different amounts.)

2. Rearrange $E = \frac{m\dot{x}^2}{2} + \frac{kx^2}{2}$ as

$$\frac{x^2}{2E/k} + \frac{\dot{x}^2}{2E/m} = 1.$$

This is an ellipse with semi-axes $\sqrt{2E/k}$ and $\sqrt{2Em}$.

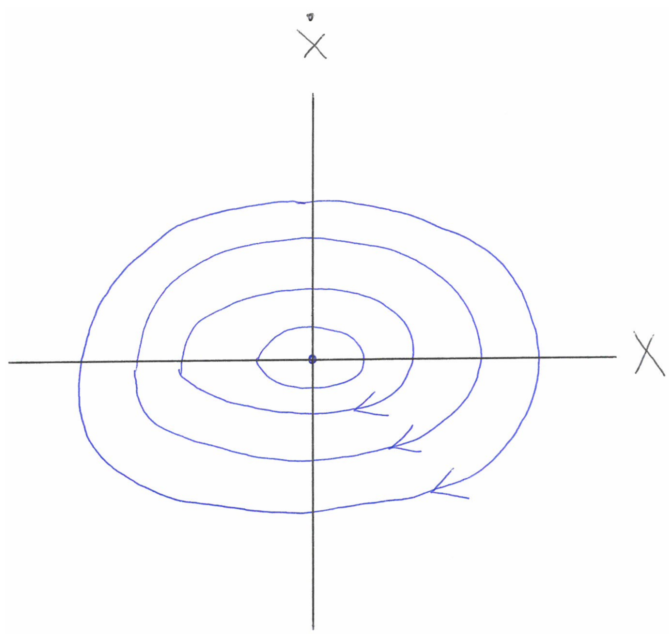
Flashcard question: In which direction is the ellipse traversed?

Possible answers:

1. Clockwise.
2. Counterclockwise.
3. It depends on the initial conditions.

Answer: Clockwise. Above the horizontal axis, $\dot{x} > 0$, which means that x is increasing.

Changing the initial conditions changes E, which changes the ellipse. The family of all such trajectories is a nested family of ellipses, the phase portrait of the system.

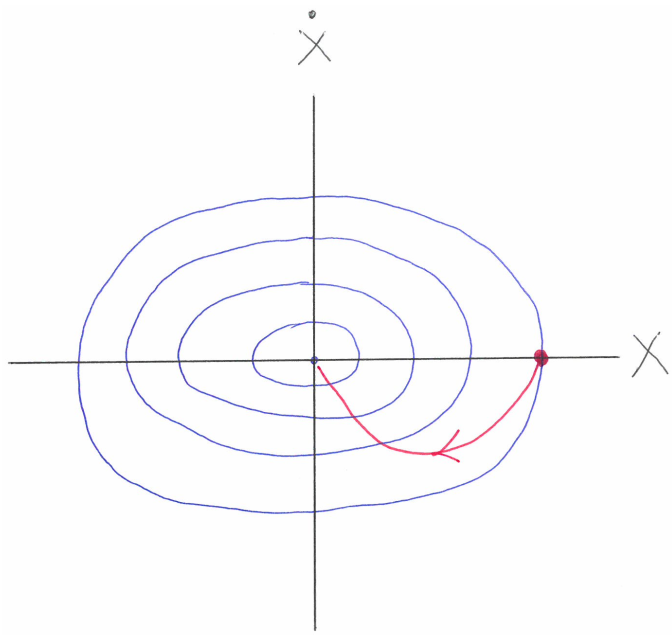


14.15.3. *Energy loss in the damped oscillator.* Now consider a damped oscillator described by $m\ddot{x} + b\dot{x} + kx = 0$. Now

$$\dot{E} = \dot{x}(m\ddot{x} + kx) = -b\dot{x}^2.$$

Energy is lost to friction. The dashpot heats up. The phase plane trajectory crosses through the equal-energy ellipses, inwards towards the origin.

It may spiral in (underdamped case), or approach the origin more directly (critically damped and overdamped cases).



Midterm 2 covers everything up to here.

15. Introduction to linear systems

15.1. **Why?** Recall that complicated real-world problems often give rise to a system of linear ODEs. We package the unknown functions into a vector-valued function $\mathbf{x}(t)$ and write the system in matrix form, such as $\dot{\mathbf{x}} = A\mathbf{x}$. To find a basis of solutions (when the matrix A is constant), we compute the eigenvalues and eigenvectors of A. Finding the eigenvectors involves solving a system of linear equations

$$(A - \lambda I)\mathbf{v} = \mathbf{0},$$

in which A and λ are known and $\mathbf{v}$ is unknown. In the 2×2 case, this is easy, but to handle more complicated situations, we'll need better methods for solving linear systems.

Solving a linear system comes up also when we have a general solution to an ODE (or system of ODEs) and want to use given initial conditions to find the coefficients in the particular solution.

And there are many other applications of solving linear systems in all branches of science and engineering (e.g., balancing a chemical equation).

15.2. **Intersecting lines in $\mathbb{R}^2$.** Given a system of two equations in two unknowns, each equation describes a line (assuming that the equation is not just constant=constant). The solution to the system is the intersection of the two lines.

From geometry, you know that there are three possibilities:

- The lines intersect at one $\boxed{\text{point}}$: one solution. (This is what happens most of the time.) Example:

$$\begin{aligned} x + y &= 1 \\ x - 2y &= 2. \end{aligned}$$

- The lines are the same, so their intersection is a $\boxed{\text{line}}$: infinitely many solutions. Example:

$$\begin{aligned} x + y &= 1 \\ 2x + 2y &= 2. \end{aligned}$$

- The lines are parallel, so their intersection is $\boxed{\text{empty}}$: no solutions. Example

$$\begin{aligned} x + y &= 1 \\ x + y &= 0. \end{aligned}$$

With more equations and more unknowns, there are more possibilities, and we want to describe them all. For this, we need to develop more linear algebra.

15.3. **Functions corresponding to matrices.** A linear system

$$\begin{aligned} 2x + 5y + 7z &= 15 \\ x \qquad + \ \ z &= 1 \end{aligned}$$

can be packaged as an equality of column vectors:

$$\begin{pmatrix} 2x + 5y + 7z \\ x \qquad + \ \ z \end{pmatrix} = \begin{pmatrix} 15 \\ 1 \end{pmatrix}.$$

What kind of thing is the left hand side? It's a vector-valued function $\mathbf{f}$ of an unknown vector $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$: the definition of $\mathbf{f}$ is

$$\mathbf{f} \underbrace{\begin{pmatrix} x \\ y \\ z \end{pmatrix}}_{\text{input}} := \underbrace{\begin{pmatrix} 2x + 5y + 7z \\ x \qquad + \ \ z \end{pmatrix}}_{\text{output}}$$

Problem 15.1. What is $\mathbf{f}\begin{pmatrix} 100 \\ 10 \\ 1 \end{pmatrix}$?

Solution: Plug in $x = 100$, $y = 10$, $z = 1$ to get $\begin{pmatrix} 257 \\ 101 \end{pmatrix}$. $\square$

Here $\mathbf{f}$ is a function from $\mathbb{R}^3$ to $\mathbb{R}^2$ (input has 3 coordinates, output has 2 coordinates).

Problem 15.2. Find the 2×3 matrix A such that $\mathbf{f}(\mathbf{x}) = A\mathbf{x}$.

Solution:

$$\mathbf{f}\begin{pmatrix} x \\ y \\ z \end{pmatrix} := \begin{pmatrix} 2 & 5 & 7 \\ 1 & 0 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix},$$

so $A = \begin{pmatrix} 2 & 5 & 7 \\ 1 & 0 & 1 \end{pmatrix}$. $\square$

One says that A is the matrix representing the function $\mathbf{f}$, and that $\mathbf{f}$ is the function defined by the matrix A.

Key point: In general, $\boxed{\text{a function } \mathbf{f} \text{ from } \mathbb{R}^n \text{ to } \mathbb{R}^m}$ such that each coordinate of the output is a linear combination of the input coordinates is represented by $\boxed{\text{an } m \times n \text{ matrix } A}$. (Warning: m and n get reversed, as in our 2×3 example.)

15.4. **Writing a system in matrix form.** Our system can be written as

$$\mathbf{f}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 15 \\ 1 \end{pmatrix}.$$

or in matrix form $A\mathbf{x} = \mathbf{b}$:

$$\underbrace{\begin{pmatrix} 2 & 5 & 7 \\ 1 & 0 & 1 \end{pmatrix}}_{A}\underbrace{\begin{pmatrix} x \\ y \\ z \end{pmatrix}}_{\mathbf{x}} = \underbrace{\begin{pmatrix} 15 \\ 1 \end{pmatrix}}_{\mathbf{b}}.$$

15.5. **Geometry of a function defined by a matrix.**

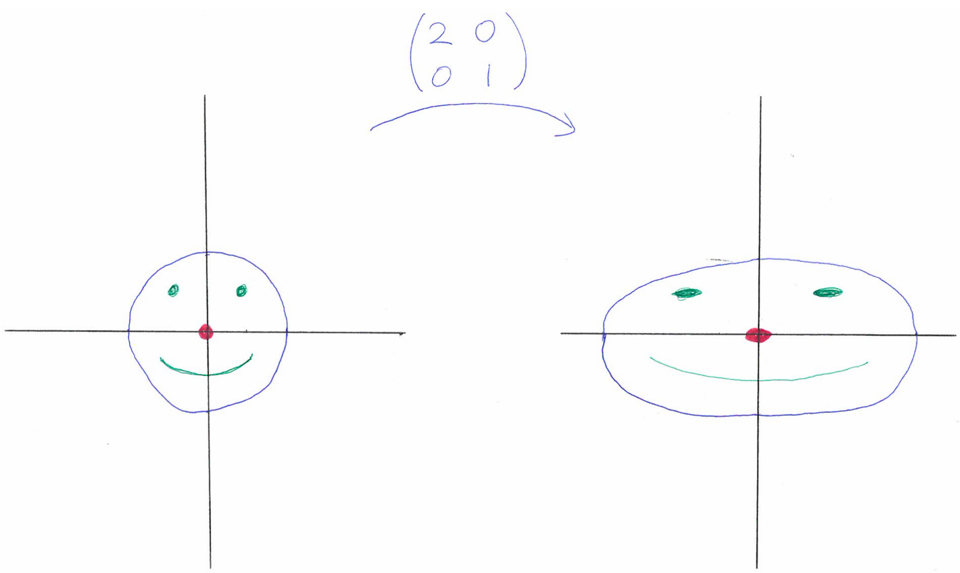
15.5.1. *Depicting a function.* Imagine evaluating $\mathbf{f}$ on every vector in the input space to get vectors in the output space. To visualize it, draw a shape in the input space, apply $\mathbf{f}$ to every point in the shape, and plot the output points in the output space.

Problem 15.3. The matrix $\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$ represents a linear transformation $\mathbf{f}$. Depict $\mathbf{f}$ by showing what it does to the standard basis vectors $\mathbf{i}, \mathbf{j}$ of $\mathbb{R}^2$ and to the unit smiley. What is the area scaling factor?

Solution: We have

$$\mathbf{f}\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}2 & 0\\0 & 1\end{pmatrix}\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}2\\0\end{pmatrix}$$
$$\mathbf{f}\begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}2 & 0\\0 & 1\end{pmatrix}\begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}0\\1\end{pmatrix}$$

and the unit smiley is stretched horizontally into a fat smiley of the same height.



For a 2×2 matrix, the area scaling factor is the absolute value of the determinant:

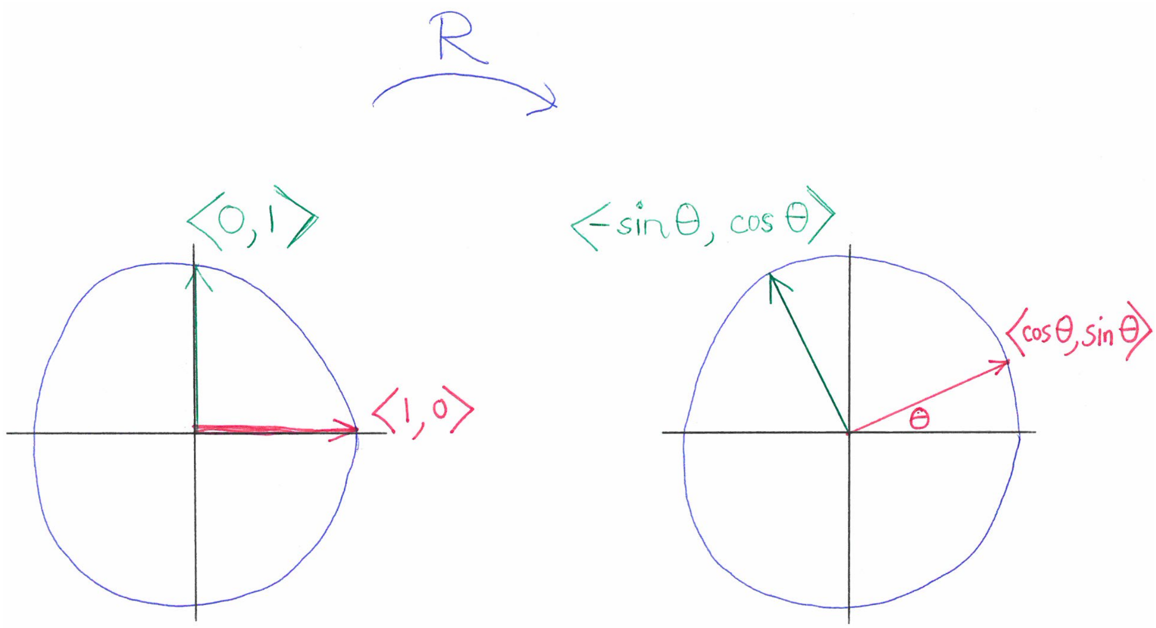
$$\left|\det\begin{pmatrix}2 & 0\\0 & 1\end{pmatrix}\right| = |2| = 2. \quad \square$$

Try the "Matrix Vector" mathlet

`http://mathlets.org/mathlets/matrix-vector/`

15.5.2. *Going from a function to a matrix.*

Problem 15.4. Let $\mathbf{f}$ be the function that rotates each vector in $\mathbb{R}^2$ counterclockwise by the angle θ. What is the corresponding matrix R?



Solution: The rotation maps $\begin{pmatrix}1\\0\end{pmatrix}$ to $\begin{pmatrix}\cos\theta\\\sin\theta\end{pmatrix}$ and $\begin{pmatrix}0\\1\end{pmatrix}$ to $\begin{pmatrix}-\sin\theta\\\cos\theta\end{pmatrix}$. Thus

$$\text{(first column of } R) = R\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}\cos\theta\\\sin\theta\end{pmatrix}$$
$$\text{(second column of } R) = R\begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}-\sin\theta\\\cos\theta\end{pmatrix},$$

so

$$R = \begin{pmatrix}\cos\theta & -\sin\theta\\\sin\theta & \cos\theta\end{pmatrix}.$$

15.6. **Augmented matrix.** A linear system

$$\begin{aligned}2x + 5y + 7z &= 15\\ x \qquad + \ z &= 1\end{aligned}$$

can be written in matrix form $A\mathbf{x} = \mathbf{b}$:

$$\underbrace{\begin{pmatrix}2 & 5 & 7\\1 & 0 & 1\end{pmatrix}}_{A}\underbrace{\begin{pmatrix}x\\y\\z\end{pmatrix}}_{\mathbf{x}} = \underbrace{\begin{pmatrix}15\\1\end{pmatrix}}_{\mathbf{b}}$$

and can be represented by the augmented matrix

$$\left(\begin{array}{ccc|c}2 & 5 & 7 & 15\\1 & 0 & 1 & 1\end{array}\right)$$

(augmented with an extra column containing the right hand sides). Each row corresponds to an equation. Each column except the last one corresponds to a variable.

The four definitions below will be discussed on March 21.

A linear system is homogeneous if the right hand sides (the constants) are all zero, and inhomogeneous otherwise. So a linear system is homogeneous if and only if the zero vector is a solution.

16. Solving linear systems

16.1. **Equation operations.** A good way to solve a linear system is to perform the following operations repeatedly, in some order:

- Multiply an equation by a nonzero number.
- Interchange two equations.
- Add a multiple of one equation to another equation.

The solution set is unchanged at each step.

16.2. **Row operations.** The equation operations correspond to operations on the augmented matrix, called elementary row operations:

- Multiply a row by a nonzero number.
- Interchange two rows.
- Add a multiple of one row to another row (while leaving the first row as it was).

16.3. **Overview of method.**

Overview of how to solve a linear system $A\mathbf{x} = \mathbf{b}$:

1. Use row operations (*Gaussian elimination*) to convert the augmented matrix to a particularly simple form, called *row-echelon form*.
2. Solve the new system by *back-substitution*.

March 16

16.4. **Review.**

16.4.1. *How to check your answers.* To check…

- that a function is a solution to a DE: plug it in (and check initial conditions too).
- that a list of functions is a basis of solutions to a (system of) linear ODE:
  - check that each function is a solution,
  - check that they are linearly independent (no linear combination of them is 0, except for the combination in which all coefficients are 0), and
  - check that the number of functions is as predicted by the dimension theorem.

- that λ is an eigenvalue of A: plug it into the characteristic polynomial to check that it gives 0, or check that $\det(A - \lambda I) = 0$, or try to find an eigenvector (the resulting system should have a redundant equation).
- that $\mathbf{v}$ is an eigenvector of A: evaluate $A\mathbf{v}$ and check that it is a scalar multiple of $\mathbf{v}$.
- a phase portrait for $\dot{\mathbf{x}} = A\mathbf{x}$: compute the velocity vector at a point or two by evaluating $A\mathbf{x}$ at specific points $\mathbf{x}$ (see problem below for an example). Also, check that the point $(\operatorname{tr} A, \det A)$ in the trace-determinant is in the region for the type you expected.

Problem 16.1. Given $A := \begin{pmatrix} 1 & -4 \\ 1 & 1 \end{pmatrix}$, sketch the phase portrait for $\dot{\mathbf{x}} = A\mathbf{x}$.

Solution: We have $\operatorname{tr} A = 1 + 1 = 2$ and $\det A = 1 - (-4) = 5$, so the characteristic polynomial is $\lambda^2 - 2\lambda + 5 = (\lambda - 1)^2 + 4$, and the eigenvalues are $1 \pm 2i$. Since the eigenvalues are not real, and have positive real part, the phase portrait is a repelling spiral (spiral source).

Check: The point $(\operatorname{tr} A, \det A) = (2, 5)$ lies above the parabola $\det = \frac{1}{4}\operatorname{tr}^2$ and to the right of the vertical axis.

Do the trajectories go clockwise or counterclockwise? It's complicated to see this in terms of eigenvalues and eigenvectors, but the velocity vector at $\mathbf{x} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ is

$$\dot{\mathbf{x}} = A\mathbf{x} = \begin{pmatrix} 1 & -4 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

so the trajectories are counterclockwise.

Further question: Is the system stable?

Solution 1: No, because the real parts of the eigenvalues are not negative.

Solution 2: No, because the coefficients of the characteristic polynomial are not all positive.

Solution 3: No, because the $(\operatorname{tr} A, \det A) = (2, 5)$ is in the first quadrant, not in the interior of the second quadrant.

Further question: Is the system structurally stable?

Solution 1: Yes, it is not a boundary case, since the eigenvalues are distinct and nonzero.

Solution 2: Yes, because the point $(\operatorname{tr} A, \det A) = (2, 5)$ is not on the horizontal axis, the positive vertical axis, or the curve $\det = \frac{1}{4}\operatorname{tr}^2$.

Solution 3: Yes, repelling spiral is one of the structurally stable types.

Further question: What kind of function is the second coordinate $y(t)$ of $\mathbf{x}(t)$?

Solution: If $\mathbf{v}$ is a nonzero eigenvector associated to $1+2i$, then $\overline{\mathbf{v}}$ is a nonzero eigenvector associated to $1-2i$, so the vector-valued functions

$$e^{(1+2i)t}\mathbf{v}, \quad e^{(1-2i)t}\overline{\mathbf{v}}$$

form a basis of the space of solutions. Therefore $\mathbf{x}(t)$ is a linear combination of these. Taking second coordinates shows that $y(t)$ is a linear combination of $e^{(1+2i)t}$ and $e^{(1-2i)t}$. We can replace the latter two functions by the real and imaginary parts of the first of them, so $y(t)$ is also a linear combination of $e^t\cos(2t)$ and $e^t\sin(2t)$. In other words, $y(t)$ is e^t times a sinusoidal function with $\omega = 2$:

$$y(t) = e^t A\cos(2t-\phi)$$

for some A and ϕ.

Further question: If $\mathbf{x}(0)$ is on the positive x-axis, what is the next time t that $\mathbf{x}(t)$ lies on the x-axis?

Solution 1: We have

$$y(t) = c_1 e^t\cos(2t) + c_2 e^t\sin(2t)$$

for some real numbers c_1 and c_2. Since $\mathbf{x}(0)$ is on the positive x-axis, $y(0) = 0$. Plugging $t = 0$ into the formula for $y(t)$ gives $0 = c_1$, so

$$y(t) = c_2 e^t\sin(2t).$$

The first positive time at which $y(t) = 0$ is when the angle $2t$ equals π, so $t = \pi/2$.

Solution 2: The sinusoidal function $y(t)$ crosses 0 twice within each period. The period is $P = 2\pi/\omega = \pi$, so the time interval between crossings is $P/2 = \pi/2$.

Further question: At that first time when it crosses the x-axis again, what is its distance to the origin compared to its initial distance to the origin?

Solution: The y-coordinate is 0 at both $t = 0$ and $t = \pi/2$, so we need only study $x(t)$. The same argument as for $y(t)$ shows that $x(t) = e^t A'\cos(2t-\phi')$ for some $A' > 0$ and some ϕ'. When t goes from 0 to $\pi/2$, the angle $2t-\phi'$ increases by π, so the cosine changes sign, while e^t increases from 1 to $e^{\pi/2}$. Thus the distance is multiplied by $e^{\pi/2}$.

Problem 16.2. Estimate the angular frequency ω for which the steady-state solution to

$$(D^3 + D^2 + 4D + 3.9)x = \cos(\omega t)$$

has largest amplitude.

Solution: Let $p(r) = r^3 + r^2 + 4r + 3.9$. The complex replacement ODE

$$p(D)\, z = e^{i\omega t}$$

has a solution $\dfrac{1}{p(i\omega)}e^{i\omega t}$, by ERF, as long as $p(i\omega) \neq 0$. Thus the complex gain is $\dfrac{1}{p(i\omega)}$ and the gain is $\dfrac{1}{|p(i\omega)|}$, which is largest when $i\omega$ is close to a root of $p(r)$.

Now

$$p(r) \approx q(r) := r^3 + r^2 + 4r + 4 = (r+1)(r^2+4).$$

The roots of $q(r)$ are -1 and $\pm 2i$ and these are close to the roots of $p(r)$. In particular, the positive numbers ω such that $i\omega$ is close to a root of $p(r)$ are the numbers ω close to 2. Thus the amplitude of the solution is maximized for a value of ω close to 2.

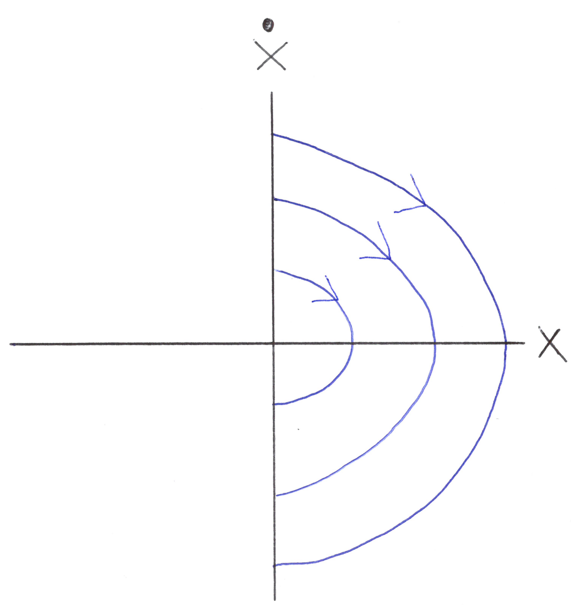
Problem 16.3. Not actually done in lecture. A ball is thrown straight upward from the ground. Let $x(t)$ be its height in meters after t seconds (up until it returns to the ground). Sketch the possible trajectories in the $(x, \dot{x})$ phase plane.

Solution: Let m be the mass of the ball. Kinetic energy is $\dfrac{m\dot{x}^2}{2}$. Declare that the potential energy is 0 at the ground. The force of gravity is a constant $-mg$, so the work done as the ball rises to height x is $-mgx$, so potential energy has increased to mgx. Total energy:

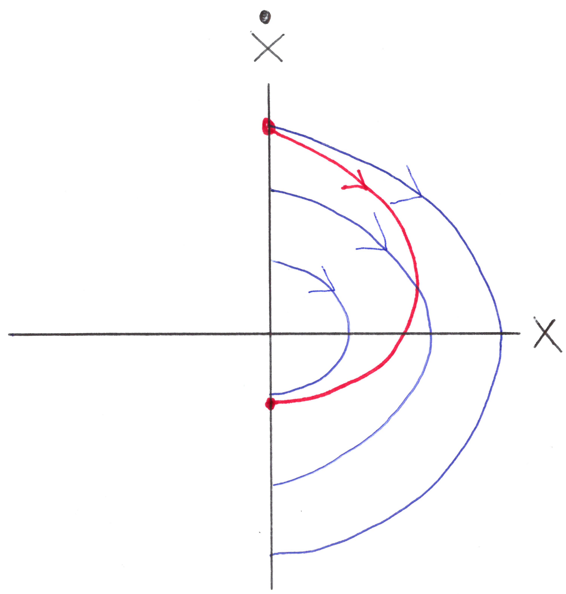
$$E = \frac{m\dot{x}^2}{2} + mgx.$$

For various positive constants E, these are the equations of the trajectories. They are parts of parabolas.

In which direction are the trajectories traversed? Downward: above the horizontal axis, $\dot{x} > 0$, which means that x is increasing.



What if there is a little bit of air resistance? Then the phase plane trajectory crosses through the equal-energy parabolas, and the ball lands with a lower speed than it started with (and its velocity is of opposite sign, of course).



Question 16.4. Not actually done in lecture. Can two different trajectories in the (x, y)-phase plane for a system $\dot{\mathbf{x}} = A\mathbf{x}$ ever intersect?

Answer: No. If a trajectory passes through a point $\mathbf{v}$, then its behavior before and after are uniquely determined, by the existence and uniqueness theorem. (They can approach the same point as $t \to \infty$, however.)

March 19

Midterm 2

March 21

16.5. **Gaussian elimination.** Gaussian elimination is an algorithm for converting any matrix into row-echelon form by performing row operations. Here are the steps:

1. Find the leftmost nonzero column, and the first nonzero entry in that column (read from the top down).
2. If that entry is not already in the first row, interchange its row with the first row.
3. Make all other entries of the column zero by adding suitable multiples of the first row to the others.

4. At this point, the first row is done, so ignore it, and repeat the steps above for the remaining submatrix (with one fewer row). In each iteration, ignore the rows already taken care of.
5. Stop when all the remaining rows consist entirely of zeros. Then the whole matrix will be in what is called *row-echelon form*.

Problem 16.5. Done in recitation. Apply Gaussian elimination to convert the 4×7 matrix

$$\begin{pmatrix} 0 & 0 & 6 & 2 & -4 & -8 & 8 \\ 0 & 0 & 3 & 1 & -2 & -4 & 4 \\ 2 & -3 & 1 & 4 & -7 & 1 & 2 \\ 6 & -9 & 0 & 11 & -19 & 3 & 0 \end{pmatrix}$$

to row-echelon form. (This example is taken from Hill, *Elementary linear algebra with applications*, p. 17.)

Solution:

Step 1. The leftmost nonzero column is the first one, and its first nonzero entry is the 2:

$$\begin{pmatrix} 0 & 0 & 6 & 2 & -4 & -8 & 8 \\ 0 & 0 & 3 & 1 & -2 & -4 & 4 \\ 2 & -3 & 1 & 4 & -7 & 1 & 2 \\ 6 & -9 & 0 & 11 & -19 & 3 & 0 \end{pmatrix}.$$

Step 2. The 2 is not in the first row, so interchange its row with the first row:

$$\begin{pmatrix} 2 & -3 & 1 & 4 & -7 & 1 & 2 \\ 0 & 0 & 3 & 1 & -2 & -4 & 4 \\ 0 & 0 & 6 & 2 & -4 & -8 & 8 \\ 6 & -9 & 0 & 11 & -19 & 3 & 0 \end{pmatrix}.$$

Step 3. To make all other entries of the column zero, we need to add -3 times the first row to the last row (the other rows are OK already):

$$\begin{pmatrix} 2 & -3 & 1 & 4 & -7 & 1 & 2 \\ 0 & 0 & 3 & 1 & -2 & -4 & 4 \\ 0 & 0 & 6 & 2 & -4 & -8 & 8 \\ 0 & 0 & -3 & -1 & 2 & 0 & -6 \end{pmatrix}.$$

Step 4. Now the first row is done. Start over with the 3×7 submatrix that remains beneath it:

$$\begin{pmatrix} 2 & -3 & 1 & 4 & -7 & 1 & 2 \\ 0 & 0 & 3 & 1 & -2 & -4 & 4 \\ 0 & 0 & 6 & 2 & -4 & -8 & 8 \\ 0 & 0 & -3 & -1 & 2 & 0 & -6 \end{pmatrix}.$$

Step 1. The leftmost nonzero column is now the third column, and its first nonzero entry is the 3:

$$\begin{pmatrix} 2 & -3 & 1 & 4 & -7 & 1 & 2 \\ 0 & 0 & 3 & 1 & -2 & -4 & 4 \\ 0 & 0 & 6 & 2 & -4 & -8 & 8 \\ 0 & 0 & -3 & -1 & 2 & 0 & -6 \end{pmatrix}.$$

Step 2. The 3 is already in the first row of the submatrix (we are ignoring the first row of the whole matrix), so no interchange is necessary.

Step 3. To make all other entries of the column zero, add -2 times the (new) first row to the (new) second row, and 1 times the (new) first row to the (new) third row:

$$\begin{pmatrix} 2 & -3 & 1 & 4 & -7 & 1 & 2 \\ 0 & 0 & 3 & 1 & -2 & -4 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -4 & -2 \end{pmatrix}.$$

Step 4. Now the first and second row of the original matrix are done. Start over with the 2×7 submatrix beneath them:

$$\begin{pmatrix} 2 & -3 & 1 & 4 & -7 & 1 & 2 \\ 0 & 0 & 3 & 1 & -2 & -4 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -4 & -2 \end{pmatrix}.$$

Step 1. The leftmost nonzero column is now the penultimate column, and its first nonzero entry is the -4 at the bottom:

$$\begin{pmatrix} 2 & -3 & 1 & 4 & -7 & 1 & 2 \\ 0 & 0 & 3 & 1 & -2 & -4 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -4 & -2 \end{pmatrix}.$$

Step 2. The -4 is not in the first row of the submatrix, so interchange its row with the first row of the submatrix:

$$\begin{pmatrix} 2 & -3 & 1 & 4 & -7 & 1 & 2 \\ 0 & 0 & 3 & 1 & -2 & -4 & 4 \\ 0 & 0 & 0 & 0 & 0 & -4 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Step 3. The other entry in this column of the submatrix is already 0, so this step is not necessary.

Now the first three rows are done. What remains below them is all zeros, so stop! The matrix is now in row-echelon form:

$$\begin{pmatrix} 2 & -3 & 1 & 4 & -7 & 1 & 2 \\ 0 & 0 & 3 & 1 & -2 & -4 & 4 \\ 0 & 0 & 0 & 0 & 0 & -4 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \quad \square$$

16.6. **Row-echelon form.** What does row-echelon form mean? Before explaining this, we need a few preliminary definitions. A zero row of a matrix is a row consisting entirely of zeros. A nonzero row of a matrix is a row with at least one nonzero entry. In each nonzero row, the first nonzero entry is called the pivot.

Example 16.6. The following 4×5 matrix has one zero row, and three pivots (shown in red):

$$\begin{pmatrix} 0 & -5 & 4 & 4 & 3 \\ 2 & 0 & 0 & 1 & 7 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 \end{pmatrix}.$$

Definition 16.7. A matrix is in row-echelon form if it satisfies both of the following conditions:

1. All the zero rows (if any) are grouped at the bottom of the matrix.
2. Each pivot lies farther to the right than the pivots of higher rows.

Warning: Some books require also that each pivot be a 1. We are not going to require this for row-echelon form, but we will require it for *reduced* row-echelon form later on.

16.7. **Back-substitution.**

Key point of row-echelon form: Matrices in row-echelon form correspond to systems that are ready to be solved immediately by back-substitution: solve for each variable in reverse order, while introducing a parameter for each variable not directly expressed in terms of later variables, and substitute values into earlier equations once they are known.

Problem 16.8. Suppose that we are solving a linear system with unknowns x, y, z, v, w. Suppose that we already wrote down the augmented matrix and used Gaussian elimination to convert it to row-echelon form, resulting in

$$\left(\begin{array}{ccccc|c} 1 & 2 & 0 & 2 & 3 & 4 \\ 0 & -1 & 2 & 3 & 1 & 5 \\ 0 & 0 & 0 & 0 & 2 & 6 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

Find the general solution to the system.

Solution: The new augmented matrix (the one shown, in row-echelon form) represents the inhomogeneous linear system

$$\begin{aligned} x + 2y \qquad + 2v + 3w &= 4 \\ -y + 2z + 3v + \ \ w &= 5 \\ 2w &= 6 \\ 0 &= 0. \end{aligned}$$

We solve for the variables in reverse order, using the equations from the bottom up. Start by solving for the last variable, w:

$$w = 3.$$

There is no equation for v in terms of the later variable w, so v can be any number; set

$$v = c_1 \quad \text{for a parameter } c_1.$$

There is no equation for z in terms of the later variables v and w, so set

$$z = c_2 \quad \text{for a parameter } c_2.$$

Substitute the values of w, v, z into the previous equation, and solve for y:

$$\begin{aligned} -y + 2c_2 + 3c_1 + 3 &= 5 \\ y &= 3c_1 + 2c_2 - 2. \end{aligned}$$

Similarly, solve for x:

$$\begin{aligned} x + 2(3c_1 + 2c_2 - 2) + 2c_1 + 3(3) &= 4 \\ x &= -8c_1 - 4c_2 - 1. \end{aligned}$$

Conclusion: The general solution is

$$\begin{aligned}
\begin{pmatrix} x \\ y \\ z \\ v \\ w \end{pmatrix} &= \begin{pmatrix} -8c_1 - 4c_2 - 1 \\ 3c_1 + 2c_2 - 2 \\ c_2 \\ c_1 \\ 3 \end{pmatrix} \\
&= \begin{pmatrix} -1 \\ -2 \\ 0 \\ 0 \\ 3 \end{pmatrix} + \begin{pmatrix} -8c_1 \\ 3c_1 \\ 0 \\ c_1 \\ 0 \end{pmatrix} + \begin{pmatrix} -4c_2 \\ 2c_2 \\ c_2 \\ 0 \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} -1 \\ -2 \\ 0 \\ 0 \\ 3 \end{pmatrix} + c_1 \begin{pmatrix} -8 \\ 3 \\ 0 \\ 1 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} -4 \\ 2 \\ 1 \\ 0 \\ 0 \end{pmatrix},
\end{aligned}$$

where c_1, c_2 are parameters. □

Suppose that a matrix is in row-echelon form. Then any column that contains a pivot is called a pivot column. A variable whose corresponding column is a pivot column is called a dependent variable or pivot variable. The other variables are called free variables. (The augmented column does not correspond to any variable.)

In the problem above, x, y, w were dependent variables, and v, z were free variables.

Warning: If your matrix is not in row-echelon form yet, don't talk about pivot columns and pivot variables!

16.8. **Reduced row-echelon form.** With even more row operations, one can simplify a matrix in row-echelon form to an even more special form:

Definition 16.9. A matrix is in reduced row-echelon form (RREF) if it satisfies all of the following conditions:

1. It is in row-echelon form.
2. Each pivot is a 1.
3. In each pivot column, all the entries are 0 except for the pivot itself.

16.9. **Gauss–Jordan elimination.** The presentation of the algorithm and the first problem below was done in recitation.

Gauss–Jordan elimination is an algorithm for converting any matrix into *reduced* row-echelon form by performing row operations. Here are the steps:

1. Use Gaussian elimination to convert the matrix to row echelon form.
2. Divide the last nonzero row by its pivot, to make the pivot 1.
3. Make all entries in that pivot's column 0 by adding suitable multiples of the pivot's row to the rows above.
4. At this point, the row in question (and all rows below it) are done. Ignore them, and go back to Step 2, but now with the remaining submatrix, above the row just completed.

Eventually the whole matrix will be in reduced row-echelon form.

Problem 16.10. Convert the 4×7 matrix

$$\begin{pmatrix} 2 & -3 & 1 & 4 & -7 & 1 & 2 \\ 0 & 0 & 3 & 1 & -2 & -4 & 4 \\ 0 & 0 & 0 & 0 & 0 & -4 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

to reduced row-echelon form.

Solution:

Step 1. The matrix is already in row-echelon form.

Step 2. The last nonzero row is the third row, and its pivot is the -4, so divide the third row by -4:

$$\begin{pmatrix} 2 & -3 & 1 & 4 & -7 & 1 & 2 \\ 0 & 0 & 3 & 1 & -2 & -4 & 4 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Step 3. To make all other entries of that pivot's column 0, add -1 times the third row to the first row, and add 4 times the third row to the second row:

$$\begin{pmatrix} 2 & -3 & 1 & 4 & -7 & 0 & 3/2 \\ 0 & 0 & 3 & 1 & -2 & 0 & 6 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Step 4. Now the last two rows are done:

$$\begin{pmatrix} 2 & -3 & 1 & 4 & -7 & 0 & 3/2 \\ 0 & 0 & 3 & 1 & -2 & 0 & 6 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Go back to Step 2, but with the 2×7 submatrix above them.

Step 2. The last nonzero row of the new matrix (ignoring the bottom two rows of the original matrix) is the second row, and its pivot is the 3, so we divide the second row by 3:

$$\begin{pmatrix} 2 & -3 & 1 & 4 & -7 & 0 & 3/2 \\ 0 & 0 & 1 & 1/3 & -2/3 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Step 3. To make the other entries of that pivot's column 0, add -1 times the second row to the first row:

$$\begin{pmatrix} 2 & -3 & 0 & 11/3 & -19/3 & 0 & -1/2 \\ 0 & 0 & 1 & 1/3 & -2/3 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Step 4. Now the last three rows are done:

$$\begin{pmatrix} 2 & -3 & 0 & 11/3 & -19/3 & 0 & -1/2 \\ 0 & 0 & 1 & 1/3 & -2/3 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Go back to Step 2, but with the 1×7 submatrix above them.

Step 2. The last nonzero row of the new matrix is the only remaining row (the first row), and its pivot is the initial 2, so we divide the first row by 2:

$$\begin{pmatrix} 1 & -3/2 & 0 & 11/6 & -19/6 & 0 & -1/4 \\ 0 & 0 & 1 & 1/3 & -2/3 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The matrix is now in reduced row-echelon form. $\square$

Problem 16.11. Suppose that we are solving a linear system for unknowns x, y, z, v, w. Suppose that we have used Gauss–Jordan elimination to put the augmented matrix in reduced row-echelon form, and the result is

$$\left(\begin{array}{ccccc|c} 1 & 0 & -2 & 0 & 7 & 3 \\ 0 & 1 & 6 & 0 & 8 & 4 \\ 0 & 0 & 0 & 1 & 9 & 5 \end{array}\right).$$

(Check: This really is in reduced row-echelon form!) Find the general solution to the system.

Solution: The system to be solved is

$$\begin{aligned} x \qquad - 2z \quad + 7w &= 3 \\ y + 6z \quad + 8w &= 4 \\ v + 9w &= 5. \end{aligned}$$

Back-substitution:

$$\begin{aligned} w &= c_1 \quad \text{(free variable)} \\ v &= -9w + 5 = -9c_1 + 5 \\ z &= c_2 \quad \text{(free variable)} \\ y &= -6z - 8w + 4 = -6c_2 - 8c_1 + 4 \\ x &= 2z - 7w + 3 = 2c_2 - 7c_1 + 3. \end{aligned}$$

(Notice that no substitution was required: we could solve for each variable directly!) Answer:

$$\begin{pmatrix} x \\ y \\ z \\ v \\ w \end{pmatrix} = \begin{pmatrix} 2c_2 - 7c_1 + 3 \\ -6c_2 - 8c_1 + 4 \\ c_2 \\ -9c_1 + 5 \\ c_1 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \\ 0 \\ 5 \\ 0 \end{pmatrix} + c_1 \begin{pmatrix} -7 \\ -8 \\ 0 \\ -9 \\ 1 \end{pmatrix} + c_2 \begin{pmatrix} 2 \\ -6 \\ 1 \\ 0 \\ 0 \end{pmatrix}. \quad \square$$

Moral: After using Gaussian elimination to put an augmented matrix into row-echelon form, there are two ways to finish solving the linear system:

- Do back-substitution.
- Do the extra row operations need to get the matrix into *reduced* row-echelon form (Gauss–Jordan elimination), and then do (a much easier) back-substitution.

You can experiment to find out which is faster for you.

Remark 16.12. Performing row operations on A in a different order than specified by Gaussian elimination and Gauss–Jordan elimination can lead to different row-echelon forms. But it turns out that row operations leading to *reduced row-echelon form* always give the same result, a matrix that we will write as $\mathrm{RREF}(A)$.

16.10. **Comparing inhomogeneous and homogeneous linear systems.** Recall that the general solution to the inhomogeneous system

$$\begin{aligned} x + 2y \qquad + 2v + 3w &= 4 \\ -y + 2z + 3v + \ \ w &= 5 \\ 2w &= 6 \\ 0 &= 0 \end{aligned}$$

was

$$\begin{pmatrix} x \\ y \\ z \\ v \\ w \end{pmatrix} = \underbrace{\begin{pmatrix} -1 \\ -2 \\ 0 \\ 0 \\ 3 \end{pmatrix}}_{\text{particular solution}} + \underbrace{c_1 \begin{pmatrix} -8 \\ 3 \\ 0 \\ 1 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} -4 \\ 2 \\ 1 \\ 0 \\ 0 \end{pmatrix}}_{\text{general homogeneous solution}},$$

where c_1, c_2 are parameters. (The labels under the braces haven't been explained yet.)

Doing Gaussian elimination and back-substitution again would show that the general solution to the associated *homogeneous* system

$$\begin{aligned} x + 2y \qquad + 2v + 3w &= 0 \\ -y + 2z + 3v + \ \ w &= 0 \\ 2w &= 0 \\ 0 &= 0 \end{aligned}$$

is

$$\begin{pmatrix} x \\ y \\ z \\ v \\ w \end{pmatrix} = c_1 \begin{pmatrix} -8 \\ 3 \\ 0 \\ 1 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} -4 \\ 2 \\ 1 \\ 0 \\ 0 \end{pmatrix},$$

where c_1, c_2 are parameters. We now want to say that the set of solutions is

$$\boxed{\text{the set of all linear combinations of the two vectors } \begin{pmatrix} -8 \\ 3 \\ 0 \\ 1 \\ 0 \end{pmatrix} \text{ and } \begin{pmatrix} -4 \\ 2 \\ 1 \\ 0 \\ 0 \end{pmatrix}},$$

so it is the span of those two vectors, so it is a vector space, except that so far we've talked about these linear algebra concepts only for functions, not for vectors. It's time to introduce these concepts for vectors too.

For a while, we are going to assume that vectors have real numbers as coordinates, and that all scalars are real numbers. This is so we can describe things geometrically in $\mathbb{R}^n$ more easily. But eventually, we will work with vectors in $\mathbb{C}^n$ whose coordinates can be complex numbers, and will allow scalar multiplication by complex numbers.

17.1. **Vector space.**

Definition 17.1. Suppose that S is a set consisting of some of the vectors in $\mathbb{R}^n$ (for some fixed value of n). Call S a vector space if all of the following are true:

0. The zero vector $\mathbf{0}$ is in S.
1. Multiplying any one vector in S by any scalar gives another vector in S.
2. Adding any two vectors in S gives another vector in S.

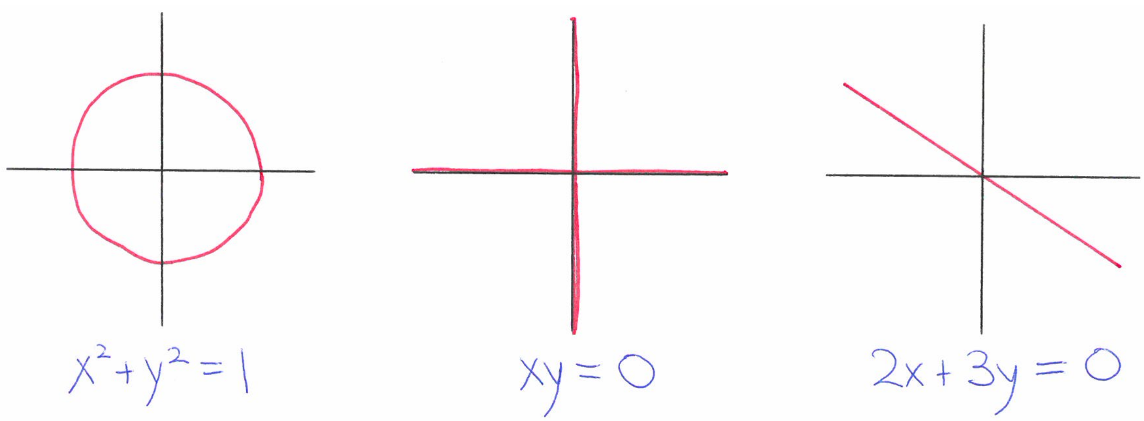
Remark 17.2. Such a set S is also called a subspace of $\mathbb{R}^n$, because $\mathbb{R}^n$ itself is a vector space, and S is a vector space contained in it.

Flashcard question: Which of the following subsets of $\mathbb{R}^2$ are vector spaces?

(a) The set of all vectors $\begin{pmatrix} x \\ y \end{pmatrix}$ satisfying $x^2 + y^2 = 1$.

(b) The set of all vectors $\begin{pmatrix} x \\ y \end{pmatrix}$ satisfying $xy = 0$.

(c) The set of all vectors $\begin{pmatrix} x \\ y \end{pmatrix}$ satisfying $2x + 3y = 0$.



Answer: Only (c) is a vector space.

Explanation: Let S be the set. For S to be a vector space, it must satisfy all three conditions.

Example (a) doesn't even satisfy condition 0, because the zero vector $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ is not in S.

Example (b) satisfies condition 0: the zero vector is in S. It satisfies condition 1 too: If $\begin{pmatrix} x \\ y \end{pmatrix}$ is one vector in S (so $xy = 0$) and c is any scalar, then the vector $c \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} cx \\ cy \end{pmatrix}$ satisfies $(cx)(cy) = c^2xy = c^2(0) = 0$. But it does not satisfy condition 2 for *every* pair of vectors in S: for example, $\begin{pmatrix} 2 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 3 \end{pmatrix}$ are in S, but their sum $\begin{pmatrix} 2 \\ 3 \end{pmatrix}$ is not in S.

Example (c) is a vector space, as we will now check. Condition 0: The zero vector is in S. Condition 1: If $\begin{pmatrix} x \\ y \end{pmatrix}$ is any element of S (so $2x+3y=0$) and c is any scalar, then multiplying the equation by c gives $2(cx)+3(cy)=0$, which shows that the vector $c \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} cx \\ cy \end{pmatrix}$ is in S. Condition 2: If $\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$ and $\begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$ are in S (so $2x_1+3y_1=0$ and $2x_2+3y_2=0$), then adding the equations shows that $2(x_1+x_2)+3(y_1+y_2)=0$, which says that the vector $\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_1+x_2 \\ y_1+y_2 \end{pmatrix}$ is in S. Thus S is a vector space. $\square$

Subspaces of $\mathbb{R}^2$ (it turns out that this is the complete list):

- $\{\mathbf{0}\}$ (the set containing only the origin)
- a line through the origin
- the whole plane $\mathbb{R}^2$.

Subspaces of $\mathbb{R}^3$ (again, the complete list):

- $\{\mathbf{0}\}$
- a line through the origin
- a plane through the origin
- the whole space $\mathbb{R}^3$.

March 23

17.2. Linear combinations.

Definition 17.3. A linear combination of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is a vector of the form $c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$ for some scalars $c_1, \ldots, c_n$.

17.3. Span.

Definition 17.4. The span of $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is the set of *all* linear combinations of $\mathbf{v}_1, \ldots, \mathbf{v}_n$:

$$\operatorname{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n) := \{\text{all vectors } c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n, \text{ where } c_1, \ldots, c_n \text{ are scalars}\}.$$

Example 17.5. If $\mathbf{v} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ in $\mathbb{R}^2$, then $\operatorname{Span}(\mathbf{v})$ is the set of all vectors $c\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ as c ranges over all real numbers, so $\operatorname{Span}(\mathbf{v})$ is the line $y = x$.

Example 17.6. Similarly, $\operatorname{Span}\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 2 \end{pmatrix}\right)$ is the set of vectors $c_1\begin{pmatrix} 1 \\ 1 \end{pmatrix} + c_2\begin{pmatrix} 2 \\ 2 \end{pmatrix}$ as c_1 and c_2 range over all real numbers, but this is still only the line $y = x$.

Example 17.7. Let $\mathbf{v} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ and $\mathbf{w} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ in $\mathbb{R}^2$. Is $\begin{pmatrix} 8 \\ 7 \end{pmatrix}$ in $\operatorname{Span}(\mathbf{v}, \mathbf{w})$? Yes, it's $3\mathbf{v} + 2\mathbf{w}$. In fact, every vector of $\mathbb{R}^2$ is in the span: $\operatorname{Span}(\mathbf{v}, \mathbf{w}) = \mathbb{R}^2$.

Example 17.8. If $\mathbf{i} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ and $\mathbf{j} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, then $\operatorname{Span}(\mathbf{i}, \mathbf{j})$ is the set of all vectors of the form

$$c_1\mathbf{i} + c_2\mathbf{j} = \begin{pmatrix} c_1 \\ c_2 \\ 0 \end{pmatrix}.$$

These form the xy-plane in $\mathbb{R}^3$, whose equation is $z = 0$.

Problem 17.9. Explain the following statement:

> If $\mathbf{v}$ and $\mathbf{w}$ are two vectors, then $\operatorname{Span}(\mathbf{v}, \mathbf{w})$ is a vector space.

Solution:

0. The zero vector $\mathbf{0}$ is in $\operatorname{Span}(\mathbf{v}, \mathbf{w})$ since $\mathbf{0} = 0\mathbf{v} + 0\mathbf{w}$.
1. Multiplying any linear combination of $\mathbf{v}$ and $\mathbf{w}$ by any scalar gives another linear combination of $\mathbf{v}$ and $\mathbf{w}$ (for example, $5(2\mathbf{v} + 3\mathbf{w}) = 10\mathbf{v} + 15\mathbf{w}$).
2. Adding any two linear combinations of $\mathbf{v}$ and $\mathbf{w}$ gives another linear combination of $\mathbf{v}$ and $\mathbf{w}$ (for example, $(2\mathbf{v} + 3\mathbf{w}) + (4\mathbf{v} + 5\mathbf{w}) = 6\mathbf{v} + 8\mathbf{w}$). □

The same argument shows that *any* span is a vector space.

17.4. **Nullspace.**

Example 17.10 (Homogeneous linear system). For

$$A = \begin{pmatrix} 1 & 2 & 0 & 2 & 3 \\ 0 & -1 & 2 & 3 & 1 \\ 0 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

we showed that

$$\{\text{all solutions to } A\mathbf{x} = \mathbf{0}\} \quad = \quad \text{Span}\left(\begin{pmatrix} -8 \\ 3 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -4 \\ 2 \\ 1 \\ 0 \\ 0 \end{pmatrix} \right).$$

This is a vector space (since any span is a vector space)!

The same happens for any matrix A:

Theorem 17.11. *For any homogeneous linear system $A\mathbf{x} = \mathbf{0}$, the set of all solutions is a vector space.*

This is analogous to the fact that the set of all solutions to a homogeneous linear ODE is a vector space!

Definition 17.12. The set of all solutions to $A\mathbf{x} = \mathbf{0}$ is called the nullspace of the matrix A, and denoted $\text{NS}(A)$.

Problem 17.13. Let $A = \begin{pmatrix} 4 & 6 \\ 2 & 3 \end{pmatrix}$. Is $\begin{pmatrix} -3 \\ 2 \end{pmatrix}$ in $\text{NS}(A)$?

Solution: The question is asking whether $\begin{pmatrix} -3 \\ 2 \end{pmatrix}$ is a solution to $A\mathbf{x} = \mathbf{0}$. Is it true that

$$\begin{pmatrix} 4 & 6 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} -3 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}?$$

Yes! $\square$

Here is a more direct way to explain why the set of all solutions to $A\mathbf{x} = \mathbf{0}$ is a vector space, without computing it as a span:

0. The zero vector $\mathbf{0}$ is a solution since $A\mathbf{0} = \mathbf{0}$.
1. Multiplying any solution $\mathbf{v}$ by any scalar c gives another solution: given that $A\mathbf{v} = \mathbf{0}$, it follows that $A(c\mathbf{v}) = c(A\mathbf{v}) = c\mathbf{0} = \mathbf{0}$.
2. Adding any solutions gives another solution: given that $A\mathbf{v} = \mathbf{0}$ and $A\mathbf{w} = \mathbf{0}$, it follows that $A(\mathbf{v} + \mathbf{w}) = A\mathbf{v} + A\mathbf{w} = \mathbf{0} + \mathbf{0} = \mathbf{0}$.

To *compute* $\text{NS}(A)$, solve the system $A\mathbf{x} = \mathbf{0}$ by using Gaussian elimination and back-substitution. (Shortcut: For a homogeneous system, there is no need to keep track of an augmented column, because it would consist of zeros, and would stay that way even after row operations.)

17.5. **Linearly dependent vectors.**

Definition 17.14. Vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly dependent if one of them is a linear combination of the others.

Definition 17.15 (equivalent)**.** Vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly dependent if there exist scalars $c_1, \ldots, c_n$ *not all zero* such that $c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n = 0$.

We haven't yet talked about dimension for a vector space of vectors, but intuitively, the dimension of $\operatorname{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ will be n if the vectors are linearly independent, and less than n if they are linearly dependent.

Algorithm to test whether given vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly dependent:

1. Create a matrix A whose columns are $\mathbf{v}_1, \ldots, \mathbf{v}_n$.
2. Compute $\mathrm{NS}(A)$.
3. • If $\mathrm{NS}(A)$ contains any nonzero vector, then $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly dependent. In fact, any nonzero vector $\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$ in $\mathrm{NS}(A)$ gives a linear dependence
$$a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n = \mathbf{0}.$$
   • If $\mathrm{NS}(A) = \{\mathbf{0}\}$, then $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent.

We'll soon explain why this works. But first, let's give an example:

Problem 17.16. Determine whether the vectors $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}, \begin{pmatrix} 7 \\ 8 \\ 9 \end{pmatrix}$ are linearly dependent.

Solution:

1. Create $A = \begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{pmatrix}$.
2. We convert A to row-echelon form. First add -2 times the first row to the second row, and add -3 times the first row to the third row, to get
$$\begin{pmatrix} 1 & 4 & 7 \\ 0 & -3 & -6 \\ 0 & -6 & -12 \end{pmatrix}.$$
Now the first row is done. Add -2 times the second column to the third column to get
$$\begin{pmatrix} 1 & 4 & 7 \\ 0 & -3 & -6 \\ 0 & 0 & 0 \end{pmatrix},$$

which is in row-echelon form. Solve the corresponding system

$$\begin{aligned} x + 4y + 7z &= 0 \\ -3y - 6z &= 0 \end{aligned}$$

by back-substitution: $z = c_1$, $y = -2c_1$, $x = -4(-2c_1) - 7c_1 = c_1$, so the general solution to $A\mathbf{x} = \mathbf{0}$ is $\begin{pmatrix} c_1 \\ -2c_1 \\ c_1 \end{pmatrix}$,

$$\mathrm{NS}(A) = \mathrm{Span}\left(\begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}\right).$$

3. Since $\mathrm{NS}(A)$ contains a nonzero vector, the three given vectors are linearly dependent. □

Actually, to answer the stated question, we could have stopped after finding the row-echelon form and seeing that there was a non-pivot column, since a non-pivot column means that there is a free variable, which means that there will be nonzero vectors in $\mathrm{NS}(A)$.

The advantage of doing the back-substitution to actually *find* a nonzero vector in $\mathrm{NS}(A)$ is that it tells us *which* combination of the three vectors is $\mathbf{0}$. In the example, $\begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$ was in $\mathrm{NS}(A)$, so

$$1\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + (-2)\begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} + 1\begin{pmatrix} 7 \\ 8 \\ 9 \end{pmatrix} = \mathbf{0}.$$

Also, we can then solve for one of the vectors as a linear combination of the others:

$$\begin{pmatrix} 7 \\ 8 \\ 9 \end{pmatrix} = (-1)\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + 2\begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}.$$

Question 17.17. Why does the algorithm work?

Answer: Checking for linear dependence is the same as searching for nonzero solutions to

$$a_1\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + a_2\begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} + a_3\begin{pmatrix} 7 \\ 8 \\ 9 \end{pmatrix} = \mathbf{0}.$$

By the interpretation of matrix-vector multiplication as a linear combination of columns, this equation is the same as

$$\begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{pmatrix}\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \mathbf{0},$$

so what we are really looking for is a nonzero vector $\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$ in the nullspace of A. $\square$

Example 17.18. For

$$A = \begin{pmatrix} 1 & 2 & 0 & 2 & 3 \\ 0 & -1 & 2 & 3 & 1 \\ 0 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

we found earlier that the general solution to $A\mathbf{x} = \mathbf{0}$ was

$$\mathbf{x} \quad = \quad c_1 \begin{pmatrix} -8 \\ 3 \\ 0 \\ 1 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} -4 \\ 2 \\ 1 \\ 0 \\ 0 \end{pmatrix} \quad = \quad \begin{pmatrix} -8c_1 - 4c_2 \\ 3c_1 + 2c_2 \\ c_2 \\ c_1 \\ 0 \end{pmatrix}.$$

Are there numbers c_1 and c_2 that make this combination $\mathbf{0}$? If that happens then the blue entries on the right are 0, so $c_1 = 0$ and $c_2 = 0$. Thus the two column vectors found that span $\mathrm{NS}(A)$ are linearly independent.

The same argument applies whenever we use Gaussian elimination and back-substitution to solve a homogeneous linear system: the list of solutions found is always linearly independent.

17.6. **Basis.**

Definition 17.19. A basis of a vector space S (of vectors) is a list of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots$ such that

1. $\mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2, \ldots) = S$, and
2. The vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots$ are linearly independent.

Example 17.20. Not actually mentioned in the March 23 lecture. If S is the xy-plane in $\mathbb{R}^3$, then $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ is a basis for S.

April 2

17.7. **Review: solving a homogeneous linear system.**

Problem 17.21. Find a basis of the vector space of solutions to the homogeneous linear system

$$\begin{aligned} 2x + y - 3z + 4w &= 0 \\ 4x + 2y - 2z + 3w &= 0 \\ 2x + y - 7z + 9w &= 0. \end{aligned}$$

Solution: The system is $A\mathbf{x} = \mathbf{0}$ for

$$A := \begin{pmatrix} 2 & 1 & -3 & 4 \\ 4 & 2 & -2 & 3 \\ 2 & 1 & -7 & 9 \end{pmatrix}.$$

First we convert the matrix to row-echelon form. Add -2 times the first row to the second row, and -1 times the first row to the third row to get

$$\begin{pmatrix} 2 & 1 & -3 & 4 \\ 0 & 0 & 4 & -5 \\ 0 & 0 & -4 & 5 \end{pmatrix}$$

and then add the second row to the third row to get

$$B := \begin{pmatrix} 2 & 1 & -3 & 4 \\ 0 & 0 & 4 & -5 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

which is in row-echelon form (the pivots are identified in red). Solve the corresponding system

$$\begin{aligned} 2x + y - 3z + 4w &= 0 \\ 4z - 5w &= 0 \\ 0 &= 0 \end{aligned}$$

by back-substitution:

$$\begin{aligned} w &= c_1 \\ z &= \frac{5}{4}c_1 \\ y &= c_2 \\ x &= (-y + 3z - 4w)/2 = -\frac{1}{8}c_1 - \frac{1}{2}c_2, \end{aligned}$$

General solution: $\begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} -\frac{1}{8}c_1 - \frac{1}{2}c_2 \\ c_2 \\ \frac{5}{4}c_1 \\ c_1 \end{pmatrix} = c_1 \begin{pmatrix} -1/8 \\ 0 \\ 5/4 \\ 1 \end{pmatrix} + c_2 \begin{pmatrix} -1/2 \\ 1 \\ 0 \\ 0 \end{pmatrix}$.

$\mathrm{NS}(A) = \mathrm{Span}\left(\begin{pmatrix} -1/8 \\ 0 \\ 5/4 \\ 1 \end{pmatrix}, \begin{pmatrix} -1/2 \\ 1 \\ 0 \\ 0 \end{pmatrix} \right)$.

Basis for $\mathrm{NS}(A)$: $\begin{pmatrix} -1/8 \\ 0 \\ 5/4 \\ 1 \end{pmatrix}, \begin{pmatrix} -1/2 \\ 1 \\ 0 \\ 0 \end{pmatrix}$, since these are automatically linearly independent.

(As a check, plug each vector back into the original system.) □

17.8. **Dimension.** It turns out that every basis for a vector space has the same number of vectors.

Definition 17.22. The dimension of a vector space is the number of vectors in any basis.

Example 17.23. The line $x + 3y = 0$ in $\mathbb{R}^2$ is a vector space L. The vector $\begin{pmatrix} -3 \\ 1 \end{pmatrix}$ by itself is a basis for L, so the dimension of L is 1. (Not a big surprise!)

17.9. **Dimension and basis of a nullspace.**

Theorem 17.24 (Formula for the dimension of the nullspace)**.** *Suppose that the result of putting a matrix A in row-echelon form is B. Then* $\mathrm{NS}(A) = \mathrm{NS}(B)$ *(since row reductions do not change the solutions), and*

$$\boxed{\dim \mathrm{NS}(A) = \#\text{non-pivot columns of } B}.$$

(The boxed formula holds since it is the same as $\dim \mathrm{NS}(B) = \#$free variables.)

In other words, here are the steps to find the dimension of the space of solutions to a homogeneous linear system $A\mathbf{x} = \mathbf{0}$:

1. Perform Gaussian elimination on A to convert it to a matrix B in row-echelon form.
2. Identify the pivots of B.
3. Count the number of *non-pivot* columns of B; that number is $\dim \mathrm{NS}(A)$.

Warnings:

- You must put the matrix in row-echelon form before counting non-pivot columns!
- If you have an augmented column (of zeros, since we are talking about a homogeneous system), then *do not include it* in the count of non-pivot columns. (The augmented

column does not correspond to a free variable, or any variable at all for that matter, so it should not be counted.)

And here are the steps to find a basis of the space of solutions to a homogeneous linear system $A\mathbf{x} = \mathbf{0}$:

1. Perform Gaussian elimination on A to convert it to a matrix B in row-echelon form.
2. Use back-substitution to find the general solution to $B\mathbf{x} = \mathbf{0}$.
3. The general solution will be expressed as the general linear combination of a list of vectors; that list is a basis of $\mathrm{NS}(A)$.

18. Inhomogeneous linear systems

A linear system is called consistent if it has at least one solution, and inconsistent if there are no solutions.

Problem 18.1. Find the general solution to the system with augmented matrix $\left(\begin{array}{ccc|c} 2 & 3 & 5 & 7 \\ 0 & 1 & 4 & 6 \\ 0 & 0 & 0 & 9 \end{array}\right)$.

Solution: The last equation says

$$0x + 0y + 0z = 9,$$

i.e., $0 = 9$, which cannot be satisfied. So there are no solutions! The system is inconsistent. □

18.1. **Algorithm to test if a linear system is consistent (has a solution).** Consider a linear system of m equations in n variables.

1. Construct the $m \times (n+1)$ augmented matrix.
2. Put it in row-echelon form. Call this row-echelon form B.
3. Look for a row that is all zero except for a nonzero entry in the augmented column.

4a. If B has such a row, that row corresponds to an equation

$$0x_1 + \cdots + 0x_n = \underset{\text{nonzero number}}{b},$$

so the linear system is inconsistent.

4b. Otherwise, we can solve the system by back-substitution, so the linear system is consistent. In a solution, the free variables may take any values, but in terms of these one can solve for the dependent variables in reverse order, so

$$\boxed{\#\text{parameters in general solution} = \underbrace{\#\text{non-pivot columns } \textit{excluding the augmented column}}_{\#\text{free variables}}}.$$

18.2. **Inhomogeneous linear systems: theory.** For an inhomogeneous linear system $A\mathbf{x} = \mathbf{b}$, there are two possibilities:

1. There are no solutions.
2. There exists a solution. In this case, if $\mathbf{x}_p$ is a particular solution to $A\mathbf{x} = \mathbf{b}$, and $\mathbf{x}_h$ is the *general* solution to the homogeneous system $A\mathbf{x} = \mathbf{0}$, then $\mathbf{x} := \mathbf{x}_p + \mathbf{x}_h$ is the general solution to $A\mathbf{x} = \mathbf{b}$.

Here is why: Suppose that a solution exists; let $\mathbf{x}_p$ be one, so $A\mathbf{x}_p = \mathbf{b}$. If $\mathbf{x}_h$ satisfies $A\mathbf{x}_h = 0$, adding the two equations gives $A(\mathbf{x}_p + \mathbf{x}_h) = \mathbf{b}$, so adding $\mathbf{x}_p$ to $\mathbf{x}_h$ produces a solution $\mathbf{x}$ to the inhomogeneous equation. Every solution $\mathbf{x}$ to $A\mathbf{x} = \mathbf{b}$ arises this way from some $\mathbf{x}_h$ (specifically, from $\mathbf{x}_h := \mathbf{x} - \mathbf{x}_p$, which satisfies $A\mathbf{x}_h = A\mathbf{x} - A\mathbf{x}_p = \mathbf{b} - \mathbf{b} = \mathbf{0}$).

Remark 18.2. To *solve* $A\mathbf{x} = \mathbf{b}$, however, don't use $\mathbf{x} = \mathbf{x}_p + \mathbf{x}_h$. Instead use Gaussian elimination and back-substitution. The above is just to describe the shape of the solution.

18.3. **Column space.**

Problem 18.3. For which vectors $\mathbf{b} \in \mathbb{R}^2$ does the inhomogeneous linear system

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \mathbf{b}$$

have a solution?

Answer: The left hand side can be rewritten as

$$\begin{pmatrix} x_1 + 2x_2 + 3x_3 \\ 2x_1 + 4x_2 + 6x_3 \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 2 \end{pmatrix} + x_2 \begin{pmatrix} 2 \\ 4 \end{pmatrix} + x_3 \begin{pmatrix} 3 \\ 6 \end{pmatrix}$$

Thus, saying that the system has a solution is the same as saying that

$$\mathbf{b} \text{ is a linear combination of } \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 4 \end{pmatrix}, \begin{pmatrix} 3 \\ 6 \end{pmatrix},$$

or equivalently, that

$$\mathbf{b} \text{ is in the span of } \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 4 \end{pmatrix}, \begin{pmatrix} 3 \\ 6 \end{pmatrix}. \quad \square$$

Definition 18.4. The column space of a matrix A is the span of its columns. The notation for it is $\mathrm{CS}(A)$. (It is also called the image of A, and written $\mathrm{im}(A)$; the reason will be clearer when we talk about the geometric interpretation.)

Since $\mathrm{CS}(A)$ is a span, it is a vector space.

Here is what happens in general for (possibly inhomogeneous) linear systems (the explanation is the same as in the example above):

Theorem 18.5. *The linear system $A\mathbf{x} = \mathbf{b}$ has a solution if and only if $\mathbf{b}$ is in* $\mathrm{CS}(A)$*.*

For the matrix $A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{pmatrix}$ in the problem above,

$$\mathrm{CS}(A) = \text{the span of } \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 4 \end{pmatrix}, \begin{pmatrix} 3 \\ 6 \end{pmatrix},$$

which is the line $y = 2x$ in $\mathbb{R}^2$, a 1-dimensional vector space.

Steps to compute a basis for $\mathrm{CS}(A)$:

1. Perform Gaussian elimination to convert A to a matrix B in row-echelon form.
2. Identify the pivot columns of B.
3. The corresponding columns of A are a basis for $\mathrm{CS}(A)$.

Here is a summary of why this works (not discussed in lecture). Let C be the reduced row-echelon form of A. If

$$\text{fifth column} = 3(\text{first column}) + 7(\text{second column})$$

is true for a matrix, it will remain true after any row operation. Similarly, any linear relation between columns is preserved by row operations. So the linear relations between columns of A are the same as the linear relations between columns of C. The condition that certain numbered columns (say the first, second, and fourth) of a matrix form a basis is expressible in terms of which linear relations hold, so if certain columns form a basis for $\mathrm{CS}(C)$, the same numbered columns will form a basis for $\mathrm{CS}(A)$. Also, performing Gauss–Jordan reduction on B to obtain C in reduced row-echelon form does not change the pivot locations. Thus it will be enough to show that the pivot columns of C form a basis of $\mathrm{CS}(C)$. Since C is in reduced row-echelon form, the pivot columns of C are the first r of the m standard basis vectors for $\mathbb{R}^m$, where r is the number of nonzero rows of C. These columns are linearly independent, and every other column is a linear combination of them, since the entries of C below the first r rows are all zeros. Thus the pivot columns of C form a basis of $\mathrm{CS}(C)$.

In particular,

$$\boxed{\dim \mathrm{CS}(A) = \#\text{pivot columns of } B}.$$

Warning: Usually $\mathrm{CS}(A) \neq \mathrm{CS}(B)$.

Problem 18.6. Let A be the 3×5 matrix $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ -1 & -2 & 9 & 10 & 11 \\ 1 & 2 & 9 & 11 & 13 \end{pmatrix}$.

(a) Find a basis for $\mathrm{CS}(A)$.

(b) What are $\dim \mathrm{NS}(A)$ and $\dim \mathrm{CS}(A)$?

Solution:

(a) First we must find a row-echelon form. Add the first row to the second, and add -1 times the first row to the third:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 0 & 0 & 12 & 14 & 16 \\ 0 & 0 & 6 & 7 & 8 \end{pmatrix}.$$

Add $-1/2$ times the second row to the third:

$$B := \begin{pmatrix} \color{red}{1} & 2 & 3 & 4 & 5 \\ 0 & 0 & \color{red}{12} & 14 & 16 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

This is in row-echelon form.

Basis for $\mathrm{CS}(B)$: first and third columns (the pivot columns) of B, i.e., $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 \\ 12 \\ 0 \end{pmatrix}$.

This is not what was asked for!

Basis for $\mathrm{CS}(A)$: first and third columns of A, i.e., $\boxed{\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 9 \\ 9 \end{pmatrix}}$.

(b)

$$\dim \mathrm{NS}(A) = \#\text{non-pivot columns of } B = 3$$
$$\dim \mathrm{CS}(A) = \#\text{pivot columns of } B = 2. \quad \square$$

18.4. **Rank.**

Definition 18.7. The rank of A is defined by

$$\mathrm{rank}(A) := \dim \mathrm{CS}(A).$$

Rank-nullity theorem. *For any $m \times n$ matrix A,*

$$\boxed{\dim \mathrm{NS}(A) + \mathrm{rank}(A) = n}.$$

Proof.

$$\begin{aligned} \dim \mathrm{NS}(A) + \mathrm{rank}(A) &= (\#\text{non-pivot columns of } B) + (\#\text{pivot columns of } B) \\ &= \#\text{columns of } B \\ &= n. \end{aligned}$$

$\square$

18.5. **Computing a basis for a span.**

Problem 18.8. Given vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{R}^m$, how can one compute a basis of $\operatorname{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$?

Solution:

1. Form the matrix A whose columns are $\mathbf{v}_1, \ldots, \mathbf{v}_n$.
2. Find a basis for $\mathrm{CS}(A)$ as above (columns of A corresponding to pivot columns of B).

18.6. **Example: a projection.** Let $\mathbf{f}$ be the function from $\mathbb{R}^3$ to $\mathbb{R}^3$ that projects all of $\mathbb{R}^3$ onto the xy-plane:

$$\mathbf{f}\begin{pmatrix} x \\ y \\ z \end{pmatrix} := \begin{pmatrix} x \\ y \\ 0 \end{pmatrix}.$$

Problem 18.9. What is the matrix A that represents f?

Solution: The matrix A is a 3×3 matrix such that

$$\begin{aligned}
(\text{first column of } A) = \mathbf{f}\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} &= \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \\
(\text{second column of } A) = \mathbf{f}\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} &= \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \\
(\text{third column of } A) = \mathbf{f}\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.
\end{aligned}$$

Thus $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$. $\square$

NS(A) is a subspace of the *input* space:

$$\begin{aligned}
\mathrm{NS}(A) &= \{\text{solutions to } A\mathbf{x} = \mathbf{0}\} \\
&= \left\{\text{solutions to } \mathbf{f}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \mathbf{0}\right\} \\
&= \left\{\begin{pmatrix} 0 \\ 0 \\ z \end{pmatrix} : z \in \mathbb{R}\right\} \\
&= \text{the } z\text{-axis in the input space } \mathbb{R}^3.
\end{aligned}$$

The image CS(A) is a subspace of the *output space*:

$$\begin{aligned}
\mathrm{CS}(A) &= \{\text{values of } A\mathbf{x}\} \\
&= \left\{\text{values of } \mathbf{f}\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right\} \\
&= \left\{\begin{pmatrix} x \\ y \\ 0 \end{pmatrix} : x, y \in \mathbb{R}\right\} \\
&= \text{the } xy\text{-plane in the output space } \mathbb{R}^3.
\end{aligned}$$

Here $\mathrm{rank}(A) = \dim \mathrm{CS}(A) = 2$.

The linear transformation $\mathbf{f}$ crushes NS(A) to the point $\mathbf{0}$ in the output space, and it flattens the whole input space $\mathbb{R}^3$ onto CS(A) in the output space. Of the 3 input dimensions, 1 is crushed, so $3 - 1 = 2$ dimensions are left.

(Mathematically, "there are m crushed dimensions" means just that NS(A) is m-dimensional.)

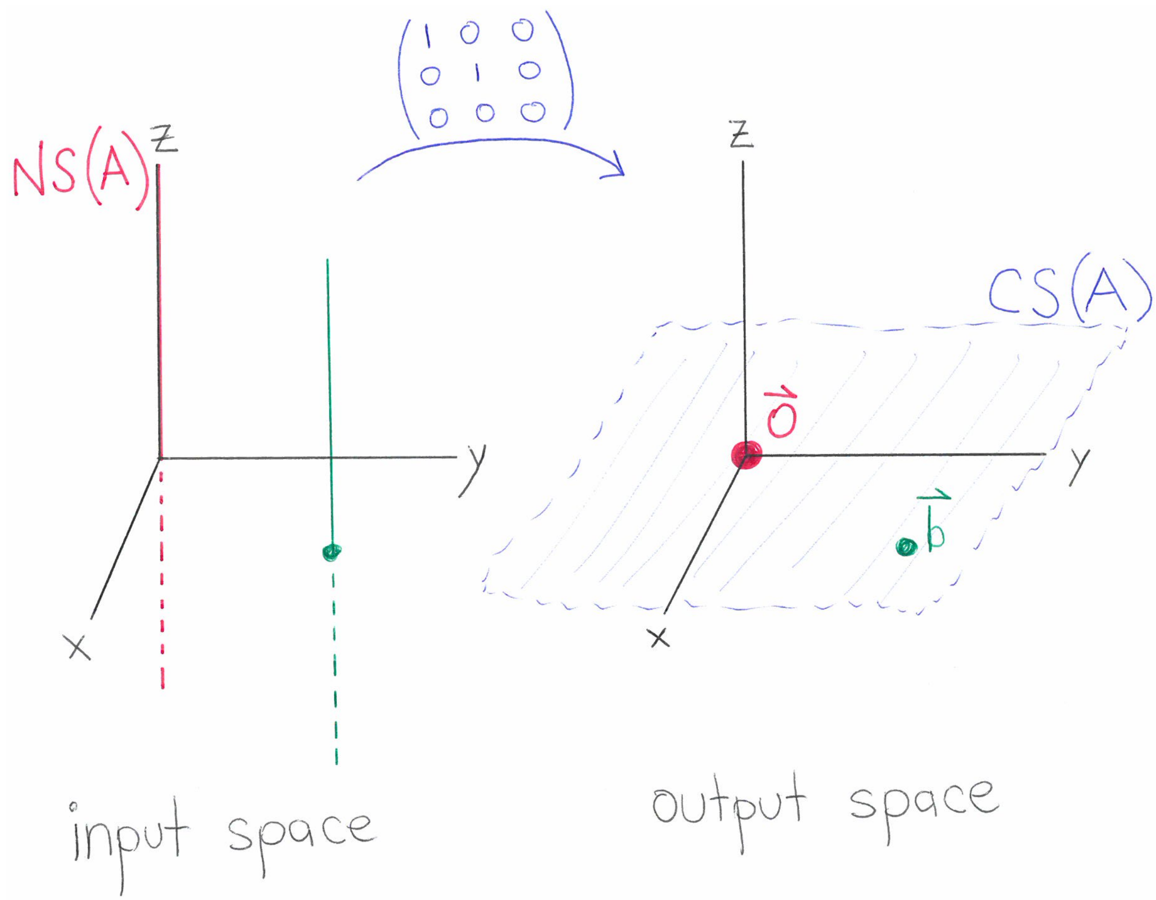
In general, for a linear transformation $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m$ represented by an $m \times n$ matrix A, of the n input dimensions, $\dim \mathrm{NS}(A)$ of them are crushed, leaving an image of dimension $n - \dim \mathrm{NS}(A)$. This explains geometrically why

$$\dim \mathrm{CS}(A) = n - \dim \mathrm{NS}(A),$$

which is the same as the rank-nullity theorem

$$\dim \mathrm{NS}(A) + \mathrm{rank}(A) = n$$

we stated earlier.



Back to the example: What does the solution set to $A\mathbf{x} = \mathbf{b}$ look like?

- If $\mathbf{b}$ is not in $\mathrm{CS}(A)$, then there are no solutions.
- If $\mathbf{b}$ is in $\mathrm{CS}(A)$, say $\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ 0 \end{pmatrix}$, then

$$
\begin{aligned}
\{\text{solutions to } A\mathbf{x} = \mathbf{b}\} &= \left\{ \text{solutions to } \mathbf{f}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ 0 \end{pmatrix} \right\} \\
&= \left\{ \begin{pmatrix} b_1 \\ b_2 \\ z \end{pmatrix} : z \in \mathbb{R} \right\} \\
&= \text{a vertical line parallel to the } z\text{-axis in the input space } \mathbb{R}^3.
\end{aligned}
$$

The set of solutions to the homogeneous system $A\mathbf{x} = \mathbf{0}$ is the line $\mathrm{NS}(A)$. To get from $\mathrm{NS}(A)$ to the set of solutions to $A\mathbf{x} = \mathbf{b}$, choose a particular solution vector to $A\mathbf{x} = \mathbf{b}$ and add it to every vector in $\mathrm{NS}(A)$.

Problem 18.10. Not done in lecture. What is the volume scaling factor?

Solution 1: It's the absolute value of the determinant:

$$\left|\det\begin{pmatrix}1&0&0\\0&1&0\\0&0&0\end{pmatrix}\right| = |0| = 0.$$

Solution 2: The linear transformation $\mathbf{f}$ takes any unit cube of volume 1 to a flat object of volume 0, so volume is getting multiplied by 0.

April 4

18.7. **Composition and matrix multiplication.**

Problem 18.11. Consider the following two functions from $\mathbb{R}^2$ to $\mathbb{R}^2$:

- $90°$ counterclockwise rotation about the origin, represented by $\begin{pmatrix}0&-1\\1&0\end{pmatrix}$.
- reflection across the x-axis, represented by $\begin{pmatrix}1&0\\0&-1\end{pmatrix}$.

We can compose these two functions, first rotating and then reflecting, to get a third function. What matrix represents this third function?

Solution: The composition maps

$$\begin{pmatrix}1\\0\end{pmatrix} \xrightarrow{\text{rotate}} \begin{pmatrix}0\\1\end{pmatrix} \xrightarrow{\text{reflect}} \begin{pmatrix}0\\-1\end{pmatrix}$$

and

$$\begin{pmatrix}0\\1\end{pmatrix} \xrightarrow{\text{rotate}} \begin{pmatrix}-1\\0\end{pmatrix} \xrightarrow{\text{reflect}} \begin{pmatrix}-1\\0\end{pmatrix};$$

the answer is the matrix having these outputs as the first and second columns:

$$\begin{pmatrix}0&-1\\-1&0\end{pmatrix}. \quad \square$$

Matrix multiplication is defined so that it corresponds to composition. In particular, we *define*

$$\underbrace{\begin{pmatrix}1&0\\0&-1\end{pmatrix}}_{\text{reflection}}\underbrace{\begin{pmatrix}0&-1\\1&0\end{pmatrix}}_{\text{rotation}} := \begin{pmatrix}0&-1\\-1&0\end{pmatrix}.$$

Question 18.12. Why did we write the rotation second?

Answer: Applying this matrix product to a vector $\mathbf{v}$ means that you apply rotation first (just as when computing $f(g(x))$, you apply g first).

In general, if A is an $m \times n$ matrix and B is an $n \times p$ matrix, then the composition

$$\mathbb{R}^p \xrightarrow{\text{left multiplication by } B} \mathbb{R}^n \xrightarrow{\text{left multiplication by } A} \mathbb{R}^m$$

is represented by an $m \times p$ matrix called AB. It turns out that the (i, j)-entry of AB is the dot product

$$(i^{\text{th}} \text{ row of } A) \cdot (j^{\text{th}} \text{ column of } B).$$

Remark 18.13. The product AB is defined only when

$$\#\text{columns of } A = \#\text{rows of } B.$$

(In our general example, this was $n = n$.)

19. Square matrices

19.1. **Determinants.** Done in recitation.

To each *square* matrix A is associated a number called the determinant:

$$\det \begin{pmatrix} a \end{pmatrix} := a$$

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} := ad - bc$$

$$\det \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix} := a_1b_2c_3 + a_2b_3c_1 + a_3b_1c_2 - c_1b_2a_3 - c_2b_3a_1 - c_3b_1a_2.$$

Alternative notation for determinant: $|A| = \begin{vmatrix} a & b \\ c & d \end{vmatrix}$. (This is a scalar, not a matrix!)

Geometric meaning: The *absolute value* of $\det A$ is the area scaling factor (or volume scaling factor or...).

Laplace expansion (along the first row) for a 3×3 determinant:

$$\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = +a_1 \begin{vmatrix} b_2 & b_3 \\ c_2 & c_3 \end{vmatrix} - a_2 \begin{vmatrix} b_1 & b_3 \\ c_1 & c_3 \end{vmatrix} + a_3 \begin{vmatrix} b_1 & b_2 \\ c_1 & c_2 \end{vmatrix}.$$

The general rule leading to the formula above is this:

(1) Move your finger along the entries in a row.

(2) At each position, compute the minor, defined as the smaller determinant obtained by crossing out the row and the column through your finger; then multiply the minor by the number you are pointing at, and adjust the sign according to the checkerboard pattern

$$\begin{matrix} + & - & + \\ - & + & - \\ + & - & + \end{matrix}$$

(the pattern always starts with $+$ in the upper left corner).

(3) Add up the results.

There is a similar expansion for a determinant of any size, computed along any row or column.

The diagonal of a matrix consists of the entries a_{ij} with $i = j$.

A diagonal matrix is a matrix that has zeros everywhere outside the diagonal:

$$\begin{pmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{pmatrix}$$

(it may have some zeros along the diagonal too).

An upper triangular matrix is a matrix whose entries strictly below the diagonal are all 0:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{pmatrix}$$

(the entries on or above the diagonal may or may not be 0).

Example 19.1. Any square matrix in row-echelon form is upper triangular.

Theorem 19.2. *The determinant of an upper triangular matrix equals the product of the diagonal entries.*

For example,

$$\det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{pmatrix} = a_{11}\, a_{22}\, a_{33}.$$

Why is the theorem true? The Laplace expansion along the first column shows that the determinant is a_{11} times a upper triangular minor with diagonal entries $a_{22}, \ldots, a_{nn}$.

Properties of determinants:

1. Interchanging two rows changes the sign of $\det A$.
2. Multiplying an entire row by a scalar c multiples $\det A$ by c.
3. Adding a multiple of a row to another row does not change $\det A$.

4. If one of the rows is all 0, then $\det A = 0$.
5. $\det(AB) = \det(A)\det(B)$ (assuming that A, B are square matrices of the same size).

In particular, row operations multiply $\det A$ by nonzero scalars, but do not change whether $\det A = 0$.

Question 19.3. Suppose that A is a 3×3 matrix such that $\det A = 5$. Doubling every entry of A gives a matrix $2A$. What is $\det(2A)$?

Solution: Each time we multiply a row by 2, the determinant gets multiplied by 2. We need to do this three times to double the whole matrix A, so the determinant gets multiplied by $2 \cdot 2 \cdot 2 = 8$. Thus $\det(2A) = 8\det(A) = 40$.

19.2. **Identity matrix.** Done in recitation.

The function $f\colon \mathbb{R}^3 \to \mathbb{R}^3$ that does nothing to its input,

$$\mathbf{f}\begin{pmatrix} x \\ y \\ z \end{pmatrix} := \begin{pmatrix} x \\ y \\ z \end{pmatrix},$$

is called the identity. The corresponding matrix, the 3×3 identity matrix I, has

$$\text{1st column} = \mathbf{f}\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \text{2nd column} = \mathbf{f}\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \text{3rd column} = \mathbf{f}\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

so

$$I := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

(The $n \times n$ identity matrix is similar, with 1s along the diagonal.)

It has the property that $AI = A$ whenever the matrix multiplication is defined, because doing nothing and then applying A is the same as applying A. Similarly, $IB = B$ whenever the matrix multiplication is defined.

Example 19.4. Suppose that A is a square matrix and $\det A \neq 0$. Then RREF(A) has nonzero determinant too, but is now upper triangular, so its diagonal entries are nonzero. In fact, these diagonal entries are 1 since they are pivots of a RREF matrix. Moreover, all non-diagonal entries are 0, by definition of RREF. So RREF(A) $= I$.

Now imagine solving $A\mathbf{x} = \mathbf{b}$. Gauss–Jordan elimination converts the augmented matrix $[A|\mathbf{b}]$ to $[I|\mathbf{c}]$ for some vector $\mathbf{c}$. Thus $A\mathbf{x} = \mathbf{b}$ has the same solutions as $I\mathbf{x} = \mathbf{c}$; the unique solution is $\mathbf{c}$.

What if instead we wanted to solve many equations with the same A, say, $A\mathbf{x}_1 = \mathbf{b}_1, \ldots, A\mathbf{x}_p = \mathbf{b}_p$? Use many augmented columns! Gauss–Jordan elimination converts $[A|\mathbf{b}_1 \ldots \mathbf{b}_p]$ to $[I|\mathbf{c}_1 \ldots \mathbf{c}_p]$, and $\mathbf{c}_1, \ldots, \mathbf{c}_p$ are the solutions.

In other words, to solve an equation $AX = B$ to find the unknown matrix X, convert $[A|B]$ to RREF $[I|C]$. Then C is the solution. □

The reason we talked about this is in order to compute inverse matrices; let's define these now.

19.3. **Inverse matrices.**

Definition 19.5. The inverse of an $n \times n$ matrix A is an $n \times n$ matrix A^{-1} such that

$$AA^{-1} = I \qquad \text{and} \qquad A^{-1}A = I.$$

It exists if and only if $\det A \neq 0$.

Suppose that A represents the linear transformation $\mathbf{f}$. Then A^{-1} exists if and only if an inverse function $\mathbf{f}^{-1}$ exists; in that case, A^{-1} represents $\mathbf{f}^{-1}$.

Problem 19.6. Does the rotation matrix $R := \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ have an inverse? If so, what is it?

Solution: The inverse linear transformation is rotation by $-\theta$, so

$$R^{-1} = \begin{pmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

(As a check, try multiplying R by this matrix, in either order.) □

Problem 19.7. Does the projection matrix $A := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ have an inverse? If so, what is it?

Solution: The associated linear transformation $\mathbf{f}$ is not a 1-to-1 correspondence, because it maps more than one vector to $\mathbf{0}$ (it maps the whole z-axis to $\mathbf{0}$). Thus $\mathbf{f}^{-1}$ does not exist, so A^{-1} does not exist. □

Suppose that $\det A \neq 0$. In 18.02, you learned one algorithm to compute A^{-1}, using the cofactor matrix. Now that we know how to compute RREF, we can give a faster algorithm (faster for big matrices, at least):

New algorithm to compute A^{-1}:

1. Form the $n \times 2n$ augmented matrix $[A|I]$.
2. Convert to RREF; the result will be $[I|B]$ for some $n \times n$ matrix B.
3. Then $A^{-1} = B$.

This is a special case of Example 19.4 since A^{-1} is the solution to $AX = I$.

19.4. **Conditions for invertibility.** There are two types of square matrices A:

- those with $\det A \neq 0$ (called nonsingular or invertible), and
- those with $\det A = 0$ (called singular).

The answer to the one question "Is $\det A = 0$?" determines a lot about the geometry of A and about solving systems $A\mathbf{x} = \mathbf{b}$, as we'll now explain.

19.4.1. *Nonsingular matrices.*

Theorem 19.8. *For a square matrix A, the following are equivalent:*

1. $\det A \neq 0$ *(scaling factor is positive)*
2. $\mathrm{NS}(A) = \{\mathbf{0}\}$ *(the only solution to $A\mathbf{x} = \mathbf{0}$ is $\mathbf{0}$)*
3. $\mathrm{rank}(A) = n$ *(image is n-dimensional)*
4. $\mathrm{CS}(A) = \mathbb{R}^n$ *(image is the whole space $\mathbb{R}^n$)*
5. *For each vector $\mathbf{b}$, the system $A\mathbf{x} = \mathbf{b}$ has exactly one solution.*
6. *A^{-1} exists.*
7. $\mathrm{RREF}(A) = I$.

So if you have a matrix A for which one of these conditions holds, then all of the conditions hold for A.

Let's explain the consequences of $\det A \neq 0$. Suppose that $\det A \neq 0$. Then the volume scaling factor is not 0, so the input space $\mathbb{R}^n$ is not flattened by A. This means that there are no "crushed dimensions", so $\mathrm{NS}(A) = \{\mathbf{0}\}$. Since no dimensions were crushed, the image $\mathrm{CS}(A)$ has the same dimension as the input space, namely n. By definition, $\mathrm{rank}(A) = \dim \mathrm{CS}(A) = n$. (Alternatively, this follows from $\dim \mathrm{NS}(A) + \mathrm{rank}(A) = n$.) The only n-dimensional subspace of $\mathbb{R}^n$ is $\mathbb{R}^n$ itself, so $\mathrm{CS}(A) = \mathbb{R}^n$. Thus every $\mathbf{b}$ is in $\mathrm{CS}(A)$, so $A\mathbf{x} = \mathbf{b}$ has a solution for every $\mathbf{b}$. The system $A\mathbf{x} = \mathbf{b}$ has the same number of solutions as $A\mathbf{x} = \mathbf{0}$ (they are just shifted by adding a particular solution $\mathbf{x}_p$); that number is 1 (the only solution to $A\mathbf{x} = \mathbf{0}$ is $\mathbf{0}$). To say that $A\mathbf{x} = \mathbf{b}$ has exactly one solution for each $\mathbf{b}$ means that the associated linear transformation $\mathbf{f}$ is a 1-to-1 correspondence, so $\mathbf{f}^{-1}$ exists, so A^{-1} exists. (Moreover, we showed how to find A^{-1} by applying Gauss–Jordan elimination to $[A|I]$.) We have $\mathrm{RREF}(A) = I$ as explained earlier, since I is the only RREF square matrix with nonzero determinant.

19.4.2. *Singular matrices.* The same theorem can be stated in terms of the opposite conditions (it's essentially the same theorem, so this is really just review):

Theorem 19.9. *For a square matrix A, the following are equivalent:*

1. $\det A = 0$ *(scaling factor is 0)*
2. $\mathrm{NS}(A)$ *is larger than* $\{\mathbf{0}\}$ *(i.e., $A\mathbf{x} = \mathbf{0}$ has a nonzero solution)*

3. $\operatorname{rank}(A) < n$ *(image has dimension less than n)*
4. $\mathrm{CS}(A)$ *is smaller than* $\mathbb{R}^n$ *(image is not the whole space* $\mathbb{R}^n$*)*
5. *The system* $A\mathbf{x} = \mathbf{b}$ *has no solutions for some vectors* $\mathbf{b}$*, and infinitely many solutions for other vectors* $\mathbf{b}$*.*
6. A^{-1} *does not exist.*
7. $\mathrm{RREF}(A) \neq I$*.*

Now let's explain the consequences of $\det A = 0$. Done in recitation.

Suppose that $\det A = 0$. Then the volume scaling factor is 0, so the input space is flattened by A. This means that some input dimensions are getting crushed, so $\mathrm{NS}(A)$ is larger than $\{\mathbf{0}\}$ (at least 1-dimensional), and the image is smaller than the n-dimensional input space: $\operatorname{rank}(A) < n$. In particular, the image $\mathrm{CS}(A)$ is not all of $\mathbb{R}^n$.

- If $\mathbf{b} \notin \mathrm{CS}(A)$, then $A\mathbf{x} = \mathbf{b}$ has no solution.
- If $\mathbf{b} \in \mathrm{CS}(A)$, then $A\mathbf{x} = \mathbf{b}$ has the same number of solutions as $A\mathbf{x} = \mathbf{0}$, i.e., infinitely many since $\dim \mathrm{NS}(A) \geq 1$.

The associated linear transformation $\mathbf{f}$ is not a 1-to-1 correspondence (it maps many vectors to $\mathbf{0}$); thus $\mathbf{f}^{-1}$ does not exist, so A^{-1} does not exist. Row operations do not change the condition $\det A = 0$, so $\det \mathrm{RREF}(A) = 0$, so definitely $\mathrm{RREF}(A) \neq I$. (In fact, $\mathrm{RREF}(A)$ must have at least one 0 along the diagonal.)

Problem 19.10. Devise a test for deciding whether a homogeneous square system $A\mathbf{x} = \mathbf{0}$ has a nonzero solution.

Solution: Compute $\det A$. If $\det A = 0$, there exists a nonzero solution. If $\det A \neq 0$, then $A\mathbf{x} = \mathbf{0}$ has only the zero solution. □

19.5. **Characteristic polynomial, eigenvalues, eigenvectors, again.** Suppose that A is an $n \times n$ matrix.

Characteristic polynomial: $\det(\lambda I - A)$. (If instead you use $\det(A - \lambda I)$, you will need to change the sign when n is odd.)

Expanding out the determinant shows that the characteristic polynomial has the form

$$\lambda^n - (\operatorname{tr} A)\lambda^{n-1} + \cdots \pm \det A$$

(the $\pm$ is $+$ if n is even, and $-$ if n is odd).

Eigenvalue: a scalar λ such that $A\mathbf{v} = \lambda\mathbf{v}$ for some nonzero vector $\mathbf{v}$.

To compute the eigenvalues, find the roots of the characteristic polynomial.

If $\lambda_1, \ldots, \lambda_n$ are all the (possibly complex) eigenvalues, listed with multiplicity, then the characteristic polynomial is

$$(\lambda - \lambda_1) \cdots (\lambda - \lambda_n) = \lambda^n - (\lambda_1 + \cdots + \lambda_n)\lambda^{n-1} + \cdots \pm \lambda_1 \cdots \lambda_n.$$

Comparing coefficients with the previous displayed equation shows that

$$\operatorname{tr} A = \lambda_1 + \lambda_2 + \cdots + \lambda_n \qquad \text{(the sum of the eigenvalues)}$$
$$\det A = \lambda_1 \lambda_2 \cdots \lambda_n \qquad \text{(the product of the eigenvalues).}$$

Problem 19.11. Find the eigenvalues of the upper triangular matrix $A := \begin{pmatrix} 2 & 3 & 5 \\ 0 & 2 & 7 \\ 0 & 0 & 6 \end{pmatrix}$.

Solution: The characteristic polynomial is

$$\det(\lambda I - A) = \det \begin{pmatrix} \lambda - 2 & -3 & -5 \\ 0 & \lambda - 2 & -7 \\ 0 & 0 & \lambda - 6 \end{pmatrix} = (\lambda - 2)(\lambda - 2)(\lambda - 6),$$

so the eigenvalues, listed with multiplicity, are 2, 2, 6. □

In general, for any upper triangular or lower triangular matrix, the eigenvalues are the diagonal entries.

Eigenvector associated to λ: a vector $\mathbf{v}$ such that $A\mathbf{v} = \lambda \mathbf{v}$.

19.6. **Eigenspaces.** For each eigenvalue λ of A, define

$$\begin{aligned} \text{eigenspace of } \lambda &:= \{\text{all eigenvectors associated to } \lambda\} \\ &= \{\text{all solutions to } (A - \lambda I)\mathbf{v} = \mathbf{0}\} \\ &= \operatorname{NS}(A - \lambda I). \end{aligned}$$

There is one eigenspace for each eigenvalue. Each eigenspace is a vector space, so it can be described as the span of a basis. To compute the eigenspace of λ, compute $\operatorname{NS}(A - \lambda I)$ by Gaussian elimination and back-substitution.

Problem 19.12. (Skipped) Find the eigenvalues and eigenvectors of the 90° counterclockwise rotation matrix $R = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$.

Solution: Since $\operatorname{tr} R = 0$ and $\det R = 1$, the characteristic polynomial of R is $\lambda^2 + 1$. Its roots are i and $-i$; these are the eigenvalues.

The eigenspace of i is $\operatorname{NS}(R - iI)$. Converting

$$R - iI = \begin{pmatrix} -i & -1 \\ 1 & -i \end{pmatrix}$$

to row-echelon form gives $\begin{pmatrix} -i & -1 \\ 0 & 0 \end{pmatrix}$, so we solve $-ix - y = 0$ by back-substitution and get the general solution $c\begin{pmatrix} i \\ 1 \end{pmatrix}$. Thus the eigenvectors having eigenvalue i are the nonzero scalar multiples of $\begin{pmatrix} i \\ 1 \end{pmatrix}$.

Applying complex conjugation to the entire universe shows that the eigenvectors having eigenvalue $-i$ are the nonzero scalar multiples of $\begin{pmatrix} -i \\ 1 \end{pmatrix}$. $\square$

19.6.1. *Dimension of an eigenspace.*

Theorem 19.13. *Let λ be an eigenvalue of an $n \times n$ matrix A. Let m be the multiplicity of λ (as a root of the characteristic polynomial). Then*

$$1 \quad \leq \quad (\text{dimension of eigenspace of } \lambda) \quad \leq \quad m.$$

Given λ, the dimension of the eigenspace of λ is also the maximum number of linearly independent eigenvectors of eigenvalue λ that can be found. This dimension is at least 1 since A has at least one nonzero eigenvector of eigenvalue λ (otherwise λ would not have been an eigenvalue). That this dimension is at most m requires more work to prove, and we're not going to do it in this class.

Problem 19.14. A 9×9 matrix has characteristic polynomial $(\lambda - 2)^3(\lambda - 5)^6$. What are the possibilities for the dimension of the eigenspace of 2?

Solution: In this case, $m = 3$, so the dimension is 1, 2, or 3. $\square$

Definition 19.15. The eigenspace of λ is called complete if its dimension equals the multiplicity m of λ, and deficient if its dimension is less than m. Warning: Different authors use different terminology here.

Example 19.16. If the multiplicity is 1, then the dimension of the eigenspace is sandwiched between 1 and 1, so the eigenspace is complete.

Definition 19.17. A matrix is complete if *all* its eigenspaces are complete. A matrix is deficient if at least one of its eigenspaces is deficient.

For the application to solving linear systems of ODEs, given an $n \times n$ matrix A we will want to find as many linearly independent eigenvectors as possible. To do this, we choose a basis of each eigenspace, and concatenate these lists of eigenvectors; it turns out that the resulting long list is linearly independent.

How many eigenvectors are in this list? Well,

$$\begin{aligned}
\#\text{eigenvectors} &= \sum_\lambda (\#\text{eigenvectors from eigenspace of } \lambda) \\
&= \sum_\lambda \dim(\text{eigenspace of } \lambda) \\
&\le \sum_\lambda (\text{multiplicity of } \lambda) \\
&= \#\text{roots of the characteristic polynomial counted with multiplicity} \\
&= \deg(\text{characteristic polynomial}) \qquad \text{(by the fundamental theorem of algebra)} \\
&= n,
\end{aligned}$$

and the $\le$ is $=$ if and only if all the eigenspaces are complete.

Conclusion:

- If A is complete, we get n eigenvectors forming a basis of $\mathbb{R}^n$ (or $\mathbb{C}^n$ if complex).
- If A is deficient, we get less than n eigenvectors, not enough for a basis.

Why does concatenating the bases produce a linearly independent list? The vectors within each basis are linearly independent, and there are no linear relations involving eigenvectors from different eigenspaces because of the following:

Theorem 19.18. *Fix a square matrix A. Eigenvectors corresponding to distinct eigenvalues are linearly independent.*

Proof. Suppose that $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are eigenvectors corresponding to distinct eigenvalues $\lambda_1, \ldots, \lambda_n$. Suppose that there were a linear relation

$$c_1\mathbf{v}_1 + \cdots c_n\mathbf{v}_n = \mathbf{0}.$$

Apply $A - \lambda_1 I$ to both sides; this gets rid of the first summand on the left. Next apply $A - \lambda_2 I$, and so on, up to $A - \lambda_{n-1} I$. This shows that some nonzero number times $c_n\mathbf{v}_n$ equals $\mathbf{0}$. But $\mathbf{v}_n \neq \mathbf{0}$, so $c_n = 0$. Similarly each c_i must be 0. Thus only the trivial relation between $\mathbf{v}_1, \ldots, \mathbf{v}_n$ exists, so they are linearly independent. $\square$

April 6

19.6.2. *Examples.* Here are three examples showing all the situations that can arise for a 2×2 matrix.

Example 19.19. Let $A := \begin{pmatrix} 1 & -2 \\ -1 & 0 \end{pmatrix}$. The characteristic polynomial is $(\lambda - 2)(\lambda + 1)$, so the eigenvalues (2 and -1) each have multiplicity 1, so the eigenspaces are automatically complete. Calculation shows that the eigenspace of 2 has basis $\begin{pmatrix} -2 \\ 1 \end{pmatrix}$ and the eigenspace of -1 has basis $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$; together, these two vectors form a basis for $\mathbb{R}^2$.

Example 19.20. Let $B = \begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix}$. Since B is upper triangular (even diagonal), the eigenvalues are 5, 5. The eigenspace of 5 is $\mathrm{NS}(B - 5I)$, which is the set of solutions to $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \mathbf{x} = \mathbf{0}$, which is the entire space. Its dimension (namely, 2) matches the multiplicity of the eigenvalue 5, so this eigenspace is complete. Every vector is an eigenvector with eigenvalue 5. So it is easy to find two linearly independent eigenvectors: for example, take $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

Example 19.21. Let $C = \begin{pmatrix} 5 & 3 \\ 0 & 5 \end{pmatrix}$. Again the eigenvalues are 5, 5. The eigenspace of 5 is $\mathrm{NS}(C - 5I)$, which is the set of solutions to $\begin{pmatrix} 0 & 3 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{0}$. This system consists of a single nontrivial equation $3y = 0$. Thus the eigenspace is the set of vectors of the form $\begin{pmatrix} c \\ 0 \end{pmatrix}$; it is only 1-dimensional, even though the multiplicity of the eigenvalue 5 is still 2. This means that this eigenspace is deficient, and hence C is deficient. It is impossible to find two linearly independent eigenvectors.

19.7. **Diagonalization.** Suppose that A is a 2×2 matrix with a basis of eigenvectors $\mathbf{v}_1, \mathbf{v}_2$ having eigenvalues λ_1, λ_2 (in other words, we are assuming that A is complete). (Warning: For what we are about to do, the eigenvectors must be listed in the same order as their eigenvalues.)

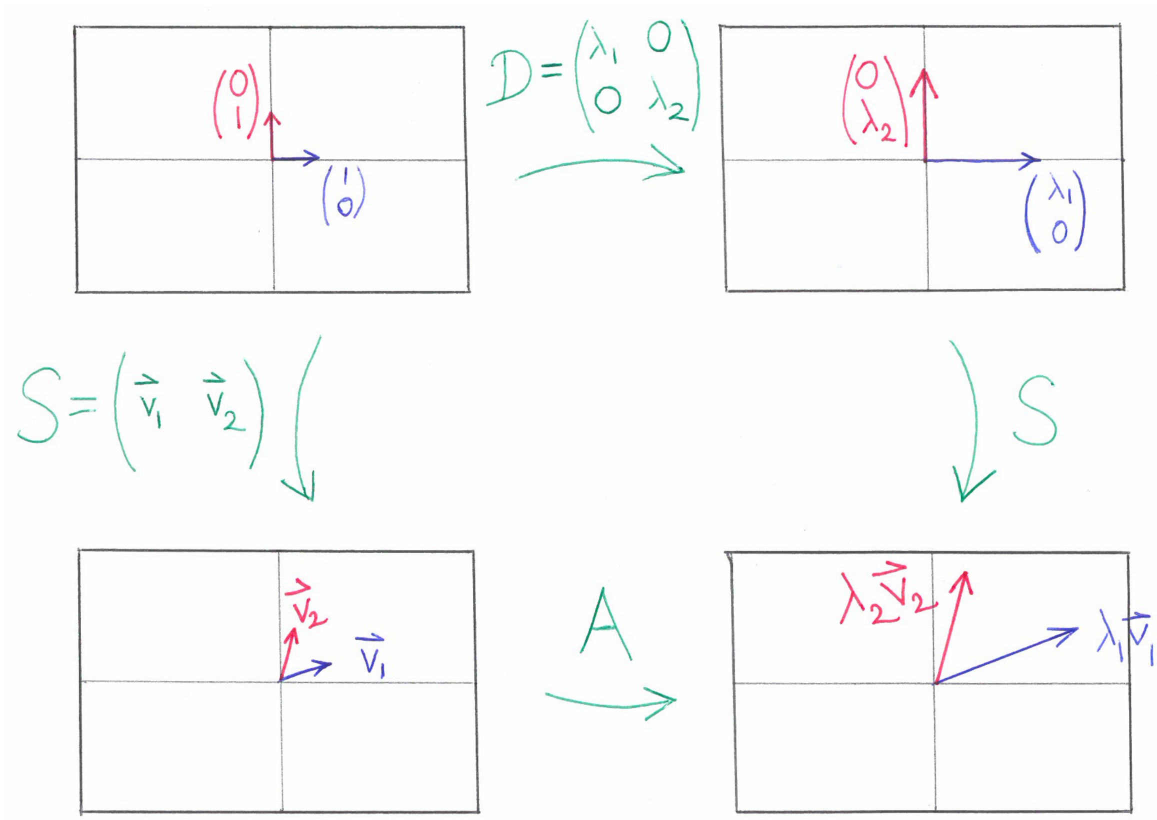
Use the eigenvalues to define a diagonal matrix

$$D := \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

The matrix mapping $\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ to $\mathbf{v}_1, \mathbf{v}_2$, respectively, is the matrix

$$S := \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{pmatrix}$$

whose columns are the eigenvectors $\mathbf{v}_1, \mathbf{v}_2$.



Question 19.22. What is the 2×2 matrix that maps $\begin{pmatrix}1\\0\end{pmatrix}$, $\begin{pmatrix}0\\1\end{pmatrix}$ to $\lambda_1\mathbf{v}_1$, $\lambda_2\mathbf{v}_2$, respectively?

Answer 1: AS, because applying AS means that $\begin{pmatrix}1\\0\end{pmatrix}$, $\begin{pmatrix}0\\1\end{pmatrix}$ are mapped by S to $\mathbf{v}_1, \mathbf{v}_2$, which are then mapped by A to $\lambda_1\mathbf{v}_1$, $\lambda_2\mathbf{v}_2$.

Answer 2: SD, because $\begin{pmatrix}1\\0\end{pmatrix}$, $\begin{pmatrix}0\\1\end{pmatrix}$ are mapped by D to $\begin{pmatrix}\lambda_1\\0\end{pmatrix}$, $\begin{pmatrix}0\\\lambda_2\end{pmatrix}$ which are then mapped by S to $\lambda_1\mathbf{v}_1$, $\lambda_2\mathbf{v}_2$.

Conclusion: $\boxed{AS = SD}$. (Memory aid: Look at where A, S, D are on your keyboard.)

Multiply by S^{-1} on the right to get another way to write it: $\boxed{A = SDS^{-1}}$.

Writing the matrix A like this is called diagonalizing A. Think of S as a "coordinate-change matrix" or "change-of-basis matrix" that

- relates the standard basis $\begin{pmatrix}1\\0\end{pmatrix}$, $\begin{pmatrix}0\\1\end{pmatrix}$ to the basis of eigenvectors of A, and
- relates the easy matrix D scaling the standard basis vectors to the original matrix A scaling the original eigenvectors.

Diagonalization of an $n \times n$ matrix A is possible if and only if A is complete (we need A to have n independent eigenvectors).

Steps to diagonalize a $n \times n$ matrix A (will succeed if and only if A is complete):

1. Find the eigenvalues of A, and list them with multiplicity: $\lambda_1, \ldots, \lambda_n$.
2. Find a basis of each eigenspace.
3. If any eigenspace is deficient, then A is not diagonalizable.
4. Otherwise, we have found a total of n linearly independent eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$, enough to form a basis of $\mathbb{R}^n$ (or $\mathbb{C}^n$). Make sure that they are ordered so that $\mathbf{v}_i$ is associated to λ_i.
5. Set $D := \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$.
6. Set $S := \begin{pmatrix} \vdots & & \vdots \\ \mathbf{v}_1 & & \mathbf{v}_n \\ \vdots & & \vdots \end{pmatrix}$ (the matrix whose columns are the eigenvectors).
7. Write $A = SDS^{-1}$.

Remark 19.23 (Using diagonalization to compute matrix powers). Suppose that $A = SDS^{-1}$. Then

$$A^3 = SD\underbrace{S^{-1}S}_{\text{cancels}}D\underbrace{S^{-1}S}_{\text{cancels}}DS^{-1} = SD^3S^{-1}.$$

More generally, for any integer $n \geq 0$,

$$\boxed{A^n = SD^nS^{-1}}. \tag{9}$$

Later we'll use diagonalization also to compute e^A for a matrix A!

20. Homogeneous linear systems of ODEs, again

20.1. Solving a homogeneous linear system of ODEs.

Steps to find a basis of solutions to $\dot{\mathbf{x}} = A\mathbf{x}$, given a complete $n \times n$ constant matrix A:

1. Compute the eigenvalues (the roots of the characteristic polynomial $\det(\lambda I - A)$).
2. For each eigenvalue λ,
   - compute a basis of the eigenspace $\mathrm{NS}(A - \lambda I)$;
   - for each eigenvector $\mathbf{v}$ in this basis, write down the vector-valued function $e^{\lambda t}\mathbf{v}$.

The total number of functions written down is the sum of the dimensions of the eigenspaces, which is n, provided that A really was complete. These functions form the basis.

Remark 20.1. These n solutions will automatically be linearly independent, since their values at $t = 0$ are the eigenvectors, which are linearly independent. (The chosen eigenvectors within each eigenspace are linearly independent, and there is no linear dependence between eigenvectors with different eigenvalues.)

Remark 20.2. If some of the eigenvalues are complex, they must be included (if you ignore them, you won't find enough eigenvectors). In this case, you may wish to find a new basis of real-valued functions (replace each pair $\boxed{\mathbf{x}, \overline{\mathbf{x}}}$ in the basis by $\boxed{\operatorname{Re}\mathbf{x}, \operatorname{Im}\mathbf{x}}$).

Remark 20.3. If some eigenspace has dimension less than the multiplicity of the eigenvalue (that is, A is deficient), then this method fails: it does not produce enough functions to form a basis.

20.2. **Fundamental matrix.**

20.2.1. *Definition.* Consider a homogeneous linear system of n ODEs $\dot{\mathbf{x}} = A\mathbf{x}$. (We'll assume that A is constant, but everything in this section remains true even if A is replaced by a matrix-valued function $A(t)$.) The dimension theorem says that the set of solutions is an n-dimensional vector space. Let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be a basis of solutions. Write $\mathbf{x}_1, \dots, \mathbf{x}_n$ as column vectors side-by-side to form a matrix

$$X(t) := \begin{pmatrix} \vdots & & \vdots \\ \mathbf{x}_1 & & \mathbf{x}_n \\ \vdots & & \vdots \end{pmatrix}.$$

(It's really a matrix-valued *function*, since each $\mathbf{x}_i$ is a vector-valued function of t.) Any such $X(t)$ is called a fundamental matrix for $\dot{\mathbf{x}} = A\mathbf{x}$. (There are many possible bases, so there are many possible fundamental matrices.)

20.2.2. *General solution in terms of a fundamental matrix.*

What is the point of putting the solutions in a fundamental matrix?

The general solution to $\dot{\mathbf{x}} = A\mathbf{x}$ is $c_1\mathbf{x}_1 + \cdots + c_n\mathbf{x}_n = \begin{pmatrix} \vdots & & \vdots \\ \mathbf{x}_1 & & \mathbf{x}_n \\ \vdots & & \vdots \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$.

Conclusion: If $X(t)$ is a fundamental matrix, then the general solution is $X(t)\,\mathbf{c}$, where $\mathbf{c}$ ranges over constant vectors.

20.2.3. *Solving a homogeneous system of ODEs with initial conditions.*

Problem 20.4. The matrix $A = \begin{pmatrix} 1 & 2 \\ -1 & 4 \end{pmatrix}$ has

- an eigenvector $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$ with eigenvalue 2 and
- an eigenvector $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ with eigenvalue 3.

(a) Find a fundamental matrix for $\dot{\mathbf{x}} = A\mathbf{x}$.

(b) Use it to find the solution to $\dot{\mathbf{x}} = A\mathbf{x}$ satisfying the initial condition $\mathbf{x}(0) = \begin{pmatrix} 4 \\ 5 \end{pmatrix}$.

Solution:

(a) The functions

$$e^{2t}\begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 2e^{2t} \\ e^{2t} \end{pmatrix} \quad \text{and} \quad e^{3t}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} e^{3t} \\ e^{3t} \end{pmatrix}$$

are a basis of solutions, so one fundamental matrix is

$$X(t) = \begin{pmatrix} 2e^{2t} & e^{3t} \\ e^{2t} & e^{3t} \end{pmatrix}.$$

(b) The solution will be $X(t)\,\mathbf{c}$ for some constant vector $\mathbf{c}$. Thus

$$\mathbf{x} = \begin{pmatrix} 2e^{2t} & e^{3t} \\ e^{2t} & e^{3t} \end{pmatrix}\begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

for some c_1, c_2 to be determined. Set $t = 0$ and use the initial condition to get

$$\begin{pmatrix} 4 \\ 5 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} c_1 \\ c_2 \end{pmatrix}.$$

In other words,

$$\begin{aligned} 2c_1 + c_2 &= 4 \\ c_1 + c_2 &= 5. \end{aligned}$$

Solving leads to $c_1 = -1$ and $c_2 = 6$, so

$$\mathbf{x} = \begin{pmatrix} 2e^{2t} & e^{3t} \\ e^{2t} & e^{3t} \end{pmatrix}\begin{pmatrix} -1 \\ 6 \end{pmatrix} = \begin{pmatrix} -2e^{2t} + 6e^{3t} \\ -e^{2t} + 6e^{3t} \end{pmatrix}. \quad \square$$

20.2.4. *Criterion for a matrix to be a fundamental matrix.* To say that each column of $X(t)$ is a solution is the same as saying that $\dot{X} = AX$, because the matrix multiplication can be done column-by-column.

For a $n \times n$ matrix whose columns are solutions, to say that the columns form a basis is equivalent to saying that they are linearly independent (the space of solutions is n-dimensional, so if n solutions are linearly independent, their span is the entire space). By the existence and uniqueness theorem, linear independence of solutions is equivalent to linear independence of their initial values at $t = 0$, i.e., to linear independence of the columns of $X(0)$. So it is equivalent to say that $X(0)$ is a nonsingular matrix.

Conclusion:

Theorem 20.5. *A matrix-valued function $X(t)$ is a fundamental matrix for $\dot{\mathbf{x}} = A\mathbf{x}$ if and only if*

- $\dot{X} = AX$ *and*
- *the matrix* $X(0)$ *is nonsingular.*

April 9

20.3. **Matrix exponential.**

20.3.1. *Definition.* Inspired by the formula for a real (or complex) number x,

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots ,$$

define, for any square matrix A,

$$e^A := I + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \cdots .$$

So e^A is another matrix of the same size as A.

20.3.2. *Properties.*

- $\boxed{e^{\mathbf{0}} = I}$ (here $\mathbf{0}$ is the zero matrix)
  (Proof: $e^{\mathbf{0}} = I + \mathbf{0} + \frac{\mathbf{0}^2}{2!} + \cdots = I$.)
- $\boxed{\dfrac{d}{dt} e^{At} = Ae^{At}}$
  (Proof: Take the derivative of e^{At} term by term.)
- $\boxed{\text{If } AB = BA, \text{ then } e^{A+B} = e^A e^B.}$ (Warning: This can fail if $AB \neq BA$.)
  (Proof: Skipped.)
- $\boxed{\text{If } D = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}, \text{ then } e^D = \begin{pmatrix} e^{\lambda_1} & 0 \\ 0 & e^{\lambda_2} \end{pmatrix}.}$
  (Proof: $D^2 = \begin{pmatrix} \lambda_1^2 & 0 \\ 0 & \lambda_2^2 \end{pmatrix}$, $D^3 = \begin{pmatrix} \lambda_1^3 & 0 \\ 0 & \lambda_2^3 \end{pmatrix}$, and so on. Thus

$$e^D = I + D + \frac{D^2}{2!} + \cdots = \begin{pmatrix} 1 + \lambda_1 + \frac{\lambda_1^2}{2!} + \cdots & 0 \\ 0 & 1 + \lambda_2 + \frac{\lambda_2^2}{2!} + \cdots \end{pmatrix} = \begin{pmatrix} e^{\lambda_1} & 0 \\ 0 & e^{\lambda_2} \end{pmatrix}.$$

  A similar statement holds for diagonal matrices of any size.)

20.3.3. *Exponential of a diagonalizable matrix.*

Suppose that A is diagonalizable: $A = SDS^{-1}$. Substituting $A^n = SD^nS^{-1}$ for each term in the power series definition of e^A leads to

$$\boxed{e^A = Se^DS^{-1}}.$$

Use this formula to compute e^A! (It works whenever A is diagonalizable (complete).)

20.3.4. *Matrix exponential and systems of ODEs.*

Theorem 20.6. *The function e^{At} is a fundamental matrix for the system $\dot{\mathbf{x}} = A\mathbf{x}$.*

Proof. The function e^{At} satisfies $\dot{X} = AX$ and $\underbrace{\text{its value at } 0}_{I}$ is nonsingular. ☐

Consequence: $\boxed{\text{The general solution to } \dot{\mathbf{x}} = A\mathbf{x} \text{ is } e^{At}\,\mathbf{c}.}$

Compare:
The solution to $\dot{x} = ax$ satisfying the initial condition $x(0) = c$ is $\boxed{e^{at}c}$.
The solution to $\dot{\mathbf{x}} = A\mathbf{x}$ satisfying the initial condition $\mathbf{x}(0) = \mathbf{c}$ is $\boxed{e^{At}\,\mathbf{c}}$.

Question 20.7. If the solution is as simple as $e^{At}\,\mathbf{c}$, why did we bother with the method involving eigenvalues and eigenvectors?

Answer: Because computing e^{At} is usually hard! (In fact, the standard method for computing it involves finding the eigenvalues and eigenvectors of A.)

Problem 20.8. Use the matrix exponential to find the solution to the system

$$\begin{aligned} \dot{x} &= 2x + y \\ \dot{y} &= \phantom{2x +{}} 2y \end{aligned}$$

satisfying $x(0) = 5$ and $y(0) = 7$.

Solution: This is $\dot{\mathbf{x}} = A\mathbf{x}$ with $A := \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix} = \underbrace{\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}}_{D} + \underbrace{\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}}_{N}$. Then $N^2 = 0$, so

$$e^{Nt} = I + Nt + 0 + 0 + \cdots = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}.$$

Also, Dt and Nt commute (a scalar times I commutes with any matrix of the same size), so

$$\begin{aligned} e^{At} &= e^{Dt+Nt} \\ &= e^{Dt}e^{Nt} \\ &= \begin{pmatrix} e^{2t} & 0 \\ 0 & e^{2t} \end{pmatrix} \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} e^{2t} & te^{2t} \\ 0 & e^{2t} \end{pmatrix} \\ \begin{pmatrix} x \\ y \end{pmatrix} &= e^{At} \begin{pmatrix} 5 \\ 7 \end{pmatrix} \\ &= \begin{pmatrix} e^{2t} & te^{2t} \\ 0 & e^{2t} \end{pmatrix} \begin{pmatrix} 5 \\ 7 \end{pmatrix} \\ &= \begin{pmatrix} 5e^{2t} + 7te^{2t} \\ 7e^{2t} \end{pmatrix}. \quad \square \end{aligned}$$

21. Inhomogeneous linear systems of ODEs

21.1. **Diagonalization and decoupling.**

21.1.1. *Solving a decoupled system.* The system $\dot{\mathbf{x}} = \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix} \mathbf{x}$ is the same as

$$\begin{aligned} \dot{x} &= 3x \\ \dot{y} &= \qquad 2y. \end{aligned}$$

This is a decoupled system, consisting of two ODEs that can be solved separately. More generally, if D is a *diagonal* matrix of any size, the inhomogeneous system

$$\dot{\mathbf{x}} = D\mathbf{x} + \mathbf{q}(t)$$

consists of first-order linear ODEs that can be solved separately.

Plan: develop a method to transform other systems into this form.

21.1.2. *Decoupling.* Here is a slightly silly way to solve

$$\dot{\mathbf{x}} = A\mathbf{x}$$

for a complete matrix $A = SDS^{-1}$. Substitute $\mathbf{x} = S\mathbf{y}$, and rewrite the system in terms of $\mathbf{y}$:

$$\begin{aligned} S\dot{\mathbf{y}} &= AS\mathbf{y} && \text{(since } S \text{ is constant)} \\ S\dot{\mathbf{y}} &= SD\mathbf{y} && \text{(since } AS = SD\text{)} \\ \dot{\mathbf{y}} &= D\mathbf{y} && \text{(we multiplied by } S^{-1} \text{ on the left).} \end{aligned}$$

This is decoupled! So solve for each coordinate function of $\mathbf{y}$, and then compute $\mathbf{x} = S\mathbf{y}$. $\square$

Why is that silly? Because we already know how to solve $\dot{\mathbf{x}} = A\mathbf{x}$ when we have a basis of eigenvectors (and their eigenvalues).

But... the same decoupling method also lets us solve an *inhomogeneous* linear system, and that's not silly:

Steps to solve $\dot{\mathbf{x}} = A\mathbf{x} + \mathbf{q}(t)$ by decoupling (when A is complete):

1. Find the eigenvalues of A (with multiplicity), and put them in a diagonal matrix D.
2. Find a basis of each eigenspace. If the total number of independent eigenvectors found is less than n, then a more complicated method (not discussed here) is required. Put the eigenvectors as columns of a matrix S.
3. Substitute $\mathbf{x} = S\mathbf{y}$ to get

$$\begin{aligned} S\dot{\mathbf{y}} &= AS\mathbf{y} + \mathbf{q}(t) \\ S\dot{\mathbf{y}} &= SD\mathbf{y} + \mathbf{q}(t) \\ \dot{\mathbf{y}} &= D\mathbf{y} + S^{-1}\mathbf{q}(t). \end{aligned}$$

   (You may skip to the last of these equations.) This is a decoupled system of inhomogeneous linear ODEs.
4. Solve for each coordinate function of $\mathbf{y}$.
5. Compute $S\mathbf{y}$; the result is $\mathbf{x}$.

Problem 21.1. Find a particular solution to $\dot{\mathbf{x}} = A\mathbf{x} + \mathbf{q}$, where $A := \begin{pmatrix} -4 & -3 \\ 6 & 5 \end{pmatrix}$ and $\mathbf{q} = \begin{pmatrix} 0 \\ \cos t \end{pmatrix}$.

Solution: We will solve it instead with $\mathbf{q} = \begin{pmatrix} 0 \\ e^{it} \end{pmatrix}$ (complex replacement), and take the real part of the solution at the very end.

Step 1. We have $\operatorname{tr} A = 1$ and $\det A = -20 - (-18) = -2$.
Characteristic polynomial: $\lambda^2 - \lambda - 2 = (\lambda - 2)(\lambda + 1)$.
Eigenvalues: $2, -1$. Therefore define

$$D := \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix}.$$

Step 2. Calculating eigenspaces in the usual way leads to corresponding eigenvectors $\begin{pmatrix} 1 \\ -2 \end{pmatrix}$, $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$, so define

$$S := \begin{pmatrix} 1 & 1 \\ -2 & -1 \end{pmatrix}.$$

Now $A = SDS^{-1}$.

Step 3. The result of substituting $\mathbf{x} = S\mathbf{y}$ is

$$\dot{\mathbf{y}} = D\mathbf{y} + S^{-1}\mathbf{q}.$$

We have

$$\begin{aligned} S^{-1}\mathbf{q} &= \frac{1}{\det S} \begin{pmatrix} -1 & -1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ e^{it} \end{pmatrix} \\ &= \begin{pmatrix} -e^{it} \\ e^{it} \end{pmatrix} \end{aligned}$$

so the decoupled system is

$$\begin{aligned} \dot{y}_1 &= 2y_1 - e^{it} \\ \dot{y}_2 &= -y_2 + e^{it}. \end{aligned}$$

Step 4. Solving with ERF gives particular solutions

$$\begin{aligned} y_1 &= \frac{-1}{i-2} e^{it} = \left(\frac{2}{5} + \frac{1}{5} i \right) e^{it} \\ y_2 &= \frac{1}{i+1} e^{it} = \left(\frac{1}{2} - \frac{1}{2} i \right) e^{it}. \end{aligned}$$

Step 5.

$$\begin{aligned} \mathbf{x} &= S\mathbf{y} \\ &= \begin{pmatrix} 1 & 1 \\ -2 & -1 \end{pmatrix} \begin{pmatrix} \frac{2}{5} + \frac{1}{5} i \\ \frac{1}{2} - \frac{1}{2} i \end{pmatrix} e^{it} \\ &= \begin{pmatrix} \frac{9}{10} - \frac{3}{10} i \\ -\frac{13}{10} + \frac{1}{10} i \end{pmatrix} (\cos t + i \sin t). \end{aligned}$$

Final step: Take the real part to get a particular solution to the original system:

$$\mathbf{x} = \begin{pmatrix} \frac{9}{10} \cos t + \frac{3}{10} \sin t \\ -\frac{13}{10} \cos t - \frac{1}{10} \sin t \end{pmatrix}.$$

21.2. **Variation of parameters.**

Long ago we learned how to use variation of parameters to solve inhomogeneous linear ODEs

$$\dot{y} + p(t)y = q(t).$$

Now we're going to use the same idea to solve an inhomogeneous linear *system* of ODEs such as

$$\dot{\mathbf{x}} = A\mathbf{x} + \mathbf{q},$$

where $\mathbf{q}$ is a vector-valued function of t. First find a basis of solutions to the corresponding homogeneous system

$$\dot{\mathbf{x}} = A\mathbf{x},$$

and put them together to form a fundamental matrix X (a matrix-valued function of t). We know that $X\mathbf{c}$, where c ranges over constant vectors, is the general solution to the homogeneous equation. Replace $\mathbf{c}$ by a vector-valued function $\mathbf{u}$: try $\mathbf{x} = X\mathbf{u}$ in the original system:

$$\begin{aligned}
\dot{\mathbf{x}} &= A\mathbf{x} + \mathbf{q} \\
\dot{X}\mathbf{u} + X\dot{\mathbf{u}} &= AX\mathbf{u} + \mathbf{q} \\
AX\mathbf{u} + X\dot{\mathbf{u}} &= AX\mathbf{u} + \mathbf{q} \\
X\dot{\mathbf{u}} &= \mathbf{q} \\
\dot{\mathbf{u}} &= X^{-1}\mathbf{q}
\end{aligned}$$

Steps to solve $\dot{\mathbf{x}} = A\mathbf{x} + \mathbf{q}$ by variation of parameters:

1. Find a fundamental matrix X for the homogeneous system $\dot{\mathbf{x}} = A\mathbf{x}$ (e.g., by using eigenvalues and eigenvectors to find a basis of solutions).
2. Substitute $\mathbf{x} = X\mathbf{u}$ for a vector-valued function $\mathbf{u}$; this eventually leads to

   $$\dot{\mathbf{u}} = X^{-1}\mathbf{q}$$

   (and you may jump right to this if you want).
3. Compute the right hand side and integrate each component function to find $\mathbf{u}$.
   (The indefinite integral will have a $+\mathbf{c}$.)
4. Then $\mathbf{x} = X\mathbf{u}$ is the general solution to the inhomogeneous equation.
   (It is a *family* of vector-valued functions because of the $+\mathbf{c}$ in $\mathbf{u}$.)

Remark 21.2. Not mentioned in lecture. One choice of X is e^{At}, in which case $\dot{\mathbf{u}} = e^{-At}\mathbf{q}$ and

$$\mathbf{x} = e^{At}\mathbf{u} = e^{At}\int e^{-At}\mathbf{q}\,dt.$$

April 11

22. Coordinates

22.1. Coordinates with respect to a basis.

Problem 22.1. The vectors $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$, $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$ form a basis for $\mathbb{R}^2$, so any vector in $\mathbb{R}^2$ is a linear combination of them. Find c_1 and c_2 such that

$$c_1 \begin{pmatrix} 2 \\ 1 \end{pmatrix} + c_2 \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \end{pmatrix}$$

(These c_1, c_2 are called the coordinates of $\begin{pmatrix} 2 \\ 4 \end{pmatrix}$ with respect to the basis. There is only one solution, since if there were two different linear combinations giving $\begin{pmatrix} 2 \\ 4 \end{pmatrix}$, subtracting them would give a nontrivial linear combination giving $\mathbf{0}$, which is impossible since the basis vectors are linearly independent.)

Solution: Multiply it out to get

$$\begin{aligned} 2c_1 - c_2 &= 2 \\ c_1 + c_2 &= 4 \end{aligned}$$

or equivalently, in matrix form,

$$\begin{pmatrix} 2 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \end{pmatrix}.$$

Solving gives $c_1 = 2$ and $c_2 = 2$.

"Coordinates with respect to a basis" make sense also in vector spaces of *functions*.

Problem 22.2. Let V be the vector space with basis consisting of the three functions 1, $t-3$, $(t-3)^2$. Find the coordinates of the function t^2 with respect to this basis.

Solution: We need to find c_1, c_2, c_3 such that

$$t^2 = c_1(1) + c_2(t-3) + c_3(t-3)^2.$$

If two polynomials are equal *as functions* (i.e., for all values of t), then their coefficients must match. Equating constant terms gives

$$0 = c_1 - 3c_2 + 9c_3.$$

Equating coefficients of t gives

$$0 = c_2 - 6c_3.$$

Equating coefficients of t^2 gives

$$1 = c_3.$$

Solving this system of three equations leads to $(c_1, c_2, c_3) = (9, 6, 1)$. $\square$

22.2. **Orthogonal basis and orthonormal basis.**

Consider a basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of $\mathbb{R}^n$.

- If each $\mathbf{v}_i$ is orthogonal (perpendicular) to every other $\mathbf{v}_j$, then the basis is called an orthogonal basis.
- If in addition each $\mathbf{v}_i$ has length 1, then the basis is called an orthonormal basis. (The vectors in it are neither abnormally long nor abnormally short!)

Example 22.3. The vectors $\mathbf{i}, \mathbf{j}, \mathbf{k}$ form an orthonormal basis for $\mathbb{R}^3$.

Flashcard question: What is true of the list of vectors $\mathbf{v}_1 := \begin{pmatrix} 2 \\ 2 \end{pmatrix}$ and $\mathbf{v}_2 := \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ in $\mathbb{R}^2$?

Possible answers:

- This is not a basis of $\mathbb{R}^2$.
- This is a basis, but not an orthogonal basis.
- This is an orthogonal basis, but not an orthonormal basis.
- This is an orthonormal basis.

Answer: It is an orthogonal basis, but not an orthonormal basis. To test, use dot products: $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$, so they are orthogonal. But $\mathbf{v}_1 \cdot \mathbf{v}_1 \neq 1$, so $\mathbf{v}_1$ does not have length 1. $\square$

22.3. **Shortcuts for finding coordinates.**

Question 22.4. Suppose that $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is an *orthonormal* basis of $\mathbb{R}^n$. How can we find the coordinates $c_1, \ldots, c_n$ of a vector $\mathbf{w}$ with respect to this basis?

Answer: We need to solve

$$\mathbf{w} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$$

for $c_1, \ldots, c_n$. Trick: dot both sides with $\mathbf{v}_1$ to get

$$\mathbf{w} \cdot \mathbf{v}_1 = c_1(1) + 0 + \cdots + 0.$$

Get

$$\boxed{c_1 = \mathbf{w} \cdot \mathbf{v}_1}, \quad \cdots \quad , \boxed{c_n = \mathbf{w} \cdot \mathbf{v}_n}. \quad \square$$

Question 22.5. Suppose that $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is only an *orthogonal* basis of $\mathbb{R}^n$. How can we find the coordinates $c_1, \ldots, c_n$ of a vector $\mathbf{w}$ with respect to this basis?

Answer: The same trick leads to

$$\mathbf{w} \cdot \mathbf{v}_1 = c_1 \mathbf{v}_1 \cdot \mathbf{v}_1,$$

so we get

$$\boxed{c_1 = \frac{\mathbf{w} \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}}, \quad \cdots \quad, \boxed{c_n = \frac{\mathbf{w} \cdot \mathbf{v}_n}{\mathbf{v}_n \cdot \mathbf{v}_n}}. \quad \square$$

23. Introduction to Fourier series

23.1. Periodic functions.

Because

$$\sin(t + 2\pi) = \sin t, \quad \cos(t + 2\pi) = \cos t,$$

hold for all t, the functions $\sin t$ and $\cos t$ are called periodic with period 2π. In general, "$f(t)$ is periodic of period P" means that $f(t + P) = f(t)$ for all t (or at least all t for which either side is defined).

There are many such functions beyond the sinusoidal functions. To construct one, divide the real line into intervals of length P, start with any function defined on one such interval $[t_0, t_0 + P)$, and then copy its values in the other intervals. The entire graph consists of horizontally shifted copies of the width P graph.

Today: $P = 2\pi$, interval $[-\pi, \pi)$.

Question 23.1. Is $\sin 3t$ periodic of period 2π?

Answer: The *shortest* period is $2\pi/3$, but $\sin 3t$ is also periodic with period any positive integer multiple of $2\pi/3$, including $3(2\pi/3) = 2\pi$:

$$\sin(3(t + 2\pi)) = \sin(3t + 6\pi) = \sin 3t.$$

So the answer is yes.

23.2. Square wave.

Define

$$\operatorname{Sq}(t) := \begin{cases} 1, & \text{if } 0 < t < \pi, \\ -1 & \text{if } -\pi < t < 0. \end{cases}$$

and extend it to a periodic function of period 2π, called a square wave. The function $\operatorname{Sq}(t)$ has jump discontinuities, for example at $t = 0$. If you must define $\operatorname{Sq}(0)$, compromise between the upper and lower values: $\operatorname{Sq}(0) := 0$. The graph is usually drawn with vertical segments at the jumps (even though this violates the vertical line test).

It turns out that

$$\operatorname{Sq}(t) = \frac{4}{\pi}\left(\sin t + \frac{\sin 3t}{3} + \frac{\sin 5t}{5} + \cdots\right).$$

We'll explain later today where this comes from.

Try the "Fourier Coefficients" mathlet

http://mathlets.org/mathlets/fourier-coefficients/

23.3. **Fourier series.** A linear combination like $2\sin 3t - 4\sin 7t$ is another periodic function of period 2π.

Definition 23.2. A Fourier series is a linear combination of the infinitely many functions $\cos nt$ and $\sin nt$ as n ranges over integers:

$$\begin{aligned} f(t) = \frac{a_0}{2} &+ a_1 \cos t + a_2 \cos 2t + a_3 \cos 3t + \cdots \\ &+ b_1 \sin t + b_2 \sin 2t + b_3 \sin 3t + \cdots \end{aligned}$$

(Terms like $\cos(-2t)$ are redundant since $\cos(-2t) = \cos 2t$. Also $\sin 0t = 0$ produces nothing new. But $\cos 0t = 1$ is included; the first term is the coefficient $a_0/2$ times the function 1. We'll explain later why we write $a_0/2$ instead of a_0.)

Written using sigma-notation,

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nt + \sum_{n=1}^{\infty} b_n \sin nt.$$

Recall that, for example, $\sum_{n=1}^{\infty} b_n \sin nt$ means the sum of the series whose n^{th} term is obtained by plugging in the positive integer n into the expression $b_n \sin nt$, so

$$\sum_{n \geq 1} b_n \sin nt = b_1 \sin t + b_2 \sin 2t + b_3 \sin 3t + \cdots .$$

Any Fourier series as above is periodic of period 2π. (Later we'll extend to the definition of Fourier series to include functions of other periods.) The numbers a_n and b_n are called the Fourier coefficients of f. Each summand ($a_0/2$, $a_n \cos nt$, or $b_n \sin nt$) is called a Fourier component of f.

Fourier's theorem. *"Every" periodic function f of period 2π "is" a Fourier series, and the Fourier coefficients are uniquely determined by f.*

(The word "Every" has to be taken with a grain of salt: The function has to be "reasonable". Piecewise differentiable functions with jump discontinuities are reasonable, as are virtually all other functions that arise in physical applications.

The word "is" has to be taken with a grain of salt: If f has a jump discontinuity at τ, then the Fourier series might disagree with f there; the value of the Fourier series at τ is always the average of the left limit $f(\tau^-)$ and the right limit $f(\tau^+)$, regardless of the actual value of $f(\tau)$.)

In other words, the functions

$$1,\ \cos t,\ \cos 2t,\ \cos 3t,\ \ldots,\ \sin t,\ \sin 2t,\ \sin 3t,\ \ldots$$

form a basis for the vector space of "all" periodic functions of period 2π.

Question 23.3. Given f, how do you find the Fourier coefficients a_n and b_n?

In other words, how do you find the coordinates of f with respect to the basis of cosines and sines?

23.4. **A "dot product" for functions.** If $\mathbf{v}$ and $\mathbf{w}$ are vectors in $\mathbb{R}^n$, then

$$\mathbf{v}\cdot\mathbf{w} := \sum_{i=1}^{n} v_i w_i.$$

Can one define the dot product of two functions? Sort of.

Definition 23.4. If f and g are real-valued periodic functions with period 2π, then their inner product is

$$\langle f, g\rangle := \int_{-\pi}^{\pi} f(t)g(t)\,dt$$

(It acts like a dot product $f\cdot g$, but don't write it that way because $\cdot$ could be misinterpreted as multiplication.)

Example 23.5. By definition,

$$\langle 1, \cos t\rangle = \int_{-\pi}^{\pi} \cos t\,dt = 0.$$

Thus the functions 1 and $\cos t$ are orthogonal.

In fact, calculating all the inner products shows that

$$\boxed{1, \cos t, \cos 2t, \cos 3t, \ldots, \sin t, \sin 2t, \sin 3t, \ldots \quad \text{is an } \textit{orthogonal} \text{ basis!}}$$

Question 23.6. Is it an orthonormal basis?

Answer: No, since $\langle 1, 1\rangle = \int_{-\pi}^{\pi} 1\,dt = 2\pi \neq 1$. □

Example 23.7.

$$\langle \sin t, \sin t\rangle = \int_{-\pi}^{\pi} \sin^2 t\,dt = \ ?$$

$$\langle \cos t, \cos t\rangle = \int_{-\pi}^{\pi} \cos^2 t\,dt = \ ?$$

Since $\cos t$ is just a shift of $\sin t$, the answers are going to be the same. Also, the two answers add up to

$$\int_{-\pi}^{\pi} \underbrace{(\sin^2 t + \cos^2 t)}_{\text{this is 1}}\,dt = 2\pi,$$

so each is π.

The same idea works to show that

$$\boxed{\langle \cos nt, \cos nt\rangle = \pi} \qquad \text{and} \qquad \boxed{\langle \sin nt, \sin nt\rangle = \pi}$$

for each positive integer n.

23.5. **Fourier coefficient formulas.** Given f, how do you find the a_n and b_n such that

$$\begin{aligned} f(t) = \frac{a_0}{2} &+ a_1 \cos t + a_2 \cos 2t + a_3 \cos 3t + \cdots \\ &+ b_1 \sin t + b_2 \sin 2t + b_3 \sin 3t + \cdots ? \end{aligned}$$

By the shortcut formulas in Section 22.3,

$$a_n = \frac{\langle f, \cos nt \rangle}{\langle \cos nt, \cos nt \rangle} = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \cos nt \, dt,$$

and the coefficient of 1 is

$$\frac{a_0}{2} = \frac{\langle f, 1 \rangle}{\langle 1, 1 \rangle} = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) \, dt.$$

so multiplying by 2 gives

$$a_0 = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \cos 0t \, dt.$$

(The unexplained point of why we write the constant term of a Fourier series as $a_0/2$ is now explained: it ensures that the formula for a_n for $n > 0$ works also for $n = 0$.) A similar formula holds for b_n.

Conclusion: Given f, its Fourier coefficients can be calculated as follows:

$$\boxed{a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \cos nt \, dt} \qquad \text{for all } n \geq 0,$$

$$\boxed{b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \sin nt \, dt} \qquad \text{for all } n \geq 1.$$

23.6. **Meaning of the constant term.** The constant term of the Fourier series of f is

$$\frac{a_0}{2} = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) \, dt,$$

which is the average value of f on $(-\pi, \pi)$.

23.7. **Even and odd symmetry.**

- A function $f(t)$ is even if $f(-t) = f(t)$ for all t.
- A function $f(t)$ is odd if $f(-t) = -f(t)$ for all t.

If

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nt + \sum_{n=1}^{\infty} b_n \sin nt,$$

then substituting $-t$ for t gives

$$f(-t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nt + \sum_{n=1}^{\infty} (-b_n) \sin nt.$$

The right hand sides match if and only if $b_n = 0$ for all n.

Conclusion: The Fourier series of an even function f has only cosine terms (including the constant term):

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nt.$$

Similarly, the Fourier series of an odd function f has only sine terms:

$$f(t) = \sum_{n=1}^{\infty} b_n \sin nt.$$

Example 23.8. The square wave $\mathrm{Sq}(t)$ is an odd function, so

$$\mathrm{Sq}(t) = \sum_{n=1}^{\infty} b_n \sin nt$$

for some numbers b_n. The Fourier coefficient formula says

$$\begin{aligned}
b_n &= \frac{1}{\pi}\int_{-\pi}^{\pi} \underbrace{\mathrm{Sq}(t)\sin nt}_{\text{even}}\,dt \\
&= \frac{2}{\pi}\int_{0}^{\pi} \mathrm{Sq}(t)\sin nt\,dt \qquad \text{(the two halves of the integral are equal, by symmetry)} \\
&= \frac{2}{\pi}\int_{0}^{\pi} \sin nt\,dt \qquad \text{(since } \mathrm{Sq}(t) = 1 \text{ whenever } 0 < t < \pi\text{)} \\
&= \left.\frac{2(-\cos nt)}{\pi n}\right|_0^{\pi} \\
&= \frac{2}{\pi n}(-\cos n\pi + \cos 0) \\
&= \begin{cases} \dfrac{4}{\pi n}, & \text{if } n \text{ is odd,} \\ 0, & \text{if } n \text{ is even.} \end{cases}
\end{aligned}$$

Thus

$$b_1 = \frac{4}{\pi}, \quad b_3 = \frac{4}{3\pi}, \quad b_5 = \frac{4}{5\pi}, \ldots$$

and all other Fourier coefficients are 0.

Conclusion:

$$\mathrm{Sq}(t) = \frac{4}{\pi}\left(\sin t + \frac{\sin 3t}{3} + \frac{\sin 5t}{5} + \cdots\right). \quad \square$$

April 13

23.8. **Finding a Fourier series representing a function on an interval.**

Problem 23.9. Suppose that $f(t)$ is a (reasonable) function defined only on the interval $(0, \pi)$. Find numbers $a_0, a_1, \ldots$ such that

$$f(t) = \frac{a_0}{2} + a_1 \cos t + a_2 \cos 2t + \cdots$$

for all $t \in (0, \pi)$.

Solution: For any a_i, the right hand side will define an even periodic function of period 2π (if the series converges). So begin by extending $f(t)$ to a function of the same type:

- Extend $f(t)$ to an even function on $(-\pi, \pi)$ by *defining* $f(-t) := f(t)$ for all $t \in (-\pi, 0)$ (and then define $f(0)$ and $f(-\pi)$ arbitrarily).
- Shift the graph of f horizontally by integer multiples of 2π to get a period 2π function defined on all of $\mathbb{R}$.

Define

$$a_n := \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \cos nt \, dt = \frac{2}{\pi} \int_0^{\pi} f(t) \cos nt \, dt.$$

Then

$$f(t) = \frac{a_0}{2} + a_1 \cos t + a_2 \cos 2t + \cdots$$

holds for all $t \in \mathbb{R}$, so in particular it holds for $t \in (0, \pi)$ (possibly excluding points of discontinuity). $\square$

Remark 23.10. The same function $f(t)$ on $(0, \pi)$ can be extended to an odd periodic function of period 2π, in order to obtain

$$f(t) = b_1 \sin t + b_2 \sin 2t + \cdots ,$$

where

$$b_n := \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \sin nt \, dt = \frac{2}{\pi} \int_0^{\pi} f(t) \sin nt \, dt.$$

23.9. **Solving ODEs with Fourier series.**

23.9.1. *Warm-ups.*

Review problem 1: Given a positive integer n, what periodic function $x(t)$ of period 2π is a solution to

$$\ddot{x} + 50x = e^{int} \, ?$$

Solution: The characteristic polynomial is $p(r) := r^2 + 50$. ERF gives

$$x(t) = \frac{1}{p(in)} e^{int} = \frac{1}{(in)^2 + 50} e^{int} = \frac{1}{50 - n^2} e^{int}.$$

(This is periodic of period 2π.) $\square$

Review problem 2: Given a positive integer n, what periodic function $x(t)$ of period 2π is a solution to

$$\ddot{x} + 50x = \sin nt \text{ ?}$$

Solution: Take imaginary parts of the previous solution to get

$$x(t) = \operatorname{Im}\left(\frac{1}{50-n^2}e^{int}\right) = \frac{1}{50-n^2}\operatorname{Im}\left(e^{int}\right) = \frac{1}{50-n^2}\sin nt. \ \square$$

23.9.2. *System response to a periodic input signal.*

Steps to solve $p(D)x = f(t)$, where $f(t)$ is a periodic function of period 2π:

1. Find the Fourier series of $f(t)$. (Use the Fourier coefficient formulas and use even/odd symmetry as a shortcut if possible.)
2. Find the periodic solution to $p(D)x = \cos nt$ and/or $p(D)x = \sin nt$ for each n, as needed, by using complex replacement and ERF.
3. Use superposition: Because $f(t)$ is a linear combination of cosines and sines, the periodic solution to $p(D)x = f(t)$ will be the corresponding linear combination of the periodic solutions to the ODEs $p(D)x = \cos nt$ and/or $p(D)x = \sin nt$.
4. To get the general solution, add in the general solution to the associated homogeneous equation $p(D)x = 0$.

OK, now we're ready for the main event:

Problem 23.11. Suppose that $f(t)$ is an odd periodic function of period 2π. What periodic function $x(t)$ of period 2π is a solution to

$$\ddot{x} + 50x = f(t) \text{ ?}$$

Solution: Since f is odd, the Fourier series of f is a linear combination of the shape

$$f(t) = b_1 \sin t + b_2 \sin 2t + b_3 \sin 3t + \cdots .$$

By the superposition principle, the system response to $f(t)$ is

$$x(t) = b_1 \frac{1}{49}\sin t + b_2 \frac{1}{46}\sin 2t + b_3 \frac{1}{41}\sin 3t + \cdots .$$

Note that each Fourier component $\sin nt$ has a different gain: the gain depends on the frequency.

One could write the answer using sigma-notation:

$$\boxed{x(t) = \sum_{n\geq 1} \frac{1}{50-n^2} b_n \sin nt}.$$

This is better since it shows precisely what every term in the series is (no need to "guess the pattern"). $\square$

Think of $f(t)$ as the input signal, and the solution $x(t)$ as the system response (output signal). Summary of the solution:

input signal	system response
e^{int}	$\dfrac{1}{50-n^2}e^{int}$
$\sin nt$	$\dfrac{1}{50-n^2}\sin nt$
$\sin t$	$\dfrac{1}{49}\sin t$
$\sin 2t$	$\dfrac{1}{46}\sin 2t$
$\sin 3t$	$\dfrac{1}{41}\sin 3t$
$\vdots$	$\vdots$
$\sum_{n\geq 1} b_n \sin nt$	$\sum_{n\geq 1} \dfrac{1}{50-n^2} b_n \sin nt$

23.9.3. *Near resonance.*

Problem 23.12. For which input signal $\sin nt$ is the gain the largest?

Solution: The complex gain is $\dfrac{1}{50-n^2}$. The gain is $\left|\dfrac{1}{50-n^2}\right|$, which is largest when $|50-n^2|$ is smallest. This happens for $n=7$. $\square$

The gain for $\sin 7t$ is 1, and the next largest gain, occurring for $\sin 6t$ and $\sin 8t$, is $\frac{1}{14}$. Thus the system approximately filters out all the Fourier components of $f(t)$ except for the $\sin 7t$ term.

Problem 23.13. Let $x(t)$ be the periodic solution to

$$\ddot{x} + 50x = \frac{\pi}{4}\operatorname{Sq}(t).$$

Which Fourier coefficient of $x(t)$ is largest? Which is second largest?

Solution: The input signal

$$\frac{\pi}{4}\operatorname{Sq}(t) = \sum_{n\geq 1,\ \text{odd}} \frac{\sin nt}{n}$$

elicits the system response

$$\begin{aligned} x(t) &= \sum_{n\geq 1,\text{ odd}} \left(\frac{1}{50-n^2}\right)\frac{\sin nt}{n} \\ &\approx 0.020 \sin t + 0.008 \sin 3t + 0.008 \sin 5t + 0.143 \sin 7t - 0.003 \sin 9t - (\text{even smaller terms}) \end{aligned}$$

so the coefficient of $\sin 7t$ is largest, and the coefficient of $\sin t$ is second largest. (This makes sense since the Fourier coefficient $\dfrac{1}{(50-n^2)n}$ is large only when one of n or $50-n^2$ is small.) $\square$

Remark 23.14. Even though the system response is a complicated Fourier series, with infinitely many terms, only one or two are significant, and the rest are negligible.

23.9.4. *Pure resonance.* What happens if we change 50 to 49 in the ODE?

Flashcard question: Which of the following is true of the ODE

$$\ddot{x} + 49x = \frac{\pi}{4}\operatorname{Sq}(t)?$$

Possible answers:

- There are no solutions.
- There is exactly one solution, but it is not periodic.
- There is exactly one solution, and it is periodic.
- There are infinitely many solutions, but none of them are periodic.
- There are infinitely many solutions, but only one of them is periodic.
- There are infinitely many solutions, and all of them are periodic.

Answer: There are infinitely many solutions, but none of them are periodic. Here is why: For $n \neq 7$, we can solve $\ddot{x} + 49x = \sin nt$ using complex replacement and ERF since in is not a root of $r^2 + 49$. For $n = 7$, we can still solve $\ddot{x} + 49x = \sin 7t$ (the existence and uniqueness theorem guarantees this), but the solution requires generalized ERF, and involves t, and hence is not periodic: it turns out that one solution is $-\frac{t}{14}\cos 7t$.

For the input signal $\operatorname{Sq}(t)$, we can find a solution x_p by superposition: most of the terms will be periodic, but one of them will be $\frac{1}{7}\left(-\frac{t}{14}\cos 7t\right)$, and this makes the whole solution x_p non-periodic.

There are infinitely many other solutions, namely $x_p + c_1\cos 7t + c_2 \sin 7t$ for any c_1 and c_2, but these solutions still include the $\frac{1}{7}\left(-\frac{t}{14}\cos 7t\right)$ term and hence are not periodic. $\square$

Remark 23.15. If the ODE had been

$$\ddot{x} + 36x = \frac{\pi}{4}\operatorname{Sq}(t)$$

then all solutions would have been periodic, because $\frac{\pi}{4}\operatorname{Sq}(t)$ has no $\sin 6t$ term in its Fourier series.

In general, for a periodic function f, the ODE $p(D)\,x = f(t)$ has a periodic solution if and only if for each term $\cos\omega t$ or $\sin\omega t$ appearing with a nonzero coefficient in the Fourier series of f, the number $i\omega$ is not a root of $p(r)$.

23.9.5. *Resonance with damping.*

Problem 23.16. Describe the steady-state solution to

$$\ddot{x} + \underbrace{0.1\dot{x}}_{\text{damping term}} + 49x = \frac{\pi}{4}\operatorname{Sq}(t).$$

Recall: The steady-state solution is the periodic solution. (Other solutions will be a sum of the steady-state solution with a transient solution solving the homogeneous ODE

$$\ddot{x} + 0.1\dot{x} + 49x = 0;$$

these transient solutions tend to 0 as $t \to \infty$, because the coefficients of the characteristic polynomial are positive (in fact, this is an underdamped system).

Solution: First, let's solve a complex replacement ODE

$$\ddot{z} + 0.1\dot{z} + 49z = e^{int}.$$

The characteristic polynomial is $p(r) = r^2 + 0.1r + 49$. ERF gives

$$z = \frac{1}{p(in)}e^{int} = \frac{1}{(49 - n^2) + (0.1n)i}e^{int},$$

with complex gain $\dfrac{1}{(49 - n^2) + (0.1n)i}$ and gain $g_n := \dfrac{1}{|(49 - n^2) + (0.1n)i|}$.

Next, take imaginary parts of z to get that the solution to

$$\ddot{x} + 0.1\dot{x} + 49x = \sin nt$$

is

$$x = \operatorname{Im}\left(\frac{1}{(49 - n^2) + (0.1n)i}e^{int}\right);$$

this is a sinusoid of amplitude g_n, so $x = g_n\cos(nt - \phi_n)$ for some ϕ_n.

Finally, the input signal

$$\frac{\pi}{4}\operatorname{Sq}(t) = \sum_{n\ge1,\text{ odd}} \frac{\sin nt}{n},$$

elicits the system response

$$\begin{aligned} x(t) &= \sum_{n\ge1,\text{ odd}} g_n\frac{\cos(nt - \phi_n)}{n} \\ &\approx 0.020\cos(t - \phi_1) + 0.008\cos(3t - \phi_3) + 0.008\cos(5t - \phi_5) \\ &\qquad + 0.204\cos(7t - \phi_7) + 0.003\cos(9t - \phi_9) + (\text{even smaller terms}). \end{aligned}$$

Conclusion: The system response is almost indistinguishable from a pure sinusoid of angular frequency 7.

23.10. **Listening to Fourier series.**

You are not required to know the material in this section for exams.

Try the "Fourier Coefficients: Complex with Sound" mathlet

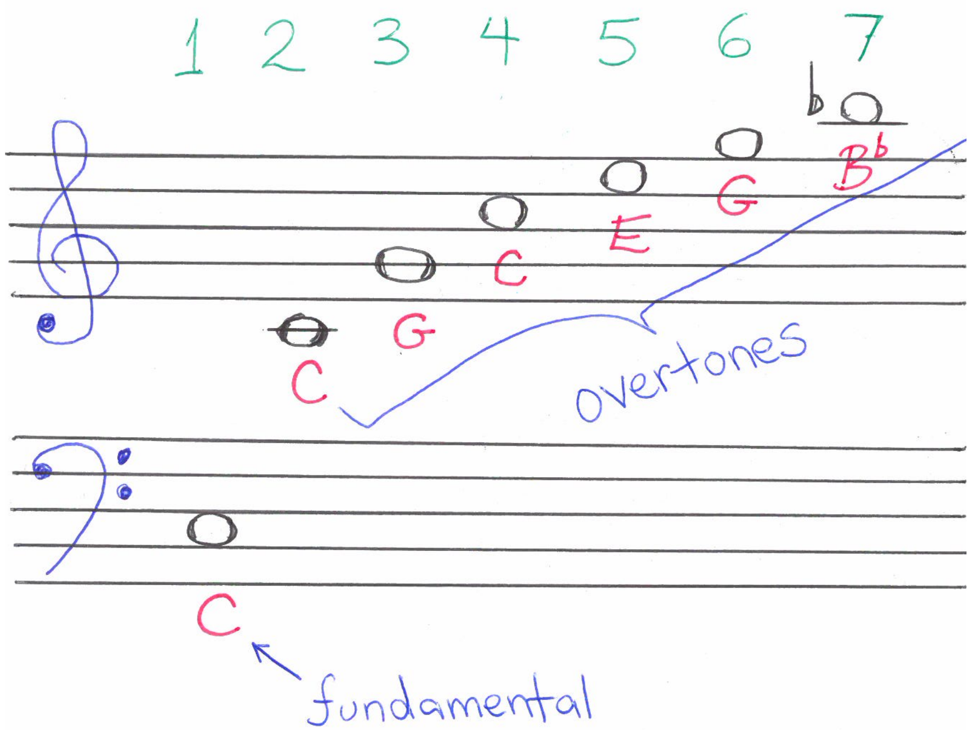
`http://mathlets.org/mathlets/fourier-coefficients-complex/`

If using headphones, start with a low volume, since pure sine waves carry more energy than they seem to, and can damage your hearing after sustained listening.

Your ear is capable of decomposing a sound wave into its Fourier components of different frequencies. Each frequency corresponds to a certain pitch. Increasing the frequency produces a higher pitch. More precisely, multiplying the frequency by a number greater than 1 increases the pitch by what in music theory is called an interval. For example, multiplying the frequency by 2 raises the pitch by an octave, and multiplying by 3 raises the pitch an octave plus a perfect fifth.

When an instrument plays a note, it is producing a periodic sound wave in which typically many of the Fourier coefficients are nonzero. In a general Fourier series, the combination of the first two nonconstant terms ($a_1 \cos t + b_1 \sin t$, if the period is 2π) is a sinusoid of some frequency ν, and the next combination (e.g., $a_2 \cos 2t + b_2 \sin 2t$) has frequency 2ν, and so on: the frequencies are the positive integer multiples of the lowest frequency ν. The note corresponding to the frequency ν is called the fundamental, and the notes corresponding to frequencies 2ν, 3ν, ... are called the overtones.

The musical staffs below show these for $\nu \approx 131\,\text{Hz}$ (the C below middle C), with the integer multiplier shown in green.



Question 23.17. Can you guess what note corresponds to 9ν?

Can you hear the phases of the sinusoids? No.

April 18

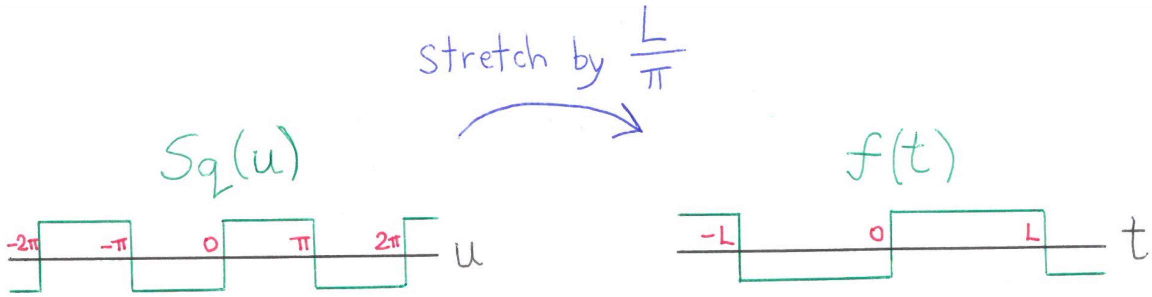
23.11. **Fourier series of arbitrary period.** Everything we did with periodic functions of period 2π can be generalized to periodic functions of other periods.

Problem 23.18. Define

$$f(t) := \begin{cases} 1, & \text{if } 0 < t < L, \\ -1 & \text{if } -L < t < 0. \end{cases}$$

and extend it to a periodic function of period $2L$. Express this new square wave $f(t)$ in terms of Sq.

Solution: To avoid confusion, let's use u as the variable for Sq. Stretching the graph of $\mathrm{Sq}(u)$ horizontally by a factor L/π produces the graph of $f(t)$.



In other words, if t and u are related by $t = \frac{L}{\pi}u$ (so that $u = \pi$ corresponds to $t = L$), then $f(t) = \operatorname{Sq}(u)$. In other words, $u = \frac{\pi t}{L}$, so $\boxed{f(t) = \operatorname{Sq}\left(\frac{\pi t}{L}\right)}$. $\square$

Similarly we can stretch any function of period 2π to get a function of different period. Let L be a positive real number. Start with "any" periodic function

$$g(u) = \frac{a_0}{2} + \sum_{n\geq 1} a_n \cos nu + \sum_{n\geq 1} b_n \sin nu,$$

of period 2π. Stretching horizontally by a factor L/π gives a periodic function $f(t)$ of period $2L$, and "every" f of period $2L$ arises this way. By the same calculation as above,

$$\begin{aligned} f(t) &= g\left(\frac{\pi t}{L}\right) \\ &= \frac{a_0}{2} + \sum_{n\geq 1} a_n \cos \frac{n\pi t}{L} + \sum_{n\geq 1} b_n \sin \frac{n\pi t}{L}. \end{aligned}$$

The substitution $u = \frac{\pi t}{L}$ (and $du = \frac{\pi}{L}\,dt$) also leads to Fourier coefficient formulas for period $2L$:

$$\begin{aligned} a_n &= \frac{1}{\pi}\int_{-\pi}^{\pi} g(u)\cos nu\,du \\ &= \frac{1}{\pi}\int_{-L}^{L} g\left(\frac{\pi t}{L}\right)\cos\left(\frac{n\pi t}{L}\right)\frac{\pi}{L}\,dt \\ &= \frac{1}{L}\int_{-L}^{L} f(t)\cos\left(\frac{n\pi t}{L}\right)dt. \end{aligned}$$

A similar formula gives b_n in terms of f.

23.11.1. *The inner product for periodic functions of period $2L$.* Adapt the definition of the inner product to the case of functions f and g of period $2L$:

$$\langle f, g\rangle := \int_{-L}^{L} f(t)g(t)\,dt.$$

(This conflicts with the earlier definition of $\langle f, g\rangle$, for functions for which both make sense, so perhaps it would be better to write $\langle f, g\rangle_L$ for the new inner product, but we won't bother to do so.)

The same calculations as before show that the functions

$$1, \cos\frac{\pi t}{L}, \cos\frac{2\pi t}{L}, \cos\frac{3\pi t}{L}, \ldots, \sin\frac{\pi t}{L}, \sin\frac{2\pi t}{L}, \sin\frac{3\pi t}{L}, \ldots$$

form an orthogonal basis for the vector space of "all" periodic functions of period $2L$, with

$$\begin{aligned} \langle 1, 1\rangle &= 2L \\ \left\langle \cos\frac{n\pi t}{L}, \cos\frac{n\pi t}{L} \right\rangle &= L \\ \left\langle \sin\frac{n\pi t}{L}, \sin\frac{n\pi t}{L} \right\rangle &= L \end{aligned}$$

(the average value of $\cos^2 \omega t$ is $1/2$ for any ω, and the average value of $\sin^2 \omega t$ is $1/2$ too).

This gives another way to derive the Fourier coefficient formulas for functions of period $2L$.

23.11.2. *Summary.*

- Fourier's theorem: "Every" periodic function f of period $2L$ is a Fourier series

$$\boxed{f(t) = \frac{a_0}{2} + \sum_{n\ge1} a_n \cos\frac{n\pi t}{L} + \sum_{n\ge1} b_n \sin\frac{n\pi t}{L}}.$$

- Given f, the Fourier coefficients a_n and b_n can be computed using:

$$\boxed{a_n = \frac{1}{L}\int_{-L}^{L} f(t)\cos\frac{n\pi t}{L}\,dt} \qquad \text{for all } n \ge 0,$$

$$\boxed{b_n = \frac{1}{L}\int_{-L}^{L} f(t)\sin\frac{n\pi t}{L}\,dt} \qquad \text{for all } n \ge 1.$$

- If f is even, then only the cosine terms (including the $a_0/2$ term) appear.
- If f is odd, then only the sine terms appear.

Problem 23.19. Define

$$s(t) := \begin{cases} 8, & \text{if } 0 < t < 5, \\ 2, & \text{if } -5 < t < 0, \end{cases}$$

and extend it to a periodic function of period 10. Find the Fourier series for $s(t)$.

Solution: One way would be to use the Fourier coefficient formulas directly. But we will instead obtain the Fourier series for $s(t)$ from the Fourier series for $\operatorname{Sq}(u)$, by stretching and shifting.

First, stretch horizontally by a factor of $5/\pi$ to get

$$\operatorname{Sq}\left(\frac{\pi t}{5}\right) = \begin{cases} 1, & \text{if } 0 < t < 5, \\ -1, & \text{if } -5 < t < 0. \end{cases}$$

Here the difference between the upper and lower values is 2, but for $s(t)$ we want a difference of 6, so multiply by 3:

$$3\,\mathrm{Sq}\left(\frac{\pi t}{5}\right) = \begin{cases} 3, & \text{if } 0 < t < 5, \\ -3, & \text{if } -5 < t < 0. \end{cases}$$

Finally add 5:

$$5 + 3\,\mathrm{Sq}\left(\frac{\pi t}{5}\right) = \begin{cases} 8, & \text{if } 0 < t < 5, \\ 2, & \text{if } -5 < t < 0. \end{cases}$$

Since

$$\mathrm{Sq}(u) = \frac{4}{\pi} \sum_{n \ge 1, \text{ odd}} \frac{1}{n} \sin nu,$$

we get

$$\begin{aligned} s(t) &= 5 + 3\,\mathrm{Sq}\left(\frac{\pi t}{5}\right) \\ &= 5 + 3\left(\frac{4}{\pi}\right) \sum_{n \ge 1, \text{ odd}} \frac{1}{n} \sin \frac{n\pi t}{5} \\ &= 5 + \sum_{n \ge 1, \text{ odd}} \frac{12}{n\pi} \sin \frac{n\pi t}{5}. \end{aligned}$$

23.12. **Convergence of a Fourier series.**

Definition 23.20. A periodic function f of period $2L$ is called piecewise differentiable if

- there are at most finitely many points in $[-L, L)$ where $f'(t)$ does not exist, and
- at each such point τ, the left limit $f(\tau^-) := \lim_{t\to\tau^-} f(t)$ and right limit $f(\tau^+) := \lim_{t\to\tau^+} f(t)$ exist (although they might be unequal, in which case we say that f has a jump discontinuity at τ).

Theorem 23.21. *If f is a piecewise differentiable periodic function, then the Fourier series of f (with the a_n and b_n defined by the Fourier coefficient formulas)*

- *converges to $f(t)$ at values of t where f is continuous, and*
- *converges to $\dfrac{f(t^-) + f(t^+)}{2}$ where f has a jump discontinuity.*

Example 23.22. The left limit $\mathrm{Sq}(0^-) = -1$ and right limit $\mathrm{Sq}(0^+) = 1$ average to 0. The Fourier series

$$\frac{4}{\pi}\left(\sin t + \frac{\sin 3t}{3} + \frac{\sin 5t}{5} + \cdots\right)$$

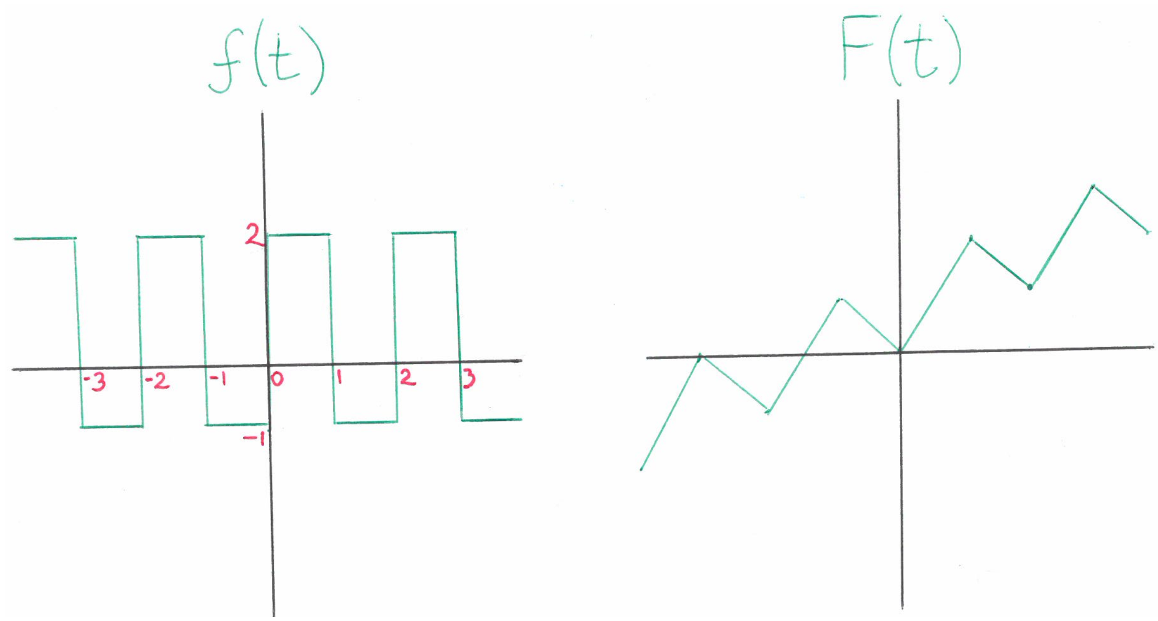
evaluated at $t = 0$ converges to 0 too.

23.13. **Antiderivative of a Fourier series.** Suppose that f is a piecewise differentiable periodic function. For any number C, the formula $F(t) := \int_0^t f(\tau)\,d\tau + C$ defines an antiderivative of f in the sense that $F'(t) = f(t)$ at any t where f is continuous. (If f has jump discontinuities, then at the jump discontinuities F will be only continuous, not differentiable.)

The function F is not necessarily periodic! For example, if f is a function of period 2 such that

$$f(t) := \begin{cases} 2, & \text{if } 0 < t < 1, \\ -1 & \text{if } -1 < t < 0, \end{cases}$$

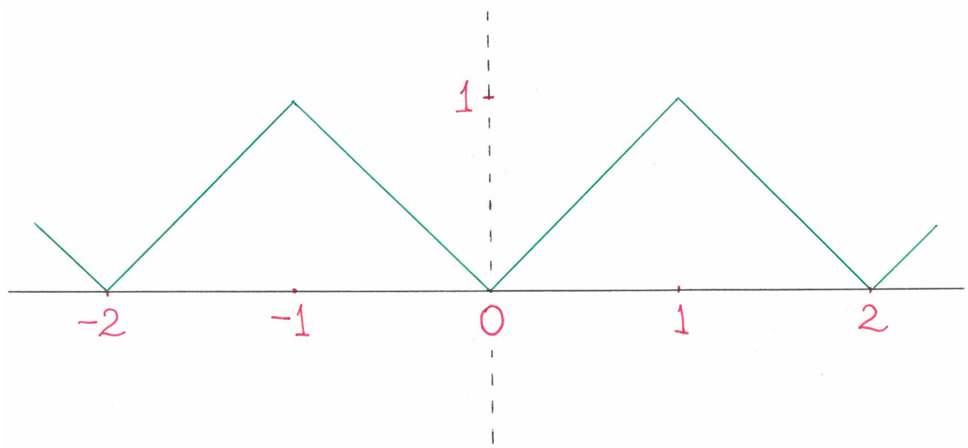
then $F(t)$ creeps upward over time.



An even easier example: if $f(t) = 1$, then $F(t) = t + C$ for some C, so $F(t)$ is not periodic.

But if the constant term $a_0/2$ in the Fourier series of f is 0, then F *is* periodic, and its Fourier series can be obtained by taking the simplest antiderivative of each cosine and sine term, and adding an overall $+C$, where C is the average value of F.

Problem 23.23. Let $T(t)$ be the periodic function of period 2 such that $T(t) = |t|$ for $-1 \le t \le 1$; this is called a triangle wave. Find the Fourier series of $T(t)$.



Solution: We could use the Fourier coefficient formula. But instead, notice that $T(t)$ has slope -1 on $(-1,0)$ and slope 1 on $(0,1)$, so $T(t)$ is an antiderivative of the period 2 square wave

$$\operatorname{Sq}(\pi t) = \sum_{n \geq 1,\ \text{odd}} \frac{4}{n\pi} \sin n\pi t.$$

Taking an antiderivative termwise (and using that the average value of $T(t)$ is $1/2$) gives

$$\begin{aligned} T(t) &= \frac{1}{2} + \sum_{n \geq 1,\ \text{odd}} \frac{4}{n\pi}\left(\frac{-\cos n\pi t}{n\pi}\right) \\ &= \frac{1}{2} - \sum_{n \geq 1,\ \text{odd}} \frac{4}{n^2\pi^2} \cos n\pi t. \quad \square \end{aligned}$$

April 20

Warning: If a periodic function f is not continuous, it will not be an antiderivative of any piecewise differentiable function, so you cannot find the Fourier series of f by integration.

Remark 23.24. A Fourier series of a piecewise differentiable periodic function f can also be *differentiated* termwise, but the result will often fail to converge. For example, the termwise derivative of the Fourier series $\operatorname{Sq}(t)$ gives a nonsensical value at $t = 0$. (Here is one good case, however: If f is continuous and piecewise *twice* differentiable, then the derivative series converges.)

Midterm 3 covers everything up to here.

23.14. **Review.**

23.14.1. *Game: What is the matrix?* The following two matrices exhibit different phenomena:

$$\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, \qquad \begin{pmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{pmatrix}.$$

You take the blue matrix, and it's nonsingular, so things look pretty nice, just a little distorted.
You take the red matrix, and dimensions get crushed!
Remember: all I'm offering is the truth.

- Which matrix satisfies $\det A = 0$? The red one.
- Which matrix has area scaling factor 0? The red one.
- For which matrix are there nonzero vectors in $\mathrm{NS}(A)$? The red one.
- Which matrix has rank 2? The blue one.
- Which matrix has column space equal to the whole space $\mathbb{R}^2$? The blue one.
- Which matrix defines a function $\mathbb{R}^2 \to \mathbb{R}^2$ that is a 1-to-1 correspondence? The blue one.
- For which matrix A is it true that for every vector $\mathbf{b}$ in $\mathbb{R}^2$, the system $A\mathbf{x} = \mathbf{b}$ is solvable? The blue one.
- Which matrix has an inverse? The blue one.
- Which matrix has $\mathrm{RREF}(A) = I$? The blue one.

The point is that the answer to the question "Is $\det A = 0$?" can tell you a lot about a matrix.

23.14.2. *Solving an ODE using Fourier series.*

Problem 23.25. Define $f(t) = |t|$ for $-\pi \le t \le \pi$ and extend $f(t)$ to a periodic function of period 2π. Find the periodic solution to $\dot{x} + x = f(t)$.

Solution: Either using the Fourier coefficient formulas or integrating $\mathrm{Sq}(t)$ shows that

$$\begin{aligned} f(t) &= \frac{\pi}{2} - \frac{4}{\pi}\left(\cos t + \frac{\cos 3t}{9} + \frac{\cos 5t}{25} + \cdots\right) \\ &= \frac{\pi}{2} - \sum_{n\ge 1,\ \text{odd}} \frac{4}{\pi n^2}\cos nt. \end{aligned}$$

(The constant term is $\frac{\pi}{2}$ since that is the average value of f.)

The strategy is to first solve

$$\dot{x} + x = \cos nt$$

and then take a linear combination to solve the original ODE. For this, first solve the complex replacement

$$\dot{z} + z = e^{int}$$

by using ERF: the characteristic polynomial is $p(r) := r+1$, so ERF gives a periodic solution

$$\begin{aligned} z &= \frac{1}{p(in)} e^{int} \\ &= \frac{1}{1+in} e^{int} \\ &= \frac{1}{1+in} \left(\frac{1-in}{1-in} \right) e^{int} \\ &= \frac{1-in}{1+n^2} (\cos nt + i \sin nt) \\ &= \frac{1}{1+n^2} ((\cos nt + n \sin nt) + i(-n \cos nt + \sin nt)). \end{aligned}$$

Taking the real part gives the periodic solution to $\dot{x} + x = \cos nt$:

$$x = \frac{1}{1+n^2} (\cos nt + n \sin nt).$$

In particular, the $n = 0$ case of this says that the periodic solution to $\dot{x} + x = 1$ is $x = 1$ (kind of obvious, in hindsight). Taking a linear combination gives the answer to the original problem:

$$x = \frac{\pi}{2} - \sum_{n \geq 1, \text{ odd}} \frac{4}{\pi n^2} \left(\frac{1}{1+n^2} \right) (\cos nt + n \sin nt).$$

23.14.3. *Resonance and Fourier series.*

Problem 23.26. Let $f(t)$ be the same function as above. For which angular frequencies $\omega > 0$ will the ODE

$$\ddot{x} + \omega^2 x = f(t)$$

exhibit pure resonance (fail to have a periodic solution)?

Solution: The solutions to the harmonic oscillator $\ddot{x} + \omega^2 x = 0$ are the linear combinations of $\cos \omega t$ and $\sin \omega t$; it has natural frequency ω. Pure resonance will occur when one of the Fourier components in the right hand side has a matching angular frequency. There is a nonzero Fourier component of frequency n for each odd integer $n \geq 1$. Thus pure resonance occurs exactly when ω is one of the numbers

$$1,\ 3,\ 5,\ \ldots.$$

23.14.4. *Solving an inhomogeneous system of ODEs.* What are the methods to solve an inhomogeneous system $\dot{\mathbf{x}} = A\mathbf{x} + \mathbf{q}$?

1. Convert to a higher-order ODE involving only one unknown function.
2. Decouple (assuming that A is complete). (Theory: If $A = SDS^{-1}$, substitute $\mathbf{x} = S\mathbf{y}$ to get $S\dot{\mathbf{y}} = (SDS^{-1})S\mathbf{y} + \mathbf{q} = SD\mathbf{y} + \mathbf{q}$, and multiply by S^{-1} on the left to get $\dot{\mathbf{y}} = D\mathbf{y} + S^{-1}\mathbf{q}$.) What you actually do:

- Calculate eigenvalues to get the diagonal matrix D.
- Calculate a basis of each eigenspace, and make all these eigenvectors into columns of a matrix S,
- Compute S^{-1} by converting $[S|I]$ to RREF $[I|?]$.
- Compute $S^{-1}\mathbf{q}$.
- Solve the decoupled system $\dot{\mathbf{y}} = D\mathbf{y} + S^{-1}\mathbf{q}$ for y_1 and y_2 separately.
- Compute $\mathbf{x} = S\mathbf{y}$.

3. Variation of parameters. (Theory: If X is a fundamental matrix for $\dot{\mathbf{x}} = A\mathbf{x}$, then substitute $\mathbf{x} = X\mathbf{u}$ to get

$$\begin{aligned}\dot{X}\mathbf{u} + X\dot{\mathbf{u}} &= AX\mathbf{u} + \mathbf{q}\\ AX\mathbf{u} + X\dot{\mathbf{u}} &= AX\mathbf{u} + \mathbf{q}\\ X\dot{\mathbf{u}} &= \mathbf{q}\\ \dot{\mathbf{u}} &= X^{-1}\mathbf{q},\end{aligned}$$

which can be solved for $\mathbf{u}$.) What you actually do:
   - Find eigenvalues, and a basis for each eigenspace.
   - For each pair (λ, v), write down the solution $e^{\lambda t}\mathbf{v}$.
   - Form the matrix X whose columns are these vector-valued functions.
   - Compute X^{-1}.
   - Compute $X^{-1}\mathbf{q}$.
   - Integrate to find $\mathbf{u}$. (There will be a $+\mathbf{c}$.)
   - Compute $X\mathbf{u}$; this is the general solution.
4. Find one particular solution somehow, and add it to the general solution to the homogeneous system.

There was not time for the rest of the review topics below during the April 20 lecture.

If there are initial conditions, first find the general solution to the inhomogeneous system, and then use the initial conditions to solve for the unknown parameters (and plug them back in at the end).

23.14.5. *Matrix exponential.* How do you compute e^A?

- If $D = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$, then $e^D = \begin{pmatrix} e^{\lambda_1} & 0 \\ 0 & e^{\lambda_2} \end{pmatrix}$, and the same idea works for diagonal matrices of any size.
- If $A = SDS^{-1}$ for some diagonal matrix D and nonsingular matrix S (so A is diagonalizable), then $e^A = Se^DS^{-1}$ (and we just saw how to compute e^D, so this lets you compute e^A).

- If $A^2 = 0$, then $e^A = I + A$ (and $e^{At} = I + At$).
  If $A^3 = 0$, then $e^A = I + A + \frac{A^2}{2!}$ (and $e^{At} = I + At + \frac{A^2}{2!}t^2$).
  The same idea works for all nilpotent matrices, i.e., matrices having a power that is 0.
- If $A = B + N$ where $BN = NB$, then $e^A = e^B e^N$.

It turns out that every square matrix can be written as $B + N$ where B is diagonalizable and N is nilpotent and $BN = NB$; so in principle, e^A can always be computed.

Other facts about e^A and the matrix-valued function e^{At}:

- The derivative of e^{At} is Ae^{At}.
- The matrix-valued function e^{At} is the fundamental matrix (for $\dot{\mathbf{x}} = A\mathbf{x}$) whose value at $t = 0$ is I.
- The solution to $\dot{\mathbf{x}} = A\mathbf{x}$ satisfying the initial condition $\mathbf{x}(0) = \mathbf{c}$ is $e^{At}\,\mathbf{c}$.

April 23

Midterm 3

April 25

24. Boundary value problems

24.1. Review: an initial value problem.

Problem 24.1. Find all solutions to

$$\begin{aligned} v''(x) &= -9\,v(x) \qquad &&\text{(ODE)} \\ v(0) &= 0 &&\text{(initial condition)} \\ v'(0) &= 0 &&\text{(initial condition).} \end{aligned}$$

(The two conditions are initial conditions since they are the value and derivative at the *same* x-value, namely $x = 0$.)

Solution: The function $v(x) = 0$ is one solution. The uniqueness part of the existence and uniqueness theorem says that there is *only* one solution, so 0 is the only solution.

24.2. **New examples: boundary value problems.**

Problem 24.2. Find all functions $v(x)$ on $[0,\pi]$ satisfying

$$\begin{aligned} v''(x) &= -9\,v(x) \qquad \text{(ODE)} \\ v(0) &= 0 \qquad \text{(boundary condition)} \\ v(\pi) &= 0 \qquad \text{(boundary condition)}. \end{aligned}$$

(The two conditions are boundary conditions since they are at *different* x-values.)

Warning: There is no existence and uniqueness theorem for boundary value problems!

Although 0 is still a solution, there is no guarantee that there are not others. In fact, we'll see soon that this particular problem has other solutions, namely $v(x) = b\sin 3x$ for any constant b.

Problem 24.3. Find all functions $v(x)$ on $[0,\pi]$ satisfying

$$\begin{aligned} v''(x) &= -10\,v(x) \qquad \text{(ODE)} \\ v(0) &= 0 \qquad \text{(boundary condition)} \\ v(\pi) &= 0 \qquad \text{(boundary condition)}. \end{aligned}$$

This time it will turn out that 0 is the only solution.

24.3. **Solving a family of boundary value problems.** Let's solve a whole family of boundary value problems like these at once.

Problem 24.4. For each real number λ, find all functions $v(x)$ on $[0,\pi]$ satisfying

$$\begin{aligned} v''(x) &= \lambda\,v(x) \qquad \text{(ODE)} \\ v(0) &= 0 \qquad \text{(boundary condition)} \\ v(\pi) &= 0 \qquad \text{(boundary condition)}. \end{aligned}$$

For which values of λ do nonzero solutions exist?

Solution: This is a homogeneous linear ODE with characteristic polynomial $r^2 - \lambda$, whose roots are $\pm\sqrt{\lambda}$.

Case 1: $\lambda > 0$. Then the general solution is $ae^{\sqrt{\lambda}x} + be^{-\sqrt{\lambda}x}$, and the boundary conditions say

$$\begin{aligned} a + b &= 0 \\ ae^{\sqrt{\lambda}\pi} + be^{-\sqrt{\lambda}\pi} &= 0. \end{aligned}$$

Since $\det \begin{pmatrix} 1 & 1 \\ e^{\sqrt{\lambda}\pi} & e^{-\sqrt{\lambda}\pi} \end{pmatrix} \neq 0$, the only solution to this linear system is $(a, b) = (0, 0)$. Thus the only solution to the boundary value problem is $v = 0$.

Case 2: $\lambda = 0$. Then the general solution is $a + bx$, and the boundary conditions say

$$\begin{aligned} a &= 0 \\ a + b\pi &= 0. \end{aligned}$$

Again the only solution to this linear system is $(a, b) = (0, 0)$. Thus the only solution to the boundary value problem is $v = 0$.

Case 3: $\lambda < 0$. The roots of the characteristic polynomial are again $\pm\sqrt{\lambda}$, but the number λ is negative, so each root will be a real number times i. In order to simplify the formula for these roots, define ω to be the positive real number such that $\lambda = -\omega^2$; then the roots $\pm\sqrt{\lambda}$ are simply $\pm i\omega$. Now the functions $e^{i\omega x}$ and $e^{-i\omega x}$ form a basis of solutions to $v''(x) = \lambda v(x)$. The functions $\cos \omega x$ and $\sin \omega x$ form a real-valued basis for the same vector space. Therefore the general solution is $a \cos \omega x + b \sin \omega x$. The first boundary condition says $a = 0$, so $v = b \sin \omega x$. The second boundary condition then says $b \sin \omega \pi = 0$, which says different things about b, depending on whether ω is an integer:

- If ω is not an integer, then $\sin \omega\pi \neq 0$, so the second condition implies $b = 0$.
- If ω is an integer n, then $\sin \omega\pi = 0$, so b can be anything. In this case, $\lambda = -n^2$ for some positive integer n (positive since $n = \omega > 0$), and $v(x)$ can be $b \sin nx$ for any constant b.

Final answer to Problem 24.4:

- If λ is one of $-1, -4, -9, \ldots$, so $\lambda = -n^2$ for some positive integer n, then the solutions are the functions $\boxed{b \sin nx}$ as b varies.
- For all other values of λ, the only solution is $\boxed{0}$. □

We will use this answer as one step in the solution of the heat equation.

24.4. **Analogy with eigenvalue-eigenvector problems.** To describe a function $v(x)$, one needs to give infinitely many numbers, namely its values at all the different input x-values. Thus $v(x)$ is like a vector with infinitely many coordinates.

The linear differential operator $\dfrac{d^2}{dx^2}$ maps each function to a function, just as a 2×2 matrix defines a linear transformation mapping each vector in $\mathbb{R}^2$ to another vector in $\mathbb{R}^2$. Thus $\dfrac{d^2}{dx^2}$ is like an $\infty \times \infty$ matrix.

The ODE $\dfrac{d^2}{dx^2} v = \lambda v$ (with boundary conditions) amounts to an infinite system of equations: the ODE consists of one equality of numbers at each $x \in (0, \pi)$, and boundary conditions are equalities at the endpoints. Thus the ODE with boundary conditions is like a system of equations $A\mathbf{v} = \lambda\mathbf{v}$. Nonzero solutions $v(x)$ to $\frac{d^2}{dx^2}v = \lambda v$ exist only for special values of λ, namely

$$\lambda = -1,\ -4,\ -9,\ \ldots,$$

just as $A\mathbf{v} = \lambda\mathbf{v}$ has a nonzero solution $\mathbf{v}$ only for special values of λ, namely the eigenvalues of λ. But the differential operator $\frac{d^2}{dx^2}$ has infinitely many eigenvalues, as one would expect for an $\infty \times \infty$ matrix.

The nonzero solutions $v(x)$ to $\frac{d^2}{dx^2}v = \lambda v$ satisfying the boundary conditions are called eigenfunctions, since they act like eigenvectors.

Summary of the analogies:

vector $\mathbf{v}$	function $v(x)$
A	the linear operator $\frac{d^2}{dx^2}$
eigenvalue-eigenvector problem $A\mathbf{v} = \lambda\mathbf{v}$	boundary value problem $\frac{d^2}{dx^2}v = \lambda v$, $v(0) = 0$, $v(\pi) = 0$
eigenvalues λ	eigenvalues $\lambda = -1, -4, -9, \ldots$
eigenvectors $\mathbf{v}$	eigenfunctions $v(x) = \sin nx$

24.5. A little lemma to be used in the solution of the heat equation.

A lemma is a statement that is used as part of an explanation of something more important.

Lemma 24.5. *Suppose that $f(x)$ and $g(t)$ are functions of independent variables x and t, respectively. If $f(x) = g(t)$ for all values of x and t, then there is a constant λ such that $f(x) = \lambda$ for all x and $g(t) = \lambda$ for all t.*

Proof. Both sides of $f(x) = g(t)$ equal the same function.

- It's $f(x)$, so it does not depend on t.
- It's $g(t)$, so it does not depend on x.

So it's a constant, which may be called λ. □

25. Heat equation

25.1. Modeling: temperature in a metal rod.

Much of the modeling and physics was skipped in lecture.

Problem 25.1. An insulated uniform metal rod with exposed ends starts at a constant temperature, but then its ends are held in ice at 0°C. Model its temperature.

Variables and functions: Define

$L:$ length of the rod

$A:$ cross-sectional area of the rod

$u_0:$ initial temperature of the rod

$x:$ position along the rod (from 0 to L)

$t:$ time

$u:$ temperature at a point of the rod at a given time

$q:$ heat flux density at a point of the rod at a given time (to be explained).

Here

- L, A, and u_0 are constants;
- x and t are independent variables; and
- $u = u(x,t)$ and $q = q(x,t)$ are functions defined for $x \in [0, L]$ and $t \geq 0$.

Physics: Each bit of the rod contains internal energy, consisting of the microscopic kinetic energy of particles (and the potential energy associated with microscopic forces). This energy can be transferred from point to point, via atoms colliding with nearby atoms. Heat flux density measures such heat transfer from left to right across a cross-section of the rod, per unit area, per unit time.

We will use three laws of physics:

1. The first law of thermodynamics (conservation of energy), in the special case in which no work is being done, states that for any bit of the rod,

$$\text{(increase in internal energy)} = \text{(net amount of heat flowing in)}.$$

2. For any bit of the rod,

$$\frac{\text{(increase in internal energy)}}{\text{volume}} \propto \text{(increase in temperature)}$$

(The symbol $\propto$ means "proportional to".) The constant of proportionality depends on the material.

3. Fourier's law of heat transfer:

$$q \propto -\frac{\partial u}{\partial x}.$$

This makes sense: If $u(x + dx, t)$ is greater than $u(x, t)$, then the heat flow at x is to the left (negative), and the rate of heat flow is proportional to the (infinitesimal) difference of temperature $u(x + dx, t) - u(x, t)$, just as in Newton's law of cooling.

Deducing the PDE: For any interior bit of rod defined by the interval $[x, x+dx]$, during a time interval $[t, t+dt]$, the first law of thermodynamics states

$$\begin{aligned}
\text{(increase in internal energy)} &= \text{(heat flowing in)} - \text{(heat flowing out)} \\
\text{(increase in temperature)(volume)} &\propto \text{(heat flowing in)} - \text{(heat flowing out)} \\
(u(x,t+dt) - u(x,t))\, A\, dx &\propto q(x,t)\, A\, dt - q(x+dx,t)\, A\, dt.
\end{aligned}$$

Divide by $A\, dx\, dt$ to get

$$\frac{\partial u}{\partial t} \propto -\frac{\partial q}{\partial x}.$$

(More correct would be to use Δx, Δt, and so on, and to take a limit, but the end result is the same.) Finally, substitute Fourier's law $q \propto -\dfrac{\partial u}{\partial x}$ into the right hand side to get the heat equation

$$\boxed{\frac{\partial u}{\partial t} = \alpha \frac{\partial^2 u}{\partial x^2}},$$

for some constant $\alpha > 0$ (called thermal diffusivity or the heat diffusion coefficient) that depends only on the material. The heat equation is a second-order homogeneous linear partial differential equation involving the unknown function $u(x,t)$.

Remark 25.2. This PDE makes physical sense, since if the temperature profile (graph of $u(x,t)$ versus x at a fixed time) is curving upward at a point ($\dfrac{\partial^2 u}{\partial x^2} > 0$), then the average of the point's neighbors is warmer than the point, so the point's temperature should increase.

Boundary conditions: $\boxed{u(0,t) = 0 \text{ and } u(L,t) = 0 \text{ for all } t \geq 0}$ (for u in degrees Celsius).
Initial condition: $\boxed{u(x,0) = u_0 \text{ for all } x \in (0,L)}$.

Try the "Heat Equation" mathlet

`http://mathlets.org/mathlets/heat-equation/`

25.2. **Solving the PDE with homogeneous boundary conditions: separation of variables; normal modes.** Let's now try to solve the PDE. For simplicity, suppose that $L = \pi$, $u_0 = 1$, and $\alpha = 1$. (The general case is similar. In fact, one could reduce to this special case by changes of variable.)

So now we are solving

$$\begin{aligned}
\frac{\partial u}{\partial t} &= \frac{\partial^2 u}{\partial x^2} \\
u(0,t) &= 0 \quad \text{for all } t \geq 0 \quad \text{(boundary condition at } x = 0) \\
u(\pi,t) &= 0 \quad \text{for all } t \geq 0 \quad \text{(boundary condition at } x = \pi) \\
u(x,0) &= 1 \quad \text{for all } x \in (0,\pi) \quad \text{(initial condition at } t = 0).
\end{aligned}$$

Temporarily forget the initial condition $u(x, 0) = 1$ (we'll impose it at the end, as we usually do with initial conditions).

Idea (separation of variables): Look for nonzero solutions of the form

$$u(x, t) := w(t)\, v(x).$$

For which pairs of functions $(v(x), w(t))$ will this be a solution? To test it, substitute it into the PDE:

$$\dot{w}(t)\, v(x) = w(t)\, v''(x)$$
$$\frac{\dot{w}(t)}{w(t)} = \frac{v''(x)}{v(x)}.$$

(at least where $w(t)$ and $v(x)$ are nonzero). By Lemma 24.5, there is a constant λ such that

$$\frac{v''(x)}{v(x)} = \lambda \quad \text{and} \quad \frac{\dot{w}(t)}{w(t)} = \lambda,$$

or in other words,

$$v''(x) = \lambda\, v(x) \quad \text{and} \quad \dot{w}(t) = \lambda\, w(t).$$

Substituting $u(x, t) = w(t)\, v(x)$ into the first boundary condition $u(0, t) = 0$ gives $w(t)v(0) = 0$ for all t, but $w(t)$ is not the zero function, so this translates into $v(0) = 0$. Similarly, the second boundary condition $u(\pi, t) = 0$ translates into $v(\pi) = 0$.

We already solved $v''(x) = \lambda\, v(x)$ subject to the boundary conditions $v(0) = 0$ and $v(\pi) = 0$: nonzero solutions $v(x)$ exist only if $\lambda = -n^2$ for some positive integer n, and in that case

$$v(x) = \sin nx \quad \text{(times any scalar)}.$$

For $\lambda = -n^2$, what is a matching possibility for w? Since $\dot{w} = -n^2 w$,

$$w(t) = e^{-n^2 t} \quad \text{(times any scalar)}.$$

Putting the $v(x)$ and the matching $w(t)$ back together gives one solution

$$u(x, t) = e^{-n^2 t} \sin nx$$

(and its scalar multiples) for each positive integer n, to the PDE with boundary conditions. Each such solution is called a normal mode. (As a check, one could plug in this $u(x, t)$ to verify that it satisfies the PDE and boundary conditions.)

The PDE and boundary conditions are homogeneous, so we can get other solutions by taking linear combinations:

$$\boxed{u(x, t) = b_1 e^{-t} \sin x + b_2 e^{-4t} \sin 2x + b_3 e^{-9t} \sin 3x + \cdots}. \tag{10}$$

This turns out to be the general solution to the PDE with the boundary conditions.

April 27

Summary of last lecture:

- We modeled an insulated metal rod with exposed ends held at 0°C.
- Using physics, we found that its temperature $u(x,t)$ was governed by the PDE
$$\frac{\partial u}{\partial t} = \alpha \frac{\partial^2 u}{\partial x^2} \quad \text{(the heat equation)}.$$
For simplicity, we specialized to the case $\alpha = 1$, length π, and initial temperature $u(x,0) = 1$.
- Trying $u = w(t)v(x)$ led to separate ODEs for v and w, leading to solutions $e^{-n^2 t} \sin nx$ for $n = 1, 2, \ldots$ to the PDE with boundary conditions.
- We took linear combinations to get the general solution
$$u(x,t) = b_1 e^{-t} \sin x + b_2 e^{-4t} \sin 2x + b_3 e^{-9t} \sin 3x + \cdots$$
to the PDE with boundary conditions.

25.3. **Initial condition.** As usual, we postponed imposing the initial condition, but now it is time to impose it.

Question 25.3. Which choices of $b_1, b_2, \ldots$ make the solution above also satisfy the initial condition $u(x,0) = 1$ for $x \in (0, \pi)$?

Set $t = 0$ in (10) and use the initial condition on the left to get
$$1 = b_1 \sin x + b_2 \sin 2x + b_3 \sin 3x + \cdots \quad \text{for } x \in (0, \pi),$$
which must be solved for $b_1, b_2, \ldots$. Section 23.8 showed how to find such b_i: the left hand side extends to an odd function of period 2π, namely $\mathrm{Sq}(x)$, so we need to solve
$$\mathrm{Sq}(x) = b_1 \sin x + b_2 \sin 2x + b_3 \sin 3x + \cdots \quad \text{for all } x \in \mathbb{R}.$$
We already know the answer:
$$\mathrm{Sq}(x) = \frac{4}{\pi} \sin x + \frac{4}{3\pi} \sin 3x + \frac{4}{5\pi} \sin 5x + \cdots.$$
Matching Fourier coefficients gives $b_1 = \frac{4}{\pi}$, $b_2 = 0$, $b_3 = \frac{4}{3\pi}$, etc. In other words, $b_n = 0$ for even n, and $b_n = \frac{4}{n\pi}$ for odd n. Substituting these b_n back into the general solution (10) gives that
$$\boxed{u(x,t) = \frac{4}{\pi} e^{-t} \sin x + \frac{4}{3\pi} e^{-9t} \sin 3x + \frac{4}{5\pi} e^{-25t} \sin 5x + \cdots}$$
is the particular solution that satisfies the initial condition. □

Question 25.4. What does the temperature profile look like when t is large?

Answer: All the Fourier components are decaying, so $u(x,t) \to 0$ as $t \to +\infty$ at every position. Thus the temperature profile approaches a horizontal segment, the graph of the zero function. But the Fourier components of higher frequency decay much faster than the first Fourier component, so when t is large, the formula

$$u(x,t) \approx \frac{4}{\pi} e^{-t} \sin x$$

is a very good approximation. Eventually, the temperature profile is indistinguishable from a sinusoid of angular frequency 1 whose amplitude is decaying to 0. This is what was observed in the mathlet. □

25.4. **Analogy between a linear system of ODEs and the heat equation.** We can continue the table of analogies from Section 24.4:

vector $\mathbf{v}$	function $v(x)$
A	the linear operator $\dfrac{d^2}{dx^2}$
eigenvalue-eigenvector problem $A\mathbf{v} = \lambda \mathbf{v}$	boundary value problem $\dfrac{d^2}{dx^2} v = \lambda v$, $v(0) = 0$, $v(\pi) = 0$
eigenvalues λ	eigenvalues $\lambda = -1, -4, -9, \ldots$
eigenvectors $\mathbf{v}$	eigenfunctions $v(x) = \sin nx$
linear system of ODEs $\dot{\mathbf{x}} = A\mathbf{x}$	heat equation with boundary conditions $\dot{u} = \dfrac{\partial^2}{\partial x^2} u, \quad u(0,t) = 0, \quad u(\pi, t) = 0$
normal modes: $e^{\lambda t}\mathbf{v}$ for an eigenvector $\mathbf{v}$ with eigenvalue λ	normal modes: $e^{\lambda t} v(x) = e^{-n^2 t} \sin nx$ for eigenfunction $v(x) = \sin nx$, eigenvalue $\lambda = -n^2$
General solution: $\mathbf{x} = \sum c_n e^{\lambda_n t} \mathbf{v}_n$	General solution: $u = \sum b_n e^{-n^2 t} \sin nx$
Solve $\mathbf{x}(0) = \sum c_n \mathbf{v}_n$ to get the c_n	Solve $u(x,0) = \sum b_n \sin nx$ to get the b_n

25.5. **Solving the PDE with inhomogeneous boundary conditions.**

Steps to solve a linear PDE with *inhomogeneous* boundary conditions:

1. Find a particular solution u_p to the PDE with the inhomogeneous boundary conditions (but without initial conditions). If the boundary conditions do not depend on t, try to find the steady-state solution $u_p(x,t)$, i.e., the solution that does not depend on t.
2. Find the general solution u_h to the PDE with the homogeneous boundary conditions.
3. Then $u := u_p + u_h$ is the general solution to the PDE with the inhomogeneous boundary conditions.
4. If initial conditions are given, use them to find the specific solution to the PDE with the inhomogeneous boundary conditions. (This often involves finding Fourier coefficients.)

Problem 25.5. Consider the same insulated uniform metal rod as before ($\alpha = 1$, length π, initial temperature 1°C), but now suppose that the left end is held at 0°C while the right end is held at 20°C. Now what is $u(x,t)$?

Solution:

1. Forget the initial condition for now, and look for a solution $u = u(x)$ that does not depend on t. Plugging this into the heat equation PDE gives $0 = \dfrac{\partial^2 u}{\partial x^2}$. The general solution to this simplified DE is $u(x) = ax + b$. Imposing the boundary conditions $u(0) = 0$ and $u(\pi) = 20$ leads to $b = 0$ and $a = 20/\pi$, so $u_p = \dfrac{20}{\pi}x$. (This is the solution whose temperature profile is an unchanging straight line from $u = 0$ at $x = 0$ up to $u = 20$ at $x = \pi$.)
2. The PDE with the homogeneous boundary conditions is what we solved earlier; the general solution is
$$u_h = b_1 e^{-t}\sin x + b_2 e^{-4t}\sin 2x + b_3 e^{-9t}\sin 3x + \cdots .$$
3. The general solution to the PDE with inhomogeneous boundary conditions is
$$u(x,t) = u_p + u_h = \frac{20}{\pi}x + b_1 e^{-t}\sin x + b_2 e^{-4t}\sin 2x + b_3 e^{-9t}\sin 3x + \cdots . \tag{11}$$
4. To find the b_n, set $t = 0$ and use the initial condition on the left:
$$1 = \frac{20}{\pi}x + b_1\sin x + b_2\sin 2x + b_3\sin 3x + \cdots \quad \text{for all } x \in (0,\pi).$$
$$1 - \frac{20}{\pi}x = b_1\sin x + b_2\sin 2x + b_3\sin 3x + \cdots \quad \text{for all } x \in (0,\pi).$$
Extend $1 - \dfrac{20}{\pi}x$ on $(0,\pi)$ to an odd periodic function $f(x)$ of period 2π. Then the b_n are the Fourier coefficients of $f(x)$; they can be calculated in two ways:
   - Use the Fourier coefficient formulas directly:
$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin nx\,dx = \frac{2}{\pi}\int_0^{\pi}\left(1 - \frac{20}{\pi}x\right)\sin nx\,dx.$$
   - Use the Fourier coefficient formulas to find the Fourier series for the odd periodic extensions of 1 and x separately, namely
$$1 = \frac{4}{\pi}\sum_{n\ge 1,\ \text{odd}} \frac{\sin nx}{n}$$
$$x = 2\sum_{n\ge 1}(-1)^{n+1}\frac{\sin nx}{n},$$
for $x \in (0,\pi)$, and take a linear combination to get $1 - \dfrac{20}{\pi}x$.

   Either way, we get
$$f(x) = -\frac{36}{\pi}\sin x + \frac{40}{2\pi}\sin 2x - \frac{36}{3\pi}\sin 3x + \frac{40}{4\pi}\sin 4x - \cdots ;$$

that is,

$$b_1 = -\frac{36}{\pi}, \quad b_2 = \frac{40}{2\pi}, \quad b_3 = -\frac{36}{3\pi}, \quad b_4 = \frac{40}{4\pi}, \quad \ldots.$$

Plug the b_n back into (11) to get

$$u(x,t) = \frac{20}{\pi}x - \frac{36}{\pi}e^{-t}\sin x + \frac{40}{2\pi}e^{-4t}\sin 2x - \frac{36}{3\pi}e^{-9t}\sin 3x + \frac{40}{4\pi}e^{-16t}\sin 4x - \cdots.$$

25.6. **Insulated ends.**

Problem 25.6. Consider the same insulated uniform metal rod as before ($\alpha = 1$, length π), but now assume that the ends are insulated too (instead of exposed and held in ice), and that the initial temperature is given by $u(x,0) = x$ for $x \in (0,\pi)$. Now what is $u(x,t)$?

Solution: As usual, we temporarily forget the initial condition, and use it only at the end.

"Insulated ends" means that there is zero heat flow through the ends, so the heat flux density function $q \propto -\frac{\partial u}{\partial x}$ is 0 when $x = 0$ or $x = \pi$. In other words, "insulated ends" means that the boundary conditions are

$$\boxed{\frac{\partial u}{\partial x}(0,t) = 0, \qquad \frac{\partial u}{\partial x}(\pi,t) = 0} \qquad \text{for all } t > 0, \tag{12}$$

instead of $u(0,t) = 0$ and $u(\pi,t) = 0$. So we need to solve the heat equation

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}$$

with the boundary conditions (12). Separation of variables $u(x,t) = v(x)\,w(t)$ leads to

$$\begin{aligned} v''(x) &= \lambda\, v(x) \qquad \text{with } v'(0) = 0 \text{ and } v'(\pi) = 0 \\ \dot{w}(t) &= \lambda\, w(t) \end{aligned}$$

for a constant λ.

Lecture actually ended here.

Looking at the cases $\lambda > 0$, $\lambda = 0$, $\lambda < 0$ (see the details in a side calculation presented after the rest of this solution), we find that

$$\lambda = -n^2 \quad \text{and} \quad v(x) = \cos nx \text{ (times a scalar)}$$

where n is one of $0, 1, 2, \ldots$ (this time it turns out that $n = 0$ also gives a nonzero function). For each such $v(x)$, the corresponding w is $w(t) = e^{-n^2 t}$ (times a scalar), and the normal mode is

$$u = e^{-n^2 t}\cos nx.$$

The case $n = 0$ is the constant function 1, so the general solution to the PDE with boundary conditions is

$$u(x,t) = \frac{a_0}{2} + a_1 e^{-t}\cos x + a_2 e^{-4t}\cos 2x + a_3 e^{-9t}\cos 3x + \cdots.$$

Finally, we bring back the initial condition: substitute $t = 0$ and use the initial condition on the left to get

$$x = \frac{a_0}{2} + a_1 \cos x + a_2 \cos 2x + a_3 \cos 3x + \cdots$$

for all $x \in (0, \pi)$. The right hand side is a period 2π even function, so extend the left hand side to a period 2π even function $T(x)$, a triangle wave, which is an antiderivative of

$$\mathrm{Sq}(x) = \frac{4}{\pi}\left(\sin x + \frac{\sin 3x}{3} + \frac{\sin 5x}{5} + \cdots\right).$$

Integration gives

$$T(x) = \frac{a_0}{2} - \frac{4}{\pi}\left(\cos x + \frac{\cos 3x}{9} + \frac{\cos 5x}{25} + \cdots\right),$$

and the constant term $a_0/2$ is the average value of $T(x)$, which is $\pi/2$. Thus

$$\begin{aligned} T(x) &= \frac{\pi}{2} - \frac{4}{\pi}\left(\cos x + \frac{\cos 3x}{9} + \frac{\cos 5x}{25} + \cdots\right) \\ u(x,t) &= \frac{\pi}{2} - \frac{4}{\pi}\left(e^{-t}\cos x + e^{-9t}\frac{\cos 3x}{9} + e^{-25t}\frac{\cos 5x}{25} + \cdots\right). \end{aligned}$$

This answer makes physical sense: when the entire bar is insulated, its temperature tends to a constant equal to the average of the initial temperature. □

Here is the boundary value problem whose solution we used above:

Problem 25.7. For each real number λ, find all functions $v(x)$ on $[0, \pi]$ satisfying

$$\begin{aligned} v''(x) &= \lambda\, v(x) && \text{(ODE)} \\ v'(0) &= 0 && \text{(boundary condition)} \\ v'(\pi) &= 0 && \text{(boundary condition).} \end{aligned}$$

Solution: This is a homogeneous linear ODE with characteristic polynomial $r^2 - \lambda$.

Case 1: $\lambda > 0$. Then the general solution is $ae^{\sqrt{\lambda}x} + be^{-\sqrt{\lambda}x}$, and the boundary conditions say

$$\begin{aligned} \sqrt{\lambda}a - \sqrt{\lambda}b &= 0 \\ a\sqrt{\lambda}e^{\sqrt{\lambda}\pi} - b\sqrt{\lambda}e^{-\sqrt{\lambda}\pi} &= 0. \end{aligned}$$

Divide each equation by $\sqrt{\lambda}$. Since $\det \begin{pmatrix} 1 & -1 \\ e^{\sqrt{\lambda}\pi} & -e^{-\sqrt{\lambda}\pi} \end{pmatrix} \neq 0$, the only solution to this linear system is $(a, b) = (0, 0)$. Thus the only solution to the boundary value problem is $v = 0$.

Case 2: $\lambda = 0$. Then the general solution is $a + bx$, and the boundary conditions say

$$b = 0$$
$$b = 0.$$

Thus the solutions to the boundary value problem are the constant functions a.

Case 3: $\lambda < 0$. We can write $\lambda = -\omega^2$ for some $\omega > 0$. Then the general solution is $a \cos \omega x + b \sin \omega x$. The first boundary condition says $b\omega = 0$, so $b = 0$, so $v = a \cos \omega x$. The second boundary condition then says $-a\omega \sin \omega\pi = 0$, which says different things about a, depending on whether ω is an integer:

- If ω is not an integer, then $\sin \omega\pi \neq 0$, so the second condition implies $a = 0$.
- If ω is an integer n, then $\sin \omega\pi = 0$, so a can be anything. In this case, $\lambda = -n^2$ for some positive integer n (positive since $n = \omega > 0$), and $v(x)$ can be $a \cos nx$ for any constant a.

Setting $n = 0$ into the result of case 3 gives the result of case 2, so we combine these cases in the following:

Final answer to Problem 25.7:

- If λ is one of $0, -1, -4, -9, \ldots$, so $\lambda = -n^2$ for some nonnegative integer n, then the solutions are the functions $\boxed{a \cos nx}$ as a varies.
- For all other values of λ, the only solution is $\boxed{0}$. □

Remark 25.8. The kind of boundary conditions we had earlier, specifying the *values* on the boundary, are called Dirichlet boundary conditions. But the kind we have now, specifying the *derivative* values on the boundary, are called Neumann boundary conditions.

April 30

26. Wave equation

The wave equation is a PDE that models light waves, sound waves, waves along a string, etc.

26.1. **Modeling: vibrating string.**

We skipped much of the physics in lecture.

Problem 26.1. Model a vibrating guitar string.

Variables and functions: Define

$$
\begin{aligned}
L &: \text{ length of the string} \\
\rho &: \text{ mass per unit length} \\
T &: \textit{ magnitude} \text{ of the tension force} \\
t &: \text{ time} \\
x &: \text{ position along the string (from 0 to } L) \\
u &: \text{ vertical displacement of a point on the string}
\end{aligned}
$$

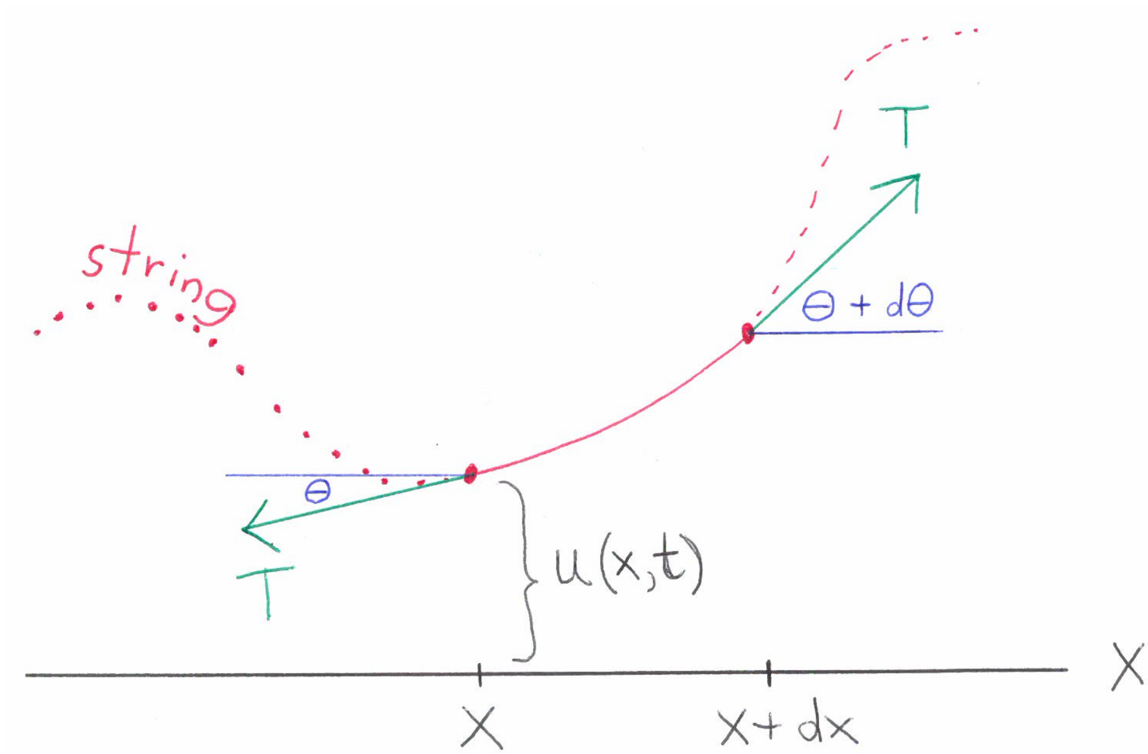
Here

- L, ρ, T are constants;
- t, x are independent variables; and
- $u = u(x,t)$ is a function defined for $x \in [0, L]$ and $t \geq 0$. The vertical displacement is measured relative to the equilibrium position in which the string makes a straight line.

At any given time t, the string is in the shape of the graph of $u(x,t)$ as a function of x.

Assumption: The string is taut, so the vertical displacement of the string is small, and the slope of the string at any point is small.

Consider the piece of string between positions x and $x + dx$. Let θ be the (small) angle formed by the string and the horizontal line at position x, and let $\theta + d\theta$ be the same angle at position $x + dx$.



Newton's second law says that $m\mathbf{a} = \mathbf{F}$. Taking the vertical component of each side gives

$$\begin{aligned} \underbrace{\rho\, dx}_{\text{mass}} \underbrace{\frac{\partial^2 u}{\partial t^2}}_{\text{acceleration}} &= \underbrace{T \sin(\theta + d\theta) - T \sin\theta}_{\text{vertical component of force}} \\ &= T\, d(\sin\theta). \end{aligned}$$

Side calculation:

$$\begin{aligned} d(\sin\theta) &= \cos\theta\, d\theta \\ d(\tan\theta) &= \frac{1}{\cos^2\theta}\, d\theta, \end{aligned}$$

but $\cos\theta = 1 - \frac{\theta^2}{2!} + \cdots \approx 1$, so up to a factor that is very close to 1 we get

$$d(\sin\theta) \approx d(\underbrace{\tan\theta}_{\text{slope of string}}) = d\left(\frac{\partial u}{\partial x}\right).$$

Substituting this in gives

$$\rho\, dx \frac{\partial^2 u}{\partial t^2} \approx T\, d\left(\frac{\partial u}{\partial x}\right).$$

Divide by $\rho\, dx$ to get

$$\begin{aligned}\frac{\partial^2 u}{\partial t^2} &\approx T\rho^{-1}\frac{d\left(\frac{\partial u}{\partial x}\right)}{dx}\\ &\approx T\rho^{-1}\frac{\partial^2 u}{\partial x^2}.\end{aligned}$$

If we define a new constant $c := \sqrt{T\rho^{-1}}$, then this becomes the

wave equation: $$\boxed{\frac{\partial^2 u}{\partial t^2} = c^2\,\frac{\partial^2 u}{\partial x^2}}.$$

This makes sense intuitively, since at places where the graph of the string is concave up $\left(\frac{\partial^2 u}{\partial x^2} > 0\right)$ the tension pulling on both sides should combine to produce an upward force, and hence an upward acceleration.

Comparing units of both sides of the wave equation shows that the units for c are m/s. The physical meaning of c as a velocity will be explained later.

The ends of a guitar string are fixed, so we have boundary conditions

$$\begin{aligned} u(0,t) &= 0 \quad \text{for all } t \ge 0\\ u(L,t) &= 0 \quad \text{for all } t \ge 0.\end{aligned}$$

26.2. **Separation of variables in PDEs; normal modes.** For simplicity, suppose that $c = 1$ and $L = \pi$. So now we are solving the PDE with boundary conditions

$$\begin{aligned}\frac{\partial^2 u}{\partial t^2} &= \frac{\partial^2 u}{\partial x^2}\\ u(0,t) &= 0\\ u(\pi,t) &= 0.\end{aligned}$$

As with the heat equation, we try separation of variables. In other words, try to find normal modes of the form

$$u(x,t) = v(x)w(t),$$

for some nonzero functions $v(x)$ and $w(t)$. Substituting this into the PDE gives

$$\begin{aligned} v(x)\ddot{w}(t) &= v''(x)w(t)\\ \frac{\ddot{w}(t)}{w(t)} &= \frac{v''(x)}{v(x)}.\end{aligned}$$

As usual, a function of t can equal a function of x only if both are equal to the same constant, say λ, so this breaks into two ODEs:

$$\ddot{w}(t) = \lambda\, w(t), \qquad v''(x) = \lambda\, v(x).$$

Moreover, the boundary conditions become $v(0) = 0$ and $v(\pi) = 0$.

We already solved the eigenfunction equation $v''(x) = \lambda\, v(x)$ with the boundary conditions $v(0) = 0$ and $v(\pi) = 0$: nonzero solutions exist only when $\lambda = -n^2$ for some positive integer n, and in this case $v = \sin nx$ (times a scalar). What is different this time is that w satisfies a *second*-order ODE

$$\ddot{w}(t) = -n^2\, w(t).$$

The characteristic polynomial is $r^2 + n^2$, which has roots $\pm in$, so

$$w(t) := \cos nt \quad \text{and} \quad w(t) := \sin nt$$

are possibilities (and all the others are linear combinations). Multiplying each by the $v(x)$ with the matching λ gives the normal modes

$$\cos nt \sin nx, \qquad \sin nt \sin nx.$$

Any linear combination

$$\boxed{u(x,t) = \sum_{n\geq 1} a_n \cos nt \sin nx + \sum_{n \geq 1} b_n \sin nt \sin nx}$$

is a solution to the PDE with boundary conditions, and this turns out to be the general solution.

26.3. **Initial conditions.** To specify a unique solution, give two initial conditions: not only the initial position $u(x,0)$, but also the initial velocity $\dfrac{\partial u}{\partial t}(x,0)$, at each position of the string. (That *two* initial conditions are needed is related to the fact that the PDE is *second*-order in the t variable.)

For a plucked string, it is reasonable to assume that the initial velocity is 0, so one initial condition is $\dfrac{\partial u}{\partial t}(x,0) = 0$. What condition does this impose on the a_n and b_n? Well, for the general solution above,

$$\begin{aligned} \frac{\partial u}{\partial t} &= \sum_{n\geq 1} -n a_n \sin nt \sin nx + \sum_{n\geq 1} n b_n \cos nt \sin nx \\ \frac{\partial u}{\partial t}(x,0) &= \sum_{n \geq 1} n b_n \sin nx, \end{aligned}$$

so the initial condition says that $b_n = 0$ for every n; in other words,

$$u(x,t) = \sum_{n\geq 1} a_n \cos nt \sin nx.$$

If we also knew the initial position $u(x,0)$, we could solve for the a_n by extending to an odd, period 2π function of x and using the Fourier coefficient formula.

26.4. **D'Alembert's solution: traveling waves.** D'Alembert figured out another way to write down solutions, in the case when $u(x,t)$ is defined for all real numbers x instead of just $x \in [0, L]$. Then, for any reasonable function f,

$$u(x,t) := f(x - ct)$$

is a solution to the PDE, as shown by the chain rule:

$$\frac{\partial u}{\partial t} = (-c)f'(x - ct) \qquad\qquad \frac{\partial u}{\partial x} = f'(x - ct)$$
$$\frac{\partial^2 u}{\partial t^2} = (-c)^2 f''(x - ct) \qquad\qquad \frac{\partial^2 u}{\partial x^2} = f''(x - ct),$$

so

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2}.$$

What is the physical meaning of this solution? At $t = 0$, we have $u(x, 0) = f(x)$, so $f(x)$ is the initial position. For any number t, the position of the wave at time t is the graph of $f(x - ct)$, which is the graph of f shifted ct units to the right. Thus the wave travels at constant speed c to the right, maintaining its shape.

The function $u(x,t) := g(x + ct)$ (for any reasonable function $g(x)$) is a solution too, a wave moving to the left. It turns out that the general solution is a superposition

$$\boxed{u(x,t) = f(x - ct) + g(x + ct)}.$$

There is a tiny bit of redundancy: one can add a constant to f and subtract the same constant from g without changing u.

Try the "Wave equation" mathlet

`http://mathlets.org/mathlets/wave-equation/`

Problem 26.2. Suppose that $c = 1$, that the initial position is $I(x)$, and that the initial velocity is 0. What does the wave look like?

Solution: The initial conditions $u(x,0) = I(x)$ and $\frac{\partial u}{\partial t}(x, 0) = 0$ become (after dividing the second one by c)

$$f(x) + g(x) = I(x)$$
$$-f'(x) + g'(x) = 0.$$

The second equation says that $g(x) = f(x) + C$ for some constant C; equivalently, $g(x) - C/2 = f(x) + C/2$. If we replace $f(x)$ by $f(x) + C/2$ and replace $g(x)$ by $g(x) - C/2$, then the new f and g produce the same sum $f(x - t) + g(x + t)$ as before, but the new functions are now equal, $f(x) = g(x)$, instead of differing by a constant. For these new f and g, the first equation yields $f(x) = I(x)/2$ and $g(x) = I(x)/2$. So the wave

$$\boxed{u(x,t) = I(x-t)/2 + I(x+t)/2}$$

consists of two equal waveforms, one traveling to the right and one traveling to the left. $\square$

26.5. **Wave fronts.** Define the step function

$$s(x) := \begin{cases} 1, & \text{if } x < 0 \\ 0, & \text{if } x > 0, \end{cases}$$

and consider the solution $u(x,t) = s(x-t)$ to the wave equation with $c = 1$. This is a "cliff-shaped" wave traveling to the right. (You would be right to complain that this function is not differentiable and therefore cannot satisfy the PDE in the usual sense, but you can imagine replacing $s(x)$ with a smooth approximation, a function with very steep slope. The smooth approximation also makes more sense physically: a physical wave would not actually have a jump discontinuity.)

Another way to plot the behavior is to use a space-time diagram, in a plane with axes x (space) and t (time). (Usually one draws only the part with $t \geq 0$.) Divide the (x,t)-plane into regions according to the value of u. The boundary between the regions is called the wave front.

In the example above, $u(x,t) = 1$ for points to the left of the line $x - t = 0$, and $u(x,t) = 0$ for points to the right of the line $x - t = 0$. So the wave front is the line $x - t = 0$.

A different example:

Flashcard question: Suppose that the initial position is $s(x)$, but the initial velocity is 0 (and still $c = 1$). Into how many regions is the $t \geq 0$ part of the space-time diagram divided?

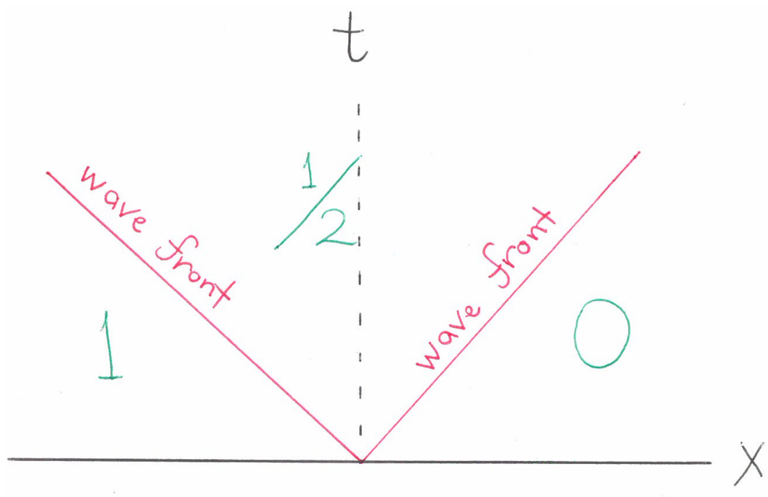
Answer: 3. According to the previous problem,

$$u(x,t) = s(x-t)/2 + s(x+t)/2.$$

Consider $t \geq 0$.

- If $x < -t$, then $u(x,t) = 1/2 + 1/2 = 1$.
- If $-t < x < t$, then $u(x,t) = 1/2 + 0 = 1/2$.
- If $x > t$, then $u(x,t) = 0 + 0 = 0$.

So the upper half of the plane is divided by a V-shaped wave front (the graph of $|x|$) into three regions, with values 1 on the left, 1/2 in the middle, and 0 on the right. $\square$



Lecture actually ended here.

Remark 26.3. We have talked about waves moving in one space dimension, but waves exist in higher dimensions too.

- In one dimension, a disturbance creates wave fronts moving to the left and right, and the space-time diagram of the wave front is shaped like a V, as we just saw.
- In two dimensions, the disturbance caused by a pebble dropped in a still pond creates a circular wave front that moves outward in all directions. The space-time diagram of this wave front is shaped like an ice cream cone (without the ice cream).
- In three dimensions, the wave front created by a disturbance at a point is an expanding sphere.

26.6. **Real-life waves.** In real life, there is always damping. This introduces a new term into the wave equation:

$$\text{damped wave equation:} \qquad \boxed{\frac{\partial^2 u}{\partial t^2} + b\frac{\partial u}{\partial t} = c^2\,\frac{\partial^2 u}{\partial x^2}}.$$

Separation of variables still works, but in each normal mode, the $w(t)$ is a damped sinusoid involving a factor $e^{-bt/2}$ (in the underdamped case).

May 2

27. Graphical methods

The final part of this course concerns *nonlinear* DEs. The sad fact is that we can hardly ever find formulas for the solutions to nonlinear DEs. Instead we try to understand the qualitative behavior of solutions, or use approximations.

27.1. **Solution curves.** We are going to consider nonlinear ODEs

$$\dot{y} = f(t, y)$$

where f is a given function and we are to solve for the unknown function $y(t)$. A solution curve (or integral curve) is the graph of one such solution in the (t, y)-plane.

Problem 27.1. Draw the solution curves to $\dot{y} = y^2$. (This is the special case $f(t, y) := y^2$.)

Solution: Even though this DE is nonlinear, it can be solved exactly by separation of variables:

$$\begin{aligned} \frac{dy}{dt} &= y^2 \\ y^{-2}\,dy &= dt \qquad \text{Oops} \\ \int y^{-2}\,dy &= \int dt \\ \frac{y^{-1}}{-1} &= t + c \qquad \text{(for some constant } c\text{)} \\ y &= -\frac{1}{t+c} \qquad \text{(for some constant } c\text{).} \end{aligned}$$

When $c = 0$, the formula describes the hyperbola $ty = -1$, which consists of two solution curves (one defined for $t < 0$ and one defined for $t > 0$). For other values of c, the solution curves are the same half-hyperbolas except shifted c units to the left.

Oops: We divided by y^2, which is not valid if $y = 0$. The constant function $y = 0$ is a solution too. So in addition to the half-hyperbolas above, there is one more solution curve: the t-axis. $\square$

Solution curves are graphs of functions, so they must satisfy the vertical line test (at most one point on each vertical line).

Problem 27.2. Consider the solution to $\dot{y} = y^2$ satisfying the initial condition $y(0) = 1$. Is there a solution $y(t)$ defined for all real numbers t?

Solution: If this were a *linear* ODE, then the existence and uniqueness theorem would guarantee a YES answer.

But here the answer is NO, as we'll now explain. Setting $t = 0$ in the general solution above and using the initial condition leads to

$$\begin{aligned} 1 &= -\frac{1}{0+c} \\ c &= -1, \end{aligned}$$

so

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

As t increases towards 1, the value of $y(t)$ tends to $+\infty$, so one says that the solution blows up in finite time. It is impossible to extend $y(t)$ to a solution defined and continuous at $t = 1$ or beyond; the largest open interval containing the starting point 0 on which a solution exists is $(-\infty, 1)$. $\square$

27.2. **Existence and uniqueness.** For nonlinear ODEs, there is still an existence and uniqueness theorem, but the solutions it provides are not necessarily defined for all t.

Existence and uniqueness theorem for a nonlinear ODE. *Consider a nonlinear ODE*

$$\dot{y} = f(t, y) \quad \text{with initial condition } y(t_0) = y_0$$

Assume that f and $\frac{\partial f}{\partial y}$ are continuous on the entire (t, y)-plane. Then

(a) *There exists a solution $y(t)$ defined on some open interval containing t_0. The largest such open interval is called the domain of validity of the solution; call it I.*
(b) *The solution on I is unique.*
(c) *If $I = (a, b)$ and b is finite, then as $t \to b^-$, the function $y(t)$ becomes unbounded. (A similar statement holds as t approaches the* left *endpoint of I.)*

Remark 27.3. If there are points in the (t, y)-plane where f or $\frac{\partial f}{\partial y}$ fails to be continuous, changes are needed. Let U be the largest open region in the (t, y)-plane on which f and $\frac{\partial f}{\partial y}$ are continuous.

(a) If $(t_0, y_0) \in U$, then a solution exists on some open interval containing t_0. There is a largest such interval I such that the solution curve stays in U.
(b) On that I, the solution is unique.
(c) If $I = (a, b)$ and b is finite, then as $t \to b^-$, either $y(t)$ becomes unbounded or else $(t, y(t))$ reaches points arbitrarily close to the boundary of U.

What does the theorem mean graphically?

(a) Through each point (t_0, y_0) there is *exactly one* solution curve. If you ever draw two solution curves that cross or even touch at a point, you are in big trouble! (Exception: They *might* meet at a point where f or $\frac{\partial f}{\partial y}$ fails to be continuous, because the theorem does not apply there.)
(b) The solution curve keeps going (both to left and right) unless it becomes unbounded or approaches a point outside U.

To see these principles in action, try the "Solution Targets" mathlet

`http://mathlets.org/mathlets/solution-targets/`

Here is an example where the hypotheses of the theorem fail.

Problem 27.4. Draw the solution curves for $t\dot{y} = 2y$.

Solution: Solving for $\dot{y}$ leads to $\dot{y} = \dfrac{2y}{t}$, and the right hand side is undefined when $t = 0$, so things might go wrong along the vertical line $t = 0$, and in fact they *do* go wrong.

Solve the ODE by separation of variables:

$$\begin{aligned}
t\frac{dy}{dt} &= 2y \\
\frac{dy}{y} &= \frac{2\,dt}{t} \qquad \text{(assuming that } t \text{ and } y \text{ are not } 0) \\
\int \frac{dy}{y} &= \int \frac{2\,dt}{t} \\
\ln|y| &= 2\ln|t| + C \qquad \text{(for some constant } C) \\
y &= \pm e^{2\ln|t|+C} \\
y &= \pm |t|^2 e^C \\
y &= ct^2,
\end{aligned}$$

where $c := \pm e^C$, which can be any nonzero real number. To bring back the solution $y = 0$, allow $c = 0$ too. The solution curves on the interval where $t < 0$ or on the interval where $t > 0$ are the curves $y = cx^2$: half-parabolas and halves of the horizontal line $y = 0$.

Weird behavior happens along $t = 0$ (where the theorem does not apply):

- Through $(0, 0)$, there are *infinitely many* solution curves.
- Through $(0, 1)$, there is *no* solution curve. (Same for $(0, b)$ for any nonzero b.)

But the rest of the plane is covered with good solution curves, one through each point, none touching or crossing the others.

27.3. **Slope field.** We are now going to introduce concepts to help with drawing solution curves to an ODE $\dot{y} = f(t, y)$. The slope field is a diagram in which at each point (t, y), you draw a short segment whose slope is the *value* $f(t, y)$.

Problem 27.5. Sketch the slope field for $\dot{y} = y^2 - t$.

Solution: Let $f(t, y) := y^2 - t$. Then

$$\begin{gathered}
f(1,2) = 3, \text{ so at } (1,2) \text{ draw a short segment of slope } 3; \\
f(0,0) = 0, \text{ so at } (0,0) \text{ draw a short segment of slope } 0; \\
f(1,0) = -1, \text{ so at } (1,0) \text{ draw a short segment of slope } -1; \\
f(0,1) = 1, \text{ so at } (0,1) \text{ draw a short segment of slope } 1; \\
\vdots
\end{gathered}$$

The diagram of all these short segments is the slope field. □

A computer can do the job more quickly: try the "Isoclines" mathlet

`http://mathlets.org/mathlets/isoclines/`

Warning: The slope field is not the same as the graph of f: in drawing the graph of f, the value of f is used as a height, but in drawing a slope field, the value of f is used as the slope of a little segment.

Why draw a slope field? The ODE is telling us that the slope of the solution curve at each point is the value of $f(t, y)$, so the short segment there is, to first approximation, a little piece of the solution curve. To get an entire solution curve, follow the segments!

27.4. **Isoclines.** Even with the computer display, it's hard to tell what is going on. To understand better, we introduce a new concept: If m is a number, the m-isocline is the set of points in the (t, y)-plane such that the solution curve through that point has slope m. (Isocline means "same incline", or "same slope".)

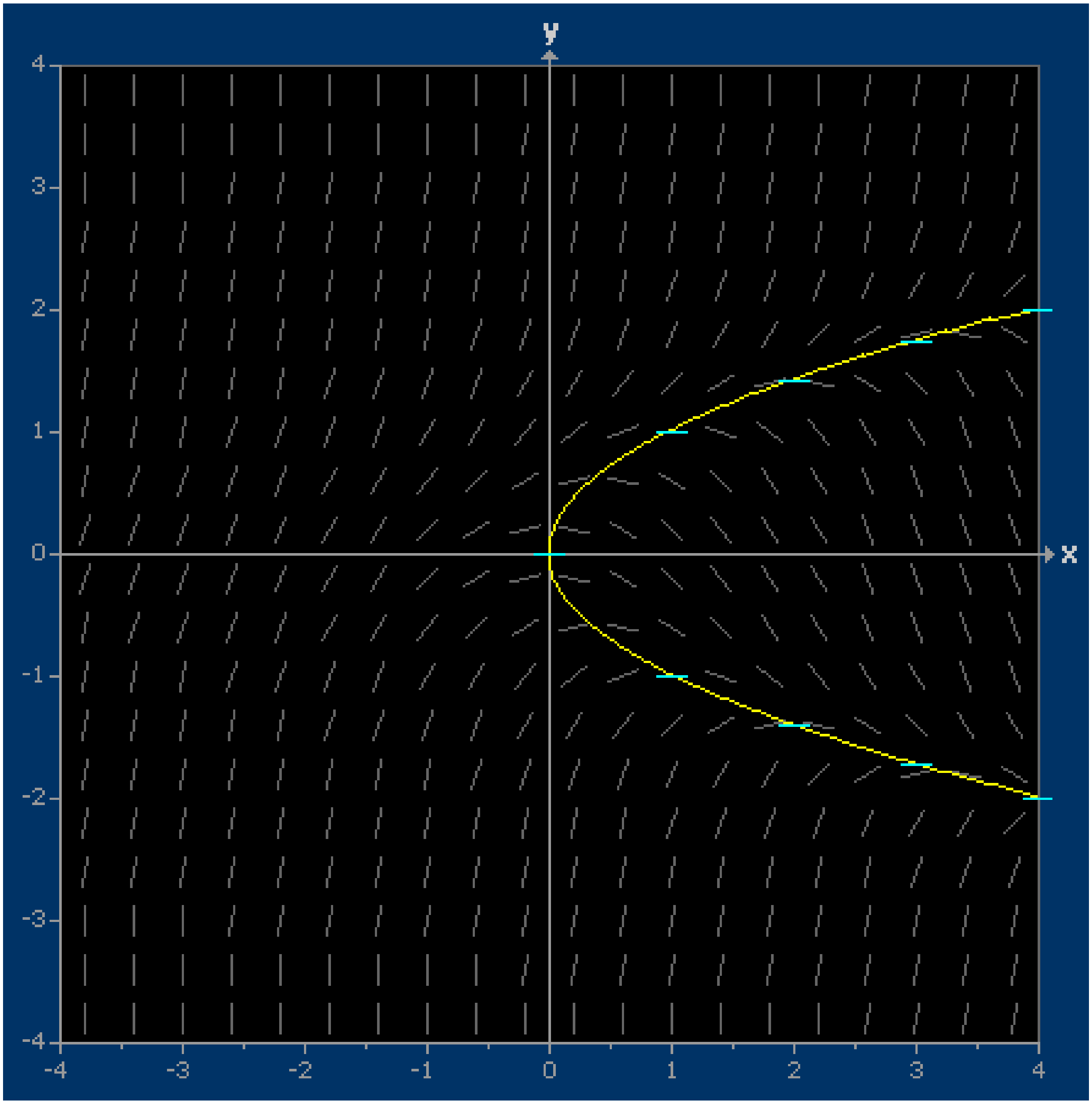
Question 27.6. What is the equation for the m-isocline?

Solution: The ODE says that the slope of the solution curve through a point (t, y) is $f(t, y)$, so the equation of the m-isocline is $f(t, y) = m$. □

Finding the isoclines will help organize the slope field. The 0-isocline is especially helpful.

Problem 27.7. For $\dot{y} = y^2 - t$, what is the 0-isocline?

Solution: Here $f(t, y) := y^2 - t$, so the 0-isocline is the curve $y^2 - t = 0$, which is a parabola concave to the right. At every point of this parabola, the slope of the solution curve is 0. □



Problem 27.8. For $\dot{y} = y^2 - t$, where are all the points at which the slope of the solution curve is positive?

Solution: This will be the region in which $f(t, y) > 0$. The 0-isocline $f(t, y) = 0$ divides the plane into regions, and $f(t, y)$ has constant sign on each region. To test the sign, just check one point in each region. For $f(t, y) := y^2 - t$, we have $f(t, y) > 0$ in the region to the left of the parabola (since $f(0, 1) > 0$), and $f(t, y) < 0$ in the region to the right of the parabola (since $f(1, 0) < 0$). On the left region, solution curves slope upward; on the right region, solution curves slope downward. The answer is: in the region to the left of the parabola. $\square$

The solution curve through $(0,0)$ increases for $t < 0$ and decreases for $t > 0$, so it reaches its maximum at $(0,0)$. How did we know that the solution for $t > 0$ does not cross the lower part of the parabola, $y = -\sqrt{t}$, back into the upward sloping region? Answer: If it crossed somewhere, its slope would have to be negative there, but the DE says that the slope is 0 everywhere along $y = -\sqrt{t}$. Thus $y = -\sqrt{t}$ acts as a fence that solution curves already inside the parabola cannot cross.

27.5. **Example: The logistic equation.** The simplest model for population $x(t)$ is the ODE $\dot{x} = ax$ for a positive growth constant a: the rate of population growth is proportional to the current population. But realistically, if $x(t)$ gets too large, then because of competition for food and space, the population will grow less quickly. In a better model, the growth rate a would become smaller as the population grows. In the simplest model of this type, the growth rate is a linearly decreasing function of the population, $a - bx$, where b is another positive constant; then the DE is $\dot{x} = (a - bx)x$ instead of $\dot{x} = ax$. In other words, the new DE is

$$\dot{x} = ax - bx^2,$$

where a and b are positive constants. This is a nonlinear ODE, called the logistic equation.

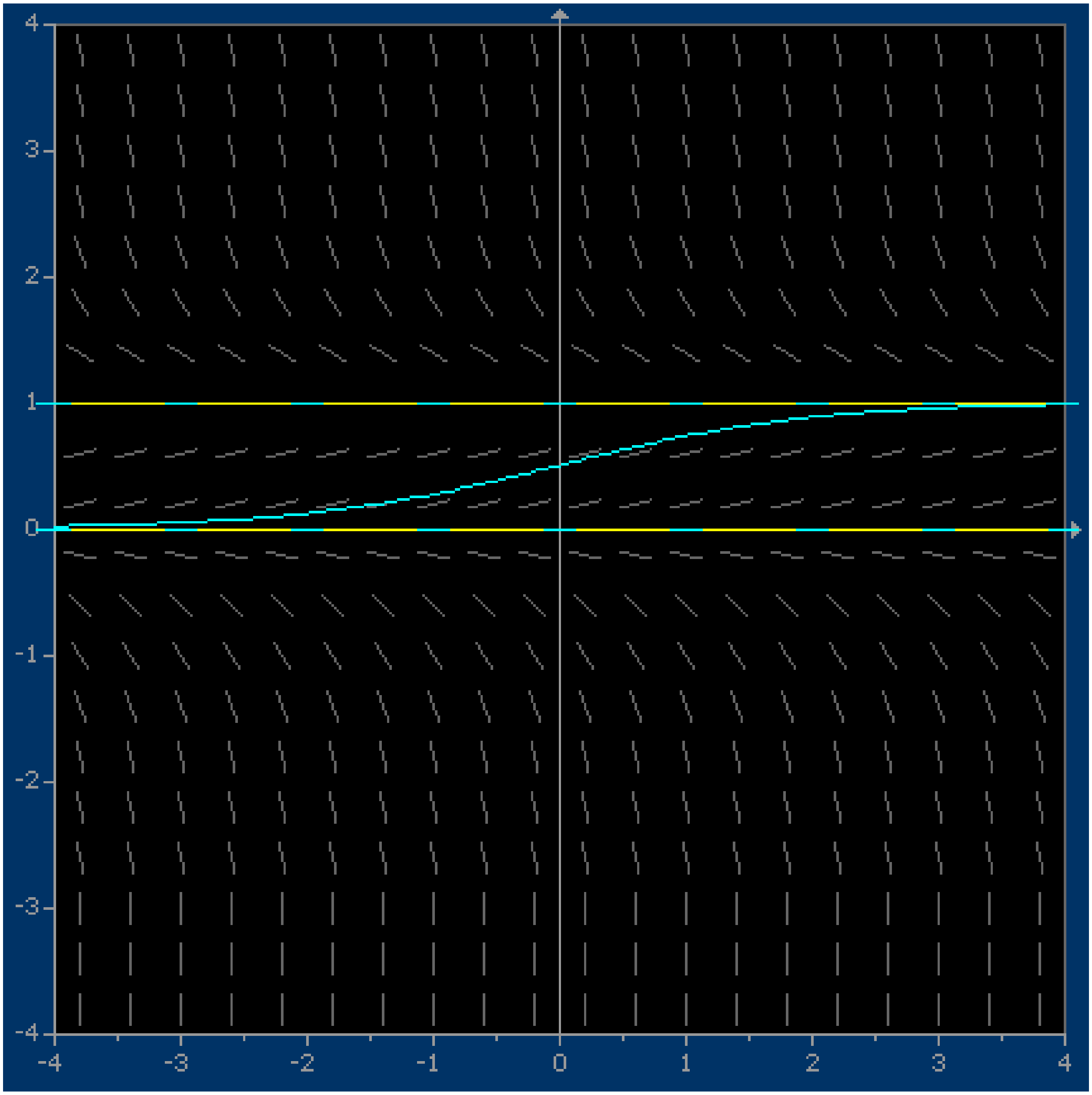
Let's consider the simplest case, in which $a = 1$ and $b = 1$:

Problem 27.9. Draw the solution curves for $\dot{x} = x - x^2$ in the (t,x)-plane.

Solution: The first step is always to find the 0-isocline. Here $f(t,x) := x - x^2$, so the 0-isocline is $x - x^2 = 0$, which consists of the horizontal lines $x = 0$ and $x = 1$. Each of these two lines has slope 0, matching the slope specified for the solution curve at each point of the line, so each line itself is a solution curve! (Warning: This is not typical. An isocline is not usually a solution curve.)

The 0-isocline divides the (t,x)-plane into three regions: in the horizontal strip $0 < x < 1$, we have $f(t,x) = x - x^2 = x(1-x) > 0$, so solutions slope upward. In the regions below and above, solutions slope downward.

The diagram below shows the slope field (gray segments), the 0-isocline (yellow line), and the solution curve with initial condition $x(0) = 1/2$ (blue curve).



May 4

28. Autonomous equations

An autonomous equation is a differential equation that is time-invariant: $\dot{x} = f(x)$ instead of $\dot{x} = f(x,t)$.

(Why is this called autonomous? In ordinary English, a machine or robot is called autonomous if it operates without human input. A differential equation is called autonomous if the coefficients of the problem are not changed over time, such as might happen if a human adjusted a dial on a machine.)

28.1. **Properties.** For an autonomous equation,

- If $x(t)$ is a solution, then so is $x(t-a)$ for any constant a.
  (Proof: If $x'(t) = f(x(t))$ holds for all t, then it holds also with t replaced by $t-a$, so $x'(t-a) = f(x(t-a))$, and by the chain rule the left hand side is the same as the derivative of $x(t-a)$, so this says that $x(t-a)$ is a solution.)
- Each isocline (in the (t,x)-plane) consists of horizontal lines.
- For the 0-isocline, these horizontal lines are also solution curves, corresponding to constant solutions.

28.2. **Phase line.**

Problem 28.1. Describe the solutions to $\dot{x} = 3x - x^2$.
(This is a special case of the logistic equation $\dot{x} = ax - bx^2$.)

Solution: Let $f(x) := 3x - x^2$. First find the 0-isocline by solving $3x - x^2 = 0$. This leads to $x(3-x) = 0$, so $x = 0$ or $x = 3$. These are horizontal lines. They are also solution curves, corresponding to the constant functions $x(t) = 0$ and $x(t) = 3$.

As in last lecture, the 0-isocline divides the plane into "up" regions and "down" regions. These are the region $x < 0$, the region $0 < x < 3$, and the region $x > 3$. To find out which are up and which are down, test one point in each:

- Since $f(-1) < 0$, the region $x < 0$ is a down region.
- Since $f(1) > 0$, the region $0 < x < 3$ is an up region.
- Since $f(4) < 0$, the region $x > 3$ is a down region.

The phase line is a plot of the x-axis that summarizes this information:

$$-\infty \quad \longleftarrow \quad \underset{\text{unstable}}{0} \quad \longrightarrow \quad \underset{\text{stable}}{3} \quad \longleftarrow \quad +\infty$$

(The labels unstable and stable will be explained later. Sometimes the phase line is drawn vertically instead, with $+\infty$ at the top.)

What happens to solutions as time passes?

- If $x(0) = 0$, then the solution will be $x(t) = 0$ for all t. (We said this already.)
- If $x(0) = 3$, the solution will be $x(t) = 3$ for all t.
- Suppose that the initial condition is that $x(0)$ is a number strictly *between* 0 and 3. Then $x(t)$ will increase. But it will never reach 3, because the solution curve cannot cross or touch the solution curve at height 3. Could it be that $x(t)$ tends to a limit less than 3? No, because then $\dot{x}(t) = 3x - x^2$ would tend to a positive limit, but $\dot{x}(t)$

must tend to 0 as the solution curve levels off. Conclusion: $x(t)$ increases, tending to 3 as $t \to +\infty$ (but never actually reaching 3).

- Similarly, if $x(0) > 3$, then $x(t)$ decreases, tending to 3 without actually reaching 3.
- Finally, if $x(0) < 0$, then $x(t)$ decreases, and $x(t) \to -\infty$ as t grows. (With more work, one could show that it tends to $-\infty$ in finite time.)

Flashcard question: If $x(0)$ is strictly between 0 and 3, what is $\lim_{t\to-\infty} x(t)$?

Answer: To run time backwards, reverse the arrows in the phase line. As $t \to -\infty$, we have $x(t) \to 0$.

Warning: Using a phase line makes sense only if the DE is autonomous!

Try the "Phase Lines" mathlet

`http://mathlets.org/mathlets/phase-lines/`

28.3. **Stability.** In general, for $\dot{x} = f(x)$, the real x-values such that $f(x) = 0$ are called critical points. Warning: Only *real* numbers can qualify as critical points.

A critical point is called

- stable if solutions starting near it move towards it,
- unstable if solutions starting near it move away from it,
- semistable if the behavior depends on *which side* of the critical point the solution starts.

In the case of the differential equation $\dot{x} = 3x - x^2$ studied above, the critical points are 0 and 3. The phase line shows that 0 is unstable, and 3 is stable.

Remark 28.2. An unstable critical point is also called a separatrix because it separates solutions having very different fates.

Example 28.3. For $\dot{x} = 3x - x^2$, a solution starting just *below* 0 tends to $-\infty$, while a solution starting just *above* 0 tends to 3: very different fates! □

To summarize:

Steps for understanding solutions to $\dot{x} = f(x)$ qualitatively:

1. Solve $f(x) = 0$ to find the critical points.
2. Write down

$$-\infty \qquad \text{(critical points in increasing order)} \qquad \infty.$$

Each space in between represents an open interval of x-values.

3. In each interval, choose an x-value and check whether $f(x)$ is positive or negative there to find out whether solutions starting in the interval are increasing or decreasing; draw an arrow to the right or left, accordingly, in the space.
4. Interpretation:
    - Solutions starting at a critical point are constant.
    - Solutions starting elsewhere tend, as t increases, to the limit that the arrow points to. (To run time backwards, to see the behavior of the solution as t decreases, reverse the arrows.)

    Usually $x(t)$ is defined for all $t \in \mathbb{R}$, and the source and target of the arrow indicate $\lim_{t\to-\infty} x(t)$ and $\lim_{t\to\infty} x(t)$. But if the target is ∞ or $-\infty$, then there might be a *finite* time T_{end} such that $x(t) \to \pm\infty$ as t increases towards T_{end}; in this case, $x(t)$ is undefined for $t \geq T_{\text{end}}$ (the solution blows up in finite time). Similarly, if the source is ∞ or $-\infty$, there might be a finite time T_{start} such that $x(t) \to \pm\infty$ as t decreases towards T_{start}, and $x(t)$ is undefined for $t \leq T_{\text{start}}$.

28.4. **Harvesting models and bifurcation diagrams.**

Problem 28.4. Frogs grow in a pond according to a logistic equation with growth constant 3 month^{-1}. The population reaches an equilibrium of 3000 frogs, but then the frogs are harvested at a constant rate. Model the population of frogs.

Variables and functions:

$$\begin{aligned} t &: \text{ time (months)} \\ x &: \text{ size of population (kilofrogs)} \\ h &: \text{ harvest rate (kilofrogs/month)} \end{aligned}$$

Equation: *Without* harvesting,

$$\dot{x} = 3x - bx^2$$

for some constant $b > 0$. Since the population settles at $x = 3$ (three thousand frogs), $0 = \dot{x} = 3x - bx^2$ at $x = 3$; thus $b = 1$.

With harvesting, $x(0) = 3$ and

$$\boxed{\dot{x} = 3x - x^2 - h}. \quad \square$$

This is an *infinite family* of autonomous equations, one for each value of h, and *each has its own phase line*. If in the (h, x)-plane, we draw each phase line vertically in the vertical line corresponding to a given value of h, and plot the critical points for each h, then we get a diagram called a bifurcation diagram. In this diagram, color the critical points according to whether they are stable, unstable, or semistable.

Example 28.5. If $h = 2$, then $\dot{x} = 3x - x^2 - 2$. Since $3x - x^2 - 2 = -(x-2)(x-1)$, the critical points are 1 and 2, and the phase line is

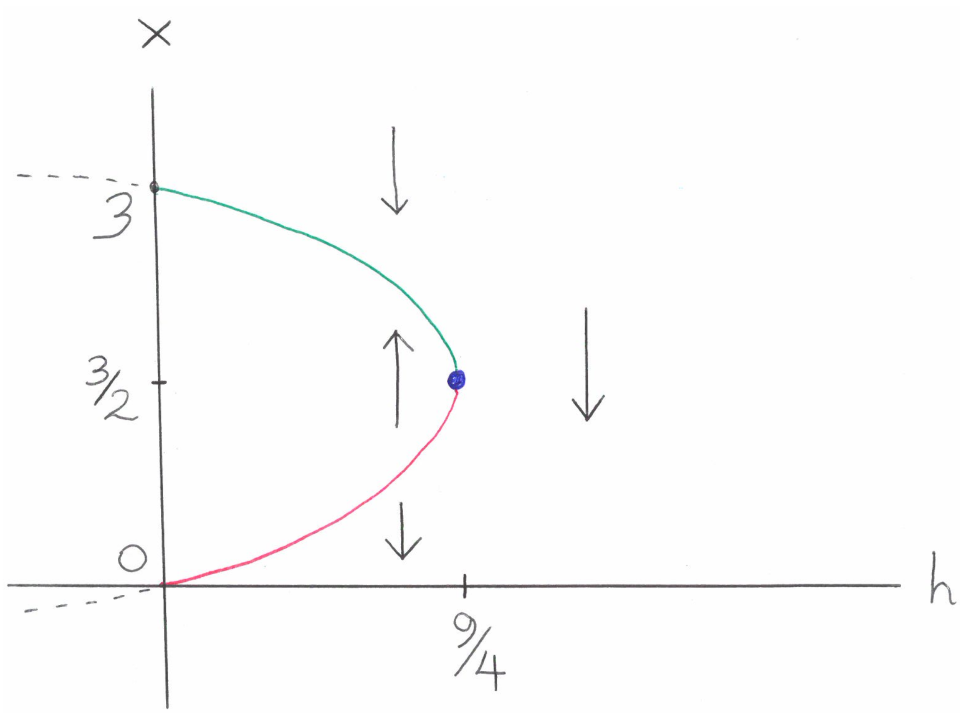
$$-\infty \quad \longleftarrow \quad \underset{\text{unstable}}{1} \quad \longrightarrow \quad \underset{\text{stable}}{2} \quad \longleftarrow \quad +\infty. \quad \square$$

For each other value of h, the critical points are the real roots of $3x - x^2 - h$. We *could* use the quadratic formula to find these roots

$$r_1(h) = \frac{3 - \sqrt{9-4h}}{2}, \quad r_2(h) = \frac{3 + \sqrt{9-4h}}{2}$$

(assuming that $9 - 4h \geq 0$), and then graph both functions to get the bifurcation diagram.

But we don't need to do this! The equation $3x - x^2 - h = 0$ is the same as $h = 3x - x^2$. The graph of this in the (x, h)-plane is a downward parabola; to get the bifurcation diagram in the (h, x)-plane, interchange the axes by reflecting in the line $h = x$.



Checking one point inside the parabola (like $(h, x) = (0, 1)$) shows that $3x - x^2 - h$ is positive there, and similarly $3x - x^2 - h$ is negative outside the parabola. Thus the upper branch $x = r_2(h)$ consists of stable critical points, and the lower branch $x = r_1(h)$ consists of unstable critical points, at least when $9 - 4h > 0$.

Question 28.6. What happens when $9 - 4h = 0$, i.e., when $h = 9/4$?

Answer: Then $3x - x^2 - 9/4 = -(x - 3/2)^2$, so the phase line is

$$-\infty \quad \longleftarrow \quad \underset{\text{semistable}}{3/2} \quad \longleftarrow \quad +\infty.$$

Does this mean that a solution $x(t)$ can go all the way from $+\infty$ through $3/2$ to $-\infty$? No, because it can't cross the constant solution $x = 3/2$. Instead there are three possible behaviors:

- If $x(0) > 3/2$, then $x(t) \to 3/2$ as $t \to +\infty$.
- If $x(0) = 3/2$, then $x(t) = 3/2$ for all t.
- If $x(0) < 3/2$, then $x(t)$ tends to $-\infty$ (we interpret this as a *population crash*: the frog population reaches 0 in finite time; the part of the trajectory with $x < 0$ is not part of the population model).

May 7

Problem 28.7. What is the maximum sustainable harvest rate?

(*Sustainable* means that the harvesting does not cause the population to crash to 0, but that instead $\lim_{t\to+\infty} x(t)$ is positive, so that the harvesting can continue indefinitely.)

Solution: $h = 9/4$, i.e., 2250 frogs/month. Why?

- For $h < 9/4$, the phase line is

$$-\infty \quad \longleftarrow \quad \underset{\text{unstable}}{r_1(h)} \quad \longrightarrow \quad \underset{\text{stable}}{r_2(h)} \quad \longleftarrow \quad +\infty$$

  and $x(0) = 3 > r_2(h)$, so $x(t) \to r_2(h)$.
- For $h = 9/4$, the phase line is

$$-\infty \quad \longleftarrow \quad \underset{\text{semistable}}{3/2} \quad \longleftarrow \quad +\infty$$

  and $x(0) = 3 > 3/2$, so $x(t) \to 3/2$.
- For $h > 9/4$, the phase line is

$$-\infty \quad \longleftarrow \quad +\infty$$

  so a population crash is inevitable (overharvesting). $\square$

Remark 28.8. Harvesting at exactly the maximum rate is a little dangerous, however, because if after a while x becomes very close to $3/2$, and a little kid comes along and takes one more frog out of the pond, the whole frog population will crash!

One student suggested the following, which seems appropriate:

`http://www.poemhunter.com/poem/death-of-a-naturalist/`

28.5. **Linear approximation in 1D.** Let's return to $\dot{x} = \underbrace{3x - x^2}_{f(x)}$, and study the solutions with $0 < x < 3$. (This particular ODE could be solved exactly, but for more complicated ODEs one cannot hope to find an exact formula, so we'll want to illustrate the general method.)

Numerical results of a computer simulation can give us a clear picture of the part of the solutions in the range $0.1 < x < 2.9$, but not enough detail when x is near the critical points 0 and 3 (as happens as $t \to -\infty$ or $t \to +\infty$, respectively). Linear approximations will show us what happens near the critical points.

- Consider $x \approx 0$, which is of interest when $t \to -\infty$, that is, when studying the origins of the population. Then
$$\dot{x} = 3x - x^2 \approx 3x.$$
Thus we can expect
$$x \approx ae^{3t}$$
for some constant a. That is, when the population is getting started, solutions to the logistic equation obey approximately exponential growth, until the competition for food or space implicit in the $-x^2$ term becomes too large to ignore.
- Consider $x \approx 3$, which is of interest when $t \to +\infty$. To measure deviations from 3, define $X := x - 3 \approx 0$, so $x = 3 + X$. The best linear approximation to $f(x)$ for $x \approx 3$ is
$$\begin{aligned} f(x) &\approx f(3) + f'(3)(x-3) \\ &= 0 + (-3)(x-3) \\ &= -3X \end{aligned}$$
so
$$\dot{X} = \dot{x} = f(x) \approx -3X.$$
Thus we can expect
$$X \approx be^{-3t}$$
and
$$x = 3 + X \approx 3 + be^{-3t}$$
for some constant b, as $t \to +\infty$. (Since we are looking at solutions with $0 < x(t) < 3$, we must have $b < 0$.)

The "big picture" combines numerical results for $0.1 < x < 2.9$ with linear approximations near 0 and 3.

29. Autonomous systems

Now we study a system of two autonomous equations in two unknown functions $x(t)$ and $y(t)$:

$$\begin{aligned}\dot{x} &= f(x,y)\\ \dot{y} &= g(x,y)\end{aligned}$$

for some functions f and g that do not depend on t.

Example 29.1. If $x(t)$ is deer population (in thousands), and $y(t)$ is wolf population (in hundreds), then the system

$$\begin{aligned}\dot{x} &= 3x - x^2 - xy\\ \dot{y} &= y - y^2 + xy\end{aligned}$$

is a reasonable model: each population obeys the logistic equation, except that there is an adjustment depending on xy, which is proportional to the number of deer-wolf encounters. Such encounters are bad for the deer, but good for the wolves!

29.1. **Phase plane.** Solution curves would now exist in 3-dimensional (t, x, y)-space, so they are hard to draw. Instead, forget t, and draw the motion in the (x, y) phase plane. At each point (x, y), the system says that the velocity vector there is the value of $\begin{pmatrix} f(x,y) \\ g(x,y) \end{pmatrix}$.

Problem 29.2. In the deer-wolf example above, what is the velocity vector at $(x, y) = (3, 2)$?

Solution:

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} 9-9-6 \\ 2-4+6 \end{pmatrix} = \begin{pmatrix} -6 \\ 4 \end{pmatrix}.$$

Draw this velocity vector with its foot at $(3, 2)$. $\square$

The velocity vectors at all points together make a vector field. If you draw them all to scale, you will wreck your picture! Mostly what we care about is the direction, so it is OK to shorten them. Or better yet, don't draw them at all, and instead just draw arrowheads along the phase plane trajectories in the direction of motion.

There is an existence and uniqueness theorem for systems of nonlinear ODEs similar to that for a single ODEs. For an autonomous system it implies that there is a unique trajectory through each point (in a region in which the partial derivatives of f and g are continuous):

Trajectories never cross or touch!

(But see the "exception" in Remark 29.4.)

29.2. **Critical points.** A critical point for an autonomous system is a point in the (x, y)-plane where the velocity vector is $\mathbf{0}$. To find all the critical points, solve

$$f(x, y) = 0$$
$$g(x, y) = 0.$$

Problem 29.3. Find the critical points for the deer-wolf system.

Solution: We need to solve

$$3x - x^2 - xy = 0$$
$$y - y^2 + xy = 0.$$

Each polynomial factors, so we get

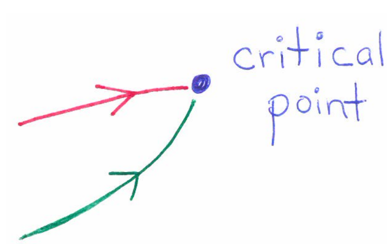
$$x = 0 \quad \text{or} \quad 3 - x - y = 0$$
$$y = 0 \quad \text{or} \quad 1 - y + x = 0.$$

Intersecting each of the first two lines with each of the last two lines gives the four points

$$(0, 0), \quad (0, 1), \quad (3, 0), \quad (1, 2). \quad \square$$

Critical points are also called stationary points, because each such point corresponds to a solution in which $x(t)$ and $y(t)$ are constant.

Remark 29.4. We said earlier that trajectories never cross. While it is true that no two trajectories can have a point in common, it *is* possible for two trajectories to have the same limit as $t \to +\infty$ or $t \to -\infty$, so they can *appear* to come together. For a trajectory to have a finite limiting position, the velocity must tend to 0, so the limiting position must be a critical point.



Conclusion: It is only at a critical point that trajectories can appear to come together.

29.3. **Linear approximation in 2D.**

If you remember nothing else from 18.01, remember this:

> If a problem you are trying to solve is too difficult because it involves a nonlinear function $f(x)$, use the best linear approximation near the most relevant x-value a: that approximation is
>
> $$f(a) + f'(a)\,(x - a)$$

since this linear polynomial has the same value and same derivative at a as $f(x)$.

If you remember nothing else from 18.02, remember this:

If a problem you are trying to solve is too difficult because it involves a nonlinear function $f(x, y)$, use the best linear approximation near the most relevant point (a, b): that approximation is

$$f(a,b) + \frac{\partial f}{\partial x}(a,b)\,(x-a) + \frac{\partial f}{\partial y}(a,b)\,(y-b)$$

since this linear polynomial has the same value and same partial derivatives at (a, b) as $f(x)$.

(We used green for numbers here.)

29.3.1. *Warm-up: linear approximation at* $(0, 0)$. To understand the behavior of the deer-wolf system near $(0, 0)$, use

$$\begin{aligned} \dot{x} &= 3x - x^2 - xy \approx 3x \\ \dot{y} &= \;\; y - y^2 + xy \approx y. \end{aligned}$$

In matrix form,

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} \approx \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

The eigenvalues are 3 and 1, so this describes a repelling node at $(0, 0)$.

29.3.2. *Linear approximation via change of coordinates (method 1).* To understand the deer-wolf system near the critical point $(1, 2)$, reduce to the previously solved case of $(0, 0)$ by making the change of variable

$$\begin{aligned} x &= 1 + X \\ y &= 2 + Y \end{aligned}$$

so that $(x, y) = (1, 2)$ is $(X, Y) = (0, 0)$ in the new coordinate system. Then

$$\begin{aligned} \dot{X} &= \dot{x} = 3(1+X) - (1+X)^2 - (1+X)(2+Y) = -X - Y - X^2 - XY \approx -X - Y \\ \dot{Y} &= \dot{y} = \;\; (2+Y) - (2+Y)^2 + (1+X)(2+Y) = 2X - 2Y - Y^2 + XY \approx 2X - 2Y \end{aligned}$$

when (X, Y) is close to $(0, 0)$. In matrix form,

$$\begin{pmatrix} \dot{X} \\ \dot{Y} \end{pmatrix} \approx \begin{pmatrix} -1 & -1 \\ 2 & -2 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}.$$

29.3.3. *Linear approximation via Jacobian matrix (method 2).*

Definition 29.5. The Jacobian matrix of the vector-valued function $\begin{pmatrix} f(x,y) \\ g(x,y) \end{pmatrix}$ is the matrix-valued function

$$J(x,y) := \begin{pmatrix} \partial f/\partial x & \partial f/\partial y \\ \partial g/\partial x & \partial g/\partial y \end{pmatrix}.$$

The Jacobian determinant is the *determinant* of the Jacobian matrix. In 18.02, you learned that the absolute value of the Jacobian determinant is the area scaling factor when doing a change of variable in a double integral.

The Jacobian matrix is also called the derivative of the multivariable function $\begin{pmatrix} f(x,y) \\ g(x,y) \end{pmatrix}$. The function has 2-variable input $\begin{pmatrix} x \\ y \end{pmatrix}$ and 2-variable output $\begin{pmatrix} f(x,y) \\ g(x,y) \end{pmatrix}$; this leads to each value of the Jacobian matrix being a 2×2 matrix.

The best linear approximations to f and g at (a,b) are

$$\begin{aligned} f(x,y) &\approx f(a,b) + \frac{\partial f}{\partial x}(a,b)\,(x-a) + \frac{\partial f}{\partial y}(a,b)\,(y-b) \\ g(x,y) &\approx g(a,b) + \frac{\partial g}{\partial x}(a,b)\,(x-a) + \frac{\partial g}{\partial y}(a,b)\,(y-b). \end{aligned}$$

These can be combined into one equation:

$$\boxed{\begin{pmatrix} f(x,y) \\ g(x,y) \end{pmatrix} \approx \underbrace{\begin{pmatrix} f(a,b) \\ g(a,b) \end{pmatrix}}_{\text{value at } (a,b)} + \underbrace{J(a,b)}_{\text{derivative at } (a,b)} \begin{pmatrix} x-a \\ y-b \end{pmatrix}}.$$

Special case where (a,b) is a critical point for the system: Then $f(a,b) = 0$ and $g(a,b) = 0$, so this linear approximation simplifies to

$$\begin{pmatrix} f(x,y) \\ g(x,y) \end{pmatrix} \approx J(a,b) \begin{pmatrix} x-a \\ y-b \end{pmatrix}.$$

Making the change of variable $X := x - a$ and $Y := y - b$ leads to

$$\begin{pmatrix} \dot{X} \\ \dot{Y} \end{pmatrix} = \begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} f(x,y) \\ g(x,y) \end{pmatrix} \approx J(a,b) \begin{pmatrix} X \\ Y \end{pmatrix}.$$

Conclusion: At a critical point (a,b), if $X := x - a$ and $Y := y - b$, then

$$\boxed{\begin{pmatrix} \dot{X} \\ \dot{Y} \end{pmatrix} \approx J(a,b) \begin{pmatrix} X \\ Y \end{pmatrix}}.$$

Problem 29.6. Find the behavior of the deer-wolf system near the critical point $(1,2)$.

Solution: We have

$$J(x,y) := \begin{pmatrix} \partial f/\partial x & \partial f/\partial y \\ \partial g/\partial x & \partial g/\partial y \end{pmatrix} = \begin{pmatrix} 3-2x-y & -x \\ y & 1-2y+x \end{pmatrix}.$$

Plug in $x = 1$ and $y = 2$ to get

$$J(1,2) = \begin{pmatrix} -1 & -1 \\ 2 & -2 \end{pmatrix}.$$

Thus, if we measure deviations from the critical point by defining $X := x - 1$ and $Y := y - 2$, we have

$$\begin{pmatrix} \dot{X} \\ \dot{Y} \end{pmatrix} \approx \begin{pmatrix} -1 & -1 \\ 2 & -2 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}$$

(the same as what we got using method 1). The matrix has trace -3 and determinant 4, so the characteristic polynomial is $\lambda^2 + 3\lambda + 4$, and the eigenvalues are $\dfrac{-3 \pm \sqrt{-7}}{2}$. These are complex numbers with negative real part, so this describes an attracting spiral. $\square$

May 9

29.4. **Structural stability.** Recall: We were studying an autonomous system

$$\dot{x} = f(x,y)$$
$$\dot{y} = g(x,y).$$

To understand the behavior near a critical point (a,b), we made a change of variable

$$x = a + X$$
$$y = b + Y$$

to move the critical point to $(0,0)$, and we replaced $f(x,y)$ and $g(x,y)$ by their best linear approximations to get the linear system

$$\begin{pmatrix} \dot{X} \\ \dot{Y} \end{pmatrix} \approx A \begin{pmatrix} X \\ Y \end{pmatrix},$$

where A is $J(a,b)$, the value of the Jacobian matrix at the critical point $(x,y) = (a,b)$.

Question 29.7. When is it OK to say that the original system behaves like the linear system?

Approximation principle. *If an autonomous system is approximated near a critical point (a,b) by*

$$\begin{pmatrix} \dot{X} \\ \dot{Y} \end{pmatrix} \approx A \begin{pmatrix} X \\ Y \end{pmatrix},$$

and if the system $\dot{\mathbf{x}} = A\mathbf{x}$ is structurally stable (saddle, repelling/attracting node, repelling/attracting spiral), then the phase portrait for the original system looks near (a, b) like the phase portrait for

$$\begin{pmatrix} \dot{X} \\ \dot{Y} \end{pmatrix} = A \begin{pmatrix} X \\ Y \end{pmatrix}$$

near $(0, 0)$. (We aren't going to prove this, or even make precise what "looks like" means.) *The phase portrait may become more and more warped as one moves away from the critical point.*

These cases, as opposed to the borderline cases in which A lies on the boundary between regions in the trace-determinant plane, are called structurally stable.

Warning: Stability and structural stability are different concepts:

- *Stable* means that all nearby solutions tend to the critical point.
- *Structurally stable* means that the phase portrait type is robust, unaffected by small changes in the matrix entries.

Example 29.8. The phase portrait for

$$\begin{pmatrix} \dot{X} \\ \dot{Y} \end{pmatrix} \approx \begin{pmatrix} 0 & 4 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}$$

is a center, so trajectories are periodic. But if an autonomous system has this as its linear approximation at a critical point, it is not guaranteed that trajectories are periodic, because the slight warping might make the trajectories no longer come back to exactly the initial position after going around once.

29.5. **Big picture.**

Steps for drawing the phase portrait for an autonomous system $\dot{x} = f(x, y)$, $\dot{y} = g(x, y)$:

1. Solve the system
$$\begin{aligned} f(x, y) &= 0 \\ g(x, y) &= 0 \end{aligned}$$
to find all the critical points in the (x, y)-phase plane. There is a stationary trajectory at each critical point.
2. Calculate the Jacobian matrix
$$J(x, y) := \begin{pmatrix} \partial f/\partial x & \partial f/\partial y \\ \partial g/\partial x & \partial g/\partial y \end{pmatrix}.$$
This will be a 2×2 matrix of functions of x and y.
3. At each critical point (a, b),
   (a) Compute the numerical 2×2 matrix $A := J(a, b)$, by evaluating $J(x, y)$ at (a, b).

(b) Determine whether the critical point is stable (attracting) or not:

$$\text{stable} \iff \operatorname{tr} A < 0 \text{ and } \det A > 0.$$

Or, for a more detailed picture, find the eigenvalues of A to classify the phase portrait for the "linear approximation system" $\begin{pmatrix} \dot{X} \\ \dot{Y} \end{pmatrix} \approx A \begin{pmatrix} X \\ Y \end{pmatrix}$. For further details:

- If the eigenvalues are real, find the eigenlines. If, moreover, the eigenvalues have the same sign, also determine the slow eigenline since trajectories in the (X, Y)-plane will be tangent to that line.
- If the eigenvalues are complex (and not real), compute a velocity vector to determine whether the rotation is clockwise or counterclockwise.

(c) Mark the critical point (a, b) in the (x, y)-plane, and draw a miniature copy of the linear approximation's phase portrait shifted so that it is centered at (a, b); this is justified in the structurally stable cases (saddle, repelling node, attracting node, or spiral). Indicate with arrowheads the direction of motion on the trajectories near the critical point.

4. (Optional) Find the velocity vector at a few other points, or use a computer.
5. (Optional) Solve $f(x, y) = 0$ to find all the points where the velocity vector is vertical or $\mathbf{0}$. Similarly, one could solve $g(x, y) = 0$ to find all the points where the velocity vector is horizontal or $\mathbf{0}$.
6. Connect trajectories emanating from or approaching critical points, keeping in mind that *trajectories never cross or touch.*

Problem 29.9. Sketch the phase portrait for the deer-wolf system

$$\begin{aligned} \dot{x} &= 3x - x^2 - xy \\ \dot{y} &= y - y^2 + xy. \end{aligned}$$

Solution: We already found the critical points

$$(0,0), \quad (0,1), \quad (3,0), \quad (1,2).$$

We already found the Jacobian matrix

$$J(x,y) = \begin{pmatrix} 3 - 2x - y & -x \\ y & 1 - 2y + x \end{pmatrix}.$$

Critical point $(1, 2)$*:* We already calculated

$$J(1,2) = \begin{pmatrix} -1 & -1 \\ 2 & -2 \end{pmatrix}.$$

This has trace -3 and determinant 4, so this critical point is stable.

The characteristic polynomial is $\lambda^2 + 3\lambda + 4$, and the eigenvalues are $\dfrac{-3 \pm \sqrt{-7}}{2}$. These are complex numbers with negative real part, so this describes an attracting spiral. The velocity vector at $\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ is $\begin{pmatrix} -1 \\ 2 \end{pmatrix}$, so the spiral is counterclockwise. This is a structurally stable case, so the phase portrait for the original system near $(1,2)$ will be a counterclockwise attracting spiral too.

Critical point $(0,0)$*:*

$$J(0,0) = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}.$$

This has trace 4, so this critical point is unstable. Since the matrix is diagonal, its eigenvalues are the diagonal entries 3 and 1, and the vectors $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ are corresponding eigenvectors. The eigenvalues are distinct positive real numbers, so this describes a repelling node. The slow eigenline is the Y-axis, so most trajectories emanating from $(0,0)$ are tangent to the Y-axis. This is a structurally stable case, so the phase portrait for the original system near $(0,0)$ too will be a repelling node, and most trajectories emanating from $(0,0)$ are tangent to the y-axis.

Critical point $(0,1)$*:*

$$J(0,1) = \begin{pmatrix} 2 & 0 \\ 1 & -1 \end{pmatrix}.$$

This has trace 1, so this critical point is unstable. Since the matrix is lower triangular, its eigenvalues are the diagonal entries 2 and -1. The eigenvalues are real numbers of opposite sign, so this describes a saddle. The eigenlines for the eigenvalues 2 and -1 are $Y = \frac{1}{3}X$ and $X = 0$. This is a structurally stable case, so the phase portrait for the original system near $(0,1)$ is a saddle too.

Critical point $(3,0)$*:*

$$J(3,0) = \begin{pmatrix} -3 & -3 \\ 0 & 4 \end{pmatrix}.$$

This has trace 1, so this critical point is unstable. Since the matrix is upper triangular, its eigenvalues are the diagonal entries -3 and 4. The eigenvalues are real numbers of opposite sign, so this describes a saddle. The eigenlines for the eigenvalues -3 and 4 are $Y = 0$ and $Y = -\frac{7}{3}X$. This is a structurally stable case, so the phase portrait for the original system near $(3,0)$ is a saddle too.

At which points are the trajectories vertical?
These are the points at which the x-coordinate of the velocity vector is 0, i.e., the points where

$$3x - x^2 - xy = 0.$$

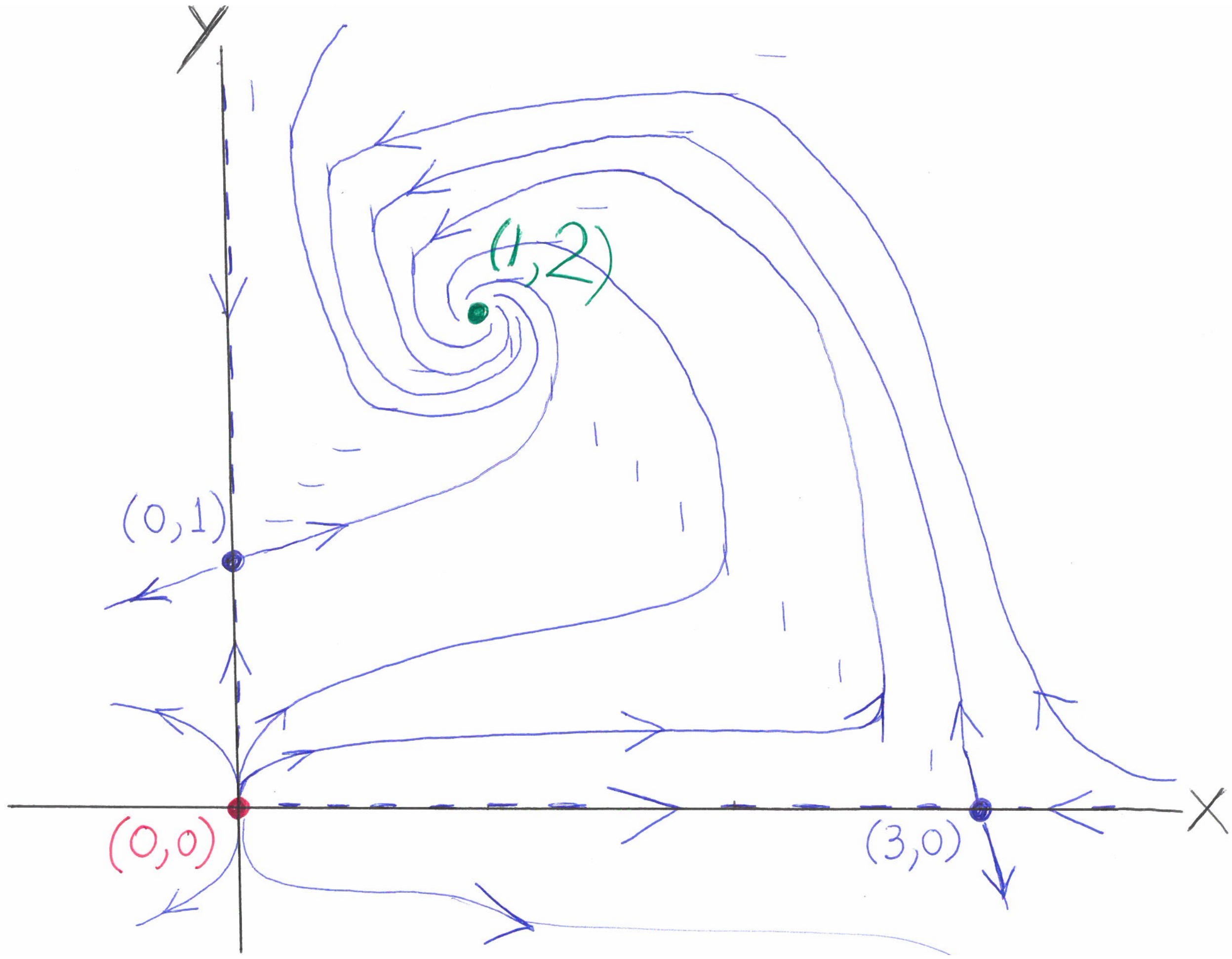
Factoring shows that these are the points on the lines $x = 0$ and $3 - x - y = 0$. So in the phase portrait we draw little vertical segments at points on these lines. In particular, there will be trajectories along $x = 0$, and we can plot them using the 1-dimensional phase line methods, by sampling the velocity vector at one point in each interval created by the critical points. The line $3 - x - y - 0$ does not contain trajectories, however, since that line has slope -1, while trajectories are vertical as they pass through these points.

At which points are the trajectories horizontal?
These are points at which

$$y - y^2 + xy = 0.$$

These are the lines $y = 0$ and $1 - y + x = 0$, so draw little horizontal segments at points on these lines. Again we can study trajectories along $y = 0$ using 1-dimensional phase line methods.

Big picture:



Try the "Vector Fields" mathlet

`http://mathlets.org/mathlets/vector-fields/`

29.6. **Changing the parameters of the system.** The big picture suggests that all trajectories in the first quadrant tend to $(1, 2)$ as $t \to +\infty$. In other words, as long as there were some deer and some wolves to begin with, eventually the populations stabilize at about 1000 deer and 200 wolves.

Problem 29.10. Suppose that we start feeding the deer so that the system becomes

$$\dot{x} = ax - x^2 - xy$$
$$\dot{y} = y - y^2 + xy$$

for some number a slightly larger than 3. What happens?

Solution: The critical points will move slightly, but they won't change their stability. The populations will end up at the stable critical point, which is the one near $(1, 2)$. To find it, solve

$$0 = ax - x^2 - xy$$
$$0 = y - y^2 + xy.$$

Since we're looking for a solution with $x > 0$ and $y > 0$, it is OK to divide the equations by x and y, respectively:

$$0 = a - x - y$$
$$0 = 1 - y + x.$$

Solving gives

$$x = \frac{a-1}{2}, \quad y = \frac{a+1}{2}.$$

For $a = 3$, this is $x = 1$ and $y = 2$. As a increases beyond 3, the deer population increases, but we also see an increase in the wolf population. By feeding the deer we have provided more food for the wolves as well!

29.7. **Fences.** In the original deer-wolf system, how can you be *sure* that all trajectories starting with $x > 0$ and $y > 0$ tend to $(1, 2)$?

Steps to *prove* that all trajectories approach the stable critical point:

(1) Find a window into which all trajectories must enter and never leave.
(2) Do a numerical simulation within the window.

Let's do step 1 for the deer-wolf system. A trajectory could escape in four ways: up, down, left, and right. We need to rule out all four.

Bottom: A trajectory that starts in the first quadrant cannot cross the nonnegative part of the x-axis, because the trajectories along the x-axis act as fences. A trajectory cannot even tend to a point on the x-axis, because such a point would be a critical point, and the phase portrait types at $(0, 0)$ and $(3, 0)$ make such an approach impossible.

Left: By the same argument, the nonnegative part of the y-axis is a fence that cannot be approached.

Right: We have

$$\dot{x} = 3x - x^2 - xy \leq 3x - x^2 < 0$$

whenever $x > 3$ (if $3x - x^2$ is negative, then $3x - x^2 - xy$ is even more negative since it has something subtracted). So all the vertical lines $x = c$ for $c > 3$ are fences that prevent trajectories from moving to the right across them. All trajectories move leftward if $x > 3$, and they can settle down only in the range $0 \leq x \leq 3$.

Top: Assuming $x \le 3$, we have

$$\dot{y} = y - y^2 + xy \le y - y^2 + 3y = 4y - y^2 < 0$$

whenever $y > 4$. Thus for $c > 4$, the horizontal segments $y = c$, $0 \le x \le 3$ are fences preventing trajectories from moving up through them.

Conclusion: All trajectories starting with $x > 0$, $y > 0$ (the only ones we care about) eventually enter the window $0 \le x \le 3$, $0 \le y \le 4$ and stay there. This is small enough that a numerical simulation can now show that all these points tend to $(1, 2)$ (step 2).

The final exam covers everything up to here, in the sense that you are not required to know anything specific below. On the other hand, the topics below serve partially as review of earlier topics that *are* covered.

May 11

29.8. **Nonlinear centers, limit cycles, etc.** Consider an autonomous system. Suppose that P is a critical point. Suppose that the linear approximation system at P is a center. What is the behavior of the original system near P? It's not necessarily a center. (This is not a structurally stable case.) In fact, there are many possibilities:

- nonlinear center, in which the trajectories are periodic (but not necessarily exact ellipses)
- repelling spiral
- attracting spiral
- hybrid situation containing a limit cycle: a periodic trajectory with an outward spiral approaching it from within and an inward spiral approaching it from outside!

For an example of a limit cycle (called the van der Pol limit cycle), set $a = 0.1$ in the system

$$\begin{aligned} \dot{x} &= y \\ \dot{y} &= a(1 - x^2)y - x \end{aligned}$$

in the "Vector Fields" mathlet

`http://mathlets.org/mathlets/vector-fields/`

30.1. **Modeling.**

Problem 30.1. Model a pendulum, consisting of a weight attached to a rod hanging from a pivot at the top.

Variables and functions: Define

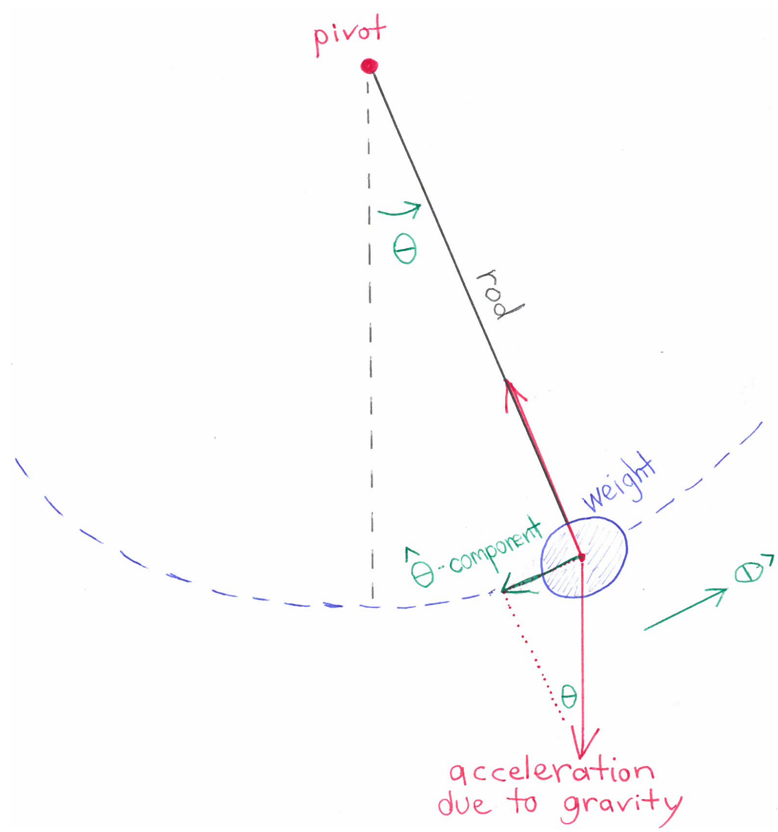
t : time

θ : angle measured counterclockwise from the rest position

Here t is the independent variable, and θ is a function of t.

Simplifying assumptions:

- The rod has length 1, so θ equals arc length and $\dot{\theta}$ is velocity.
- The rod has negligible mass.
- The rod does not bend or stretch.
- The weight has mass 1.
- The pivot is such that the motion is in a plane (no Coriolis effect).
- The local gravitational field g is a constant (the pendulum is not thousands of kilometers tall).



Equation: When the weight is at a certain position, let $\hat{\theta}$ be the unit vector in the direction that the weight moves as θ starts to increase. The $\hat{\theta}$-component of the weight's acceleration is

$$\ddot{\theta} = \underbrace{-g\sin\theta}_{\text{gravity}}. \tag{13}$$

More realistic (with friction, assumed for simplicity to be proportional to $\dot{\theta}$):

$$\boxed{\ddot{\theta} = -\underbrace{b\dot{\theta}}_{\text{friction}} \underbrace{- g\sin\theta}_{\text{gravity}}}.$$

The $\ddot{\theta}$ and $b\dot{\theta}$ terms are linear, but the $g\sin\theta$ makes the whole DE nonlinear.

Remark 30.2. If θ is very small, then it is reasonable to replace the nonlinear term by its best linear approximation at $\theta = 0$, namely $\sin\theta \approx \theta$, which leads to

$$\ddot{\theta} + b\dot{\theta} + g\theta = 0,$$

a damped harmonic oscillator.

30.2. **Converting to a first-order system.** But to get an accurate understanding even when θ is not so small, we need to analyze equation (13) in its full nonlinear glory. It is a *second-order* nonlinear ODE; we haven't developed tools for those. So instead convert it to a (still nonlinear) *system* of first-order ODEs, by introducing a new function $v := \dot{\theta}$ (velocity):

$$\begin{aligned} \dot{\theta} &= v \\ \dot{v} &= -bv - g\sin\theta. \end{aligned}$$

This is an autonomous system! So we can use all the methods we've been developing.

30.3. **Critical points.** The critical points are given by

$$\begin{aligned} v &= 0 \\ -bv - g\sin\theta &= 0. \end{aligned}$$

Substituting $v = 0$ into the second equation leads to $\sin\theta = 0$, so $\theta = \ldots, -2\pi, -\pi, 0, \pi, 2\pi, \ldots$. Thus there are infinitely many critical points:

$$\ldots, \quad (-2\pi, 0), \quad (-\pi, 0), \quad (0,0), \quad (\pi, 0), \quad (2\pi, 0), \quad \ldots.$$

But these represent only two distinct physical situations, since adding 2π to θ does not change the position of the weight.

30.4. **Phase portrait of the frictionless pendulum; energy levels.** Let's draw the phase portrait in the (θ, v)-plane when $\boxed{b=0}$ and $\boxed{g=1}$. Now the system is

$$\dot{\theta} = v$$
$$\dot{v} = -\sin\theta.$$

Flashcard question: In the frictionless case, are the critical points $(0,0)$ and $(\pi,0)$ stable?
Answer: Neither is stable.

- The point $(\pi,0)$ corresponds to a vertical rod with the weight precariously balanced at the top. If the weight is moved slightly away, the trajectory goes far from $(\pi,0)$.
- The point $(0,0)$ corresponds to a vertical rod with the weight at the bottom. If the weight is moved slightly away, the trajectory does not tend to $(0,0)$ in the limit because the pendulum oscillates forever in the frictionless case. $\square$

To analyze the behavior near each critical point, use a linear approximation. The Jacobian matrix is

$$J(\theta, v) = \begin{pmatrix} 0 & 1 \\ -\cos\theta & 0 \end{pmatrix}.$$

Critical point $(\pi,0)$*:*

$$J(\pi,0) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

(This is not a diagonal matrix: wrong diagonal.)

Eigenvalues: 1, -1

Eigenvectors: $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$

Type of the linear approximation: Saddle, with outgoing trajectories of slope 1, incoming trajectories of slope -1.

Type of the original nonlinear system: Same.

Critical point $(0,0)$*:*

$$J(0,0) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Eigenvalues: $\pm i$.

Type of the linear approximation: Center.

Type of the original nonlinear system: ??? (not a structurally stable case, so we can't tell yet whether trajectories near $(0,0)$ are periodic).

To figure out what happens near $(0,0)$, use conservation of energy! Assume mass 1. The energy function is

$$E(\theta, v) := \underbrace{\frac{1}{2}v^2}_{\text{kinetic energy}} + \underbrace{1-\cos\theta}_{\text{potential energy}}.$$

(Remember that all the constants were set to 1, so potential energy equals height, which we choose to measure relative to the rest position.)

Let's check conservation of energy:

$$\begin{aligned}\dot{E} &= v\dot{v} + (\sin\theta)\dot{\theta} \\ &= v(-\sin\theta) + (\sin\theta)v \\ &= 0.\end{aligned}$$

This means that along each trajectory, E is constant. In other words, each trajectory is contained in a level curve of E.

Energy level $E = 0$:

$$\frac{1}{2}v^2 + (1-\cos\theta) = 0.$$

Both terms on the left are nonnegative, so their sum can be 0 only if both are 0, which happens only at $(\theta, v) = (0,0)$ (and the similar points with some $2\pi n$ added to θ). The energy level $E = 0$ consists of the stationary trajectory at $(0,0)$.

Energy level $E = \epsilon$ for small $\epsilon > 0$:

$$\frac{1}{2}v^2 + (1-\cos\theta) = \epsilon.$$

Both kinetic energy and potential energy must be small, so the height is small, so θ is small, so $\cos\theta \approx 1 - \frac{\theta^2}{2}$, so the energy level is very close to

$$\frac{v^2}{2} + \frac{\theta^2}{2} = \epsilon,$$

a small circle. The trajectory goes clockwise along it since θ is increasing when $\dot{\theta} > 0$, and decreasing when $\dot{\theta} < 0$. So trajectories near $(0,0)$ are periodic ovals (approximately circles); these represent a pendulum doing a small oscillation near the bottom. The critical point is a nonlinear center.

Lecture more or less ended here. The remaining topics below were mentioned only very briefly.

Energy level $E = 2$:

$$\begin{aligned}
\frac{1}{2}v^2 + (1 - \cos\theta) = 2 \\
\frac{1}{2}v^2 &= 1 + \cos\theta \\
&= 1 + (2\cos^2\frac{\theta}{2} - 1) \\
&= 2\cos^2\frac{\theta}{2} \\
v &= \pm 2\cos\frac{\theta}{2}.
\end{aligned}$$

Does this mean that the motion is periodic, going around and around? No. This energy level contains three physical trajectories:

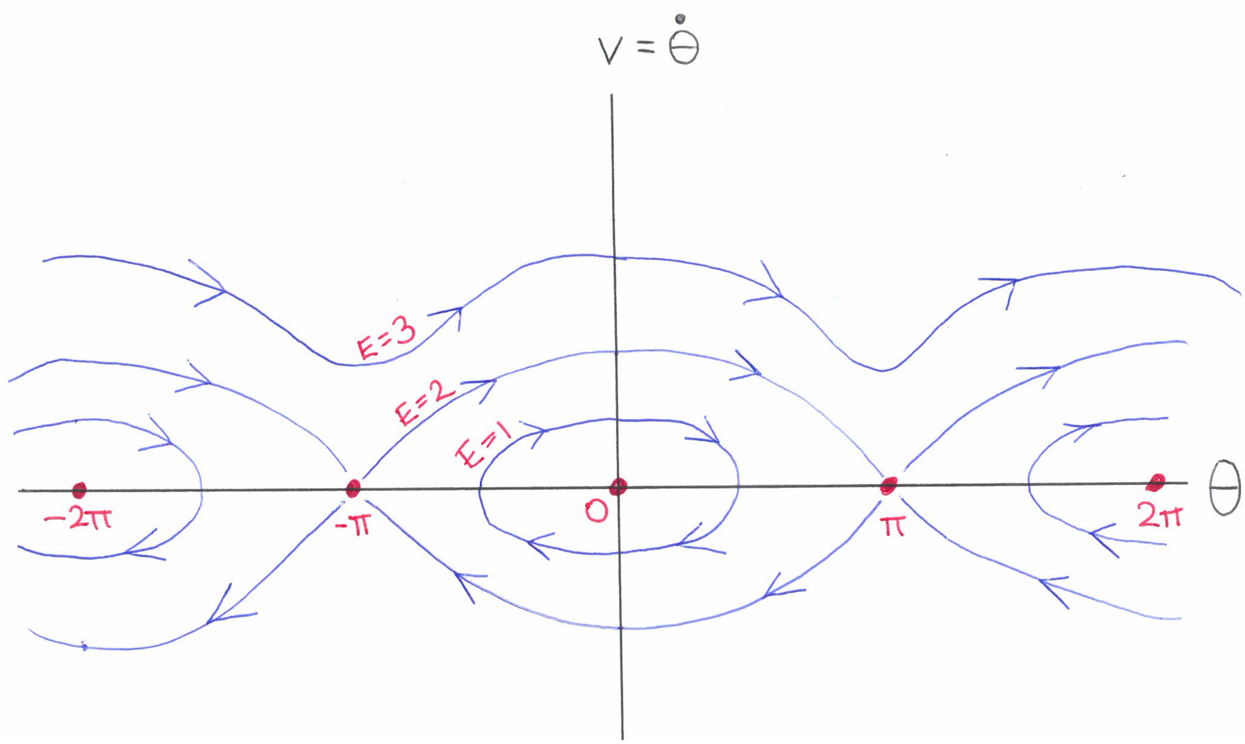
- one in which the weight is stationary at the top
- one in which the weight does one clockwise loop as t goes from $-\infty$ to ∞, slowing down as it approaches the top, taking infinitely long to get there (and infinitely long to come from there),
- the same, except counterclockwise.

In the last two cases, the weight can't actually reach the top, since its phase plane trajectory can't touch the stationary trajectory.

Energy level $E = 3$:

$$\begin{aligned}
\frac{1}{2}v^2 + (1 - \cos\theta) = 3 \\
v &= \pm\sqrt{4 + 2\cos\theta}.
\end{aligned}$$

The possibility $v = \sqrt{4 + 2\cos\theta}$ is a periodic function of θ, varying between $\sqrt{2}$ and $\sqrt{6}$. The energy level consists of two trajectories: in each, the weight makes it to the top still having some kinetic energy, so that it keeps going around (either clockwise or counterclockwise).



30.5. **Phase portrait of the damped pendulum.** Next let's draw the phase portrait when $\boxed{b > 0}$ (so there is friction) and $\boxed{g = 1}$. The system is

$$\begin{aligned} \dot{\theta} &= v \\ \dot{v} &= -bv - \sin\theta. \end{aligned}$$

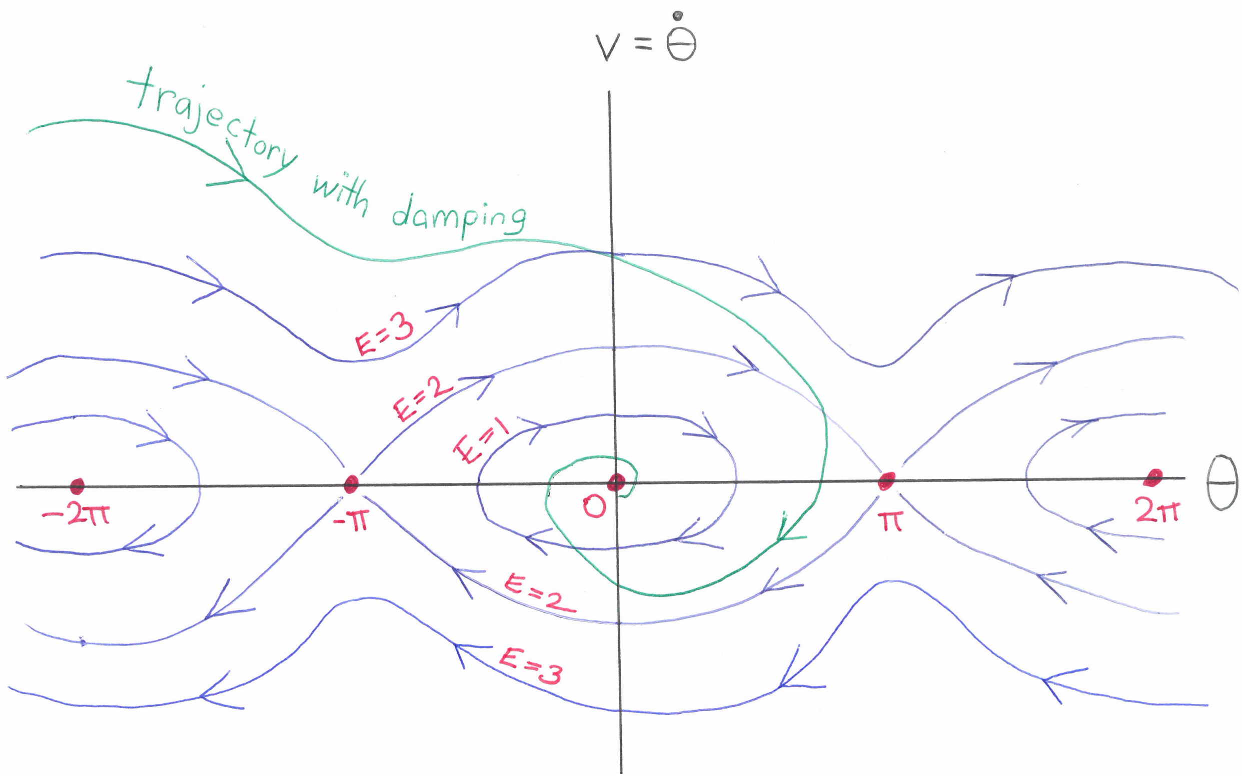
This time,

$$\begin{aligned} \dot{E} &= v\dot{v} + (\sin\theta)\dot{\theta} \\ &= v(-bv - \sin\theta) + (\sin\theta)v \\ &= -bv^2, \end{aligned}$$

so energy is lost to friction whenever the weight is moving.

- There are still the stationary trajectories at critical points.
- All other trajectories cross the energy levels, moving to lower energy: the direction field always points "inwards" towards lower energy; the energy levels serve as fences preventing phase plane motion to higher energy; the trajectory tends to a limit that must be a critical point. one of

$$\ldots, \quad (-2\pi, 0), \quad (-\pi, 0), \quad (0, 0), \quad (\pi, 0), \quad (2\pi, 0), \quad \ldots.$$



May 14

31. Numerical methods

31.1. **Euler's method.** Consider a nonlinear ODE $\dot{y} = f(t, y)$. It specifies a slope field in the (t, y)-plane, and solution curves follow the slope field.

Suppose that we are given a starting point (t_0, y_0) (here t_0 and y_0 are given numbers), and that we are trying to approximate the solution curve through it.

Question 31.1. Where, approximately, will be the point on the solution curve at a time h seconds later?

Solution: Pretend that the solution curve is a straight line segment between times t_0 and $t_0 + h$, with slope as specified by the ODE at (t_0, y_0). The ODE says that the slope at (t_0, y_0) is $f(t_0, y_0)$, so the estimated answer is (t_1, y_1) with

$$t_1 := t_0 + h$$
$$y_1 := y_0 + \underbrace{f(t_0, y_0)}_{\text{slope}}\, h. \quad \square$$

Question 31.2. Where, approximately, will be the point on the solution curve at time $t_0 + 3h$?

Solution: The stupidest answer would be to take 3 steps each using the initial slope $f(t_0, y_0)$ (or equivalently, one big step of width $3h$). The slightly less stupid answer is called Euler's method: take 3 steps, *but reassess the slope after each step, using the slope field at each successive position*:

$$\begin{aligned}
t_1 &:= t_0 + h & \qquad y_1 &:= y_0 + f(t_0, y_0)\, h \\
t_2 &:= t_1 + h & \qquad y_2 &:= y_1 + f(t_1, y_1)\, h \\
t_3 &:= t_2 + h & \qquad y_3 &:= y_2 + f(t_2, y_2)\, h.
\end{aligned}$$

The sequence of line segments from (t_0, y_0) to (t_1, y_1) to (t_2, y_2) to (t_3, y_3) is an approximation to the solution curve. The answer to the question is approximately (t_3, y_3). $\square$

Usually these calculations are done by computer, and there are round-off errors in calculations. But even if there are no round-off errors, Euler's method usually does not give the exact answer. The problem is that the actual slope of the solution curve changes between t_0 and $t_0 + h$, so following a segment of slope $f(t_0, y_0)$ for this entire time interval is not exactly correct.

To improve the approximation, use a smaller step size h, so that the slope is reassessed more frequently. The cost of this, however, is that in order to increase t by a fixed amount, more steps will be needed.

Under reasonable hypotheses on f, one can prove that as $h \to 0$, this process converges and produces an exact solution curve in the limit. This is one way to prove the existence theorem for ODEs.

Try the "Euler's Method" mathlet

`http://mathlets.org/mathlets/eulers-method/`

31.2. **Euler's method for systems.** A first-order system of ODEs can be written in vector form $\dot{\mathbf{x}} = \mathbf{f}(t, \mathbf{x})$, where $\mathbf{f}$ is a vector-valued function. Euler's method works the same way. Starting from $(t_0, \mathbf{x}_0)$, define

$$\begin{aligned}
t_1 &:= t_0 + h & \qquad \mathbf{x}_1 &:= \mathbf{x}_0 + \mathbf{f}(t_0, \mathbf{x}_0)\, h \\
t_2 &:= t_1 + h & \qquad \mathbf{x}_2 &:= \mathbf{x}_1 + \mathbf{f}(t_1, \mathbf{x}_1)\, h \\
t_3 &:= t_2 + h & \qquad \mathbf{x}_3 &:= \mathbf{x}_2 + \mathbf{f}(t_2, \mathbf{x}_2)\, h.
\end{aligned}$$

31.3. **Tests for reliability.**

Question 31.3. How can we decide whether answers obtained numerically can be trusted?

Here are some heuristic tests. ("Heuristic" means that these tests seem to work in practice, but they are not proved to work always.)

- Self-consistency: Solution curves should not cross! If numerically computed solution curves appear to cross, a smaller step size is needed. (E.g., try the mathlet "Euler's Method" with $\dot{y} = y^2 - x$, step size 1, and starting points $(0, 0)$ and $(0, 1/2)$.)
- Convergence as $h \to 0$: The estimate for $y(t)$ at a fixed later time t should converge to the true value as $h \to 0$. If shrinking h causes the estimate to change a lot, then h is probably not small enough yet. (E.g., try the mathlet "Euler's Method" with $y' = y^2 - x$ with starting point $(0, 0)$ and various step sizes.)
- Structural stability: If small changes in the DE's parameters or initial conditions change the outcome completely, the answer probably should not be trusted. One reason for this could be a separatrix, a curve such that nearby starting points on different sides lead to qualitatively different outcomes; this is not a fault of the numerical method, but is an instability in the answer nevertheless. (E.g., try the mathlet "Euler's Method" with $y' = y^2 - x$, starting point $(-1, 0)$ or $(-1, -0.1)$, and step size 0.125 or actual solution.)

31.4. **Change of variable.** Euler's method generally can't be trusted to give reasonable values when (t, y) strays very far from the starting point. In particular, the solutions it produces usually deviate from the truth as $t \to \pm\infty$, or in situations in which $y \to \pm\infty$ in finite time. Anything that goes off the screen can't be trusted.

Example 31.4. The solution to $\dot{y} = y^2 - t$ starting at $(-2, 1)$ seems to go to $+\infty$ in finite time. But Euler's method never produces a value of $+\infty$.

To see what is really happening in this example, try the change of variable $u = 1/y$. To rewrite the DE in terms of u, substitute $y = 1/u$ and $\dot{y} = -\dot{u}/u^2$:

$$\begin{aligned} -\frac{\dot{u}}{u^2} &= \frac{1}{u^2} - t \\ \dot{u} &= -1 + tu^2. \end{aligned}$$

This is equivalent to the original DE, but now, when y is large, u is small, and Euler's method can be used to estimate the time when u crosses 0, which is when y blows up.

31.5. **Runge–Kutta methods.** When computing $\int_a^b f(t)\, dt$ numerically, the most primitive method is to use the left Riemann sum: divide the range of integration into subintervals of width h, and estimate the value of $f(t)$ on each subinterval as being the value at the left endpoint. More sophisticated methods are the *trapezoid rule* and *Simpson's rule*, which have smaller errors.

There are analogous improvements to Euler's method.

Integration	Differential equation	Error
left Riemann sum	Euler's method	$O(h)$
trapezoid rule	second-order Runge–Kutta method (RK2)	$O(h^2)$
Simpson's rule	fourth-order Runge–Kutta method (RK4)	$O(h^4)$

The big-O notation $O(h^4)$ means that there is a constant C (depending on everything except for h) such that the error is at most Ch^4, assuming that h is small. The error estimates in the table are valid for reasonable functions.

The Runge–Kutta methods "look ahead" to get a better estimate of what happens to the slope over the course of the interval $[t_0, t_0 + h]$.

Here is how one step of the second-order Runge–Kutta method (RK2) goes

1. Starting from (t_0, y_0), look ahead to see where one step of Euler's method would land, say (t_1, y_1), but do not go there!
2. Instead sample the slope at the *midpoint* $\left(\frac{t_0+t_1}{2}, \frac{y_0+y_1}{2}\right)$.
3. Now move along the segment of *that* slope: the new point is

$$\left(t_0 + h, y_0 + f\left(\frac{t_0 + t_1}{2}, \frac{y_0 + y_1}{2}\right) h\right).$$

Repeat, reassessing the slope after each step. (RK2 is also called midpoint Euler.)

The fourth-order Runge–Kutta method (RK4) is similar, but more elaborate, averaging several slopes. It is probably the most commonly used method for solving DEs numerically. Some people simply call it *the* Runge–Kutta method. The mathlets use RK4 with a small step size to compute the "actual" solution to a DE.

May 16

32. Review

32.1. **Check your answers!** On the final exam, there is no excuse for getting an eigenvector wrong, since you will have plenty of time to check it! You can also check solutions to linear systems, or solutions to DEs.

32.2. **D'Alembert's solution to the wave equation.** Consider the wave equation

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2}$$

without boundary conditions. For any function f of one variable, the function $u(x,t) := f(x - ct)$ of the two variables x and t is a solution to the wave equation (just plug it in — both sides end up being $c^2 f''(x - ct)$. At time $t = 0$, the shape of the wave is the graph of $u(x, 0) = f(x)$; at time $t = 1$, the shape of the wave is the graph of $u(x, 1) = f(x - c)$, which is the same shape shifted c units to the right; and so on. The physical meaning of this solution is a wave keeping its shape but moving to the right with speed c.

Similarly, for any function g, the function $u(x,t) := g(x+ct)$ is a solution. The physical meaning of this solution is a wave moving to the left.

Since the wave equation is a *linear* PDE, the superposition

$$u(x,t) := f(x-ct) + g(x+ct)$$

is again a solution, for any choice of functions f and g. This turns out to be the general solution (it's an infinite family of solutions, since there are infinitely many possibilities for the functions f and g). This is called d'Alembert's solution. The f and g are like the parameters c_1 and c_2 in the general solution to an ODE: to find them, one must use initial conditions.

Problem 32.1. Suppose that $u(x,t)$ is the solution to the wave equation

$$\frac{\partial^2 u}{\partial t^2} = 4\frac{\partial^2 u}{\partial x^2}$$

such that the initial waveform at time 0 is given by the function

$$r(x) = \begin{cases} 0, & \text{if } t < 0, \\ 2, & \text{if } 0 < t < 1, \\ 4, & \text{if } t > 1, \end{cases}$$

and the initial velocity at each point is 0. Find a formula for $u(x,t)$.

Solution: In the given wave equation, c^2 is 4, so the speed of the waves is $c = 2$. Thus

$$u(x,t) = f(x-2t) + g(x+2t)$$

for some functions f and g to be determined. The initial conditions will constrain the possibilities for f and g. For example, evaluating both sides at $t = 0$ and plugging in the initial condition $u(x,0) = r(x)$ on the left side gives

$$r(x) = f(x) + g(x).$$

On the other hand, taking the t-derivative of both sides gives

$$\frac{\partial u}{\partial t}(x,t) = -2f'(x-2t) + 2g'(x+2t)$$

by the chain rule, and then evaluating at $t = 0$ and plugging in the initial condition $\frac{\partial u}{\partial t}(x,0) = 0$ on the left side gives

$$0 = -2f'(x) + 2g'(x),$$

which implies that $f'(x) = g'(x)$, so $f(x) = g(x) + C$ for some constant C. Solving the system

$$r(x) = f(x) + g(x)$$
$$f(x) = g(x) + C$$

for the unknown functions $f(x)$ and $g(x)$ (e.g., by substituting the second equation into the first) gives

$$f(x) = \frac{r(x) + C}{2}, \qquad g(x) = \frac{r(x) - C}{2}.$$

Plug these back into the general solution to get the particular solution

$$\begin{aligned} u(x,t) &= f(x - 2t) + g(x + 2t) \\ &= \frac{r(x - 2t) + C}{2} + \frac{r(x + 2t) - C}{2} \\ &= \frac{r(x - 2t)}{2} + \frac{r(x + 2t)}{2}, \end{aligned}$$

which is a known function, since the function r was given. $\square$.

There are at least two ways to visualize the solution $u(x,t)$ we found.

The first way is to plot the waveform at different times, to produce snapshots that if displayed in succession will be a movie of the wave. For example, what does the wave look like at time $t = 1$? The answer is the graph of

$$u(x,1) = \frac{r(x - 2)}{2} + \frac{r(x + 2)}{2},$$

but what does this look like? The function $r(x)$ jumps up at $x = 0$ and $x = 1$, so $r(x - 2)$ jumps up when $x - 2 = 0$ or $x - 2 = 1$ (that is, $x = 2$ or $x = 3$). Meanwhile, $r(x + 2)$ jumps up when $x + 2 = 0$ or $x + 2 = 1$ (that is, $x = -2$ or $x = -1$). Thus $u(x,1)$ jumps up at $x = -2, -1, 2, 3$; these values divide the real line into intervals on which $u(x,1)$ is constant. To find the values, we just need to evaluate $u(x,1)$ at a single x-value in each interval. For example, for $2 < x < 3$, the value of $u(x,1)$ equals

$$u(2.5,1) = \frac{r(0.5)}{2} + \frac{r(4.5)}{2} = \frac{2}{2} + \frac{4}{2} = 3.$$

Similar calculations eventually lead to

$$u(x,1) = \begin{cases} 0, & \text{if } t < -2, \\ 1, & \text{if } -2 < t < -1, \\ 2, & \text{if } -1 < t < 2, \\ 3, & \text{if } 2 < t < 3, \\ 4, & \text{if } 3 < t, \end{cases}$$

so the wave at $t = 1$ looks like a staircase with four steps going up.

The second way to visualize the solution $u(x,t)$ is draw its space-time diagram, in the (x,t)-plane. At $t = 0$ (the horizontal axis), mark the x-values where $u(x,0)$ jumps up ($x = 0$ and $x = 1$) and in a different color write the value of $u(x,0)$ in each interval formed (these values are 0, 1, 2). Then one can do the same for $t = 1$ (the horizontal line one unit higher):

mark the points where $u(x,1)$ jumps up $(-2,-1,2,3)$ and write the values in each interval formed. Actually, it is easier to do this for all t at once instead of one t-value at a time. That is, since $r(x)$ jumps at 0 and 1, the function

$$u(x,t) = \frac{r(x-2t)}{2} + \frac{r(x+2t)}{2}$$

jumps whenever $x - 2t = 0$, $x - 2t = 1$, $x + 2t = 0$, or $x + 2t = 1$. The parts of these four lines above the x-axis (i.e., the part where $t \geq 0$) are the wave fronts. They divide the upper half of the plane ($t \geq 0$) into regions such that $u(x,t)$ is constant on each region. To find the constant value within each region, evaluate $u(x,t)$ at one point in the region.

32.3. **Isoclines and fences.**

Problem 32.2. What happens to solutions $y(x)$ to $y' = y^3 - x$ as $x \to -\infty$?

(Try the "Isoclines" mathlet with $y' = y^3 - ay - x$ and $a = 0$.)

Solution: First draw the 0-isocline, which is where $y^3 - x = 0$. This is like the graph of $y = x^3$, except reflected across the line $y = x$. At $(0,1)$, the value of $y' = y^3 - x$ is $1^3 - 0 > 0$, so the region above the 0-isocline is an up region. Testing $(0,-1)$ similarly shows that the region below the 0-isocline is a down region.

A solution curve that starts in the up region must go down as x *decreases* but it can never cross the 0-isocline (because the solution curve's slope would have to be positive at the crossing point). Therefore as $x \to -\infty$, the solution curve decreases forever, but it cannot escape to $-\infty$ in finite time (the 0-isocline acts as a fence preventing that). Moreover, $y(x) \to -\infty$ as $x \to -\infty$ because if instead $y(x)$ tended to a finite limit c as $x \to -\infty$, then $y' \to 0$ as $x \to -\infty$, but then $y' = y^3 - x$ tends to $0 = c^3 - (-\infty)$ as $x \to \infty$, which is impossible.

A solution curve that starts in the down region must go up as x decreases, while the 0-isocline decreases to $-\infty$ as $x \to -\infty$, so eventually the solution curve must cross the 0-isocline and enter the up region, at which point, by the previous paragraph, it goes down towards $-\infty$ as $x \to -\infty$.

Conclusion: All solutions $y(x)$ have domain of validity extending to the left to $-\infty$, and all solutions eventually tend to $-\infty$ as $x \to -\infty$.

33. How Google search uses an eigenvector

Bonus section; not covered.

This section is based on

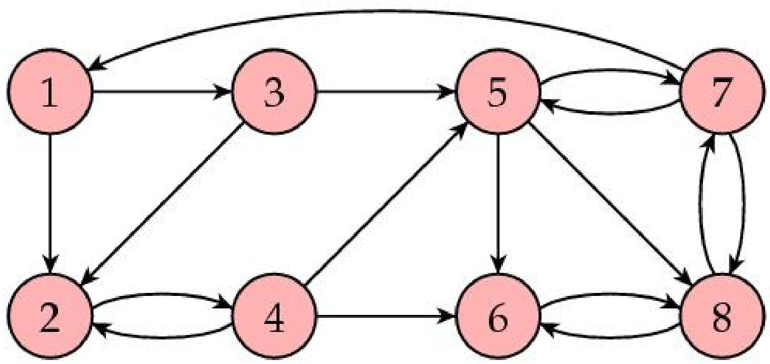
`http://www.ams.org/samplings/feature-column/fcarc-pagerank`

so go there for more details.

Eigenvalues and eigenvectors are used in many ways in science and engineering, not just for solving DEs.

Google claims that the heart of its software is PageRank: this is the algorithm for deciding how to order search results. The core idea involves an eigenvector, as we'll now explain. (The details of the algorithm are more complicated and proprietary.)

The web consists of webpages linking to each other. Number them.



(OK, maybe the web has more than 8 webpages, but you get the idea.)

Let v_i be the "importance" of webpage i.

Idea: A webpage is important if important webpages link to it. Each webpage "shares" its importance equally with all the webpages it links to.

In the example above, page 2 inherits $\frac{1}{2}$ the importance of page 1, $\frac{1}{2}$ the importance of page 3, and $\frac{1}{3}$ the importance of page 4:

$$v_2 = \frac{1}{2}v_1 + \frac{1}{2}v_3 + \frac{1}{3}v_4.$$

Yes, this is self-referential, but still it makes sense. All eight equations are encapsulated in

$$\begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \\ v_8 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 0 \\ 1/2 & 0 & 1/2 & 1/3 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 1 & 1/3 & 0 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \\ v_8 \end{pmatrix},$$

which is of the form $\mathbf{v} = A\mathbf{v}$. In other words, $\mathbf{v}$ should be an eigenvector with eigenvalue 1.

Question 33.1. How do we know that a matrix like A has an eigenvector with eigenvalue 1? Could it be that 1 is just not an eigenvalue?

Trick: use the transpose A^T.

$$\det A = \det A^T$$
$$\det(A - \lambda I) = \det(A^T - \lambda I)$$
$$\text{eigenvalues of } A = \text{eigenvalues of } A^T.$$

The equation

$$A^T \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 1/3 & 0 & 0 & 1/3 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1/3 & 0 & 0 & 0 & 1/3 & 0 & 0 & 1/3 \\ 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix},$$

shows that 1 is an eigenvalue of A^T, so 1 is an eigenvalue of A.

In the example above, the unique solution (up to multiplying by a scalar) is

$$\mathbf{v} = \begin{pmatrix} 0.0600 \\ 0.0675 \\ 0.0300 \\ 0.0675 \\ 0.0975 \\ 0.2025 \\ 0.1800 \\ 0.2950 \end{pmatrix}.$$

Google finds the eigenvector of a 50,000,000,000 $\times$ 50,000,000,000 matrix.

34. What math subject to take next?

Now that you have finished 18.02 and 18.03, there are many options open to you. Here are some of them:

- More differential equations (System functions and the Laplace transform): 18.031 (IAP)
- Linear algebra: 18.06, 18.700, or 18.701. Of these, 18.701 is for students who are already very comfortable with writing proofs.
- Probability and statistics: 18.05 (spring) or 6.041, or 18.600.
- Discrete math: 18.062/6.042, or 18.200A or **18.200** (spring).
- Real analysis: 18.100A, **18.100P** (spring), 18.100B, **18.100Q**.
- Complex analysis: 18.04 (spring).

- Continuous applied math: 18.300 (spring).

(Green means more theoretical, and red even more so. Boldface means communication-intensive.)

Special advice for potential math majors/minors, or double majors involving math:

- In general, you're probably better off taking the versions with first decimal digit 1 or higher.
- A good starting point is 18.700 or **18.100P** or **18.200**.
- **18.100P**, **18.100Q**, and **18.200** are 15 units instead of 12, and give **CI-M** credit in math because they include practice in written and oral communication; this feedback is helpful if you are learning to write proofs for the first time.
- Instead of taking 18.04, wait until you've finished 18.100 so that you can take the more advanced complex analysis 18.112.

35. Thank you

Thank you to Jennifer French for developing the MITx site for 18.03 (incorporating and adding to the content I provided, itself adapted partially from past professors' notes), to Karene Chu for helping with this, to Theresa Cummings and the rest of the staff in the Mathematics Academic Services office, to Jean-Michel Claus and the rest of the team that developed the mathlets, to David Jerison and Haynes Miller and other math professors for sharing their advice and materials from previous semesters, to the professors in engineering and other departments who coordinated with the math department to make the 18.03 content relevant to their subjects, to the MIT audio-visual staff on hand at each lecture, to Marta Manzin for cleaning the boards before each lecture, and to all 20 recitation instructors for their hard work over the semester!

And thank you to all the 18.03 students for making this class fun to teach! I hope that you all ace the final!