Linear ODE Systems

(version 12 of notes) February 26, 2018

1 Setup

An ODE is an equation involving an unknown function y and some of its derivatives. An ODE *system* is possibly *multiple* equations involving possibly *multiple* unknown functions $y_1, \ldots, y_n$ and some of their derivatives. We will be concerned with systems of *first order* ODEs. In the technical definition, these are ODE systems that have the form

$$\begin{aligned}\frac{dy_1}{dt} &= F_1(t, y_1, \ldots, y_n)\\ \frac{dy_2}{dt} &= F_2(t, y_1, \ldots, y_n)\\ &\vdots\\ \frac{dy_n}{dt} &= F_n(t, y_1, \ldots, y_n).\end{aligned}$$

A convenient way to organize multiple unknown functions is as a list of functions

$$\vec{y}(t) = \begin{bmatrix} y_1(t) \\ \vdots \\ y_n(t) \end{bmatrix}.$$

Note that the list of functions is written vertically. You could also call $\vec{y}(t)$ a "vector-valued function," meaning a function of one variable with outputs in $\mathbb{R}^n$.

Just like ODEs, ODE systems can be linear or nonlinear. Let's see what this could mean.

For ODEs, "linear" meant that the ODE had the form $\mathcal{L}(y(t)) = Q(t)$, where $\mathcal{L}$ is a linear map that operates on *functions*. The vectors in this picture are functions. For example, $\frac{dy}{dt} = t^2 y + t^3$ is a linear ODE, since it can be written as $(\frac{d}{dt} - t^2)y(t) = t^3$, and $\mathcal{L} = (\frac{d}{dt} - t^2)$ operates linearly on functions.

For ODE systems, "linear" means essentially the same thing: the ODE system has the form $\mathcal{L}(\vec{y}(t)) = \vec{Q}(t)$, where $\mathcal{L}$ is a linear map. But the vectors in this picture are *lists of functions*, aka *vector-valued functions*. For example,

$$\begin{aligned}\frac{dy_1}{dt} &= t^2 y_2 - y_1 + \sin(t)\\ \frac{dy_2}{dt} &= e^t y_1 - t y_2 + \cos(t)\end{aligned}$$

is a linear ODE system, since it can be written as

$$\left(\frac{d}{dt} - \begin{bmatrix} -1 & t^2 \\ e^t & -t \end{bmatrix}\right)\vec{y}(t) = \begin{bmatrix} \sin(t) \\ \cos(t) \end{bmatrix}.$$

(Convince yourself that those really are the same equation!) The general form for a linear first order ODE system is

$$\frac{d}{dt}\,\vec{y}(t) - \mathbf{A}(t)\,\vec{y}(t) = \vec{Q}(t),$$

where $\mathbf{A}(t)$ is a matrix of functions (or, if you prefer, a "matrix-valued function"). When $\vec{Q}(t)$ is zero, we call the linear ODE system *homogeneous*. When the matrix $\mathbf{A}(t)$ is a matrix of constants, we say the linear ODE system has *constant coefficients*.

Applying the same old concepts of linear algebra to homogeneous linear ODE systems $\mathcal{L}(\vec{y}(t)) = 0$, we have the following fact: The set of solutions is a vector space; i.e. any linear combination of solutions is a solution. Thus, one seeks a basis for this solution space, and then the general solution is an arbitrary linear combination of those basis solutions.

Our concern in these notes is to deal with first order homogeneous linear ODE systems with constant coefficients. We will focus on systems with only two unknown functions; much of the reasoning can be extended once we have that down. So the things we're interested in all have the form

$$\frac{d}{dt}\vec{y}(t) = \mathbf{A}\vec{y}(t)$$

for a constant 2×2 matrix $\mathbf{A}$. If $\mathbf{A}$ is $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, then the ODE system is

$$\begin{aligned} y_1' &= a_{11}y_1 + a_{12}y_2 \\ y_2' &= a_{21}y_2 + a_{22}y_2. \end{aligned}$$

2 Uncoupled Systems

Let's focus on a very specific case of the problem:

$$\mathbf{A} = \begin{bmatrix} a & 0 \\ 0 & -1 \end{bmatrix}.$$

The ODE system is now

$$\begin{bmatrix} y_1' \\ y_2' \end{bmatrix} = \begin{bmatrix} a & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \tag{1}$$

which becomes

$$\begin{aligned} y_1' &= ay_1 \\ y_2' &= -y_2. \end{aligned}$$

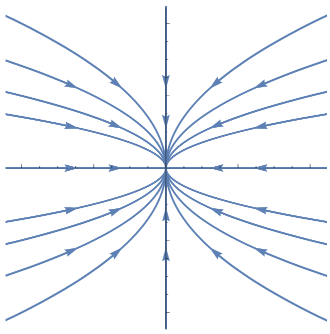
There is no y_2 in the y_1' equation, and no y_1 in the y_2' equation... these two equations are *uncoupled.* They are really just two individual ODEs, and we know how to solve them. The first has general solution $y_1 = c_1 e^{at}$ and the second has general solution $y_2 = c_2 e^{-t}$. The two-parameter family of solutions to the system can be described as

$$\vec{y}(t) = \begin{bmatrix} c_1 e^{at} \\ c_2 e^{-t} \end{bmatrix} = c_1 \begin{bmatrix} e^{at} \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ e^{-t} \end{bmatrix}.$$

The problem is technically solved now, but let's try to understand what the solution is *doing*. First, some conventions for picturing things: Think of t as time. As the system evolves in time, y_1 and y_2 change. Think of the pair (y_1, y_2) as describing the *state* of the system at any given time. That is, *points* in the y_1-y_2 plane represent possible states of the system. So a *trajectory* in the y_1-y_2 plane represents a possible time-evolution of the system. And the general solution of the system is a whole family of trajectories in the y_1-y_2 plane.

Now in order to picture what the system does, we need to know more about the constant a.

Assume that $a < -1$ in (1). The trajectories in y_1-y_2 space look like:



The horizantal axis is y_1 and the vertical axis is y_2; the exact dependence of these two functions on time is not really displayed in the picture. As $t \to \infty$, each of $y_1(t)$ and $y_2(t)$ decays to zero, hence the trajectories all heading towards the origin. As $t \to \infty$, each trajectory decays into the origin in a vertical manner, i.e. in a direction tangent to the y_2 axis. Why this shape of trajectories? It's because y_1 is decaying faster than y_2, due to our assumption that $a < -1$. A good way to understand the shape is to imagine that a is extremely negative– then y_1 would decay very quickly and we would see the curve almost slam into the y_2 axis, after which it would crawl down to the origin due to the slow decay of y_2. Looking at $t \to -\infty$, each trajectory comes in from a horizantal direction, i.e. parallel to the y_1 axis

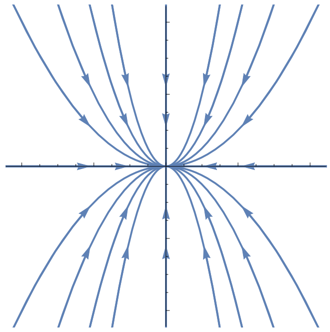
Notice the two very simple trajectories along the axes themselves. These correspond to

$$\begin{bmatrix} e^{at} \\ 0 \end{bmatrix} \text{ and } \begin{bmatrix} 0 \\ e^{-t} \end{bmatrix}.$$

The other trajectories are obtained as various linear combinations of those two simple ones.

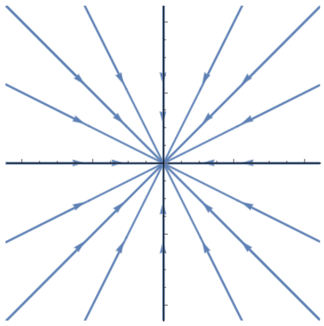
There is also an invisible trajectory which just stays put at the origin. Here, the origin is called a *stable node*. The origin is "stable" in the sense that trajectories that start near to the origin tend to the origin as $t \to \infty$.

Assume that $-1 < a < 0$ in (1). This time the trajectories in y_1-y_2 space look like:



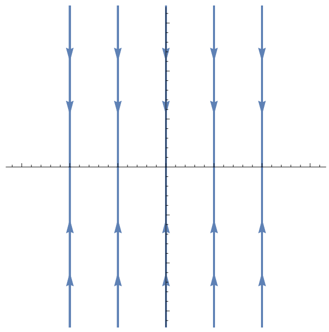
The origin here is also called a stable node, and in fact we're looking at basically the same picture as the previous one. The axes are just swapped.

Assume that $a = -1$ in (1). This is a special case, in which y_1 and y_2 decay at exactly the same rate:



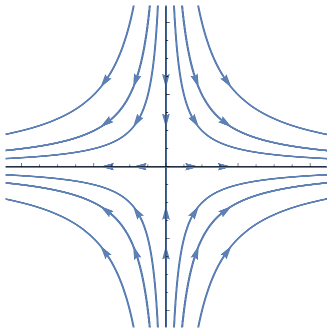
In addition to all the straight-line trajectories decaying towards the origin, there is also a fixed trajectory that stays put at the origin. The origin here is called a *stable star*.

Assume that $a = 0$ in (1). This is another special case. The solutions now all involve y_1 being some constant, with y_2 decaying as usual. So the trajectories just decay straight towards the y_1 axis:



Note that with all the previous cases, the origin was both the only location where there was a fixed trajectory, and it was the only point that *attracted* all trajectories. In this case, however, every point on the y_1 axis has a fixed trajectory on it. And every point on the y_1 axis attracts some of the trajectories near it.

Assume that $a > 0$ in (1). Now y_1 grows exponentially while y_2 decays:



The origin has become unstable. Almost all trajectories veer away from the origin towards larger and larger y_1. The exception is the trajectories that start with $y_1 = 0$; they travel precariously towards the origin along the y_2 axis. If you run time forward then you see all trajectories approaching the y_1 axis, and if you run time backward then you see all trajectories approaching the y_2 axis. In this situation, the origin is called a *saddle point*.

By the way, those pictures are called "phase portraits." In each picture the origin was a *fixed point*– it had a fixed trajectory sitting on it. We've named the type of fixed point in some of the pictures: stable node, stable star, and saddle point. There are also unstable nodes and unstable stars; they look the same as stable nodes and stable stars but with the arrows flipped around. And four more types of fixed points can occur: stable spirals, unstable spirals, centers, and degenerate nodes. Those did not make an appearance in our uncoupled example, but they are coming up.

3 Coupled Systems

Let's shift our focus back to the general problem:

$$\frac{d}{dt}\vec{y} = \mathbf{A}\vec{y}. \tag{2}$$

The obstacle here is that the equations are *coupled*; the y_1' equation can involve y_2 and the y_2' equation can involve y_1. However, we can still draw inspiration from the uncoupled situation. For one thing, there is still a fixed solution $\vec{y}(t) = \vec{0}$ at the origin. Perhaps there are also straight-line trajectories decaying towards or growing away from the origin? Let us search for such solutions; they would have the form

$$\vec{y}(t) = e^{\lambda t}\vec{\eta},$$

where $\vec{\eta}$ is a constant vector pointing in the direction of the straight line solution.

Substituting $e^{\lambda t}\vec{\eta}$ into $\frac{d\vec{y}}{dt} = \mathbf{A}\vec{y}$, we get

$$\lambda e^{\lambda t}\vec{\eta} = e^{\lambda t}A\vec{\eta}.$$

It's harmless to divide out the never-zero scalar function $e^{\lambda t}$:

$$\lambda\vec{\eta} = A\vec{\eta}.$$

Conclusion: $\vec{y}(t) = e^{\lambda t}\vec{\eta}$ is a straight-line solution to $\frac{d\vec{y}}{dt} = \mathbf{A}\vec{y}$ if and only if $\vec{\eta}$ is an eigenvector of $\mathbf{A}$ with eigenvalue λ!

This would be a good point for you to pause and review some linear algebra if you need a refresher. Here are some linear algebra facts that are good to have available:

- An $n \times n$ matrix $\mathbf{A}$ has at least one (possibly complex) eigenvalue, and it can have up to n eigenvalues.
- The characteristic polynomial of $\mathbf{A}$ is the polynomial $\det(\mathbf{A} - \lambda\mathbf{I})$ in the variable λ, where $\mathbf{I}$ is the identity matrix. The eigenvalues of $\mathbf{A}$ are the roots of this polynomial.
- The eigenvectors of $\mathbf{A}$ that have eigenvalue λ can be found by solving $(\mathbf{A} - \lambda I)\vec{\eta} = \vec{0}$. In other words, they are the nonzero vectors in the nullspace of $\mathbf{A} - \lambda I$.
- A collection of eigenvectors of $\mathbf{A}$ that have distinct eigenvalues are necessarily linearly independent (this is a theorem). So if an $n \times n$ matrix $\mathbf{A}$ has n distinct eigenvalues, then it has enough eigenvectors to make a *basis* for $\mathbb{R}^n$ out of them. But if $\mathbf{A}$ has repeated eigenvalues, then there may or may not be enough eigenvectors to make a basis.
- When there are enough eigenvectors of $\mathbf{A}$ to make a basis, then using that basis to represent $\mathbf{A}$ as a linear map yields a *diagonal* representation of $\mathbf{A}$. We then say that $\mathbf{A}$ is *diagonalizable.* Otherwise, we say that $\mathbf{A}$ is *defective.*

The possible scenarios for a 2×2 matrix $\mathbf{A}$ are then:

1. $\mathbf{A}$ has two distict and real eigenvalues λ_1, λ_2, with respective eigenvectors $\vec{\eta}_1$, $\vec{\eta}_2$.
2. Eigenvalues are repeated instead of distinct, but with there still being two linearly independent eigenvectors $\vec{\eta}_1$ and $\vec{\eta}_2$.
3. Eigenvalues are repeated instead of distinct, with a merely one-dimensional eigenspace (the "defective" case).
4. Eigenvalues are complex instead of real.

3.1 Cases 1 and 2: real eigenvalues, enough eigenvectors

In case 1, $e^{\lambda_1 t}\vec{\eta}_1$ and $e^{\lambda_2 t}\vec{\eta}_2$ are two linearly independent solutions of (2), and the general solution can be written as

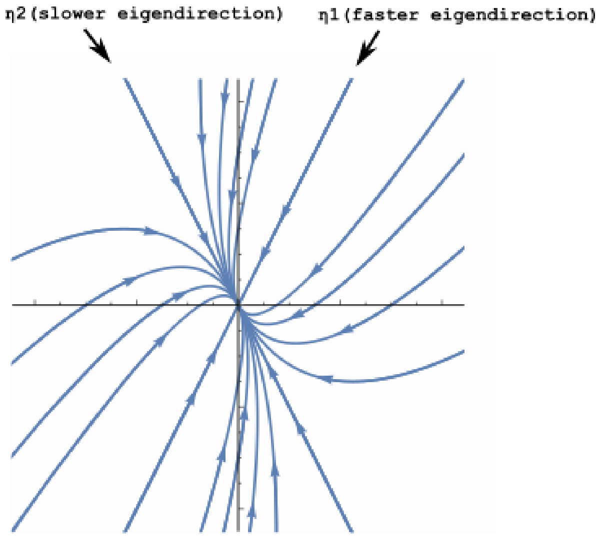
$$\vec{y}(t) = c_1 e^{\lambda_1 t}\vec{\eta}_1 + c_2 e^{\lambda_2 t}\vec{\eta}_2.$$

Case 2 is actually handled in exactly the same way; the general solution of (2) is

$$\vec{y}(t) = c_1 e^{\lambda t}\vec{\eta}_1 + c_2 e^{\lambda t}\vec{\eta}_2,$$

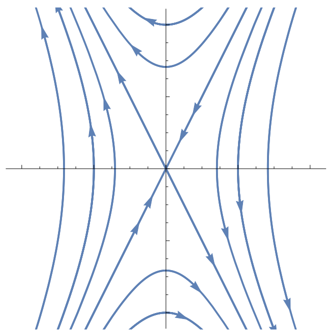
where λ is the single repeated eigenvalue.

Suppose we are in case 1, and $\lambda_1 < \lambda_2 < 0$. What is the picture? There are two "eigensolutions" with directions given by $\vec{\eta}_1$ and $\vec{\eta}_2$; they are $\vec{y} = e^{\lambda_1 t}\vec{\eta}_1$ and $\vec{y} = e^{\lambda_2 t}\vec{\eta}_2$. They both decay towards the origin, but the one along $\vec{\eta}_2$ decays more slowly. This looks a lot like the stable node we saw before. To draw the picture, use the same reasoning we used to picture the solutions of (1) in the case $a < -1$, but replace the y_1 and y_2 axes by the eigenvector directions. The following sort of phase portrait emerges:



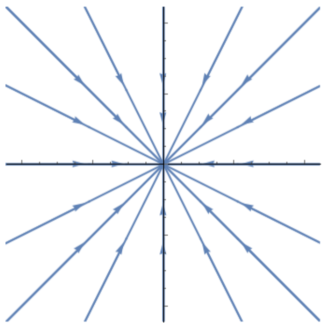
So a stable node is what you get when the eigenvalues are distinct and both negative. You can also get an unstable node– that happens when the eigenvalues are distinct and both positive.

How about when $\lambda_1 < 0 < \lambda_2$? Then one eigensolution dominates in forward time, and the other dominates in backward time. It's a saddle point! It's similar to what we we got in the phase portrait of (1) when we assumed $a > 0$. It can look something like



Can you tell which eigendirection in the picture is $\vec{\eta}_1$ and which is $\vec{\eta}_2$?

Suppose we are in case 2 from the numbered list above. Let λ be the repeating eigenvalue. In this case, *every* vector of $\mathbb{R}^2$ is an eigenvector of $\mathbf{A}$ with eigenvalue λ. Thus, $\vec{y}(t) = e^{\lambda t}\vec{v}$ is a solution for any vector $\vec{v}$ of $\mathbb{R}^2$. The phase portrait shows straight line solutions decaying towards or growing out from the origin. It's a star, stable when $\lambda < 0$ and unstable when $\lambda > 0$.



3.2 Case 4: nonreal (complex) eigenvalues

In this case, one still has the solutions $e^{\lambda_1 t}\vec{\eta}_1$ and $e^{\lambda_2 t}\vec{\eta}_2$, but they are *complex* solutions. To get real solutions out of them, we use the principle we established before in lecture:

If $\mathcal{L}$ is a linear map of real vector spaces and we find a complex solution $\vec{y} = \vec{y}_{re} + i\,\vec{y}_{im}$ to the homogeneous linear system $\mathcal{L}(\vec{y}) = \vec{0}$, then its real and imaginary parts, $\vec{y}_{re}$ and $\vec{y}_{im}$, are real solutions.

The two complex eigenvalues will come as a complex conjugate pair $\lambda_1 = \alpha + \beta i$, $\lambda_2 = \alpha - \beta i$. Similarly, the eigenvectors $\vec{\eta}_1$, $\vec{\eta}_2$ will have complex entries and will in fact come in a complex conjugate pair; so $\vec{\eta}_1 = \vec{u} + \vec{v}i$ and $\vec{\eta}_2 = \vec{u} - \vec{v}i$ for some real vectors $\vec{u}$ and $\vec{v}$. Using the identity $e^{a+bi} = e^a(\cos(b) + i\sin(b))$, the solution $e^{\lambda_1 t}\vec{\eta}_1$ is

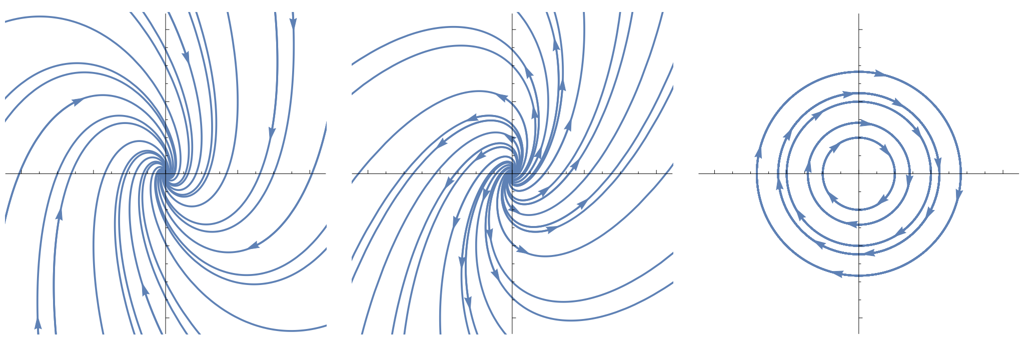
$$\begin{aligned} e^{\lambda_1 t}\vec{\eta}_1 &= e^{\alpha t}(\cos(\beta t) + i\sin(\beta t))(\vec{u} + i\vec{v}) \\ &= e^{\alpha t}(\cos(\beta t)\vec{u} - \sin(\beta t)\vec{v}) + ie^{\alpha t}(\cos(\beta t)\vec{v} + \sin(\beta t)\vec{u}). \end{aligned}$$

Picking out the real and imaginary parts, we obtain two linearly independent solutions of $\frac{d\vec{y}}{dt} = \mathbf{A}\vec{y}$:

$$\begin{aligned} \vec{y}(t) &= e^{\alpha t}(\cos(\beta t)\vec{u} - \sin(\beta t)\vec{v}), \\ \vec{y}(t) &= e^{\alpha t}(\sin(\beta t)\vec{u} + \cos(\beta t)\vec{v}). \end{aligned}$$

The solutions of $\frac{d\vec{y}}{dt} = \mathbf{A}\vec{y}$ are then linear combinations of those two.

Phase trajectories are inward spirals about the origin if $\alpha < 0$, outward spirals if $\alpha > 0$, and ellipses centered at the origin if $\alpha = 0$. Example pictures:

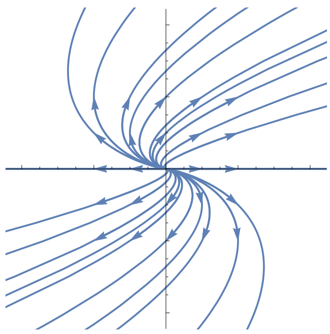


There is a fixed point at the origin in each case, and it is referred to as a *stable spiral*, *unstable spiral*, or *center* respectively.

3.3 Case 3: repeated eigenvalues, not enough eigenvectors

Recall our earlier conclusion: $\vec{y}(t) = e^{\lambda t}\vec{\eta}$ is a straight-line solution of (2) if and only if $\vec{\eta}$ is an eigenvector of $\mathbf{A}$ with eigenvalue λ. In case 3, there simply aren't enough straight-line solutions to form a basis for the solution space of (2)! There is only one, and we need two.

I'm punting on writing the general solution of (2) in this case; it will come later. You might like to see a typical phase portrait of it though:



The one eigendirection in the picture is horizantal. This picture is something between a node and a spiral. You can see it's kind of trying to be a spiral, but doesn't quite make it. The origin here is called a *degenerate node*. Degenerate nodes can be stable or unstable (the one pictured is clearly unstable).

4 Exercises

Exercise 1: For each, write the general solution and sketch and describe the phase portrait:

a. $y_1' = y_2$, $y_2' = -2y_1 - 3y_2$

b. $y_1' = 4y_1 - y_2$, $y_2' = 2y_1 + y_2$

c. $x' = -3x + 4y$, $y' = -2x + 3y$

d. $x' = 5x + 10y$, $y' = -x - y$

Exercise 2: Linear second order ODEs, which look like

$$f_2(t)y'' + f_1(t)y' + f_0(t)y = q(t),$$

can be viewed as *first order* linear ODE *systems*! Name a new variable v and let it stand for y'. Then $v = y'$ is one ODE in the new system. Determine the other ODE in the system and write out the system using a matrix.

Exercise 3: Apply the method of exercise 2 to solve the simple harmonic oscillator system

$$y'' + \omega_0^2 y = 0.$$

Sketch and describe the phase portrait.

Exercise 4: Apply the method of exercise 2 to solve the damped harmonic oscillator system

$$y'' + 2ry' + \omega_0^2 y = 0.$$

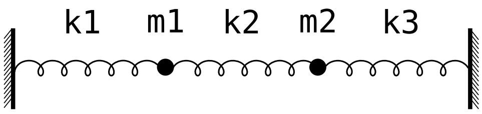
Describe the phase portrait. There should be different pictures based on the values of the constants r and ω_0.

Exercise 5: Recall that we found the equation of motion for a pendulum to be

$$\theta'' + \frac{g}{\ell}\sin(\theta) = 0$$

before we passed to a linear approximation. Even though this is a *nonlinear* second order ODE, try to use a method like in exercise 2 to write it as a first order ODE system. Can you write it out using a matrix?

Exercise 6: Consider a setup consisting of two point masses and three frictionless springs arranged as follows:



Each m_i denotes a mass and each k_i denotes a spring stiffness constant. For simplicity, assume that the distance between the fixed side walls equals the sum of the natural lengths of the springs. Model this setup with an ODE system.

Exercise 7: Consider the ODE system

$$\frac{d\vec{y}}{dt} = \begin{bmatrix} \lambda & b \\ 0 & \lambda \end{bmatrix} \vec{y}$$

with b nonzero. This should give a degenerate node, but you won't need any fancy tools to solve it.

a. Show that the matrix $\mathbf{A} = \begin{bmatrix} \lambda & b \\ 0 & \lambda \end{bmatrix}$ has a repeated eigenvalue λ with the corresponding eigenspace being only one-dimensional.

b. Solve the system. (Hint: It is almost uncoupled.)

Exercise 8: The following exercises are due to Steven Strogatz and appear in his book, *Nonlinear Dynamics and Chaos.*

a. Romeo is in love with Juliet, but Juliet is a fickle lover. The more Romeo loves her, the more Juliet wants to run away and hide. But when Romeo gets discouraged and backs off, Juliet begins to find him strangely attractive. Romeo, on the other hand, tends to echo her: he warms up when she loves him, and grows cold when she hates him. Let

$$R(t) = \text{Romeo's love/hate for Juliet at time t.}$$
$$J(t) = \text{Juliet's love/hate for Romeo at time t.}$$

Positive values of R, J signify love, and negative values signify hate. Then a model for their romance is

$$R' = aJ$$
$$J' = -bR$$

where the parameters a and b are positive. What happens?

b. Consider two identically cautious lovers. The system is

$$R' = -aR + bJ$$
$$J' = bR - aJ$$

where a and b are positive. Here a is a measure of cautiousness (they each try to avoid throwing themselves at each other), and b is a measure of responsiveness (they both get excited by the other's advances). What happens?

5 Matrix Exponentiation

Consider the 1×1 case of (2):

$$\frac{dy}{dt} = Ay.$$

Here, A is just a number. We know the solution well:

$$y = e^{At}.$$

You may be used to thinking about the exponential function e^{At} as "the $(At)^{\text{th}}$ power of the magical real number $e = \frac{1}{0!} + \frac{1}{1!} + \cdots$." Another viewpoint, one that better captures the spirit of exponentiation, is that e^{At} means "the solution of $y' = Ay$ with $y(0) = 1$." You can actually *discover* the exponential this way! First, guess a solution to $y' = Ay$ that has a taylor series at the origin:

$$y(t) = c_0 + c_1 t + c_2 t^2 + \cdots .$$

Then substitute it into the ODE to get a left hand side of

$$y'(t) = c_1 + 2c_2 t + \cdots$$

and a right hand side of

$$Ay(t) = Ac_0 + Ac_1 t + \cdots .$$

The initial condition $y(0) = 1$ forces $c_0 = 1$. Equate coefficients on the left and right hand sides to get a bunch of equations involving the c_n's. You can deduce from these equations that $c_n = \frac{A^n}{n!}$. Try it!

You can guess where this is going. *Apply the same principle to the ODE system:*

$$\frac{d\vec{y}}{dt} = \mathbf{A}\vec{y}.$$

5.1 Definition

To make the solution unique, fix an initial condition $\vec{y}(0) = \vec{y}_0$. Now guess a solution that has a taylor series at the origin. This means that each component function has a taylor series at zero, so we could write it as:

$$\vec{y}(t) = \vec{c}_0 + t\vec{c}_1 + t^2\vec{c}_2 + \cdots .$$

The initial condition gives $\vec{c}_0 = \vec{y}_0$. Substitute into the ODE system to get a left hand side of

$$\vec{y}\,'(t) = \vec{c}_1 + 2t\vec{c}_2 + 3t^2\vec{c}_3 + \cdots$$

and a right hand side of

$$\mathbf{A}\vec{y}(t) = \mathbf{A}\vec{c}_0 + t\mathbf{A}\vec{c}_1 + t^2\mathbf{A}\vec{c}_2 + \cdots .$$

Equating coefficients of the powers of t and adding in $\vec{c}_0 = \vec{y}_0$, we obtain the equations

$$\begin{aligned}
\vec{c}_0 &= \vec{y}_0 \\
\vec{c}_1 &= \mathbf{A}\vec{c}_0 \\
\vec{c}_2 &= \frac{1}{2}\mathbf{A}\vec{c}_1 \\
\vec{c}_3 &= \frac{1}{3}\mathbf{A}\vec{c}_2 \\
&\vdots \qquad .
\end{aligned}$$

From these you can deduce

$$\vec{c}_n = \frac{1}{n!}\mathbf{A}^n\vec{y}_0,$$

so the solution of $\vec{y}\,' = \mathbf{A}\vec{y}$ with initial condition $\vec{y}(0) = \vec{y}_0$ is

$$\vec{y}(t) = \vec{y}_0 + t\mathbf{A}\vec{y}_0 + \frac{1}{2!}t^2\mathbf{A}^2\vec{y}_0 + \frac{1}{3!}t^3\mathbf{A}^3\vec{y}_0 + \cdots .$$

Ripping out the $\vec{y}_0$ from the series, we can view this as a series of matrices being applied to the initial condition $\vec{y}_0$,

$$\vec{y}(t) = \left(I + t\mathbf{A} + \frac{1}{2!}t^2\mathbf{A}^2 + \frac{1}{3!}t^3\mathbf{A}^3 + \cdots\right)\vec{y}_0\,,$$

where I is the identity matrix. The matrix series in parentheses turns out to converge no matter what $\mathbf{A}$ is. Evaluate that stuff in parentheses at $t = 1$; we call the result $e^{\mathbf{A}}$:

$$e^{\mathbf{A}} = I + \mathbf{A} + \frac{1}{2!}\mathbf{A}^2 + \frac{1}{3!}\mathbf{A}^3 + \cdots . \qquad (3)$$

This is the definition of the matrix exponential. In this notation, solution to the ODE system $\vec{y}' = \mathbf{A}\vec{y}$ with initial condition $\vec{y}(0) = \vec{y}_0$ is

$$\vec{y}(t) = e^{t\mathbf{A}}\vec{y}_0.$$

We can say this is the *general* solution of $\vec{y}' = \mathbf{A}\vec{y}$ if we interpret $\vec{y}_0$ as being an arbitrary vector.

5.2 Properties

Let's be clear about the *types* of objects being tossed around here. The symbol $\mathbf{A}$ stands for a matrix. And $t\mathbf{A}$ is a scalar multiple of that matrix. The powers of $\mathbf{A}$ appearing in (3) are *matrix powers*; i.e. you would have to use matrix multiplication to compute them. Exponentials of matrices like $e^{\mathbf{A}}$ and $e^{t\mathbf{A}}$ are then themselves matrices.

The matrix exponential $e^{\mathbf{A}}$ shares some but not all properties of the ordinary exponential. The primary troublemaker in terms of getting all the familiar properties of exponentiation is the fact that matrix multiplication is *noncommutative*: the product $\mathbf{AB}$ need not equal the product $\mathbf{BA}$. Let me list some properties without proof.

- $e^0 = I$
- $e^{t\mathbf{A}}e^{s\mathbf{A}} = e^{(t+s)\mathbf{A}}$
- $e^{-\mathbf{A}}$ is the inverse of $e^{\mathbf{A}}$. (So $e^{-\mathbf{A}}e^{\mathbf{A}} = I$.)
- If $\mathbf{AB} = \mathbf{BA}$, then $e^{\mathbf{A}}e^{\mathbf{B}} = e^{\mathbf{B}}e^{\mathbf{A}} = e^{\mathbf{A}+\mathbf{B}}$.
- If $\mathbf{P}$ is invertible, then $e^{\mathbf{PAP}^{-1}} = \mathbf{P}e^{\mathbf{A}}\mathbf{P}^{-1}$

The convention here is that boldface capital letters stand for matrices, and lowercase letters stand for scalars.

And of course we have the property

- $\frac{d}{dt}e^{t\mathbf{A}} = \mathbf{A}e^{t\mathbf{A}}$

which was the whole point of all this.

5.3 Exercises

Exercise 9: Use (3) to directly show that $e^{\mathbf{PAP}^{-1}} = \mathbf{P}e^{\mathbf{A}}\mathbf{P}^{-1}$.

Exercise 10: Use (3) to directly show that $\frac{d}{dt}e^{t\mathbf{A}} = \mathbf{A}e^{t\mathbf{A}}$.

5.4 Computing $e^{\mathbf{A}}$

Computing $e^{\mathbf{A}}$ using (3) directly is an unpleasant task. Powers of matrices get pretty complicated. Try just computing

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^2$$

in case you've forgotten!

However there is one situation where matrix powers are simple: diagonal matrices. One has

$$\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}^n = \begin{bmatrix} \lambda_1^n & 0 \\ 0 & \lambda_2^n \end{bmatrix}.$$

Using this property and the definition (3), we have

$$e^{\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}} = \begin{bmatrix} e^{\lambda_1} & 0 \\ 0 & e^{\lambda_2} \end{bmatrix}. \tag{4}$$

Great, we can compute exponentials of diagonal matrices. If we can *diagonalize* a matrix, this gives us a way to compute its exponential. Recall that to diagonalize a matrix $\mathbf{A}$ is to find an invertible matrix $\mathbf{P}$ such that $\mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ is diagonal. If you are able to do this for $\mathbf{A}$, then you can compute its exponential using the fact that $e^{\mathbf{P}^{-1}\mathbf{A}\mathbf{P}} = \mathbf{P}^{-1}e^{\mathbf{A}}\mathbf{P}$. Solving $e^{\mathbf{P}^{-1}\mathbf{A}\mathbf{P}} = \mathbf{P}^{-1}e^{\mathbf{A}}\mathbf{P}$ for $e^{\mathbf{A}}$,

$$e^{\mathbf{A}} = \mathbf{P}e^{\mathbf{P}^{-1}\mathbf{A}\mathbf{P}}\mathbf{P}^{-1}. \tag{5}$$

For example suppose that

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix}.$$

Recall that diagonalization is a matter of solving the eigenvalue problem. We find eigenvalues $\lambda_1 = 2$ and $\lambda_2 = -3$, with eigenvectors

$$\vec{\eta}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \ \vec{\eta}_2 = \begin{bmatrix} -1 \\ 2 \end{bmatrix}.$$

Therefore the matrix

$$\mathbf{P} = \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}$$

serves to convert $\mathbf{A}$ into a diagonal matrix whose diagonal entries are λ_1, λ_2:

$$\mathbf{A} = \mathbf{P}\begin{bmatrix} 2 & 0 \\ 0 & -3 \end{bmatrix}\mathbf{P}^{-1}.$$

We compute $\mathbf{P}^{-1}$ to be

$$\mathbf{P}^{-1} = \frac{1}{5}\begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix}.$$

Hence

$$\begin{aligned}
e^{\mathbf{A}} &= e^{\mathbf{P}\left[\begin{smallmatrix} 2 & 0 \\ 0 & -3 \end{smallmatrix}\right]\mathbf{P}^{-1}} \\
&= \mathbf{P}e^{\left[\begin{smallmatrix} 2 & 0 \\ 0 & -3 \end{smallmatrix}\right]}\mathbf{P}^{-1} \\
&= \frac{1}{5}\begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} e^2 & 0 \\ 0 & e^{-3} \end{bmatrix}\begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix} \\
&= \begin{bmatrix} \frac{4}{5}e^2 + \frac{1}{5}e^{-3} & -\frac{2}{5}e^2 + \frac{2}{5}e^{-3} \\ -\frac{2}{5}e^2 + \frac{2}{5}e^{-3} & \frac{1}{5}e^2 + \frac{4}{5}e^{-3} \end{bmatrix}
\end{aligned}$$

Okay it is quite tedious. But doable. More importantly, you could instruct a computer to do it.

5.5 Exercises

Exercise 11: Compute $e^{\left[\begin{smallmatrix} 4 & -1 \\ -1 & 4 \end{smallmatrix}\right]}$.

Exercise 12: Compute $e^{\left[\begin{smallmatrix} 0 & t \\ -t & 0 \end{smallmatrix}\right]}$, where t is a real number. Eigenvalues are complex, but just go with it. The answer should come out real!

Exercise 13: Use matrix exponentiation to find the general solution of

$$\begin{aligned} x' &= 4x - y \\ y' &= 4y - x. \end{aligned}$$

Hint: with some modifications, you can almost use your answer to exercise 11.

Exercise 14: Find the particular solution of the ODE system in exercise 13 that satisfies the initial condition $x(0) = 1$, $y(0) = 2$.

Exercise 15: Use matrix exponentiation to find the general solution of

$$\begin{aligned} x' &= y \\ y' &= -x. \end{aligned}$$

Hint: You can use your answer to exercise 12.

Exercise 16: Find the particular solution of the ODE system in exercise 15 that satisfies the initial condition $x(0) = -2$, $y(0) = 1$.

Exercise 17: What if a matrix isn't diagonalizable? Try to use the usual technique to compute the exponential of

$$\begin{bmatrix} 5 & -1 \\ 1 & 3 \end{bmatrix}.$$

Do you see the issue? Linear algebra theory still gives us a way to compute the exponential, and we will be able to use it to deal with the situation that was deferred back in section 3.3 of these notes. We will see it in lecture soon.

6 Analyzing Nonlinear Systems

It is possible to analyze certain nonlinear systems by puzzling together a picture from their linearizations at fixed points.

Here is a compact notation for a first order ODE system:

$$\frac{d\vec{y}}{dt} = \vec{F}(\vec{y}, t).$$

We won't tackle quite this level of generality, but close:

$$\frac{d\vec{y}}{dt} = \vec{F}(\vec{y}). \tag{6}$$

We will try to analyze nonlinear systems in which there is no *explicit* time dependence. "Analyze" does not mean "solve" here. Nonlinear systems are usually hard to solve and often impossible to solve in any reasonable way. Worse, even when a solution is available it is often too complicated-looking to offer any insight. No, we are not interested in solutions, we are interested in getting at *qualitative feautres* of nonlinear systems. Chief among these telling features are fixed points and their stability type.

A *fixed point* is a constant $\vec{y}^*$ such that $\vec{y}(t) = \vec{y}^*$ works as a constant solution to (6). In other words, fixed points are those points in phase space on which there is a fixed trajectory. Substituting a constant $\vec{y}^*$ into (6), we see that $\vec{y}^*$ is a fixed point if and only if

$$\vec{F}(\vec{y}^*) = \vec{0}.$$

The technique we will use is this: Locate all the fixed points by solving $\vec{F}(\vec{y}^*) = \vec{0}$, linearize $\vec{F}$ at each fixed point to obtain homogeneous linear ODEs, use each linearization to determine the local picture near the fixed points, and finally try to glue together the local pictures to understand what is happening. We will need to understand how to work with functions like $\vec{F}(\vec{y})$.

6.1 Vector Fields

The function $\vec{F}(\vec{y})$ is a function of n variables $y_1, \ldots, y_n$, with n components. When we write $\vec{F}(\vec{y})$, it is a shorthand for

$$\begin{bmatrix} F_1(y_1, \cdots, y_n) \\ \vdots \\ F_n(y_1, \cdots, y_n) \end{bmatrix}.$$

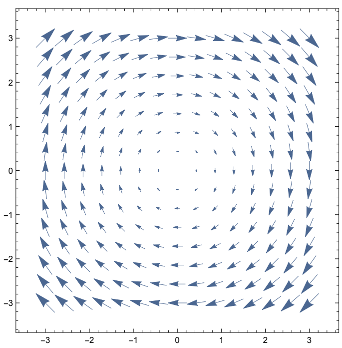
Such a function is called a *vector field.* It's domain is the *phase space of the ODE.* Input a phase point $\vec{y}$, and what comes out is a vector "at that point". Technically what comes out is a vector in $\mathbb{R}^n$, but it helps to imagine that the vector emanates from $\vec{y}$ itself in the phase space. Given a state $\vec{y}$, the vector $\vec{F}(\vec{y})$ should be thought of as the direction in which the state evolves from $\vec{y}$.

Solutions of the ODE $\frac{d\vec{y}}{dt} = \vec{F}(\vec{y})$ are nothing but trajectories $\vec{y}(t)$ whose velocity $\frac{d\vec{y}}{dt}$ is always the one prescribed by the vector field $\vec{F}$.

For example, we could picture the vector field

$$\vec{F}(y_1, y_2) = \begin{bmatrix} y_2 \\ -y_1 \end{bmatrix}$$

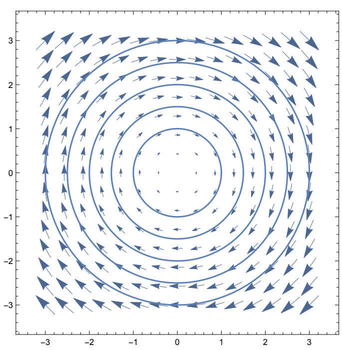
as a bunch of little velocity vectors peppered on the y_1-y_2 plane:



And solutions of

$$\begin{aligned} y_1' &= y_2 \\ y_2' &= -y_1 \end{aligned}$$

are just curves that follow the "flow" of those arrows:



Solving ODEs is nothing but figuring out the flows of vector fields.

(Actually, this idea suggests another way to analyze ODEs without solving them: plot the vector field $\vec{F}(\vec{y})$! That's always doable in two dimensions at least. If you draw enough arrows, you can kind of guess what the flow will look like.)

6.2 Derivative of $\vec{F}$

To linearly approximate a real-valued function of one variable, we take its derivative. To linearly approximate a vector-valued function of multiple variables, we also take its derivative– but what could this mean?

One way to interpret what a derivative *is* is to think of it as the *linear part of a taylor series*. Consider a nice real-valued function $f(x)$ of a single real variable x, and think about its derivative $f'(x_0)$ at a specific point x_0 in its domain. One feature of this *number* $f'(x_0)$, is that it appears in the following approximation of f near x_0:

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + [\text{higher degree terms}].$$

Think of this as *definining* the number $f'(x_0)$– it is that number which, when multiplied by $(x - x_0)$, yields $f(x) - f(x_0)$ up to terms of degree two or more in $(x - x_0)$.

Now try to apply the same thought to $\vec{F}(\vec{y})$. Focus on a point $\vec{y}_0$. Whatever $\vec{F}\,'(\vec{y}_0)$ is, it should be something that, when multiplied by the *vector* $(\vec{y} - \vec{y}_0)$, yields $\vec{F}(\vec{y}) - \vec{F}(\vec{y}_0)$ up to terms of higher degree. What *type* of thing could be multiplied by a vector $(\vec{y} - \vec{y}_0)$ to yield a vector $\vec{F}(\vec{y}) - \vec{F}(\vec{y}_0)$? The answer: an $n \times n$ matrix. The derivative of a vector-valued function $\vec{F}$ of n variables with n components is the $n \times n$ matrix $\vec{F}\,'(\vec{y}_0)$ for which

$$\vec{F}(\vec{y}) = \vec{F}(\vec{y}_0) + \vec{F}\,'(\vec{y}_0)(\vec{y} - \vec{y}_0) + [\text{higher degree terms}]. \tag{7}$$

This matrix is also called the *total derivative* or the *Jacobian derivative* of $\vec{F}$. Another great name for it is the *linearization of* $\vec{F}$, since it is giving us the closest linear map to the behavior of $\vec{F}$.

You may learn much more about this derivative later, but for now we just need to know how to compute it. The formula for a two-component function of two variables is

$$\vec{F}\,'(\vec{y}) = \begin{bmatrix} \frac{\partial F_1}{\partial y_1} & \frac{\partial F_1}{\partial y_2} \\ \frac{\partial F_2}{\partial y_1} & \frac{\partial F_2}{\partial y_2} \end{bmatrix}$$

where the round d symbol "∂" denotes a partial derivative. Partial derivatives are derivatives in which all variables are treated like constants except the one with respect to which the derivative is being taken. So $\frac{\partial F_1}{\partial y_1}$ means "Take the derivative of F_1 with respect to y_1, treating y_2 as though it were a constant."

Try to convince yourself that the derivative of

$$\vec{F}(x, y) = \begin{bmatrix} x^2y^3 + e^y \\ \sin(x)\sin(y) \end{bmatrix}$$

is

$$\vec{F}\,'(x, y) = \begin{bmatrix} 2xy^3 & 3x^2y^2 + e^y \\ \cos(x)\sin(y) & \sin(x)\cos(y) \end{bmatrix}.$$

6.3 Linearization of $\vec{y}\,' = \vec{F}(\vec{y})$

Suppose you have the ODE system $\vec{y}\,' = \vec{F}(\vec{y})$ and you compute that $\vec{y}^*$ is a fixed point; i.e. you find that $\vec{F}(\vec{y}^*) = 0$.

Shift coordinates of the state space so that the fixed point is at the origin:

$$\vec{u} = \vec{y} - \vec{y}^*.$$

Now the fixed point corresponds to $\vec{u} = \vec{0}$, the origin of the $\vec{u}$ coordinate system. Taking derivatices, we have

$$\vec{u}' = \vec{y}',$$

so they ODE system $\vec{y}' = \vec{F}(\vec{y})$ becomes

$$\vec{u}' = \vec{F}(\vec{u} + \vec{y}^*) \tag{8}$$

in terms of the new coordinates. Let's approximate $\vec{F}$ near the fixed point $\vec{y} = \vec{y}^*$. Writing out a taylor approximation as in (7), we have

$$\begin{aligned} \vec{F}(\vec{u} + \vec{y}^*) &= \vec{F}(\vec{y}^*) + \vec{F}'(\vec{y}^*)\vec{u} + [\text{higher degree terms}] \\ &= \vec{F}'(\vec{y}^*)\vec{u} + [\text{higher degree terms}] \\ &\approx \vec{F}'(\vec{y}^*)\vec{u} \end{aligned} \tag{9}$$

so (8) becomes

$$\vec{u}' \approx \vec{F}'(\vec{y}^*)\vec{u}.$$

Note that $\vec{F}'(\vec{y}^*)$ is just a constant matrix, so $\vec{u}' = \vec{F}'(\vec{y}^*)\vec{u}$ is linear homogeneous ODE with constant coefficients!

Conclusion: *The phase portrait of $\vec{y}' = \vec{F}(\vec{y})$ near a fixed point $\vec{y}^*$ looks similar to the phase portrait of $\vec{u}' = \vec{F}'(\vec{y}^*)\vec{u}$ at the origin.*

For example, if $\vec{F}'(\vec{y}^*)$ is a matrix with one positive eigenvalue and one negative eigenvalue, then the phase portrait of $\vec{y}' = \vec{F}(\vec{y})$ near $\vec{y}^*$ looks like a saddle point centered at $\vec{y}^*$.

6.3.1 Neglecting Nonlinear Terms

How safe is it to neglect the higher degree terms in (9)? Typically, the *qualitative* behavior of $\vec{y}' = \vec{F}(\vec{y})$ near fixed point is correctly predicted by the linearization. But there is a small exception with "borderline" cases. For example, a center is a borderline case between unstable and stable spirals, and one of the nonlinear terms we neglected could turn a center into a spiral. Another subtlety we've been ignoring is the case of non-isolated fixed points, fixed points that have other fixed points arbitrarily close to them (we saw this happen once while examining (1) in the case $a = 0$). The rules are:

- If the linearized system $\vec{u}' = \vec{F}'(\vec{y}^*)\vec{u}$ predicts a saddle, node, or spiral, then the fixed point really is a saddle, node, or spiral respectively.
- Not necessarily so for centers, degenerate nodes, stars, and non-isolated fixed points.

6.4 The Technique

Let's do it. We will take a nonlinear ODE of the form $\frac{d\vec{y}}{dt} = \vec{F}(\vec{y})$, locate its fixed points, classify them, and build a picture of what is going on.

do pendulum system... this still needs to be typed up... next time I teach this material