SYSTEMS OF DIFFERENTIAL EQUATIONS

ALLAN YASHINSKI

ABSTRACT. We solve systems of linear differential equations using matrix exponentials and diagonalizability. Heat diffusion on a graph is considered as a motivating example.

1. A Heat Diffusion Problem

We shall consider a discretized version of the classical continuous heat problem on a metal rod. In the continuous version, we have a long thin metal rod whose temperature is non-uniform. We consider the temperature $u(x, t)$ of the rod as a function of both the position x (a point on the rod) and of time t. One also imposes *boundary conditions* on the endpoints, for example we'll assume that the endpoints of the rod are held fixed with $u = 0$. The function $u(x, t)$ is governed by the second-order partial differential equation

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2},$$

which is called the *heat equation*. (In reality, there may be some physical constants mixed into the equation.)

We will not study this partial differential equation. Instead of considering a continuous rod, we will imagine that it has been divided into n discrete pieces, say of equal size, ordered from left to right. Our simplifying assumption is that at any given time t, each piece will have a uniform temperature. The temperatures of the pieces will be denoted $u_1(t), \ldots, u_n(t)$. We will also keep our boundary conditions, which say that the ends of the rods are held fixed at $u = 0$. For us, this does not mean that $u_1(t) = 0$ and $u_n(t) = 0$. Instead, we imagine that temperature is 0 to the left of the first piece and to the right of the n-th piece.

Let's focus on the case where $n = 2$, depicted in the figure. The main principle

$$u = 0 \mapsto \overset{\displaystyle u_1(t)}{\rule{4cm}{0.4pt}} \quad \overset{\displaystyle u_2(t)}{} \mapsto u = 0$$

is that the rate at which heat flows across a boundary (either the two ends of the rod or the boundary between the two pieces) is proportional to the temperature difference of the two regions at the boundary. For simplicity, we will assume this rate is *equal* to the temperature difference (proportional with constant 1). For example, heat leaves the left endpoint of the first region at a rate of $u_1(t)$, and it leaves the right endpoint of the first region (and enters the second) at a rate of $u_1(t) - u_2(t)$ (if this is negative, heat is actually flowing into the first region.) This leads to the differential equation

$$u_1'(t) = -u_1(t) - (u_1(t) - u_2(t)) = -2u_1(t) + u_2(t).$$

Note that we had to include negative signs because we were discussing the rate at which heat was *leaving* the first region.

Our assumption is that the temperature $u = 0$ is fixed at the ends. It may be helpful to imagine this is a region outside of the rod, held fixed at $u = 0$. Heat is flowing out of the rod and to the outside, but it is not raising the temperature outside.

In a similar way, we can consider how heat enters/leaves the second region. It flows out the right of the second region at a rate of $u_2(t)$ and it flows *in* from the first region at a rate of $u_1(t) - u_2(t)$, as we already remarked. So we obtain

$$u_2'(t) = (u_1(t) - u_2(t)) - u_2(t) = u_1(t) - 2u_2(t).$$

Altogether, we have obtained a system of differential equations

$$\begin{cases} u_1'(t) = -2u_1(t) + u_2(t) \\ u_2'(t) = u_1(t) - 2u_2(t) \end{cases}$$

The goal is to solve for the functions $u_1(t)$ and $u_2(t)$, that is, to give explicit formulas for them. We store both functions in a vector-valued function

$$\mathbf{u}(t) = \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix}.$$

The system of differential equations tells us about the derivative $\mathbf{u}'(t)$. (Note that to differentiate a vector-valued function, we just differentiate each entry). We have

$$\mathbf{u}'(t) = \begin{bmatrix} u_1'(t) \\ u_2'(t) \end{bmatrix} = \begin{bmatrix} -2u_1(t) + u_2(t) \\ u_1(t) - 2u_2(t) \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix} \mathbf{u}(t).$$

As it stands, the problem is incomplete because we need to specify an *initial condition*, which are values of the functions u_1 and u_2 when $t = 0$. For example, suppose that $u_1(0) = 10$ and $u_2(0) = 0$. Then we have an initial condition

$$\mathbf{u}(0) = \begin{bmatrix} u_1(0) \\ u_2(0) \end{bmatrix} = \begin{bmatrix} 10 \\ 0 \end{bmatrix}.$$

When we solve this *initial value problem* (this means differential equation along with initial condition), we will obtain explicit formulas for $u_1(t)$ and $u_2(t)$ which describe how heat flows as a function of time. Intuitively, we expect heat to flow out of the first region and into the second, but also out both end points. Eventually, all heat should leave the rod.

Before learning techniques to solve this, let's consider the same problem, but with a rod divided into four regions ($n = 4$). Here, we keep track of four functions

$u_1(t), u_2(t), u_3(t), u_4(t)$ for the temperatures of four regions. When heat leaves this rod, it only does so from region 1 or region 4. The differential equations are

$$\begin{cases} u_1'(t) = -2u_1(t) + u_2(t) \\ u_2'(t) = u_1(t) - 2u_2(t) + u_3(t) \\ u_3'(t) = u_2(t) - 2u_3(t) + u_4(t) \\ u_4'(t) = u_3(t) - 2u_4(t) \end{cases}$$

If we let $\mathbf{u}(t) = \begin{bmatrix} u_1(t) \\ u_2(t) \\ u_3(t) \\ u_4(t) \end{bmatrix}$, then we obtain

$$\mathbf{u}'(t) = \begin{bmatrix} -2 & 1 & 0 & 0 \\ 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 1 & -2 \end{bmatrix} \mathbf{u}(t).$$

2. Solving with Matrix Exponentials

Let A be an $n \times n$ matrix and let $\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix}$ be a vector-valued function with n "unknown" functions. We'd like to solve the general initial value problem

$$\mathbf{x}'(t) = A\mathbf{x}(t), \qquad \mathbf{x}(0) = \mathbf{x}_0,$$

where $\mathbf{x}_0$ is some given vector in $\mathbb{R}^n$ (the "initial conditions"). We can take a little inspiration from the $n = 1$ case, which we just simply have a differential equation

$$x'(t) = ax(t), \qquad x(0) = x_0,$$

where a is a constant. The function $x(t) = e^{at}$ clearly satisfies this differential equation, but more generally so does every function of the form $x(t) = Ce^{at}$, where C is an arbitrary constant. The constant C is determined by the initial condition, and in fact $x(t) = x_0 e^{at}$ is the unique function that satisfies the differential equation along with the initial condition.

Let's return to the system of differential equations $\mathbf{x}'(t) = A\mathbf{x}(t)$. Could it be possible that e^{At} is a solution? But wait, what does that even mean?

It turns out, there is something of value here, but we need to figure out what it means. Recall that for any real number x, the exponential function is given by the Taylor series

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \dots$$

More specifically, this series converges for every number x, and the value of the sum is precisely e^x. Following this, we'll define a matrix exponential.

Definition 2.1. Let A be an $n \times n$ matrix. We define the *matrix exponential* of A to be

$$e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!} = I_n + A + \frac{1}{2}A^2 + \frac{1}{6}A^3 + \frac{1}{24}A^4 + \dots$$

Note that we use the identity matrix I_n in place of 1 in the original Taylor series. The matrix A was assumed to be square, so it is possible to compute powers A^k of A. The formula appears to make sense, apart from the natural issue of convergence.

Proposition 2.2. *For any $n \times n$ matrix A, the series defining e^A converges. Hence e^A is defined for any $n \times n$ matrix A.*

Note that if A is an $n \times n$ matrix, then e^A is some other $n \times n$ matrix.

Example 2.3. Let $\mathbf{0}_{n \times n}$ denote the $n \times n$ matrix of all 0's. Then by defintion,

$$e^{\mathbf{0}_{n \times n}} = I_n + \mathbf{0}_{n \times n} + \frac{1}{2}\mathbf{0}_{n \times n}^2 + \frac{1}{6}\mathbf{0}_{n \times n}^3 + \frac{1}{24}\mathbf{0}_{n \times n}^4 + \ldots = I_n.$$

This shows that $e^{\mathbf{0}_{n \times n}} = I_n$, just like how $e^0 = 1$.

We will be interested in considering e^{tA}, where t is a real variable. Note that e^{tA} is a matrix-valued function of the real variable t. A matrix-valued function can be differentiated much like a vector-valued function (just differentiate each entry).

Proposition 2.4. *Let A be an $n \times n$ matrix. Then*

$$\frac{d}{dt}\left[e^{tA}\right] = Ae^{tA}.$$

Proof. We'll give a "proof" which ignores important analytic details. We have

$$e^{tA} = \sum_{k=0}^{\infty} \frac{(tA)^k}{k!} = I_n + tA + \frac{t^2}{2}A^2 + \frac{t^3}{6}A^3 + \frac{t^4}{24}A^4 + \ldots$$

To take the derivative, we differentiate this series, term-by-term[1], with respect to t to obtain

$$\begin{aligned}
\frac{d}{dt}\left[e^{tA}\right] &= \mathbf{0}_{n \times n} + A + tA^2 + \frac{t^2}{2}A^3 + \frac{t^3}{6}A^4 + \ldots \\
&= A\left(I_n + tA + \frac{t^2}{2}A^2 + \frac{t^3}{6}A^3 + \ldots\right) \\
&= Ae^{tA}
\end{aligned}$$

$\square$

Now we are in a position to show how matrix exponentials can solve systems of differential equations.

Theorem 2.5. *Let A be an $n \times n$ matrix. The solution to the initial value problem*

$$\mathbf{x}'(t) = A\mathbf{x}(t), \qquad \mathbf{x}(0) = \mathbf{x}_0$$

is $\mathbf{x}(t) = e^{tA}\mathbf{x}_0$.

Proof. We'll just verify that $\mathbf{x}(t) = e^{tA}\mathbf{x}_0$ is a solution. First observe that it satisfies the initial condition:

$$\mathbf{x}(0) = \mathbf{e}^{0A}\mathbf{x}_0 = \mathbf{e}^{\mathbf{0}_{n \times n}}\mathbf{x}_0 = I_n\mathbf{x}_0 = \mathbf{x}_0.$$

Next, observe that it satisfies the differential equation:

$$\mathbf{x}'(t) = \frac{d}{dt}\left[e^{tA}\mathbf{x}_0\right] = Ae^{tA}\mathbf{x}_0 = A\mathbf{x}(t).$$

$\square$

So it appears that we've figured out how to solve $\mathbf{x}'(t) = A\mathbf{x}(t)$. The problem is, how do we actually compute a matrix exponential?

---

[1]This is the part in which we are being sloppy. It turns out that this is OK to do, but we are not providing rigorous justification.

3. Diagonalization to the Rescue!

The first observation is that we can, in fact, compute e^D for a diagonal matrix D.

Proposition 3.1. *If* $D = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix}$, *then* $e^D = \begin{bmatrix} e^{\lambda_1} & 0 & \ldots & 0 \\ 0 & e^{\lambda_2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & e^{\lambda_n} \end{bmatrix}$.

Proof. For ease of notation, let's prove it in the 2×2 case (nothing different happens in the $n \times n$ case.) Suppose $D = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$. We know that $D^k = \begin{bmatrix} \lambda_1^k & 0 \\ 0 & \lambda_2^k \end{bmatrix}$. So

$$
\begin{aligned}
e^D &= I_2 + D + \frac{1}{2}D^2 + \frac{1}{6}D^3 + \frac{1}{24}D^4 + \ldots \\
&= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} \lambda_1^2 & 0 \\ 0 & \lambda_2^2 \end{bmatrix} + \frac{1}{6}\begin{bmatrix} \lambda_1^3 & 0 \\ 0 & \lambda_2^3 \end{bmatrix} + \frac{1}{24}\begin{bmatrix} \lambda_1^4 & 0 \\ 0 & \lambda_2^4 \end{bmatrix} + \ldots \\
&= \begin{bmatrix} \left(1 + \lambda_1 + \frac{1}{2}\lambda_1^2 + \frac{1}{6}\lambda_1^3 + \frac{1}{24}\lambda_1^4 + \ldots\right) & 0 \\ 0 & \left(1 + \lambda_2 + \frac{1}{2}\lambda_2^2 + \frac{1}{6}\lambda_2^3 + \frac{1}{24}\lambda_2^4 + \ldots\right) \end{bmatrix} \\
&= \begin{bmatrix} e^{\lambda_1} & 0 \\ 0 & e^{\lambda_2} \end{bmatrix}.
\end{aligned}
$$

$\square$

Next, we consider the case where $A = PDP^{-1}$.

Proposition 3.2. *Suppose that* $A = PDP^{-1}$. *Then* $e^A = Pe^D P^{-1}$.

Proof. We use the fact that $A^k = PD^kP^{-1}$ and compute

$$
\begin{aligned}
e^A &= I_n + A + \frac{1}{2}A^2 + \frac{1}{6}A^3 + \frac{1}{24}A^4 + \ldots \\
&= PP^{-1} + PDP^{-1} + \frac{1}{2}PD^2P^{-1} + \frac{1}{6}PD^3P^{-1} + \frac{1}{24}PD^4P^{-1} + \ldots \\
&= P\left(I_n + D + \frac{1}{2}D^2 + \frac{1}{6}D^3 + \frac{1}{24}D^4 + \ldots\right)P^{-1} \\
&= Pe^D P^{-1}.
\end{aligned}
$$

$\square$

Now we simply combine the previous two results:

Theorem 3.3. *Suppose that* $A = PDP^{-1}$, *where* $D = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix}$. *Then*

$$
e^A = Pe^D P^{-1} = P \begin{bmatrix} e^{\lambda_1} & 0 & \ldots & 0 \\ 0 & e^{\lambda_2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & e^{\lambda_n} \end{bmatrix} P^{-1},
$$

and more generally

$$e^{tA} = Pe^{tD}P^{-1} = P \begin{bmatrix} e^{\lambda_1 t} & 0 & \cdots & 0 \\ 0 & e^{\lambda_2 t} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{\lambda_n t} \end{bmatrix} P^{-1}.$$

Now we can solve our initial diffusion problem

$$\mathbf{u}'(t) = \begin{bmatrix} -2 & 1 \\ 1 & 2 \end{bmatrix} \mathbf{u}(t), \qquad \mathbf{u}(0) = \begin{bmatrix} 10 \\ 0 \end{bmatrix}.$$

The solution is given by $\mathbf{u}(t) = e^{tA} \begin{bmatrix} 10 \\ 0 \end{bmatrix}$, where $A = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix}$. The eigenvalues/eigenvectors of A are

$$\lambda_1 = -1, \qquad \mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

$$\lambda_2 = -3, \qquad \mathbf{v}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

So we can diagonalize A as

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

Consequently, we can compute

$$e^{tA} = Pe^{tD}P^{-1} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} e^{-t} & 0 \\ 0 & e^{-3t} \end{bmatrix} \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{bmatrix}$$

$$= \begin{bmatrix} e^{-t} & e^{-3t} \\ e^{-t} & -e^{-3t} \end{bmatrix} \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{bmatrix}$$

$$= \begin{bmatrix} (1/2)e^{-t} + (1/2)e^{-3t} & (1/2)e^{-t} - (1/2)e^{-3t} \\ (1/2)e^{-t} - (1/2)e^{-3t} & (1/2)e^{-t} + (1/2)e^{-3t} \end{bmatrix}.$$

Now we can use the initial condition to solve for $\mathbf{u}(t)$:

$$\mathbf{u}(t) = e^{tA}\mathbf{u}_0 = \begin{bmatrix} (1/2)e^{-t} + (1/2)e^{-3t} & (1/2)e^{-t} - (1/2)e^{-3t} \\ (1/2)e^{-t} - (1/2)e^{-3t} & (1/2)e^{-t} + (1/2)e^{-3t} \end{bmatrix} \begin{bmatrix} 10 \\ 0 \end{bmatrix} = \begin{bmatrix} 5e^{-t} + 5e^{-3t} \\ 5e^{-t} - 5e^{-3t} \end{bmatrix}.$$

In terms of $u_1(t), u_2(t)$, this says

$$\begin{cases} u_1(t) = 5e^{-t} + 5e^{-3t} \\ u_2(t) = 5e^{-t} - 5e^{-3t}. \end{cases}$$

We can now describe the temperature on the two regions of the rod as time goes on. For example,

$$\mathbf{u}(0) = \begin{bmatrix} 10 \\ 0 \end{bmatrix}, \quad \mathbf{u}(0.25) \approx \begin{bmatrix} 6.26 \\ 1.53 \end{bmatrix} \quad \mathbf{u}(0.5) \approx \begin{bmatrix} 4.15 \\ 1.92 \end{bmatrix} \quad \mathbf{u}(0.75) \approx \begin{bmatrix} 2.89 \\ 1.83 \end{bmatrix} \quad \mathbf{u}(1) \approx \begin{bmatrix} 2.09 \\ 1.59 \end{bmatrix}.$$

As expected, the temperature on the first region just drops. On the second region, it goes up and then starts to drop. Some of the heat from the first region entered the second, but then the heat is exiting out of the sides. We see that both $u_1(t)$ and $u_2(t)$ will go to 0 as t goes to infinity, because they are both sums of decaying exponentials. Intuitively, this makes sense.