Introduction to Dynamical Systems

Justin Werfel

9.29 Optional Lecture #7, 4/10/03

1 Linear dynamical systems

The term *dynamical system* just refers to a system that changes in time. We describe the system as being in some state x, which is in general a vector; and we're interested in how that state changes as time goes on. (If x has n dimensions, we call it an n-dimensional or nth-order system.) The simplest case is the following:

$$\dot{x} = Ax$$

So we can think of x as the position in what may be a high-dimensional vector space; then $\dot{x}$, which describes how the system will move around in that vector space, is given by some linear combination (given by A) of the coordinates of that position.

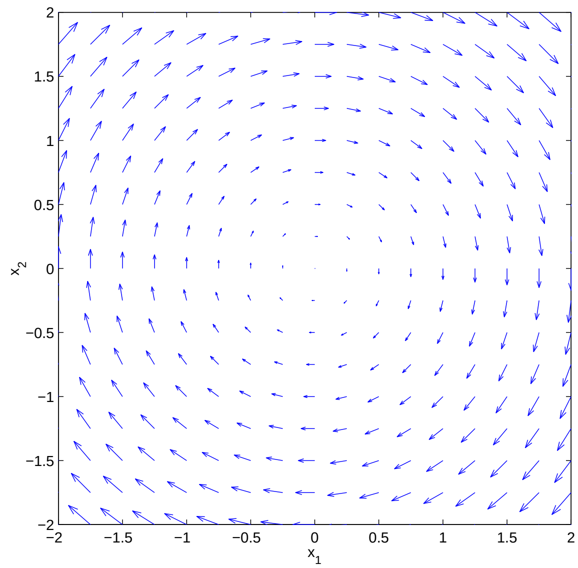
Let's take, as an example, the two-dimensional system $x = [x_1\ x_2]^T$. We can draw, for every point in the plane, an arrow originating from that point whose magnitude and direction corresponds to $\dot{x}$. (We can do that in MATLAB with the `quiver` command.) That will give us an idea of how the system will behave. Suppose A is $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. (Code:

```
>> [x,y]=meshgrid(-2.5:.25:2.5);
>> A=[0 -1; 1 0];
>> u=A(1,1)*x+A(1,2)*y; v=A(2,1)*x+A(2,2)*y;
>> quiver(x,y,u,v)
>> axis([-2 2 -2 2])
>> axis square)
```

Then you can see from the figure on the next page that starting at any point in the plane, the system will tend to circle clockwise; and the speed $|\dot{x}|$ will be greater if the distance from the center is greater, which makes sense because the velocity is just a linear transformation of the position.

The most general case for a linear dynamical system (LDS) has the following form:

$$\begin{aligned} \frac{dx}{dt} &= A(t)x(t) + B(t)u(t) \\ \frac{dy}{dt} &= C(t)x(t) + D(t)u(t) \end{aligned}$$



Here t, of course, is time; x, again, is the state of the system; u is a control variable, some input you have into the system that can affect its behavior; y is the output of the system, some kind of readout which is a transformation of the state x and input u but doesn't affect the time evolution of the system. Again, x, u, and y are in general vectors, so A, B, C, D (which all have names in control theory) are in general matrices. This is the continuous-time case; in the discrete-time case, rather than these being differential equations, the left-hand side of the equations are $x(t+1)$ and $y(t+1)$.

In most cases, A, B, C, D are time-invariant. Often there's no control input u, in which case the system is called *autonomous*. And if there's no particular output of the system that we care about distinct from its input, we get the case we started with, a continuous-time autonomous time-invariant LDS:

$$\dot{x} = Ax$$

Now, the time derivatives here only go up to first order. What if we wanted to study a case with higher-order time derivatives? Let's take everyone's favorite example, the undamped pendulum. The equation of motion is $\ddot{\theta} + \frac{g}{L}\sin(\theta) = 0$. This is actually a more complicated case for two reasons: the second derivative with respect to time, and the nonlinearity of the sine function. We'll get to the nonlinear part later; for now, let's get rid of that complication by making the usual small-angle approximation, $\sin(\theta) \approx \theta$ for $\theta \ll 1$. Now what we do is define a new state variable $x \equiv [\theta\ \dot{\theta}]^T$. Then $\dot{x} = [\dot{\theta}\ \ddot{\theta}]^T$, and we can write $\dot{x} = \begin{bmatrix} 0 & 1 \\ -g/L & 0 \end{bmatrix} x$. We can always get rid of higher-order time derivatives by increasing the dimensionality of the system. Notice also that except for a scale factor g/L, the matrix here is the same as the one in the example above; the pendulum swinging back and forth corresponds to a circular motion in the phase plane. Along the positive x-axis, angle is at a maximum, the rate of change of position is 0, and the rate of change of velocity is greatest in the direction of negative values; later, when the system reaches the negative y-axis, angle is 0, velocity is maximally negative, rate of change of angle is greatest and negative, rate of change of velocity is zero; etc. All this can be read directly off the graph (in this case which is sufficiently low-dimensional that we can graph it).

One key feature of the system we can look at is its *fixed points*, those points x^* in state space where $\dot{x} = 0$. If the system is at a fixed point, it won't move from there; hence the name. Since $\dot{x} = Ax$, that means $Ax = 0$, so the fixed points are those in the nullspace of A. With linear systems, you'll either have a single fixed point at $x = 0$ (if A is nonsingular), or an infinity of fixed points along a hyperplane passing through the origin (if A is singular).

An important question is that of the *stability* of fixed points. If the system is exactly at a fixed point, it won't change; but what happens if it's very close to a fixed point—where will it go? The standard framework for approaching this question comes from the Russian mathematician Alexandr Mikhailovich Lyapunov, who investigated nonlinear stability analysis in the late 1800s; the following definitions are due to him.

1.1 Lyapunov stability

The following all assume the fixed point x^* under consideration is at 0. For arbitrary x^*, $|x|$ should be replaced by $|x - x^*|$ in the below definitions.

- A fixed point x^* is *Lyapunov stable* if, when the system starts sufficiently close to x^*, it will stay arbitrarily close to it for all time thereafter. The formal definition is: $\forall\, R > 0\ \exists\, r > 0$ s.t. $|x(0)| < r \Rightarrow |x(t)| < R\ \forall\, t \geq 0$.

- If a fixed point is not stable in this sense, it is *unstable*. That is, there is at least one spherical neighborhood around the fixed point such that you can't get the system to stay within it forever, no matter how close you start it.

- A fixed point is *attracting* if, when you start sufficiently close to it, the system will converge to the fixed point as $t \to \infty$. Formally, $\exists\ r > 0$ s.t. $|x(0)| < r \Rightarrow \lim_{t\to\infty} x(t) = x^*$. Note that in nonlinear systems, a fixed point can be attracting without being stable in the Lyapunov sense. For instance, in the system $\dot{\theta} = 1 - \cos\theta$, the system will always go to $\theta = 0$ at infinite time, but you can start it with θ as small and positive as you want and it'll go on this extended excursion first, going to the fixed point the long way around; it won't stay within a small ball around 0.

- If a fixed point is Lyapunov stable but not attracting, it is called *marginally* or *neutrally stable*. For instance, a ball sitting on a table: it's at a fixed point, you can displace it a tiny bit, and it won't return to the original fixed point nor will it move away from it.

- If a fixed point is both Lyapunov stable and attracting, then it is called *asymptotically stable*.

- Finally, a fixed point is *exponentially stable* if nearby states converge to it faster than exponentially, i.e., $\exists\ \alpha, \lambda$ s.t. $|x(t)| \leq \alpha |x(0)| e^{-\lambda t}\ \forall\ t > 0$.

Lyapunov's work was on *nonlinear* dynamic systems; linear systems are always either asymptotically stable, marginally stable, or unstable—as we'll show now.

1.2 The return of eigenanalysis

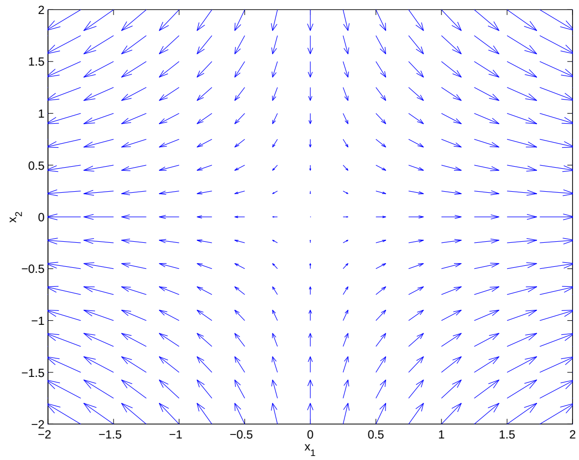
As we said, the origin will always be a fixed point for linear systems of the sort we're discussing. Let's suppose that there's some fixed vector v such that if you start the system along that vector, it will move along that vector exponentially, $x(t) = Ke^{\lambda t}v$. If there is such a vector, then you can see that $\lambda > 0$ means the system is unstable and goes to infinity exponentially; $\lambda < 0$ means the system is asymptotically stable, converging to 0 with time constant $1/\lambda$; and $\lambda = 0$ means the system is neutrally stable, it doesn't change. So we have:

$$\begin{aligned} \dot{x} &= Ax \\ \lambda K e^{\lambda t} v &= AKe^{\lambda t} v \\ \lambda v &= Av \end{aligned}$$

Thus all of the above holds if v is an eigenvector of A, with eigenvalue λ. If the eigenvalues of A are all distinct, then the eigenvectors will be linearly independent, and we can write any state x as a linear combination of the eigenvectors, $x(t) = c_1 e^{\lambda_1 t} v_1 + \ldots + c_n e^{\lambda_n t} v_n$. This is the general solution: it satisfies the equation $\dot{x} = Ax$, so it is *a* solution; and in linear differential equations we have an existence and uniqueness theorem that says that it's the *only* solution.

That means you can characterize the behavior of the system just by finding the eigenvalues; you don't even need the eigenvectors, if determining stability is your only concern. Because you've written the state as a linear combination of eigenvectors, each of which satisfies the update equation, the behavior of the total state will be the sum of the behaviors of its components. So if the first eigenvalue is positive and the second is negative, for instance, the system will blow up in the direction of the first eigenvector with that time constant, while at the same time the component in the direction of the second eigenvector will go to 0 with that time constant, and so on. Let's take a really simple 2-D example: $A = \begin{bmatrix} 3 & 0 \\ 0 & -2 \end{bmatrix}$, for which the eigenvalues are 3 and -2, and the eigenvectors are just the usual basis vectors in the x- and y-directions. If we start this somewhere, the x-component will blow up as e^{3t} while the y-component goes to 0 as e^{-2t}. You can see this in the quiver plot in MATLAB, shown on the next page.

If *all* the eigenvalues are negative, the system is asymptotically stable. If any eigenvalues are positive, the system is unstable. If there are no positive eigenvalues, and at least one 0 eigenvalue, the system is marginally stable. That's for real eigenvalues; what about for complex ones? We can write $\lambda = R + i\omega$; then $e^{\lambda t} = e^{Rt} e^{i\omega t}$. The real component has the effect on stability we just described; meanwhile, because $e^{i\omega t} = \cos(\omega t) + i \sin(\omega t)$, the complex component corresponds to circling. So you'll get growing or decaying spirals, or if the eigenvalue is purely imaginary, closed orbits. (The first example above had eigenvalues $\pm i$.)



2 Nonlinear dynamical systems

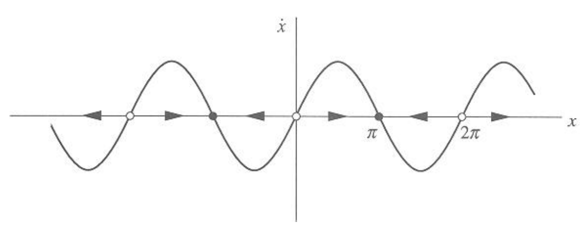
Here we get to the interesting cases. Pretty much nothing is linear in real life. The more general nonlinear form is

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

where f is an arbitrary function. What can you do to analyze this?

Let's start by just looking at some examples. One important point is that, whereas a linear dynamical system has either one fixed point or an infinity of them that that make up some hyperplane, a nonlinear dynamical system can be constructed to have an *arbitrary* number of *isolated* fixed points.

Take $\dot{x} = \sin(x)$. This actually has an exact solution—$t = \ln\left(\frac{\csc x_0 + \cot x_0}{\csc x + \cot x}\right)$—but it's not a very helpful one. Not only can you not get a closed-form solution for $x(t)$, but qualitative features of the solution, like what happens to $x(t)$ as $t \to \infty$, are hard to extract. But since we have only one dimension here, we can draw the system with x on the x-axis and $\dot{x}$ on the y-axis.



(Taken from Strogatz, *Nonlinear Dynamics and Chaos*.)

To start with, everywhere the graph crosses the x-axis is a fixed point, because $\dot{x}$ is 0 there. If the system is somewhere where $\sin(x)$ is positive, then $\dot{x}$ is positive and x

increases; and similarly where $\sin(x)$ is negative. We can draw arrows on the x-axis to show which direction the system will move in for that value of x, based on the sign of $\dot{x}$; this is the one-dimensional version of the vector fields we looked at earlier. Here it shows that we have alternating unstable and stable fixed points, spaced at intervals of π. If x is just slightly greater than 0, first the system will move to the right, with greater and greater velocity; once x passes $\pi/2$, the system continues to move to the right, but with decreasing velocity, so that at infinite time it approaches $x = \pi$.

Note that this characterization of stable vs. unstable fixed points is a *local* one. The intuition is that if you perturb the system from a fixed point, it'll move away if it's an unstable fixed point or return if it's a stable one. However, if a perturbation is large enough, the system can be moved far enough from a locally stable fixed point that it won't return, and will go somewhere else instead. For an asymptotically stable fixed point, the set of all points such that if you start the system there, it will end up at that fixed point, is called the *basin* or *domain of attraction* of that fixed point. In the sine case, the unstable fixed points separate successive basins of attraction.

In one-dimensional cases like this, we can consider the dynamics in state space as though the function $f(x)$ were a force, in the sense that it can be expressed as the negative gradient of some scalar potential function, $f(x) = -V'(x)$ or $V(x) = -\int f(x)dx$. The time evolution of the system will then always be such that V is non-increasing. In higher dimensions, it is sometimes, but not always, possible to express the dynamics as a gradient flow in this way. (Limit cycles (see below), for instance, cannot be expressed as movement following a potential gradient.)

In higher dimensions, moreover, we can't determine the stability of fixed points by the above graphical method. So what can we do? Here's the great technique of working with nonlinear systems: we pretend they're linear. Because we've got all these tools for dealing with linear systems, that we'd like to be able to use again here; so we find the fixed points (how that's done will depend on the particular $f(x)$), and then we *linearize* about each one. The idea is just to take a Taylor expansion about the fixed point x^*:

$$\dot{x} = f(x(t)) = f(x^*) + f'(x^*)(x(t) - x^*) + O(|x(t) - x^*|^2)$$

In general, this is a vector derivative[1]; here we're just talking about the one-dimensional case. Now, if we define $\delta(t) \equiv x(t) - x^*$—this is our perturbation from the fixed point—we have

$$\dot{\delta} = 0 + f'(x^*)\delta(t) + O(|\delta|^2) \approx f'(x^*)\delta(t)$$

which is just a linear dynamical system like we were talking about earlier. And this is a good approximation for x sufficiently close to the fixed point; you can always choose δ small enough that the extra factors of δ will make all higher-order terms negligible compared to the first-order one.

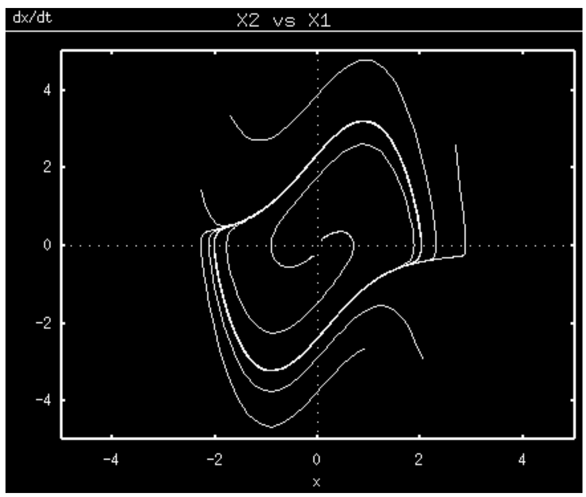
Now we can apply all the tools for stability analysis that we were talking about for the linear case. In one dimension, it's simply: if $f'(x^*) < 0$, it's asymptotically stable; if $f'(x^*) > 0$, it's unstable. Note that in the latter case, it may not continue to diverge for very long; this is only a good approximation in the immediate vicinity of the fixed point.

[1]Specifically, the Jacobian matrix $\frac{\partial f_i}{\partial x_j}$.

When $f'(x^*) = 0$, you can't say anything about the stability of the fixed point. The examples $\dot{x} = -x^3; \dot{x} = x^3; \dot{x} = x^2; \dot{x} = 0$ illustrate this. In each, 0 is a fixed point, and $f'(0) = 0$, but 0 is a different kind of fixed point for each.

In higher dimensions, these tests apply to the real parts of the eigenvalues of the Jacobian. Also in higher dimensions, systems can have all sorts of interesting behaviors. An *invariant set* is defined as a set of points such that, if a trajectory is started from within that set, it will stay within it for all future time. So, for instance, any fixed point is an invariant set; the domain of attraction of a fixed point is an invariant set; and another type of invariant set is a *limit cycle*, an isolated closed trajectory—isolated meaning that nearby trajectories are not closed, but spiral in toward or away from the limit cycle, depending on whether it's *attracting* or *unstable*.[2]

To consider an example of an attracting limit cycle, let's look briefly at the van der Pol oscillator. The equation is $\ddot{x} + \mu(x^2 - 1)\dot{x} + x = 0$; we can make that a second-order system with only first-order time derivatives by the trick we discussed before, $y \equiv [y_1\, y_2]^T = [x\, \dot{x}]^T$; then $\dot{y} = [x_2 \;\; -x_1 - \mu(x_1^2 - 1)x_2]^T$. We can numerically integrate that using the techniques Thomas talked about in the last optional lecture. Then we can look and see what the qualitative behavior of the system is. The figure below shows x on the horizontal axis, $\dot{x}$ on the vertical, with $\mu = 1.5$; and looking at trajectories that start from a variety of initial conditions, we see that there's an unstable fixed point at the origin and an attracting limit cycle of this weird shape. Trajectories starting inside or outside that limit cycle are drawn to approach it.



[2]The circular orbits we saw can occur in linear systems are *not* limit cycles; they're not isolated, because neighboring orbits are also closed. It's like the difference between asymptotic stability and neutral stability.