First order differential equations and phase lines

by Eric Cytrynbaum

November 4, 2009

First order differential equations

Antiderivatives in disguise

Suppose $g(t)$ is a continuous function that is known and $x(t)$ is an unknown function. We can formulate a simple equation that relates the two such as

$$\frac{dx}{dt} = g(t).$$

The equation above is called a differential equation because it is an equation in an unknown function that involves a derivative of the unknown function. The equation happens to also be an antiderivative in disguise in the sense that asking what function solves this equation is equivalent to asking for the antiderivative of $g(t)$. The solution to this equation can be found provided that an antiderivative of $g(t)$ can be found.

Imagine you took a road trip with a radar gun on the dash board set up to measure your velocity at every moment and dump the data to a laptop. At the end of the trip, you would have a complete record of your velocity at every moment; call it $g(t)$. You could then use an equation like the one above to determine your position along the highway, $x(t)$, at every moment.

This scenario is, in some sense, an unusual one in the context of modeling physical laws. When addressing physical laws, we are usually interested in time dependent phenomena that do not depend on the actual time but instead on the relative time. The dependence on absolute time arises whenever the derivative is equal to a specified function of time, like $g(t)$ above. If you look at your watch and it happened to be t_0, your velocity would be $g(t_0)$. If you had left an hour later than you actually did, and driven in the exact same manner, you would have to come up with a new function to describe your velocity. In particular, the equation $dx/dt = g(t-1)$ would work provided t is measured in hours. To describe a physical law, we would prefer to have the equation depend only on the state of the system (in this case, the value of x) rather than the absolute time. To illustrate, consider dropping a ball from the top of a tower. It should take the same amount of time to hit ground (i.e. relative time – the time from dropping to hitting) whether you do the experiment in the middle of the night or during the day (the absolute time).

To see this difference between absolute and relative time in terms of the equations, suppose $g(t) = -e^{-t}$ so that

$$\frac{dx}{dt} = e^{-t}$$

and suppose that $x(6) = 1$ where x is the position of your car in km from UBC campus and t is the time in Vancouver. The 6 represents the fact that you started driving at 6 in the morning. The solution to the equation with this initial condition is $x(t) = e^{-t} - e^{-6} + 1$. [*Check that this function solves both the equation and the initial condition.*]

To contrast this with a more physical-law type equation, let's invent a rule for how to drive that does not depend on the absolute time (as above) but instead depends on the current position of the car, $x(t)$. This is more like how gravity works, for example (but a bit simpler). Suppose that at any moment during the drive, you look out the window and estimate how far you are from the far end of the upcoming exit ramp. You carefully set your speed according to the rule

$$\frac{dx}{dt} = -v_0(1 - e^{-x}).$$

Notice that your speed depends on your position and not on the clock-time. I've also made the right hand side (RHS) such that when you reach the end of the exit ramp, your speed is zero and, if you are far from the exit, your speed is close to $-v_0$. Assuming v_0 is positive, the minus sign ensures that, as you approach the exit, the rate of change of the distance between you and the exit is negative but increases toward zero. This means you are approaching the exit and will eventually stop when you get to the end of the exit ramp. Unlike the previous example, this is not an antiderivative in disguise because x appears on the RHS.

General Case

A more general form of differential equation has a RHS that depends on both t and the unknown function $x(t)$:

$$\frac{dx}{dt} = g(x, t).$$

Of course, you could say that x depends on t so the entire RHS is really not so different from the "integral in disguise" above. The difference, however, is that $x(t)$ is not yet known and its dependence on t can not really be explicitly described until the equation is solved. This means that we are no longer dealing with an antiderivative in disguise. Integration tricks will nonetheless come in handy when trying to solve equations like this. Like the previous case, these kinds of equations describe physical problems in which the absolute time matters. For example, the equation for a mass-spring system immersed in a thick fluid that is connected to the shaft of a motor rotating at 20 rpm has a RHS that depends both on the position of the mass and on the time in motor's cycle. The equation describing the position of the mass would look something like

$$\frac{dx}{dt} = -kx + \sin(40\pi t)$$

For now, we are not interested in this kind of equation but instead will focus on a simpler version of it in which absolute time does not appear, like in the second example above. This is equivalent to not having a t anywhere on the RHS. Physically, we can accomplish this by disconnecting the motor from the mass-spring and just focusing on how the mass-spring system behaves. This simpler type of equation is referred to as an autonomous equation because the kind of physical systems it describes are independent of outside influence. With the motor attached, the motion of the mass is not autonomous but instead controlled by the motor.

Autonomous equations

So far, we have only discussed equations in which the first derivative of x appears. These are called first order equations. Higher order equations have higher derivatives. In physics, first order equations might be thought of as the domain of Aristotle while second order equations are the domain of Galileo and Newton. I say this because of the relationship between forces and motion. Aristotle thought that an object required a force acting on it to maintain its velocity. The equation which he might have written down, had he known about calculus, is

$$\frac{dx}{dt} = \frac{F}{\mu}$$

where x is the position of the object, F is the force applied to the object and μ is the the drag coefficient associated with the object. Galileo later recognized that velocity actually comes for free through the concept of inertia. Newton formalized this in his Second Law which states that

$$\frac{d^2x}{dt^2} = \frac{F}{m}$$

where m is the mass of the object. Although Aristotle's equation is incorrect in general, it provides a decent approximation to the motion of objects in an environment that is so viscous that motion cannot be maintained for any reasonable period of time in the absence of force (the so-called low-Reynolds-number regime). Because of their small size and slow velocities, the movement of cells and proteins is well described by Aristotle's equation. The low-Reynolds-number world is a useful place to start learning about differential equations for the simple reason that the equations describing motion in this context are first order. A great introduction to biological physics in the low-Reynolds-number world is the well-known paper by EM Purcell titled Life at Low Reynolds Number. For now, I will implicitly assume that we are talking about this inertia-less brand of physics.

When does a physical law take the form

$$\frac{dx}{dt} = g(x)?$$

Comparing this to Aristotle's equation, an equation like this arises when the force acting on an object depends on its position. Examples include a mass-spring system, a pendulum, a protein (or DNA) with a net charge sitting in an electric field (e.g. running a gel), particles in a centrifuge and many others. In all these cases, the rate at which x is changing at a particular moment is entirely determined by the value of x at that moment. That dependence is determined by the physical problem and represented by the function g.

The phase line

When dealing with differential equations, one is often faced with the problem of not being able to come up with a solution in closed form (i.e. a nice clean formula for the solution). In these cases, what information about solutions can you extract directly from the equation? To work with a concrete example, let's consider the equation:

$$\frac{dx}{dt} = x - x^2.$$

To facilitate the following explanations, let's give the function on the RHS of the equation a name: $f(x) = x - x^2$, so that the equation can be written as $dx/dt = f(x)$. Note that

I have suppressed the explicit reference to the t dependence of the function and will do so often – you must always keep mental track of which letters are parameters (constants whose values are unspecified – none in the example above) and which are variables (values that change within the dynamic context of the problem), specifically independent variables (like t) and dependent variables (like x).

Steady states solutions

These are solutions that do not change in time so they must have time-derivative equal to zero. To find these, we force $dx/dt = 0$ which occurs whenever $x - x^2 = 0$. The steady state solutions for our example equation are therefore the constant functions of time $x(t) = 0$ and $x(t) = 1$.

Stability

Once we have found the steady states, we can ask how solutions that start close to the steady state behave. If *all* solutions that start sufficiently close to a steady state eventually approach the steady state value, then the steady state is called stable[1]. If there are solutions starting arbitrarily close to the steady state that leave the area near the steady state, then the steady state is called unstable. There are two ways to determine stability, (1) qualitatively from the graph of the RHS, $x - x^2$, and (2) by studying the RHS algebraically (or "differentially", really). The former approach helps to explain the latter so I'll start with the former.

At any time t, the expression on the RHS, $x(t) - x(t)^2$ determines the instantaneous direction and amplitude of change in $x(t)$ at time t. So if $x(t) - x(t)^2 > 0$ for some value of t, then $x(t)$ is increasing at that moment. Similarly, if $x(t) - x(t)^2 < 0$, then $x(t)$ is decreasing at that moment. By drawing the graph of the RHS as a function of x, we can quickly identify the intervals on which $x(t)$ is increasing and those on which it is decreasing. The x axis in this graph is called the phase space[2] where by phase we mean a representation of the state of the system. Because in this case the state is entirely characterized by a single variable (x) we use the expression phase line.

Note that in drawing $f(x)$ as a function of x we have no graphical representation of time. This means that when looking at the phase line, we must imagine that the current state is marked by a flashing dot at the current value of x. As time goes on, the flashing dot moves according to the instructions provided by $f(x)$ which specifies the rate of change of x. The phase line for the example above is illustrated in Figure 1A.

The steady states of the equation are the values of x at which the RHS is zero so the graph must cross or, at least touch, the x axis at each steady state. To determine stability for the steady state $x(t) = 1$, we are interested in what happens at values of x close to the steady state in question. The graph of $x - x^2$ is positive immediately to the left of $x = 1$, so solutions that start just below $x = 1$ (to the left of the steady state), must increase in time

[1] Technically, the definition of stability only requires that solutions which start close to the steady state stay close to the steady state. The notion of stability used here, requiring that all nearby solutions *approach* the steady state is referred to as asymptotic stability.

[2] The expression state space is often used instead of phase space and is arguably a better term because we usually talk about the state of the system rather than the phase of the system except for systems in which the state space is periodic or well characterized by plan or previous experience (phase of the moon, phase of construction, "its just a phase he's going through"). Other expressions include phase line (one state variable), phase plane (two state variables), phase portrait (arbitrary number of state variables).

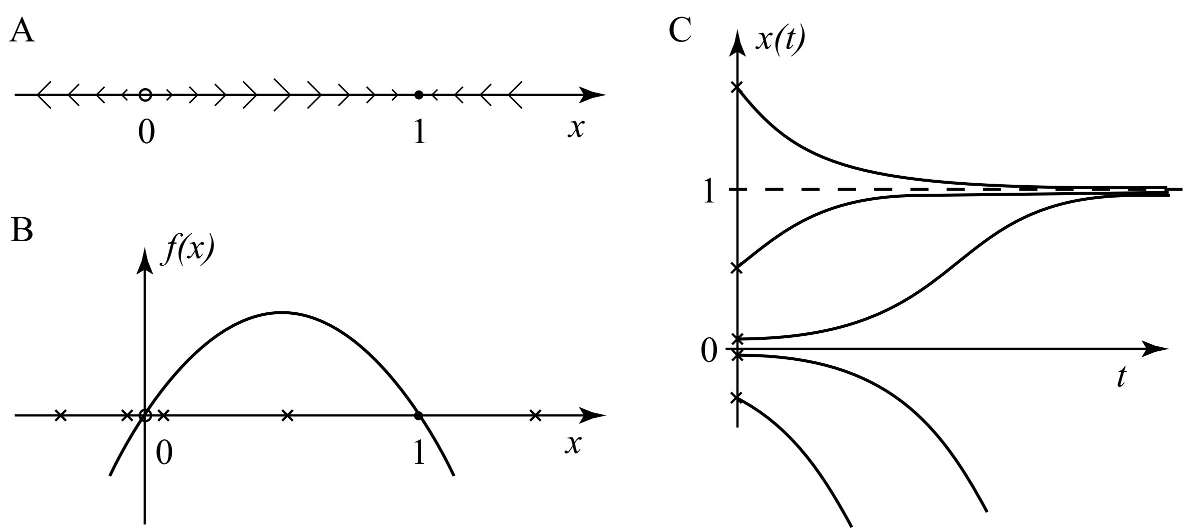

Figure 1: The phase line and interpretive aids for the equation $dx/dt = f(x)$. Note that according to a narrow definition, the phase line is really just the diagram in A. It consists of an axis for the state variable (x), steady states marked with solid/open dots (solid for stable, open for unstable) and arrows indicating how the state changes at each point on the axis. The arrows are often called a *direction field*. Panel B shows the state axis (x) annotated with a vertical axis to show the direction field as a curve instead of as arrows. The x's mark the points used as initial conditions for the sketch in C (also shown in C). To translate from A (or B) to C, imagine the x's moving in time along the axis, at every instant following the directions given by the direction field (including direction and speed).

and must continue to do so without ever getting above $x = 1$ and so they must approach 1. The steady state at $x = 1$ is starting to look stable. But we must check all nearby starting values, in particular, those above $x = 1$. Here, $x - x^2 < 0$ so any solution that starts above $x = 1$ decreases and continues to do so without ever getting below $x = 1$ and so must also approach 1. Thus, we conclude that $x = 1$ is a stable steady state.

Notice that stability arose from the fact that below the steady state the RHS is positive and above the steady state the RHS is negative (i.e. the arrows on the phase line point toward the steady state). Provided the RHS is differentiable, this is equivalent to saying that the slope of the function $f(x) = x - x^2$ is negative at the steady state. Checking this condition in the example above, we see that $f'(1) = -1$.

Similarly, instability arises when $f'(x_{ss}) > 0$ which is equivalent to saying that the arrows point away from the steady state. Note that $f'(0) = 1$ so $x = 0$ is an unstable steady state.