Ordinary Differential Equations

Slobodan N. Simić

Contents

Preface

This is a set of lecture notes for Math 133A: Ordinary Differential Equations taught by the author at San José State University in the Fall 2014. The only prerequisite for the course is multivariable calculus. The notes focus on qualitative analysis of differential equations in dimensions one and two. Since the class is mainly intended for SJSU engineering majors, the Laplace transform is also covered in some detail.

Corrections, comments, and suggestions would be greatly appreciated and should be emailed to the author at `simic@math.sjsu.edu`.

Slobodan Simić

CHAPTER 1

First order differential equations

> Newton's fundamental discovery, the one which he considered necessary to keep secret and published only in the form of an anagram, consists of the following: *Data aequatione quotcunque fluentes quantitae involvente fluxiones invenire et vice versa.* In contemporary mathematical language this means: "It is useful to solve differential equations".
>
> Vladimir Arnold

1.1. What is a differential equation?

Differential equations are arguably the most important types of equations in mathematics, the natural sciences and engineering. They are one of the basic tools of mathematics. Modeling natural phenomena leads straight to differential equations. The first person to systematically study differential equations was Isaac Newton. He considered his discovery that differential equations are useful for modeling physical phenomena one of his greatest, so he kept it secret and published it only in the form of an anagram (see the quote above).

But what is a differential equation? Here's an example:

$$\frac{dy}{dt} = y. \tag{1.1}$$

Here's another:

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0. \tag{1.2}$$

And here's an equation that's not a differential equation:

$$x^2 + 3x + 2 = 0. \tag{1.3}$$

In each example there's an unknown we are supposed to solve for, but unlike the last equation (which is just a simple quadratic equation), where the unknown is a *number*, in the first two equations the unknown is a *function*. In (1.1) the unknown is the function y which depends on a single real variable t; in (1.2), the unknown is a function u which depends on two real variables, x and y. The first two equations involve derivatives of the unknown function, which is why we call them *differential* equations. Equation (1.1) contains an "ordinary" derivative $\frac{dy}{dt}$ of the unknown function y, so we call it an ordinary differential equation. This equation arises in modeling exponential growth (e.g., population dynamics). Equation (1.2) contains partial derivatives of the unknown function u, so we call it a partial differential equation. This equation (called Laplace's equation) arises in potential theory.

Here we will only be concerned with ordinary differential equations.

So we could say that an ordinary differential equation is an equation in which the unknown is a function of *one* real independent variable and the equation involves derivatives of that function. The order of an equation is the order of the highest derivative appearing in it.

We will always denote the independent variable by t and think of it as time. Usually the unknown function of t will be denoted by the letter y. Thus a first-order ordinary differential equation is always of the form

$$F(t, y, y') = 0,$$

where F is some function of three variables and $y' = \frac{dy}{dt}$. It is often possible to solve for y' and rewrite the equation in the more useful form

$$y' = f(t, y). \tag{1.4}$$

For instance, in (1.1), $f(t, y) = y$. If the right-hand side in (1.4) does not depend on t, we call the equation autonomous. Otherwise, it is called non-autonomous. Thus (1.1) is autonomous, whereas $y' = t + y$ is not.

Similarly, most second-order ordinary differential equations can be written in the form

$$y'' = g(t, y, y'),$$

where g is some function of three variables and $y'' = \frac{d^2y}{dt^2}$.

In later Chapters 2, 3 and 4, we will be studying *systems of differential equations*, where there is not just one, but several unknown functions, all depending on the same independent variable t. We postpone the discussion of systems to Chapter 2.

Since the unknown in a differential equation is a function, solving a differential equation means (at least in principle) finding infinitely many unknown numbers. This is in general much harder then solving for just one or two numbers (like in numerical equations such as (1.3)). As a rule of thumb, differential equations are harder to solve than numerical equations, and among differential equations, partial differential equations are harder to solve than ordinary differential equations. In fact, it turns out that most differential equations cannot be solved explicitly (i.e., there's no equivalent of the quadratic formula). But more about this later.

Notation. If y is a differentiable function of an independent variable t, we will denote its derivative either by y' (when it's clear with respect to what variable the derivative is taken) or by $\frac{dy}{dt}$. Sometimes (especially if t denotes time) we will use Newton's notation $\dot{y}$. So remember:

$$y' = \dot{y} = \frac{dy}{dt}.$$

1.2. Basic examples and modeling

In this section we will introduce three differential equations which arise in modeling of population growth and harmonic oscillators. Before we do that, let us first discuss what we mean by modeling.

Essentially, all of science is a search for models for how the Universe works. But to quote George E. P. Box, a British-American statistician:

Essentially, all models are wrong, but some are useful.

That's of course because the Universe (or almost any subset of it) is too complicated to be modeled by our puny mathematical tools, since a model is just a mathematical representation of the real world (whatever that is). We can only hope to partially represent some aspects of the real phenomenon we wish to understand. So to obtain a reasonable model we need to make *simplifying assumptions*. That allows us to identify a finite (and ideally small) number of relevant variables and parameters, and then write down a set (also ideally small) of equations comprising our model. Once we have it, we can temporarily (but only temporarily) forget the "real thing" and analyze the model using any tools and methods that mathematics can offer. And that means calculus, geometry, topology, etc. After we have understood our mathematical model, we can make predictions about its future behavior and compare those predictions with real data (which we can presumably obtain by measurement). This can give us some idea about the validity and limitations of our model, which we can use to improve it, say by including more relevant variables or parameters, choosing a better set of equations, etc.

We usually start by building a model that is as simple as possible yet reflecting the basic features of the real phenomenon we are trying to represent. If the model is not good enough, we make it more complex. The goal is to have a good balance between simplicity (which often translates into tractability) and faithfulness in representing the real world.

Let's see how this works in the following examples.

1.2.1. Unlimited population growth. Assume we have the population of some species (say fruit flies), which has the following remarkable properties:

- Its members never die.
- The population has access to an unlimited supply of food.
- The species spends all of its time reproducing itself.

How do we go about modeling the change of this population with time? First, let us denote the number of individual members of the population by P and time by t. So P is a function of t. Of course, in reality P is a positive integer, but since we are hoping to study its rate of change with respect to t, we choose to treat it as a *real* number. In fact, we will assume that P is a *differentiable* function of t; in other words, we assume that $P' = \dfrac{dP}{dt}$ exists for all values of t.

What do our standing assumptions tell us about the rate of change P' of P with respect to t? It is clear that the larger the size P of the population, the faster it reproduces, i.e., the larger its rate of change P'. Thus it makes sense to assume that P' is proportional to P, which means that there exists a positive parameter k[1] such that $P' = kP$. Thus the simplest differential equation modeling the growth of the population satisfying the above properties is

$$\frac{dP}{dt} = kP. \tag{1.5}$$

This is the simplest non-trivial example of a differential equation (in which the right-hand side is not just a function of the independent variable). We will study it in two ways: qualitatively and analytically (i.e., by solving it).

Qualitative approach: First of all, what *is* qualitative analysis of a differential equation?

The short answer is: it is a type of analysis in which we do not explicitly solve the equation

[1]This parameter is related to the food supply, but we will not elaborate on that connection here.

(mostly because we can't) but instead try to understand what happens to solutions as $t \to +\infty$ and $t \to -\infty$.

Let $P(t)$ be a solution to (1.5) and set $P_0 = P(0)$. This is the initial value of the population when $t = 0$. Let's see what we can say about $P(t)$ for large t for various values of P_0. Assume first that $P_0 = 0$. If there is no population to speak of at the beginning, there will obviously be no population at any time in the future or past, so $P(t) = 0$, for all values of t. The solution is constant; we will call such solutions equilibrium solutions or equilibria.[2] Finding and understanding such solutions is an important part of qualitative analysis.

If $P_0 > 0$, then $P'(0) > 0$, so by calculus P must be an increasing function near $t = 0$. Therefore $P(t)$ is positive at least for small values of $t > 0$. But for any such value of t, $P(t) > 0$ implies that $P'(t)$ is also positive, since $P'(t) = kP(t)$ and $k > 0$. Therefore, $P(t)$ keeps increasing as t gets bigger and it is not hard to see that it never stops increasing. That is, if $P_0 > 0$, then $P(t) \to \infty$, as $t \to \infty$, which is not surprising.

Forgetting that P denotes population for a moment, let's see what happens if $P_0 < 0$. We leave it as an exercise to the reader to show, in exactly the same way as in the previous paragraph, that $P(t) \to -\infty$, as $t \to \infty$.

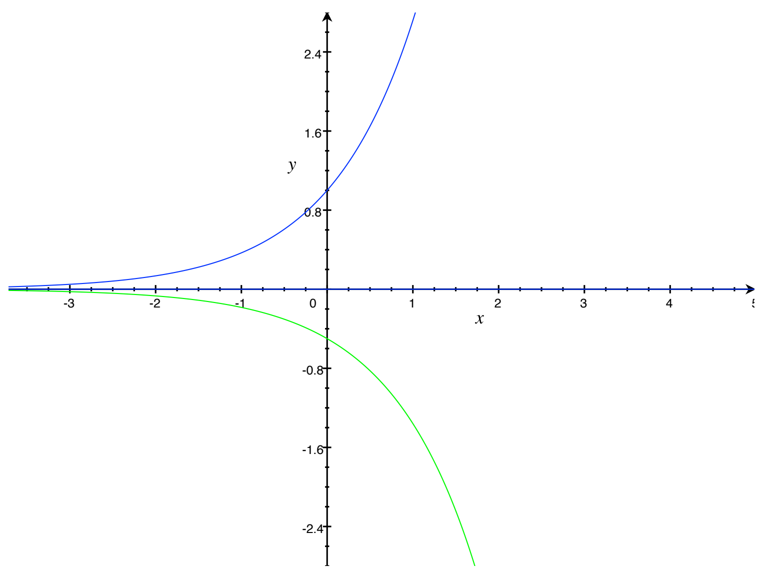


FIGURE 1.1. Three types of solutions of equation (1.5).

Therefore, *qualitatively* speaking, there are only three types of solutions: the equilibrium solution (corresponding to the not very interesting situation when $P_0 = 0$), those solutions that tend to $+\infty$ and those that tend to $-\infty$ (both as $t \to \infty$).

Analytic approach: Now let us solve (1.5). It is easy to see that one solution is $P(t) \equiv 0$[3]: just plug it into both sides of the equation and check that it "works". Assume $P \neq 0$. Divide both sides of the equation by P and, pretending that dP and dt have a meaning independent of each other, "multiply" both sides by dt. We obtain

$$\frac{dP}{P} = k\,dt.$$

[2]Equilibria is the plural of *equilibrium*.

[3]The symbol $\equiv$ means that the left-hand side equals the right-hand side for *all* values of the independent variable.

Now integrate both sides:

$$\int \frac{dP}{P} = \int k\,dt.$$

Observe that P is now a "dummy variable" in the indefinite integral on the left and t plays the same role in the integral on the right. Solve these integrals. (If you don't remember how to do that, you need to review basic integration techniques.) This gives us

$$\ln|P| = kt + C,$$

where C is a constant and ln denotes the natural logarithm. Exponentiating both sides we get

$$|P| = e^{kt+C} = e^{kt}e^{C} = Be^{kt},$$

where $B = e^C$, just another constant (for now). Since $|P| = \pm P$, taking $A = \pm B$ (just another arbitrary constant), we obtain

$$P = Ae^{kt}.$$

So every non-zero solution appears to be a constant multiple of the exponential function e^{kt}! How does the constant A relate to the initial value P_0 of P? Let's see: set $t = 0$ in the previous equation and get

$$P_0 = Ae^{k\cdot 0} = A.$$

So A equals the initial value P_0 of $P(t)$! Therefore, given $P_0 = P(0)$, our solution must be

$$P(t) = P_0e^{kt}, \tag{1.6}$$

for all values of t. The method we used for finding the solution is called separation of variables. We'll learn more about it in Section 1.4.

Observe that the explicit solution given by (1.6) is consistent with the conclusions of the qualitative analysis: even though there are infinitely many solutions (one for each initial value of $P(t)$), they fall into one of the three categories discussed above. □

1.2.2. Logistic population model. It's pretty clear that the assumptions we made in the previous model are extremely unrealistic. Indeed, if we compared the predictions of this model to real data (say, the US census) we would see that they agree only for small values of t. When t becomes large, the predictions have nothing to do with reality whatsoever.

So how can we improve our model and keep it not too complicated? Recall that the right-hand side of equation (1.5) is a very simple function of P: it is linear in P, which is why (1.5) is an example of a linear differential equation (with constant coefficients); more on those in Section 1.5. So one way to improve our model would be to use a slightly more sophisticated, non-linear function on the right-hand side of the equation. Of course, we don't want to make it too complicated either, so the most natural choice would be a *quadratic* function of P, i.e., something that looks like this: $f(P) = AP^2 + BP + C$, for some suitably chosen parameters A, B and C. But how do we choose these parameters? One way is to make better assumptions. For starters, no more unlimited quantities of food. How do we incorporate this assumption in the new model? Here's one way to do that.

(1) For small P, we still want P' to be nearly proportional to P, i.e., when the population is small we would like P' to be approximately equal to kP, where k is a parameter as above.

(2) However, if P gets too large, we will assume that P' becomes *negative*, that is, the population decreases. How large is too large? We assume that there is a sort of "ideal population" N – also called the *carrying capacity* – such that if $P < N$, then $P' > 0$ and if $P > N$, then $P' < 0$.

What is the simplest quadratic function $f(P)$ which incorporates these assumptions? Assumption (1) suggests that we should try $f(P) = (\text{something}) \cdot kP$, where "something" is close to 1 when P is small. But we need to make sure that assumption (2) is also satisfied. After thinking about this for a while, we are inescapably lead to the following solution:

$$\text{something} = 1 - \frac{P}{N}.$$

Thus our new model for population growth becomes

$$\frac{dP}{dt} = f(P), \quad \text{where} \quad f(P) = kP\left(1 - \frac{P}{N}\right). \tag{1.7}$$

The graph of the function f (with $k = 3$ and $N = 4$) is given in Figure 1.2.

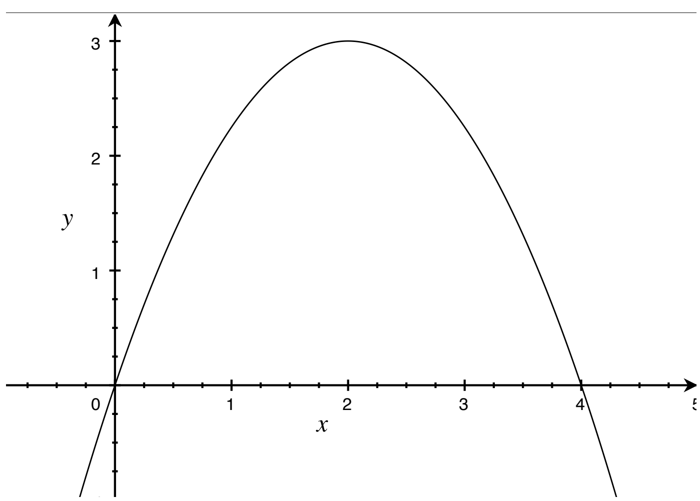


FIGURE 1.2. The graph of f with $k = 3$ and $N = 4$.

It is not hard to see that assumptions (1) and (2) are satisfied. Equation (1.7) is called a logistic equation. It is an example of a non-linear differential equation. Even though we could again separate the variables and find an explicit formula for all solutions, we will not do that here (see Section 1.4 for that approach). Instead, we employ only the qualitative approach to show its power and simplicity.

As in the previous example, let us first find the equilibrium solutions. Recall that a solution $P(t)$ is an equilibrium solution if it is constant, i.e., if $P'(t) = 0$ for all t. But if $P' = 0$, then $f(P) = 0$, so we can find equilibrium solutions by solving the *algebraic* equation $f(P) = 0$, which in our case is

$$kP\left(1 - \frac{P}{N}\right) = 0.$$

And that's a pretty easy task: the only solutions are $P = 0$ and $P = N$. Therefore, there are only two equilibrium solutions: the trivial one (when the population is always zero) and the "ideal one" when the population equals the carrying capacity.

What if the initial population P_0 is different from 0 and N? Here's how to approach this question. Suppose that at some time t the population is P. If $0 < P < N$, then $f(P) > 0$, so $P' = f(P) > 0$, which means that P is increasing. If $P > N$, then $P' = f(P) < 0$, which means that P is decreasing. Thus solutions whose value is in the interval $(0, N)$ are increasing and those in the interval (N, ∞) are decreasing. Here's a crucial question: if $0 < P < N$, can the solution increase beyond N? Or: if $P > N$, can it decrease below N? Let's see.

Assume we have a solution $P(t)$ such that $0 < P(0) < N$ but at some point in the future $P(t_+) > N$. Solutions are continuous (in fact, differentiable!) functions, so there must be some intermediate time t_0 such that $P(t_0) = N$. (What theorem of calculus guarantees this?) Therefore, when $t = t_0$, our solution and the equilibrium solution $P_*(t) \equiv N$ cross! This would mean that after time $t = t_0$ our population could "choose" to follow the solution P or it could "decide" to stay at the equilibrium N. Population growth would not be a deterministic process! This is clearly nonsense, so solutions cannot cross. (Our conclusion is related to uniqueness of solutions, which will be discussed in Section 1.6.) Thus if a solution P satisfies $0 < P(t) < N$ for some t, satisfies the same inequality for *all* t. Similarly, if $P(t) > N$ for some t, the same holds for *all* t.

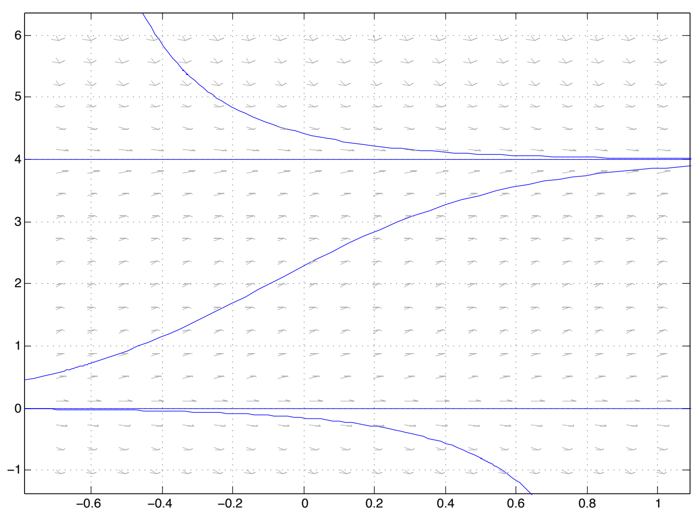


FIGURE 1.3. Graphs of several solutions of the logistic equation (1.7) with $k = 3$ and $N = 4$.

We draw the following conclusions (see Figure 1.3):

- There are two equilibrium solutions, $P = 0$ and $P = N$.
- If $0 < P < N$, the solution increases and converges to N, as $t \to \infty$. (Where does it converge as $t \to -\infty$?)
- If $P > N$, the solution decreases and converges to N, as $t \to \infty$.
- Finally, if $P < 0$ (ignoring the physical impossibility of this), then we similarly obtain that the solution must decrease to $-\infty$, as $t \to \infty$. □

1.2.3. The harmonic oscillator. There are many harmonic oscillators in nature. We will describe the so called mass-spring oscillator consisting of a mass m attached by a spring to a wall. If the mass is displaced from its natural equilibrium position, it will oscillate. Our job is to find a mathematical model for this physical system and explore its properties. We will do the former in this section and the later in Chapter 3.

Denote by $y(t)$ the displacement of the mass m from its equilibrium position at time t. Place the equilibrium at the origin of the real line and assume that the distance increases towards the right, assuming the wall is on the left as in Figure 1.4.

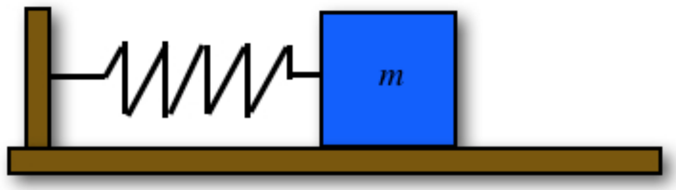


FIGURE 1.4. A mass-spring oscillator.

The motion of the oscillator is governed by NEWTON'S SECOND LAW:

The net force on an object equals its mass times its acceleration.

In mathematical language,

$$F = ma. \tag{1.8}$$

We know that a is the second derivative of the position:

$$a = \frac{d^2y}{dt^2} = y''$$

Let's figure out what the net force F is. To make the problem tractable, we need to make some simplifying assumptions such as these: the only forces that matter are friction (or damping) F_f, the force exerted by the spring, F_s, and an externally applied force F_e. Thus we assume

$$F = F_f + F_s + F_e.$$

By Hooke's Law, the force needed to extend a spring by a certain distance is proportional to that distance. Therefore,

$$F_s = -ky,$$

where $k > 0$ is a constant called the stiffness of the spring. Note the negative sign; it's there because the spring is trying to return the mass to its equilibrium position (which is at $y = 0$), so when $y > 0$, $F_s < 0$ and when $y < 0$, then $F_s > 0$.

To characterize the frictional force F_f, we will use its simplest approximation and assume that it is proportional to the velocity of the object. In other words,

$$F_f = -by',$$

where $b > 0$ is the damping constant (or the coefficient of friction).

The external force F_e cannot be specified in advance so we are only going to assume that it depends on time, i.e., $F_e = f(t)$, where f is some function.[4]

Substituting back in (1.8), we obtain

$$-ky - by' + f(t) = my'',$$

which is equivalent to

[4] $f(t)$ is often periodic though it can be any kind of function, including a discontinuous one.

$$my'' + by' + ky = f(t).$$

This is the mass-spring oscillator equation. Observe that it's a *second* order equation; we will study these types of equations in Chapter 3. □

Exercises

1. Consider the population model
$$\frac{dP}{dt} = 0.2P\left(1 - \frac{P}{135}\right),$$
where $P(t)$ is the population at time t.
   (a) For what values of P is the population in equilibrium?
   (b) For what values of P is the population increasing?
   (c) What is the carrying capacity?
   (d) For which initial values of P does the population converge to the carrying capacity as $t \to \infty$?

2. Consider the differential equation
$$y' = y^3 + 3y^2 - 10y.$$
   (a) What are the equilibrium solution?
   (b) For which values of y is $y(t)$ increasing?
   (c) For which values of y is $y(t)$ decreasing?

3. In this exercise your task is to model *radioactive decay*. Use the following notation:
   t = time (independent variable);
   $I(t)$ = amount of the radioactive isotope at time t (dependent variable);
   $-\lambda$ = decay rate, where $\lambda > 0$ (parameter).
   Write the simplest equation modeling the decay of a radioactive isotope with decay rate $-\lambda$. State the corresponding initial-value problem if the initial amount of the isotope is I_0.

4. The half-life of a radioactive isotope is the amount of time it takes for a quantity of that isotope to decay to one half of its original value.
   (a) Express the half-life of a radioactive isotope in terms of its decay rate.
   (b) The half-life of radiocarbon or Carbon 14 (C-14) is 5230 years. Determine its decay rate parameter λ.
   (c) Carbon dating is a method of determining the age of an object using the properties of radiocarbon. It was pioneered by Willard Libby and collaborators in 1949 to date archaeological, geological, and other samples. Its main idea is that by measuring the amount of radiocarbon still found in the organic mater and comparing it to the amount normally found in living matter, we can approximate the amount of time since death occurred.[5] Using the decay-rate parameter found in part (b), find the time since death if 35% of radiocarbon is still in the sample.

[5] This is based on the assumption that radiocarbon makes a constant proportion of the carbon ingested by living matter *and* that once the matter dies no new carbon is added to it.

1.3. The notion of a solution

So far we have been a little imprecise and have freely talked about solutions to various differential equations without ever defining what we meant by it. So suppose we have a first-order differential equation

$$y' = f(t, y). \tag{1.9}$$

What do we mean when we say that $y(t)$ is a solution? The answer quite natural: $y(t)$ is a solution to (1.9) if when plugged into (1.9), $y(t)$ *satisfies* it, i.e., the left-hand side becomes identically and trivially equal to the right-hand side:

$$y'(t) = f(t, y(t)),$$

for all values of t for which both sides make sense. Here are some examples.

1.1. Example. Consider $y' = 2y$. Then $y(t) = e^{2t}$ is a solution as one can easily check. So are the functions $-e^{2t}$ and $2014e^{2t}$, but not t^2, e^{-t} or $\sin t$. (Note that this is just an equation modeling unlimited population growth as in Section 1.2.1.) □

1.2. Example. The function $y(t) = \tan t$ is a solution to the equation $y' = 1+y^2$. This equation has no equilibrium solutions. (Check this!) □

1.3. Example. Define a function $z(t)$ by

$$z(t) = \begin{cases} 0 & \text{if } t \leq 0 \\ t^3 & \text{it } t > 0. \end{cases}$$

Verify that z satisfies the differential equation $z' = 3z^{2/3}$. □

We have already encountered the notion of an equilibrium solution but because of its importance we repeat it definition:

> A solution is called an **equilibrium solution** if it is constant. A point y_0 is called an **equilibrium point** of a differential equation if $y(t) \equiv y_0$ is a solution.

How do we look for equilibrium solutions? Suppose that $y(t) = y_0$ is an equilibrium solution of the differential equation $y' = f(t, y)$. Plugging $y(t)$ into the equation we obtain

$$f(t, y_0) = f(t, y(t)) = y'(t) = \frac{d}{dt} y_0 = 0,$$

for all t. Conversely, suppose we found y_0 such that $f(t, y_0) = 0$, for all values of t. Then the above calculation also shows that $y(t) = y_0$ is an equilibrium solution of the equation $y' = f(t, y)$. We conclude:

> *y_0 is an equilibrium point of the equation $y' = f(t, y)$ if and only if $f(t, y_0) = 0$, for all values of t.*

In particular, if the equation is autonomous, $y' = f(y)$, then finding equilibrium points amounts to solving the equation $f(y) = y$.

Observe that checking whether a given function is a solution is easy, but finding solutions is usually not. For instance, does the equation

$$y' = 1 + y^4 \sin^2 y + 17e^{\arctan y^3}$$

have a solution satisfying the condition $y(0) = 0$? The answer is yes, but there's no way we can tell using the tools we have available at this point.

This brings us to another notion of a solution. Often it is important to find a solution to a differential equation which satisfies an *initial condition* of the type $y(0) = y_0$ or more generally $y(t_0) = y_0$. We call the problem of finding a solution to the pair of equations

$$y' = f(t, y), \qquad y(t_0) = y_0$$

an **initial-value problem**, which we will sometimes abbreviate by IVP. A solution to an initial-value problem is sometimes called a **particular solution**. We will also use the term "general solution": an expression with parameters which describes all possible solutions to a differential equation is called the **general solution**.

1.4. Example. The general solution to the equation $\dfrac{dP}{dt} = kP$ is $P(t) = Ae^{kt}$, where A is a constant. (Recall that in fact A is just $P(0)$, the initial value of the solution.) Let us prove this.

Assume that $Q(t)$ is any solution to the above equation. Define $R(t) = Q(t)e^{-kt}$. Then by the product rule,

$$R'(t) = Q'(t)e^{-kt} - kQ(t)e^{-kt} = \{Q'(t) - kQ(t)\}e^{-kt}.$$

But $Q(t)$ is a solution to $P' = kP$, so $Q'(t) - kQ(t) = 0$. Therefore, $R'(t) = 0$, so $R(t)$ must be constant, say $R(t) \equiv A$. Multiplying both sides by e^{kt}, we obtain $Q(t) = Ae^{kt}$, as claimed. Thus *every* solution is of the form Ae^{kt}. □

Now that we have defined the notion of a solution, it's time to ask:

> *Does every differential equation have a solution? Does every initial-value problem have a solution? If so, is it unique?*

We will discuss these questions in Section 1.6. But first, let's learn how to solve some simple, though common types of differential equations.

Exercises

1. Verify the claim in Example 1.3.

2. Find a function $g(t)$ such that $y(t) = te^{t^2}$ is a solution to the differential equation
$$y' = g(t)y.$$

3. Find a function $h(y)$ such that the function $y(t) = e^{2t}$ is a solution to the equation
$$y' = 2y - t + h(y).$$

1.4. Separable equations

Recall that in a differential equation we distinguish two types of variables: the independent variable t and the dependent variable (or the unknown) y. A separable differential equation is simply an equation in which these variables can be separated, in the sense that the right-hand side can be written as a product of functions of each individual variable. So separable equations are equations of the form

$$\frac{dy}{dt} = g(t)h(y).$$

For instance, the equation $y' = t^2 y$ is separable, whereas $y' = t^3 + y^2$ is not.

Not surprisingly, the method for solving separable equations is called separation of variables. It goes as follows.

STEP 1: Find all equilibrium solutions by solving the equation $h(y) = 0$. If y_0 is a root[6] of this equation, then $y(t) \equiv y_0$ is an equilibrium solution.

STEP 2: Next we look for the non-equilibrium solutions. We therefore assume $h(y) \neq 0$, divide both sides of the equation by $h(y)$, and pretending that dy/dt is a quotient of real numbers, "multiply" both sides by dt. This gives

$$\frac{dy}{h(y)} = g(t)\, dt.$$

We have separated the variables. Both sides are expressions ready to be integrated.

STEP 3: Integrate both sides, treating y and t as "dummy variables" in their respective integrals:

$$\int \frac{dy}{h(y)} = \int g(t)\, dt.$$

Since we always assume that our differential equations are nice, $1/h(y)$ and $g(t)$ are at least continuous functions, so by calculus they have anti-derivatives, which we denote by $H(y)$ and $G(t)$. Whether these anti-derivaties are easy to find is another question. Assuming we are lucky and that we can integrate, we get

$$H(y) = G(t) + C, \tag{1.10}$$

where C is a constant. And we are done! But wait, how does this solve our differential equation? The answer is, equation (1.10) defines one or more solutions *implicitly*. That is, solutions to the differential equation can be found as solutions to equation (1.10), which is not a differential equation and is hopefully easier to solve than the original equation. Sometimes we may get even more lucky and be able to solve for y explicitly as a function of t (which is what we would ideally like), but in many cases we just won't.

REMARK. Since in the method of separation of variables we used a somewhat dubious idea of thinking of the derivative dy/dt as a quotient of differentials, let us verify that solutions to (1.10) are indeed solution to our differential equation. So assume $y(t)$ solves (1.10) in the sense that there is an interval I such that for all t in I,

$$H(y(t)) = G(t) + C.$$

Assuming we can differentiate both sides (which indeed we can, thanks to something called the Implicit Function Theorem), and using the Chain Rule, we obtain

$$H(y(t))y'(t) = G'(t).$$

Now recall that G and H are antiderivatives of g and $1/h$, respectively, i.e., $G' = g$ and $H' = 1/h$. It therefore follows that

$$\frac{1}{h(y(t))}y'(t) = g(t).$$

[6] Root means solution, though the term root is usually reserved for algebraic equations.

Multiplying both sides by $h(y(t))$, we obtain $y'(t) = g(t)h(y(t))$, i.e., $y(t)$ does satisfy the original equation, as claimed.

Here are some examples.

1.5. Example. Consider the equation

$$y^2 y' = t.$$

Since we can write $y' = t/y^2$, the equation is clearly separable with $g(t) = t$ and $h(y) = 1/y^2$. Observe that $h(y) \neq 0$, for all $y \neq 0$ ($h(0)$ is not even defined), so there are no equilibrium solutions. Separating the variables we obtain

$$\int y^2 \, dy = \int t \, dt.$$

Integrating both sides gives

$$\frac{y^3}{3} = \frac{t^2}{2} + C.$$

Multiplying both sides by 3, setting $K = 3C$ and solving for y, we obtain

$$\boxed{y = \left(\frac{3t^2}{2} + K \right)^{1/3}.}$$

In this example we were pretty lucky and were able to find an explicit formula for all solutions.

1.6. Example. Consider the ODE $(1 + y^2)y' = y$. It is clearly separable with $g(t) = 1$ and $h(y) = y/(1 + y^2)$. Following the procedure for the method of separation of variables, we first solve the equation $h(y) = 0$. This gives us the only equilibrium solution $y(t) = 0$. Next, for $y \neq 0$, we separate the variables and integrate:

$$\int \frac{1 + y^2}{y} \, dy = \int dt.$$

The integrand on the left-hand side is $\dfrac{1}{y} + y$, so integrating we obtain

$$\boxed{\ln |y| + \frac{y^2}{2} = t + C.} \tag{1.11}$$

Try solving this equation for y! Or better: don't even try. This is a hard possibly unsolvable transcendental equation. So we are stuck with an *implicit* solution. It is possible but not easy to show that for any real number C there are two solutions $y_+(t)$ and $y_-(t)$ to (1.11): $y_+(t)$ is positive and increasing, $y_+(t) \to \infty$, as $t \to \infty$, $y_+(t) \to 0$, as $t \to -\infty$, and $y_-(t) = -y_+(t)$, for all t. See Figure 1.5.

1.7. Example. Consider the equation $y^2 y' = te^{-y^3}$. It is not hard to see that it is separable and that it has *no equilibrium solutions*. Separating the variables and integrating we get

$$\int y^2 e^{y^3} \, dy = \int t \, dt.$$

You should be able to see right away that the substitution $u = y^3$ solves the integral on the left-hand side. (If you cannot see that right away, you need to review basic techniques of integration from Calculus 1 and 2.) We get $du = 3y^2 \, dy$, so

$$\int \frac{1}{3} e^u \, du = \int t \, dt.$$

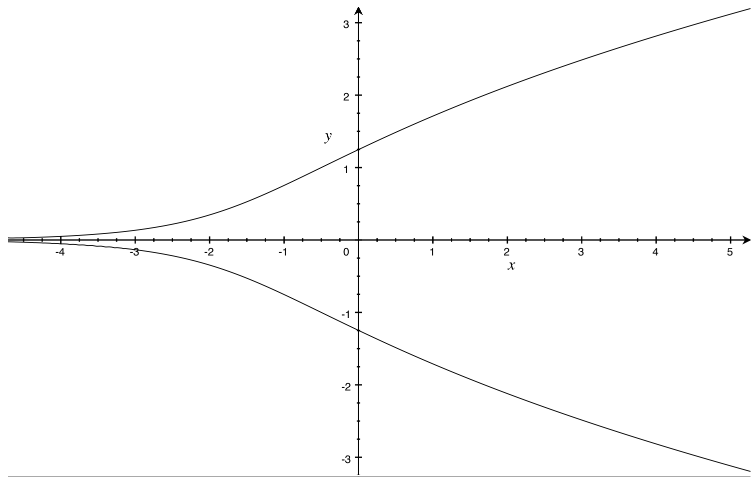


FIGURE 1.5. Graphs of solutions y_+ and y_- to (1.11).

These are easy integrals:

$$\frac{1}{3}e^u = \frac{t^2}{2} + C.$$

Of course, we want to solve for the original unknown variable y, so we substitute back:

$$\frac{1}{3}e^{y^3} = \frac{t^2}{2} + C.$$

We could stop here but we won't because this equation is solvable for y. Indeed, multiplying both sides by 3, taking the ln and cube roots of both sides, we obtain

$$\boxed{y(t) = \sqrt[3]{\frac{3t^2}{2} + K},}$$

where $K = 3C$. What if we want the solution which satisfies $y(0) = 2$? No problem: substitute $t = 0$ and $y(0) = 2$ into the above equation to obtain $2 = \sqrt[3]{K}$, which implies $K = 8$. Thus the unique solution to the given IVP is

$$y_*(t) = \sqrt[3]{\frac{3t^2}{2} + 8}.$$

1.8. EXAMPLE (Logistic equation). Let us revisit the logistic equation

$$\frac{dP}{dt} = kP\left(1 - \frac{P}{N}\right).$$

The equation is separable, so let's try to solve it. For simplicity, we choose units with respect to which $N = 1$, so we are looking at the equation

$$\frac{dP}{dt} = kP(1 - P).$$

Separating the variables and integrating we obtain

$$\int \frac{dP}{P(1-P)} = \int k\,dt.$$

To solve the integral on the left-hand side we use the method of partial fractions:

$$\begin{aligned}\int \frac{dP}{P(1-P)} &= \int \left(\frac{1}{P} + \frac{1}{1-P}\right) dt \\ &= \ln|P| - \ln|1-P| + A \\ &= \ln\left|\frac{P}{1-P}\right| + A,\end{aligned}$$

where A is an arbitrary constant. The integral on the right-hand side equals $kt + B$, where B is an arbitrary constant. Taking $C = B - A$, we obtain

$$\ln\left|\frac{P}{1-P}\right| = kt + C$$

which is equivalent to

$$\left|\frac{P}{1-P}\right| = e^C e^{kt}.$$

Recall that for any real number x, $|x| = \pm x$. So if we set $K = \pm e^C$ (just another arbitrary constant), we obtain

$$\frac{P}{1-P} = Ke^{kt}.$$

We can now solve this equation for $P = P(t)$:

$$P(t) = \frac{Ke^{kt}}{1 + Ke^{kt}}. \tag{1.12}$$

Recall that we are interested in the behavior of solutions as a function of their initial value, so let's express K in terms of $P_0 = P(0)$. To do that substitute $t = 0$ into the last equation and obtain

$$P_0 = \frac{K}{1+K}.$$

Solving for K we get

$$K = \frac{P_0}{1 - P_0}.$$

Plugging this expression into (1.12) and simplifying, we obtain an explicit solution to the logistic equation with $N = 1$:

$$P(t) = \frac{P_0 e^{kt}}{1 - P_0 + P_0 e^{kt}}. \tag{1.13}$$

Let's investigate the behavior of solutions as $t \to \infty$ for different initial values P_0:

$P_0 = 0$: In this case $P(t) \equiv 0$ so we have an equilibrium solution.

$P_0 = 1$: It is not hard to check that $P(t) \equiv 1$, which is another equilibrium solution (corresponding to the ideal population or carrying capacity).

$0 < P_0 < 1$: In this case the denominator in (1.13) is always positive, so $P(t)$ is defined for *all values of t*. Letting $t \to \infty$ we obtain $P(t) \to 1$. The population tends to the ideal population.

$P_0 > 1$: As in the previous case we can show that $P(t) \to 1$, as $t \to \infty$.

$P_0 < 0$: This case is of course not realistic, but for completeness we investigate it anyway. As in the previous two cases we could just let $t \to \infty$ and conclude that $P(t) \to 1$. But this would mean that the population starts off negative but miraculously becomes positive and then approaches the ideal population $N = 1$! Something is obviously very wrong here. Let's investigate what it is.

If $P_0 < 0$, then by setting the denominator in (1.13) to zero,

$$1 - P_0 + P_0 e^{kt} = 0$$

we see that when $t = t_0$, where

$$t_0 = \frac{1}{k} \ln \frac{P_0 - 1}{P_0}$$

the denominator equals zero. This means that as $t \to t_0$, the denominator goes to zero so the solution blows up. This phenomenon is known as finite-time blow up; in particular, solutions are not defined for all values of t: if $P_0 < 0$, then $P(t)$ is defined only for $t < t_0$ (so it does not make sense to let $t \to \infty$). With a little care (we leave this to the reader) it is possible to show that $P(t) \to -\infty$, as $t \to t_0+$ (that is, if t approaches t_0 from the right).

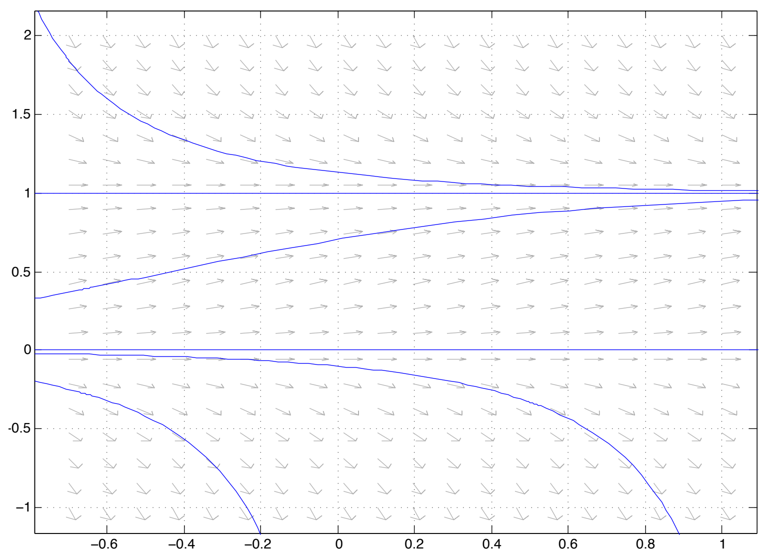


FIGURE 1.6. Typical solutions to the logistic equation with $N = 1$ and $k = 2$.

Figure 1.6 shows the behavior of several typical solutions. (Compare with Figure 1.3.) □

We conclude this section with some remarks.

REMARK.

(a) Most differential equations are *not* separable.

(b) Most separable equations, even those of the simplest form $y' = g(t)$ (with $h(y) = 1$), cannot be solved explicitly. Take for example $y' = e^{t^2}$. We know from Calculus that the integral $\int e^{t^2}\, dt$ is unsolvable, in the sense that it cannot be expressed in terms of elementary functions.[7]

[7]Recall that an elementary function is a function which can be expressed in terms of a finite number of polynomials, roots, sines, cosines, and logarithms using the basic arithmetic operations of addition, subtraction, multiplication and division, and composition of functions.

(c) Still, separation of variables *is* often quite useful!

Exercises

In Exercises 1-4 find the general solution of the given differential equations.

1. $y' = 1 - y$
2. $\dfrac{dz}{dt} = 4 + z^2$
3. $\dfrac{dy}{dt} = y^2 - 9$
4. $\dfrac{dy}{dt} = \dfrac{1}{t^2y + t^2 + y + 1}$

In Exercises 5-8 sove the given initial-value problems. equations.

5. $\dfrac{dy}{dt} = y^3, \quad y(0) = 0$
6. $\dfrac{dy}{dt} = y^3, \quad y(0) = -2$
7. $\dfrac{dy}{dt} = t(y^2 + 4), \quad y(0) = 2$
8. $\dfrac{dy}{dt} = \dfrac{1 + y^2}{y}, \quad y(0) = \sqrt{e - 1}$

9. A 10 liter bucket is full of pure water. Suppose we start adding salt to the water at a rate of 0.2 kilograms per minute. Assume also that every minute 1 liter of the salt water solution is leaking from the bucket, but that we are adding pure water to keep the bucket full. The salt water solution is always kept well mixed.

(a) Set up a differential equation modeling the amount of salt in the bucket.
(b) What is the amount of salt in the bucket after 6.93147 minutes? (Note that $6.93147 \approx 10 \ln 2$.)
(c) What is the amount of salt in the bucket after an extremely long time?

1.5. Linear equations

Recall that the general form of a first-order ordinary differential equation is

$$y' = f(t, y),$$

where f is some function of two variables, t and y. If f is a complicated function of y (or even if it is not), it may be impossible to solve this equation in closed form. However, if f is the simplest possible function of y – that is, a *linear* function – it turns out that it is always possible to write down the general solution at least in terms of integrals. Now, what does it mean for $f(t, y)$ to be a linear function of y? To answer this question, let us fix t and think of it as constant, and assume that $f(t, y)$ as a function of y can be expanded into its Taylor series:

$$f(t, y) = a_0(t) + a_1(t)y + a_2(t)y^2 + \cdots ,$$

where the Taylor coefficients $a_0(t), a_1(t), a_2(t), \cdots$ depend on t (since $f(t, y)$ in general depends on t). If this series does not go on forever but stops after the 0^{th} and first terms,

$$f(t, y) = a_0(t) + a_1(t)y,$$

we say that f is a linear function of y.[8] For the same reason a differential equation of the form

$$y' = a_0(t) + a_1(t)y$$

is called a linear differential equation of first order. (There are also linear differential equations of second, third, etc. order.) The functions $a_0(t)$ and $a_1(t)$ are called the coefficients of the equation. They are allowed to be any functions of t (though we will always assume they are at least continuous).

[8] More accurately, in this case f is called an *affine* function of y, but we don't want to use any fancy terminology.

1.9. Example. The equations

$$y' = \frac{y}{t^2} + \cos t, \qquad ty + 3 = y' - e^t y, \qquad y' = e^{t^2} y + \frac{\sin t}{1 + t^2 + t^4}$$

are linear, while

$$y' = \frac{y^2}{t^2} + \cos t, \qquad yy' = t + 1, \qquad y' = e^y + 3t$$

are not. Observe that it is the dependence on y, not on t, that matters.

It is standard to write a linear differential equation in the form

$$y' + p(t)y = q(t). \tag{1.14}$$

If $q(t) \equiv 0$, the equation is called homogeneous. To each linear equation (1.14) we can associate the homogeneous equation

$$y' + p(t)y = 0. \tag{1.15}$$

This equation is separable and easy to solve. Indeed, separating the variables and integrating, we get

$$\int \frac{dy}{y} = -\int p(t)\, dt,$$

which yields

$$\boxed{y = C \exp\left(-\int p(t)\, dt\right),}$$

where exp denotes the exponential function (i.e., $\exp(x) = e^x$) and C is an arbitrary constant. We will present two methods for solving linear equations.

Method 1. Consider the linear equation (1.14). We know how to solve the associated homogeneous equation $y' + p(t)y = 0$. Denote its general solution by y_h. Suppose now that we are able to find just one solution y_p – usually called a particular solution – to the non-homogenous equation (1.14). We claim that the general solution y_{general} to (1.14) is then of the form $y_h + y_p$. Let's see why that is the case.

Suppose y is an arbitrary solution to the non-homogeneous equation (1.14). Consider the function $z = y - y_p$. Since both y and y_p satisfy (1.14), we have:

$$\begin{aligned} z' &= y' - y_p' \\ &= -p(t)y + q(t) - \{-p(t)y_p + q(t)\} \\ &= -p(t)(y - y_p) \\ &= -p(t)z, \end{aligned}$$

so z satisfies the homogeneous equation $z' + p(t)z = 0$. Therefore, $z = y_h$, for some choice of the constant C in the general solution y_h. This implies that $y - y_p = y_h$, hence $y = y_h + y_p$, proving that an arbitrary solution y to (1.14) has to be of the form $y_h + y_p$.

So it is enough to find a particular solution y_p. We could try to guess it but that may not work so well (though in some situations it does; for example, if $p(t)$ is constant and $q(t)$ is of the form Ae^{kt} or $B \sin \omega t + C \cos \omega t$). Instead we will us use a different approach, sometimes called *variation of constants* (or *parameters*). It goes like this.

Since the general solution of the associated homogeneous equation $y' + p(t)y = 0$ is

$$y = Cu(t), \qquad \text{where} \qquad u(t) = \exp\left(-\int p(t)\, dt\right),$$

perhaps we can find a particular of the equation $y' + p(t)y = q(t)$ by *varying the constant* (!) C, i.e., by replacing it by a bona fide function of t, which we can call $c(t)$. In other words, we want to find a function $c(t)$ such that

$$y_p = c(t)u(t)$$

solves (1.14). To find $c(t)$, we plug $y_p = c(t)u(t)$ into equation (1.14). First we need to find y_p'; by the Product Rule we get

$$y_p' = c'(t)u(t) + c(t)u'(t).$$

Thus

$$\begin{aligned} y_p' + p(t)y_p &= c'(t)u(t) + c(t)u'(t) + p(t)c(t)u(t) \\ &= c'(t)u(t) + c(t)\{u'(t) + p(t)u(t)\} \\ &= c'(t)u(t) + c(t) \cdot 0 \\ &= c'(t)u(t), \end{aligned}$$

since $u' + p(t)u = 0$. We conclude that $y_p' + p(t)y_p = q(t)$ if $c'(t)u(t) = q(t)$. Dividing both sides of the equation $c'(t)u(t) = q(t)$ by $u(t)$ and integrating we obtain

$$c(t) = \int \frac{q(t)}{c(t)}\, dt = \int q(t) \exp\left(\int p(t)\, dt\right)\, dt.$$

Since we have computed $c(t)$, we now have a particular solution y_p and thus the general solution y_{general} to (1.14).

Let us summarize this method. To find the general solution to the non-homogeneous equation $y' + p(t)y = q(t)$, perform the following steps:

STEP 1: Find the general solution $y_h = Cu(t)$ of the associated homogeneous equation $y' + p(t)y = 0$, as we did above.

STEP 2: Find a particular solution y_p of the non-homogeneous equation in the form $y_p = c(t)u(t)$, where $u(t)$ is as in Step 1. To find $c(t)$ solve the equation $c'(t)u(t) = q(t)$.

STEP 3: The general solution to (1.14) is

$$y_{\text{general}} = y_h + y_p. \tag{1.16}$$

REMARK. This idea to break up the general solution of the non-homogenous equation into the sum of its particular solution and the general solution of the associated homogeneous equation will be useful later, in Chapter 3, and where it will be used to solve second order linear equations.

METHOD 2. This method should of course give us the same answer as Method 1. The main idea is the following: Even though the left-hand side of the equation may not be the derivative of any function, perhaps we can *multiply the equation by some, yet to be determined function* $\mu(t)$ *so that the left-hand side becomes the derivative of* $\mu(t)y$. We will call such a function $\mu(t)$ an integrating factor.

So let's multiply both sides of (1.14) by $\mu(t)$ and figure out what $\mu(t)$ has to be in order for the left-hand side of the new equation to be of the form $\frac{d}{dt}[\mu(t)y]$. We obtain

$$\mu(t)t' + \mu(t)p(t)y = \mu(t)q(t).$$

We require that the left-hand side of this equation be equal to the derivative of $\mu(t)y$:

$$\mu(t)t' + \mu(t)p(t)y = \frac{d}{dt}[\mu(t)y].$$

Since $\frac{d}{dt}[\mu(t)y] = \mu'(t)y + \mu(t)y'$, we get

$$\mu(t)t' + \mu(t)p(t)y = \mu'(t)y + \mu(t)y'.$$

Canceling out $\mu(t)y'$, we obtain the equation

$$\mu(t)p(t)y = \mu'(t)y.$$

This is an equation for $\mu(t)$, not y. Assuming $y \neq 0$ (otherwise, things are not very interesting), switching sides and dividing by y, we obtain

$$\mu'(t) = \mu(t)p(t).$$

Note that this is a separable equation in μ. One solution of this equation is the trivial one, $\mu(t) \equiv 0$. But this is not the solution we are looking for (what would be the point of multiplying (1.14) by zero!?), so we may as well assume that $\mu(t) \neq 0$ and divide both sides by $\mu(t)$ to obtain

$$\frac{\mu'(t)}{\mu(t)} = p(t).$$

The left-hand side is the derivative of $\ln|\mu(t)|$, but since we just want one integrating factor $\mu(t)$, we may as well assume that $\mu(t) > 0$ in which case the left-hand side is the derivative of $\ln \mu(t)$ (no absolute value signs). So we have

$$\frac{d}{dt} \ln \mu(t) = p(t),$$

which by integration (and one extra step left to the reader) implies

$$\mu(t) = \exp\left(\int p(t)\, dt\right).$$

This is our integrating factor. Since we want only one solution for $\mu(t)$, we have chosen to omit the constant of integration.

We *know* that the left-hand side of the new equation (the old equation multiplied by $\mu(t)$) equals $\frac{d}{dt}[\mu(t)y]$, so

$$\frac{d}{dt}[\mu(t)y] = \mu(t)q(t).$$

Integrating both sides we obtain

$$\mu(t)y(t) = \int \mu(t)q(t)\, dt + C.$$

Dividing both sides by $\mu(t)$, we obtain the general solution to (1.14):

$$y_{\text{general}}(t) = \frac{1}{\mu(t)}\left\{\int \mu(t)q(t)\, dt + C\right\}. \tag{1.17}$$

Remark.

(a) If $\mu(t)$ is an integrating factor, then so is $K\mu(t)$, where K is any constant.

(b) We leave it to the reader to check that the solution (1.17) obtained by Method 2 is identical to the solution (1.16) obtained by Method 1. Since we will get to practice Method 1 in Chapter 3, in the subsequent examples we will use only Method 2.

(c) To solve a linear differential equation you can either (i) compute $\mu(t)$ using the above formula, plug it into (1.17) and solve the integral in it (if possible) or (ii) go through the entire procedure outlined above, which means, compute $\mu(t)$, multiply both sides of the equation by $\mu(t)$, etc. Do what works best for you.

1.10. Example. The equation

$$y' + \frac{2}{t}y = t - 1$$

is linear, with $p(t) = 2/t$ and $q(t) = t - 1$. Solving for an integrating factor we obtain

$$\mu(t) = \exp\left(\int p(t)\,dt\right) = \int \frac{2}{t} dt = t^2,$$

so the general solution is

$$\begin{aligned} y &= \frac{1}{\mu(t)}\left\{\int \mu(t)q(t)\,dt + C\right\} \\ &= \frac{1}{t^2}\int t^2(t-1)\,dt \\ &\ \vdots \\ &= \boxed{\frac{t^2}{4} - \frac{t}{3} + \frac{C}{t^2}.} \qquad \square \end{aligned}$$

1.11. Example. Consider the equation

$$y' = t^2 y + t - 1.$$

The equation is linear with $p(t) = -t^2$ (note the sign!) and $q(t) = t - 1$. So

$$\mu(t) = \exp\left(\int (-t^2)\,dt\right) = e^{-t^3/3}.$$

Therefore the general solution is

$$y = \frac{1}{e^{-t^3/3}}\int e^{-t^3/3}(t-1)\,dt,$$

and we are stuck, since we $\int e^{-t^3/3}(t-1)$ is not an elementary function (i.e., the integral can't be solved)!

What if $q(t)$ were equal to t^2 instead of $t - 1$, i.e., what about the equation

$$y' = t^2 y + t^2?$$

We could treat this equation both as linear and as separable: $y' = t^2(y + 1)$. If we do the latter, we obtain

$$\int \frac{dy}{y+1} = \int t^2\,dt,$$

so the general solution is (check this!)

$$\boxed{y = Ce^{t^3/3} - 1.} \qquad \square$$

1.6. Existence and uniqueness of solutions

So far we have tacitly worked under the assumption that ordinary differential equations have solutions and that, if we also specify an initial condition (i.e., if we have an initial-value problem), the solution is unique. But how do we know this is true? For instance, how can we be so sure that the monstrosity of an equation

$$y' = e^{\arctan y^3} \sin(\ln(1 + y^2)) + y^{2014}$$

admits any solutions whatsoever? On the other hand, if we want to use differential equations to model physical phenomena, we better make sure that every differential equation used for modeling does have a unique solution for each initial condition! Otherwise, we cannot ever hope to predict the motion of stars and planets, trajectories of rockets, behavior of electrical circuits, etc.

But what exactly do we even mean by existence and uniqueness of solutions of differential equations? Let's look at a couple of examples.

1.12. Example. Consider the following differential equation:

$$\frac{dy}{dt} = \begin{cases} 1 & \text{if } y < 0 \\ -1 & \text{if } y \geq 0. \end{cases}$$

We claim that there exists no solution satisfying $y(0) = 0$. Assume that $y(t)$ was such a solution. Since $y'(0) = -1$, $y(t)$ is supposed to start by decreasing. However, as soon as it does that, $y(t)$ (for small $t > 0$) becomes negative and the equation demands that if $y(t) < 0$, then $y'(t) = 1$, so the solution needs to *increase*, not decrease. These two requirements are obviously incompatible, so the solution does not exist. If this slightly hand-waving argument is not convincing, here is a more rigorous one. Suppose $y(t)$ is a solution satisfying $y(0) = 0$. Since $y(t)$ must be differentiable (otherwise, how can it ever satisfy a *differential* equation?!), by Darboux's theorem from Calculus, its derivative y' has the Intermediate Value Property (IntVP).[9] Thus $y'(t)$ cannot take both -1 and 1 as values, so since $y'(0) = -1$, it follows that $y'(t)$ must always equal -1. Thus $y'(t) = -1$, for all t. Integrating and using $y(0) = 0$, we get $y(t) = -t$. However, this function is *not* a solution, as can be easily checked. Contradiction! Therefore, *the equation does not admit a solution satisfying* $y(0) = 0$. □

1.13. Example. Consider the initial-value problem

$$y' = 3y^{2/3}, \qquad y(0) = 0.$$

It is easy to check that $y(t) = t^3$ is a solution. For any non-negative real number a define the following function:

$$y_a(t) = \begin{cases} 0 & \text{if } t \leq a \\ (t-a)^3 & \text{if } t > a. \end{cases}$$

We claim that y_a is also a solution to the above IVP. Indeed, $y_a(0) = 0$ (remember that $a \geq 0$) and it is not hard to see that $y_a'(t) = 3y_a(t)^{2/3}$ if $t \neq a$ (both the zero function and the function $(t-a)^3$ satisfy the equation). Using the limit definition of derivative we can also verify that $y_a'(a) = 0$, so y_a satisfies the equation also when $t = a$. Thus there are *infinitely many solutions to this IVP.*

So it turns out that differential equations do not always have solutions and solutions to IVP's are not always unique! But observe that the right-hand side of the equation in Example 1.12 is discontinuous whereas the right-hand side in Example 1.13 although continuous is not everywhere

[9]A function f has the IntVP if for every two distinct values of f, any number in between is also a value of f. We call b a value of f if $b = f(a)$ for some a in the domain of f. Continuous functions have the IntVP but a function which has the IntVP is not necessarily continuous.

differentiable. Could that be the cause of the non-existence and non-uniqueness of solutions? It turns our that the answer to both questions is yes. This is due to the following fundamental result.

1.14. THEOREM (Existence and Uniqueness of Solutions). *Consider the differential equation*

$$y' = f(t, y).$$

(a) *Assume that f is continuous on the rectangle*

$$Q = (a, b) \times (c, d) = \{(t, y) : a < t < b, \quad c < y < d\}$$

and let (the initial point) (t_0, y_0) be in Q. Then there exists $\varepsilon > 0$ and a solution $y(t)$ defined for $t_0 - \varepsilon < t < t_0 + \varepsilon$ which satisfies the initial condition $y(t_0) = y_0$.

(b) *If both f and its partial derivatives with respect to t and y are continuous on Q, then for every $(t_0, y_0) \in Q$ the solution satisfying $y(t_0) = y_0$ is unique. That is, if both $y_1(t)$ and $y_2(t)$ are solutions satisfying the given initial condition, then*

$$y_1(t) = y_2(t),$$

for all t for which both sides are defined.

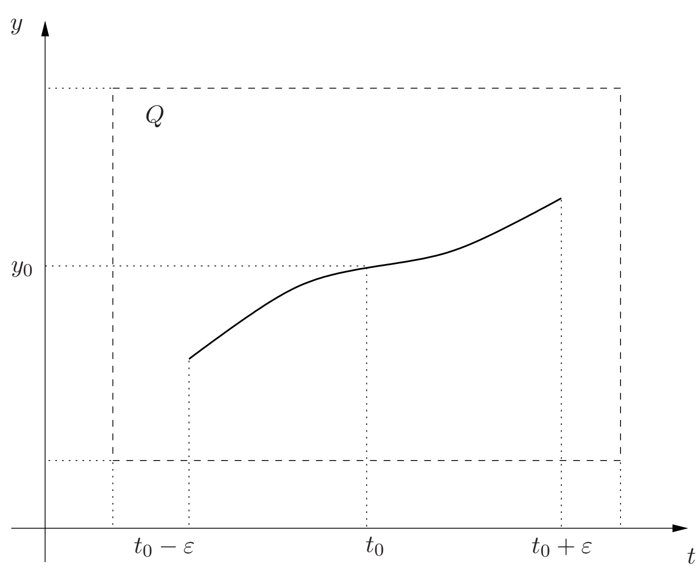


FIGURE 1.7. Solution satisfying the initial condition $y(t_0) = y_0$.

The moral of this theorem is: *if the right-hand side of the differential equation is sufficiently nice, the solutions exists and are unique* (for any given initial value). Part (a) says that the solution may be defined only for a short interval of time (with length 2ε). See Figure 1.7.

So even the nasty equation

$$y' = e^{\arctan y^3} \sin(\ln(1 + y^2)) + y^{2014} \tag{1.18}$$

admits solutions and for any initial value y_0, there is a unique solution satisfying $y(0) = y_0$! The reason is that $f(y) = e^{\arctan y^3} \sin(\ln(1 + y^2)) + y^{2014}$ although complicated, *is* a "nice" function: it has continuous partials with respect to both variables (of course, since it doesn't depend on t, its t-partial is just zero).

Think about it: we have no idea how to actually solve this equation, but we do know that its solutions *exist*. In what sense do they exist if we cannot construct them? That is a question for philosophers of mathematics, which we do not have time to address here.[10] However, we can certainly *approximate* them, which we will learn how to do in Section 1.9.

Here are some important consequences of the Existence and Uniqueness Theorem. Suppose that f satisfies the assumptions of Theorem 1.14. Then:

(1) *The graphs of distinct solutions never cross.* That is, if $y_1(t)$ and $y_2(t)$ are two solutions and $y_1(t) = y_2(t)$ for *some* t, then $y_1(t) = y_2(t)$ for *all* t for which both sides are defined. In other words, if two solutions are in the same place at the same time, then they are the same solution.

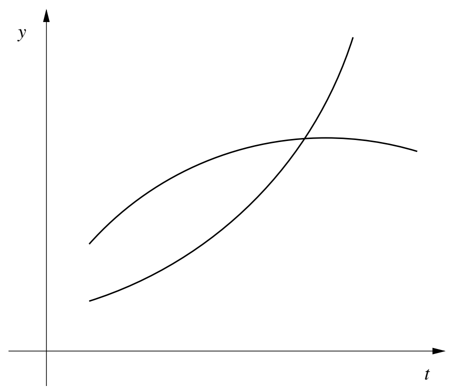


FIGURE 1.8. Solutions can't cross, so this scenario is impossible.

(2) Consider the (autonomous) equation $y' = f(y)$ and suppose that $f(y_0) = 0$. Let $y(t)$ be a solution such that $y(t_0) = y_0$ for some t_0. Then $y(t) = y_0$ for *all* t, i.e., y is an equilibrium solution.

(3) If $y_0(t), y_1(t)$ and $y_2(t)$ are solutions and $y_0(t) < y_1(t) < y_2(t)$ for *some* t, then $y_0(t) < y_1(t) < y_2(t)$ for *all* t. That is, if $y_1(t)$ is between two solutions at one instant of time, then it is trapped between these solutions for all time.

Here are some examples of applications of the Existence and Uniqueness Theorem.

1.15. EXAMPLE. Find the unique solution to (1.18) satisfying $y(0) = 0$.

SOLUTION: Let $f(y) = e^{\arctan y^3} \sin(\ln(1 + y^2)) + y^{2014}$. It is easy to check that $f(0) = 0$, so the zero function $y_*(t) \equiv 0$ is an equilibrium solution. Since $y_*(0) = 0$, by the Existence and Uniqueness Theorem $y_*(t)$ is *the* (unique) solution satisfying $y(0) = 0$. □

1.16. EXAMPLE. What can be said about the solution to the equation

$$y' = (y + 1)y(y - 2)$$

satisfying $y(0) = 1$?

SOLUTION: Let $f(y) = (y + 1)y(y - 2)$. This function does not depend on t and $f'(y)$ is clearly continuous (it's a polynomial), so the equation admits unique solutions. Since $f(-1) = f(0) =$

[10] There is a point of view called *Mathematical Platonism*, which (according to the Stanford Encyclopedia of Philosophy) is "the metaphysical view that there are abstract mathematical objects whose existence is independent of us and our language, thought, and practices. (...) Mathematical truths are therefore discovered, not invented."

$f(2) = 0$, the equilibrium solutions are $y_-(t) = -1$, $y_0(t) = 0$ and $y_+(t) = 2$. Consider the solution $y(t)$ satisfying $y(0) = 1$. Since $y_0(0) < y(0) < y_+(0)$, Remark (3) above implies that $y(t)$ is trapped between 0 and 2 forever. Moreover, since f is negative on $(0, 2)$, we have $f(y(t)) = y'(t) < 0$, so the solution is *decreasing*. Since $y(t)$ is bounded and decreasing, we know from Calculus that $y(t)$ must approach a limit, as $t \to \infty$. Let's call that limit L:

$$\lim_{t\to\infty} y(t) = L.$$

What value might L have? Well, it is clear that $0 \le L < 2$. Since $y(t)$ approaches a constant value

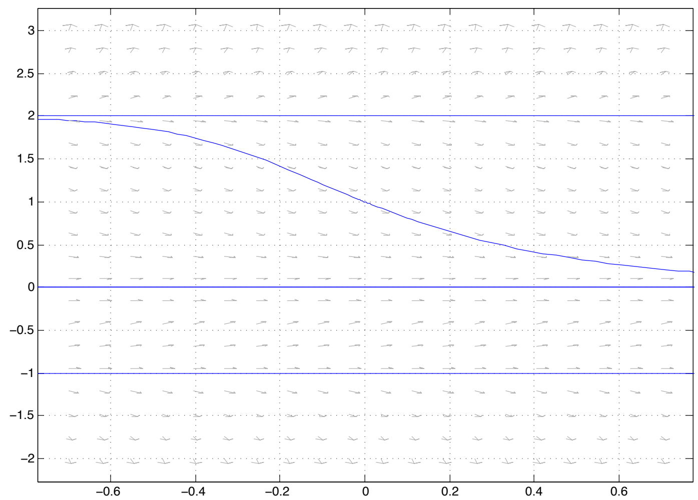


FIGURE 1.9. The graphs of equilibrium solutions $y_-(t) = -1, y_0(t) = 0$ and $y_+(t) = 2$, and the solution $y(t)$ satisfying $y(0) = 1$.

in a decreasing manner, $y'(t) \to 0$, as $t \to \infty$. So:

$$\begin{aligned} 0 &= \lim_{t\to\infty} y'(t) \\ &= \lim_{t\to\infty} f(y(t)) \\ &= f(\lim_{t\to\infty} y(t)) \\ &= f(L). \end{aligned}$$

Here we used the fact that $y(t)$ is a solution (second line) and that f is continuous (third line). Thus $f(L) = 0$ and since $0 \le L < 2$, we must have $L = 0$. In conclusion: the unique solution satisfying $y(0) = 0$ lies in the interval $(0, 2)$ for all time and approaches the equilibrium solution 0 as $t \to \infty$. In a similar way it can be shown that $y(t) \to 2$, as $t \to -\infty$. See Figure 1.9. □

REMARK. It can be shown in a similar way as in Example 1.16 that solutions to first order differential equations (in one variable) *can converge only to equilibrium solutions.*

Picard's method of successive approximations. The Existence and Uniqueness Theorem is an extremely general (hence powerful and important) theorem, so it is natural to ask: how in the world can one prove that a differential equation has solutions when we don't even know its right-hand side? Here we sketch the classical approach to this question, called Picard's method of successive approximations.[11]

[11]Émile Pickard (1856-1941) was a prominent French mathematician.

Suppose we are given an IVP

$$y' = f(t, y), \qquad y(0) = y_0, \tag{1.19}$$

and assume that f satisfies the assumptions of the Existence and Uniqueness Theorem. Denote the unique solution to the IVP by $y(t)$. At this point, we know nothing about y except that it exists in some abstract sense (living in some strange Platonic universe). However, we can construct a sequence of approximate solutions $y_0(t), y_1(t), y_2(t), \dots$ such that $y_k(t)$ converges to $y(t)$, as $k \to \infty$. This is done in the following way.

Since $y(t)$ is a solution defined on some interval $(-\varepsilon, \varepsilon)$, we have

$$y'(s) = f(s, y(s)),$$

for all $-\varepsilon < s < \varepsilon$. Let us integrate this equation with respect to s from 0 to t. We obtain

$$y(t) - y_0 = \int_0^t f(s, y(s))\, ds.$$

On the left-hand side we used the Fundamental Theorem of Calculus and the initial condition $y(0) = y_0$. Moving y_0 to the right-hand side, we see that $y(t)$ satisfies the following *integral* equation

$$y(t) = y_0 + \int_0^t f(s, y(s))\, ds. \tag{1.20}$$

Conversely, if $y(t)$ satisfies (1.20), then it satisfies the IVP (1.19). In other words, equations (1.19) and (1.20) are equivalent. That means that we may as well attempt to solve (1.20), which turns out to be better suited for constructing successive approximations.

We define our sequence $y_0(t), y_1(t), y_2(t), \dots$ recursively (or inductively) as follows:

$$\begin{aligned} y_0(t) &\equiv y_0 \\ y_{k+1}(t) &= y_0 + \int_0^t f(s, y_k(s))\, ds. \end{aligned}$$

Thus we first compute $y_0(t)$, which is just a constant function equal to the initial value y_0, then use the above formula to compute (at least theoretically: the integral may be hard) $y_1(t)$, then we plug $y_1(t)$ back into the recursive formula to find $y_2(t)$, and so on. It turns out that this works, namely, $y_k(t)$ converges to the true solution, at least for small enough t! Let's see how this plays out in a simple example.

1.17. Example. Let us apply Pickard's method to the IVP

$$y' = y, \qquad y(0) = 1.$$

Note that $f(t, y) = y$. We obtain:

$$\begin{aligned} y_0(t) &= 1 \\ y_1(t) &= 1 + \int_0^t 1\,ds = 1 + t \\ y_2(t) &= 1 + \int_0^t (1+s)\,ds = 1 + t + \frac{t^2}{2} \\ y_3(t) &= 1 + \int_0^t \left(1 + s + \frac{s^2}{2}\right) ds = 1 + t + \frac{t^2}{2} + \frac{t^3}{6} \\ &\vdots \end{aligned}$$

It is not hard to see (and even prove) that for a general $k \geq 1$, we have

$$y_k(t) = 1 + t + \frac{t^2}{2!} + \cdots + \frac{t^k}{k!},$$

where $k! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot k$. So as $k \to \infty$, $y_k(t)$ becomes an infinite series:

$$y_k(t) \to \sum_{n=0}^{\infty} \frac{t^n}{n!}.$$

We learned in calculus that the sum of this infinite series is $y(t) = e^t$, which we know is the unique solution to the above IVP. □

Completeness. In applications it is important to know if a model has solutions defined for all future (and sometimes past) time t. For if a differential equation has a solution $y(t)$ defined only up to some finite time t_0, it can undergo the phenomenon called a *finite-time blow up*. We already had two instances of that phenomenon: see Examples 1.8 and 1.1. Recall that in the latter one $y(t) = \tan t$ is the (unique) solution to the IVP $y' = 1 + y^2$, $y(0) = 0$.

If all solutions are defined for all time, the equation is called complete. We saw that even some very simple equations are not complete. The question is: under what conditions is a given equation complete? This is a non-trivial question so we will only mention the following basic result.

1.18. THEOREM. *(a) If $y(t)$ is a solution to $y' = f(y)$ and $y(t)$ is bounded (i.e., there exist numbers a, b such that $a \leq y(t) \leq b$ for all t), then $y(t)$ is defined for all t.*

(b) Let f be a differentiable function and assume its derivative f' is bounded (i.e., there is a constant $M > 0$ such that $|f'(y)| \leq M$, for all y). Then the differential equation $y' = f(y)$ is complete.

Note that in the above example $f(y) = 1 + y^2$, so $f'(y) = 2y$, which is not bounded (it goes to infinity as $y \to \infty$). However, the theorem guarantees that the equation

$$y' = (\arctan y) \sin y$$

is complete. (Check that!)

1.7. The phase line and classification of equilibria

In this section we focus on *qualitative analysis* of differential equations. This approach was pioneered by the the famous French mathematician Henri Poincaré (1854-1912), one of the leading mathematicians and scientists of his time.[12] Poincaré studied the Solar system hoping to show that it is stable (which among other things would mean that the Moon would never crash into the Earth or fly off into space). Instead what he discovered was that the Solar system, modeled by differential equations defined by Newton's law of gravitation and his second law of motion, exhibits extremely complicated behavior later named *chaotic*. (Remember Ian Malcolm, the Chaos Expert from Jurassic Park?) To understand this type of behavior of solutions (which in this case correspond to trajectories of the planets and their moons) Poincaré could not count on studying any explicit formulas, both because they were either hard to come by and because even if he had had them, they would have been too complicated to analyze. Instead he invented brand new methods and techniques which comprise what we now call qualitative analysis. In his plenary address at the 1908 International Congress of Mathematicians Poincaré said:

> *In the past an equation was only considered to be solved when one had expressed the solution with the aid of a finite number of known functions; but this is hardly possible one time in a hundred. What we can always do, or rather what we should always try to do, is to solve the qualitative problem so to speak, that is, to try to find the general form of the curve representing the unknown function.*

In other words, most differential equations cannot be solved explicitly in terms of elementary functions. What we should do instead of looking for explicit solutions is try to understand them from a qualitative point of view. This means that we should try to answer questions such as these:

- How many equilibrium solutions are there and what is their type?
- What is the limit of a "typical" solution as t goes to infinity?
- Is the given equation qualitatively the same as a simpler equation?

This type of approach to differential equations is facilitated by the concept of the *phase line*. Recall that so far we visually represented solutions via their graphs. Now we are going to look at the space in which all solutions "live". Of course if $y(\cdot)$ is a solution, then for each t, $y(t)$ is a just a real number. So solutions "live" on the real line. However, when regarded as the "living space" of solutions of a differential equation, the real line is called the phase line. To draw the phase line of an autonomous differential equation $y' = f(y)$ we procede as follows:

STEP 0 Draw the a horizontal real y-line.

STEP 1 Find the equilibria and mark them on the y-line.

STEP 2 Find the intervals on which $f(y) > 0$ and label each one by an arrow pointing to the *right*.

STEP 3 Find the intervals on which $f(y) < 0$ and label each one by an arrow pointing to the *left*.

Here are some examples.

1.19. EXAMPLE. Consider yet another logistic equation

$$y' = 2y(1 - y),$$

with $k = 2$ and $N = 1$. Let $f(y) = 2y(1 - y)$. The equilibria points are 0 and 1. Since $f > 0$ on $(0, 1)$, solutions in $(0, 1)$ are increasing. Since $f < 0$ on $(1, \infty)$ and on $(-\infty, 0)$ solutions contained

[12] Among his many accomplishments, Poincaré discovered special theory of relativity independently of Einstein and more or less at the same time as him.

in these intervals are decreasing. The phase line of the logistic equation therefore looks like this:

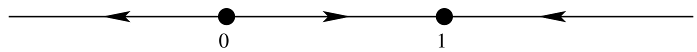


FIGURE 1.10. The phase line of the logistic equation.

It is instructive to plot the phase line and the graphs of typical solutions from Fig. 1.6 side by side:

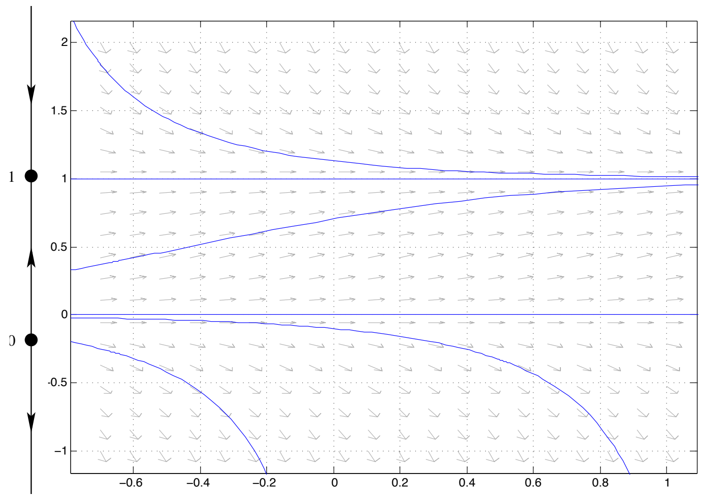


FIGURE 1.11. The phase line of the logistic equation and the graphs of typical solutions.

Observe that in the above example solutions which have a positive initial value are all pulled toward the equilibrium solution $y = 1$. We call such an equilibrium a sink. More generally:

An equilibrium point y_* of a differential equation $y' = f(y)$ is called a sink if there exists an open interval (a, b) containing y_* such that every solution $y(t)$ with $y(0) \in (a, b)$ converges to y_*, as $t \to +\infty$.

In other words, y_* is a sink if every solution which starts sufficiently close to y_* converges to it in *forward* time. Sinks *attract* nearby solutions.

An equilibrium point y_* is called a source if it is a sink for the *reverse* equation $y' = -f(y)$. Equivalently:

An equilibrium point y_* of a differential equation $y' = f(y)$ is called a source if there exists an open interval (a, b) containing y_* such that every solution $y(t)$ with $y(0) \in (a, b)$ converges to y_*, as $t \to -\infty$.

In other words, y_* is a source if every solution which starts sufficiently close to y_* converges to it in *backward* time. Sources *repel* nearby solutions. In the previous example $y_* = 0$ is a source, with $(a, b) = (-\infty, 1)$.

If an equilibrium attracts nearby solutions on one side and repels nearby solutions on the other side we will call it a node.

1.20. Example. Consider the differential equation

$$y' = y^3 - y^2.$$

Write $f(y) = y^3 - y^2 = y^2(y-1)$. The equilibria are $y_0 = 0$ and $y_1 = 1$. Since $f(y) < 0$ for $0 < y < 1$ and $f(y) > 0$ for $y > 1$, y_1 is a source. Since we also have $f(y) < 0$ for $y < 0$, y_0 is a node. Observe also that $f'(0) = 0$. Here is the phase line for this equation.

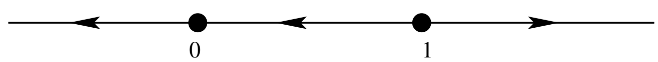


Figure 1.12. The phase line of $y' = y^3 - y^2$ with a node at zero.

1.21. Example. To appreciate the power of qualitative analysis, consider the differential equation $y' = f(y)$ and assume that all we know about a function f is a sketch of its graph given in Figure 1.13.

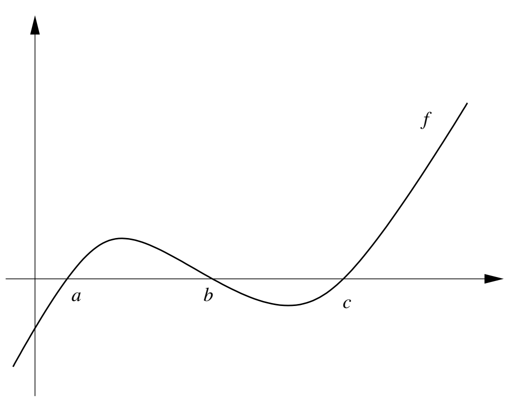
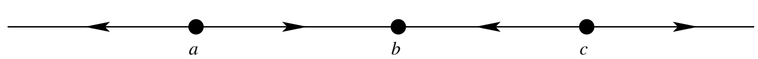


Figure 1.13. The graph of f.

We see that $f(y) = 0$ if $y = a, b, c$. Furthermore, if $y < a$ or $b < y < c$, then $f(y) < 0$ and if $a < y < b$ or $y > c$, then $f(y) > 0$. The phase line of $y' = f(y)$ therefore looks like this:



Note that even though we do not have any details about the differential equation, we can for instance derive an important fact that any solution which starts between a and b converges to b, as $t \to \infty$.

Remark. We should emphasize that we are always speaking of *isolated* equilibria. An equilibrium is called isolated if there is an open interval around it which contains no other equilibria. The equation

$$y' = \begin{cases} y^2 \sin \dfrac{1}{y} & \text{if } y \neq 0 \\ 0 & \text{if } y = 0 \end{cases}$$

has an equilibrium at zero which is not isolated, since there is a sequence of equilibria converging to zero. The graph of the function on the right-hand side of the equation is pictured below.

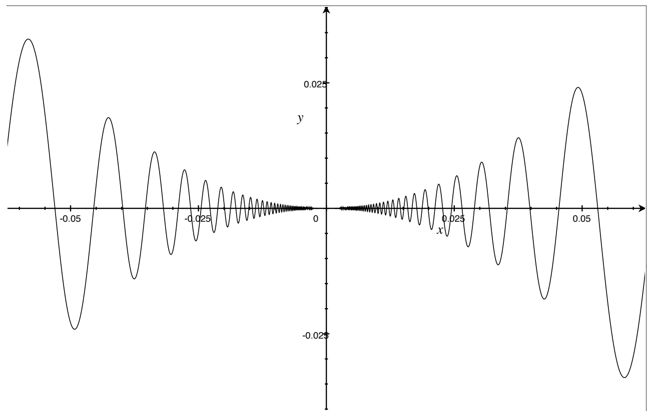


FIGURE 1.14. The graph of $F(x) = x^2 \sin \frac{1}{x^2}$ $(x \neq 0)$, $F(0) = 0$.

In this course we will not be concerned with non-isolated equilibria, which are in a certain sense rare.

1.8. Linearization

It is pretty clear from what we've learned so far that linear differential equations are the easiest to deal with: they are easy to solve and do qualitative analysis with. They are certainly easier than most non-linear equations. This is true in mathematics in general: linear objects (functions, shapes, spaces, etc.) are always easier to understand than non-linear ones. But most things in the "real world" are not linear so we have no choice but to find ways of studying nonlinear equations. Luckily, we have can use Calculus. Remember one of its central ideas:

To understand something non-linear, approximate it by something linear.

For instance, to undertand the shape of the graph of the non-linear function

$$f(x) = x^5 - 2x + 1$$

near the point $(1, 0)$ we calculate $f'(1) = 3$ and find the equation of its tangent line

$$y = 3(x - 1).$$

The tangent is the best linear approximation to f near $x = 1$. When x is close to 1, $f(x)$ is fairly close to $3(x - 1)$. See Figure 1.15.

Following the same idea, we will try to understand non-linear equations by *linearizing* them. This works particularly well near equilibrium points. Here's how to do it.

Suppose we have a non-linear (autonomous) differential equation

$$y' = f(y)$$

and that y_0 is an equilibrium point (i.e., $f(y_0) = 0$). Let $a = f'(y_0)$. Then:

The equation $z' = az$ is called the linearization of $y' = f(y)$ at the equilibrium point y_0.

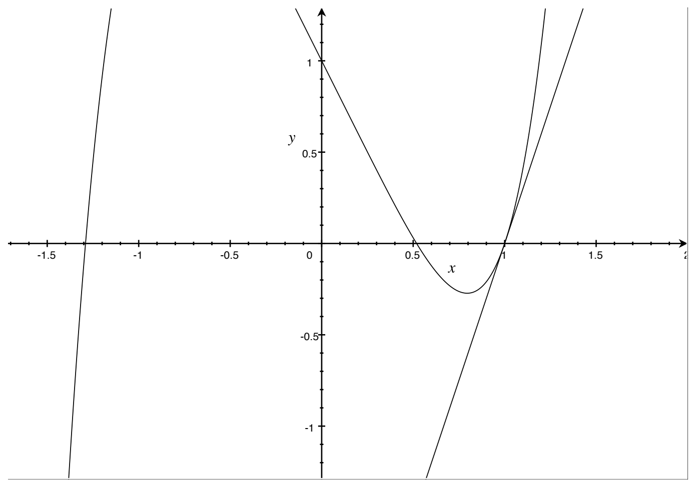


FIGURE 1.15. The tangent approximates the non-linear function well near the point $(1, 0)$.

Of course, we know how to solve $z' = az$: the general solution is

$$z(t) = z(0)e^{at}.$$

Note that when $a < 0$, the origin is a sink for the linearized equation and if $a > 0$, it is a source.

The question is, what does the linearized equation tell us about the original, non-linear equation? The answer is given by the following result.

1.22. THEOREM (Linearization Theorem). *Let y_0 be an equilibrium point of the equation $y' = f(y)$ and set $a = f'(y_0)$.*

(a) *If $a < 0$, then y_0 is a sink. equilibrium.*
(b) *If $a > 0$, then y_0 is a source.*
(c) *If $a = 0$, then the linearized equation does not give us any information about the non-linear equation.*

To get an idea why the Linearization Theorem is true, let's assume $a = f'(y_0) < 0$. That means that in some small interval around the point y_0 the function f is *decreasing*. In other words, there exists a number $\varepsilon > 0$ such that f is decreasing on $(y_0 - \varepsilon, y_0 + \varepsilon)$. Recall that $f(y_0) = 0$, since y_0 is an equilibrium. Putting the two things together, we get that $f(y) > 0$ for $y_0 - \varepsilon < y < y_0$ and $f(y) < 0$ for $y_0 < y < y_0 + \varepsilon$. That means that any solution starting in the interval $(y_0 - \varepsilon, y_0)$ is increasing toward the equilibrium whereas solutions starting in the interval $(y_0, y_0 + \varepsilon)$ are decreasing toward the equilibrium. Thus y_0 is sink.

The case $a > 0$ is dealt with in a similar way.

1.23. EXAMPLE. Consider the differential equation

$$y' = y^5 - 2y + 1.$$

Let $f(y) = y^5 - 2y + 1$. Since $f(1) = 0$, the equation has an equilibrium at $y_0 = 1$. We have $f'(1) = 3 > 0$, so the Linearization Theorem tells us that $y_0 = 1$ is a source. By inspecting Figure 1.15 we see that there are also equilibrium points $y_- < 0$ (between -1.5 and -1) and $y_+ > 0$ (between 0 and 1). We also see that $f'(y_-) > 0$ and $f'(y_+) < 0$, so y_- is a source and y_+ is a sink. □

1.24. Example. Consider the differential equation $y' = f(y)$, where

$$f(y) = (y^2 - y)\left\{\cos(y^7 - 3y) - \frac{\pi}{e}y^8\right\}.$$

It is easy to check that $f(0) = 0$, so zero is an equilibrium. Is it a sink, source or a node? We leave it as an exercise to check that $f'(0) = -1$. Therefore by the Linearization Theorem zero is a sink.

1.25. Example. Consider the equation $y' = -y^3$ and let $f(y) = -y^3$. Since $f(0) = 0$, zero is the (only) equilibrium. However, $f'(0) = 0$, so the Linearization Theorem does not apply. On the other hand, we see that $f(y) > 0$ for $y < 0$ and $f(y) < 0$ for $y > 0$, so negative solutions increase and positive solutions decrease, which means that 0 is a sink. It is not hard to see that 0 is a source for the equation $y' = y^3$. Note that for both equations, $f'(0) = 0$.

1.26. Example. Zero is the only equilibrium of the equation $y' = y^2 = f(y)$. Since $f'(0) = 0$, the Linearization Theorem is not applicable. However, since $y^2 > 0$, for all $y \neq 0$, we see that all non-equilibrium solutions are increasing, so zero is a node.

The last two examples show that if y_0 is an equilibrium of the equation $y' = f(y)$ and $f'(y_0) = 0$, then y_0 may be of any type whatsoever, confirming the assertion in part (c) of the Linearization Theorem.

Remark. It can also be shown that if y_0 is an equilibrium point of $y' = f(y)$ and $a = f'(y_0) \neq 0$, then near y_0, solutions of the linear equation $y' = a(y - y_0)$ approximate solutions of $y' = f(y)$.

1.9. Numerical methods

Qualitative analysis is useful but what if we need some *quantitative* information about solutions of a differential equation which we cannot solve? In that case we have no choice but to work with approximate solutions. In other words, we try to construct a sequence of not-quite-solutions $y_k(t)$ such that as $k \to \infty$, $y_k(t)$ converges to a bona fide solution $y(t)$. Ideally, we also want to have some information about the error we would be making if we were to replace the actual solution $y(t)$ by an approximate one $y_k(t)$.

There are many methods – called *numerical methods* – for constructing approximate solutions. Here we will only discuss one of them called Euler's Method.[13] It is ultimately based on the following idea. Assume $y(t)$ is is a solution to the equation $y' = f(t, y)$. Since

$$y'(t) = \lim_{h \to 0} \frac{y(t+h) - y(t)}{h},$$

we can approximate the derivative $y'(t)$ by the difference quotient:

$$y'(t) \approx \frac{y(t+h) - y(t)}{h},$$

where h is small. Multiplying both sides by h and rearranging the terms, we obtain

$$y(t+h) \approx y(t) + hy'(t) = y(t) + hf(t, y(t)).$$

So the value of the solution at some future time $t+h$ can be approximated using its "present" value $y(t)$ and its "present" slope $f(t, y(t))$.

Before we describe Euler's method, we need the notion of a slope field. Suppose we are interested in the solution $y(t)$ which at time $t = t_0$ takes some value y_0. At $t = t_0$ the graph of $y(t)$ has slope $y'(t_0) = f(t_0, y_0)$. Even if we don't know $y(t)$, we can always draw the tangent line at the point (t_0, y_0). If we do this at every point, we obtain the slope field. More precisely:

[13] Named after Leonhard Euler (1707-1783), one of the greatest mathematicians of all time, who published the method in 1768. Euler was so enormously productive that the publication of his collected works is still in progress.

The **slope field** of a differential equation $y' = f(t, y)$ is the assignment to every point (t_0, y_0) in the plane of a slope equal to $s_0 = f(t_0, y_0)$. This is graphically represented by drawing a short line segment through (t_0, y_0) with slope s_0.

In practice, we usually draw the slope field only on a finite rectangular array of points in the plane. The following slope fields were plotted using the website `http://slopefield.nathangrigg.net` created by Nathan Grigg.

1.27. Example. Figure 1.16 shows the slope field of the logistic equation $y' = 2y(1 - y)$. The solution satisfying $y(0) = 1/2$ is plotted in blue.

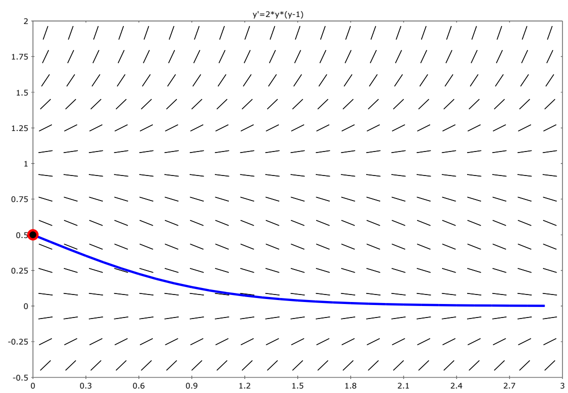


Figure 1.16. The slope field of $y' = 2y(1-y)$ with the solution satisfying $y(0) = 1/2$ in blue.

Note that if the equation is autonomous (i.e., of the form $y' = f(y)$), then the slope field stays the same if we shift it horizontally by any amount. Similarly, if the equation is of the form $y' = f(t)$, then its slope field is invariant with respect to vertical shifts.

1.28. Example. The slope field of the non-linear non-separable equation $y' = y^2 - t$ is shown in Figure 1.17 with the solution satisfying $y(-1) = 0$ shown in blue.

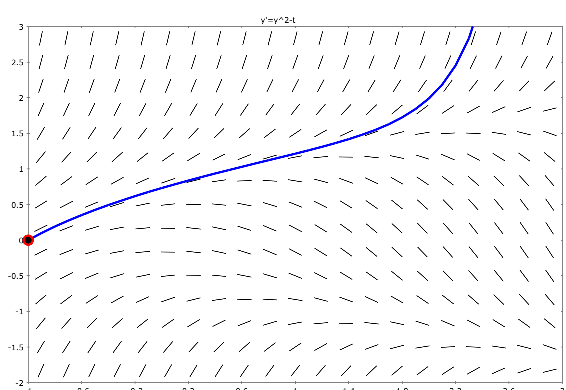


Figure 1.17. The slope field of $y' = y^2 - t$ with the solution satisfying $y(-1) = 0$ in blue.

The basic idea behind slope fields is: they are easy to draw and by visually inspecting a slope field we may gain some insight in the behavior of solutions of the corresponding differential equation.

Now we can formulate the main idea of Euler's Method, which is:

Glue together little pieces of the slope field to obtain a curve which approximates the graph of an actual solution.

Joining together little pieces of the slope field produces a polygonal line. That means that our goal is to construct a *polygonal approximation* to the actual solution of the IVP

$$y' = f(t, y), \qquad y(t_0) = y_0.$$

A polygonal line consists of finitely many straight line segments. That means we are looking for a piecewise linear approximate solution $u(t)$, in the sense that for some sequence of times $t_0 < t_1 < t_2 < \cdots < t_{n-1} < t_n$, $u(t)$ is a linear function[14] between any two consecutive t_k's. That is, for $t_k \le t \le t_{k+1}$, $u(t)$ should be given by

$$u(t) = a_k t + b_k$$

for some constants a_k and b_k. The question is, how should we choose t_k's, a_k's and b_k's? We will see that we can make some pretty natural choices. Here's how we do that.

First, choose a *step size* Δt and define

$$t_k = t_0 + k\Delta t,$$

for $k = 1, 2, 3, \ldots$. Thus $t_{k+1} - t_k = \Delta t$, for all k. (It is also possible to choose the t_k's which are not uniformly spaced.) Next, let us choose a_0 and b_0. We want our approximate solution $u(t)$ to have the following properties:

(a) It should satisfy the initial condition, i.e., $u(t_0) = y_0$.
(b) It should have the same slope at $t = t_0$ as the actual solution, i.e., $u'(t_0) = f(t_0, y_0)$.
(c) It should be linear for $t_0 \le t \le t_1$.

These three properties force us to define

$$u(t) = y_0 + f(t_0, y_0)(t - t_0),$$

for all t between t_0 and t_1. This is the first linear segment of the approximate solution. It terminates at $y_1 = u(t_1) = y_0 + f(t_0, y_0)\Delta t$.

How do we continue? Well, $u(t)$ should be linear between t_1 and t_2 and it makes sense to take its slope at $t = t_1$ to be equal to the slope of the actual solution passing through the point (t_1, y_1). Since that slope is $f(t_1, y_1)$, we are again forces to define

$$u(t) = y_1 + f(t_1, y_1)(t - t_1),$$

for all t between t_1 and t_2. This is the second linear segment of the approximate solution. It terminates at $y_2 = u(t_2) = y_1 + f(t_1, y_1)\Delta t$.

Continuing this process in the same way we obtain the approximate solution $u(t)$. Between t_k and t_{k+1}, for any k, $u(t)$ is given by

$$u(t) = y_k + f(t_k, y_k)(t - t_k).$$

The sequence $y_1, y_2, \ldots$ is defined recursively by

[14] Or more precisely, affine. A function is called linear if it is of the form $L(t) = ct$ and it is technically called affine if it is of the form $A(t) = at + b$.

$$y_{k+1} = y_k + f(t_k, y_k)\Delta t. \tag{1.21}$$

See Figure 1.18.

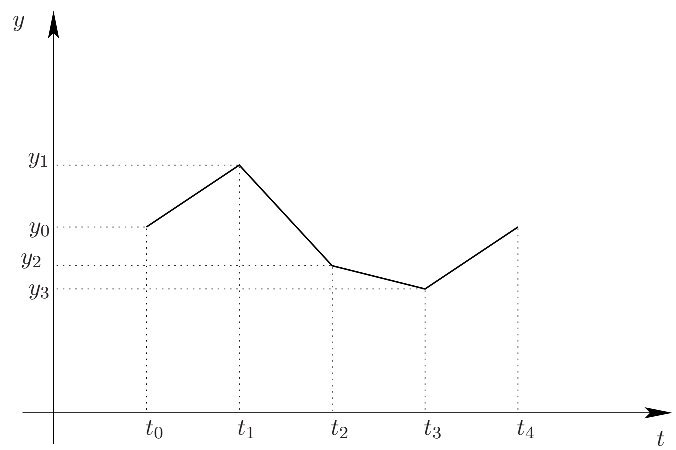


FIGURE 1.18. A piecewise linear solution constructed using Euler's method.

Euler's method can be summarized as follows:

STEP 1 Choose a step size Δt.
STEP 2 For $k = 0, 1, 2, \ldots$, plot the line segment joining (t_k, y_k) and (t_{k+1}, y_{k+1}), where y_{k+1} is defined by (1.21).

1.29. EXAMPLE (Taken from [1], p. 59). Using Euler's method approximate $y(2)$, where $y(t)$ is the solution to the initial-value problem

$$y' = -2ty^2, \qquad y(0) = 1.$$

Of course, the equation is separable and we can easily solve it explicitly to obtain

$$y(t) = \frac{1}{1+t^2}.$$

So the exact value of $y(2) = 1/5 = 0.2$.

Choose $\Delta t = 1/2$ to be the initial step size. So $t_0 = 0$, $t_1 = 1/2$, $t_2 = 1$, $t_3 = 3/2$, and $t_4 = 2$. The recursive formula for Euler's method in this case is

$$y_{k+1} = y_k + f(t_k, y_k)\Delta t = y_k - 2t_k y_k^2 \Delta t.$$

Applying Euler's method we obtain the following table:

k	t_k	y_k	$f(t_k, y_k)$
0	0	1	0
1	1/2	1	−1
2	1	1/2	−1/2
3	3/2	1/4	−3/16
4	2	5/32	

The resulting approximation of $y(2)$ we obtain with this step size is therefore $y_4 = 5/32 = 0.15625$. The absolute error is $|0.2 - 0.15625| = 0.04375$, while the relative error is $0.04375/0.2 = 0.21875$, or 21.875%, which is huge!

Let's try to do a little better by choosing the step size to be $\Delta t = 0.1$. To reach 2 starting from zero with this step size we need $k = 20$ steps. Euler's method done to four decimal places yields the following table:

k	t_k	y_k
0	0	1
1	0.1	1.000
2	0.2	0.9800
3	0.3	0.9416
⋮	⋮	⋮
19	1.9	0.2101
20	2.0	0.1933

This time we obtain $y_{20} = 0.1933$ as the approximation of $y(2)$. The absolute error is 0.0067 while the relative error is $0.0067/0.2 = 0.0335$ or 3.35%, which is a lot better! See Figure 1.19 for a comparison between the two approximate solutions.[15]

Making the step size even smaller, say $\Delta t = 0.001$, would further decrease the absolute and relative error: the former would drop to 0.000063 and the latter to 0.0315%.

Remark.

(a) Euler's method is a *first-order method*, which means that the error at a given time (also known as the global error) is proportional to the step size. It can be shown that the global

[15] Plotted using `http://www.mathscoop.com/calculus/differential-equations/euler-method-calculator.php`.

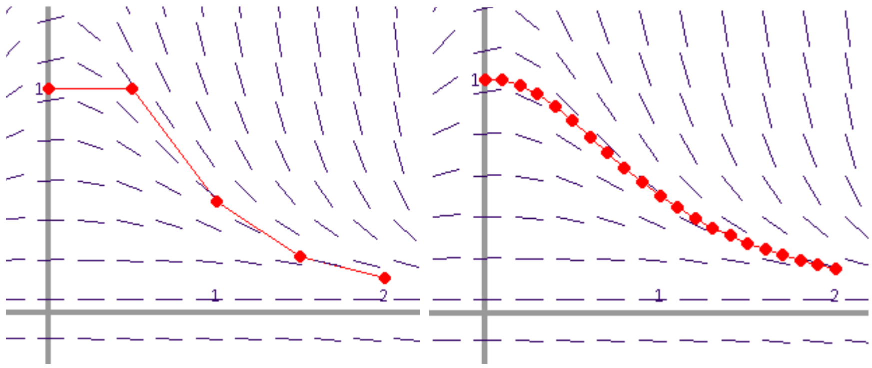


FIGURE 1.19. Comparison between step sizes $\Delta t = 1/2$ and $\Delta t = 1/10$.

error at time t is at most

$$\frac{(\Delta t)M}{2L}\left\{e^{L(t-t_0)} - 1\right\},$$

where M is an upper bound on the second derivative of the solution $y(\cdot)$ on the interval $[t_0, t]$ and L is an upper bound of the first partial derivatives of f. So if we make the step size ten times shorter, the error becomes approximately ten times smaller.

(b) There are many other, more efficient and accurate numerical methods for approximately solving ordinary differential equations, such as so called multiderivative methods (which use not only the function f on the right-hand side of the differential equation but also its derivatives), Runge-Kutta methods, ets.

(c) Just like computers can fail, so can numerical methods, especially if they are not used prudently. Take for example the IVP

$$y' = e^t \cos y, \qquad y(0) = 1/2.$$

As t increases, the right-hand side of the differential equation quickly gets very large, making the slope field almost vertical and confusing the computer. Here's the "solution" we get

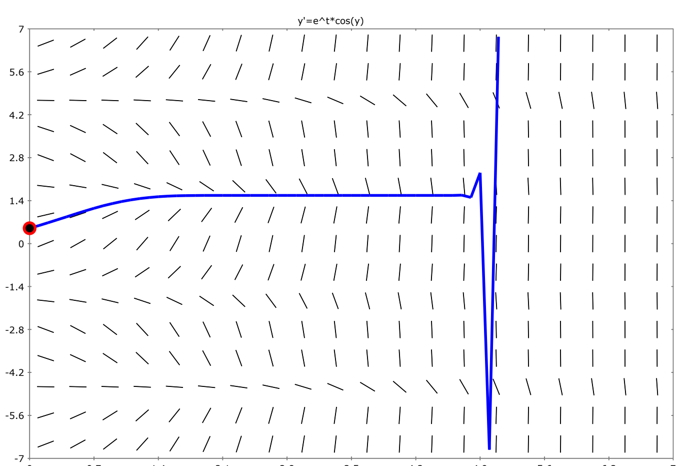


FIGURE 1.20. Numerical solution is wrong for large t.

Qualitative analysis tells us that the blue curve cannot be a solution: it crosses the equilibrium solution $y(t) = \pi/2$ (as well as the equilibrium $y(t) = -\pi/2$), which by the Existence and Uniqueness Theorem is not permitted.

CHAPTER 2

First order systems

CHAPTER 3

Linear systems

CHAPTER 4

Nonlinear systems

CHAPTER 5

Laplace transform

Appendix

Bibliography

1. Paul Blanchard, Robert L. Devaney, and Glen R. Hall, *Differential equations*, third ed., Thomson, 2006.