Existence and Uniqueness of Solutions to First Order Ordinary Differential Equations

Jesse Ratzkin

February 1, 2007

1 Introduction

These notes will prove that there is a unique solution to the initial value problem for a wide range of first order ordinary differential equations (ODEs). The initial value problem we consider is

$$\frac{du}{dx} = F(x, u(x)), \qquad u(a) = b, \tag{1}$$

where F is a given function, and a and b are given real numbers. A solution to this problem is a function $u(x)$ satisfying the differential equation $u' = F(x, u)$ with the proper initial condition $u(a) = b$. The result we will prove is

Theorem 1 *If F and $\frac{\partial F}{\partial u}$ are continuous at (a, b) then there is an $\epsilon > 0$ such that there is a unique solution to (1) on the interval $a - \epsilon < x < a + \epsilon$.*

The rest of these notes are structured as follows. Section 2 contains some preliminary material regarding sequences of functions. Section 3 transforms the initial value problem (1) into an integral equation and introduces the sequence of functions which will converge to the unique solution. Section 4 proves teh contraction mapping principle, which is interesting in its own right. We prove existence and uniqueness in Section 5 by applying the contraction mapping principle to the sequence we defined in Section 3. Finally, we discuss some examples in Section 6.

I have based these notes mostly on course notes I took as a student [P, B]. One can find a user–friendly, if brief, sketch of Picard iteration in section 2.11 of [BP], and a more complete proof in [CL] (which is the standard reference for the theory of ODEs).

2 Preliminaries

This section contains some basic tools we will need: the notion of distance between two functions, notions of convergence of a sequence of functions (both pointwise and uniform), and some consequences of such convergence.

First we discuss the distance between two functions. Let $u(x)$ and $v(x)$ be function on the interval $0 \leq x \leq 1$. What should it mean that u and v are close? It should mean that $u(x)$ is close to $v(x)$ for all $x \in [0, 1]$. This motivates the following definition.

Definition 1 *Let $u(x)$ and $v(x)$ be continuous functions on $[0, 1]$. Then the C^0 (or L^∞) distance between u and v is*

$$\|u - v\|_\infty := \max_{0 \leq x \leq 1} (|u(x) - v(x)|).$$

There are several comments. First observe that $|u - v|$ is a continuous function, so it achieves its maximum on the closed interval $[0, 1]$. Also, there is nothing special about $[0, 1]$; we can make a similar definition for any closed interval $[a, b]$ so long as $a < b$. However, it is important that we consider a *closed* interval, as the example $u(x) = 1/x$, $v(x) = 1$ on the open interval $(0, 1)$ illustrates. Third, this distance function satisfies all the usual properties of distance functions you're used to.

Lemma 2
- *For any two continuous functions $u(x)$ and $v(x)$, $\|u - v\|_\infty \geq 0$, with equality if and only if $u \equiv v$.*
- *Symmetry holds: $\|u - v\|_\infty = \|v - u\|_\infty$.*
- *For any three continuous functions u, v, w, we have the triangle inequality:*

$$\|u - v\|_\infty \leq \|u - w\|_\infty + \|w - v\|_\infty.$$

Proof: The first property holds because the absolute value of a nonzero number is positive, and only the absolute value of 0 is 0. The second property holds because for any two numbers a and b, $|a - b| = |b - a|$. To see the triangle equality, let $|u - v|$ be maximized at the point $x_0 \in [0, 1]$. Then

$$\begin{aligned} \|u - v\|_\infty &= |u(x_0) - v(x_0)| \\ &\leq |u(x_0) - w(x_0)| + |w(x_0) - v(x_0)| \\ &\leq \max_{0 \leq x \leq 1} (|u(x) - w(x)|) + \max_{0 \leq x \leq 1} (|w(x) - v(x)|) \\ &= \|u - w\|_\infty + \|w - v\|_\infty \end{aligned}$$

□

From calculus, you're familiar with the idea of a convergent sequence. A sequence of numbers $\{a_n\}$ converges to a, written $\lim_{n\to\infty} a_n = a$, if for any $\epsilon > 0$ there exists N such that $n > N$ implies $|a_n - a| < \epsilon$. We have two notions of a convergent sequence of functions.

Definition 2 *Let $\{u_n(x)\}$ be a sequence of continuous functions on the interval $[0, 1]$. We say u_n converges to $u(x)$ pointwise if $lim_{n\to\infty} u_n(x) = u(x)$ for all $x \in [0, 1]$. We say u_n converges to u uniformly if $\lim_{n\to\infty} \|u_n - u\|_\infty = 0$.*

Observe that the pointwise limit of a sequence of differentiable functions does not need to be even continuous. Indeed, let $u_n(x) = x^n$ for $0 \leq x \leq 1$. Then pointwise $u_n \to u$ where

$$u(x) = \begin{cases} 0 & x < 1 \\ 1 & x = 1 \end{cases}$$

We summarized some important properties in the following lemma.

Lemma 3 *Let the sequence $\{u_n\}$ of continuous functions on $[0, 1]$ converge uniformly to u. Then $\{u_n\}$ also converges to u pointwise, and u is also continuous.*

Before proving this lemma we make two remarks. First, observe that pointwise convergence does **not** imply uniform convergence. (Consider the example listed before the lemma.) Second, one can strengthen this lemma (using essentially the same proof) to show that the uniform limit of differentiable functions is differentiable.

Proof: We first show that uniform convergence implies pointwise convergence. Let $x_0 \in [0, 1]$. Then

$$|u_n(x_0) - u(x_0)| \leq \max_{0 \leq x \leq 1} |u_n(x) - u(x)| = \|u_n - u\|_\infty \to 0,$$

which implies $u_n(x_0) \to u(x_0)$. Next we show u is continuous. Choose $x_0 \in [0, 1]$ and let $\epsilon > 0$. Choose N so that if $n > N$ then $\|u_n - u\|_\infty < \epsilon/3$. Choose an $n > N$, and let $\delta > 0$ be such that if $|x - x_0| < \delta$ then $|u_n(x) - u_n(x_0)| < \epsilon/3$. Now apply the triangle inequality twice, using that $|u_n(x) - u(x)| < \epsilon/3$ for all x:

$$\begin{aligned} |u(x) - u(x_0)| &\leq |u(x) - u_n(x)| + |u_n(x) - u_n(x_0)| + |u_n(x_0) - u(x_0)| \\ &\leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon. \end{aligned}$$

□

3 Transformation to an integral equation and Picard iteration

There are two basic steps to the proof of Theorem 1. The first step is to transform the initial value problem (1) into an integral equation. The second step is to construct a sequence of functions which converges uniformly to a solution of this integral equation. We define this sequence iteratively, using a scheme called Picard iteration (named after Emile Picard, 1856–1941).

First we consider the integral equation

$$u(x) = b + \int_a^x F(t, u(t))dt. \tag{2}$$

Lemma 4 *The function u satisfies the initial value problem (1) if and only if it satisfies the integral equation (2).*

Proof: First suppose u satisfies (2). Then

$$u(a) = b + \int_a^a F(t, u(t))dt = b$$

and, by the fundamental theorem of calculus,

$$\frac{du}{dx} = \frac{d}{dx}\left[b + \int_a^x F(t, u(t))dt\right] = \frac{d}{dx}\int_a^x F(t, u(t))dt = F(x, u(x)).$$

Now let u satisfy (1). Then, again by the fundamental theorem of calculus,

$$u(x) = c + \int_a^x F(t, u(t))dt$$

for some constant c. Evaluating at $x = a$, we see this constant c must be b.

□

Next we define the sequence of functions which will converge to our solution of (2), at least for x sufficiently close to a. We start with $u_0 \equiv b$, and define the iteration scheme

$$u_{n+1}(x) := b + \int_a^x F(t, u_n(t))dt.$$

It is useful to think of this sequence as a certain operation applied repeatedly to the original function $u_0 \equiv b$. This operation is given by

$$\Phi(u) := b + \int_a^x F(t, u(t))dt.$$

Thus we have defined Φ, which is a function *on the space of functions*. Such an object, a function on the space of functions, is usually called a functional. This functional Φ has some very nice properties, which we will explore in the next two sections.

4 The contraction mapping principle

Let (X, d) be a space equipped with a distance function d. That is, the function $d = d(x, y)$ on pairs of points $x, y \in X$ satisfies the properties listed in Lemma 2:

- $d(x, y) \geq 0$ for all x, y, with equality if and only if $x = y$.
- $d(x, y) = d(y, x)$
- $d(x, y) \leq d(x, z) + d(z, y)$.

Definition 3 *A sequence of points $\{x_n\}$ in such a space is called a Cauchy sequence if, given $\epsilon > 0$ there is a number N such that*

$$n, m > N \Rightarrow d(x_n, x_m) < \epsilon.$$

Also, the space (X, d) with a distance function is called complete if all Cauchy sequences converge.

The real line is complete, with the absolute value function as its distance function, is complete. However, the open interval $(0, 1)$ is not complete, as the example $\{x_n = 1/n\}$ illustrates. The space of continuous functions on a closed interval $[a, b]$, equipped with the C^0 distance function defined above, is complete. The proof of completeness is rather technical, see chapter 7 of [R].

Roughly speaking a contraction is a map which decreases distances.

Definition 4 *Let (X, d) be a space equipped with a distnance function d. A function $\Phi : X \to X$ from X to itself is a contraction if there is a number $k \in (0, 1)$ such that for any pair of points $x, y \in X$ we have*

$$d(\Phi(x), \Phi(y)) \leq k d(x, y).$$

It is important that the constant k is strictly less than one.

The following proposition is known as the contraction mapping principle. It is widely used in the fields of analysis, differential equations, and geometry.

Proposition 5 *Let Φ be a contraction on a complete space (X, d) equipped with a distance function. Then Φ has a unique fixed point, that is a unique solution to the equation $\Phi(x) = x$.*

Proof: Pick any $x_0 \in X$ and consider the sequence of iterates defined by $x_{n+1} = \Phi(x_n)$. We will show that $\{x_n\}$ is a Cauchy sequence. Let's first see that the proposition follows once we establish that $\{x_n\}$ is Cauchy. Indeed, by completeness this sequence must converge to something, which we call

$$\bar{x} = \lim_{n\to\infty} x_n = \lim_{n\to\infty} \Phi^n(x_0).$$

Here Φ^n means we have applied the function Φ n times. Now apply Φ to both sides:

$$\Phi(\bar{x}) = \lim_{n\to\infty} \Phi^{n+1}(x_0) = \lim_{n\to\infty} \Phi^n(x_0) = \bar{x}.$$

Here we have used the fact that if limits exist then they are unique. Moreover, suppose there is another fixed point, that is another $\hat{x}$ such that $\Phi(\hat{x}) = \hat{x}$. Then, using the fact that Φ is a contraction and both $\bar{x}$ and $\hat{x}$ are fixed points,

$$d(\hat{x}, \bar{x}) = d(\Phi(\hat{x}), \Phi(\bar{x})) \leq k d(\hat{x}, \bar{x}).$$

Since $k < 1$, this is only possible if $d(\hat{x}, \bar{x}) = 0$, *i.e.* $\hat{x} = \bar{x}$.

We complete the proof by showing $\{x_n\}$ is a Cauchy sequence. First observe that, for any n,

$$d(x_n, x_{n+1}) = d(\Phi(x_{n-1}), \Phi(x_n)) = k d(x_{n-1}, x_n).$$

Applying this inequality n times, we see

$$d(x_n, x_{n+1}) \leq k^n d(x_0, x_1).$$

Now pick $\epsilon > 0$ and let $M := d(x_0, x_1)$. For $n > m$ we have

$$\begin{aligned}
d(x_n, x_m) &\leq d(x_m, x_{m+1}) + d(x_{m+1}, x_{m+2}) + \cdots + d(x_{n-1}, x_n) \\
&\leq k^m d(x_0, x_1) + k^{m+1} d(x_0, x_1) + \cdots k^{n-1} d(x_0, x_1) \\
&= M \sum_{j=m}^{n-1} k^j \leq M k^m \sum_{j=0}^{\infty} k^j \\
&= \frac{M k^m}{1-k}.
\end{aligned}$$

Becaude M is a fixed number and $k < 1$, we can pick an N such that $m > N$ forces

$$\frac{M k^m}{1-k} < \epsilon,$$

which in turn implies $d(x_n, x_m) < \epsilon$, which is the criterion for the sequence $\{x_n\}$ to be a Cauchy sequence.

□

5 Convergence of the Picard sequence

In this section we will show the Picard sequence converges using the contraction mapping principle. Recall that we defined our function $\{u_n\}$ by the iteration scheme

$$u_0 \equiv b, \qquad u_{n+1} = \Phi(u_n) := b + \int_a^x F(t, u_n(t)) dt.$$

The heart of the proof is a series of estimates, showing that Φ decreases distances, when distances are computed using the C^0 distance function on a closed interval, provided the interval is small enough.

Let $u_1(x)$ and $u_2(x)$ be continuous functions, and consider $\|\Phi(u_1) - \Phi(u_2)\|_\infty$, as compared to $\|u_1 - u_2\|_\infty$. We begin with some pointwise estimates:

$$\begin{aligned}
|\Phi(u_1)(x) - \Phi(u_2)(x)| &= |\int_a^x F(t, u_1(t)) dt - \int_a^x F(t, u_2(t)) dt| \\
&= |\int_a^x (F(t, u_1(t)) - F(t, u_2(t))) dt| \\
&\leq \int_a^x |F(t, u_1(t)) - F(t, u_2(t))| dt \\
&\leq \int_a^x \max |\frac{\partial F}{\partial u}| |u_1(t) - u_2(t)| dt \\
&\leq |x - a| \max |\frac{\partial F}{\partial u}| \|u_1 - u_2\|_\infty.
\end{aligned}$$

The consequence of this inequality is, if we restrict to $|x - a| < \epsilon$ and let M be the maximum of $|\partial F / \partial u|$, the estimate

$$\|\Phi(u_1) - \Phi(u_2)\|_\infty \leq \epsilon M \|u_1 - u_2\|_\infty.$$

Because $\partial F / \partial u$ is continuous, we can choose ϵ small enough so that $\epsilon M \leq 1/2$, so that we have

$$\|\Phi(u_1) - \Phi(u_2)\|_\infty \leq \frac{1}{2} \|u_1 - u_2\|_\infty. \tag{3}$$

We are finally ready to put everything together for the proof of Theorem 1. First observe that a solution of (2) satisfies $\Phi(u) = u$, that is a solution is a fixed point of Φ. Next, provided we restrict to $|x - a| \leq \epsilon$ where $\epsilon > 0$ is sufficiently small, the inequality (3) shows that Φ is a contraction on the space of continuous functions on $[a-\epsilon, a+\epsilon]$. Because the space of continuous functions on $[a-\epsilon, a+\epsilon]$ is complete, we can apply the contraction mapping principle to conclude that Φ has a unique fixed point. In other words, the integral equation (2) has a unique solution $u(x)$ on the inteval $[a - \epsilon, a + \epsilon]$.

□

We conclude with some remarks. First, we see that it is necessary to restrict to a small interval $|x - a| < \epsilon$ in order for the proof to work. This restriction is called short time existence. In other words, we have shown that solutions to an initial value problem exist for a short time before and after the time of the initial condition. Is this a real restriction, or is it an artifact of the proof? We will address this question in the next section. Next we remark that this is the typical behavior of solutions which do not exist for all x. More precisely, suppose $u(x)$ solves (1) for $a \leq x < x_1$, but not on any larger interval. Then we must have $|u(x)| \to \infty$ as $x \to x_1$. (This fact is sometime called the escape lemma, and its proof is beyond the scope of these notes.) Finally, we note that essentially the same proof works for first order systems, as well as the scalar equations described above.

6 Examples

We end these notes with some examples illustrating when an initial value problem has solutions, unique solutions, or long time solutions.

For our first example, consider

$$u' = 2\sqrt{u},$$

choosing the positve branch of the square root function. Notice that the right hand side $F(x, u) = 2\sqrt{u}$ doesn't have a continuous derivative at $u = 0$, and is not even well-defined for $u < 0$. As a conseqence, we must have $u \geq 0$, which (from the equation) implies $u' \geq 0$ as well. This is a separable equations, and the general solution has the form

$$u(x) = (x + c)^2.$$

However, this solution is only valid for $x \geq -c$, where $u' \geq 0$. There is also an equilibrium solution, namely $u(x) \equiv 0$. Putting everything together, we see that the initial value problem has a unique solution if $u(a) > 0$, multiple solutions if $u(a) = 0$ (the constant solution and the parabola), and no solutions if $u(a) < 0$. This example illustrates we can't relax the hypotheses of Theorem 1. The right hand side is continuous at $u = 0$, but it doesn't have a continuous derivative. At that point, there are solutions to the initial value problem but they are not unique.

For out next example, we consider the equation

$$u' = u^2.$$

This time the right hand side $F(x, u) = u^2$ is a polynomial, so the initial value problem always has a unique solution. The general solution of this equation has the form $u(x) = (c - x)^{-1}$ for some constant c, and there is a constant solution $u(x) \equiv 0$. This time we see that the solution does not exist for all x. Indeed, as $x \to -c$ the solution becomes unbounded, that is $|u(x)| \to \infty$. This is the typical behavior of a solution to an ODE which does not exist for all values of x: the solution must blow up near a finite value of x if it's not a solution for all values of x.

References

[BP] W. Boyce and R. DiPrima. *Elementary Differential Equations and Boundary Value Problems.* John Wiley & Sons, 1992.

[B] K. Bube *Lecture Notes in Linear Analysis*. unpublished, 1996/7.

[CL] E. Coddington and N. Levinson. *Theory of Ordinary Differential Equations*. McGraw–Hill, 1984.

[P] C. Pugh. *Lecture Notes in Real Analysis*. unpublished, spring 1994.

[R] W. Rudin. *Principles of Mathematical Analysis*. McGraw–Hill, 1976.