Chapter 2

An introduction to shooting methods

2.1. Introduction

This chapter starts with some simple examples and continues through a variety of types of shooting, presented in considerable detail. Those who have some familiarity with the method may wish to start with Section 2.4, or merely take a look at some of the exercises, particularly the last three.

What has come to be called the shooting method has its origins in a more sophisticated technique, due mainly to Ważewski [**250**]. This method makes use of a topological lemma which, in R^n, is related to Brouwer's fixed point theorem. An exposition of Ważewski's method, and some applications to differential equations, may be found in the classic ode text by Hartman [**83**]. We will discuss this method further in the final section of this chapter.

While Ważewski's result has seen continued application over the years, examples where its hypotheses can be verified are relatively difficult to find. Shooting may be thought of as including Ważewski's method, but also simpler topological arguments involving only connectedness. These will form the bulk of the applications of shooting in this book. We will see that a wide range of existence problems can be handled with this technique. In the last two sections of the chapter, and in Chapter 13, we discuss topological principles similar to Ważewski's which may be viewed as extending the range of the shooting method.

2.2. A first order example

We begin by returning to the second example from Chapter 1, namely the equation

$$x' = x^3 + \sin t. \tag{2.1}$$

As before, our goal is to prove that there is a solution with period 2π, and in Exercise 1.4 you were asked to show that it is sufficient to find a solution such that $x(0) = x(2\pi)$.

For each real number α, let $x(t, \alpha)$ be the unique solution of (2.1) satisfying

$$x(0) = \alpha.$$

In the shooting method we consider α as an unknown and try to show that the equation

$$x(2\pi, \alpha) - \alpha = 0$$

has at least one solution.

We apply the intermediate value theorem from basic analysis. Define a function $F : R^1 \to R^1$ by letting

$$F(\alpha) = x(2\pi, \alpha) - \alpha.$$

We wish to show that the equation $F(\alpha) = 0$ has a solution.

To use the intermediate value theorem we look for numbers α_1 and α_2 such that $F(\alpha_1) > 0$ and $F(\alpha_2) < 0$. We also have to verify that F is defined and continuous on an interval that contains α_1 and α_2. Unfortunately, in this problem the second step is not as straightforward as we would like, because of possible "blowup". As was pointed out in footnote 1.1 of Chapter 1, the solution $x(t, \alpha)$ might not exist on the entire interval $0 \leq t \leq 2\pi$. For the moment, let us ignore this complication. In Chapter 1, when discussing equation (2.1), we pointed out that if $x > 1$ then $x' > 0$. So if $\alpha > 1$ then the solution increases for as long as it exists.[1] For example, if $x(t, 2)$ exists on $[0, 2\pi]$ then $F(2)$ is defined, and $F(2) = x(2\pi, 2) - 2 > 0$. Similarly, $x' < 0$ if $x < -1$, and so if $F(-2)$ is defined then it is negative. Therefore, if F is continuous on $[-2, 2]$ then there is an $\alpha^* \in [-2, 2]$ with $F(\alpha^*) = 0$. The corresponding solution $x(t, \alpha^*)$ satisfies $x(0) = x(2\pi)$ and so is periodic with period 2π.

This simple argument is the essence of the shooting method. The reader should not proceed without understanding it. It is helpful to sketch two possible solutions of (2.1) on the same graph, one starting above 1, and hence increasing, and one starting below -1, and hence decreasing. Then

[1] We will often use "as long as" arguments. To be more precise, if u exists on an interval $[0, T]$ then u is increasing in this interval.

ask yourself what must be true about at least one of the solutions in between, assuming that they all exist on the interval $[0, 2\pi]$.

To complete the existence proof, we must deal with the question of blow-up. It is usually an easy issue to dispose of, by truncating the equation outside the region where the solutions you want can lie. Let $I = [-2, 2]$. We will assume that $x(0) \in I$. The idea is that once a solution leaves I it is monotonic, because $x' \neq 0$, and so $x(t)$ does not return to I. Such a solution is not periodic, and so we do not care how it behaves. We are only interested in solutions such that $x(t) \in [-2, 2]$ for $0 \leq t \leq 2\pi$. Therefore, it is not necessary to assume that x satisfies the same differential equation when $|x| > 2$. Replace (2.1) with a different ode, by setting

$$f(t, y) = \begin{cases} y^3 + \sin t & \text{if } -2 \leq y \leq 2, \\ 8 + \sin t & \text{if } y \geq 2, \\ -8 + \sin t & \text{if } y \leq -2, \end{cases}$$

and considering the initial value problem

$$\begin{aligned} y' &= f(t, y), \\ y(0) &= \alpha. \end{aligned} \tag{2.2}$$

Since $|f(t, y)| \leq 9$ for all (t, y), the solution $y(t, \alpha)$ exists on $[0, 2\pi]$ for every α. Hence the function

$$G(\alpha) = y(2\pi, \alpha) - \alpha$$

is defined for all α. Since the function $f(t, y)$ is Lipschitz continuous in y, the solution $y(t, \alpha)$ is a continuous function of α and thus G is continuous. As before, $y' > 0$ if $y > 1$ and $y' < 0$ if $y < -1$, and so $G(2) > 0$, $G(-2) < 0$. By the intermediate value theorem there is an $\alpha^* \in [-2, 2]$ such that $G(\alpha^*) = 0$.

The corresponding solution $y(t, \alpha^*)$ remains in $[-2, 2]$ on $[0, 2\pi]$, for otherwise $G(\alpha^*) \neq 0$. Hence, from the definition of f, $y(t, \alpha^*) = x(t, \alpha^*)$ and $F(\alpha^*) = 0$. We have therefore found an initial condition α^* such that $x(2\pi) = x(0)$, and the solution x is 2π-periodic, as desired.

In this argument, the interval $[-2, 2]$ was chosen for simplicity. It is clear that a periodic solution must lie in the smaller interval $[-1, 1]$, because every solution which enters the region $|x| > 1$ is monotonic thereafter. We may wish to know if a periodic solution x can satisfy $|x(t_0)| = 1$ for some t_0. The technique for answering this arises fairly frequently in shooting methods. We must analyze higher derivatives.

For example, if $x(t_0) = 1$ then $x'(t_0) = 0$, because $x(t) \le 1$ for all t. From (2.1) we see that $\sin t_0 = -1$. Differentiate (2.1) to obtain

$$x''(t_0) = 3x(t_0)^2 x'(t_0) + \cos t_0 = \cos t_0 = 0,$$
$$x'''(t_0) = -\sin t_0 = 1.$$

Therefore $x'''(t_0) > 0$, which implies that the solution enters the region $x > 1$ and so is not periodic. A similar argument applies if $x(t_0) = -1$, and so every periodic solution lies in the open interval $-1 < x < 1$.

We are not aware of any other method for demonstrating the existence of a periodic solution to (2.1). We remark also that shooting usually does not provide any information about the uniqueness of the desired solution. Uniqueness is often more difficult to prove than existence. For (2.1), however, a simple comparison method shows that there is only one periodic solution. In particular, there is no solution of period $2k\pi$, with $k > 1$, except for the solution (with **least** period 2π) found above. This is implied by a more general result which you are asked to prove in Exercise 2.2 below.

2.2.1. An alternative formulation of shooting. In this subsection we show that shooting is essentially a topological method, by looking again at the existence of periodic solutions of (2.1). We will describe our approach to this problem somewhat abstractly. Our goal is to prove the existence of solutions with a certain property, say "property P". In this example, an initial choice might be that a solution has property P if it is periodic. We identify a parameter which we are free to choose within some nonempty set Ω. In the example the parameter is α, and $\Omega = R^1$. There must be a topology defined on Ω, and in all of our examples, Ω is connected in this topology.

We don't know if there are any points α in Ω such that $x(t,\alpha)$ has property P, but if not, then obviously there are values of α such that $x(t,\alpha)$ does not have this property. Let

$$\Lambda = \{\alpha \in \Omega \mid x(t,\alpha) \text{ does not have property } P\}.$$

In cases where there is only one "shooting parameter" (i.e. $\Omega \subset R^1$), we usually show that $\Lambda \ne \Omega$ by partitioning Λ into two nonempty, disjoint, open subsets A and B, sometimes called "bad sets". The connectedness of Ω then implies the desired existence result. More complicated problems, as in Section 2.6 below, require more sophisticated topological principles than connectedness.

For (2.1) we again consider the truncated initial value problem (2.2), so that all solutions exist on $[0, 2\pi]$. The property of "being periodic" is too vague, because a division of Λ into two disjoint nonempty open sets is not obvious. Instead, we will say that a solution $y(t,\alpha)$ has property

P if $y(2\pi, \alpha) = \alpha$. We saw earlier that this version of property P implies periodicity with period 2π, and Exercise 2.2 implies that if there is a solution with property P, then it is the only periodic solution.

Let $\Omega = R^1$, and define the bad sets A and B as follows:

$$A = \{\alpha \mid y(2\pi, \alpha) > \alpha\},$$
$$B = \{\alpha \mid y(2\pi, \alpha) < \alpha\}.$$

A solution $y(t, \alpha)$ has property P if and only if $\alpha \notin \Lambda = A \cup B$. The sets A and B are obviously disjoint, simply from their definitions. Our argument above shows that $2 \in A$ and $-2 \in B$, so they are each nonempty. To show that A is open, suppose that $\alpha_1 \in A$. Then $y(2\pi, \alpha_1) > \alpha_1$. Since $y(2\pi, \alpha)$ is a continuous function of α, there is a $\delta > 0$ such that if $|\alpha - \alpha_1| < \delta$ then $y(2\pi, \alpha) > \alpha$. Hence A is open, and similarly B is open. This proves that $A \cup B$ is disconnected, and so there is some $\alpha^* \in R^1$ which is not in $A \cup B$. Then $y(t, \alpha^*)$ is a periodic solution of (2.1).

2.2.2. A problem on $[0, \infty)$. We now look at a generalization of (2.1), by considering

$$x' = x^3 + f(t), \tag{2.3}$$

where we assume only that f is continuous and bounded on $[0, \infty)$. Without assuming periodicity of f we cannot expect there to be a periodic solution. Instead, we ask if there is a solution which is bounded on $[0, \infty)$. To search for such a solution we again consider an initial value problem, now consisting of (2.3) and the initial condition

$$x(0) = \alpha, \tag{2.4}$$

with solution $x(t, \alpha)$. We look for an α^* such that $x(t, \alpha^*)$ is bounded on $[0, \infty)$.

We use the topological formulation of shooting which was introduced in the previous subsection. The point here is to show that the choice of property P is not always obvious. For this problem, one's first guess might be to say that a solution has property P if it exists on $[0, \infty)$ and is bounded.

Proceeding with this suggestion, let M be an upper bound for $|f|$ over $[0, \infty)$, and set $m = M^{1/3}$. From (2.3), if $x(t) > m$ then $x'(t) > 0$, while if $x(t) < -m$ then $x'(t) < 0$. It follows that if $x(t_1) > m$ for some $t_1 \geq 0$, then $x(t) > m$ for all $t \geq t_1$ as long as the solution exists, and a similar remark applies if $x(t_2) < -m$ for some $t_2 \geq 0$.[2] If a solution enters $x > m$ then it is increasing from there on, with $x' \geq x^3 - m^3$, and it easily follows

[2]One further explanation of "as long as": There is no $\tau > t_1$ such that the solution exists on $[t_1, \tau]$, with $x(t_1) > m$ and $x(\tau) \leq m$. If there were such a τ, then there would have to be a $t \in (t_1, \tau)$ with $x(t) > m$ and $x'(t) < 0$, which is impossible from (2.3) and the definition of m.

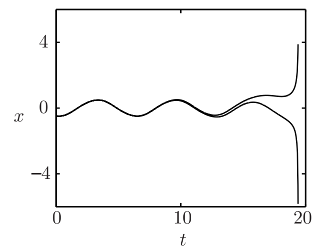

Figure 2.1

that x is unbounded on its maximal interval of existence. Similarly, every solution entering $x < -m$ is unbounded.

These considerations suggest certain choices for the sets A and B. We note that because of the x^3 term on the right of (2.3), unbounded solutions blow up in finite time.[3] To allow for this we let

(2.5)
$$A = \{\alpha \mid x \text{ exists on some interval } [0,\omega) \text{ and } \lim_{t\to\omega^-} x(t) = \infty\},$$
$$B = \{\alpha \mid x \text{ exists on some interval } [0,\omega) \text{ and } \lim_{t\to\omega^-} x(t) = -\infty\},$$

where in each case, $x = x(t,\alpha)$ and ω denotes either a positive number or ∞. We show solutions corresponding to each set in Figure 2.1.

We now consider the three properties required of A and B, namely that they are nonempty, disjoint, and open subsets of R^1. The easiest is disjointedness, for the definitions in (2.5) themselves rule out the existence of a point α which is in both A and B.

We saw above that if $\alpha > m$ then $\alpha \in A$, while if $\alpha < -m$ then $\alpha \in B$, so that A and B are nonempty. But the third property required of A and B is not so straightforward. It is not immediately clear that either is an open set.

Suppose, for example, that $\alpha_1 \in A$, so that $\lim_{t\to\omega^-} x(t,\alpha_1) = \infty$ for some $\omega \le \infty$. Must this be true for all α in some open neighborhood of α_1? The property of "tending to infinity" is not intrinsically an "open property". As an example, for each α let $\phi(\cdot,\alpha)$ be the function defined by

$$\phi(t,\alpha) = \frac{t}{1+\alpha^2 t} \quad \text{for } t \ge 0. \tag{2.6}$$

The function $\phi(\cdot,0)$ is unbounded, but for $\alpha \ne 0$, $\phi(t,\alpha)$ is a bounded function of t. See Exercise 2.1 for related examples.

For this reason, our initial choice for property P and the resulting definitions of the sets A and B are not optimal in this problem. Recall that we chose a fixed $m > 0$ such that $m^3 \ge \sup_{t\ge 0} |f(t)|$. Recognizing that a

[3]See Exercise 1.1.

solution which takes values only in $[-m, m]$ is bounded, we will now say that $x(t, \alpha)$ has property P if $|x(t, \alpha)| \leq m$ for all $t \geq 0$. We then set

$$A = \{\alpha | x(t, \alpha) > m \text{ for some } t \geq 0\},$$
$$B = \{\alpha | x(t, \alpha) < -m \text{ for some } t \geq 0\}.$$

If $\alpha \notin A \cup B$ then the solution $x(t, \alpha)$ can be extended to $[0, \infty)$ and is bounded on that interval.

From the definitions of A and B it is clear that if $\alpha > m$ then $\alpha \in A$, while if $\alpha < -m$ then $\alpha \in B$, so that both sets are nonempty. They are disjoint because, as remarked earlier, once a solution enters the region $|x(t)| > m$, it remains in this region and so does not cross from $x > m$ to $x < -m$ or vice versa.

The final property needed is that A and B are open subsets of Ω. Suppose that $\alpha_0 \in A$. Then there is a t_1 with $x(t_1, \alpha_0) > m$. Since $x(t_1, \alpha)$ is a continuous function of α, there is an open neighborhood $\mathcal{O}$ of α_0 such that if $\alpha \in \mathcal{O}$ then $x(t_1, \alpha) > m$. Therefore A is open, and similarly, B is open. Hence $A \cup B$ does not equal the connected set $\Omega = (-\infty, \infty)$, and so there is an α^* such that $x(t, \alpha^*)$ is bounded on $[0, \infty)$.

We will not go into such detail about these properties again. Usually the openness of the bad sets, such as A and B, follows from continuity of a solution to an initial value problem with respect to initial conditions or parameters. We have proved

Theorem 2.1. *If f is continuous and bounded on $[0, \infty)$, with $|f(t)| \leq m$, then the equation* (2.3) *has a solution x which exists on this interval and satisfies $-m \leq x(t) \leq m$ for all $t \geq 0$.*

In Exercise 2.2 you are asked to show that the bounded solution is unique.

2.3. Some second order examples

2.3.1. A linear problem. We start with the linear second order scalar equation

$$x'' + q(t)x = f(t), \tag{2.7}$$

and we seek a solution satisfying the boundary conditions

$$\begin{aligned} x(0) &= 0, \\ x(1) &= 0. \end{aligned} \tag{2.8}$$

The theory of linear boundary value problems, especially the Fredholm alternative, tells us that there is a solution to (2.7)–(2.8) if and only if f

satisfies the orthogonality condition $\int_0^1 f(t)\,\phi(t)\ dt = 0$ for every solution $\phi(t)$ to the homogeneous equation

$$x'' + q(t)\,x = 0 \tag{2.9}$$

which satisfies (2.8).[4] If the only solution to (2.9) and (2.8) is $\phi = 0$, then (2.7)–(2.8) has a unique solution, which can be found by constructing a Green's function $G(t, \tau)$. We then have the formula

$$x(t) = \int_0^1 G(t, \tau)\, f(\tau)\ d\tau.$$

Looking at the problem another way, let x_p denote the solution to (2.7) such that $x(0) = x'(0) = 0$, and let x_h denote the solution to the corresponding homogeneous equation (2.9) such that $x(0) = 0$, $x'(0) = 1$. Then the general solution to (2.7) with $x(0) = 0$ is

$$x(t, \alpha) = x_p(t) + \alpha x_h(t). \tag{2.10}$$

The boundary value problem (2.7)–(2.8) has a solution if and only if there is an α such that $x(1, \alpha) = 0$. If $x_h(1) \neq 0$ then $x_p(1) + \alpha x_h(1) = 0$ can be solved for a unique α.

To put this in the language of shooting, observe that $x(t, \alpha)$ is the unique solution of the **initial value problem** consisting of (2.7) and the initial conditions

$$x(0) = 0, \qquad x'(0) = \alpha. \tag{2.11}$$

For example, if $x_h(1) > 0$ then from (2.10), $x(1, \alpha) > 0$ if α is large and positive and $x(1, \alpha) < 0$ if α is large and negative. The intermediate value theorem can then be applied as before.

Shooting is not a particularly efficient method for this linear problem. To see its true power we need to consider nonlinear equations.

2.3.2. A nonlinear problem. Our next example is a forced pendulum equation in which the dependent variable x is the angle that a swinging pendulum makes with the downward vertical and the applied force is a continuous function f. With appropriate units the equation becomes[5]

$$x'' + \sin x = f(t). \tag{2.12}$$

We again look for a solution satisfying the boundary conditions

$$x(0) = x(1) = 0. \tag{2.13}$$

[4] See [**201**] or [**82**, p. 146].

[5] This is not the usual linearized pendulum equation seen in elementary ode books. The equation is valid even with large oscillations. However, a first derivative term accounting for friction is omitted.

Once again consider an initial value problem. As in the linear example above, the natural set of initial conditions to choose is

$$\begin{aligned} x(0) &= 0, \\ x'(0) &= \alpha, \end{aligned} \tag{2.14}$$

and we again denote the unique solution by $x(t,\alpha)$.

Because f is continuous and $\sin x$ is bounded, $x(t,\alpha)$ is defined for $0 \leq t \leq 1$ and every α. We wish to find an α^* such that $x(1,\alpha^*) = 0$. Then $x(t,\alpha^*)$ will be a solution to (2.12)–(2.13).

To use shooting, we look for α_1 and α_2 such that $x(1,\alpha_1) > 0$ and $x(1,\alpha_2) < 0$ and apply the intermediate value theorem to the continuous map $\alpha \to x(1,\alpha)$. To show that α_1 and α_2 exist, let

$$M = \max_{0\leq t\leq 1} |f(t)|.$$

Lemma 2.2. *If* $\alpha > \frac{M+1}{2}$ *then* $x(1,\alpha) > 0$. *If* $\alpha < -\frac{M+1}{2}$ *then* $x(1,\alpha) < 0$.

Proof. Suppose that $\alpha > \frac{M+1}{2}$. Then on $[0,1]$,

$$\begin{aligned} x'' &= -\sin x + f(t) \geq -\ (M+1), \\ x' &\geq \alpha - (M+1)t > \frac{M+1}{2} - (M+1)t, \\ x &> \frac{M+1}{2}t - (M+1)\frac{t^2}{2}, \end{aligned}$$

and $x(1,\alpha) > 0$. A similar argument applies if $\alpha < -\frac{M+1}{2}$. □

This lemma and the continuity of $x(1,\alpha)$ with respect to α lead immediately to the existence of at least one α^* such that $x(1,\alpha^*) = 0$, giving the desired solution of (2.12)–(2.13).

We can also phrase this in the topological language of Section 2.2.1, letting

$$\begin{aligned} A &= \{\alpha | x(1,\alpha) > 0\}, \\ B &= \{\alpha | x(1,\alpha) < 0\}. \end{aligned}$$

As in Section 2.2.1, the continuity of $x(1,\alpha)$ as a function of α implies that A and B are open sets. Further, they are disjoint by their definitions. Lemma 2.2 implies that A and B are each nonempty. Since R^1 is connected, $A \cup B \neq R^1$, and hence there is an α^* in R^1 which is not in $A \cup B$. From the definitions of A and B it is seen that $x(1,\alpha^*) = 0$.

2.3.3. Airy's equation on $[0,\infty)$. Airy's equation in its simplest form is

$$y'' - xy = 0. \tag{2.15}$$

It was introduced by Airy, a distinguished nineteenth-century British scientist and mathematician, during his studies of optics [**6**], [**2**]. Mathematically it is important as the simplest example of a "turning point", because the nature of the solutions changes dramatically at $x = 0$. Using the Sturm oscillation theorem, it is easy to see that as x decreases from zero all solutions oscillate more and more frequently, while no solution has more than one critical point in the region $x > 0$. The solutions of this equation are linear combinations of what are called Airy functions, which are among the "special functions" of mathematical physics. We will see one context where they are important in Chapter 3.

Here we will be concerned with a relatively simple problem, chosen to illustrate an important feature of most shooting arguments which has not been encountered in our previous examples. We look for a solution of (2.15) satisfying the conditions

$$\begin{gathered} y\,(0) = 1, \\ \lim_{x\to\infty} y\,(x) = 0. \end{gathered} \tag{2.16}$$

To find such a solution we consider, as usual, an initial value problem, letting $y\,(x) = y\,(x,\alpha)$ denote the solution of (2.15) such that $y\,(0) = 1$, $y'\,(0) = \alpha$. Further, let

$$\begin{aligned} A &= \{\alpha < 0 |\ y\,(x) < 0 \text{ for some } x > 0\}, \\ B &= \left\{\alpha < 0 |\ y'\,(x) > 0 \text{ for some } x > 0\right\}. \end{aligned}$$

We hope it is obvious that these sets are open, by the sort of argument used near the end of the previous subsection, since solutions depend continuously on α. To show that A and B are disjoint, suppose that there is an $\alpha \in A \cap B$. Since $\alpha < 0$ and $y'\,(x) > 0$ for some $x > 0$, there is an $x_1 > 0$ with $y'\,(x_1) = 0$. There is also an $x_2 > 0$ with $y\,(x_2) = 0$. Choose the smallest possible x_1 and x_2, so that $y > 0$ on $[0, x_2)$ and $y' < 0$ on $[0, x_1)$.

We now point out the feature of most interest to us in this example. Suppose that $x_1 = x_2$. Then

$$y\,(x_1) = y'\,(x_1) = 0. \tag{2.17}$$

However the uniqueness theorem for initial value problems then implies that y is the zero solution of (2.15), which is a contradiction because we chose $y\,(0) = 1$. Hence, no nontrivial solution of (2.15) is tangent to the x-axis. Such a "nontangency condition" appears in most shooting arguments, and often establishing nontangency is the key step in the proof.

We have shown that either $x_2 < x_1$ or $x_2 > x_1$. Suppose that $x_2 < x_1$. Then $y(x_2) = 0$ and $y'(x_2) < 0$. Therefore $y < 0$ and $y' < 0$ in some interval $(x_2, x_2 + \varepsilon)$. However if $y < 0$ and $x > 0$, then $y'' < 0$. Therefore $y'(x) < 0$ for all $x > x_2$ where the solution exists, and so x_1 is not defined. A similar contradiction is obtained if $x_1 < x_2$, because if $y > 0$ and $x > 0$, then $y'' > 0$, and so if y' vanishes while $y > 0$, then y is increasing from there on and cannot become negative. Hence, x_2 is not defined, completing the proof that A and B are disjoint.

Finally, we show that these sets are nonempty. Suppose that $\alpha < -3$. Then on some interval $[0, x_0)$, $y' < -2$. Suppose that there is an $x_1 \in (0,1)$ with $y' < -2$ on $(0, x_1)$ and $y'(x_1) = -2$. In the interval $(0, x_1)$, $y < 1$ and so $y'' < 1$. But then, $y'(x_1) < -3 + x_1 < -2$, a contradiction. Therefore, $y' < -2$ on $[0,1]$. Since $y(0) = 1$, we conclude that $y(1) < 0$. Hence $\alpha \in A$, so that A is nonempty.[6]

On the other hand, if $y(0) = 1$ and $y'(0) = 0$, then $y''(0) = 0$ and $y'''(0) = y(0) > 0$ and so for some $x_1 > 0$, $y'(x_1) > 0$. Hence, $y'(x_1) > 0$ for sufficiently small $|\alpha|$. Therefore, small negative numbers are in B, so B is nonempty.

We have shown that A and B are open, nonempty, and disjoint, and so there is an $\alpha < 0$ which is not in $A \cup B$. For such an α, $y > 0$ and $y' < 0$ on $[0, \infty)$. Therefore $L = \lim_{x\to\infty} y(x)$ exists and is nonnegative. If $L > 0$ then y'' becomes large, which clearly causes y' to become positive. Hence $L = 0$.

2.4. Heteroclinic orbits and the FitzHugh-Nagumo equations

2.4.1. Heteroclinic orbits. We begin with

Definition 2.3. Suppose that an autonomous system

$$X' = F(X) \tag{2.18}$$

has at least two distinct equilibrium points, say X_1 and X_2. (Thus, $F(X_1) = F(X_2) = 0$.) Suppose that there is a solution such that $\lim_{t\to-\infty} X(t) = X_1$ and $\lim_{t\to\infty} X(t) = X_2$. Then the trajectory of this solution is called a "heteroclinic" orbit.

Note that if Γ is a heteroclinic orbit then it is an invariant set for (2.18), meaning that if $X(0) \in \Gamma$ then $X(t) \in \Gamma$ for all t where the solution is defined. The existence of compact invariant sets is a fundamental problem in dynamical systems.

[6] It is not necessary to choose the best possible estimates in the argument of this paragraph. Something weaker than $\alpha < -3$ would work, but why work harder than we have to?

Chapter 5

A linear example

5.1. Statement of the problem and a basic lemma

Most of the problems discussed in this book are nonlinear. However to illustrate further the distinction between finite and infinite dimensional approaches to proving the existence of solutions to a boundary value problem, we will discuss a linear system of equations and boundary conditions where several methods work, each of which is important in the right context.

The model comes from mathematical biology and results from the study of molecular motors, which are nanoscale devices for moving material around within biological cells. See [**55**] or [**36**] for further details. Much of the material in this section is taken from [**35**] or [**87**]. We thank David Kinderlehrer for introducing us to this problem.

A relatively simple form of the equations is the system

$$\begin{aligned}(\rho_1' + f_1\rho_1)' &= \alpha_1\rho_1 - \alpha_2\rho_2,\\ (\rho_2' + f_2\rho_2)' &= -\alpha_1\rho_1 + \alpha_2\rho_2,\end{aligned} \tag{5.1}$$

where α_1 and α_2 are continuously differentiable functions on $[0,1]$ which are positive and have period one. Also, f_1 and f_2 are continuously differentiable, with period one and mean value zero. Thus

$$\int_0^1 f_i(x)\ dx = 0, \quad i = 1,2. \tag{5.2}$$

The ρ_i may be viewed as partial probability density functions on $[0,1]$, with the requirement that $\rho_i \geq 0$ for $i = 1,2$ and

$$\int_0^1 (\rho_1(x) + \rho_2(x))\ dx = 1. \tag{5.3}$$

There are several possible boundary value problems for these equations, as noted in [**87**], but here we prove only the existence and uniqueness of positive solutions of period one. These were the solutions considered in [**55**]. By "positive solution", we mean a solution such that $\rho_1 > 0$ and $\rho_2 > 0$ on R^1. Because the problem is linear, every positive solution can be multiplied by a positive constant to achieve the normalization condition (5.3).

We present three existence proofs, one using Schauder's fixed point theorem, one using a "continuation" method in an infinite dimensional space, and one using a combination of elementary linear algebra and a finite dimensional continuation (or "homotopy") argument. Understanding the first two methods requires some background in functional analysis. In our discussion we will state the required abstract results, such as Schauder's theorem, but will not give their proofs or all of the precise definitions required for a full understanding. The reader without this background is encouraged to try to read through the proofs anyway as a beginning towards mastering these ideas. Although they turn out not to be necessary for the problem at hand, they are indispensable in other contexts, especially when dealing with inherently infinite dimensional problems. We will cite sources for the details of this material. The third proof is fully within our category of classical methods. Each of the sections containing these proofs can be read independently of the others. At the end we refer briefly to still another (classical) proof which, while rather special to this problem, is perhaps the shortest of the four proofs we discuss.

All four proofs require a maximum principle type lemma which is itself an important tool, of much wider applicability. We begin the discussion with one version of this result.

Lemma 5.1. *Let a and b be continuously differentiable functions in a closed interval $\bar{J} = [x_0, x_1]$, with $a(x) > 0$ and $b(x) \geq 0$. Let u'' be continuous on $\bar{J}$ and suppose that in the open interval $J = (x_0, x_1)$,*

$$\frac{d}{dx}\left(a(x)\frac{du}{dx}\right) - b(x)\,u \leq 0. \tag{5.4}$$

Then

(i) *the function u does not attain a negative minimum in the open interval J unless b is identically zero and u is constant;*

(ii) *if $u \geq 0$ on J and u is not identically zero on J, then $u > 0$ on J;*

(iii) *if u is not identically zero and $\inf_J u = 0$, then $u = 0$ at either x_0 or x_1, and if $u(x_0) = 0$ then $u'(x_0) > 0$ while if $u(x_1) = 0$ then $u'(x_1) < 0$.*

Proof. First suppose that the inequality in (5.4) is strict. ($<$ instead of $\leq$. The result in this case is sometimes called a "weak maximum principle".)

Then u does not have a minimum in (x_0, x_1) at which $u \leq 0$, proving (i) and (ii) in this case, and we observe also that u is not constant. For (iii), if for example $u(x_0) = 0$ and $u'(x_0) = 0$, then $u''(x_0) < 0$, so that $u < 0$ just to the right of x_0, a contradiction. Similarly, $u'(x_0) < 0$ is ruled out, as is the case $u(x_1) = 0, u'(x_1) \geq 0$, implying (iii).

To prove the full result, with $\leq$ in (5.4), suppose first that u has a negative minimum, say at $x = \zeta \in J$. In every interval $[\zeta, \zeta + \delta)$ where u is negative, the quantity $a(x) u'(x)$ is nonincreasing, so that $u' \leq 0$ to the right of ζ, and likewise, $u' \geq 0$ to the left of ζ. Since u has a minimum at ζ, u must be constant.

For (ii), suppose that $u \geq 0$ in J and for some $\zeta \in J$, $u(\zeta) = \min u = 0$. If u is not constant then choose $\delta \neq 0$ so that $\zeta + \delta \in J$ and $u(\zeta + \delta) > 0$. Consider the case $\delta > 0$, the case $\delta < 0$ being similar. Let

$$v(x) = 1 - e^{\kappa(x-\zeta)}$$

where κ is chosen large enough that

$$\frac{d}{dx}\left(a(x)\frac{dv}{dx}\right) - b(x) v < 0 \tag{5.5}$$

on J. Let $w(x) = u(x) + \varepsilon v(x)$, where $\varepsilon > 0$ is chosen so small that $w(\zeta + \delta) > 0$. From (5.4) and (5.5) we see that $\frac{d}{dx}\left(a(x)\frac{dw}{dx}\right) - b(x) w < 0$ on J. Note that $w(\zeta) = u(\zeta) = 0$ and $w'(\zeta) < 0$. Hence w has a negative minimum in $(\zeta, \zeta + \delta)$. But the weak maximum principle applies to w, giving a contradiction which proves (ii). The proof of (iii) is left as Exercise 5.1. □

5.2. Uniqueness

In this brief section we will illustrate the application of Lemma 5.1 to prove a uniqueness result for our problem. This result is important in itself, and also useful in the second and third existence proofs below, bringing out the important point that often uniqueness implies existence. This principle is not limited to differential equations. For example, the Fredholm alternative, in both finite and infinite dimensional settings, has the consequence that if a linear homogeneous problem of the form $Lu = 0$ has only the trivial (zero) solution, then a nonhomogeneous problem $Lu = f$ has a solution for every allowable f [**201**]. As we will see in Chapter 12, uniqueness for nonlinear problems is often more difficult to prove than existence. But in this case, the uniqueness will be used in proving existence.

Theorem 5.2. *If* f_1 *and* f_2 *are continuously differentiable functions on* R^1 *and there is a positive period one solution* (ρ_1, ρ_2) *of* (5.1) *which also* *satisfies* (5.3)*, then there is no other solution of* (5.1)–(5.3) *(positive or not) with period one.*

Proof. Suppose that (σ_1, σ_2) is a second period one solution of (5.1) which satisfies (5.3). For sufficiently small μ, $(\rho_1 - \mu\sigma_1, \rho_2 - \mu\sigma_2)$ is a positive periodic solution. However, for $\mu > 1$, $\int_0^1 \rho_1 + \rho_2 - \mu(\sigma_1 + \sigma_2)\ dx < 0$ and so $(\rho_1 - \mu\sigma_1, \rho_2 - \mu\sigma_2)$ is not positive. Let

$$\mu^* = \sup\{\mu \mid (\rho_1 - \mu\sigma_1, \rho_2 - \mu\sigma_2) \text{ is a positive solution}\}.$$

Then $(\rho_1 - \mu^*\sigma_1, \rho_2 - \mu^*\sigma_2)$ is a nonnegative solution such that some component $\rho_i - \mu^*\sigma_i$ vanishes at some $\zeta \in [0,1]$. Suppose for convenience that $i = 1$. Let $F_1(x) = \int_0^x f_1(s)\ ds$ and notice that $u = e^{F_1}(\rho_1 - \mu^*\sigma_1)$ satisfies the equation

$$\left(e^{-F_1}u'\right)' - \alpha_1 e^{-F_1}u = -\alpha_2(\rho_2 - \mu^*\sigma_2) \le 0$$

on $[0,1]$. By Lemma 5.1 either u is identically zero or ζ is one of the end points 0 or 1. In the first case, $\int_0^1 (\rho_1 + \rho_2)\ dx = \mu^* \int_0^1 (\sigma_1 + \sigma_2)\ dx = 1$, implying that $\mu^* = 1$ and $(\rho_1, \rho_2) = (\sigma_1, \sigma_2)$. In the second case, periodicity implies that $u(0) = u(1) = 0$ and that $u'(0) = u'(1)$. By (iii) of Lemma 5.1, $u'(0) \neq 0$. However $u > 0$ on $(0,1)$ implies that $u'(0) \ge 0$ and $u'(1) \le 0$ and a contradiction results. □

5.3. Existence using Schauder's fixed point theorem

Theorem 5.3. *Suppose that* f_1, f_2, α_1*, and* α_2 *are continuously differentiable on* R^1 *and have period one. Suppose also that* α_1 *and* α_2 *are positive and that* f_1 *and* f_2 *have mean value zero on* $[0,1]$*. Then the system* (5.1) *has a unique positive solution with period one satisfying* (5.3).

Uniqueness was proved above. We will outline the proof of existence using Schauder's theorem, which is stated below. First we show how a fixed point theorem is relevant.

Rewrite the system (5.1) as follows:

$$\left(\rho_1' + f_1\rho_1\right)' - \alpha_1\rho_1 = -\alpha_2\rho_2,$$
$$\left(\rho_2' + f_2\rho_2\right)' - \alpha_2\rho_2 = -\alpha_1\rho_1.$$

Looking at the individual equations of this system motivates consideration of a single equation of the form

$$(\rho' + f\rho)' - \alpha\rho = -\beta\sigma, \tag{5.6}$$

where σ is assumed to be a known periodic function, β is a positive function, and we look for a periodic solution ρ. (Every periodic function will be assumed to have period one, though this may not be the least period of the function.) Let $F(x) = \int_0^x f(s)\,ds$. If f is periodic and has zero mean over a period, then F is also periodic. Let $\eta = e^F \rho$. Then

$$(e^{-F}\eta')' - \alpha e^{-F}\eta = -\beta\sigma. \tag{5.7}$$

This equation is in selfadjoint form. The corresponding homogeneous equation has no periodic solution, because η has no positive maximum or negative minimum. Hence by the Fredholm alternative theory (theory of Green's functions in this case), (5.7) has a unique periodic solution for every continuous periodic σ (e.g. [**41**, Chapter 7, Theorem 2.2]). Setting

$$\rho = e^{-F}\eta \tag{5.8}$$

gives a mapping T_0 from σ to a unique periodic solution ρ of (5.6).

Next, consider a pair of periodic functions (σ_1, σ_2) and the system

$$\begin{aligned} (\rho_1' + f_1\rho_1)' - \alpha_1\rho_1 &= -\alpha_2\sigma_2, \\ (\rho_2' + f_2\rho_2)' - \alpha_2\rho_2 &= -\alpha_1\sigma_1. \end{aligned} \tag{5.9}$$

Applying the mapping T_0 to each equation, with f replaced successively by f_1 and f_2, α by α_1 and α_2, and β by α_2 and α_1, we obtain a map T taking (σ_1, σ_2) to the unique pair (ρ_1, ρ_2) of periodic solutions to the equations in (5.9). By Lemma 5.1, if $\sigma_i \geq 0$ for $i = 1, 2$ and neither σ_i is identically zero, then $\rho_i > 0$ for $i = 1, 2$. We look for fixed points of T, that is, pairs (σ_1, σ_2) of continuous periodic functions such that

$$T(\sigma_1, \sigma_2) = (\sigma_1, \sigma_2).$$

Clearly a fixed point for T is a periodic solution to (5.1). We now state the Schauder fixed point theorem.

Theorem 5.4. *Let B be a Banach space and K a closed convex subset of B. Let $T : K \to B$ be a continuous transformation from K into B such that the image $T(K) \subset K$ and the closure of $T(K)$ is compact. Then T has a fixed point in K.*

For the needed definitions, and a proof, see [**62**]. To apply this theorem to the system (5.1) we must specify the Banach space B and the closed convex set K. To motivate these definitions, first look in more detail at the periodic solution of (5.6). This solution is constructed by using a Green's function $G(x, y)$, which is bounded and has bounded first partial derivatives at every (x, y) with $y \neq x$. From the equation for η in terms of G and σ, namely

$$\eta(x) = \int_0^1 G(x,y)\,\sigma(y)\,dy, \tag{5.10}$$

one obtains bounds on η and η'. These bounds can be stated in several norms, but it will be seen below that the $\mathcal{L}^1$ norm on $[0,1]$, denoted in this section by $||\ ||$, is useful here. (See [**62**], Appendix D for background.)

We need to know that $T(K)$ is compact for an appropriate Banach space B and closed convex set K. This will follow if the mapping T_0 is compact for every f and α, meaning that if K_0 is a closed bounded subset of $\mathcal{L}^1([0,1])$ then the image $T_0(K_0)$ is a compact set. Equivalently, we need to show that every sequence of functions in $T_0(K_0)$ has a convergent subsequence.[1]

The Ascoli-Arzela lemma tells us that it is sufficient for $T_0(K_0)$ to be an equicontinuous uniformly bounded set of functions. Neither of these properties, equicontinuity and uniform boundedness, is natural in the space $\mathcal{L}^1([0,1])$. They are best defined for subsets of the space $C([0,1])$ of continuous functions on $[0,1]$. Fortunately, the formulas (5.8) and (5.10) give us the bounds which are needed. From these equations and the properties of G it follows that there is a constant M such that if $\rho = T_0(\sigma)$ then for every $x \in [0,1]$,

$$\begin{aligned} |\rho(x)| &\le M\,||\beta\sigma||, \\ |\rho'(x)| &\le M\,||\beta\sigma||. \end{aligned} \tag{5.11}$$

Since β is bounded and both β and σ are nonnegative, there are uniform estimates for ρ and ρ' in terms of the $\mathcal{L}^1$ norm of σ, and therefore the mapping $T_0 : \mathcal{L}^1([0,1]) \to \mathcal{L}^1([0,1])$ is compact. Each function $T_0(\sigma)$ is in $\mathcal{L}^1([0,1])$ and has a representative ρ which is continuous and differentiable in $[0,1]$, with $|\rho|$ and $|\rho'|$ bounded in terms of $||\sigma||$. Also, from (5.11),

$$||\rho|| = \int_0^1 |\rho(x)|\,dx \le M\,||\beta\sigma||, \tag{5.12}$$

so that the linear mapping T_0 is continuous.

To apply the Schauder fixed point theorem, B is chosen to be the space $\mathcal{L}^1([0,1]) \times \mathcal{L}^1([0,1])$, with norm $||(\sigma_1,\sigma_2)||_B = ||\sigma_1|| + ||\sigma_2||$.

[1]The reader should think carefully about what sort of convergence is needed and whether this type of convergence is proved.

Let K be the set of pairs $(\sigma_1, \sigma_2) \in B$ such that (a representative of) each σ_i is periodic and the following additional properties hold for $i = 1, 2$:

$$\begin{gathered}\sigma_i \geq 0 \text{ on } [0,1] \text{ for } i = 1,2,\\ \int_0^1 \left(\alpha_1(x)\,\sigma_1(x) + \alpha_2(x)\,\sigma_2(x)\right)\, dx = 1,\\ \int_0^1 \left(\alpha_1(x)\,\sigma_1(x) - \alpha_2(x)\,\sigma_2(x)\right) dx = 0.\end{gathered} \tag{5.13}$$

It is easily seen that K is a closed convex set in B. The compactness of $T(K)$ follows from the compactness of T_0, and continuity of T is also inherited in this way. To complete the proof of Theorem 5.3 it must also be shown that T maps K into K. We leave this as Exercise 5.2.

5.4. Existence using a continuation method

Continuation methods have been used in a variety of types of problems, including ode's, pde's, and integral equations [**61**], [**160**]. They require more functional analysis than was needed for the Schauder theorem, and we will be content here to outline a proof of Theorem 5.3 assuming this background, with appropriate citations where it can be found. This method also relies on Lemma 5.1 and Theorem 5.2. Continuation is closely related to Leray-Schauder degree theory, which allows extension of the method to partial differential equations. See [**163**] for an application to swirling flow in fluid mechanics.

A continuation method begins by finding a system similar to the given problem but for which it is known that a unique solution of the desired type exits. In the case of (5.1) we start with the system

$$\begin{aligned}(\rho_1' + f_1\rho_1)' &= \alpha_1^0\rho_1 - \alpha_2^0\rho_2,\\ (\rho_2' + f_2\rho_2)' &= -\alpha_1^0\rho_1 + \alpha_2^0\rho_2,\end{aligned}$$

where $\alpha_i^0(x) = e^{F_i}$ for $i = 1, 2$. A positive periodic solution to this system is given by $\rho_i = e^{-F_i}$.

The method then continues by considering systems of the form

$$\begin{aligned}(\rho_1' + f_1\rho_1)' &= \alpha_1^\lambda\rho_1 - \alpha_2^\lambda\rho_2,\\ (\rho_2' + f_2\rho_2)' &= -\alpha_1^\lambda\rho_1 + \alpha_2^\lambda\rho_2,\end{aligned} \tag{5.14}$$

where for $i = 1, 2$, $\alpha_i^\lambda(x) = \lambda\alpha_i(x) + (1-\lambda)\,\alpha_i^0(x)$.

We will show that the solution $\rho^0 = \left(\rho_1^0, \rho_2^0\right)$ is part of a continuous family of solutions of (5.14) obtained as λ increases from 0 to 1. The method is often challenging for a nonlinear equation, but the current problem is linear, which simplifies the analysis.

We now suppose that a positive periodic solution $\left(\rho_1^{\lambda_0}, \rho_2^{\lambda_0}\right)$, necessarily unique, has been found for some $\lambda_0 \geq 0$ and show how to continue this solution to a nearby λ_1. Denote the solution which we seek at $\lambda = \lambda_1$ by $\left(\rho_1^{\lambda_1}, \rho_2^{\lambda_1}\right)$. If this solution exists then $\phi = \rho^{\lambda_1} - \rho^{\lambda_0}$ satisfies

$$\phi_i'' + (f_i\phi_i)' - \alpha_i^{\lambda_0}\phi_i = -\alpha_j^{\lambda_0}\phi_j + S_i \tag{5.15}$$

for $i = 1, 2$ and $j \neq i$, where $S_i = \left(\alpha_i^{\lambda_1} - \alpha_i^{\lambda_0}\right)\rho_i^{\lambda_1} - \left(\alpha_j^{\lambda_1} - \alpha_j^{\lambda_0}\right)\rho_j^{\lambda_1}$. The unknown function ρ^{λ_1} appears on both sides of (5.15) . Therefore we consider

$$\phi_i'' + (f_i\phi_i)' - \alpha_i^{\lambda_0}\phi_i = -\alpha_j^{\lambda_0}\phi_j + S_i^* \tag{5.16}$$

where $S_i^* = \left(\alpha_i^{\lambda_1} - \alpha_i^{\lambda_0}\right)\rho_i^* - \left(\alpha_j^{\lambda_1} - \alpha_j^{\lambda_0}\right)\rho_j^*$ for some known pair $\rho^* = (\rho_1^*, \rho_2^*)$ of positive continuous periodic functions on $[0, 1]$. The plan is to show that (5.16) has a unique periodic solution ϕ for each ρ^*, giving a map from ρ^* to $\rho^{\lambda_1} = \phi + \rho^{\lambda_0}$, and showing that this map has a fixed point. This fixed point will solve (5.14) for $\lambda = \lambda_1$. A fixed point exists because, for $|\lambda_0 - \lambda_1|$ sufficiently small, the map $\rho^* \to \rho^{\lambda_1}$ is a contraction in an appropriate space.

Write (5.16) symbolically as

$$-T_i\phi_i = -\alpha_j^{\lambda_0}\phi_j + S_i^*. \tag{5.17}$$

The operator T_i can be put in the form of a selfadjoint operator on the Hilbert space $\mathcal{P}$ of periodic functions in $\mathcal{L}^2[0, 1]$, by setting

$$\eta_i = e^{F_i}\phi_i.$$

Then for η in $C^2([0, 1])$,

$$T_i\phi_i = -\frac{d}{dx}\left(e^{-F_i(x)}\eta_i'\right) + \alpha_i^{\lambda_0}e^{-F_i}\eta_i = M_i\eta_i, \text{ say,} \tag{5.18}$$

where M_i has the standard form for a selfadjoint second order differential operator in $L^2(0, 1)$; see Chapter 7 of [**41**]. Further, M_i is what is called a positive operator, meaning that for all $\eta_i \neq 0$, $\int_0^1 \eta_i(M_i\eta_i)\,ds > 0$. To show this we integrate by parts:

$$\int_0^1 \eta_i(M_i\eta_i)\,ds = -e^{-F_i(x)}\eta_i'\eta_i|_0^1 + \int_0^1 e^{-F_i}\eta_i'^2\,ds + \int_0^1 \alpha_i^{\lambda_0}e^{-F_i}\eta_i^2 ds,$$

which is positive because the boundary terms cancel due to periodicity.

The general theory of positive operators is invoked at this stage. See [**127**]. This theory tells us that the inverse M_i^{-1} exists in $\mathcal{P}$, and (5.17) becomes

$$\eta_i = M_i^{-1}\left(\alpha_j^{\lambda_0} e^{-F_j}\eta_j\right) - M_i^{-1}S_i^*. \tag{5.19}$$

M_i^{-1} is an integral operator with a symmetric kernel (a Green's function). We then see that M_i^{-1} is a compact operator, meaning (as in the Schauder theorem) that if Ω is a closed bounded set in $\mathcal{P}$ then the image $M_i^{-1}(\Omega)$ is a compact set, and sequences in this set always have convergent subsequences. The compactness of M_i^{-1} follows ultimately from the Ascoli-Arzela theorem.

The operator M_i^{-1} is positive (from the general theory), but it has the additional property that it takes positive functions to positive functions. (Notice that this is not the same as being a positive operator, which is a property involving inner products.) To see this, observe that if $M_i\eta_i$ is positive then $\frac{d}{dx}\left(e^{-F_i}\eta_i'\right) - \alpha_i^{\lambda_0}e^{-F_i}\eta_i < 0$. Also, $\eta_i(0) > 0$ and $\eta_i(1) > 0$, for otherwise η_i could not be periodic. Then $\eta_i > 0$ in $[0,1]$ by Lemma 5.1.

Now consider the pair of equations (5.19) with $i = 1, 2$ and $j \neq i$, which we can write as

$$\eta = P\eta + QS^*, \tag{5.20}$$

where η is in the space H of periodic functions in $L^2(0,1) \times L^2(0,1)$. Here

$$(QS^*)_i = -M_i^{-1}\left(\left(\alpha_i^{\lambda_1} - \alpha_i^{\lambda_0}\right)\rho_i^* - \left(\alpha_j^{\lambda_1} - \alpha_j^{\lambda_0}\right)\rho_j^*\right). \tag{5.21}$$

The linear operator P is necessarily compact and takes pairs of positive functions to pairs of positive functions. Furthermore, P has 1 as an eigenvalue, since a corresponding eigenfunction has the positive components $\eta_i = e^{F_i}\rho_i^{\lambda_0}$. It is then a further standard result from the theory of positive operators that the eigenvalue 1 is simple, meaning that there are no other eigenfunctions [**127**]. This fact is deducible also from the uniqueness result that we proved earlier. Further, this eigenvalue is the eigenvalue of largest modulus. Thus if, in (5.20), QS^* is orthogonal to the unique positive eigenfunction corresponding to the eigenvalue 1 for the adjoint operator P^*, then by the Fredholm alternative [**201**] we can solve (5.20) for the η_i, and the solution is unique modulo an additive multiple of $\left(e^{F_1}\rho_1^{\lambda_0}, e^{F_2}\rho_2^{\lambda_0}\right)$.

To complete the proof, we need to investigate P^*. By definition,

$$(P\psi)_i = M_i^{-1}\left(\alpha_j^{\lambda_0}e^{-F_j}\psi_j\right).$$

To find the adjoint P^*, we recall that on H the inner product is $\langle\psi,\omega\rangle = \int_0^1 (\psi_1\omega_1 + \psi_2\omega_2)\,dx$ and the adjoint P is defined by $\langle P\psi,\omega\rangle = \langle\psi, P^*\omega\rangle$.

We write

$$\int_0^1 \left(\omega_1 \left(P\psi\right)_1 + \omega_2 \left(P\psi\right)_2\right) dx$$

$$= \int_0^1 \left(\omega_1 M_1^{-1}\left(\alpha_2^{\lambda_0} e^{-F_2}\psi_2\right) + \omega_2 M_2^{-1}\left(\alpha_1^{\lambda_0} e^{-F_1}\psi_1\right)\right) dx$$

$$= \int_0^1 \left(\alpha_2^{\lambda_0} e^{-F_2}\psi_2 M_1^{-1}\left(\omega_1\right) + \alpha_1^{\lambda_0} e^{-F_1}\psi_1 M_2^{-1}\left(\omega_2\right)\right) dx,$$

since M_i and M_i^{-1} are selfadjoint. Hence,

$$\left(P^*\omega\right)_i = \alpha_i^{\lambda_0} e^{-F_i} M_j^{-1}\left(\omega_j\right).$$

It follows that the solution to $\omega = P^*\omega$ is given by

$$\omega_i = M_i\left(1\right), \tag{5.22}$$

as is seen by substitution. From (5.18), $\omega_i = \alpha_i^{\lambda_0} e^{-F_i}$.

We now have to check that QS^* in (5.20) is orthogonal to ω. From equation (5.21), we must show that

$$\sum_{\substack{i=1\\ j\neq i}}^{2} \int_0^1 M_i^{-1}\left(\left(\alpha_i^{\lambda_1} - \alpha_i^{\lambda_0}\right)\rho_i^* - \left(\alpha_j^{\lambda_1} - \alpha_j^{\lambda_0}\right)\rho_j^*\right)\alpha_i^{\lambda_0} e^{-F_i} dx = 0.$$

Since M_i is selfadjoint, this is

$$\int_0^1 \sum_{\substack{i=1\\ j\neq i}}^{2} \left(\left(\alpha_i^{\lambda_1} - \alpha_i^{\lambda_0}\right)\rho_i^* - \left(\alpha_j^{\lambda_1} - \alpha_j^{\lambda_0}\right)\rho_j^*\right) M_i^{-1}\left(\alpha_i^{\lambda_0} e^{-F_i}\right) dx = 0,$$

but as we saw in (5.22),

$$M_i^{-1}\left(\alpha_i^{\lambda_0} e^{-F_i}\right) = 1$$

for $i = 1, 2$, which gives the result because the integrand is now zero.

We now have a solution (η_1, η_2) to

$$\eta = P\eta + QS^*$$

for each ρ^*, and it is unique modulo an additive multiple of $\left(e^{F_1}\rho_1^{\lambda_0}, e^{F_2}\rho_2^{\lambda_0}\right)$. We specify the multiple by insisting that if $\phi_i = e^{-F_i}\eta_i$ and $\rho^1 = \phi + \rho^0$, then

$$\int_0^1 \left(\rho_1^1 + \rho_2^1\right) = \int_0^1 \left(\rho_1^0 + \rho_2^0\right) = 1.$$

We thus have a mapping $L : \rho^* \to \rho^1$. From (5.21) it can be shown that for sufficiently small $|\lambda_1 - \lambda_0|$ this map is a contraction on the orthogonal complement of the eigenspace corresponding to the eigenvalue 1, giving the desired solution ρ^{λ_1}.

The final step is to show that the set Λ of λ for which a solution exists includes the entire interval $[0,1]$. This step follows by showing that Λ is both open and closed in $[0,1]$ and nonempty. We will give a similar argument in the next section (in a finite dimensional setting) and so omit it here. This completes our outline of a proof by continuation for Theorem 5.3.

As we are seeing in this chapter, there are several existence proofs, and the continuation method as outlined above requires the most sophisticated functional analysis of the three we will give. However, in more global problems, such as integral equations and differential-difference equations, this degree of sophistication is required, and it is perhaps in these areas that the method of continuation really comes into its own, as in [**61**] and [**160**].

5.5. Existence using linear algebra and finite dimensional continuation

This proof of Theorem 5.3 is divided into two parts. It turns out that existence can be proved independently of the uniqueness and positivity of the solution. To do so, we use a standard ode approach and write our system (5.1) of two second order equations as a first order system of four equations. Define new independent variables ϕ_1 and ϕ_2, by setting

$$\phi_i = \rho_i' + f_i \rho_i, \quad i = 1, 2.$$

The system (5.1) is then equivalent to

$$\begin{aligned} \rho_1' &= \phi_1 - f_1 \rho_1, \\ \rho_2' &= \phi_2 - f_2 \rho_2, \\ \phi_1' &= \alpha_1 \rho_1 - \alpha_2 \rho_2, \\ \phi_2' &= \alpha_2 \rho_2 - \alpha_1 \rho_1. \end{aligned} \tag{5.23}$$

This system is nonautonomous and linear with continuous coefficients on $[0,1]$ and so has a fundamental solution Z, a 4×4 matrix function whose columns form a fundamental set of solutions for the system. Further, we can assume that Z satisfies $Z(0) = I$, the 4×4 identity matrix.

Every solution of (5.23) can be written as

$$\begin{pmatrix} \rho_1 \\ \rho_2 \\ \phi_1 \\ \phi_2 \end{pmatrix} = Z\mathbf{c} \tag{5.24}$$

where $\mathbf{c}$ is a constant vector. A solution is periodic if and only if

$$Z(1)\,\mathbf{c} = \mathbf{c}. \tag{5.25}$$

This system of four algebraic equations in four unknowns $c_1, \ldots, c_4$ has a nontrivial solution if and only if the matrix

$$Z(1) - I$$

is singular.

From (5.23) it is seen that $\phi_1' + \phi_2' = 0$. Hence, the sum of the last two rows of $Z'(x)$ is zero. Since $Z(0) = I$, the last two rows of $Z(0) - I$ are zero. Therefore the sum of the last two rows of $Z(x) - I$ is zero for all x, which implies that $Z(x) - I$ is singular for each x. Therefore $Z(1) - I$ is singular, and so (5.1) has a nontrivial periodic solution.

Remark 5.5. This proof shows that the existence of a periodic solution to (5.23) is a problem in linear algebra. The signs of the terms in (5.1) play no role. In contrast, it was found in Section 5.2 that the uniqueness of a positive solution depends on these signs, and we will see now that existence of such a solution is also dependent on these signs.

To prove existence of a positive solution we observe as in the previous section that this is easy in the special case

$$\alpha_1^0 = e^{F_1}, \qquad \alpha_2^0 = e^{F_2} \tag{5.26}$$

where

$$F_i(x) = \int_0^x f_i(s)\,ds.$$

A positive periodic solution of (5.23) with (5.26) is given by

$$\rho_i^0 = e^{-F_i}, \qquad \phi_i^0 = \rho_i' + f_i\rho_i \tag{5.27}$$

for $i = 1, 2$. The corresponding solution to (5.23) is $\left(e^{-F_1}, e^{-F_2}, 0, 0\right)^T$, and by (5.24) a solution $\mathbf{c}^0$ to (5.25) is $(1, 1, 0, 0)^T$. Theorem 5.2 implies that if α_1 and α_2 in (5.23) are positive, then a positive periodic solution is the only periodic solution up to a constant multiple, and so the solution space of (5.25) is one dimensional.

It is convenient now to use a different normalization from (5.3), by considering solutions in the form (5.24) with

$$\|\mathbf{c}\| := \max_{i=1,\ldots,4} |c_i| = 1. \tag{5.28}$$

If (5.23) has a positive periodic solution, then there is a unique solution to (5.25) with $c_1 > 0$ and satisfying (5.28).

We prove existence of a positive solution for given positive, continuous, and periodic functions α_1 and α_2 by showing that (5.25) is satisfied by a nonzero vector $\mathbf{c}$ such that the solution with initial condition $Z(0)\,\mathbf{c}$ is positive on $[0, 1]$. It is useful to phrase the problem in terms of a homotopy between the pairs $\left(\alpha_1^0, \alpha_2^0\right)$ and (α_1, α_2). Introduce a parameter λ taking values in the interval $[0,1]$. Let $\alpha_i^\lambda = \lambda\alpha_i + (1-\lambda)\,\alpha_i^0$, $i = 1,2$. For each λ we consider the system

$$\begin{aligned} \rho_1' &= \phi_1 - f_1\rho_1, \\ \rho_2' &= \phi_2 - f_2\rho_2, \\ \phi_1' &= \alpha_1^\lambda\rho_1 - \alpha_2^\lambda\rho_2, \\ \phi_2' &= \alpha_2^\lambda\rho_2 - \alpha_1^\lambda\rho_1. \end{aligned} \tag{5.29}$$

There is a corresponding system of linear equations

$$Z_\lambda(1)\,\mathbf{c}^\lambda = \mathbf{c}^\lambda, \tag{5.30}$$

where Z_λ is the fundamental matrix solution of (5.29). We have shown that for every $\lambda \in [0,1]$ there is a nontrivial solution to (5.30). We can assume that for this solution $c_1 \geq 0$ and (5.28) holds. If the corresponding solution of (5.29) is positive, then this solution is unique and $c_1 > 0$.

We now show that if a periodic solution of (5.29) and (5.28) is positive, and therefore unique, when $\lambda = \lambda_0$, then there is a positive periodic solution to these two equations for λ in a neighborhood of λ_0, and this solution (necessarily unique) is continuous in λ at λ_0. In fact, all this follows if we prove that for every sequence $\{\lambda_i\} \subset [0,1]$ converging to λ_0 and corresponding nonzero solutions $\mathbf{c}^{\lambda_i}$ of (5.30) with $\left\|\mathbf{c}^{\lambda_i}\right\| = 1$ and $c_1 \geq 0$, it is the case that $\lim_{i\to\infty}\mathbf{c}^{\lambda_i} = \mathbf{c}^{\lambda_0}$.

Otherwise there is a sequence $\{\lambda_i\} \subset [0,1]$ which converges to λ_0 such that some sequence $\left\{\mathbf{c}^{\lambda_i}\right\}$ of solutions to (5.30) satisfying (5.28) and $c_1 \geq 0$ does not converge to $\mathbf{c}^{\lambda_0}$. It follows that some subsequence of $\left\{\mathbf{c}^{\lambda_i}\right\}$ converges to some $\mathbf{d} \neq \mathbf{c}^{\lambda_0}$ which also satisfies (5.30) with $\lambda = \lambda_0$, (5.28), and $d_1 \geq 0$. However $Z_{\lambda_0}(t)\,\mathbf{d}$ is a periodic solution of (5.29) for $\lambda = \lambda_0$, and $\mathbf{d} \neq -\mathbf{c}^{\lambda_0}$, which contradicts the uniqueness of $Z_{\lambda_0}(t)\,\mathbf{c}^{\lambda_0}$. The idea that uniqueness implies continuity is rather common in analysis.

Now let Λ be the set of $\lambda \in [0,1]$ such that (5.29) has a positive periodic solution. Clearly $0 \in \Lambda$. We will show that $1 \in \Lambda$ by showing that Λ is both open and closed within $[0,1]$.

To show that Λ is open, recall that for each λ, the matrix $Z_\lambda(1) - I$ is singular. If $\lambda_0 \in \Lambda$ then the null space of $Z_{\lambda_0}(1) - I$ has dimension 1, since the nontrivial solution of (5.30) is unique to within a multiplicative constant. Hence the rank of $Z_{\lambda_0}(1) - I$ is three. In other words, there is a nonzero three-by-three subdeterminant of this matrix. Since $Z_\lambda(1)$ is a continuous function of λ, the same subdeterminant is nonzero for λ in some neighborhood of λ_0, while $Z_\lambda(1) - I$ remains singular. Hence the set of solutions of (5.30) is still one dimensional. There is a unique solution to (5.30) satisfying (5.28) and with $c_1^\lambda > 0$ for λ in this neighborhood. Hence, in this neighborhood $\mathbf{c}^\lambda$ is a continuous function of λ, implying that $\left(\rho_1^\lambda, \rho_2^\lambda\right)$ remains positive in some neighborhood of λ_0.

To show that Λ is closed, suppose that λ_1 is a limit point of Λ which is not in Λ. The set of all functions which are periodic solutions of (5.29) for some λ and with initial conditions satisfying (5.28) is equicontinuous and uniformly bounded, since both the solutions and their derivatives must be bounded by a constant independent of λ. Hence corresponding to λ_1 there is a nontrivial periodic solution of (5.29) which is a limit of positive solutions. For this solution $\rho_i \geq 0$ for $i = 1, 2$, and at least one of the ρ_i is zero somewhere in $[0, 1]$. Applying Lemma 5.1 to $\eta = e^{F_i}\rho_i$ shows that ρ_i is identically zero and it follows easily from (5.29) that both components are identically zero. But then, $\phi_1 = \phi_2 = 0$, so that $\mathbf{c}^{\lambda_1} = 0$, a contradiction.

Since Λ is open and closed in $[0, 1]$ and nonempty, $\Lambda = [0, 1]$, and in particular, $1 \in \Lambda$. This completes the proof of Theorem 5.3.

5.6. A fourth proof

In [**35**] there is still another proof, entirely classical and perhaps shorter than the Schauder proof above.[2] It is based on the observation that in (5.23), $\phi_1' + \phi_2' = 0$. Hence, $\phi_2 = -\phi_1 + c$, and the system reduces to a set of three equations with an extra parameter. As with the other proofs, important use is made of Lemma 5.1 and Theorem 5.2. However we did not give this proof here, because it is rather special to the particular system (5.1). In a subsequent paper, [**87**], the theory was extended to a system of n equations, in variables $\rho_1, \ldots, \rho_n$. Each of the three methods for proving existence which we gave above applies to the larger system, but the special fourth proof used in [**35**] does not.

Finally, we mention that the main emphasis in both [**35**] and [**87**] is not on the existence and uniqueness of solutions, but rather on physically important properties of the solutions. We refer the reader to these papers, particularly [**87**], for details.

5.7. Exercises

Exercise 5.1. Complete the proof of Lemma 5.1 by proving (iii) more generally.

Exercise 5.2. Complete the proof of existence by Schauder's fixed point theorem by showing that if (σ_1, σ_2) is in K and (ρ_1, ρ_2) is the unique periodic solution to (5.9), then (ρ_1, ρ_2) is in K. Hint: Show that in K, neither σ_1 nor σ_2 is identically zero. Also, note from (5.15) that $\int_0^1 \alpha_i(x)\,\rho_i(x)\,dx \leq 1$ for $i = 1, 2$, and use (5.11).

[2] Different boundary conditions are treated in [**35**], but the proof can be adapted to the case of periodic solutions.

Shooting with more parameters

13.1. A problem from the theory of compressible flow

In Chapter 2 we gave an example of shooting in which two initial conditions had to be adjusted to satisfy the boundary conditions. In this chapter we discuss a more difficult example of shooting with two parameters than the model problem treated earlier. The setting is flow of a compressible fluid past a fixed boundary surface. The differential equations in the fluid boundary layer were derived by Stewartson [**227**]. They are

$$\begin{gathered} f''' + ff'' + \lambda\left(h - f'^2\right) = 0, \\ h'' + fh' = 0, \end{gathered} \tag{13.1}$$

where f' represents the velocity tangential to the surface and h is proportional to the total energy of the system.

The boundary conditions in the compressible flow case (13.1) involve a constant h_0, which we will assume satisfies

$$0 < h_0 < 1. \tag{13.2}$$

Then we want

$$\begin{gathered} f(0) = f'(0) = 0, \quad f'(\infty) = 1, \\ h(0) = h_0, \quad h(\infty) = 1. \end{gathered} \tag{13.3}$$

In this section we prove the existence of a solution to (13.1) and (13.3) when (13.2) holds. The proof is based on one given by McLeod and Serrin in [**166**], where the physical background is discussed and further references given. McLeod and Serrin treat a more general problem, but the essence of the shooting method they use is seen in the problem (13.1)–(13.3) with condition (13.2). A uniqueness result for this problem is given in [**85**]. Existence when $h_0 > 1$ is also proved in [**166**].

Note that we obtain a family of solutions to (13.1) by setting $h = 1$ and letting f be a solution of the Falkner-Skan equation, which was discussed in Chapter 9. However, these do not satisfy the boundary conditions (13.3).

13.1.1. Existence of a solution. The following proof differs in some details from the one given in [**166**]. We take advantage of the relative simplicity of the problem we are treating, and also of insights obtained from the efficient and easy to use numerical ode packages now available. We used the program XPP [**55**], of G. B. Ermentrout, which we find particularly convenient for shooting experiments, but other programs could be used instead.

Theorem 13.1. *The equation* (13.1) *has a solution satisfying* (13.3) *for each* $h_0 \in (0, 1)$.

Proof. As expected in a shooting proof, we consider the initial value problem consisting of (13.1) and

$$\begin{gathered} f(0) = 0, \quad f'(0) = 0, \quad f''(0) = \alpha, \\ h(0) = h_0, \quad h'(0) = \beta. \end{gathered} \tag{13.4}$$

There are two shooting parameters, α and β, and two conditions to be satisfied at infinity. We use the following result from plane topology. This is similar to Lemma 2.8.

Lemma 13.2. *Let* J *be a closed rectangle* $a \le x \le b, c \le y \le d$, *and let* S^+ *and* S^- *be disjoint open subsets of* R^2 *such that* S^- *contains the left edge* $x = a$ *of* J *and* S^+ *contains the right edge* $x = b$. *Then the complement* D *in* J *of* $(S^+ \cup S^-) \cap J$ *contains a continuum* γ *connecting the lines* $y = c$ *and* $y = d$ *in* J.

Proof. A short elementary proof is given in [**166**], and a similar result is proved in [**180**]. Note that S^- and S^+ need not be connected sets. They only need to have connected components which contain the left and right sides of the rectangle. The theorem proved by degree theory in the appendix to this chapter includes this result. □

We will take a moment to discuss the use of computations to help find appropriate sets S^+ and S^-. In panels A–F of Figure 13.1 we plot h and f'^2 for a variety of initial conditions. In each case $h(0) = \frac{1}{2}$. These graphs guided us to the choices we made of S^- and S^+, which are different from the sets used in [**166**]. At the time of the original proof, the available numerical packages were much more cumbersome than what is now available, making it more time consuming to search in this way, especially when two shooting

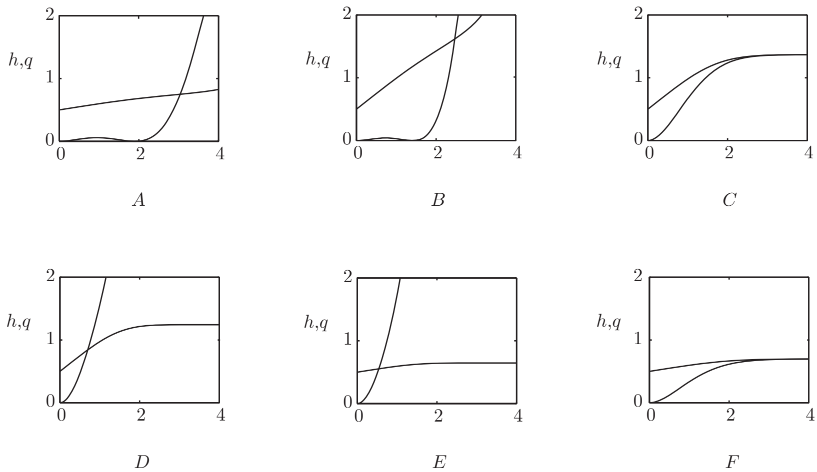

Figure 13.1. Graphs of h and $q = (f')^2$; note that $q(0) = 0$.

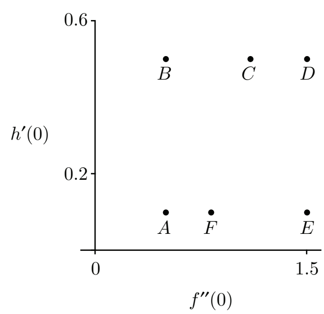

Figure 13.2. Initial conditions for solutions above.

parameters were involved. The locations in the (α, β)-plane of the initial conditions which gave each of these pairs of graphs is shown in Figure 13.2.

We now define S^+ and S^-. In these definitions, (f, h) denotes the unique solution to (13.1) and (13.4), with the dependence on (α, β) understood. From (13.4) we see that if $\alpha > 0$ and $\beta > 0$ then on some initial interval $(0, \delta)$, $f'' > 0$ and $h - f'^2 > 0$. Let R_+^2 denote the closed quadrant $\alpha \geq 0, \beta \geq 0$. Then set

$$S^- = \{(\alpha, \beta) \in R_+^2 \mid \text{there is a } t_1 > 0 \text{ such that } f''(t_1) < 0 \text{ and } h - f'^2 > 0 \text{ on } [0, t_1]\},$$

$$S^+ = \Big\{(\alpha, \beta) \in R_+^2 \mid \text{there is a } t_2 > 0 \text{ such that } h(t_2) - f'(t_2)^2 < 0 \text{ and } f'' > 0 \text{ on } [0, t_2]\Big\}.$$

By referring to Figure 13.1, with panels labeled A through F, it can be seen that the points labeled A and B in Figure 13.2 are initial conditions in S^-, while D and E are in S^+. Also, C and F are not in $S^+ \cup S^-$, and so these two points may be in γ.

Since $h_0 > 0$, we see that if $\alpha = 0$ then $f'''(0) < 0$, and so $(0, \beta) \in S^-$ for every $\beta \geq 0$. The sets S^+ and S^- are clearly open, because the solutions of (13.1) are continuous with respect to the initial conditions. Also, they are disjoint from their definitions. The following lemma restricts the behavior of (f, h) when $(\alpha, \beta) \notin S^- \cup S^+$.

Lemma 13.3. *None of the following are possible for a solution of* (13.1) *and* (13.4) *with* $\alpha \geq 0$, $\beta \geq 0$, *and* $(\alpha, \beta) \notin S^+ \cup S^-$.

(i) $f'' > 0$ *on* $[0, \tau)$, $f''(\tau) = 0$, $h - f'^2 > 0$ *in* $[0, \tau]$, *and* $f'''(\tau) = 0$.

(ii) $f'' > 0$ *on* $[0, \tau]$, $h - f'^2 > 0$ *in* $[0, \tau)$, $h(\tau) = f'(\tau)^2$ *and* $\left(h - f'^2\right)' = 0$ *at* τ.

(iii) $f'' > 0$ *on* $[0, \tau)$, $h - f'^2 > 0$ *on* $[0, \tau)$, *and* $f''(\tau) = h(\tau) - f'(\tau)^2 = 0$.

Proof. (i) If $f'' = 0$ then $f''' = -\lambda\left(h - f'^2\right)$, which is negative at τ. (Thus, $f'' < 0$ just after τ, and $(\alpha, \beta) \in S^-$.)

(ii) An easy computation shows that at τ,

$$\left(h - f'^2\right)'' = -2f''(\tau)^2 < 0,$$

and so $h - f'^2$ must be negative somewhere in $(0, \tau)$, a contradiction.

(iii) Note that $h'(t) = \beta e^{-\int_0^t f(s)ds}$. Under the given conditions,

$$\left(h - f'^2\right)'(\tau) = h'(\tau) \geq 0.$$

If $\beta > 0$, then $h'(\tau) > 0$ and $h - f'^2$ is negative somewhere in $(0, \tau)$, again a contradiction. If $\beta = 0$, then $h \equiv h_0$, $f' \equiv \sqrt{h_0}$, and the solution does not satisfy $f'' > 0$ on $[0, \tau)$. □

Corollary 13.4. *If* $\alpha \geq 0$, $\beta \geq 0$ *and* $(\alpha, \beta) \notin S^- \cup S^+$, *then the solution exists on* $[0, \infty)$, *with* $f'' > 0$ *and* $h - f'^2 > 0$. *Further,* $\lim_{x\to\infty} h(x)$ *exists.*

Proof. If $\alpha = 0$ and $\beta \geq 0$, then $f'''(0) < 0$ and $(\alpha, \beta) \in S^-$. Therefore, the lemma implies that if $\alpha \geq 0$, $\beta > 0$ and $(\alpha, \beta) \notin S^- \cup S^+$, then $f'' > 0$ and $h - f'^2 > 0$ as long as the solution exists. Thus, f' is increasing and f is unbounded. Integrating the equation for h in (13.1) shows that h is bounded, and since $f'^2 < h$, f' is also bounded. Thus the solution exists on $[0, \infty)$ and h, which is increasing, approaches a limit $h(\infty)$. □

We will choose J to be a rectangle of the form

$$J = \{(\alpha, \beta) \mid 0 \le \alpha \le \alpha_2, \beta_1 \le \beta \le \beta_2\},$$

for some $\alpha_2 > 0$ and $\beta_2 > \beta_1 > 0$.

Lemma 13.5. *The numbers α_2, β_1, and β_2 can be chosen so that S^- contains the left side of J and S^+ contains the right side of J and so that if (α, β) is on the top edge $\beta = \beta_2$ of J and not in $S^+ \cup S^-$, then $h(t) > 1$ for some $t > 0$, while if (α, β) is on the bottom edge $\beta = \beta_1$ of J and not in $S^+ \cup S^-$, then $\sup_{t \ge 0} h(t) < 1$.*

Proof. We identify several steps.

(a) Recall that $(0, \beta) \in S^-$ for every $\beta \ge 0$ and that S^- is open. It follows that there is a continuous positive function $\alpha(\beta)$ defined for $\beta \ge 0$ such that if $0 \le \alpha < \alpha(\beta)$ then $(\alpha, \beta) \in S^-$.

(b) We use the following result.

Lemma 13.6. *If $\beta = \sqrt{\frac{2}{\pi}}$ and $(\alpha, \beta) \notin S^+ \cup S^-$, then $h(\infty) > 1$.*

Proof. We saw earlier that $h(\infty)$ exists. Suppose that $h(\infty) \le 1$. Since $f'^2 < h$ on $(0, \infty)$, $f' < 1$ and $f < x$. Hence,

$$h'(x) = \beta e^{-\int_0^x f ds} \ge \beta e^{-\frac{x^2}{2}}, \tag{13.5}$$

$$h(x) \ge h_0 + \beta \int_0^x e^{-\frac{s^2}{2}} ds,$$

and a contradiction results if $\beta = \sqrt{\frac{2}{\pi}}$. □

Hence, we take $\beta_2 = \sqrt{\frac{2}{\pi}}$.

(c) If $\beta \le \beta_2$ then $h \le h_0 + \sqrt{\frac{2}{\pi}} x$. Then an argument used in Chapter 9, in the proof of Theorem 9.1, shows that α_2 can be chosen independently of β so that if $\alpha = \alpha_2$ then $f'(x)^2 > h(x)$ for some $x \in (0, 1]$.

(d) This step requires the most thought. Choosing α_2 from part (c), we have to find a positive number β_1 such that if $0 < \alpha \le \alpha_2$ and $(\alpha, \beta_1) \notin S^+ \cup S^-$, then $h(\infty) < 1$. It was shown above that $(0, 0) \in S^-$. Hence there is a $\rho > 0$ such that the square σ_ρ of points (α, β) such that $0 \le \alpha \le \rho$ and $0 \le \beta \le \rho$ is contained in S^-. We will choose $\beta_1 \in (0, \rho)$.

To give further conditions on β_1 we assume that $(h_0, \beta_1) \in J \setminus (S^- \cup S^+)$, which implies that $(\alpha, \beta_1) \notin \sigma_\rho$, and so $\rho < \alpha \le \alpha_2$. Initially, therefore, $f'' > h'$.

Lemma 13.7. *We can choose β_1 so small that if $\beta = \beta_1$ and $f'' > h'$ on an interval $[0, T)$, then on this interval, $h' \le \rho$ and $h \le \frac{1+h_0}{2}$.*

Proof. Choosing $\beta_1 < \rho$ ensures that $h' < \rho$ on $[0,\infty)$, since $h'' \le 0$. In the interval $[0,T)$,

$$f' > h - h_0,$$
$$f(x) > \int_0^x (h(s) - h_0)\,ds,$$
$$\int_0^x f(s)\,ds > \int_0^x (x-s)(h(s)-h_0)\,ds.$$

Hence, $h'(x) = \beta_1 e^{-\int_0^x f(s)ds} \le \beta_1 e^{-\int_0^x (x-s)(h(s)-h_0)ds} ds.$

If $h \le \frac{2h_0+1}{3}$ on $[0,T)$ then (13.2) implies the result. Suppose that $h(\tau) = \frac{2h_0+1}{3}$ for some $\tau < T$. Then on $[\tau, T]$,

$$h'(t) \le \beta_1 e^{-\int_\tau^t (t-s)\frac{1-h_0}{3} ds} \le \beta_1 e^{-\frac{1-h_0}{3}\frac{(t-\tau)^2}{2}}.$$

Integrating from τ to ∞, we get

$$h(t) \le \frac{2h_0+1}{3} + \sqrt{\frac{3}{2}}\beta_1 \frac{\sqrt{\pi}}{\sqrt{1-h_0}}.$$

We can thus choose β_1 so small (and independent of T) that $h(t) \le \frac{1+h_0}{2}$ on $[0,T)$. This proves the lemma. □

Choose β_1 as in Lemma 13.7, and assume that $(\alpha,\beta_1) \notin S^- \cup S^+$. Further, suppose that $f''(T) = h'(T)$, with $f'' > h'$ on $[0,T)$. Then $h(T) \le \frac{1+h_0}{2}$ and $f''(T) = h'(T) \le \beta_1$. Recalling the earlier stage of the argument where we considered $\alpha = \beta = 0$, we see that if β_1 is sufficiently small, independent of T, then f'' turns negative, contradicting our assumption that $(\alpha,\beta_1) \notin S^- \cup S^+$. Hence, if $(\alpha,\beta_1) \in \gamma$ then $f'' > h'$ on $[0,\infty)$ and $h(\infty) < 1$, as desired. This completes the proof of Lemma 13.5. □

Lemmas 13.2 and 13.5 imply that there is a continuum γ connecting $\beta = \beta_2$ and $\beta = \beta_1$ and lying in J such that $\gamma \cap (S^- \cup S^+)$ is empty. Suppose that $(\alpha,\beta) \in \gamma$. Then $\alpha > 0$. Using Lemma 13.3, we see that $f'' \ge 0$ and $f'^2 \le h$ as long as the solution exists. Also, if $x > 0$ and $f(x)$ exists, then $f'(x) > 0$ and $f(x) > 0$. Since $\beta > 0$, it follows from (13.1) that $h' > 0, h'' < 0$, and so $h' \le \beta$, $h \le h_0 + \beta x$. Thus, $f' \le \sqrt{h_0 + \beta x}$, from which it follows that the solution exists in $[0,\infty)$.

We also see that for $x \ge 1$, $f' \ge f'(1) > 0$, and so

$$h'(x) = \beta e^{-\int_0^x f} \le c e^{-\mu x}$$

for some $\mu > 0, c > 0$. Thus, $\lim_{x\to\infty} h(x) = h_\infty$ exists. Also, $f'^2 < h_\infty$, and $f'' > 0$, so that $f'(\infty)$ exists and is positive. If $p = h(\infty) - f'(\infty)^2 > 0$ then for large x,

$$f''' + ff'' \le -\lambda \frac{p}{2}.$$

Since $f > 0, f'' > 0$, it follows that for large x, $f'''(x) \leq -\lambda \frac{p}{2}$, which implies that f'' becomes negative, a contradiction. Hence,

$$\lim_{x\to\infty} \left(h(x) - f'(x)^2 \right) = 0.$$

We must now show that $h(\infty)$ is a continuous function of (α, β) on γ. In other words, suppose $(\bar{\alpha}, \bar{\beta}) \in \gamma$ and $\varepsilon > 0$. Then we must show that there is a $\delta > 0$ such that if $(\alpha, \beta) \in \gamma$ and

$$|\alpha - \bar{\alpha}| + \left|\beta - \bar{\beta}\right| < \delta, \tag{13.6}$$

then

$$\left|h(\infty) - \bar{h}(\infty)\right| < \varepsilon, \tag{13.7}$$

where $(\bar{f}, \bar{h})$ is the solution with initial conditions $(\bar{\alpha}, \bar{\beta})$. Since $\bar{\beta} \geq \beta_1 > 0$, this is a fairly routine step, which is mostly left to the reader in Exercise 13.1. Given ε, we choose $T > 0$ such that $\int_T^\infty \bar{\beta} e^{-\int_0^s \bar{f}(s)ds} < \frac{\varepsilon}{4}$, for example. We can then choose δ so small that if (13.6) holds and $(\alpha, \beta) \in \gamma$, then $\left|h(T) - \bar{h}(T)\right| < \frac{\varepsilon}{2}$ and $\beta \int_T^\infty e^{-\int_0^s f(s)ds} < \frac{\varepsilon}{2}$, implying (13.7) and proving the continuity of $h(\infty)$. The existence of an $(\alpha, \beta) \in \gamma$ such that $h(\infty) = 1$ then follows from Lemma 13.5. □

13.2. A result of Y.-H. Wan

The McLeod-Serrin two-parameter shooting principle, and similar topological results such as Lemma 2.8, can be extended to higher dimensions. In this section we state such a generalization. It is an intriguing result which one feels should be useful in some circumstances, and it was helpful in one three-dimensional setting [**88**]. However we have never encountered another problem where it could be applied. We state this result in the hope that a reader may find it of interest. It is due to our former colleague, Y.-H. Wan, whom we thank for allowing us to include it here. Only the three-dimensional version was published previously, for lack of other applications. We should also mention a difficult four-dimensional shooting proof by Dunbar [**53**]. We have not investigated whether Wan's result would simplify Dunbar's proof.

Theorem 13.8. *Let $B = [-1,1]^n \times [-1,1]$. For $i = 1, \ldots, n$ let $\Omega_i^\pm$ be open sets in B. Further, let $\Delta_1 = B$, and for $j = 2, \ldots, n+1$, let*

$$\Delta_j = B \backslash \left\{ \bigcup_{i=1}^{j-1} \left(\Omega_i^+ \cup \Omega_i^- \right) \right\}. \tag{13.8}$$

Chapter 19

Three unsolved problems

Statements of Problems

There may be readers who turn to this chapter first, being most interested in what is unknown. In this short final chapter we mention three other problems which have occupied a number of researchers in recent decades, and yet have so far failed to reveal their secrets.[1] Space constraints keep us from giving many details about the various approaches which have been tried, but for each problem we cite at least one paper which in turn has a number of citations to earlier work in the field. We would be thrilled if one or more of these problems were to be solved by researchers who first learned of them in this book. We begin by succinctly stating the problems and then give some background for each. The last section contains a new proof, based partly on work of W. C. Troy, of an important piece of the background for the third problem.

19.1. Homoclinic orbit for the equation of a suspension bridge

Prove that if $0 < q < 2$ then *the equation*

$$u^{(iv)} + qu'' + e^u - 1 = 0 \tag{19.1}$$

has a nontrivial solution u *such that* $\lim_{s\to-\infty} u(s) = \lim_{s\to\infty} u(s) = 0$.

[1]The prudent researcher may wish to search recent literature to be sure that these problems haven't been solved since the book went to press!

19.2. The nonlinear Schrödinger equation

Show that for some $a > 0$ the system

$$\begin{aligned} x'' + \tfrac{2}{\xi}x' + x\left(x^2 + y^2 - 1\right) - ay - a\xi y' = 0, \\ y'' + \tfrac{2}{\xi}y' + y\left(x^2 + y^2 - 1\right) + ax + a\xi x' = 0 \end{aligned} \tag{19.2}$$

has a solution $(x(\xi), y(\xi))$ on $(0, \infty)$ such that

$$\begin{aligned} \lim_{\xi\to 0^+} y(\xi) = \lim_{\xi\to 0^+} x'(\xi) = \lim_{\xi\to 0^+} y'(\xi) = 0, \\ \lim_{\xi\to\infty} \xi x'(\xi) + x(\xi) - \tfrac{1}{a}y(\xi) = 0, \\ \lim_{\xi\to\infty} \xi y'(\xi) + \tfrac{1}{a}x(\xi) + y(\xi) = 0. \end{aligned} \tag{19.3}$$

19.3. Uniqueness of radial solutions for an elliptic problem

Prove that for each positive integer k the boundary value problem

$$u'' + \frac{2}{r}u' - u + u^3 = 0, \tag{19.4}$$

$$\lim_{r\to 0^+} u'(r) = 0, \quad \lim_{r\to\infty} u(r) = 0 \tag{19.5}$$

has at most one solution which is initially positive and has exactly k zeros in $(0, \infty)$.

References and some background

19.4. Comments on the suspension bridge problem

This is the newest and least studied of these problems, and also the most specialized, but it has challenged a number of researchers for the past two decades. The equations of a suspension bridge were developed largely by McKenna and coworkers in [**146**], [**33**], [**162**] and references therein. To explain the key physical feature, let $u = u(x,t)$ denote the vertical displacement downward of the bridge surface from its rest position at distance x along the bridge and time t. When $u > 0$, the cables near that location are stretched and exert an upward restoring force on the bridge. But when $u < 0$, there is no corresponding downward force. Thus, cables are different from springs. The force term $e^u - 1$ is a smooth function showing a version of this asymmetry. It changes sign at $u = 0$, but $|e^u - 1|$ is much larger then $|e^{-u} - 1|$ when u is large and positive. The specific model above is found in [**33**]. It was studied in the monograph [**190**], which is where we found the specific formulation (19.1). A number of interesting solutions are given there but the traveling wave problem is left as a conjecture.

The book [**190**] is also a good source for a number of other references for the problem. We mention particularly the paper [**220**], which gives a

Painlevé II in Chapter 3. A third equivalent version can be found by using polar coordinates for the dependent variable $Q = x + iy$. It appears from numerical computation that the problem has a solution for each d with $2 < d \leq 4$. However the only existence proofs are those by Kopell and Landman [**123**] and by Rottshäfer and Kaper [**205**], each of which covers a range $2 < d < 2 + \delta$ for a small δ.[3]

19.6. Comments on the elliptic problem and a new existence proof

Our final problem can be obtained from the nonlinear Schrödinger equation as well as a number of other problems in partial differential equations. Starting with (19.6), we ask if there are standing wave solutions. These are solutions of the form

$$\Phi(x,t) = e^{i\lambda^2 t}\psi(x).$$

Substituting this into (19.6) and looking for real solutions for ψ, we get

$$\Delta\psi - \lambda^2\psi + \psi^3 = 0. \tag{19.10}$$

Then, we look for radially symmetric solutions $\psi(x) = u(r)$ where $r = |x|$. Setting $\lambda = 1$ and taking the spatial dimension n to be 3 gives equation (19.4):

$$u'' + \frac{2}{r}u' - u + u^3 = 0.$$

The boundary conditions (19.5), namely

$$u'(0) = 0, \quad u(\infty) = 0,$$

are obtained from the physical problems where these equations arise [**251**].

19.6.1. Existence and uniqueness of solutions. The first result we know about was by Nehari [**179**], who proved that the problem (19.4)–(19.5) has at least one positive solution. Positive solutions are the most important physically, since they are solutions of minimum energy and are the most stable [**251**]. In [**42**] Coffman proved that there is only one positive solution, and we repeated his proof in Chapter 12. This solution is called the "ground state", because it is the solution of lowest energy in an appropriate sense.

Solutions which have sign changes, called "bound states", are also important, as localized finite energy solutions [**252**]. In [**207**] it was shown by Ryder that for each $k \geq 1$ there is a solution to (19.4)–(19.5) with exactly k zeroes in $(0, \infty)$. The problem posed above is whether these solutions are unique. There are no results for this problem. At the end of this section, and this book, we give a new proof of the existence of the bound states.

[3] The range of d in [**205**] is slightly larger than that in [**123**].

19.6.2. Extensions. Following Coffman's paper, K. McLeod and Serrin [**168**] considered equations of the form

$$u'' + \frac{n-1}{r}u' + |u|^{p-1}\,u - u = 0 \tag{19.11}$$

with boundary conditions

$$u'(0) = 0, \quad u(\infty) = 0. \tag{19.12}$$

It is known from work of Pohozaev [**193**] that for $n > 2$ there are no solutions if $p \geq \frac{n+2}{n-2}$. Coffman's result was for the case $p = n = 3$. In [**168**] uniqueness of the ground state was shown under a variety of conditions on p and n. In particular, the proof covers those (n,p) such that $2 < n \leq 4$ and $1 < p < \frac{n+2}{n-2}$. In [**134**], M.-K. Kwong completely solved the ground state uniqueness problem by removing the restriction on n, showing that for every $n > 2$ there is a unique positive solution if $1 < p < \frac{n+2}{n-2}$. His proof uses techniques of Coffman and Sturm oscillation theory. The result was extended, and the proof shortened, by K. McLeod [**167**], and a further shortening, using geometric methods, was given by Clemons and Jones in [**39**]. The latter proof made use of an "Emden-Fowler" transformation, and it would be interesting to see if a shorter classical proof could be found using this transformation.

There are also solutions to problem (19.11)–(19.12) which are not positive. Once again the condition $1 < p < \frac{n+2}{n-2}$ is imposed, and then, for every $k \geq 1$ there is a solution with exactly k zeros. This result was proved in [**111**] by a dynamical systems method and in [**169**] by a classical method.[4] The uniqueness of these solutions is unknown for any $k > 0$. In [**241**] Troy obtained a uniqueness result for $k = 1, 2$ in the case where the nonlinear term is a piecewise linear function mimicking $u - u^3$ [**241**].

Remark 19.1. Equation (19.10) is the basis of both our second and third unsolved problems. It is a special case of

$$\Delta u + f(u) = 0$$

where $f(u) \sim |u|^{p-1}\,u$ as $|u| \to \infty$. This equation has been intensively studied in R^n for arbitrary positive integers n, and the properties of solutions depend markedly on the values of n and p. The literature is too vast to summarize here, or to list a significant number of references. One way to search for these would be to look for citations of the landmark paper of Pohozaev [**193**]. For papers more closely related to the problem (12.44)–(12.45) and its extensions, look for citations of [**42**].

[4] The classical proof is considerably shorter.