Chapter 4

Nonlinear Systems: Global Theory

Theorem 3.2.1 guarantees that a solution to the initial value problem exists for what might be an extremely short time. Typically, the ODEs that arise in physical applications possess solutions for much longer times than can be deduced with the contraction-mapping principle, and in this chapter we introduce methods for demonstrating this behavior. The technique is based on extending a short-time solution obtained from Theorem 3.2.1. The main tool for such extensions, Theorem 4.1.2, is proved in Section 4.1. In Section 4.2, this theorem is used to derive two theoretical results that guarantee global existence.

In Sections 4.3 and 4.4 we introduce techniques, including *nullclines*, for verifying the hypotheses of these global existence theorems, and we illustrate the techniques by applying them to specific ODEs drawn from various fields. These equations will reoccur frequently in later chapters.

We regard the introduction of meaningful applications to illustrate the theory as one of the attractive features of this book. In the present chapter we consider the ODEs, without motivation, from a purely mathematical point of view; this analysis completes the theoretical treatment of IVPs begun in Chapters 2 and 3. (In Chapter 5 we introduce models in their original form, including interpretation of the variables and underlying physical assumptions. Central to that chapter, we study scaling as a systematic technique to simplify the original equations to forms more convenient for analysis, as considered in the present chapter.)

In the last two sections of this chapter we show that the solution of an IVP depends continuously (Section 4.5) and even differentiably (Section 4.6) on its initial conditions. These results are a fundamental part of the theory.

An appendix is devoted to Euler's method, the simplest numerical approximation for solutions of an IVP. In particular, we prove that as the step size tends to zero, the

D.G. Schaeffer, J.W. Cain, *Ordinary Differential Equations: Basics and Beyond*, Texts in Applied Mathematics, DOI 10.1007/978-1-4939-6389-8_4

approximations converge to the true solution. This proof closely mimics the proof in Section 4.5 that the solution of an IVP depends continuously on its initial data.

4.1 The Maximal Interval of Existence

Our first result asserts that there is a maximal interval for which the solution of an IVP exists, a sort of "gold standard" for solutions.

Proposition 4.1.1. *Let $\mathbf{F} : \mathcal{U} \to \mathbb{R}^d$ be locally Lipschitz on $\mathcal{U} \subset \mathbb{R}^d$. Given $\mathbf{b} \in \mathcal{U}$, there is a solution $\mathbf{x}_* : (-\alpha_*, \beta_*) \to \mathcal{U}$ of the IVP*

$$\mathbf{x}' = \mathbf{F}(\mathbf{x}), \qquad \mathbf{x}(0) = \mathbf{b} \tag{4.1}$$

that is maximal in the following sense: if another function $\mathbf{x}$ solves (4.1) for t in some open interval $\mathcal{I}$, then

$$\textit{(i)}\ \mathcal{I} \subset (-\alpha_*, \beta_*) \qquad \textit{and} \qquad \textit{(ii)}\ \mathbf{x}(t) = \mathbf{x}_*(t)\ \textit{for}\ t \in \mathcal{I}. \tag{4.2}$$

Remark: It often happens that either α_* or β_*, or both, equals infinity. Note that the maximal interval of existence is always open, even if α_* or β_* is finite.

Proof. We focus only on β_* and $t \geq 0$, leaving the analogous treatment of α_* and $t \leq 0$ for the dedicated reader. Let

$$\beta_* = \sup\{\beta : \text{ IVP (4.1) is solvable for } 0 \leq t < \beta\}.$$

Of course, by Theorem 3.2.1, $\beta_* > 0$. For $n = 1, 2, \ldots$, choose solutions $\mathbf{x}_n$ of (4.1) that exist for times $t \in [0, \beta_n)$, where $\beta_n \to \beta_*$, finite or infinite. To define $\mathbf{x}_*$, given $t \in [0, \beta_*)$ choose any n such that $\beta_n > t$ and let

$$\mathbf{x}_*(t) = \mathbf{x}_n(t). \tag{4.3}$$

By Theorem 3.3.4, the uniqueness result, the definition (4.3) does not depend on the choice of n, and moreover, $\mathbf{x}_*$ is a solution of (4.1). It is readily checked *(do so!)* that every solution $\mathbf{x}$ of (4.1) on some interval $\mathcal{I}$ satisfies properties (i) and (ii) of (4.2). □

Although it may not be apparent, the following result is extremely useful in extending solutions to larger times. Of course there is an analogous result for negative time.

Theorem 4.1.2. *Suppose, regarding the maximal solution $\mathbf{x}_* : (-\alpha_*, \beta_*) \to \mathbb{R}^d$, that $\beta_* < \infty$. Then for every compact set $\mathcal{K} \subset \mathcal{U}$, there is an $\varepsilon > 0$ such that $\mathbf{x}_*(t) \notin \mathcal{K}$ for $\beta_* - \varepsilon < t < \beta_*$.*

Proof. By Corollary 3.3.3, there is a compact set $\mathcal{K}'$ and a $\delta > 0$ such that for all $\mathbf{x} \in \mathcal{K}$, the closed ball $\overline{B(\mathbf{x}, \delta)}$ is contained in $\mathcal{K}'$. Let $M = \max_{\mathcal{K}'} |\mathbf{F}(\mathbf{x})|$, let L be a Lipschitz constant for $\mathbf{F}$ on $\mathcal{K}'$, and choose $\varepsilon < \min\{\delta/M, 1/L\}$.

We claim that $\mathbf{x}_*(t) \notin \mathcal{K}$ if $t > \beta_* - \varepsilon$. Suppose to the contrary that there is a time $t_0 > \beta_* - \varepsilon$ such that $\mathbf{x}_*(t_0) \in \mathcal{K}$. It follows from Corollary 3.2.8 that the IVP

$$\mathbf{y}' = \mathbf{F}(\mathbf{y}), \qquad \mathbf{y}(t_0) = \mathbf{x}_*(t_0)$$

has a solution on $(t_0 - \varepsilon, t_0 + \varepsilon)$. Applying Lemma 3.2.9, we conclude that the original solution $\mathbf{x}_*$ may be defined on $[0, t_0 + \varepsilon)$. But $t_0 + \varepsilon > \beta_*$, which contradicts the hypothesis that β_* was maximal. □

4.2 Two Sufficient Conditions for Global Existence

4.2.1 Linear Growth of the RHS

Our first result[1] gives existence for all times, positive and negative.

Theorem 4.2.1. *If $\mathbf{F} : \mathbb{R}^d \to \mathbb{R}^d$ is locally Lipschitz and if there exist nonnegative constants B, K such that*

$$|\mathbf{F}(\mathbf{x})| \le K|\mathbf{x}| + B, \qquad \mathbf{x} \in \mathbb{R}^d, \tag{4.4}$$

then the solution $\mathbf{x}(t)$ of (4.1) exists for all time, $-\infty < t < \infty$, and moreover,

$$|\mathbf{x}(t)| \le |\mathbf{b}| e^{K|t|} + \frac{B}{K}(e^{K|t|} - 1), \qquad -\infty < t < \infty. \tag{4.5}$$

Proof. This proof will use the generalization of Gronwall's lemma given in Exercise 3.8(a). We consider only forward time, $t \ge 0$; negative time can be handled with trivial modifications of the argument. Suppose (4.1) has a solution for $t \in [0, \beta)$, which of course satisfies the integral equation

$$\mathbf{x}(t) = \mathbf{b} + \int_0^t \mathbf{F}(\mathbf{x}(s))\, ds, \qquad 0 \le t < \beta.$$

Defining $g(t) = |\mathbf{x}(t)|$, we deduce that

$$g(t) \le |\mathbf{b}| + \int_0^t [Kg(s) + B]\, ds, \qquad 0 \le t < \beta.$$

Hence by the generalized Gronwall lemma, $\mathbf{x}$ satisfies the estimate (4.5) for its entire domain of existence, $0 \le t < \beta$.

[1] We alert you one final time: the same symbol $\mathbf{x}$ may simply denote a point in $\mathbb{R}^d$ (as in (4.4)) or may denote a vector-valued function of time (as in (4.5)).

Now let $\mathbf{x}_*, \beta_*$ be the maximal solution of (4.1), and suppose $\beta_* < \infty$. According to (4.5), $\mathbf{x}_*(t)$ belongs to the compact ball

$$\mathcal{K} = \{\mathbf{z} \in \mathbb{R}^d : |\mathbf{z}| \le |\mathbf{b}|e^{K\beta_*} + \frac{B}{K}(e^{K\beta_*} - 1)\}$$

for all $t \in [0, \beta_*)$. This estimate contradicts Theorem 4.1.2, so we must have β_* infinite. □

This result may be easily extended to nonautonomous equations that satisfy a linear-growth estimate. (See Exercise 5(a).)

4.2.2 Trapping Regions[2]

Our second result, which gives global existence in forward time only, is more widely applicable but also requires more explanation.

(a) An introductory example

Rewriting Duffing's equation (1.28) as a first-order system, we obtain

$$\begin{aligned} x' &= y \\ y' &= -\beta y + x - x^3. \end{aligned} \tag{4.6}$$

Let us repeat the calculation from Section 1.4.1 that the energy

$$E(x, y) = y^2/2 - x^2/2 + x^4/4 \tag{4.7}$$

decreases along trajectories of (4.6). Indeed, by the chain rule,

$$\frac{dE}{dt} = \frac{\partial E}{\partial x}\frac{dx}{dt} + \frac{\partial E}{\partial y}\frac{dy}{dt} = \langle \nabla E, \mathbf{F} \rangle,$$

where $\langle \cdot, \cdot \rangle$ is the inner product on $\mathbb{R}^2$. Obtaining ∇E from (4.7) and $\mathbf{F}$ from (4.6), we calculate that

$$\langle \nabla E, \mathbf{F} \rangle = -\beta y^2 \le 0. \tag{4.8}$$

For a constant E_0, consider the sublevel set

$$\mathcal{K} = \{(x, y) \in \mathbb{R}^2 : E(x, y) \le E_0\}, \tag{4.9}$$

[2]Trapping regions represent a first hint of a shift toward more geometric thinking in this book. In this connection, you may be amused by the aphorism, "Geometry is the art of reasoning well from badly drawn figures." The oldest citation that we can give for this is from an article by Poincaré in 1895 [64], but apparently the quotation was old even at that time, for Poincaré introduces it with the remark, "It is worth repeating ...."

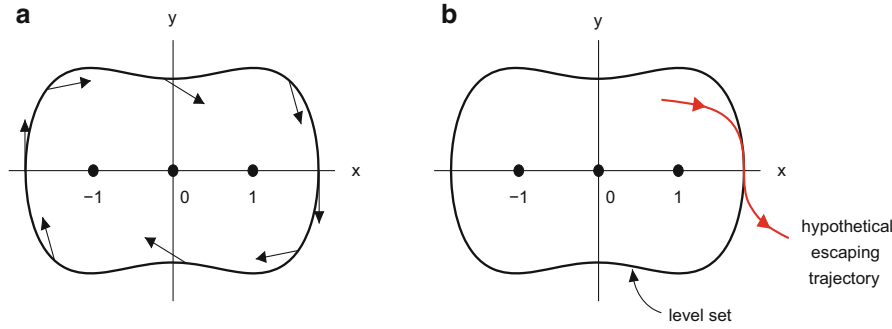

Figure 4.1: *(a) The level set $E(x, y) = 1$, where the energy $E(x, y)$ is given by (4.7), and the flow direction for Duffing's equation (4.6) at selected points on the curve, assuming $\beta = 1$. The flow is strictly inward except for $y = 0$, where it is tangential. (b) Hypothetical escaping trajectory for the Duffing system. Were such a trajectory possible, it would have to cross the level set tangentially at one of the two points where the flow is not strictly inward (i.e., along the x-axis).*

whose boundary is the level set

$$\partial\mathcal{K} = \{(x, y) \in \mathbb{R}^2 : E(x, y) = E_0\}. \tag{4.10}$$

Provided[3] that $E_0 > 0$, formula (4.10) defines a smooth closed curve (see Figure 4.1). Now ∇E is normal to the level set; since ∇E points in the direction of increasing E, the *inward* normal is given by $\mathbf{N} = -\nabla E$. Thus, we may rewrite (4.8) as

$$\langle \mathbf{N}_\mathbf{x}, \mathbf{F}(\mathbf{x}) \rangle \geq 0 \qquad (\mathbf{x} \in \partial\mathcal{K}).$$

In words, the direction of the flow of the ODE at $\partial\mathcal{K}$ is inward, or at worst tangential or zero.

As we now show, if an inequality of this type holds on the boundary of a region, then the solution of the IVP (for positive times) is trapped inside that region.

(b) Statement and discussion of the result

Consider the IVP for an ODE $\mathbf{x}' = \mathbf{F}(\mathbf{x})$, where $\mathbf{F}$ is defined on an open subset $\mathcal{U}$ of $\mathbb{R}^d$. Let $\mathcal{K}$ be a closed subset of $\mathcal{U}$ whose boundary is a $\mathcal{C}^1$ surface (see Section B.3.3 for definitions), and for $\mathbf{x} \in \partial\mathcal{K}$, let $\mathbf{N}_\mathbf{x}$ be an inward normal to $\partial\mathcal{K}$ at $\mathbf{x}$. We shall

[3]The topology of the set (4.9) changes if $E_0 < 0$. For simplicity, we sidestep this complication.

call $\mathcal{K}$ a *trapping region*[4] for $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ if

$$(\forall \mathbf{x} \in \partial\mathcal{K})\ \langle \mathbf{N}_{\mathbf{x}}, \mathbf{F}(\mathbf{x}) \rangle \geq 0. \tag{4.11}$$

Theorem 4.2.2. *Suppose that $\mathbf{F} : \mathcal{U} \to \mathbb{R}^d$ is $\mathcal{C}^1$ and that $\mathcal{K}$ is a compact trapping region for $\mathbf{x}' = \mathbf{F}(\mathbf{x})$. If the initial data $\mathbf{b}$ lies in the interior of $\mathcal{K}$, then the solution $\mathbf{x}$ to equation (4.1) exists for all positive time and moreover lies in the interior of $\mathcal{K}$.*

If the inequality (4.11) is replaced by strict inequality,

$$(\forall \mathbf{x} \in \partial\mathcal{K})\ \langle \mathbf{N}_{\mathbf{x}}, \mathbf{F}(\mathbf{x}) \rangle > 0, \tag{4.12}$$

only a few lines suffice[5] to prove this result:

Proof of Theorem 4.2.2 assuming (4.12). By Theorem 4.1.2, the solution may cease to exist only if it first leaves $\mathcal{K}$. If, coming from inside $\mathcal{K}$, the trajectory $\mathbf{x}(t)$ reaches a point on $\partial\mathcal{K}$, then at that point the (outward) normal velocity must be nonnegative, and this contradicts the trapping hypothesis (4.12). □

If we have only the weaker hypothesis (4.11), our proof must rule out the possibility illustrated in Figure 4.1(b) for Duffing's equation (4.6): could a trajectory escape tangentially from $\mathcal{K}$ at a point where the inner product (4.8) vanishes?[6] This proof is somewhat technical and does not contain fundamental new ideas.[7] You may safely postpone reading it, but since the full result is useful, we will feel free to invoke it below in studying specific examples.

(c) Proof of Theorem 4.2.2. By Theorem 4.1.2, the solution may cease to exist only if it first leaves $\mathcal{K}$. Let

$$t_* = \sup\{t : (\forall s \leq t)\ \mathbf{x}(s) \in \text{Int } \mathcal{K}\}. \tag{4.13}$$

---

[4]Some authors reserve the word "region" to describe an open set. Note that we are not following that convention here.

[5]If you feel that these remarks are too sketchy to constitute a real proof, we urge you to revisit them after reading the proof of the full result.

[6]For Duffing's equation, we can argue that such an escape is not possible, since the function $E(x, y)$ is nonincreasing along orbits. However, this argument uses the fact that $E(x, y)$ is defined in a neighborhood of $\partial\mathcal{K}$, while we want to prove the theorem using only information derived from the fact that (4.11) holds on $\partial\mathcal{K}$.

[7]Indeed, the argument is only a minor extension of what you already were asked to do in Exercise 1.16.

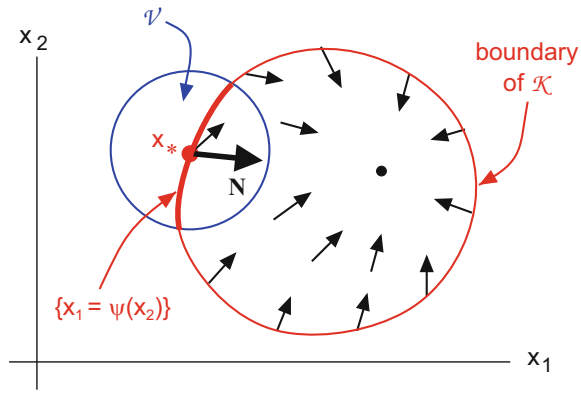

Figure 4.2: *Schematic of a trapping region $\mathcal{K}$ and the data of (4.14). Remark: Note that the vector field* **F** *has an equilibrium point inside the trapping region; in two dimensions, this cannot be avoided if $\mathcal{K}$ is simply connected.*

By the local existence theorem, $t_* > 0$. We suppose $t_* < \infty$ and look for a contradiction. Of course at the supremum (4.13), $\mathbf{x}(t_*) \in \partial\mathcal{K}$. For brevity we write $\mathbf{x}_* = \mathbf{x}(t_*)$.

Since $\partial\mathcal{K}$ is a $\mathcal{C}^1$ surface, near the hypothetical exit point $\mathbf{x}_*$, one of the coordinates, say x_1, may be expressed as a function of the others. That is, there exist a neighborhood $\mathcal{V} \subset \mathcal{U}$ of $\mathbf{x}_*$ and a function $\psi : \mathcal{V} \to \mathbb{R}$, independent of x_1, such that

$$\partial\mathcal{K} \cap \mathcal{V} = \{\mathbf{x} \in \mathcal{V} : x_1 = \psi(\tilde{\mathbf{x}})\}, \tag{4.14}$$

where $\tilde{\mathbf{x}}$ is shorthand for $(x_2, \ldots, x_d)$. Moreover, reversing signs of both x_1 and ψ if necessary, we may assume that $x_1 > \psi(\tilde{\mathbf{x}})$ on the interior of $\mathcal{K}$, or $\mathbf{N} = (1, -\tilde{\nabla}\psi)$ is an *inward* normal along $\partial\mathcal{K}$. (Cf. Figure 4.2.)

By continuity, there is an interval $[t_* - \delta, t_*]$ such that $\mathbf{x}(t) \in \mathcal{V}$ for t in this interval. Let

$$g(t) = x_1(t) - \psi(\tilde{\mathbf{x}}(t)), \qquad t_* - \delta \le t \le t_*. \tag{4.15}$$

Then $g(t) > 0$ for $t_* - \delta \le t < t_*$, while $g(t_*) = 0$. On the other hand, we claim that

$$g'(t) \ge -Kg(t), \qquad t_* - \delta \le t \le t_* \tag{4.16}$$

for some constant K. Given (4.16), it follows that $(d/dt)[e^{Kt}g] \ge 0$, so

$$e^{Kt_*}g(t_*) \ge e^{K(t_*-\delta)}g(t_* - \delta).$$

But $g(t_* - \delta) > 0$, and hence this inequality implies that $g(t_*) > 0$, which is a contradiction that will prove Theorem 4.2.2.

It remains to prove the claim, (4.16). Applying the chain rule to (4.15), we

calculate that $g'(t) = G(\mathbf{x}(t))$, where

$$G(\mathbf{x}) = F_1(\mathbf{x}) - \sum_{j=2}^{d} \frac{\partial \psi}{\partial x_j}(\tilde{\mathbf{x}}) F_j(\mathbf{x}). \tag{4.17}$$

This function need not be $\mathcal{C}^1$, because ψ may not have the requisite smoothness, but it is continuously differentiable *with respect to the first component*, x_1. Also, if $\mathbf{x} \in \partial\mathcal{K}$, then

$$G(\mathbf{x}) = \langle \mathbf{N}_{\mathbf{x}}, \mathbf{F}(\mathbf{x}) \rangle \geq 0. \tag{4.18}$$

To estimate G at a general point $(x_1, \tilde{\mathbf{x}})$ in $\mathcal{K} \cap \mathcal{V}$, we add and subtract G evaluated at a nearby point $(\psi(\tilde{\mathbf{x}}), \tilde{\mathbf{x}})$ on $\partial\mathcal{K}$:

$$G(x_1, \tilde{\mathbf{x}}) = [G(x_1, \tilde{\mathbf{x}}) - G(\psi(\tilde{\mathbf{x}}), \tilde{\mathbf{x}})] + G(\psi(\tilde{\mathbf{x}}), \tilde{\mathbf{x}}). \tag{4.19}$$

Invoking (4.18) to drop the second term, we have

$$G(x_1, \tilde{\mathbf{x}}) \geq G(x_1, \tilde{\mathbf{x}}) - G(\psi(\tilde{\mathbf{x}}), \tilde{\mathbf{x}}). \tag{4.20}$$

We apply the fundamental theorem of calculus to rewrite the RHS of (4.20) as

$$G(x_1, \tilde{\mathbf{x}}) - G(\psi(\tilde{\mathbf{x}}), \tilde{\mathbf{x}}) = \int_{\psi(\tilde{\mathbf{x}})}^{x_1} \frac{\partial G}{\partial x_1}(s, \tilde{\mathbf{x}})\, ds.$$

Since $\partial G / \partial x_1$ is continuous, by compactness it is bounded on $\mathcal{K} \cap \overline{\mathcal{V}}$, say by the constant K. Therefore,

$$G(x_1, \tilde{\mathbf{x}}) - G(\psi(\tilde{\mathbf{x}}), \tilde{\mathbf{x}}) \geq -K[x_1 - \psi(\tilde{\mathbf{x}})],$$

which we may substitute into the RHS of (4.20). Thus,

$$g'(t) = G(\mathbf{x}(t)) \geq -K[x_1(t) - \psi(\tilde{\mathbf{x}}(t))] = -Kg(t),$$

as claimed in (4.16). The proof of Theorem 4.2.2 is now complete. □

(d) Various generalizations of the theorem

Theorem 4.2.2 may be generalized to trapping regions whose boundary is only piecewise smooth, which provides a *much* more versatile tool. Here is such a result for two-dimensional problems, which we ask you to prove in Exercise 2. Note that the trapping condition (4.11) need not be explicitly imposed at corner points[8] of the boundary; information derived from continuity suffices to handle the corners.

[8]If this term is unclear, see Section B.3.3(b), where the distinction between regular points and corner points on the boundary is defined.

Theorem 4.2.3. *Suppose that* $\mathbf{F} : \mathcal{U} \to \mathbb{R}^2$ *is* $\mathcal{C}^1$ *on a domain* $\mathcal{U} \subset \mathbb{R}^2$ *and that* $\mathcal{K} \subset \mathcal{U}$ *is a compact region with a piecewise smooth boundary such that* (4.11) *holds at all regular points of* $\partial\mathcal{K}$. *If the initial data* $\mathbf{b}$ *lies in the interior of* $\mathcal{K}$, *then the solution* $\mathbf{x}$ *to equation* (4.1) *exists for all positive time and moreover lies in the interior of* $\mathcal{K}$.

Of course analogous results hold in dimensions higher than two, but it is rather tedious to deal carefully with all possible cases, and not much is learned in doing so. We do not formulate any such generalization.

With a minor revision of the proof of Theorem 4.2.2, you may show that even if a trapping region is not compact, the following conclusion still holds. (*Check this result! We will use it below.*)

Corollary 4.2.4. *Suppose that* $\mathbf{F} : \mathcal{U} \to \mathbb{R}^d$ *is* $\mathcal{C}^1$ *and that* $\mathcal{K}$ *is a trapping region for* $\mathbf{x}' = \mathbf{F}(\mathbf{x})$. *If the initial data* $\mathbf{b}$ *lies in the interior of* $\mathcal{K}$, *then the solution* $\mathbf{x}$ *to equation* (4.1) *lies in the interior of* $\mathcal{K}$ *for as long as this solution continues to exist.*

Two further generalizations: (i) If the initial data $\mathbf{b}$ of an IVP belongs to the *boundary* of a compact trapping region, global existence may still be deduced. (ii) Theorem 4.2.2 may be extended to nonautonomous equations. In practice, neither of these results turns out to be terribly useful, and we do not pursue them.

4.3 Level Sets and Trapping Regions

4.3.1 Introduction via Duffing's Equation

In Section 4.2.2(a), we observed that sublevel sets (4.9) of the energy function (4.7) are trapping regions for Duffing's equation (4.6). This fact provides an easy global existence proof for the IVP for this equation. For every E_0, the set (4.9) is a compact trapping region. Given initial conditions $\mathbf{b} \in \mathbb{R}^2$, choose E_0 large enough that $\mathbf{b} \in \mathcal{K}$. By invoking Theorem 4.2.2, we obtain existence for all positive time.

For many other ODEs as well, level sets of some auxiliary function(s) may be used to construct trapping regions; we present two such examples.

4.3.2 The Chemostat

The chemostat is described by the scaled ODEs

$$\begin{aligned} x' &= \frac{y}{y+1}x - \rho x, \\ y' &= -\frac{y}{y+1}x - \rho(y-\sigma), \end{aligned} \tag{4.21}$$

where ρ, σ are positive constants. The variables x and y are concentrations, so they are nonnegative. The linear terms $-\rho x$ and $-\rho y$ represent decay, and the constant term $\rho\sigma$ in the second equation represents replenishment of y. The assumptions leading to the nonlinear terms will be explained in Section 5.5. Even without understanding the basis for these terms, we can see that they tend to increase x at the expense of y. Since these terms differ only by a minus sign, we may add the equations and deduce a third ODE

$$x' + y' = -\rho(x + y) + \rho\sigma \tag{4.22}$$

that will be useful in analyzing global existence.

Let's seek a triangular trapping region of the form

$$\mathcal{K} = \{(x, y) \in \mathbb{R}^2 : x \geq 0,\ y \geq 0,\ x + y \leq A\}, \tag{4.23}$$

where A is a constant. Along the sloping face of $\partial\mathcal{K}$, the inward normal is given by $\mathbf{N} = (-1, -1)$, and we observe from (4.22) that $\langle \mathbf{N}, \mathbf{F} \rangle = -x' - y' = \rho(A - \sigma)$. In particular, provided $A \geq \sigma$, we have $\langle \mathbf{N}, \mathbf{F} \rangle \geq 0$ along this face, which shows that flow is inward here. Regarding the other two sides, you may check that the flow of (4.21) is inward along the x-axis and is tangential along the y-axis. Therefore, $\mathcal{K}$ is a compact trapping region.

For every initial condition $\mathbf{b}$ in the first quadrant, the constant A in (4.23) may be chosen large enough that $\mathbf{b}$ belongs to the trapping region $\mathcal{K}$. Therefore, we may invoke Theorem 4.2.3 to obtain global existence for the IVP for (4.21).

The lesson to take away from this example is that we have used level sets of the linear function $L(x, y) = x + y$ in constructing trapping regions for (4.21).

4.3.3 The Torqued Pendulum and ODEs on Manifolds[9]

Consider a pendulum, as illustrated in Figure 4.3, that is subjected to a "torque" μ, which tends to twist the unperturbed pendulum away from its stable, straight-down equilibrium. If friction is modeled by linear damping, then after appropriate scaling,

[9]Although we use the general term "manifold" here, in fact we need only a couple of special cases (like the circle S^1) with which you are probably already familiar. A manifold is a topological space in which each point has a neighborhood homeomorphic to a ball in Euclidean space, subject to some compatibility conditions. If you want precise definitions, you may find these in Section 2.7 of [63] or look online, but we expect that most readers will not need to consult other sources to read the present section.

this problem may be described by the first-order system[10]

$$\begin{aligned} x' &= y, \\ y' &= -\sin x - \beta y + \mu, \end{aligned} \tag{4.24}$$

where $\beta > 0$ and μ are constants. Without loss of generality, it suffices to consider the case $\mu \geq 0$. *(Why?)*

We seek a trapping region derived from the total energy, kinetic plus potential, of the pendulum, which is given by

$$E(x, y) = y^2/2 - \cos x. \tag{4.25}$$

Thus, given a constant E_0, we consider the sublevel set

$$\mathcal{K} = \{(x, y) \in \mathbb{R}^2 : \ y^2/2 - \cos x \leq E_0\}. \tag{4.26}$$

Substitution into (4.24) yields

$$\frac{dE}{dt} = \langle \nabla E, \mathbf{F} \rangle = -\beta y^2 + \mu y.$$

In general, dE/dt may have either sign, but if $|y| > \mu/\beta$, then $dE/dt = -\beta y(y - \mu/\beta)$ is negative. At the boundary of (4.26), $y = \pm\sqrt{2(E_0 + \cos x)}$, so $|y| > \mu/\beta$, provided E_0 is sufficiently large. Doing the calculation, we see that (4.26) is a trapping region for (4.24), provided $E_0 \geq (\mu/\beta)^2/2 + 1$. Increasing E_0 if necessary, we may also assume that $\mathcal{K}$ contains any proposed initial data.

However, our attempt to apply Theorem 4.2.2 is thwarted by the fact that $\mathcal{K}$ is not compact (see Figure 4.4a). In principle, a solution of (4.24) might stay inside $\mathcal{K}$ while x marches off to infinity in finite time. In fact, this does not happen: the RHS of (4.24) satisfies the hypothesis (4.4) of Theorem 4.2.1, so the solution exists for all $t \in \mathbb{R}$.

The trapping-region existence proof may be revived with the addition of some geometry. Note that the RHS of (4.24) is 2π-periodic in x. Therefore, rather than considering (4.24) as an ODE on the Euclidean space $\mathbb{R} \times \mathbb{R}$, we may regard it as an ODE on the cylinder[11] $S^1 \times \mathbb{R}$, where $S^1 = \mathbb{R}/2\pi\mathbb{Z}$ is the circle. If $\mathcal{K}$ is considered a subset of $S^1 \times \mathbb{R}$, then this set *is* compact (see Figure 4.4b). Moreover, as you will show in Exercise 6, Theorem 4.2.2 generalizes to trapping regions on $S^1 \times \mathbb{R}$, and thus we may obtain a different, in our view more elegant, proof of global existence in forward time for (4.24). As an added bonus, using this approach, you will find

[10]Incidentally, the equations (4.24) also describe the behavior of an electrical device known as the *Josephson junction*. See Strogatz [81], Section 4.6, for details.

[11]You have already encountered an equation on a cylinder: in using polar coordinates to study an ODE in the plane, you obtain an ODE in which r, θ belong to the manifold $(0, \infty) \times S^1$.

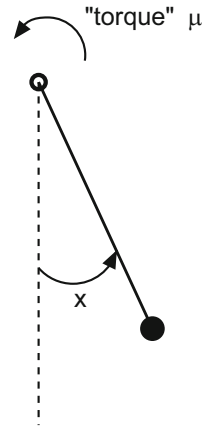

Figure 4.3: *Schematic diagram of the torqued pendulum.*

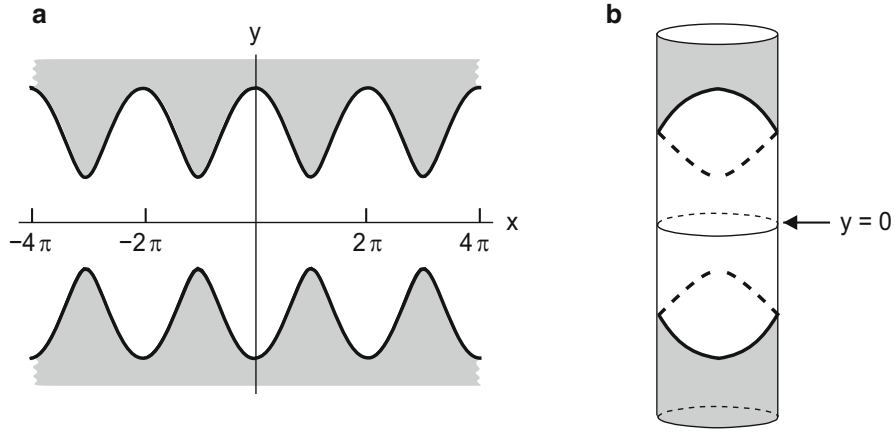

Figure 4.4: *Sketch of the trapping region $\mathcal{K}$ for the torqued pendulum, (4.26). In Panel (a), $\mathcal{K}$ is the unbounded region that lies between the two curves. In Panel (b), the x, y-plane is wrapped around a cylinder, so that values of x differing by 2π are identified.*

that the solution grows at worst linearly as $t \to \infty$. This conclusion is stronger than what can be deduced using Theorem 4.2.1.

In this book we do not attempt to extend Theorem 4.2.2 to ODEs on general manifolds, even though given the right technical background, this is not difficult. Rather, we consider ODEs only on two specific manifolds, the cylinder $S^1 \times \mathbb{R}$ and the torus $\mathbb{T}^2 = S^1 \times S^1$. For these two special cases, the generalization of Theorem 4.2.2 may be proved with ad hoc arguments using (multivalued) Euclidean coordinates, as in Exercise 6.

4.4 Nullclines and Trapping Regions

4.4.1 Nullclines in the Chemostat

The term *nullcline* refers to a curve[12] where one of the components of the velocity vector vanishes. For example, consider the chemostat (4.21). We have that x' vanishes if

$$x = 0 \qquad \text{or} \qquad y = \frac{\rho}{1-\rho}, \tag{4.27}$$

while y' vanishes if

$$x = -\rho \, \frac{(y-\sigma)(y+1)}{y}. \tag{4.28}$$

The portions of both curves that lie in the first quadrant are graphed in Figure 4.5, in brown for the x-nullclines (4.27) and in cyan[13] for the y-nullcline (4.28). Intersections of the nullclines at

$$(x, y) = (0, \sigma) \quad \text{and} \quad \left(\sigma - \frac{\rho}{1-\rho}, \frac{\rho}{1-\rho} \right) \tag{4.29}$$

are equilibrium solutions of the ODE. In the figure we show the more interesting case, for which the second equilibrium lies in the first quadrant, i.e., parameters such that $\rho < 1$ and $\sigma > \rho/(1-\rho)$.

To extract information from the nullclines with minimal pain, it is helpful to proceed in the three stages represented in Figure 4.5. Whenever you need to graph nullclines, *we urge you to follow this three-stage procedure*. Being systematic in this way reduces (slightly) the opportunities for making careless mistakes, of which there are plenty.

---

[12] "Curve" is the appropriate word for two-dimensional systems, which is the usual context in which nullclines are studied. For higher-dimensional systems, one should say the "set where ..." or "surface where ...."

[13] Whenever we plot nullclines, we will follow this color convention.

- In Figure 4.5(a), vertical lines are drawn along the x-nullclines (4.27) because the flow is vertical there, i.e., $x' = 0$. Similarly, horizontal lines have been drawn along the y-nullcline (4.28).

- Figure 4.5(b) augments the previous figure by specifying along the x-nullclines whether the flow is up or down; and along the y-nullcline whether the flow is to the right or left. Here is the thinking behind the construction of this figure. The orientation of the flow along (4.27) changes from up to down[14] whenever this curve crosses a nullcline of the other family. Alternatively put, on every segment of (4.27) that does *not* intersect (4.28), the orientation of the flow does *not* change. It may be seen from (4.21b) that y' is positive at the origin. As shown in the figure, the flow remains upward as one moves away from the origin until points where (4.28) is crossed, causing the direction to reverse itself. Similarly, for the other nullcline, start by observing from (4.21a) that far out on (4.28), near the x-axis, x' is negative. Then the other horizontal arrows along (4.28) in Figure 4.5(b) may be constructed by reversing direction whenever (4.27) is crossed.

- The nullclines partition the first quadrant into regions. Within one region the flow $\mathbf{F}(x, y)$ points into one of the four quadrants, $\{\pm x > 0, \pm y > 0\}$, and the quadrant remains the same if (x, y) moves within this region. The quadrant of the flow in each of the regions is indicated by a thick black arrow in Figure 4.5(c). Again, one can complete this figure by analyzing a special case (e.g., along the x-axis, the flow points into the second quadrant) and making appropriate reversals on crossing nullclines.

We will call the completed Figure 4.5(c) a *flow-quadrant* diagram.

Especially for two-dimensional ODEs, nullclines are an invaluable aid in sketching trajectories. We shall see that this technique is most effective when used in conjunction with information from Chapter 6 about flow near equilibria. For now we turn our attention to using nullclines to construct trapping regions.

4.4.2 An Activator–Inhibitor System

In this example x and y evolve according to the ODEs

$$\begin{aligned} (a) \quad x' &= \sigma \frac{1}{1+y} \frac{x^2}{1+x^2/\kappa^2} - x, \\ (b) \quad y' &= \rho \left[\frac{x^2}{1+x^2/\kappa^2} - y \right], \end{aligned} \tag{4.30}$$

[14] Well, *generically* one expects the orientation of the flow arrows to change, but here is a cautionary example: $x' = y^2$, $y' = -x$. Along the y-nullcline (the y-axis), all arrows are oriented in the direction of increasing x. Of course, this example is concocted with nongeneric behavior in mind: Generically, polynomials change sign at every root, whereas y^2 does not.

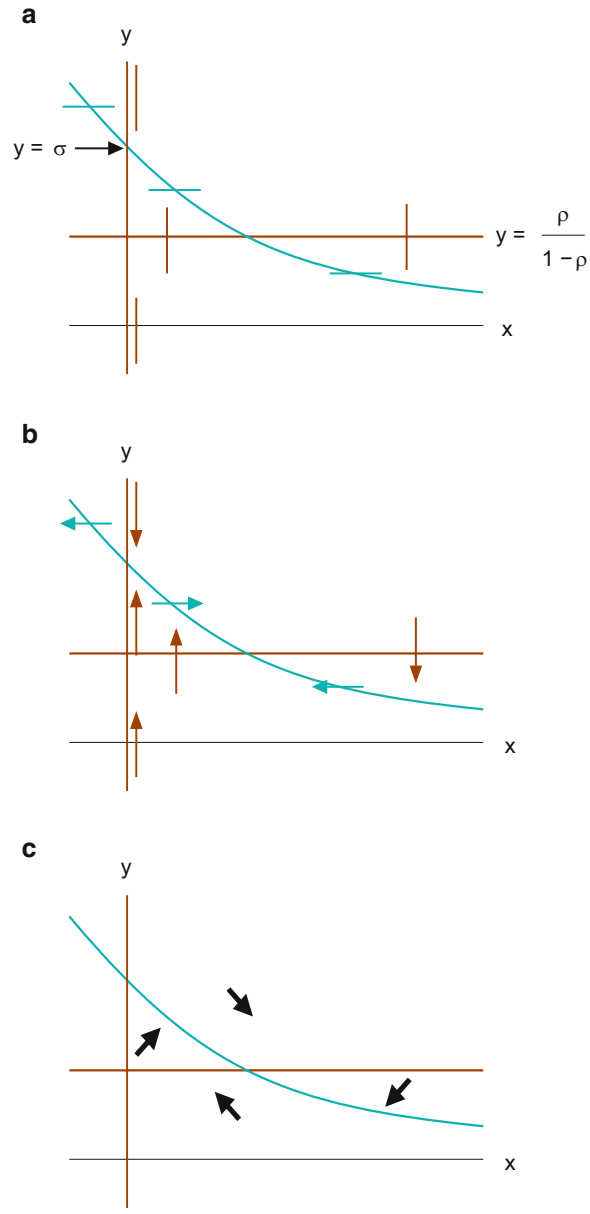

Figure 4.5: *Nullclines for the chemostat equations* (4.21) *with* $\rho = 1/2$ *and* $\sigma = 2$. *See text for a detailed description of the panels and the color conventions.*

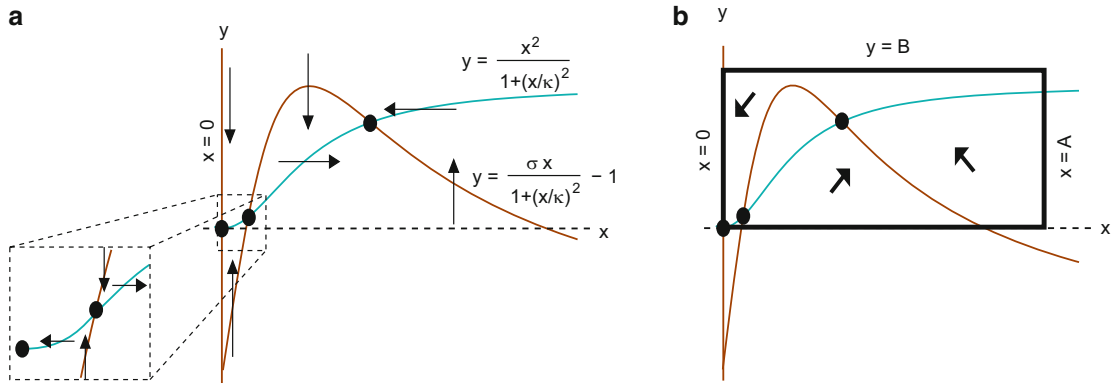

Figure 4.6: *(a) Nullclines for the activator–inhibitor equations* (4.30) *with* $\sigma = 4$ *and* $\kappa = 1$*, in which case there are three equilibria (bold dots). The inset shows a blowup of a region near the origin that contains two of the three equilibria. (b) A rectangular trapping region in the first quadrant, with flow quadrants indicated.*

where σ, ρ, κ are positive parameters; κ is often quite large. The variables x and y, which represent concentrations, are nonnegative. Linear terms in each equation describe decay. The physical basis for the nonlinear terms will be discussed in Section 5.6.

Panel (a) in Figure 4.6 shows the nullclines of (4.30) *(Check them!)*, and Panel (b) superimposes a rectangular region

$$\mathcal{K} = \{(x, y) : 0 \le x \le A,\ 0 \le y \le B\} \tag{4.31}$$

on the flow-quadrant diagram. Just from the figure, we may deduce that $\mathcal{K}$ is a trapping region. We see from Panel (a) that the flow is tangential along its left side, the y-axis, because this is an x-nullcline; and we see from the flow quadrant vectors in Panel (b) that the flow is inward along the other three sides, *provided* A and B are appropriately large. (You need to fill in the flow quadrant for the underresolved triangular region near the origin.) It is good analytical practice to determine explicit estimates for A, B to ensure that the flow is inward.

To derive global existence: given initial data $x(0) = a,\ y(0) = b$, where (a, b) lies in the first quadrant, choose A, B large enough that $\mathcal{K}$ is a trapping region and (a, b) belongs to $\mathcal{K}$, and then apply Theorem 4.2.3.

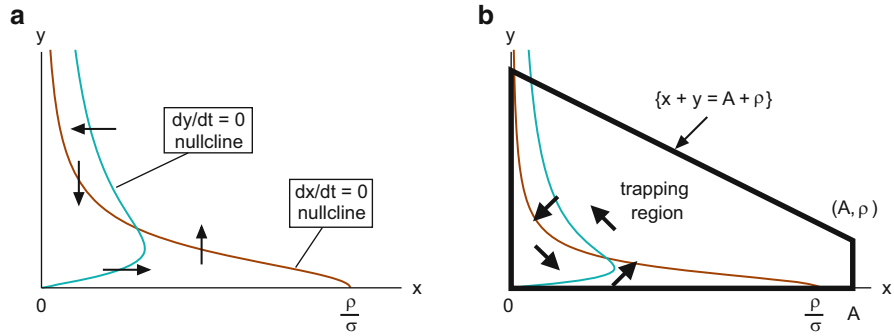

Figure 4.7: *(a) Nullclines of the Sel'kov model (4.32). (b) A trapezoidal trapping region. The diagonal line segment has slope -1 and connects the y-axis to the point (A, ρ), where $A > \rho/\sigma$. Along this diagonal line it is not apparent from the flow quadrant diagram that the flow is inward; a calculation is needed for this.*

4.4.3 Sel'kov's Model for Glycolysis

The Sel'kov model, which will be introduced in Section 5.1.3, is given by the equations

$$\begin{aligned} x' &= \rho - \sigma x - xy^2, \\ y' &= -y + \sigma x + xy^2, \end{aligned} \tag{4.32}$$

where ρ, σ are positive constants. The variables x and y are concentrations and must be nonnegative. As in the chemostat (4.21), the sole nonlinear terms in the two equations are equal in magnitude and opposite in sign. Imitating that example, let us seek a triangular trapping region like (4.23). The flow is inward along both coordinate axes. Regarding the sloping side of (4.23), we add the equations to derive

$$x' + y' = \rho - y.$$

Unfortunately, the flow along this side is inward only if $y \geq \rho$. Thus, no region of the form (4.23) will trap solutions of (4.32).

Nullclines provide an easy fix for this minor difficulty. Nullclines for (4.32) are shown in Panel (a) of Figure 4.7, and in Panel (b) a trapezoidal region is superimposed on the flow-quadrant diagram. Regarding the right side of the trapezoid, we require that $A > \rho/\sigma$. Then, because the x-nullcline of (4.32) crosses the x-axis at $x = \rho/\sigma < A$, we see from the flow-quadrant diagram that the flow is inward along this side. We already know that the flow is inward along the other three sides. Hence every trapezoid with A sufficiently large is a trapping region, which may be used to prove global existence.

To conclude, we have found trapping regions with a combination of level sets and nullclines.

4.4.4 Van der Pol's Equation

Next we consider the van der Pol system,

$$\begin{array}{lrcl} (a) & x' & = & y, \\ (b) & y' & = & -\beta(x^2-1)y - x. \end{array} \tag{4.33}$$

Recall that the rate of change of the energy-like function $E(x,y) = (x^2+y^2)/2$ along a trajectory is given by

$$\frac{dE}{dt} = -\beta(x^2-1)y^2.$$

Thus, the flow is inward along *most* of a large (circular) level set $\{E(x,y) = E_0\}$, but not within the vertical strip $\{-1 < x < 1\}$.

To handle this difficulty, we modify a sublevel set $\{E(x,y) \le E_0\}$, as indicated in Panel (b) of Figure 4.8, by deleting slices PQRP from the top and STUS from the bottom of this disk. Taking advantage of the fact that the flow (4.33) is odd under the reflection $(x,y) \mapsto (-x,-y)$, we construct the region to be invariant under this reflection; thus, we need specify only PQR, the boundary of the upper deletion. Let QR be the line segment given by the equation

$$y = A + 2\beta x, \qquad -2 \le x \le 2, \tag{4.34}$$

where A is a large constant to be chosen below. Once A is chosen, the region is completely specified, as follows: The point R has coordinates $(2, A+4\beta)$, which determine the radius $\sqrt{2E_0}$ of the circle; i.e.,

$$2E_0 = 2^2 + (A+4\beta)^2.$$

The horizontal line starting at Q, along which $y = A - 4\beta$, meets the circle at P, which has coordinates $(-X, A-4\beta)$, where

$$(-X)^2 + (A-4\beta)^2 = 2E_0.$$

Finally, S, T, and U are located by symmetry.

We claim that provided A is sufficiently large, PQRSTUP is a trapping region. By symmetry we need to show that the flow is inward only along half of the boundary, say along PQRS. (i) Along the circular arc RS, we already know that the flow is inward. (ii) Regarding QR, by dividing the two equations in (4.33), we calculate that the flow direction has slope

$$\frac{dy}{dx} = \beta(1-x^2) - \frac{x}{y} \le \beta - \frac{x}{y}. \tag{4.35}$$

Choose A large enough that along QR, the second term on the RHS of (4.35) satisfies $|x/y| \le \beta$; then the flow direction along QR has slope $dy/dx \le 2\beta$. But QR has

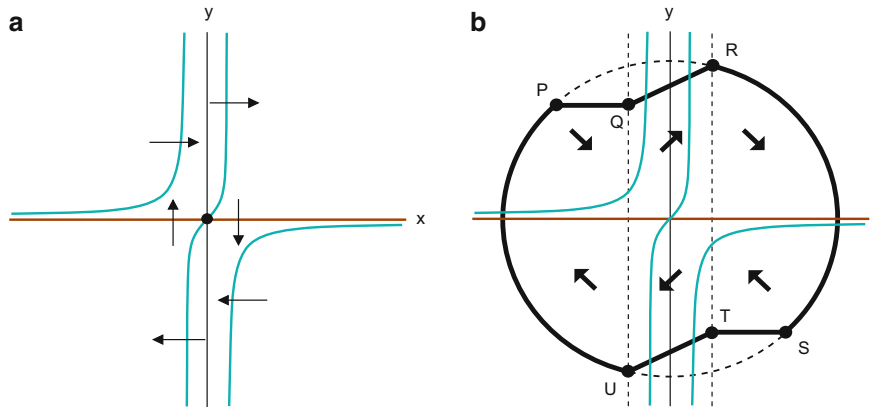

Figure 4.8: *(a) Nullclines for the van der Pol system (4.33). The y-nullclines have vertical asymptotes at $x = \pm 1$. (b) A trapping region and flow-quadrant diagram as described in the text. The dashed vertical lines are located at $x = \pm 2$.*

slope exactly 2β, so for such large A, the flow is inward along QR. (iii) Along PQ, we see from the flow-quadrant diagram Figure 4.8(b) that the flow is inward, which finishes the proof of the claim.

Using these trapping regions, we derive global existence for van der Pol's equation. As in the previous example, the successful strategy combined level sets and nullclines.

4.4.5 Michaelis–Menten Kinetics

The above examples provide an adequate introduction to nullclines and their usefulness in finding trapping regions, and more examples are given in the exercises. However, we present one more example because of its scientific interest.

Michaelis–Menten kinetics arises in modeling the concentrations (thus $x, y \geq 0$) of certain chemical species in an enzyme-mediated reaction (see Section 5.7). Applying suitable scaling, the equations take the form

$$\begin{array}{lrcl} (a) & x' & = & -x(1-y)+y, \\ (b) & \varepsilon\, y' & = & x(1-y)-(1+\kappa)y, \end{array} \tag{4.36}$$

where ε, κ are positive parameters. Typically ε is very small indeed. Reflecting this, we shall call (4.36) a *fast–slow* system, because at least away from the y-nullcline, the second variable evolves much more rapidly than x.

Global existence for (4.36) is easily demonstrated. By adding the equations, one sees that the derivative of $x + \varepsilon y$ is negative. Any triangular region bounded by the x-axis, the y-axis, and a line $\{x + \varepsilon y = A\}$, where $A > 0$, can serve as a trapping region.

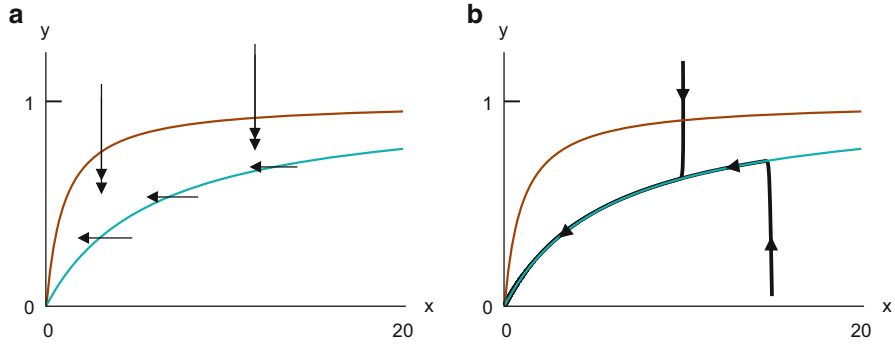

Figure 4.9: *(a) Nullclines for the scaled Michaelis–Menten equations* (4.36) *form a trapping region* ($\kappa = 5$ *in this figure). We use double arrowheads to indicate fast flow in the vertical direction (for small* ε*); if we attempted to represent lengths accurately, these vectors would be absurdly long. (b) Two trajectories for* (4.36)*, starting outside the trapping region. During a brief initial transient, motion is nearly vertical, after which trajectories hug the y-nullcline while both variables decay to zero.*

However, the region between nullclines

$$y = \frac{x}{x+1} \qquad \text{and} \qquad y = \frac{x}{x+1+\kappa}$$

is a much more interesting trapping region (see Figure 4.9(a)). The fast equation (4.36b) drives (x, y) into this trapping region, after which both variables tend slowly to zero *while staying inside the trapping region*, as sketched in Figure 4.9(b). Since ε is small, the solution hugs the y-nullcline, the lower boundary of the trapping region.

In this and other fast–slow systems, it is natural to consider the approximation[15] of setting $\varepsilon = 0$ in (4.36b). In this approximation, we may then solve (4.36b), *now an algebraic equation*, to obtain $y = x/(x+1+\kappa)$; substituting into (4.36a), we derive

$$\frac{dx}{dt} = -\frac{\kappa x}{x+1+\kappa}. \tag{4.37}$$

This equation is the (scaled) Michaelis–Menten approximation for the enzymatic reaction rate arising from (4.36).

We hope that you are worried about the violent approximation from which (4.37) is derived; indeed, the approximation changes the differential equation (4.36b) to

[15]Chemists call this approximation "letting the fast reaction go to completion" or the "quasi-steady-state" assumption. Modifying this language, we shall speak of "letting the fast equation go to equilibrium."

an algebraic equation! However, *the above argument with nullclines supports the approximation*. It indicates that after a brief initial transient, the exact solution follows the approximation rather closely.

4.5 Continuity Properties of the Solution

Your authors are fond of the "proof through pictures" style of the last two sections. We will encounter a lot more of this style of mathematics in later chapters, but for now it's back to hard analysis without the relief provided by pictures.

4.5.1 The Main Issue: Continuous Dependence on Initial Conditions

Theorem 4.5.1. *Suppose* $\mathbf{F} : \mathcal{U} \to \mathbb{R}^d$ *is locally Lipschitz, and let* $\mathbf{x}_0(t)$, $0 \le t < \beta_0$, *be a solution in forward time of* $\mathbf{x}_0' = \mathbf{F}(\mathbf{x}_0)$ *with initial condition* $\mathbf{x}_0(0) = \mathbf{b}_0$. *(i) For every positive* $T < \beta_0$, *there is a neighborhood* $\mathcal{V}$ *of* $\mathbf{b}_0$ *such that if* $\mathbf{b} \in \mathcal{V}$, *the IVP*

$$\mathbf{x}' = \mathbf{F}(\mathbf{x}), \qquad \mathbf{x}(0) = \mathbf{b} \tag{4.38}$$

has a solution for $0 \le t < T$. *(ii) Moreover, there is a constant* L *such that for all* $\mathbf{b} \in \mathcal{V}$,

$$|\mathbf{x}(t) - \mathbf{x}_0(t)| \le |\mathbf{b} - \mathbf{b}_0|\, e^{Lt}, \qquad 0 \le t < T. \tag{4.39}$$

In words, Conclusion (ii) asserts that if the initial data for an IVP are altered slightly, then the perturbed solution diverges from the original solution no faster than at a controlled exponential rate. Conclusion (i), which guarantees that the perturbed solution exists for nearly as long as $\mathbf{x}_0$, gives the estimate more significance.

Incidentally, $\beta_0 = \infty$ is allowed in Theorem 4.5.1, but the condition $T < \beta_0$ means that T must be finite. Such issues are clarified by examining the theorem in the context of an example, the scalar IVP

$$x' = x^3, \quad x(0) = b.$$

The solution $x_0(t) \equiv 0$ with initial condition $b = b_0 = 0$ exists for all $t \ge 0$, but for every $b \ne 0$, the IVP is solvable only for a finite interval. If T in the theorem is increased, the neighborhood $\mathcal{V}$ must be shrunk in compensation. An example in the reverse direction: it is possible for $\mathbf{x}_0$ to blow up in finite time while most nearby solutions exist for infinite times (see Exercise 4(c)).

Proof of Theorem 4.5.1. Let $\mathcal{K}_0$ be the image of $[0, T]$ under $\mathbf{x}_0$, which is a compact subset of $\mathcal{U}$. By Corollary 3.3.3, there exist a larger compact subset $\mathcal{K} \subset \mathcal{U}$ and a $\delta > 0$ such that

$$(\forall t \le T) \quad \overline{B(\mathbf{x}_0(t), \delta)} \subset \mathcal{K}. \tag{4.40}$$

By Proposition 3.3.2, $\mathbf{F}|\mathcal{K}$ is Lipschitz continuous, say with Lipschitz constant L. It is technically convenient to assume[16] $L > 0$, so that $e^{-L\varepsilon} < 1$ for every $\varepsilon > 0$.

Let $\mathcal{V}$ be the ball $B(\mathbf{b}_0, e^{-LT}\delta)$. If $\mathbf{b} \in \mathcal{V}$, let $\mathbf{x}$ be the solution in forward time of (4.38), extended to the maximal interval $0 \le t < \beta$, and define

$$t_* = \sup\{t \in [0,T] \cap [0,\beta) \,:\, (\forall s \le t)\, |\mathbf{x}(s) - \mathbf{x}_0(s)| \le \delta\}. \tag{4.41}$$

Since $|\mathbf{x}(0) - \mathbf{x}_0(0)| \le e^{-LT}\delta < \delta$, we know that $t_* > 0$. On the other hand, it follows from (4.41) that $\mathbf{x}(t)$ remains in $\mathcal{K}$ for $0 \le t < t_*$, so by Theorem 4.1.2 we have $t_* < \beta$. To derive Conclusion (i) we will show that $t_* = T$.

Let $g(t) = |\mathbf{x}(t) - \mathbf{x}_0(t)|$. From subtracting the integral equations for $\mathbf{x}$ and $\mathbf{x}_0$, we deduce that

$$g(t) \le |\mathbf{b} - \mathbf{b}_0| + \int_0^t |\mathbf{F}(\mathbf{x}(s)) - \mathbf{F}(\mathbf{x}_0(s))|\, ds. \tag{4.42}$$

If $t \le t_*$, then both $\mathbf{x}(s)$ and $\mathbf{x}_0(s)$ in the integrand belong to $\overline{B(\mathbf{x}_0(s), \delta)} \subset \mathcal{K}$, so Lipschitz continuity gives us

$$|\mathbf{F}(\mathbf{x}(s)) - \mathbf{F}(\mathbf{x}_0(s))| \le L|\mathbf{x}(s) - \mathbf{x}_0(s)|. \tag{4.43}$$

Substituting into (4.42), we have

$$g(t) \le |\mathbf{b} - \mathbf{b}_0| + L\int_0^t g(s)\, ds,$$

and hence by Gronwall's lemma,

$$g(t) \le |\mathbf{b} - \mathbf{b}_0| e^{Lt}, \qquad 0 \le t \le t_*. \tag{4.44}$$

To complete the proof we show that $t_* = T$, and thus (4.44) gives us (4.39). We see from the definition (4.41) that $t_* \le T$. But we have from (4.44) that

$$g(t_*) \le (e^{-LT}\,\delta)\, e^{Lt_*} = e^{-L(T-t_*)}\delta.$$

If t_* were strictly less than T, then we would have $g(t_*) < \delta$. By continuity, $g(t) = |\mathbf{x}(t) - \mathbf{x}_0(t)|$ would remain less than δ for some interval beyond t_*, and this would contradict the definition of t_* as a supremum. □

For use below let us formulate a refinement of this result under the stronger hypothesis that $\mathbf{F} \in \mathcal{C}^1(\mathcal{U})$. Given $\mathbf{x}_0(t)$ as in the theorem and $T < \beta_0$, choose $\delta > 0$

[16] Rigor can be such a pain! We are guarding against a triviality, since L could vanish only if $\mathbf{F}(\mathbf{x})$ were constant.

and a compact set $\mathcal{K} \subset \mathcal{U}$ for which (4.40) holds, and let

$$L = \max_{\mathbf{x}\in\mathcal{K}} \|\mathbf{DF}(\mathbf{x})\|. \tag{4.45}$$

Corollary 4.5.2. *Under these hypotheses, if* $\mathbf{b} \in B(\mathbf{b}_0, e^{-LT}\delta)$*, the IVP* (4.38) *is solvable for* $0 \le t < T + \eta$*, where* $\eta > 0$*,*

$$\mathbf{x}(t) \in \overline{B(\mathbf{x}_0(t), \delta)} \subset \mathcal{K}, \qquad 0 \le t \le T, \tag{4.46}$$

and (4.39) *holds with the constant* (4.45)*.*

Proof. A problem in adapting the proof of Theorem 4.5.1 to the present case: we cannot assume via Corollary 3.2.4 that (4.45) is a Lipschitz constant for $\mathcal{K}$, since this set is not necessarily convex. The saving point is that L need not be a Lipschitz constant for the entire set $\mathcal{K}$; it suffices if (4.43) is satisfied for all $s \in [0, T]$. By Corollary 3.2.4, given a value of s , (4.43) is satisfied for that s if

$$L \ge \max_{\mathbf{x}\in\overline{B(\mathbf{x}_0(s),\delta)}} \|\mathbf{DF}(\mathbf{x})\|,$$

and it is satisfied for all s if

$$L \ge \max_{0\le s\le T} \max_{\mathbf{x}\in\overline{B(\mathbf{x}_0(s),\delta)}} \|\mathbf{DF}(\mathbf{x})\|.$$

This inequality holds for the constant (4.45). □

Of course, the above results have analogues in backward time, which we invite you to formulate. Less trivially, the solution of the IVP for a nonautonomous equation $\mathbf{x}' = \mathbf{G}(\mathbf{x}, t)$ depends continuously on the initial condition, but we do not pursue this generalization.

4.5.2 Some Associated Formalism

Sometimes, when it is instructive to focus on how the solution of an IVP depends on the initial data, we shall use the flow notation. Specifically, we shall write

$$\boldsymbol{\varphi}(t, \mathbf{b}) = \mathbf{x}(t) \tag{4.47}$$

for the solution $\mathbf{x}(t)$ of the IVP (4.38). This *solution operator* or *flow function* is a mapping $\boldsymbol{\varphi} : \Omega \to \mathcal{U}$, where its domain is given by

$$\Omega = \{(t, \mathbf{b}) \in (-\infty, \infty) \times \mathcal{U} \ : \ t \in \text{ maximal interval of existence for (4.38)}\}. \tag{4.48}$$

It follows from Theorem 4.5.1 that $\boldsymbol{\varphi}$ is locally Lipschitz with respect to its second argument, $\mathbf{b}$, provided $\mathbf{F}(\mathbf{x})$ is locally Lipschitz. Trivially, $\boldsymbol{\varphi}$ is in fact locally Lipschitz with respect to both arguments simultaneously. *(Why is this trivial?)*

The solution operator[17] satisfies the following relation, which is known as the *semigroup property*:

Proposition 4.5.3. *If* $(s, \mathbf{b}) \in \Omega$ *and if* $(t, \boldsymbol{\varphi}(s, \mathbf{b})) \in \Omega$, *then* $(s + t, \mathbf{b}) \in \Omega$ *and*

$$\boldsymbol{\varphi}(t, \boldsymbol{\varphi}(s, \mathbf{b})) = \boldsymbol{\varphi}(s + t, \mathbf{b}). \tag{4.49}$$

This result follows easily from Lemma 3.2.9. (*Show this!*)

4.5.3 Continuity with Respect to Parameters

In Theorem 4.5.1 we proved that the solution of an IVP depends continuously on its initial data. It is also true that the solution "depends continuously on the equation." The most straightforward version of such a result addresses how the solution of a parametrized family of IVPs depends on the parameters. For simplicity, in the following theorem we address this issue only in the case of linear equations, which suffices for our needs below. Thus, suppose $A(t, \alpha_1, \ldots, \alpha_m)$ is an m-parameter family of $d \times d$ matrices, defined for $t \in (T_1, T_2)$, where this interval contains zero and, using an obvious vector notation, for $\boldsymbol{\alpha} \in \mathcal{V}$, where $\mathcal{V} \subset \mathbb{R}^m$ is open. Let $\mathbf{w}(t, \boldsymbol{\alpha})$ be the solution of

$$\mathbf{w}' = A(t, \boldsymbol{\alpha})\mathbf{w}, \qquad \mathbf{w}(0, \boldsymbol{\alpha}) = \mathbf{b}. \tag{4.50}$$

Theorem 4.5.4. *If* $A(t, \boldsymbol{\alpha})$ *is continuous on* $(T_1, T_2) \times \mathcal{V}$, *then* $\mathbf{w}(t, \boldsymbol{\alpha})$ *is continuous on this set.*

In Exercise 7 we provide hints to help you prove this result.

4.6 Differentiability Properties of the Solution

4.6.1 Dependence on Initial Conditions

In the previous section we proved that the flow $\boldsymbol{\varphi}(t, \mathbf{b})$ is Lipschitz continuous in $\mathbf{b}$. In this section we show that $\boldsymbol{\varphi}$ is in fact $\mathcal{C}^1$, provided of course that $\mathbf{F}$ is $\mathcal{C}^1$.

The above phrasing is a concise summary of the results of this section. However, let us restate the conclusion in the more discursive language of perturbation theory, since we believe that this makes the discussion more intuitive. We suppose $\mathbf{x}_0(t)$, $0 \leq t < \beta_0$, is a solution in forward time of $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ with initial condition $\mathbf{x}_0(0) = \mathbf{b}_0$,

[17] A function $\boldsymbol{\varphi}$ satisfying (4.49) is sometimes called a *dynamical system*.

and we ask how the solution changes if the initial condition is perturbed. Specifically, let $\mathbf{x}(t, \varepsilon)$ be the solution of

$$\mathbf{x}' = \mathbf{F}(\mathbf{x}), \qquad \mathbf{x}(0, \varepsilon) = \mathbf{b}_0 + \varepsilon \mathbf{b}_1. \tag{4.51}$$

We look for an expansion[18] of this solution in powers of ε,

$$\mathbf{x}(t, \varepsilon) = \mathbf{x}_0(t) + \varepsilon \mathbf{x}_1(t) + \dots \ . \tag{4.52}$$

The size of the neglected terms, which are represented by the dots, will be estimated in Theorem 4.6.1. For the moment we proceed formally. Substituting (4.52) into (4.51), we obtain

$$\mathbf{x}_0'(t) + \varepsilon \mathbf{x}_1'(t) + \dots = \mathbf{F}(\mathbf{x}_0(t) + \varepsilon \mathbf{x}_1(t) + \dots).$$

Using a Taylor series to expand[19] the RHS of this equation in powers of ε, we calculate

$$\mathbf{x}_0'(t) + \varepsilon \mathbf{x}_1'(t) + \dots = \mathbf{F}(\mathbf{x}_0(t)) + \varepsilon \mathbf{DF}(\mathbf{x}_0(t)) \cdot \mathbf{x}_1 + \dots . \tag{4.53}$$

Equation (4.53) must hold for all values of ε, i.e., the two power series on either side of the equation define the same functions of ε. Thus the coefficients of each power of ε must be equal. Matching corresponding powers of ε in(4.53), we obtain

$$\begin{array}{lrcl} \text{(a) } \mathcal{O}(\varepsilon^0): & \mathbf{x}_0' & = & \mathbf{F}(\mathbf{x}_0), \\ \text{(b) } \mathcal{O}(\varepsilon^1): & \mathbf{x}_1' & = & \mathbf{DF}(\mathbf{x}_0(t)) \cdot \mathbf{x}_1. \end{array} \tag{4.54}$$

(The letter $\mathcal{O}$ is a mnemonic for "order.") The $\mathcal{O}(\varepsilon^0)$-equation merely repeats the ODE for our original solution. The $\mathcal{O}(\varepsilon^1)$-equation gives new information, i.e., an ODE for $\mathbf{x}_1$, which is a linear homogeneous system with time-dependent coefficients

$$\mathbf{x}_1' = A(t)\mathbf{x}_1, \tag{4.55}$$

where the coefficient matrix is given in (4.54b). Similarly, matching powers of ε in the initial conditions (4.51) gives

$$\begin{array}{lrcl} \text{(a) } \mathcal{O}(\varepsilon^0): & \mathbf{x}_0(0) & = & \mathbf{b}_0, \\ \text{(b) } \mathcal{O}(\varepsilon^1): & \mathbf{x}_1(0) & = & \mathbf{b}_1. \end{array} \tag{4.56}$$

The $\mathcal{O}(\varepsilon^0)$-equation here is nothing new, but the $\mathcal{O}(\varepsilon^1)$-equation provides an initial

[18] We may describe (4.52) as an *ansatz*, i.e., an assumed form for the solution of a problem. This term, which comes from German, is a useful one to add to your (mathematical) vocabulary.

[19] Fortunately, it suffices for our purposes to carry the expansion only through the first power. Although higher-order terms of a multivariable Taylor series can be handled efficiently with the multi-index notation (cf. Section 10.2 of [74]), it would be a distraction to introduce it here.

condition for the ODE (4.55). Theorem 3.4.1 guarantees that the IVP (4.55), (4.56b) has a unique solution $\mathbf{x}_1(t)$ for t in the same interval $[0, \beta_0)$ on which $\mathbf{x}_0$ is defined.

Here is the main result of Section 4.6. Of course analogous results hold for negative times.

Theorem 4.6.1. *Let $\mathbf{F} : \mathcal{U} \to \mathbb{R}^d$ be $\mathcal{C}^1$. In the above notation, for every $t \in [0, \beta_0)$,*

$$\lim_{\varepsilon \to 0} \frac{|\mathbf{x}(t, \varepsilon) - \mathbf{x}_0(t) - \varepsilon \mathbf{x}_1(t)|}{\varepsilon} = 0. \tag{4.57}$$

Moreover, if $T < \beta_0$, the limit is uniform for $0 \le t \le T$.

You may find the proof of this result tough going. We postpone the proof until Section 4.6.5, first exploring related ideas that among other things, make the proof easier to read.

4.6.2 The Perspective of Differentiability

Consider differentiating the flow map $\boldsymbol{\varphi} : \Omega \to \mathcal{U}$ with respect to the initial condition $\mathbf{b}$; i.e., consider

$$\frac{\partial \boldsymbol{\varphi}}{\partial b_j}(t, \mathbf{b}_0) = \lim_{\varepsilon \to 0} \frac{\boldsymbol{\varphi}(t, \mathbf{b}_0 + \varepsilon \mathbf{e}_j) - \boldsymbol{\varphi}(t, \mathbf{b}_0)}{\varepsilon}. \tag{4.58}$$

In our notation above, we have $\boldsymbol{\varphi}(t, \mathbf{b}_0) = \mathbf{x}_0(t)$, and if in (4.51) we define $\mathbf{b}_1 = \mathbf{e}_j$, then $\boldsymbol{\varphi}(t, \mathbf{b}_0 + \varepsilon \mathbf{e}_j) = \mathbf{x}(t, \varepsilon)$. Theorem 4.6.1 implies that the limit in (4.58) exists, so $\boldsymbol{\varphi}(t, \mathbf{b})$ is in fact differentiable with respect to b_j; indeed, $\partial \boldsymbol{\varphi} / \partial b_j(t, \mathbf{b}_0)$ equals the appropriate solution of the IVP (4.55), (4.56b), which we repeat as

$$\mathbf{w}_j' = A(t)\mathbf{w}_j, \qquad \mathbf{w}_j(0) = \mathbf{e}_j, \tag{4.59}$$

where $A(t) = \mathbf{DF}(\mathbf{x}_0(t))$. It is noteworthy that the partial derivatives of $\boldsymbol{\varphi}$ with respect to the various components b_j all satisfy the *same* ODE; they differ only in their initial conditions.

Let us motivate how the IVP (4.59) arises. Regarding the initial condition, at time zero, (4.58) reduces to the triviality

$$\frac{\partial \boldsymbol{\varphi}}{\partial b_j}(0, \mathbf{b}_0) = \lim_{\varepsilon \to 0} \frac{(\mathbf{b}_0 + \varepsilon \mathbf{e}_j) - (\mathbf{b}_0)}{\varepsilon} = \mathbf{e}_j. \tag{4.60}$$

For $t > 0$, the flow $\boldsymbol{\varphi}(t, \mathbf{b})$ satisfies the ODE

$$\frac{\partial}{\partial t} \boldsymbol{\varphi}(t, \mathbf{b}) = \mathbf{F}(\boldsymbol{\varphi}(t, \mathbf{b})). \tag{4.61}$$

We differentiate this equation without worrying about justification; this will be provided by the proof of Theorem 4.6.1. Specifically, take the derivative of (4.61) with

respect to b_j using the chain rule,

$$\frac{\partial}{\partial b_j}\frac{\partial \boldsymbol{\varphi}}{\partial t}(t,\mathbf{b}) = \mathbf{DF}(\boldsymbol{\varphi}(t,\mathbf{b}))\,\frac{\partial \boldsymbol{\varphi}}{\partial b_j}(t,\mathbf{b}), \tag{4.62}$$

interchange the order of the t and b derivatives to obtain

$$\frac{\partial}{\partial t}\frac{\partial \boldsymbol{\varphi}}{\partial b_j}(t,\mathbf{b}) = \mathbf{DF}(\boldsymbol{\varphi}(t,\mathbf{b}))\,\frac{\partial \boldsymbol{\varphi}}{\partial b_j}(t,\mathbf{b}),$$

and set $\mathbf{b} = \mathbf{b}_0$ to argue that $\partial\boldsymbol{\varphi}/\partial b_j(t,\mathbf{b}_0)$ should satisfy the ODE in (4.59).

4.6.3 Examples

The IVP (4.59) provides a beautiful characterization of $\partial\boldsymbol{\varphi}/\partial b_j$, which is used frequently in theoretical analysis of ODEs. Unfortunately, cases in which (4.59) can be solved explicitly are the exception rather than the rule.

One important special case in which the IVP can be solved occurs when $\mathbf{b}_0$ is an equilibrium of $\mathbf{x}' = \mathbf{F}(\mathbf{x})$. In this case $\boldsymbol{\varphi}(t,\mathbf{b}_0) \equiv \mathbf{b}_0$, so the coefficient matrix in (4.59) is independent of time, $A = \mathbf{DF}(\mathbf{b}_0)$. In other words, the system (4.59) has constant coefficients. Incidentally, this approximation, which is known as the *linearization* of $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ at the equilibrium, figures heavily in the qualitative theory of ODEs.

In case you would find it helpful to see the theorem in action, here is a less trivial example in which an explicit solution of (4.59) is possible. Consider the IVP for the Lotka–Volterra equations,

$$\begin{array}{llcll} (a) & x' & = & x - xy, & x(0) = b_1, \\ (b) & y' & = & \rho(xy - y), & y(0) = b_2. \end{array} \tag{4.63}$$

If $b_2 = 0$, then (4.63) has the explicit solution $\boldsymbol{\varphi}(t,(b_1,0)) = (b_1 e^t, 0)$; i.e., without predators, the prey grow exponentially. If a small population of predators were introduced, how would their numbers evolve? For small b_2, we have the approximation (for the predator population)

$$\varphi_2(t,(b_1,b_2)) \approx b_2 \cdot \frac{\partial \varphi_2}{\partial b_2}(t,(b_1,0)).$$

In Exercise 12, we ask you to solve the appropriate version of (4.59) to show that

$$\frac{\partial \varphi_2}{\partial b_2}(t,(b_1,0)) = \exp\left\{\rho[b_1(e^t - 1) - t]\right\}. \tag{4.64}$$

Thus, to lowest order, the number of predators grows very rapidly indeed as t increases, an exponential of an exponential. Rapid growth is hardly surprising, since

the predator's food supply is increasing without bound. Of course, the above approximation must become inaccurate as t increases. Indeed, we know from Exercise 3 in Chapter 1 that the solution of (4.63) is periodic if $b_2 > 0$, so the population must remain bounded. To interpret this discussion in the language of Theorem 4.6.1, the larger you want to make T, the smaller you must take ε in the limit (4.57).

Other examples of differentiation with respect to initial conditions, including in the above example calculating the effect of the predators on the prey, are given in the exercises.

4.6.4 The Order Notation

To prepare for the proof of Theorem 4.6.1, we introduce the order notation,[20] i.e., big-$\mathcal{O}$ and little-o. This notation makes an otherwise messy proof relatively clean.

The rigorous use of big-$\mathcal{O}$, the simpler concept, is as follows: Given a quantity that depends on a parameter, say $f(\varepsilon)$, where $0 < \varepsilon < \varepsilon_0$ and f may be either a vector or scalar quantity, we say that f is order-ε, written $f(\varepsilon) = \mathcal{O}(\varepsilon)$, if

$$(\exists C)(\exists \varepsilon_1 > 0) \text{ such that } 0 < \varepsilon < \varepsilon_1 \implies |f(\varepsilon)| \leq C\varepsilon.$$

(The formula $f(\varepsilon) = \mathcal{O}(\varepsilon)$ may also be read "f is big-$\mathcal{O}$ of ε.") The same notation is also used in several more complicated contexts. If ε can assume either sign, i.e., if $f(\varepsilon)$ is defined for $0 < |\varepsilon| < \varepsilon_0$, we write $f(\varepsilon) = \mathcal{O}(|\varepsilon|)$ to mean $|f(\varepsilon)| \leq C|\varepsilon|$, provided $|\varepsilon|$ is sufficiently small. More generally, if $\phi(\varepsilon) > 0$ is some function that tends to zero as $\varepsilon \to 0$, for example $\phi(\varepsilon) = |\varepsilon|^p$, we interpret the formula $f(\varepsilon) = \mathcal{O}(\phi(\varepsilon))$ with the obvious inequality. Also, the notation is generalized to estimate quantities that depend on multiple parameters. For example, in Theorem 4.5.1 the solution $\mathbf{x}(t) = \boldsymbol{\varphi}(t, \mathbf{b})$ depends on the d parameters of the initial data, $b_1, \ldots, b_d$, and we may paraphrase the conclusion of the theorem as

$$|\boldsymbol{\varphi}(t, \mathbf{b}) - \boldsymbol{\varphi}(t, \mathbf{b}_0)| = \mathcal{O}(|\mathbf{b} - \mathbf{b}_0|). \tag{4.65}$$

In the notation of Theorem 4.6.1, we have from Theorem 4.5.1

$$|\mathbf{x}(t, \varepsilon) - \mathbf{x}_0(t)| = \mathcal{O}(\varepsilon). \tag{4.66}$$

[20] There is an unfortunate ambiguity in the use of the symbol $\mathcal{O}$. In (4.54) and (4.56), the symbol $\mathcal{O}(\varepsilon^p)$ is merely intended as a placeholder, something to indicate the terms in a power series that are proportional to ε^p. This might be called the informal usage. By contrast, we are now going to describe a rigorous, technical meaning of this symbol. Both usages appear in the literature. Although some authors introduce separate notations to distinguish between the two meanings, we prefer to avoid this proliferation of notation. We believe that once you have been alerted to the issue, you will be able to determine from context which usage is intended.

Indeed, we say that (4.66) holds *uniformly* for $t \in [0, T]$, because the inequality

$$|\mathbf{x}(t, \varepsilon) - \mathbf{x}_0(t)| \leq e^{LT} |\mathbf{b}_1| \, \varepsilon \tag{4.67}$$

holds for all t in this interval with the same constant $C = e^{LT} |\mathbf{b}_1|$. Alternatively, we may hide the explicit constants in (4.67) and rewrite this inequality as

$$\sup_{0 \leq t \leq T} |\mathbf{x}(t, \varepsilon) - \mathbf{x}_0(t)| = \mathcal{O}(\varepsilon). \tag{4.68}$$

Little-o is a more delicate concept. If f is defined for $0 < \varepsilon < \varepsilon_0$, we say that f is little-o of ε, written $f = o(\varepsilon)$, if

$$(\forall \eta > 0)(\exists \varepsilon_1 > 0) \text{ such that } 0 < \varepsilon < \varepsilon_1 \implies |f(\varepsilon)| \leq \eta \varepsilon.$$

Of course, this definition is equivalent to[21]

$$\lim_{\varepsilon \to 0} \frac{|f(\varepsilon)|}{\varepsilon} = 0.$$

The little-o notation, like big-$\mathcal{O}$, is also used in more complicated contexts. For example, if the vector-valued function $\mathbf{F}(\mathbf{x})$ is continuously differentiable, then we may write[22]

$$\mathbf{F}(\mathbf{z}) = \mathbf{F}(\mathbf{z}_0) + \mathbf{DF}(\mathbf{z}_0) \cdot (\mathbf{z} - \mathbf{z}_0) + o(|\mathbf{z} - \mathbf{z}_0|) \tag{4.69}$$

for $\mathbf{z}$ near $\mathbf{z}_0$.

The following facts are part of the order-notation liturgy:

$$\begin{array}{l} \text{(a) If } f(\varepsilon) = o(\varepsilon), \text{ then } Cf(\varepsilon) = o(\varepsilon) \text{ for every constant } C. \\ \text{(b) If } f(\varepsilon) = o(\phi(\varepsilon)) \text{ and if } \phi(\varepsilon) = \mathcal{O}(\varepsilon), \text{ then } f(\varepsilon) = o(\varepsilon). \end{array} \tag{4.70}$$

We ask you to derive (4.70) in Exercise 10. Even though these facts may be viewed as just a reworking of familiar properties of limits, it is worth your while to *do the exercise before reading the proof of Theorem 4.6.1*. Order notation allows you to focus on higher-level issues in the proof than repeatedly rederiving properties of limits. Incidentally, (4.70) can be generalized in numerous ways, but this limited version suffices for our needs.

---

[21] Incidentally, big-$\mathcal{O}$ may be similarly characterized:

$$f(\varepsilon) = \mathcal{O}(\varepsilon) \iff \limsup_{\varepsilon \to 0} |f(\varepsilon)/\varepsilon| < \infty.$$

[22] We use $\mathbf{z}$ rather than $\mathbf{x}$ in order to reserve the latter for the functions defined by (4.54), (4.56).

To return to Theorem 4.6.1, note that our desired conclusion simply asserts that

$$\sup_{0\le t\le T} |\mathbf{x}(t,\varepsilon) - \mathbf{x}_0(t) - \varepsilon \mathbf{x}_1(t)| = o(\varepsilon). \tag{4.71}$$

4.6.5 Proof of Theorem 4.6.1

Given $T < \beta_0$, choose $\delta > 0$ and a compact set $\mathcal{K} \subset \mathcal{U}$ for which (4.40) holds, and let L be given by (4.45). For all sufficiently small ε, the perturbed initial condition $\mathbf{b}_0 + \varepsilon \mathbf{b}_1$ belongs to $B(\mathbf{b}_0, e^{-LT}\delta)$, so by Corollary 4.5.2,

$$\mathbf{x}(t,\varepsilon) \in \overline{B(\mathbf{x}_0(t), \delta)} \subset \mathcal{K}, \qquad 0 \le t \le T. \tag{4.72}$$

Forming a linear combination of the integral relations

$$\begin{aligned}
\mathbf{x}(t,\varepsilon) &= \mathbf{b}_0 + \varepsilon \mathbf{b}_1 + \int_0^t \mathbf{F}(\mathbf{x}(s,\varepsilon))\, ds, \\
\mathbf{x}_0(t) &= \mathbf{b}_0 + \int_0^t \mathbf{F}(\mathbf{x}_0(s))\, ds, \\
\mathbf{x}_1(t) &= \mathbf{b}_1 + \int_0^t A(s)\mathbf{x}_1(s)\, ds,
\end{aligned}$$

we deduce that $g(t,\varepsilon) = |\mathbf{x}(t,\varepsilon) - \mathbf{x}_0(t) - \varepsilon \mathbf{x}_1(t)|$ satisfies

$$g(t,\varepsilon) \le \int_0^t |\mathbf{F}(\mathbf{x}(s,\varepsilon)) - \mathbf{F}(\mathbf{x}_0(s)) - \varepsilon A(s)\mathbf{x}_1(s)|\, ds.$$

Add and subtract $A(s)\{\mathbf{x}(s,\varepsilon) - \mathbf{x}_0(s)\}$ in the integral and use the triangle inequality to obtain

$$g(t,\varepsilon) \le \mathcal{I}_1(t,\varepsilon) + \mathcal{I}_2(t,\varepsilon), \tag{4.73}$$

where

$$\begin{aligned}
&\text{(a) } \mathcal{I}_1(t,\varepsilon) = \int_0^t |\mathbf{F}(\mathbf{x}(s,\varepsilon)) - \mathbf{F}(\mathbf{x}_0(s)) - A(s)\{\mathbf{x}(s,\varepsilon) - \mathbf{x}_0(s)\}|\, ds, \\
&\text{(b) } \mathcal{I}_2(t,\varepsilon) = \int_0^t |A(s)\{\mathbf{x}(s,\varepsilon) - \mathbf{x}_0(s) - \varepsilon \mathbf{x}_1(s)\}|\, ds.
\end{aligned} \tag{4.74}$$

The integrand in $\mathcal{I}_2(t,\varepsilon)$ is bounded by $\|A(s)\|\, g(s,\varepsilon)$, and since $\mathbf{x}_0(s) \in \mathcal{K}$, we have $\|A(s)\| = \|\mathbf{DF}(\mathbf{x}_0(s))\| \le L$. Therefore,

$$g(t,\varepsilon) \le \mathcal{I}_1(t,\varepsilon) + L \int_0^t g(s,\varepsilon)\, ds.$$

Thus, by Gronwall's inequality,

$$g(t,\varepsilon) \le e^{Lt} \sup_{0\le s\le T} \mathcal{I}_1(s,\varepsilon),$$

and taking the supremum over t yields

$$\sup_{0\le t\le T} g(t,\varepsilon) \le e^{LT} \sup_{0\le s\le T} \mathcal{I}_1(s,\varepsilon). \tag{4.75}$$

The following lemma gives control over the RHS of (4.75). (In Exercise 25 we ask you to prove the lemma, based on showing that the pointwise estimate (4.69) is uniform over an appropriate compact set.)

Lemma 4.6.2. *As ε tends to zero,*

$$\sup_{0\le s\le T} \mathcal{I}_1(s,\varepsilon) = o(\varepsilon). \tag{4.76}$$

Applying the lemma to (4.75), we may rewrite this equation as

$$\sup_{0\le t\le T} g(t,\varepsilon) = C\, o(\varepsilon),$$

where $C = e^{LT}$. Thus, our desired conclusion (4.71) follows from (4.70a). This completes the proof.

4.6.6 Tying Up Loose Ends

We claim that the flow map $\boldsymbol{\varphi}(t,\mathbf{b})$ is $\mathcal{C}^1$. By Theorem 4.6.1, the partial derivatives $\partial\boldsymbol{\varphi}/\partial b_j$ exist, and of course $\partial\boldsymbol{\varphi}/\partial t$ also exists. We need to show that these partial derivatives are continuous. Trivially, from the ODE $\partial\boldsymbol{\varphi}/\partial t = \mathbf{F}(\boldsymbol{\varphi}(t,\mathbf{b}))$, the t-derivative is continuous. The derivative with respect to b_j was characterized by the IVP (4.59), which we rewrite indicating the dependence of the coefficient matrix on $\mathbf{b}$:

$$\frac{\partial}{\partial t}\frac{\partial\boldsymbol{\varphi}}{\partial b_j} = \mathbf{DF}(\boldsymbol{\varphi}(t,\mathbf{b}))\frac{\partial\boldsymbol{\varphi}}{\partial b_j}, \qquad \frac{\partial\boldsymbol{\varphi}}{\partial b_j}(0) = \mathbf{e}_j. \tag{4.77}$$

It follows from Theorem 4.5.4 that $\partial\boldsymbol{\varphi}/\partial b_j$ is continuous. This proves the claim.

In Section 4.6.2 we motivated the formula (4.59) for $\partial\boldsymbol{\varphi}/\partial b_j$, modulo an interchange of the order of differentiation that needs to be justified. According to Theorem B.3.2 in Appendix B, it suffices to show that one of the mixed partials exists and is continuous. It follows from (4.77) that $\partial^2\boldsymbol{\varphi}/\partial t\partial b_j$ is continuous, the RHS of this ODE being a product of continuous functions.

4.6.7 Generalizations

There are more results about the differentiability of the solution of IVPs than either you or we care to explore fully, but a few highlights need to be mentioned. We refrain

from giving formal statements of these results in the hopes of making the text more readable. A careful treatment of such results is given in Section 1.7 of [15].

First let us generalize Theorem 4.6.1 to nonautonomous IVPs. It can be shown that the flow operator $\boldsymbol{\varphi}(t, t_0, \mathbf{b})$ obtained by solving

$$\mathbf{x}' = \mathbf{G}(\mathbf{x}, t), \qquad \mathbf{x}(t_0) = \mathbf{b} \tag{4.78}$$

is $\mathcal{C}^1$ with respect to all variables, provided $\mathbf{G}$ is $\mathcal{C}^1$. Moreover, $\partial\boldsymbol{\varphi}/\partial b_j(t, t_0, \mathbf{b})$ satisfies a linear homogeneous IVP

$$\frac{d\mathbf{w}}{dt} = \mathbf{DG}(\boldsymbol{\varphi}(t, t_0, \mathbf{b}), t)\mathbf{w}, \qquad \mathbf{w}(t_0) = \mathbf{e}_j, \tag{4.79}$$

where $\mathbf{DG}$ denotes the $d \times d$ matrix of partial derivatives $\partial\mathbf{G}/\partial x_j$, not including the t derivative.

Next, we consider "differentiability with respect to the equation" through a parametrized family of IVPs, say

$$\mathbf{x}' = \mathbf{G}(\mathbf{x}, \boldsymbol{\alpha}), \qquad \mathbf{x}(0) = \mathbf{b}, \tag{4.80}$$

where $\boldsymbol{\alpha} \in \mathcal{V} \subset \mathbb{R}^m$; to simplify the notation, we assume that (4.80) is autonomous. The flow map $\boldsymbol{\varphi}(t, \mathbf{b}, \boldsymbol{\alpha})$ is $\mathcal{C}^1$ with respect to all its arguments, and $\partial\boldsymbol{\varphi}/\partial\alpha_j$ satisfies the linear *inhomogeneous* IVP

$$\frac{d\mathbf{w}}{dt} = \mathbf{DG}(\boldsymbol{\varphi}(t, \mathbf{b}, \boldsymbol{\alpha}), \boldsymbol{\alpha})\mathbf{w} + \frac{\partial\mathbf{G}}{\partial\alpha_j}(\boldsymbol{\varphi}(t, \mathbf{b}, \boldsymbol{\alpha}), \boldsymbol{\alpha}), \quad \mathbf{w}(0) = \mathbf{0}. \tag{4.81}$$

As above, $\mathbf{DG}$ denotes the $d \times d$ matrix of derivatives of $\mathbf{G}$ with respect to x_j; the derivative with respect to the parameter is written out explicitly. In fact, this result does not require separate proof; it may be derived easily with a "reduce it to the previous case" ruse, as we discuss in Exercise 24.

Finally, our third generalization addresses higher-order derivatives: it can be shown that if an ODE is of class $\mathcal{C}^k$, then the solution operator also has k continuous derivatives with respect to all variables. In particular, under the hypotheses of Theorem 4.6.1, if $\mathbf{F} \in \mathcal{C}^2$, then

$$\mathbf{x}(t, \varepsilon) - \mathbf{x}_0(t) - \varepsilon\mathbf{x}_1(t) = \mathcal{O}(\varepsilon^2).$$

4.7 Exercises

After the core exercises, there are subsections on applying the differentiation results, on cleaning up some loose ends from previous chapters, and on computing.

4.7.1 Core Exercises

The primary purposes of the core exercises are as follows:

Proof of a minor generalization	1
Unfinished business	2, 5–7, 10
Use of trapping regions to prove global existence	3
Counterexamples to clarify the theory	4
A first look at asymptotic behavior	8, 9

1. (a) Prove the following variant of Theorem 4.1.2: If $\mathbf{F}$ is bounded on $\mathcal{U}$ and if $\beta_* < \infty$, then $\mathbf{x}_*(t)$ tends to a point on $\partial\mathcal{U}$ as $t \to \beta_*$.

   (b) Construct an IVP for a scalar ODE $x' = f(x)$, with f bounded, whose solution has a maximal interval of existence (α_*, β_*) with both endpoints finite.

2. Prove Theorem 4.2.3.

   Hint: You must show that the solution cannot leave $\mathcal{K}$. Rule out crossing $\partial\mathcal{K}$ at a regular point by the same argument used to prove Theorem 4.2.2. At a corner point, say $\mathbf{P}$, recall the notation of Section B.3.3 to represent $\partial\mathcal{K}$ near $\mathbf{P}$ as the intersection of two smooth curves, $\{\phi_1 = 0\}$ and $\{\phi_2 = 0\}$. Take limits of regular points to conclude that

   $$\langle \nabla\phi_k(\mathbf{P}), \mathbf{F}(\mathbf{P})\rangle \geq 0, \qquad k = 1, 2.$$

   Use the hypothesis that $\nabla\phi_1(\mathbf{P})$ and $\nabla\phi_2(\mathbf{P})$ are linearly independent to argue that one of these inequalities must be strict, as in (4.12). Rule out crossing $\partial\mathcal{K}$ at $\mathbf{P}$ by the simple argument used to derive Theorem 4.2.2 from (4.12).

3. Use trapping regions to analyze global existence for the following equations:

   (a) The Lorenz equations:

   $$\begin{aligned} x' &= \sigma(y - x), \\ y' &= \rho x - y - xz, \\ z' &= -\beta z + xy, \end{aligned}$$

   where σ, ρ, β are positive constants (arbitrary initial conditions).

   Hint: Deduce global existence by showing that a set of the form

   $$\mathcal{K} = \{(x, y, z) : x^2 + y^2 + (z - \rho - \sigma)^2 \leq A^2\}$$

   is a trapping region if A is sufficiently large; i.e., the trapping region is bounded by the level set of a carefully chosen quadratic function.

   (b) Equations for the evolution of two interacting species:

   $$x' = x\,(1 - x - by)\,, \qquad y' = \rho y\,(1 - y - cx)\,,$$

   where ρ, b, c are constants with $\rho > 0$ (both $x(0)$, $y(0)$ nonnegative).

Discussion: Note that each species has logistic growth that is modified by the presence of the other. If b, c are both positive, then the interaction is competitive; if they are both negative, it is symbiotic. The most interesting cases, for which there are coexistence equilibria in the (open) first quadrant, are (i) $1 < b, c$, (ii) $0 < b, c < 1$, and (iii) $-1 < b, c < 0$.

Hint: In Cases (i) and (ii), draw the nullclines to show that there are triangular trapping regions of the form (4.23). A slightly more complicated region is needed for Case (iii).

Discussion: Incidentally, we invite you to compute a few typical solutions for an equation belonging to Case (i) and for one belonging to Case (ii). You will find that they behave differently as $t \to \infty$. This behavior may seem a little mysterious at present, but the ideas from Chapter 6 will clarify it.

Challenge: A harder, but educational, problem is to prove that the solution may blow up in finite time when $b, c < -1$. *Try it!* As preparation for this harder problem, you might first try to show blowup for the simpler system $x' = xy,\ y' = xy$.

(c) A simplified activator–inhibitor system

$$\begin{aligned} (a) \quad x' &= \sigma \tfrac{1}{1+y} x^2 - x, \\ (b) \quad y' &= \rho\,[x^2 - y], \end{aligned} \tag{4.82}$$

i.e., the limit of equations (4.30) as $\kappa \to \infty$ (both $x(0),\ y(0)$ nonnegative).

Remark: It was straightforward to construct a trapping region for (4.30); you'll have to work considerably harder to do so for (4.82). But below we will use the simpler equations (4.82) in calculations about the properties of solutions of activator–inhibitor models.

(d) The "repressilator"

$$\begin{aligned} x' &= \frac{\mu}{1+y^n} - x, \\ y' &= \frac{\mu}{1+z^n} - y, \\ z' &= \frac{\mu}{1+x^n} - z, \end{aligned}$$

where μ and n are positive constants (initial values of all variables nonnegative).

Hint: To prove global existence for this system, which will be introduced and analyzed in Section 8.7.2, construct a trapping region of the form

$$\mathcal{K} = \{(x, y, z) : x \geq 0,\ y \geq 0,\ z \geq 0,\ x + y + z \leq A\};$$

i.e., prove that the flow is inward on each of the four faces of this simplex if A is sufficiently large. Note that $\mathcal{K}$ is a three-dimensional region whose boundary is only piecewise smooth. (We have shied away from actually defining piecewise smooth in higher dimensions, but however this phrase is defined, it surely applies to this set.) Hence Theorem 4.2.2 is not immediately applicable to proving global existence. If you are bothered by this gap, you can close it by mimicking Problem 2.

4. (a) Regarding Theorem 4.1.2, give an example to show that if $\beta_* = \infty$, then the solution of an IVP can stay inside a compact set for all time.

   (b) Give an example of a $\mathcal{C}^1$ function that satisfies the linear-growth estimate (4.4) but is not (globally) Lipschitz.

   (c) For the system

$$\begin{aligned} x' &= x^2 - y^2, \\ y' &= 2xy, \end{aligned} \tag{4.83}$$

   find a solution that blows up in finite time but most nearby solutions exist for all time.

   Hint: First show that with initial conditions $x(0) = 1, y(0) = 0$, the solution of (4.83) blows up in finite time. Although you can't solve (4.83) for other trajectories, you can locate the solution curves, i.e., find the orbits, as follows. Along an orbit you have the ODE

$$\frac{dy}{dx} = \frac{dy/dt}{dx/dt} = \frac{2xy}{x^2 - y^2}.$$

   Multiply the equation by $(x/y)^2 - 1$ and manipulate the result into the form[23]

$$\frac{d}{dx}\left(\frac{x^2}{y} + y\right) = 0,$$

   from which you may deduce that the solution curves are circles through the origin, $x^2 + (y - C)^2 = C^2$, where C is an arbitrary constant. Argue from this information that the solution of (4.83) with initial conditions $x(0) = 1, y(0) = b$, where $b \neq 0$, exists for all time.

5. (a) Prove the following generalization of Theorem 4.2.1, referring to Exercise 8(a) in Chapter 3.

   Theorem 4.7.1. *If $\mathbf{G} : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d$ is locally Lipschitz in $\mathbf{x}$ and t and if there exist nonnegative continuous functions $B(t), K(t)$ such that*

$$|\mathbf{G}(\mathbf{x}, t)| \leq K(t)|\mathbf{x}| + B(t), \qquad (\mathbf{x}, t) \in \mathbb{R}^d \times \mathbb{R}, \tag{4.84}$$

   then the solution of the IVP

$$\mathbf{x}' = \mathbf{G}(\mathbf{x}, t), \qquad \mathbf{x}(0) = \mathbf{b}$$

[23] A scalar ODE of the form $(d/dx)f(x, y) = 0$ is called *exact*. In this example we have made the equation exact, and therefore solvable, by means of an *integrating factor*. This is another solution technique, one that we did not cover in Section 1.3; to learn more, see Section 1.9 of [10].

exists for all time, $-\infty < t < \infty$. Moreover, for every finite $T > 0$,

$$|\mathbf{x}(t)| \le |\mathbf{b}|e^{K_{\max}|t|} + \frac{B_{\max}}{K_{\max}}(e^{K_{\max}|t|} - 1), \qquad -T \le t \le T, \tag{4.85}$$

where

$$B_{\max} = \max_{|s|\le T} |B(s)|, \qquad K_{\max} = \max_{|s|\le T} |K(s)|.$$

Discussion: It will be useful below to have Theorem 4.7.1 explicitly formulated, and it is useful training to adapt the proof of Theorem 4.2.1 to handle the nonautonomous case. For a complete proof, you would need to prove extensions of Proposition 4.1.1 and Theorem 4.1.2 to nonautonomous equations. We invite you to skip this not very rewarding task and regard the extensions of those two results as given.

Incidentally, every linear equation $\mathbf{x}' = A(t)\mathbf{x} + \mathbf{g}(t)$ satisfies the above hypotheses, provided of course that $A(t)$ and $\mathbf{g}(t)$ are continuous. Thus, this theorem provides another proof of global existence for linear equations, an alternative to the approach based on Picard iteration that was outlined in Exercise 16 in Chapter 3.

(b) Use Theorem 4.7.1 to prove global existence for

$$\begin{aligned} x' &= y, \\ y' &= -(1/4 + \beta\cos t)x. \end{aligned}$$

Remark: Recall your computations from Exercise 13 in Chapter 3, in which you found the surprising fact that solutions of this equation may grow exponentially with time; from this problem you may conclude that solutions grow no faster than exponentially.

6. Extend Theorem 4.2.2 to ODEs on the cylinder $\mathbb{R}/2\pi\mathbb{Z} \times \mathbb{R}$.

Hint: First, some notation: If $\mathcal{K}$ is a subset of $\mathbb{R}/2\pi\mathbb{Z} \times \mathbb{R}$, let $\mathcal{K}_{\text{lift}} \subset \mathbb{R}^2$ be defined by[24]

$$\mathcal{K}_{\text{lift}} = \{(x, y) \in \mathbb{R}^2 : \Pi \cdot (x, y) \in \mathcal{K}\},$$

where $\Pi : \mathbb{R}^2 \to \mathbb{R}/2\pi\mathbb{Z} \times \mathbb{R}$ is the natural projection. In less formal terms, the projection "wraps the plane around the cylinder," and $\mathcal{K}_{\text{lift}}$ is the set obtained from $\mathcal{K}$ if the cylinder is "unwrapped." (Cf. Figure 4.4.)

By an ODE on $\mathbb{R}/2\pi\mathbb{Z} \times \mathbb{R}$ we mean a two-dimensional ODE

$$\begin{bmatrix} \theta' \\ y' \end{bmatrix} = \mathbf{F}(\theta, y), \tag{4.86}$$

where $\mathbf{F} : \mathbb{R}^2 \to \mathbb{R}^2$ is 2π-periodic in its first argument, i.e., $\mathbf{F}(\theta+2\pi, y) = \mathbf{F}(\theta, y)$. We propose to prove global existence for the IVP for (4.86) by applying the theory

[24]The subscript "lift" refers to the idea that $\mathbb{R}^2$ is a simply connected covering space for $\mathbb{R}/2\pi\mathbb{Z} \times \mathbb{R}$ that lies "above" $\mathbb{R}/2\pi\mathbb{Z} \times \mathbb{R}$.

of this chapter to the equivalent "lifted" planar IVP

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \mathbf{F}(x,y), \quad \begin{bmatrix} x(0) \\ y(0) \end{bmatrix} = \mathbf{b}, \tag{4.87}$$

where x may vary over $(-\infty, \infty)$.

Suppose $\mathcal{K} \subset \mathbb{R}/2\pi\mathbb{Z} \times \mathbb{R}$ is a compact trapping region for (4.86); then $\mathcal{K}_{\text{lift}}$ is a trapping region for (4.87). By Corollary 4.2.4, if $\mathbf{b} \in \mathcal{K}_{\text{lift}}$, then the solution of (4.87) remains in $\mathcal{K}_{\text{lift}}$ for as long as it exists, say $0 \le t < t_*$. From compactness we have

$$C = \max_{(x,y)\in\mathcal{K}_{\text{lift}}} |\mathbf{F}(x,y)| = \max_{(\theta,y)\in\mathcal{K}} |\mathbf{F}(\theta,y)| < \infty.$$

Therefore, $|\mathbf{x}(t)| \le |\mathbf{b}| + Ct$ for $0 \le t < t_*$. If t_* were finite, this estimate would contradict Theorem 4.1.2, so (4.87) must have a global solution.

Remark: A similar argument works for ODEs on the torus $\mathbb{T}^2$.

7. Prove Theorem 4.5.4.

Hint: You need to show that $|\mathbf{w}(t_1, \boldsymbol{\alpha}_1) - \mathbf{w}(t_2, \boldsymbol{\alpha}_2)|$ can be made arbitrarily small by making t_1, t_2 and $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2$ sufficiently close to one another. It suffices to consider only positive times $t_k \le T$, where $T < T_2$, and to restrict $\boldsymbol{\alpha}_k$ to a compact subset $\mathcal{N}$ of the parameter set $\mathcal{V}$. Let

$$M = \max_{0\le t\le T} \max_{\boldsymbol{\alpha}\in\mathcal{N}} \|A(t,\boldsymbol{\alpha})\|. \tag{4.88}$$

It is convenient to abbreviate $\mathbf{w}(t, \boldsymbol{\alpha}_k)$ to $\mathbf{w}_k(t)$; i.e., you need to show that $|\mathbf{w}_1(t_1) - \mathbf{w}_2(t_2)|$ is small. Add and subtract a term $\mathbf{w}_2(t_1)$ to estimate

$$|\mathbf{w}_1(t_1) - \mathbf{w}_2(t_2)| \le |\mathbf{w}_1(t_1) - \mathbf{w}_2(t_1)| + |\mathbf{w}_2(t_1) - \mathbf{w}_2(t_2)|. \tag{4.89}$$

To estimate the second term here, first apply Gronwall's inequality to the ODE $\mathbf{w}_2' = A(t, \boldsymbol{\alpha}_2)\mathbf{w}_2$ to conclude that $|\mathbf{w}_2(t)| \le |\mathbf{b}|e^{Mt}$ and then integrate this ODE to obtain

$$|\mathbf{w}_2(t_1) - \mathbf{w}_2(t_2)| \le M|\mathbf{b}|e^{MT}|t_1 - t_2|.$$

To estimate the first term in (4.89), subtract the ODEs for $\mathbf{w}_k(t)$ and then add and subtract $A(t, \boldsymbol{\alpha}_1)\mathbf{w}_2(t)$ to deduce that

$$\left|\frac{d}{dt}(\mathbf{w}_1 - \mathbf{w}_2)\right| \le \|A(t,\boldsymbol{\alpha}_1)\|\,|\mathbf{w}_1 - \mathbf{w}_2| + \|A(t,\boldsymbol{\alpha}_1) - A(t,\boldsymbol{\alpha}_2)\|\,|\mathbf{w}_2|.$$

Integrate this inequality and apply the generalization of Gronwall's inequality in Exercise 3.8(a) to conclude that

$$|(\mathbf{w}_1 - \mathbf{w}_2)(t)| \le B\frac{e^{Mt} - 1}{M},$$

where M is defined by (4.88) and

$$B = \max_{0\le t\le T} |\mathbf{b}|e^{MT}\,\|A(t,\boldsymbol{\alpha}_1) - A(t,\boldsymbol{\alpha}_2)\|.$$

Combine these inequalities with the fact that $A(t, \boldsymbol{\alpha})$ is uniformly continuous in $\boldsymbol{\alpha}$ to complete your proof.

Remark: Theorem 4.5.4 may be generalized to a (Lipschitz continuous) nonautonomous nonlinear equation, say $\mathbf{x}' = \mathbf{G}(\mathbf{x}, t, \boldsymbol{\alpha})$. The most convenient proof of such a result uses the trick proposed in Exercise 24.

8. For the chemostat equations (4.21) with $0 < \sigma < \rho/(1-\rho)$, use nullclines to show that every solution with initial conditions in the first quadrant tends to the equilibrium $(0, \sigma)$ as $t \to \infty$.

   Hint: Recall that a monotone function has a limit as $t \to \infty$.

   Remark: This problem and the next address an issue that figures heavily in the second half of this book, i.e., the asymptotic behavior of solutions of an ODE. In these problems the issue can be resolved with ad hoc methods. Starting in Chapter 6, we introduce more effective tools for investigating such asymptotic behavior.

9. (a) Prove the following lemma.

   Lemma 4.7.2. *Suppose $\mathbf{F} : \mathcal{U} \to \mathbb{R}^d$ is continuous. If the solution $\mathbf{x}(t)$ of the ODE $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ tends to a point $\mathbf{b}_* \in \mathcal{U}$ as $t \to \infty$, then $\mathbf{b}_*$ is an equilibrium of this equation.*

   Hint: Use the ODE to show that $\mathbf{x}'(t)$ has a limit as $t \to \infty$ and then argue that $\lim \mathbf{x}'(t) = \mathbf{0}$. Incidentally, satisfying an ODE is an essential part of this exercise; Exercise 3 in Appendix B concerns an example of a $\mathcal{C}^1$ bounded, monotone increasing function $g(t)$ (which must have a limit as $t \to \infty$) whose derivative does not converge.

   (b) For the chemostat equations (4.21) as shown in Figure 4.5, i.e., with $\sigma > \rho/(1-\rho) > 0$, show that every solution with initial conditions in the open first quadrant tends to the equilibrium $(\sigma - \rho/(1-\rho), \rho/(1-\rho))$ as $t \to \infty$.

   Hint: Let $x(t), y(t)$ be a solution of (4.21) with $x(0), y(0) > 0$. By solving (4.22), deduce that $x(t)+y(t)$ tends to σ as $t \to \infty$. Now refer to Figure 4.10, in which the first quadrant is divided into six regions by the nullclines and the line $\{x + y = \sigma\}$. First suppose that $(x(0), y(0)) \in R_1 \cup R_2 \cup R_3$. Argue that the sets R_3, $R_2 \cup R_3$, and $R_1 \cup R_2 \cup R_3$ are all trapping regions. Thus, one of the following must hold:

$$\begin{array}{ll} (\exists T) \text{ s.t. } (\forall t \in (T, \infty)) & (x(t), y(t)) \in R_3, \\ (\exists T) \text{ s.t. } (\forall t \in (T, \infty)) & (x(t), y(t)) \in R_2, \\ (\forall t \in (0, \infty)) & (x(t), y(t)) \in R_1. \end{array} \tag{4.90}$$

   In each of the cases, both components of $(x(t), y(t))$ are monotonic functions for large t. *(Why?)* Therefore, they have limits; i.e., $(x(t), y(t))$ tends to some point $\mathbf{b}_*$ in the first quadrant. By the lemma, $\mathbf{b}_*$ must be an equilibrium, and there is only one equilibrium.

   If $(x(0), y(0)) \in R_4 \cup R_5 \cup R_6$, you may truncate these sets as in Section 4.3.2,

$$\tilde{R}_k = R_k \cap \{(x, y) : x + y \le A\}, \qquad k = 4, 5, 6,$$

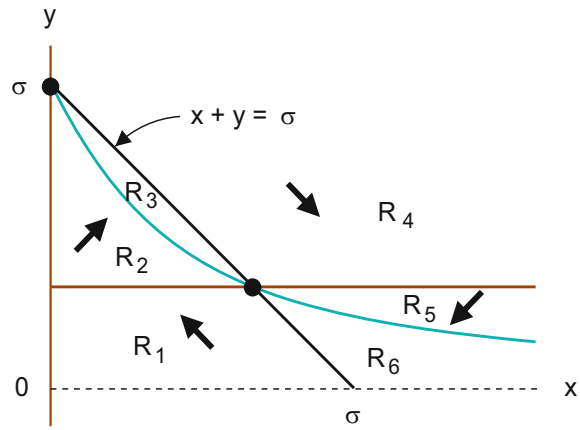

Figure 4.10: *Illustrating the regions described in Exercise 9(b).*

for an appropriate constant A and proceed similarly. Truncation is needed to show that no trajectories escape to infinity.

10. Prove the two claims regarding the order notation made in (4.70).

4.7.2 Applying the Differentiation Theorems

11. *Introduction:* This exercise is intended as a confidence builder; you verify formulas (4.79) and (4.81) in a specific example by explicit solutions. It is also useful preparation for some calculations below.

    (a) Use separability to solve the IVP

    $$\frac{dx}{dt} = \frac{\alpha + \cos t}{x}, \quad x(0) = b, \tag{4.91}$$

    where $b \neq 0$.

    (b) Differentiate your solution to compute $\partial x/\partial b$.

    (c) Define $G(x, t, \alpha) = (\alpha + \cos t)/x$ as in (4.91), calculate DG, write out the IVP (4.79), and show that your answer to Part (b) satisfies this IVP.

    (d) Differentiate your solution to Part (a) to compute $\partial x/\partial \alpha$.

    (e) Write out the IVP (4.81) and show that your answer to Part (d) satisfies this IVP.

    Remark: It might be more consistent to use the flow notation in this problem, but when explicit calculations are involved, we usually find it more intuitive to use x instead of φ. Note that what we are calling x depends on t, b, and α, but we are deliberately sloppy about not indicating all these dependencies explicitly.

12. For the Lotka–Volterra system (4.63), verify formula (4.64) for $\partial\varphi_2/\partial b_2(t,(b_1,0))$.

    Hint: Recall that $\boldsymbol{\varphi}(t,(b_1,0)) = (b_1e^t,0)$. Thus $\partial\boldsymbol{\varphi}/\partial b_j$ satisfies an ODE $\mathbf{w}' = A(t)\mathbf{w}$, where

    $$A(t) = \mathbf{DF}(\boldsymbol{\varphi}(t,(b_1,0))) = \begin{bmatrix} 1 & -b_1e^t \\ 0 & \rho(b_1e^t-1) \end{bmatrix}. \tag{4.92}$$

    As a warmup exercise, calculate from the explicit solution that $\partial\varphi_1/\partial b_1(t,(b_1,0)) = e^t$ and check that $\mathbf{w}(t) = (e^t,0)$ satisfies this ODE with initial condition $\mathbf{w}(0) = (1,0)$. Then solve the ODE with $\mathbf{w}(0) = (0,1)$ and derive (4.64). As a bonus, you may also calculate $\partial\varphi_1/\partial b_2(t,(b_1,0))$, from which you can estimate the effect of a small number of predators on the prey.

13. *Introduction:* Let $x_0(t,\mu), y_0(t,\mu)$ be the solution of the IVP for the torqued pendulum (4.24) subject to initial conditions $x(0) = 0$, $y(0) = 0$, and similarly let $x_\pi(t,\mu), y_\pi(t,\mu)$ be the solution with $x(0) = \pi$, $y(0) = 0$. If $\mu = 0$, these two initial conditions are equilibria for (4.24), so we have

    $$\begin{bmatrix} x_0(t,0) \\ y_0(t,0) \end{bmatrix} \equiv \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} x_\pi(t,0) \\ y_\pi(t,0) \end{bmatrix} \equiv \begin{bmatrix} \pi \\ 0 \end{bmatrix}.$$

    Calculate the partial derivatives of $x_0(t,\mu), y_0(t,\mu)$ and $x_\pi(t,\mu), y_\pi(t,\mu)$ with respect to μ at $\mu = 0$; i.e., determine the ODE (4.81) that these functions of t satisfy and solve the appropriate IVP.

    Food for thought: Solutions of (4.81) for the derivative of x_0, y_0 involve only decaying exponentials, while solutions of (4.81) for the derivative of x_π, y_π may include a growing exponential. How does this different behavior relate to differences between the equilibria $x = 0$ and $x = \pi$?

14. *Introduction:* Let $\boldsymbol{\varphi}(t,\mathbf{b})$ be the solution of the IVP

    $$\begin{aligned} x' &= x - y - (x^2+y^2)x, & x(0) &= b_1, \\ y' &= x + y - (x^2+y^2)y, & y(0) &= b_2. \end{aligned} \tag{4.93}$$

    With the particular initial condition $\mathbf{b}_* = (1,0)$, the IVP has the solution $\boldsymbol{\varphi}(t,\mathbf{b}_*) = (\cos t, \sin t)$, which is periodic. In this exercise we invoke polar coordinates to find the solution of the ODE (4.59) for $\partial\boldsymbol{\varphi}/\partial b_j(t,\mathbf{b}_*)$. The primary challenge of the exercise is calculational; to rephrase this more positively, the exercise offers useful practice with calculations in multivariable calculus.

    (a) Write down the ODE (4.59) for $\partial\boldsymbol{\varphi}/\partial b_j(t,\mathbf{b}_*)$.

    Discussion: Your equation will be a linear system with variable coefficients, and it's not clear how to solve this equation. Let's exploit the fact that (4.93) becomes much simpler if it is written in polar coordinates:

    $$r' = r - r^3, \qquad \theta' = 1. \tag{4.94}$$

Let $\boldsymbol{\varphi}_{\text{polar}}(t, \mathbf{c})$ be the solution of (4.94) with initial conditions $r(0) = c_1$, $\theta(0) = c_2$. (For clarity we'll write $\boldsymbol{\varphi}_{\text{cart}}$—"cart" for Cartesian—for the solution of (4.93).)

We introduce the notation

$$\boldsymbol{\Psi} : (0, \infty) \times \mathbb{R} \to \mathbb{R}^2, \qquad \boldsymbol{\Psi}(r, \theta) = \begin{bmatrix} r\cos\theta \\ r\sin\theta \end{bmatrix}$$

for the transformation from polar to Cartesian coordinates. Note that the inverse $\boldsymbol{\Psi}^{-1}$ is multivalued, but we may define it uniquely in a neighborhood of $\mathbf{b}_*$, with $\boldsymbol{\Psi}^{-1}(\mathbf{b}_*) = (1, 0) = \mathbf{b}_*$. Then for $\mathbf{b}$ near $\mathbf{b}_*$ we have the representation

$$\boldsymbol{\varphi}_{\text{cart}}(t, \mathbf{b}) = \boldsymbol{\Psi} \circ \boldsymbol{\varphi}_{\text{polar}}(t, \boldsymbol{\Psi}^{-1}(\mathbf{b})). \tag{4.95}$$

This equation is valid for all t, even though the orbit circles the origin multiple times. We will differentiate (4.95) with the chain rule. The formulas will be simpler with a bit of notation: let $\mathbf{D}\boldsymbol{\varphi}_{\text{cart}}$ be the 2×2 matrix with columns $\partial\boldsymbol{\varphi}_{\text{cart}}/\partial b_j$ and define $\mathbf{D}\boldsymbol{\varphi}_{\text{polar}}$ similarly with columns $\partial\boldsymbol{\varphi}_{\text{polar}}/\partial c_j$.

(b) Show that

$$\mathbf{D}\boldsymbol{\varphi}_{\text{cart}}(t, \mathbf{b}_*) = \mathbf{D}\boldsymbol{\Psi}(\boldsymbol{\varphi}_{\text{polar}}(t, \mathbf{b}_*)) \cdot \mathbf{D}\boldsymbol{\varphi}_{\text{polar}}(t, \mathbf{b}_*), \tag{4.96}$$

where of course

$$\mathbf{D}\boldsymbol{\Psi}(r, \theta) = \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix}.$$

Remark: From the chain rule you would expect a third factor on the right in (4.96) from differentiation of $\boldsymbol{\Psi}^{-1}$, but $\mathbf{D}\boldsymbol{\Psi}^{-1}(\mathbf{b}_*)$ equals the identity matrix.

(c) Using the fact that $\boldsymbol{\varphi}_{\text{polar}}(t, \mathbf{b}_*) = (1, t)$, apply (4.59) to show that

$$\mathbf{D}\boldsymbol{\varphi}_{\text{polar}}(t, \mathbf{b}_*) = e^{tA}, \qquad \text{where } A = \begin{bmatrix} -2 & 0 \\ 0 & 0 \end{bmatrix}.$$

(d) Calculate $\mathbf{D}\boldsymbol{\varphi}_{\text{cart}}(t, \mathbf{b}_*)$ from (4.96).

(e) Verify that the columns of $\mathbf{D}\boldsymbol{\varphi}_{\text{cart}}(t, \mathbf{b}_*)$, i.e., $\partial\boldsymbol{\varphi}_{\text{cart}}/\partial b_j$, satisfy your equation in Part (a).

4.7.3 Some Mopping-Up Exercises

15. Prove the comparison result from Chapter 1, Theorem 1.8.1.

Hint: Make an autonomous system out of the ODE for x in the theorem,

$$\begin{bmatrix} x_1' \\ x_2' \end{bmatrix} = \begin{bmatrix} f(x_1, x_2) \\ 1 \end{bmatrix}.$$

Use the function $u(t)$ to define a set

$$\mathcal{K} = \{(x_1, x_2) \in \mathbb{R}^2 : u(x_2) \leq x_1\}.$$

Show that $\mathcal{K}$ is a trapping region for this system.

16. In the context of Corollary 3.2.8, strengthen that result by showing that the IVP is solvable on $(-\eta, \eta)$, provided merely that $\eta < \delta/K$.

Hint: Imitate the proof of Theorem 4.2.1.

17. Rewrite the equation $\varepsilon x'' + x' + x = 0$ as a first-order system, draw the nullclines, and make a flow-quadrant diagram, observing the double-arrow convention (cf. Figure 4.9) for the fast-flow direction. Argue from your figure that after a brief transient, a typical trajectory hugs the y-nullcline as it decays to zero.

Discussion: In Exercise 1.14, working with explicit solutions, you showed that the approximation of setting $\varepsilon = 0$ in this equation gives decent results. Using nullclines you can understand geometrically why the approximation works. For a more complete understanding of these issues, repeat the exercise for $\varepsilon x'' - x' - x = 0$.

4.7.4 Computing Exercise

18. In a programming language of your choosing, apply Euler's method (which is introduced in the appendix of this chapter) to solve approximately the IVP for (4.93) with $\mathbf{b} = (0.1, 0)$, say for $0 \leq t \leq 0.5$. Choose various mesh sizes $h = 10^{-n/2}$, $n = 2, 3, \ldots, 10$. Compare your calculations with the exact solution by making a log-log plot of the errors in $x(0.5)$ and $y(0.5)$ as a function of h over this range.

Remark: You will find that, as suggested by Theorem 4.9.1, the error in Euler's method is roughly proportional to h as the mesh size tends to zero.

The simplest improvement over Euler's method is the unimaginatively named "improved Euler method." In this method also, one computes approximations $\mathbf{y}_n$ to $\mathbf{x}(nh)$, the solution of an IVP at integer multiples of the step size. In the improved Euler method, advancing to the next approximation is a two-step process: given $\mathbf{y}_n$, let

$$\text{(a) } \mathbf{y}_{n+1/2} = \mathbf{y}_n + (h/2)\mathbf{F}(\mathbf{y}_n), \qquad \text{(b) } \mathbf{y}_{n+1} = \mathbf{y}_n + h\mathbf{F}(\mathbf{y}_{n+1/2}). \tag{4.97}$$

The only change from Euler's method is that $\mathbf{y}_{n+1/2}$, the crude initial estimate for the solution at the intermediate time $(n + 1/2)h$, is used in (4.97b). (After substitution into (4.97b), $\mathbf{y}_{n+1/2}$ is discarded.) Remarkably, the simple fudge (4.97) gives a more accurate method; as $h \to 0$, the error tends to zero as h^2. We invite you to verify this claim on the above example with your own computations.

4.7.5 PHD Exercises

19. In the context of Theorem 4.2.2, show that if a compact trapping region $\mathcal{K}$ satisfies (4.12), the distance from the solution $\mathbf{x}(t)$ to $\partial\mathcal{K}$ remains bounded away from zero as $t \to \infty$.

    Advice: You might as well assume that $\partial\mathcal{K}$ is smooth; handling the more general case of a piecewise smooth boundary would add only technical complications.

20. Prove global existence for Duffing's equation if the sign of friction is reversed,

$$\begin{aligned} x' &= y, \\ y' &= +\beta y + x - x^3, \end{aligned}$$

    where $\beta > 0$.

    Hint: You can't get anywhere on this problem using trapping regions, because the solution grows without bound. Likewise, Theorem 4.2.1 is useless, because the cubic term in the force violates (4.4). Here's a strategy that works: Calculate the rate at which energy grows. Manipulate your result to show that $dE/dt \leq \beta E$, provided (x, y) lies outside some large circle, say $x^2 + y^2 > C^2$. (In fact, $C = 2$ is sufficient.) Then extract your conclusion from this information.

21. Use the order notation to make a completely rigorous proof out of your heuristic proof from Exercise 9 of Chapter 3 that (in the notation of that problem)

$$\phi'(t) - [\mathrm{tr}A(t)]\,\phi = 0.$$

22. *Introduction:* The equations

$$\begin{aligned} x' &= x - xy, \\ y' &= \rho(\sigma + xy - y), \end{aligned} \tag{4.98}$$

    where ρ and σ are positive parameters, are like the Lotka–Volterra equations, except that even without predation, *new predators appear at a small background rate* (normalized to $\rho\sigma$). This assumption is pretty hokey when applied to foxes and rabbits, but variants of these equations arise in certain models for the evolution of a viral infection; see [62] for details. In this application, x measures the total population of virus cells in the patient's body, while y represents the body's immune mechanism; specifically, y measures the population of what are called *effector* cells.

    (a) Interpret in words each term in these equations.

    (b) Draw nullclines for these equations.

    (c) Use the nullclines to construct trapping regions to prove global existence in forward time for the IVP, assuming initial data in the first quadrant.

    Remarks: Setting $\sigma = 0$ in (4.98) yields the Lotka–Volterra equations (4.63). Recall that all orbits of (4.63) are periodic. Thus, the only trapping regions for

Lotka–Volterra are regions bounded by the periodic orbits themselves. While we were able to find explicit formulas for the orbits for the Lotka–Volterra equations, we can't do likewise for (4.98).

Hint: Try to construct a trapping region for (4.98) bounded in part by an orbit of the Lotka–Volterra equations, or at least something close to an orbit. Tolerances are small, and you have to be careful in carrying out the construction. If you find it helpful, assume that σ is small.

23. (a) Use software to solve the IVP

$$\begin{array}{lrcll} \text{(a)} & x' & = & y, & x(0)=2, \\ \text{(b)} & y' & = & -x-zy, & y(0)=0, \\ \text{(c)} & \varepsilon z' & = & -\left(z-x^2+1\right), & z(0)=1, \end{array} \tag{4.99}$$

for $t \in [0,5]$, say for $\varepsilon = 10^{-3}, 10^{-6}, 10^{-9}$.

Warning: Depending on what method you choose, you may encounter trouble for very small ε.

(b) Show that the fast–slow approximation of setting $\varepsilon = 0$ reduces the three-dimensional problem to the van der Pol system.

(c) Compare your solution in Part (a) (in cases in which you were able to get a solution) with solutions of the van der Pol equation.

Remark: In the language of Section 10.3, (4.99) is a *stiff* ODE.

24. *Introduction:* The differentiability of the solution of the IVP (4.80) with respect to the parameters can be proved with minimal effort using a trick of augmenting the system with m "fake" variables corresponding to the parameters $\alpha_1, \ldots, \alpha_m$. Specifically consider an auxiliary variable $\mathbf{y} \in \mathbb{R}^m$, and let $(\mathbf{x}, \mathbf{y})$ evolve according to the system

$$\begin{array}{rcll} \mathbf{x}' & = & \mathbf{G}(\mathbf{x},\mathbf{y}), & \mathbf{x}(0)=\mathbf{b}, \\ \mathbf{y}' & = & \mathbf{0}, & \mathbf{y}(0)=\boldsymbol{\alpha}. \end{array}$$

Apply Theorem 4.6.1 to this system to show that the solution of (4.80) is continuously differentiable with respect to α and to derive (4.81).

25. (a) *Introduction:* If $\mathbf{F} \in \mathcal{C}^1(\mathcal{U})$, let $\Delta(\mathbf{z}, \mathbf{z}_0) = \mathbf{F}(\mathbf{z}) - \mathbf{F}(\mathbf{z}_0) - \mathbf{DF}(\mathbf{z}_0) \cdot (\mathbf{z} - \mathbf{z}_0)$ be the error in the first-order Taylor series approximation for $\mathbf{F}(\mathbf{z})$ based at $\mathbf{z}_0$. In this notation, (4.69) may be rephrased as

$$\Delta(\mathbf{z}, \mathbf{z}_0) = o(|\mathbf{z} - \mathbf{z}_0|).$$

Given a compact set $\mathcal{K} \subset \mathcal{U}$, show that the above estimate is uniform for $\mathbf{z}_0 \in \mathcal{K}$; i.e., show that for every $\eta > 0$, there is a $\delta > 0$ such that for all $\mathbf{z}_0 \in \mathcal{K}$ and all $\mathbf{z}$ with $|\mathbf{z} - \mathbf{z}_0| < \delta$,

$$|\Delta(\mathbf{z}, \mathbf{z}_0)| \leq \eta |\mathbf{z} - \mathbf{z}_0|. \tag{4.100}$$

Hint: By Corollary 3.3.3, there exist a larger compact set $\mathcal{K}' \subset \mathcal{U}$ and a δ_0 such that for every $\mathbf{z}_0 \in \mathcal{K}$, the ball $\overline{B(\mathbf{z}_0, \delta_0)}$ is contained in $\mathcal{K}'$. Apply calculus to conclude that if $|\mathbf{z} - \mathbf{z}_0| < \delta_0$, then

$$\Delta(\mathbf{z}, \mathbf{z}_0) = \left\{ \int_0^1 \left[\mathbf{DF}(\mathbf{z}_0 + s(\mathbf{z} - \mathbf{z}_0)) - \mathbf{DF}(\mathbf{z}_0) \right] ds \right\} \cdot (\mathbf{z} - \mathbf{z}_0).$$

Use the fact that $\mathbf{DF}$ is uniformly continuous on $\mathcal{K}'$ to complete the argument.

(b) Prove Lemma 4.6.2.

Hint: Since $A(s) = \mathbf{DF}(\mathbf{x}_0(s))$, the definition (4.74) of $\mathcal{I}_1(t, \varepsilon)$ may be rewritten as

$$\mathcal{I}_1(t, \varepsilon) = \int_0^t \Delta(\mathbf{x}(s, \varepsilon), \mathbf{x}_0(s)) ds.$$

Apply Part (a) with $\mathcal{K}$ equal to the image of $[0, T]$ under the base solution $\mathbf{x}_0$, $\mathbf{z} = \mathbf{x}(s, \varepsilon)$, and $\mathbf{z}_0 = \mathbf{x}_0(s)$; specifically, show that

$$\sup_{0 \le s \le T} |\Delta(\mathbf{x}(s, \varepsilon), \mathbf{x}_0(s))| = o(\phi(\varepsilon)),$$

where

$$\phi(\varepsilon) = \sup_{0 \le s \le T} |\mathbf{x}(s, \varepsilon) - \mathbf{x}_0(s)|. \tag{4.101}$$

Show that $\phi(\varepsilon) = \mathcal{O}(\varepsilon)$ and invoke (4.70) to finish the proof.

4.8 Pearls of Wisdom

While Theorem 4.5.1 gives control over the solution of an IVP for every finite time, infinite times are beyond our reach: the limit of $\boldsymbol{\varphi}(t, \mathbf{b})$ as $t \to \infty$ may be discontinuous in $\mathbf{b}$. For example, the solution of the IVP for Duffing's equation

$$\begin{aligned} x' &= y, & x(0) &= b, \\ y' &= -\beta y + x - x^3, & y(0) &= 0, \end{aligned}$$

converges to $(1, 0)$ as $t \to \infty$ if $b > 0$ and to $(-1, 0)$ if $b < 0$, at least provided that $|b|$ is not too large. (You could probably prove this now; in any case, you will see it proved in Chapter 6.)

As we observed in Section 4.6.3 (but it bears repeating), if $\mathbf{b}_0$ is an equilibrium of $\mathbf{x}' = \mathbf{F}(\mathbf{x})$, we may approximate $\boldsymbol{\varphi}(t, \mathbf{b}_0 + \varepsilon \mathbf{b}_1)$, the solution of an IVP with initial conditions near $\mathbf{b}_0$, by $\mathbf{b}_0 + \varepsilon \mathbf{w}(t)$, where $\mathbf{w}(t)$ solves the linear constant-coefficient IVP

$$\mathbf{w}' = A\mathbf{w}, \qquad \mathbf{w}(0) = \mathbf{b}_1$$

with the coefficient matrix $A = \mathbf{DF}(\mathbf{b}_0)$. This approximation, known as the *linearization* of $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ at the equilibrium, typically determines the qualitative behavior of solutions of the full nonlinear equation near the equilibrium. (Cf. Chapter 6.)

There is an efficient procedure for handling the linearized equations (4.59) in numerical solutions. Given a d-dimensional autonomous ODE $\mathbf{x}' = \mathbf{F}(\mathbf{x})$, consider the greatly enlarged system of dimension $d + d^2$ with unknowns the d-dimensional vector $\mathbf{x}(t)$ and a $d \times d$ matrix $X(t)$:

$$\begin{array}{rcll} \mathbf{x}' & = & \mathbf{F}(\mathbf{x}), & \mathbf{x}(0) = \mathbf{b}, \\ X' & = & \mathbf{DF}(\mathbf{x})X, & X(0) = I, \end{array} \tag{4.102}$$

where I is the identity matrix. Then the jth column of $X(t)$ reproduces (4.59); thus, the jth column of $X(t)$ equals $\partial\boldsymbol{\varphi}/\partial b_j(t, \mathbf{b})$. To conclude: although in discussing the theory, it is more natural to first solve the ODE and then differentiate with respect to initial conditions, in computations it works better to attack both issues at the same time.

4.9 Appendix: Euler's Method

4.9.1 Introduction

Since it is rarely possible to produce explicit solutions of ODEs, we often resort to *numerical methods*[25] in order to obtain approximate solutions. In this section we introduce the simplest numerical method, known as *Euler's method*. We do not propose to actually use this method as a practical source of information about solutions of ODEs, since software employs methods that are far more accurate than this, and their automated control of step size makes them a joy to use. Rather, we study Euler's method for cultural reasons, namely, it provides useful insight into numerical methods in general, and its simplicity allows the conceptual issues to come through more easily.

Euler's method is an iterative process for approximating the solution of an IVP, say

$$\mathbf{x}' = \mathbf{F}(\mathbf{x}), \qquad \mathbf{x}(0) = \mathbf{b}. \tag{4.103}$$

For simplicity, let's consider a set of evenly spaced t-values: given $h > 0$, we calculate the approximations $\mathbf{y}_n$ for $\mathbf{x}(nh)$ recursively according to the rule

$$\mathbf{y}_0 = \mathbf{b}; \qquad \mathbf{y}_{n+1} = \mathbf{y}_n + h\mathbf{F}(\mathbf{y}_n), \quad n = 0, 1, \ldots . \tag{4.104}$$

Although $\mathbf{y}_n$ depends on h, here we follow the usual convention of not indicating this dependence.

As an illustration, consider the time-honored scalar IVP $x' = x$, $x(0) = 1$, which has exact solution $x(t) = e^t$. Let's approximate the solution on the interval $t \in [0, 1]$

[25] Perturbation methods, some of which are discussed in Sections 7.5 and 7.6 (among other places), offer another valuable way of approximating solutions.

with Euler's method, say using a step size of $h = 1/N$, where N is a positive integer. Starting from $y_0 = 1$, we use (4.104) to generate the subsequent iterates recursively:

$$y_{n+1} = y_n + hy_n = \left(1 + \frac{1}{N}\right) y_n, \qquad n = 0, 1, 2, \ldots,$$

so $y_n = (1 + 1/N)^n$. To test the accuracy of the approximation, consider y_n with $n = N$, which should approximate x at time $t = Nh = 1$: as $h \to 0$ (and thus $N \to \infty$), we have

$$y_N = (1 + 1/N)^N \to e = x(1),$$

as desired.

More generally, y_n provides an approximation for e^t for all t, but the formulation of this behavior is made awkward by two issues: (i) the number of steps needed to reach time t scales up as $1/h$ as $h \to 0$ and (ii) any specific time t need not belong to the set of grid points $\{nh : n = 0, 1, \ldots\}$ for which the approximations are computed. Thus, on the time interval $0 \le t \le T$, the convergence result for this example takes the somewhat clumsy form

$$\lim_{h \to 0} \max_{0 \le n \le T/h} |e^{nh} - y_n| = 0.$$

A convergence result for the general case is given in Theorem 4.9.1 below.

4.9.2 Theoretical Basis for the Approximation

We offer three motivations[26] for Euler's method. All three motivations begin with the limited goal of understanding the first step in (4.104), which we may rephrase as

$$x(h) \approx x(0) + hF(x(0)). \tag{4.105}$$

(For simplicity, we assume temporarily that we are solving a scalar equation.)

Motivation 1: *(Tangent line)* Interpreting the derivative geometrically (see Figure 4.11), we see from the ODE that the slope of the solution curve through $(0, x(0))$ equals $F(x(0))$. Thus, we may estimate $x(h)$ by following the tangent line, resulting in the approximation (4.105).

Motivation 2: *(Finite differences)* Using the difference quotient approximation

$$\frac{x(h) - x(0)}{h} \approx x'(0) = F(x(0)),$$

[26]For Euler's method, all three motivations produce the same formula, but for advanced numerical methods, different approximation formulas may result from starting with one or another of these three points of view.

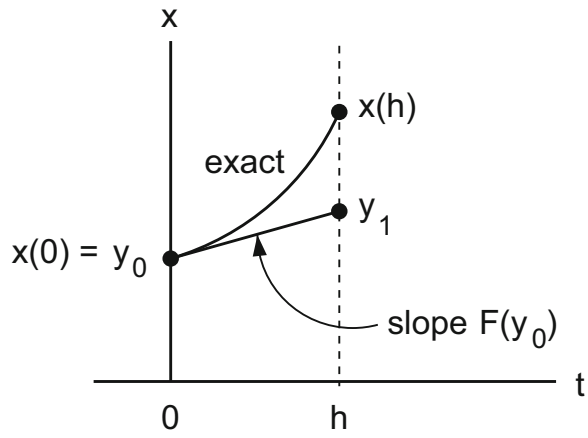

Figure 4.11: *Schematic illustration of one iteration of Euler's method. Note the discrepancy between the exact solution $x(h)$ and its Euler's method approximation y_1.*

we again obtain (4.105).

Motivation 3: *(Integral equation)* Reformulating the IVP as an integral equation,

$$x(h) \;=\; x(0) + \int_0^h F(x(s))\, ds,$$

we derive (4.105) from a one-term Riemann-sum approximation for the integral.

The continuation of Euler's method may seem like an act of desperation. It is extremely unlikely that the point (h, y_1) will lie on the exact solution curve (see Figure 4.11). Nevertheless, it is the best information we have about the solution. Therefore, we will use that point as the starting point for another iteration of Euler's method; i.e., we let y_2 equal the Euler approximation to the solution of $x' = F(x)$ through the point (h, y_1). All subsequent steps are derived similarly. One may well wonder about an approximation in which each step is based on increasingly faulty information, especially since as $h \to 0$, more and more steps are required to advance a finite time. However, as we show in the next subsection, the accumulated error in the numerical solution actually tends to zero with h.

4.9.3 Convergence of the Numerical Solution

If $\mathbf{F}$ is defined everywhere, then the definition (4.104) of $\mathbf{y}_n$ remains meaningful for arbitrarily large n, even if the solution $\mathbf{x}$ that is being approximated blows up in finite time. However, if $\mathbf{F}$ is defined only on a subset $\mathcal{U} \subset \mathbb{R}^d$, then some iterate $\mathbf{y}_n$ may lie outside $\mathcal{U}$, so the iteration would halt. This possibility is addressed in Conclusion (i) of the following theorem, which has much in common with Theorem 4.5.1.

Theorem 4.9.1. *Suppose $\mathbf{F} : \mathcal{U} \to \mathbb{R}^d$ is locally Lipschitz, and let $\mathbf{x}(t)$, $0 \le t < \beta$, be a solution in forward time of $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ with initial condition $\mathbf{x}(0) = \mathbf{b}$. (i) For every positive $T < \beta$, there exists a positive constant h_0 such that if $h < h_0$, then the iterates $\mathbf{y}_n$ are defined for all n such that $nh \le T$. (ii) There are constants C, L such that if $h < h_0$, then*

$$|\mathbf{x}(nh) - \mathbf{y}_n| \le Che^{Lnh}, \qquad \text{for } 0 \le nh \le T.$$

Remark: Note that Conclusion (ii) implies the uniform error estimate

$$|\mathbf{x}(nh) - \mathbf{y}_n| \le Che^{LT}.$$

Proof. Choose a compact subset $\mathcal{K} \subset \mathcal{U}$ and a constant $\delta > 0$ such that

$$(\forall t \in [0, T]) \ \ \overline{B(\mathbf{x}(t), \delta)} \subset \mathcal{K} \subset \mathcal{U}.$$

We define the constants: let $C = \max_{\mathcal{K}} |\mathbf{F}|$, let L be a Lipschitz constant for $\mathbf{F}|\mathcal{K}$, and let $h_0 = e^{-LT}\delta/C$. We compute $\mathbf{y}_n$ for as many iterations as $nh \le T$ and $\mathbf{y}_n \in \mathcal{K}$, say $n \le N$. Note that $\mathbf{y}_N \in \mathcal{K}$, so that it is possible to calculate at least one more iterate, $\mathbf{y}_{N+1}$.

The solution $\mathbf{x}$ satisfies the integral equation

$$\mathbf{x}(t) = \mathbf{b} + \int_0^t \mathbf{F}(\mathbf{x}(s))\, ds.$$

In order to derive an analogous integral equation for the approximate solution, we construct a piecewise constant function on $[0, (N+1)h]$ as follows: for $t < (N+1)h$, let

$$\mathbf{y}^{(h)}(t) = \mathbf{y}_n \quad \text{for} \quad nh \le t < (n+1)h,$$

and at the right-hand endpoint define $\mathbf{y}^{(h)}((N+1)h) = \mathbf{y}_{N+1}$. Note that

$$\int_{nh}^{(n+1)h} \mathbf{F}(\mathbf{y}^{(h)}(s))\, ds = h\mathbf{F}(\mathbf{y}_n),$$

the integrand being constant. Therefore, at the grid points,

$$\mathbf{y}^{(h)}(nh) = \mathbf{b} + \int_0^{nh} \mathbf{F}(\mathbf{y}^{(h)}(s))\, ds\,, \qquad n = 0, 1, 2, \ldots, N+1.$$

Moving off grid points, we obtain the desired integral equation

$$\mathbf{y}^{(h)}(t) = \mathbf{b} + \int_0^t \mathbf{F}(\mathbf{y}^{(h)}(s))\, ds - \int_{nh}^t \mathbf{F}(\mathbf{y}^{(h)}(s))\, ds,$$

where n is the largest integer such that $nh \le t$.

For $0 \le t \le \min\{T, (N+1)h\}$, let $g(t) = |\mathbf{x}(t) - \mathbf{y}^{(h)}(t)|$. Subtracting the integral equations for $\mathbf{x}(t)$ and $\mathbf{y}^{(h)}(t)$, we deduce that

$$g(t) \le \int_0^t |\mathbf{F}(\mathbf{x}(s)) - \mathbf{F}(\mathbf{y}^{(h)}(s))|\, ds + \int_{nh}^t |\mathbf{F}(\mathbf{y}^{(h)}(s))|\, ds,$$

where again n is the largest integer such that $nh \le t$. Note that in the integrands, we have $\mathbf{x}(s)$, $\mathbf{y}^{(h)}(s) \in \mathcal{K}$. By Lipschitz continuity, the first term here satisfies

$$\int_0^t |\mathbf{F}(\mathbf{x}(s)) - \mathbf{F}(\mathbf{y}^{(h)}(s))|\, ds \le L \int_0^t g(s) ds,$$

and by the definition of C, the second satisfies

$$\int_{nh}^t |\mathbf{F}(\mathbf{y}^{(h)}(s))|\, ds \le Ch.$$

Thus, by Gronwall's inequality (extended to piecewise continuous functions),

$$g(t) \le Che^{Lt}, \qquad 0 \le t \le \min\{T, (N+1)h\}. \tag{4.106}$$

Regarding Conclusion (i): If $(N+1)h < T$, then taking $t = (N+1)h$ in (4.106), we see that

$$|\mathbf{y}_{N+1} - \mathbf{x}((N+1)h)| \le Che^{L(N+1)h} < Che^{LT} < \delta,$$

so $\mathbf{y}_{N+1} \in \mathcal{K}$; i.e, the iteration would continue. Inequality (4.106) verifies Conclusion (ii). □

By (4.106), the errors produced by Euler's method may be expected to be on the order of h to the first power. This apparently good news is actually bad news. Contrast this error estimate with, for example, the error in the fourth-order Runge–Kutta algorithm (RK4), which is of order h^4. Thus, if the step size is halved, the error in Euler's method is merely halved, while the error in the RK4 method[27] is decreased by a factor of 16. To achieve high accuracy with Euler's method, the step size must be chosen painfully small; this wastes computational time and also raises round-off issues [65].

Much effort has gone into devising highly accurate numerical methods for solving ODEs. In Exercise 18, we describe one simple improvement over Euler's method, but for serious further study see the cautionary examples in Section 10.3 and the references in that section.

[27] Typically, in software h is chosen automatically, so the convergence rate is not readily apparent to the user.

http://www.springer.com/978-1-4939-6387-4

Ordinary Differential Equations: Basics and Beyond
Schaeffer, D.; Cain, J.W.
2016, XXX, 542 p. 139 illus., 61 illus. in color.,
Hardcover
ISBN: 978-1-4939-6387-4