5 Stability Theory

Without stability there is no rest and no peace in this life.
—Marsha Carol Watson Gandy

Not all models using dynamical systems are well behaved, and not all solutions of dynamical systems are well behaved, and even when the model or the solution is well behaved, some algorithms we try to use to solve them numerically might not be well behaved. We generally need our models, our solutions, and our numerical algorithms to all be *stable* if we want to get meaningful results.

There are several types of stability that matter when working with dynamical systems. Two types of stability (*stable equilibrium* and *asymptotically stable equilibrium*) have to do with an *equilibrium point* of an autonomous system $\dot{\mathbf{x}} = f(\mathbf{x}(t))$, that is, a point $\overline{\mathbf{x}}$ with the property that $f(\overline{\mathbf{x}}) = 0$. This defines a constant function $\mathbf{x}(t) = \overline{\mathbf{x}}$ for all t, which we (confusingly) also write as $\overline{\mathbf{x}}$. Such a function $\overline{\mathbf{x}}$ is always a solution of the autonomous system. An equilibrium that is stable or asymptotically stable is one that other solutions stay close to or get close to. Stable equilibria are essentially independent of initial conditions, if we wait long enough.

Sometimes a solution does not settle down to a single value, but cycles around in some periodic or cyclic way. Such solutions are called *limit cycles*. All nearby initial conditions will move onto this pattern as $t \to \infty$, but the underlying structure will still capture the cyclical behavior we saw in the predator–prey dynamics. These limit cycles can be either stable (as indicated in Figure 5.1) or unstable, in which case we would never expect to see the solution either numerically or physically.

Another important type of stability discussed in this chapter has to do with the stability of models: *structurally stable dynamical system*. A structurally stable dynamical system is one whose fundamental qualitative dynamics don't change due to small changes in the model itself. Section 1.2.5 shows how the Lotka–Volterra system is not structurally stable.

The last type of stability we are interested in is *numerical stability*, which is all about numerical algorithms for approximating solutions of a dynamical system. This is completely different from the stability of the system. Numerical stability is about how well the algorithm approximates solutions of the system, whereas stability of the system is about how the true solutions behave.

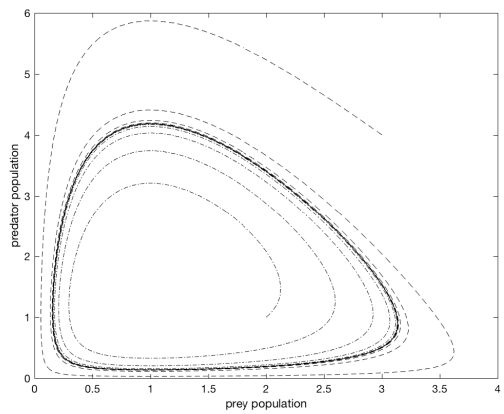


Figure 5.1: The evolution of the predator and prey populations for a variety of initial conditions for the Lotka–Volterra third-order model (see Section 5.6)

5.1 Stability of Linear Systems and Equilibria

This section begins by formally quantifying stability for linear systems and for equilibrium solutions of autonomous systems. The results for linear systems give information about nonlinear systems through the tool of *linearization*, that is local approximation of a nonlinear system by a linear system. Throughout the section we focus on autonomous systems because, as noted earlier, non-autonomous systems can always be recast as autonomous ones.

5.1.1 Stability of Linear Operators

Before getting to the formal quantified definitions of stability and asymptotic stability for ODEs, let's consider the case of a homogeneous linear system $\dot{\mathbf{x}} = A\mathbf{x}$. The origin $\mathbf{0}$ is a special point for this ODE because the constant function $\mathbf{x}(t) = \mathbf{0}$ is a solution of $\dot{\mathbf{x}} = A\mathbf{x}$ with initial value $\mathbf{0}$. We call the point $\mathbf{0}$ (and the corresponding constant function $\mathbf{x}(t) = \mathbf{0}$) an *equilibrium* of the system. No other constant function has this property if A is nonsingular, because a constant solution must have $\mathbf{0} = \dot{\mathbf{x}}(t) = A\mathbf{x}$, and $\mathbf{0}$ is the only solution to this equation for nonsingular A. Therefore, no other point is an equilibrium of the linear system—any solution $\mathbf{x}(t)$ passing through a nonequilibrium point cannot remain at that point because it's derivative is nonzero.

As the next example illustrates, the long-term behavior of the solutions of the linear system $\dot{\mathbf{x}} = A\mathbf{x}$ depends only on the spectrum of the linear operator A.

Example 5.1.1 (2D linear systems classification). Consider the generic 2D linear system:

$$\dot{\mathbf{x}} = A\mathbf{x}, \qquad \text{where } A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

The long-term behavior of the system depends on just a few main properties of the eigenvalues λ_1, λ_2 of A. Here we consider the main possibilities in the generic setting that $\lambda_1 \neq \lambda_2$.

(i) If both eigenvalues are real and $\lambda_1,\ \lambda_2 < 0$, then all solutions converge toward the origin as $t \to \infty$. In this case we say that the origin is an *asymptotically stable equilibrium*); see also Theorem 5.1.2.

(ii) If both eigenvalues are real and $\lambda_1 < 0 < \lambda_2$, then there is a stable direction along the eigenspace of λ_1 and an unstable direction along the eigenspace of λ_2. The origin is an equilibrium, but it is not stable—almost all solutions that pass through nearby points don't converge to $\mathbf{0}$ as $t \to \infty$, and, worse, they don't even stay near $\mathbf{0}$.

(iii) If both eigenvalues are real and $\lambda_1,\ \lambda_2 > 0$, then the origin is still an equilibrium, but it is unstable in all directions: all solutions except the constant solution $\mathbf{0}$ tend to infinity in all directions as $t \to \infty$.

(iv) When the eigenvalues are both imaginary, this is the case of the undamped harmonic oscillator (when $b = 0$ in the previous example), then the solutions are time periodic, orbiting around the origin. In some sense the origin is stable, because solutions that start near $\mathbf{0}$ stay near $\mathbf{0}$. But it is not *asymptotically stable*, because solutions do not converge to the origin—they orbit around it, instead.

(v) If the eigenvalues are complex with positive real part, then the origin is an unstable equilibrium. All nonzero solutions form an unstable spiral, with oscillatory behavior but ever-growing amplitude.

(vi) If the eigenvalues are complex with with negative real part, then the origin is a stable equilibrium. All solutions spiral inward to $\mathbf{0}$, with oscillatory behavior and decreasing amplitude. This is the case for the damped oscillator with $\omega^2 > \frac{b^2}{4m^2}$.

A phase plot for each of the different cases is given in Figure 5.3, and the evolution of each coordinate over time is given in Figure 5.2.

The previous example suggests that if the eigenvalues of A all have negative real part, then all solutions of the homogeneous linear system $\dot{\mathbf{x}} = A\mathbf{x}$ approach $\mathbf{0}$ as $t \to \infty$. This is a corollary (Corollary 5.1.5) of the Theorem 5.1.2 below.

Theorem 5.1.2. *For $A \in M_n(\mathbb{C})$, if the real part $\Re(\lambda)$ of every eigenvalue λ of A is negative, then there exist $\eta > 0$ and $C > 0$ such*

$$\|\exp(tA)\| \leq C\exp(-\eta t) \quad \textit{for all } t \geq 0.$$

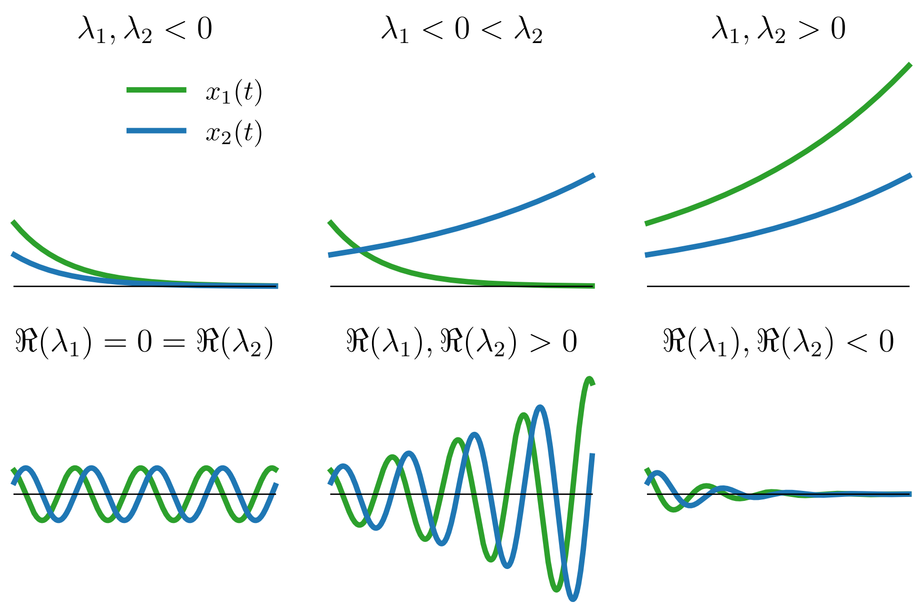


Figure 5.2: Time evolution of each of the six different cases of generic two-dimensional linear systems, as discussed in Example 5.1.1. When the real part of an eigenvalue is positive, it leads to exponential growth and the system moves rapidly away from the origin in the direction of the corresponding eigenvector.

Proof. For each $t \in [0, \infty)$ the function $f(tz) = \exp(tz)$ is entire (holomorphic on the whole complex plane). Since there are only finitely many eigenvalues, there exists $\eta > 0$ such that $\Re(\lambda) \leq -\eta$ for all eigenvalues λ of A. Let Γ be a contour lying strictly in the left half of $\mathbb{C}$ and containing $\sigma(A)$; see Figure 5.4.

The Spectral Resolution Theorem (Theorem 4.2.2) guarantees that

$$\exp(tA) = f(A) = \frac{1}{2\pi i} \oint_\Gamma f(z) R_A(z)\, dz = \frac{1}{2\pi i} \oint_\Gamma \exp(tz) R_A(z)\, dz,$$

for each $t \in [0, \infty)$. On the contour Γ, letting $z = x + iy$ gives

$$|\exp(zt)| = |\exp((x+iy)t)| = |\exp(xt)(\cos(yt)+i\sin(yt))| = |\exp(xt)| \leq \exp(-\eta t).$$

Set

$$C = \frac{1}{2\pi} \oint_\Gamma \|R_A(z)\|\, |dz| > 0.$$

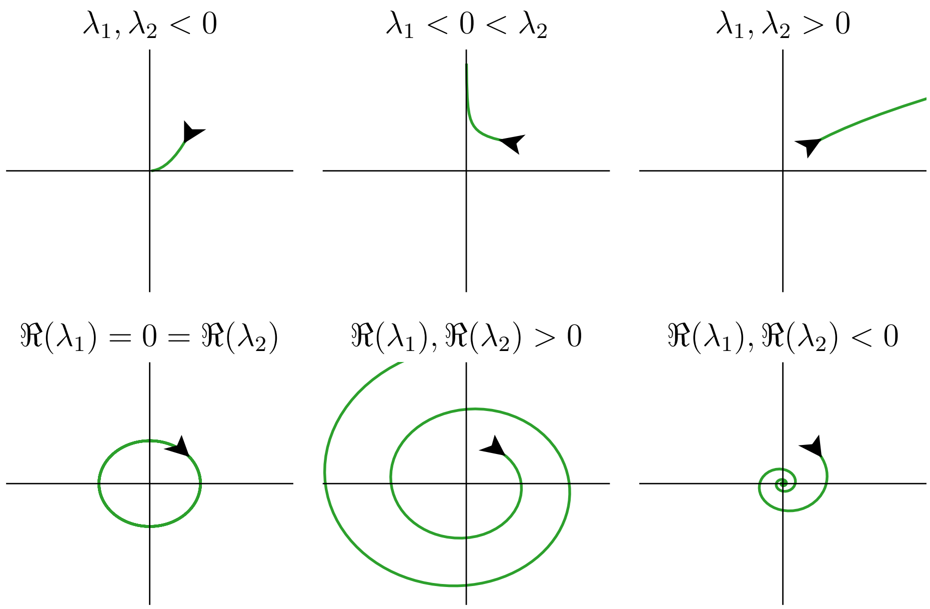


Figure 5.3: Plots in the phase plane of each of the six different cases of generic two-dimensional linear systems, as discussed in Example 5.1.1. In each case the system starts at the black arrowhead and proceeds along the curve. When the real part of at least one eigenvalue is positive, it leads to exponential growth, and the system moves rapidly away from the origin in the direction of the corresponding eigenvector. If all the eigenvalues have negative real part, then the solutions converge to the origin.

Then

$$
\begin{aligned}
\|\exp(tA)\| &= \left\| \frac{1}{2\pi i} \oint_\Gamma \exp(tz) R_A(z)\ dz \right\| \\
&\leq \frac{1}{2\pi} \oint_\Gamma |\exp(tz)|\ \|R_A(z)\|\ |dz| \\
&\leq \frac{1}{2\pi} \oint_\Gamma \exp(-\eta t) \|R_A(z)\|\ |dz| \\
&= \frac{\exp(-\eta t)}{2\pi} \oint_\Gamma \|R_A(z)\|\ |dz| \\
&= C \exp(-\eta t). \qquad \square
\end{aligned}
$$

Definition 5.1.3. *A linear operator A is* exponentially stable *when there exists $\eta > 0$ such that $\Re(\lambda) \leq -\eta$ for all eigenvalues λ of A. The largest such η is called the* spectral gap. *It is the distance between the spectrum of A and the imaginary axis.*

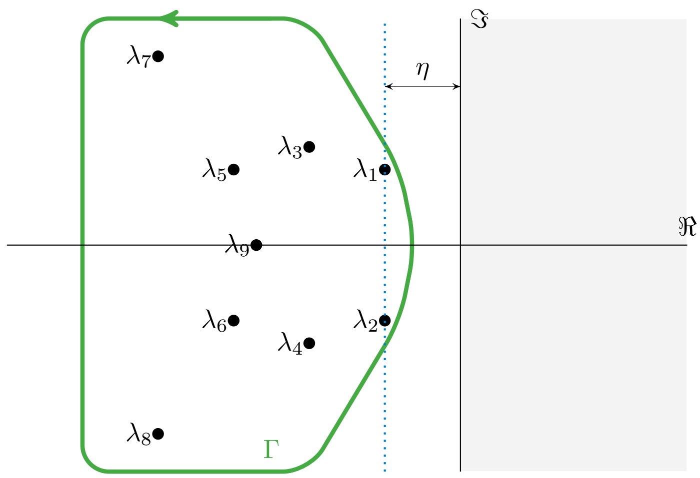


Figure 5.4: Illustration of the location of the spectrum in the complex plane for the statement of Theorem 5.1.2. All eigenvalues lie in the left half plane (white). The spectral gap η is the distance from the spectrum of A to the imaginary axis. The green curve Γ, which is used in the proof of the theorem, encircles the spectrum and remains entirely in the left half plane.

Remark 5.1.4. When A is a finite-dimensional linear operator (hence representable by a matrix), the condition of the eigenvalues having negative real part is sufficient for A to be exponentially stable; but this is not the case when A is an infinite-dimensional linear operator, because there could be infinitely many eigenvalues with negative real part and they could accumulate at the imaginary axis.

Corollary 5.1.5. *For $A \in M_n(\mathbb{C})$, if every eigenvalue of A has negative real part, then every solution of the IVP*

$$\dot{\mathbf{x}} = A\mathbf{x}, \qquad \mathbf{x}(0) = \mathbf{x}_0,$$

converges to $\mathbf{0}$ as $t \to \infty$.

Proof. The solution of the IVP is $\mathbf{x}(t) = \exp(tA)\mathbf{x}_0$. Since every eigenvalue of A has negative real part, Theorem 5.1.2 guarantees there are constants $\eta > 0$ and $C > 0$ such that $\|\exp(tA)\| \leq C\exp(-\eta t)$ for all $t \in [0, \infty)$. This gives

$$\|\exp(tA)\mathbf{x}_0\| \leq \|\exp(tA)\|\,\|\mathbf{x}_0\| \leq C\|\mathbf{x}_0\|\exp(-\eta t) \quad \text{for all} \quad t \geq 0,$$

which implies that $\|\exp(tA)\mathbf{x}_0\| \to 0$ as $t \to \infty$. $\square$

Remark 5.1.6. The constant C in the bound allows for the possibility of some transitory polynomial growth of a solution $\mathbf{x}(t) = \exp(tA)\mathbf{x}_0$ before the exponential decay dominates as $t \to \infty$. This can be seen from the expression for the principal fundamental matrix solution

$$\exp(tA) = \sum_{\lambda \in \sigma(A)} e^{\lambda t} \left(P_\lambda + \sum_{k=1}^{m_\lambda - 1} \frac{t^k D_\lambda^k}{k!} \right).$$

When the geometric multiplicity of λ is strictly smaller than the algebraic multiplicity, the nilpotent D_λ is nonzero for some $\lambda \in \sigma(A)$. This transitory behavior may be pertinent to a phenomenon being modeled, and for small t and large C, the solution may move far from $\mathbf{0}$, but as $t \to \infty$ the exponential dominates that constant and the solution converges to $\mathbf{0}$.

Example 5.1.7. Recall that the damped oscillator can be written as the system:

$$\dot{\mathbf{x}} = A\mathbf{x}, \quad \text{where} \quad A = \begin{bmatrix} 0 & 1 \\ -\omega^2 & -\frac{b}{m} \end{bmatrix}.$$

The characteristic polynomial of this matrix is

$$\lambda^2 + \frac{b}{m}\lambda + \omega^2 = 0,$$

with corresponding eigenvalues

$$\lambda = -\frac{b}{2m} \pm \sqrt{\frac{b^2}{4m^2} - \omega^2}.$$

Thus as long as $\omega^2 > \frac{b^2}{4m^2}$ then $\Re(\lambda) < 0$ for both eigenvalues, and the solutions always decay to $\mathbf{0}$ as $t \to \infty$. Hence, damped harmonic oscillators are asymptotically (exponentially) stable.

Remark 5.1.8. Given any solution $\mathbf{x}_1(t)$ of an inhomogeneous linear system $\dot{\mathbf{x}} = A\mathbf{x} + \mathbf{b}$, all other solutions can be written as $\mathbf{x}(t) = \mathbf{x}_1(t) + \boldsymbol{\xi}(t)$, where $\boldsymbol{\xi}(t)$ is a solution of the homogeneous system $\dot{\boldsymbol{\xi}} = A\boldsymbol{\xi}$. If A is exponentially stable, then every solution $\boldsymbol{\xi}(t)$ of the homogeneous system converges exponentially to $\mathbf{0}$ as $t \to \infty$, and thus every solution of the inhomogeneous system converges exponentially to $\mathbf{x}_1(t)$ as $t \to \infty$. Thus if A is exponentially stable, then all solutions of the inhomogeneous system converge toward a common solution as $t \to \infty$.

5.1.2 Uniformly Stable Linear Operators

Having all the solutions of the homogeneous system converge to $\mathbf{0}$ is a very strong sort of stability that is only achieved by the strong requirement that the spectrum of A lie in the left half plane. Weakening this requirement to allow the spectrum to touch the imaginary axis gives a weaker, but still useful, form of stability, but only if the geometric multiplicity is equal to the algebraic multiplicity for any eigenvalue on the imaginary axis. The following theorem guarantees this and also provides a sufficient condition on the spectrum of A for $\exp(tA)$ to be uniformly bounded on $[0, \infty)$.

Theorem 5.1.9. *For $A \in M_n(\mathbb{C})$, if $\Re(\lambda) \le 0$ for all $\lambda \in \sigma(A)$, and the geometric multiplicity of each $\lambda \in \sigma(A)$ with $\Re(\lambda) = 0$ equals its algebraic multiplicity, then there exists $M > 0$ such that $\|\exp(tA)\| \le M$ for all $t \in [0, \infty)$.*

Proof. Split the spectrum of A into two parts,

$$\begin{aligned}\sigma_-(A) &= \{\lambda \in \sigma(A) : \Re(\lambda) < 0\},\\ \sigma_0(A) &= \{\lambda \in \sigma(A) : \Re(\lambda) = 0\}.\end{aligned}$$

Using this splitting of the spectrum of A, the assumption on the geometric multiplicity equaling the algebraic multiplicity when $\Re(\lambda) = 0$ implies that the eigennilpotent D_λ is 0 for each $\lambda \in \sigma_0(A)$. The spectral decomposition of A becomes

$$A = \sum_{\lambda \in \sigma_-(A)} (\lambda P_\lambda + D_\lambda) + \sum_{\lambda \in \sigma_0(A)} \lambda P_\lambda.$$

Hence the principal fundamental matrix is

$$\exp(tA) = \sum_{\lambda \in \sigma_-(A)} e^{\lambda t} \left(P_\lambda + \sum_{k=1}^{m_\lambda - 1} \frac{t^k D_\lambda^k}{k!} \right) + \sum_{\lambda \in \sigma_0(A)} e^{\lambda t} P_\lambda.$$

For the eigenvalues in $\sigma_-(A)$ there exists $\eta > 0$ such that $|e^{t\lambda}| \le e^{-t\eta}$ for all $t \ge 0$. For all the eigenvalues $\lambda \in \sigma_0(A)$, we know that $|e^{\lambda t}| = 1$ for all $t \ge 0$.

Applying the norm to the principal fundamental solution matrix gives

$$\begin{aligned}\|\exp(tA)\| &\le \sum_{\lambda \in \sigma_-(A)} |e^{\lambda t}| \left\| P_\lambda + \sum_{k=1}^{m_\lambda - 1} \frac{t^k D_\lambda^k}{k!} \right\| + \sum_{\lambda \in \sigma_0(A)} |e^{\lambda t}| \|P_\lambda\| \\ &\le e^{-t\eta} \left(\sum_{\lambda \in \sigma_-(A)} \|P_\lambda\| + \frac{t^k}{k!} \sum_{k=1}^{m_\lambda - 1} \|D_\lambda\|^k \right) + \sum_{\lambda \in \sigma_0(A)} \|P_\lambda\|.\end{aligned}$$

The first term goes to zero as $t \to \infty$ and so is bounded on $[0, \infty)$, and the second term is constant.

Thus there exists $M > 0$ such that $\|\exp(tA)\| \le M$ for all $t \ge 0$. $\square$

Definition 5.1.10. *A matrix $A \in M_n(\mathbb{C})$ that satisfies the hypotheses of Theorem 5.1.9 is called* uniformly stable.

Corollary 5.1.11. *If $A \in M_n(\mathbb{C})$ is uniformly stable, then each solution of the IVP*

$$\dot{\mathbf{x}} = A\mathbf{x} \qquad \text{and} \qquad \mathbf{x}(0) = \mathbf{x}_0,$$

is bounded on $t \geq 0$.

Proof. The proof is Exercise 5.3 □

When $\Re(\lambda) \leq 0$ for every $\lambda \in \sigma(A)$ but the geometric multiplicity is less than the algebraic multiplicity for some $\lambda \in \sigma_0(A)$, then there are solutions of $\dot{\mathbf{x}} = A\mathbf{x}$ with polynomial or subexponential growth because the principal fundamental matrix solution has the form

$$\exp(tA) = \sum_{\lambda \in \sigma_-(A)} e^{\lambda t} \left(P_\lambda + \sum_{k=1}^{m_\lambda - 1} \frac{t^k D_\lambda^k}{k!} \right) + \sum_{\lambda \in \sigma_0(A)} e^{\lambda t} \left(P_\lambda + \sum_{k=1}^{m_\lambda - 1} \frac{t^k D_\lambda^k}{k!} \right).$$

This is the first manifestation of an instability in the solutions of $\dot{\mathbf{x}} = A\mathbf{x}$.

In the most general case where the eigenvalues of A have no restriction, split the spectrum of A into three parts $\sigma_-(A)$, $\sigma_0(A)$, and

$$\sigma_+(A) = \{\lambda \in \sigma(A) : \Re(\lambda) > 0\}.$$

The principal fundamental matrix solution has the representation

$$\begin{aligned} \exp(tA) = &\sum_{\lambda \in \sigma_-(A)} e^{\lambda t} \left(P_\lambda + \sum_{k=1}^{m_\lambda - 1} \frac{t^k D_\lambda^k}{k!} \right) \\ &+ \sum_{\lambda \in \sigma_0(A)} e^{\lambda t} \left(P_\lambda + \sum_{k=1}^{m_\lambda - 1} \frac{t^k D_\lambda^k}{k!} \right) \\ &+ \sum_{\lambda \in \sigma_+(A)} e^{\lambda t} \left(P_\lambda + \sum_{k=1}^{m_\lambda - 1} \frac{t^k D_\lambda^k}{k!} \right). \end{aligned}$$

If $\sigma_+(A) \neq \emptyset$, then solutions $\exp(tA)\mathbf{x}_0$ with $\mathbf{x}_0 \in \mathscr{R}(P_\lambda) \setminus \{\mathbf{0}\}$ for any $\lambda \in \sigma_+(A)$ grow exponentially as $t \to \infty$, giving an instability.

5.1.3 Stability of Equilibria

We now have enough of the theory for linear systems in place that we can formally define some notions of stability for equilibria of general, not-necessarily-linear ODEs. Throughout this section assume that $\mathbf{f} : \Omega \to \mathbb{R}^n$ is a C^1 vector field on an open subset $\Omega \subset \mathbb{R}^n$.

Definition 5.1.12. *A point $\overline{\mathbf{x}} \in \Omega$ is an* equilibrium solution *of the autonomous ODE $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ if $\mathbf{f}(\overline{\mathbf{x}}) = \mathbf{0}$. We sometimes call such an $\overline{\mathbf{x}}$ an equilibrium solution* for $\mathbf{f}$.

We think of $\overline{\mathbf{x}}$ not just as a single point of Ω but rather as a constant function $\overline{\mathbf{x}}(t)$ taking the value $\overline{\mathbf{x}}$ for all $t \in \mathbb{R}$. The derivative satisfies $\frac{d}{dt}\bar{x}(t) = \mathbf{0} = \mathbf{f}(\overline{\mathbf{x}})$, so the function $\overline{\mathbf{x}}$ is a solution of the ODE, and its maximal open interval of existence is $\mathbb{R}$.

Definition 5.1.13. *Let $\mathbf{x}(t)$ be the unique solution of the IVP*

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}), \qquad \text{and} \qquad \mathbf{x}(0) = \mathbf{x}_0 \in \Omega.$$

The solution $\mathbf{x}(t)$ has a maximal open interval of existence I that depends on the initial condition.

(*i*) *The graph $\{\mathbf{x}(t) \mid t \in I\} \subset \Omega$ is called the* orbit of $\mathbf{f}$ *containing the initial condition $\mathbf{x}_0$. Note that different initial conditions can have the same orbit.*

(*ii*) *The collection of all orbits of $\mathbf{f}$ in Ω is called the* phase portrait *of $\mathbf{f}$.*

Remark 5.1.14. A fundamental objective of dynamical systems theory is to describe the qualitative features of the phase portrait of a given $\mathbf{f}$. In low-dimensional systems this can often be accomplished by sketching enough of the orbits to get the overall sense of the phase portrait of the vector field $\mathbf{f}$.

Example 5.1.15. Figure 5.5 shows plots of the linear vector field $\mathbf{f}(\mathbf{x}) = A\mathbf{x}$ and several orbits from the phase portrait of that vector field for each of the six types of systems in Example 5.1.1. We often call such a plot a phase portrait of the system. Of course the full phase portrait would include all the orbits, which would make the plots too cluttered to be useful. But the few orbits that are plotted here are enough to give a good sense of the dynamics of each system.

The origin is always an equilibrium solution. Unless A has an eigenvalue equal to 0, this is the unique equilibrium solution. If A is exponentially stable (upper left and bottom right panels), all the orbits converge to this equilibrium. When the eigenvalues are distinct and nonzero, but purely imaginary (bottom left panel), the matrix is uniformly bounded, and all the solutions that start close to the equilibrium at the origin remain close to it (but they never converge to it). When at one eigenvalue has a strictly positive real part (upper center panel), then most of the orbits diverge away from the equilibrium, but the orbits that contain a point on the x-axis (in this example, that's the eigenspace corresponding to the eigenvalue with negative real part) do converge to the origin. Finally, orbits for those systems that have two eigenvalues with positive real part (bottom center and upper right panels) all diverge away from the equilibrium at the origin.

Remark 5.1.16. Dealing with maximal open intervals of existence that are not $\mathbb{R}$ can be problematic. There is a way to guarantee the solution of every IVP has $\mathbb{R}$ as its maximal open interval of existence, namely, replace the vector field $\mathbf{f}(\mathbf{x})$ with the following scalar multiple

$$\frac{\mathbf{f}(\mathbf{x})}{1 + \|\mathbf{f}(\mathbf{x})\|}.$$

This scaled vector has the same phase portrait as the original $\mathbf{f}(\mathbf{x})$, but guarantees that it makes sense to talk about limits of solutions $\mathbf{x}(t)$ as $t \to \infty$.

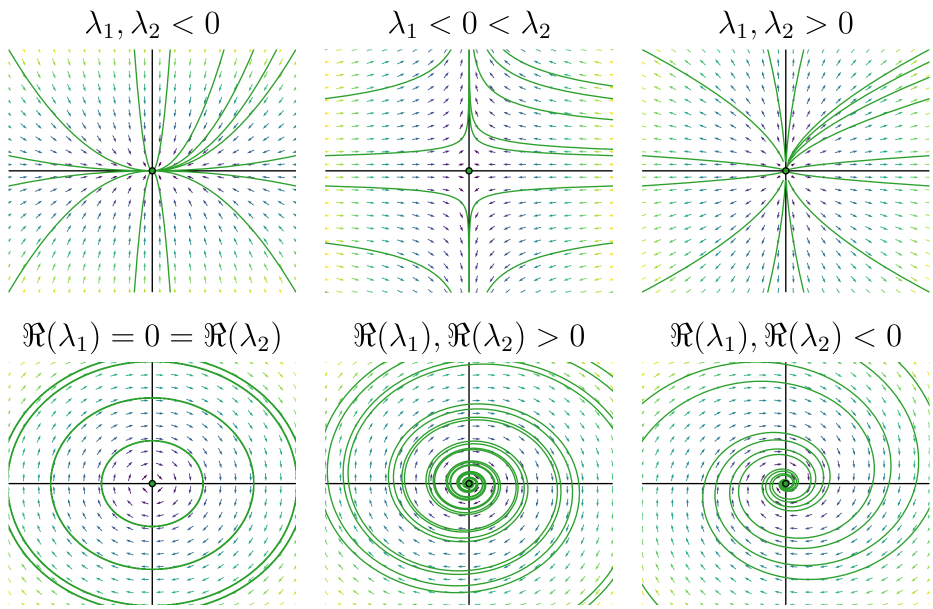


Figure 5.5: Plot of the linear vector field $\mathbf{f}(\mathbf{x}) = A\mathbf{x}$ and several orbits from the phase portrait of that vector field for each of the six types of systems in Example 5.1.1, as discussed in Example 5.1.15. The vectors in this plot have been scaled as discussed in Remark 5.1.16. The true length of each vector is indicated by the color of the corresponding arrow (purple indicates shorter and yellow indicates longer vectors).

Definition 5.1.17. *An equilibrium solution* $\overline{\mathbf{x}}$ *of the IVP*

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t)), \qquad \textit{and} \qquad \mathbf{x}(t_0) = \mathbf{x}_0, \tag{5.1}$$

is said to be stable *if for every* $\varepsilon > 0$ *there is a* $\delta > 0$ *such that for every initial condition* $\mathbf{x}_0$ *satisfying* $\|\mathbf{x}_0 - \overline{\mathbf{x}}\| < \delta$*, the solution of* (5.1) *exists for all* $t > t_0$ *and satisfies* $\|\mathbf{x}(t) - \overline{\mathbf{x}}\| < \varepsilon$*.*

Example 5.1.18. Of the systems in Example 5.1.15, plotted in Figure 5.5, all have a unique equilibrium solution at the origin, and that equilibrium is stable in the sense of Definition 5.1.17 when the system is defined by an exponentially stable matrix (upper left and lower right panels of Figure 5.5) or a uniformly stable matrix (lower left panel of the figure). For that lower left panel, solutions near the origin do not get closer to the equilibrium over time, but they also do not get farther away, so if the initial value is close to the equilibrium, all future values of the corresponding solution will also be close to the equilibrium.

Remark 5.1.19. Stating that an equilibrium $\overline{\mathbf{x}}$ is stable in this sense means that if we want a solution (possibly time-dependent) be near the stable equilibrium $\overline{\mathbf{x}}$, then we can do this by taking an initial condition near $\overline{\mathbf{x}}$. This is not a very strong form of stability—it does not guarantee that solutions starting near $\overline{\mathbf{x}}$ converge to $\overline{\mathbf{x}}$, but rather only that they stay close to $\overline{\mathbf{x}}$. This type of stability also does not occur as frequently as one might expect. For example, solutions of linear systems $\mathbf{f}(\mathbf{x}) = A\mathbf{x}$ are only stable in this sense if A is uniformly stable. Otherwise the solutions are oscillating or exponentially growing.

Proposition 5.1.20. *If A is uniformly stable (See Definition 5.1.10), then the equilibrium solution at the origin is stable.*

Proof. The proof is Exercise 5.4 □

A stronger condition for an equilibrium solution $\overline{\mathbf{x}}$ than the stability of Definition 5.1.17 is to insist that solutions starting near $\overline{\mathbf{x}}$ converge to $\overline{\mathbf{x}}$, that is, solutions $\mathbf{x}(t)$ with initial conditions $\mathbf{x}_0$ near $\overline{\mathbf{x}}$ must satisfy $\lim_{t\to\infty} \mathbf{x}(t) = \overline{\mathbf{x}}$. There are two possibilities for how solutions starting near such an equilibrium solution $\overline{\mathbf{x}}$ can behave: (i) all the nearby solutions stay nearby and converge to the equilibrium (such an equilibrium is called *asymptotically stable*), or (ii) there exists at least one solution $\mathbf{x}(t)$ that wanders far away before returning and eventually converging to the equilibrium. You might think this second possibility is impossible, but it does happen. Such an equilibrium is called *attracting*. The formal definitions are given below.

Definition 5.1.21. *An equilibrium solution $\overline{\mathbf{x}}$ of $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ is said to be* asymptotically stable *if both of the following conditions are satisfied:*

(i) *The equilibrium $\overline{\mathbf{x}}$ is stable.*

(ii) *There exists a $\delta > 0$ such that if $\mathbf{x}(t)$ is a solution of (5.1) with $\|\mathbf{x}(0)-\overline{\mathbf{x}}\| < \delta$, then $\lim_{t\to\infty} \mathbf{x}(t) = \overline{\mathbf{x}}$.*

An equilibrium solution $\overline{\mathbf{x}}$ is called attracting *if it is not stable but there exists $\nu > 0$ such that for all $\mathbf{x}_0$ satisfying $\|\mathbf{x}_0 - \overline{\mathbf{x}}\| < \nu$ the solution $\mathbf{x}(t)$ of (5.1) satisfies $\mathbf{x}(t) \to \overline{\mathbf{x}}$ as $t \to \infty$.*

If a matrix A is exponentially stable, that is, every eigenvalue of A has negative real part, then the equilibrium solution at the origin for the homogeneous linear system $\dot{\mathbf{x}} = A\mathbf{x}$ is asymptotically stable, as shown by the following proposition.

Proposition 5.1.22. *If every eigenvalue of A has negative real part, then the equilibrium at the origin for $\dot{\mathbf{x}} = A\mathbf{x}$ is asymptotically stable.*

Proof. As shown in the proof of Corollary 5.1.5 there exists $\eta > 0$ and $C > 0$ such that

$$\|\exp(tA)\mathbf{x}_0\| \leq C\|\mathbf{x}_0\|\exp(-\eta t) \quad \text{for} \quad \text{all } t \geq 0.$$

To see that this implies stability note that $\exp(-\eta t) \leq 1$ for all $t \geq 0$, so

$$\|\exp(tA)\mathbf{x}_0\| \leq C\|\mathbf{x}_0\| \quad \text{for} \quad \text{all } t \geq 0.$$

For $\varepsilon > 0$ choose $\delta = \varepsilon/C$, so that for all $\mathbf{x}_0$ satisfying $\|\mathbf{x}_0\| < \delta$ we have

$$\|\exp(tA)\mathbf{x}_0\| \leq C\mathbf{x}_0 < C\frac{\varepsilon}{C} = \varepsilon.$$

This shows that the equilibrium at the origin is stable.

For $\nu = 1$ (we can choose any positive number for ν in the linear case) we have for $\mathbf{x}_0$ satisfying $\|\mathbf{x}_0\| < \nu$ that

$$\|\exp(tA)\mathbf{x}_0\| \leq C\|\mathbf{x}_0\|\exp(-\eta t) < C\nu\exp(-\eta t) = C\exp(-\eta t) \to 0 \quad \text{as} \quad t \to \infty.$$

Thus the equilibrium at the origin is asymptotical stable. □

5.2 Linearization, a Stability Theorem, and Phase Portraits

With the rigorous definition of stability for equilibrium solutions in place, we proceed to the first nonlinear asymptotic stability result for equilibrium solutions.

5.2.1 Linearization

One of the most fundamental tools in applied mathematics is linearization. Reality is mostly nonlinear but nonlinear things are hard, mathematically, so we pretend that what we are considering is linear; that is, we look locally where things are almost linear and use a linear approximation to try to understand the nonlinear system. In the case of dynamical systems, it is most useful to linearize near an equilibrium solution.

Suppose that $\overline{\mathbf{x}}$ is an equilibrium solution of $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$, where $\mathbf{f} : \Omega \to \mathbb{R}^n$ is C^3 for open $\Omega \subset \mathbb{R}^n$. Define the *variation* of a solution $\mathbf{x}$ from the equilibrium $\overline{\mathbf{x}}$ to be

$$\mathbf{v} = \mathbf{x} - \overline{\mathbf{x}}.$$

We have

$$\dot{\mathbf{v}} = \dot{\mathbf{x}} - \dot{\overline{\mathbf{x}}} = \dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) = \mathbf{f}(\overline{\mathbf{x}} + \mathbf{v}).$$

Expanding $\mathbf{f}(\overline{\mathbf{x}}+\mathbf{v})$ about $\mathbf{v}=\mathbf{0}$, using the C^3 assumption on $\mathbf{f}$ and the fact that $\mathbf{f}(\overline{\mathbf{x}})=0$ gives

$$\mathbf{f}(\overline{\mathbf{x}}+\mathbf{v}) = \mathbf{f}(\overline{\mathbf{x}}) + D\mathbf{f}(\overline{\mathbf{x}})\mathbf{v} + \frac{1}{2}\mathbf{v}^{\mathsf{T}} D^2\mathbf{f}(\overline{\mathbf{x}})\mathbf{v} + \cdots = D\mathbf{f}(\overline{\mathbf{x}})\mathbf{v} + Q(\mathbf{v}).$$

where

$$Q(\mathbf{v}) = \frac{1}{2}\mathbf{v}^{\mathsf{T}} D^2\mathbf{f}(\overline{\mathbf{x}})\mathbf{v} + \cdots = O(\|\mathbf{v}\|^2). \tag{5.2}$$

Thus the variation $\mathbf{v}$ satisfies the differential equation

$$\dot{\mathbf{v}} = D\mathbf{f}(\overline{\mathbf{x}})\mathbf{v} + Q(\mathbf{v}). \tag{5.3}$$

Since the nonlinear term $Q(\mathbf{v})$ is $O(\|\mathbf{v}\|^2)$ it satisfies

$$\lim_{\|\mathbf{v}\|\to 0} \frac{\|Q(\mathbf{v})\|}{\|\mathbf{v}\|} = 0. \tag{5.4}$$

This partially motivates the following definition.

Definition 5.2.1. *The* linearization *of $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ at an equilibrium solution $\overline{\mathbf{x}}$ is the linear differential equation*

$$\dot{\mathbf{v}} = D\mathbf{f}(\overline{\mathbf{x}})\mathbf{v},$$

which is a linear approximation of (5.3)*. The equilibrium $\overline{\mathbf{x}}$ of $\mathbf{f}$ is* hyperbolic[26] *if the spectrum of the linearization $A = D\mathbf{f}(\overline{\mathbf{x}})$ satisfies $\Re(\lambda) \neq 0$ for all $\lambda \in \sigma(A)$.*

5.2.2 Asymptotic Stability

The linearization of an ODE at an equilibrium gives a lot of information about the stability of the equilibrium. Our first main result in this direction is Theorem 5.2.2, below. It follows from from Duhamel's Principle, which allows us to control the effect of the nonlinear term in (5.3).

Theorem 5.2.2 (Asymptotic Stability). *Suppose $\overline{\mathbf{x}}$ is a hyperbolic equilibrium of a C^3 vector field $\mathbf{f} : \Omega \to \mathbb{R}^n$ for some open $\Omega \subset \mathbb{R}^n$. If $\sigma(D\mathbf{f}(\overline{\mathbf{x}}))$ lies in the left-hand side of $\mathbb{C}$, i.e., $\Re(\lambda) < 0$ for all $\lambda \in \sigma(D\mathbf{f}(\overline{\mathbf{x}}))$, then the equilibrium solution $\overline{\mathbf{x}}$ is asymptotically stable for $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$.*

Proof. The variation $\mathbf{v} = \mathbf{x} - \overline{\mathbf{x}}$ satisfies $\mathbf{v}' = D\mathbf{f}(\overline{\mathbf{x}})\mathbf{v} + Q(\mathbf{v})$ where $Q(\mathbf{v})$ satisfies (5.4). The linearized equation $\dot{\mathbf{y}} = A\mathbf{y}$, with $A = D\mathbf{f}(\overline{\mathbf{x}})$, has the principal fundamental matrix solution $\Phi(t) = \exp(tA)$. By Duhamel's Principle, the variation $\mathbf{v}(t)$ satisfies the implicit equation

$$\begin{aligned}
\mathbf{v}(t) &= \exp(tA)\mathbf{v}(0) + \exp(tA)\int_0^t \exp(sA)^{-1} Q(\mathbf{v}(s))\, ds \\
&= \exp(tA)\mathbf{v}(0) + \int_0^t \exp(tA)\exp(-sA) Q(\mathbf{v})(s))\, ds \\
&= \exp(tA)\mathbf{v}(0) + \int_0^t \exp((t-s)A) Q(\mathbf{v})(s))\, ds,
\end{aligned}$$

[26] This name comes from the fact that solutions of some systems near hyperbolic equilibria can look like hyperbolae. But unfortunately not all systems look like hyperbolae near hyperbolic equilibria. As Strogatz says, "hyperbolic is an unfortunate name...but it has become standard."

where I is the maximal open interval of existence which contains 0.

From the limit in (5.4) for each $\eta > 0$ there exists $\alpha > 0$ such that when $\|\mathbf{v}\| < \alpha$,

$$\frac{\|Q(\mathbf{v})\|}{\|\mathbf{v}\|} < \eta \qquad \text{which implies} \qquad \|Q(\mathbf{v})\| \leq \eta\|\mathbf{v}\|. \tag{5.5}$$

Because the spectrum of A satisfies $\Re(\lambda) < 0$ for all $\lambda \in \sigma(A)$, Theorem 5.1.2 guarantees there exists a $\zeta > 0$ and $K > 0$ such that

$$\|\exp(tA)\| \leq K\exp(-\zeta t) \text{ for all } t \geq 0.$$

Applying the norm to the implicit equation that the variation satisfies and assuming that $\|\mathbf{v}\| < \alpha$ so that $\|Q(\mathbf{v}(s))\| < \eta\|\mathbf{v}(s)\|$, we obtain

$$\begin{aligned} \|\mathbf{v}(t)\| &\leq \|\exp(tA)\|\,\|\mathbf{v}(0)\| + \int_0^t \|\exp((t-s)A\|\,\|Q(\mathbf{v}(s))\|\,ds \\ &\leq K\|\mathbf{v}(0)\|\exp(-\zeta t) + \int_0^t \eta K\exp((-\zeta(t-s))\|\mathbf{v}(s)\|\,ds. \end{aligned}$$

Multiply both sides by $\exp(\zeta t)$ to get

$$\exp(\zeta t)\|\mathbf{v}(t)\| \leq K\|\mathbf{v}_0\| + \int_0^t \eta K\exp(\zeta s)\|\mathbf{v}(s)\|\,ds.$$

Applying Gronwall's inequality (Lemma 2.2.11) with $u(t) = \exp(\zeta t)\|\mathbf{v}(t)\|$ and $\phi(t) = \eta$ gives

$$\exp(\zeta t)\|\mathbf{v}(t)\| \leq K\|\mathbf{v}(0)\|\exp\int_0^t \eta K\,ds = K\|\mathbf{v}(0)\|\exp(\eta K t).$$

Multiplying both sides by $\exp(-\zeta t)$ gives

$$\|\mathbf{v}(t)\| \leq K\|\mathbf{v}(0)\|\exp((\eta K - \zeta)t),\ \ t \in I, \tag{5.6}$$

as long as $\|\mathbf{v}(t)\| < \alpha$.

Now, we have freedom to choose $\eta > 0$ (which determines the value of $\alpha > 0$). Hence choose η so that $0 < \eta < \zeta/K$. Then the quantity $\eta K - \zeta$ that appears in the exponential bounding $\mathbf{v}(t)$ satisfies

$$\eta K - \zeta < \zeta - \zeta = 0.$$

This choice of η implies that a variation $\mathbf{v}(t)$ starting with $\|\mathbf{v}(0)\| < \alpha$ is bounded by the inequality (5.6) (no finite time blowup) and hence is defined for all $t \geq 0$.

For the α corresponding to the η selected as described above, choose $\varepsilon > 0$ satisfying $0 < \varepsilon < \alpha$ and choose $\delta = \varepsilon/K$. It follows that when $\|\mathbf{v}(0)\| < \delta$, then

$$\|\mathbf{v}(t)\| \leq K\|\mathbf{v}(0)\|\exp((\eta K - \zeta)t) \leq K\|\mathbf{v}(0)\| < K\delta = K\frac{\varepsilon}{K} = \varepsilon.$$

This shows that the equilibrium $\overline{\mathbf{x}}$ for $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ is stable.

To establish asymptotic stability, note that for $\|\mathbf{v}(0)\| < \delta$ it follows that

$$\|\mathbf{v}(t)\| \leq K\|\mathbf{v}(0)\|\exp((\eta K - \zeta)t) \leq K\delta\exp((\eta K - \zeta)t) \to 0 \ \text{ as } \ t \to \infty.$$

This shows that the equilibrium $\overline{\mathbf{x}}$ for $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ is asymptotically stable. $\square$

5.2.3 Unstable Linearization

The Hartman-Grobman theorem (which we do not prove here) says that at a hyperbolic equilibrium a nonlinear system is qualitatively the same as its linearization at that equilibrium. The asymptotic stability theorem is a sort of special case, which guarantees that a hyperbolic equilibrium is asymptotically stable if the linearization is.

Proposition 5.2.3. *Any hyperbolic equilibrium solution* $\overline{\mathbf{x}}$ *of* $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ *must be isolated.*

Proof. Translate the equilibrium solution $\overline{\mathbf{x}}$ to the origin, by constructing the function

$$\mathbf{g}(\mathbf{x}) = \mathbf{f}(\mathbf{x} + \overline{\mathbf{x}}),$$

which is the composition of the function $\mathbf{f}$ with the translation $\mathbf{x} \to \mathbf{x} + \overline{\mathbf{x}}$. The map $\mathbf{g}$ takes $\mathbf{0}$ to $\mathbf{0}$ because

$$\mathbf{g}(\mathbf{0}) = \mathbf{f}(\mathbf{0} + \overline{\mathbf{x}}) = \mathbf{0}.$$

Because $\overline{\mathbf{x}}$ is a hyperbolic equilibrium, the derivative $D\mathbf{f}(\overline{\mathbf{x}}) = D\mathbf{g}(\mathbf{0})$ (by the chain rule) cannot have a zero eigenvalue, which means that $D\mathbf{g}(\mathbf{0})$ is nonsingular. The Inverse Function Theorem guarantees $\mathbf{g}$ is locally invertible, meaning no point near $\mathbf{0}$ except $\mathbf{0}$ gets mapped by $\mathbf{g}$ to $\mathbf{0}$. This means that no point near $\overline{\mathbf{x}}$ except $\overline{\mathbf{x}}$ gets mapped by $\mathbf{f}$ to $\mathbf{0}$, which says that the equilibrium solution $\overline{\mathbf{x}}$ is isolated. □

The previous proposition implies that if an equilibrium is not isolated, it is not hyperbolic, and so the asymptotic stability theorem does not apply to nonisolated equilibria. We need other techniques for this case. And, unfortunately, there are examples of systems with a stable (nonhyperbolic) equilibrium where the linearization is unstable. This can happen when the instability for the linearization is polynomial.

To see where these concepts of stability actually apply, we get to the very exciting part of this Chapter, lots of pictures. For finite dimensional autonomous systems, a picture will often teach us almost everything we need to know about the dynamics of the system.

5.2.4 One-dimensional (Scalar) Models

One-dimensional autonomous dynamical systems can be completely understood qualitatively. We are considering $\dot{x} = f(x)$, $x \in \Omega$, for $f : \Omega \to \mathbb{R}$ with Ω an open interval in $\mathbb{R}$ and f being smooth enough so that we can apply the appropriate qualitative results.

To analyze $\dot{x} = f(x)$ first find the zeros $\bar{x}$ of f, which correspond to the equilibrium solutions. For a generic f these zeros are isolated (which we assume) and partition the open interval Ω into subintervals on which f is either all positive or all negative. Where $f = \dot{x}$ is positive, the solution $x(t)$ is an increasing function of t (moving right), and where f is negative, the solution $x(t)$ is a decreasing function of t (moving left). This gives a complete phase portrait for $\dot{x} = f(x)$ on the open interval Ω.

Remark 5.2.4. Sometimes phase portraits of $\dot{x} = f(x)$ are drawn vertically, in which case increasing is up and decreasing is down.

Remark 5.2.5. The equation $\dot{x} = f(x)$ can always be solved implicitly by separation of variables,

$$\frac{dx}{dt} = f(x) \ \Rightarrow \ \frac{dx}{f(x)} = dt \ \Rightarrow \ \int \frac{dx}{f(x)} = t + C.$$

This is called an *implicit* solution, because x is not computed explicitly. Only in rare cases can the antiderivative of $\frac{1}{f(x)}$ be inverted to obtain explicit solutions. But the implicit solutions are typically of little help in understanding the behavior of the solutions.

Example 5.2.6. We begin with a classic example that quickly illustrates the need to graphically investigate solutions of dynamical systems rather than relying on explicit algebraically formulated solutions. Not only are explicit solutions most often not available, but often even when such solutions are available they are not particularly useful.

Consider the ODE

$$\dot{x} = \sin(x), \quad x(0) = a.$$

Separation of variables and a bit of calculus gives the implicit solution

$$t = \log\left|\frac{\csc a + \cot a}{\csc x + \cot x}\right|,$$

which tells us almost nothing about the behavior of $x(t)$.

To get a better idea how the solutions behave, consider Figure 5.6. The panel on the left shows a plot of $\sin(x)$ where the horizontal axis is x and the vertical axis represents $y = \sin(x)$. The arrows on the x-axis indicate how the solution evolves. If $\sin(x) > 0$, then $x(t)$ is increasing; and if $\sin(x) < 0$, then $x(t)$ is decreasing. The fixed points, where $\dot{x} = \sin(x) = 0$, occur at $x_k = k\pi$, corresponding to points where $x(t)$ is neither decreasing nor increasing, but instead remains fixed (stationary). In a neighborhood of x_k for odd k, all solutions tend toward x_k, indicating that x_k is a linearly asymptotically stable fixed point. For even k, all solutions near x_k tend to move away from the fixed point, indicating that x_k is unstable.

This is further illustrated in the right panel of Figure 5.6. The horizontal axis in this panel is time t, and the vertical axis and corresponding plots represent different solutions x of the ODE. Three different fixed points are illustrated, two stable, and one unstable. Here the solutions are indicated by the curves with arrows showing the tendency of the solution to move away from the unstable fixed points toward the stable ones.

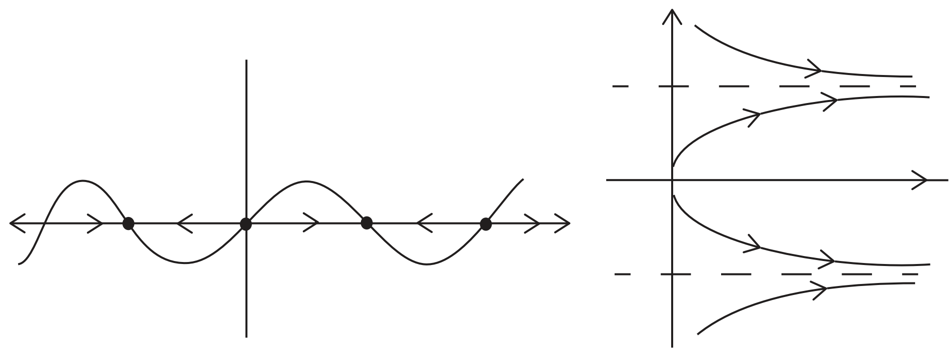


Figure 5.6: Phase space (left) and configuration space (right) for the ODE $\dot{x} = \sin(x)$ as described in Example 5.2.6.

The phase portrait depicted in the left panel of Figure 5.6 shows that all points $x_k = k\pi$ for k even are unstable fixed points, and all points x_k for k odd are asymptotically stable. To verify this algebraically, consider the linearization of this ODE:

$$\dot{v} = \cos(x_k) = (-1)^k.$$

This clearly gives the same stability result that we obtained graphically.

Example 5.2.7. Here is an example where the only information is graphical. Consider

$$\dot{x} = f(x), \quad x(0) = x_0,$$

for the function $f(x)$ whose graph is shown in Figure 5.7. The long-term (often called time-asymptotic) behavior of solutions satisfies

$$\begin{aligned}
&\text{if } x_0 < x_1, \quad \text{then } x(t) \to -\infty \\
&\quad x_1 < x_0 < x_2, \quad \text{then } x(t) \to x_2 \\
&\quad x_2 < x_0 < x_3 \Rightarrow x(t) \to x_2 \\
&\quad x_3 < x_0 < x_4 \Rightarrow x(t) \to x_4 \\
&\quad x_4 < x_0 \Rightarrow x(t) \to x_4.
\end{aligned}$$

Remark 5.2.8. This illustrates again the usefulness of a pictorial representation of the system, even without any explicit formula for the solution. Often even the right hand side of an ODE is not representable as a simple function, but can be approximated well numerically. Without solving the actual ODE, we can get an understanding of the behavior of solutions by looking at a graph such as the one seen here.

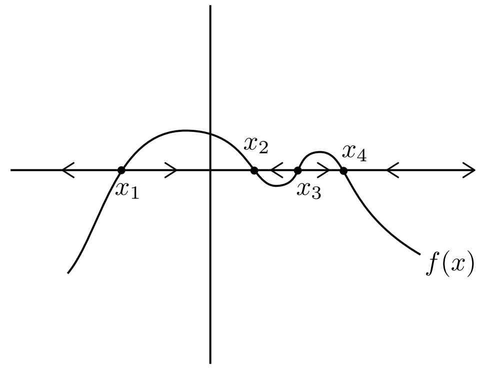


Figure 5.7: The phase portrait for Example 5.2.7

Interestingly, we need not calculate the linearization at each fixed point to see what happened here, but if we could we would come to the same conclusion. This is because the linearization is simply giving the slope of the tangent line at each of the equilibria, which indicates whether the phase portrait is crossing from negative to positive (positive slope) or positive to negative (negative slope).

Example 5.2.9. We consider $\dot{x} = f(x)$ where $f(x) = -x^3$ for $f : \mathbb{R} \to \mathbb{R}$. The equation $\dot{x} = -x^3$ has one equilibrium $\bar{x} = 0$ which partitions $\mathbb{R}$ into two subintervals $x > 0$ and $x < 0$.

(i) On $x > 0$ the value of $f(x) = -x^3 < 0$ so solutions decrease towards 0.

(ii) On $x < 0$ the value of $f(x) = -x^3 > 0$ so solutions increase towards 0.

These qualitative features of the solutions show that the equilibrium $\bar{x}$ is stable and that solutions starting nearby converge to $\bar{x}$; hence $\bar{x}$ is asymptotically stable.

The "matrix" of the linearization of f at the equilibrium $\bar{x} = 0$ is the 1×1 matrix $Df(\bar{x}) = -3(0)^2 = 0$ which has the eigenvalue 0; this does not have negative real part, and in fact because the real part of the eigenvalue is 0 this is not a hyperbolic equilibrium and hence the theory developed above does not apply.

To further justify this behavior, we note that for $x(0) = x_0 \neq 0$:

$$x(t) = \frac{x_0}{\sqrt{1 + t x_0^2}}, \quad t > -\frac{1}{2x_0^2}.$$

These solutions monotonically converge to 0, which gives asymptotic stability. The point of analyzing the graph and behavior of $f(x) = -x^3$ is that we didn't need to find the explicit solutions to determine the long-term behavior of the solutions.

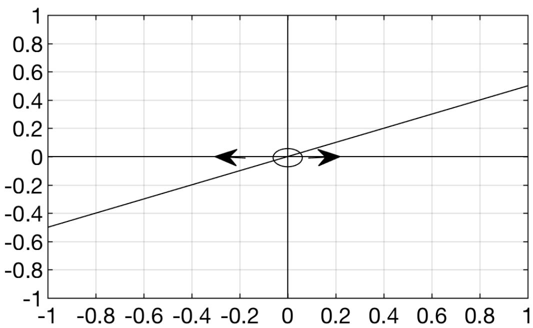


Figure 5.8: Phase portrait of $\dot{x} = \frac{1}{2}x$ as described in Example 5.2.10.

Example 5.2.10. Draw a phase portrait of the one-dimensional system $\dot{x} = \frac{1}{2}x$.

Solution: The only fixed point is $x = 0$. When the initial value is positive, $\dot{x} = \frac{1}{2}x > 0$ so that the trajectory moves to the right (growing). When the initial point is negative, $\dot{x} < 0$ so the trajectory moves to the left. In Figure 5.8 we draw the phase portrait on the horizontal axis. We use an open dot for the fixed point to demonstrate that solutions move away from the origin.

5.2.5 Planar Examples

Sketching the phase portrait for planar systems $\mathbf{x} = \mathbf{f}(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^2$, starts with finding the zeros of the vector field $\mathbf{f}$. This is not as simple as the one-dimensional situation becuase the vector field $\mathbf{f}$ has two components, i.e.,

$$\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x})).$$

For higher dimensional vector fields $\mathbf{f} : \Omega \to \mathbb{R}^n$, the function $\mathbf{f}$ has n components

$$\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_n(\mathbf{x})).$$

Definition 5.2.11. *A* nullcline *of* $\mathbf{f} : \Omega \to \mathbb{R}^n$ *for the coordinate* x_k *is the set*

$$\{\mathbf{x} \in \Omega : f_k(\mathbf{x}) = 0\}.$$

Remark 5.2.12. In two dimensions, the nullclines effectively separate phase space (the plane) into different regions and equilibria appear at intersections of nullclines. In higher dimensions nullclines are not used as frequently, probably because they are more difficult to represent graphically.

Example 5.2.13. Consider the ODE given by

$$\begin{aligned}\dot{x} &= x + e^{-y}\\ \dot{y} &= -y.\end{aligned}$$

The x nullcline is given by $x = -e^{-y}$ and the y nullcline is given by $y = 0$. The only fixed point of the system occurs at the intersection of these two curves, at $(-1, 0)$. To determine the complete phase portrait of this system and get a qualitative picture for the evolution of the full system, determine the sign of $\dot{x}$ and $\dot{y}$ in each of the four regions in the plane defined by these nullclines. As indicated in Figure 5.9, the nullclines are where $\dot{x}$ or $\dot{y}$ change sign, and divide the plane into regions where the signs cannot change. Noticing which direction the vector field is pointed at a given point in the plane helps indicate what a solution looks like at that point.

What can we say about the single fixed point of this system? Without delving into εs and δs Figure 5.9 shows that this equilibrium is not stable in any sense that we have discussed so far. It is interesting to note, however, that the line $y = 0$ appears to be 'stable' in the sense that all solutions tend to decay toward it (but they grow in x).

Example 5.2.14. Consider the linear system

$$\dot{x} = A\mathbf{x}, \qquad \text{where} \qquad A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

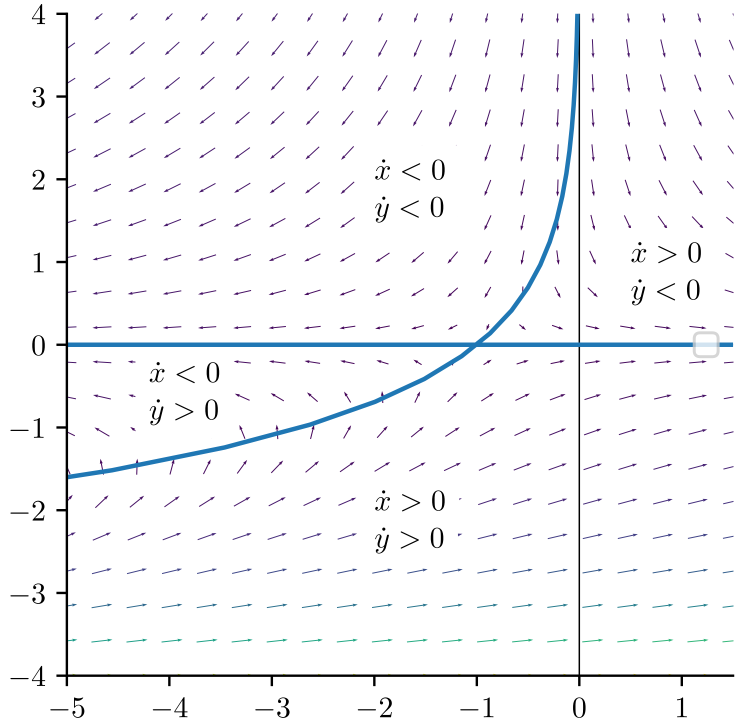


Figure 5.9: Nullclines (blue) and flow field (arrows) for Example 5.2.13. The arrows indicate the motion of the vector field with the length of the arrow indicating the strength of the vector field at that point. Solutions can only cross the $\dot{x} = 0$ nullcline vertically, and the $\dot{y} = 0$ nullcline horizontally. In this case that means the y nullcline cannot be crossed, separating the plane into two regions in which solutions will always remain.

This is depicted in the bottom left panel of Figure 5.5. In this case the nullclines are $x = 0$ and $y = 0$ for $\dot{y}$ and $\dot{x}$ respectively. Solutions must cross the x- and y-axes orthogonally, and solutions rotate about the origin, the only true equilibrium of the linear system. This is a classic example of a fixed point that is stable, but not asymptotically stable. Solutions near the origin will not move further away, but will also not move closer.

In this last example, the origin is often referred to as a *center*, because the flow stays centered about the origin. We have seen this type of behavior before in the first-order Lotka–Volterra model (there the center was about the coexistence fixed point—not the origin—and the system was certainly *not* linear) and in the undamped harmonic oscillator. One can check that the eigenvalues for a center are purely imaginary.

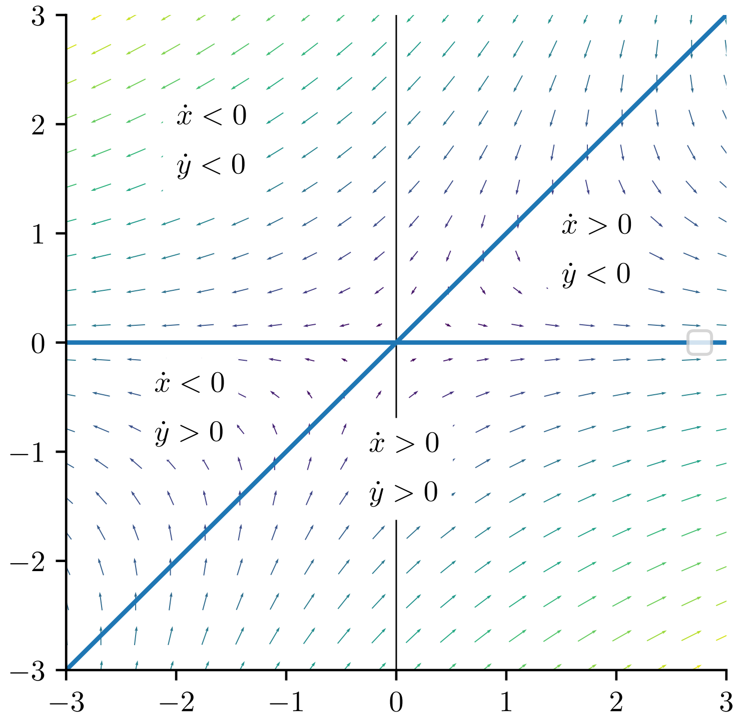


Figure 5.10: Nullclines (blue) and flow field (arrows) for Example 5.2.15. The arrows indicate the motion of the vector field, with the length and color of the arrow indicating the strength of the vector field at that point. Solutions can only cross the $\dot{x} = 0$ nullcline vertically, and the $\dot{y} = 0$ nullcline horizontally. As in Figure 5.9, that means the y nullcline cannot be crossed, separating the plane into two regions in which solutions will always remain.

Example 5.2.15. Consider now the linear homogeneous system given by

$$\begin{aligned} \dot{x} &= x - y \\ \dot{y} &= -y. \end{aligned}$$

Just as in the previous example, the only fixed point is $x = y = 0$ the origin (this is generically true for all linear homogeneous systems). However, as shown in Figure 5.10, the nullclines are $x = y$ for $\dot{x} = 0$ and $y = 0$ for $\dot{y}$, meaning that somewhat like Example 5.2.13, solutions can not cross the x-axis, and must cross the line $x = y$ vertically. This leads to solutions that approach the x axis, meaning that y decays, but x grows without bound. In this case the origin is neither stable nor attracting.

Example 5.2.16. Consider the vector field

$$\mathbf{f}(x, y) = (-x^3 + xy^4, -y^3 - x^4 y).$$

To find the nullclines set the components of the vector to zero and solve:

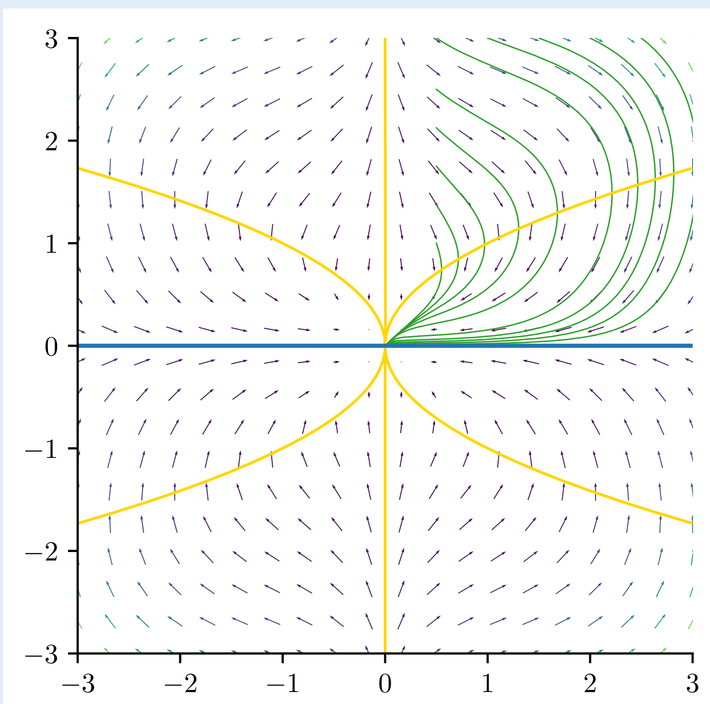
$$0 = -x^3 + xy^4 = x(-x^2 + y^4) \quad \text{and} \quad 0 = -y^3 - x^4 y = -y(y^2 + x^4).$$

(i) The nullcline for $\dot{x}$ is $x = 0$ (the y-axis) and the curve $-x^2 + y^4 = 0$ which is $y^2 = |x|$.

(ii) The nullcline for $\dot{y}$ is $y = 0$ (the x-axis) and the point $y^2 + x^4 = 0$, i.e., $x = 0$, $y = 0$.

The $\dot{x}$ and $\dot{y}$ nullclines intersect at the origin giving one equilibrium solution $\overline{\mathbf{x}} = (0, 0)$.

The nullclines partition $\mathbb{R}^2$ into eight sectors in which the signs of the $\dot{x}$ and $\dot{y}$ do not change. Here is a picture of the vector field (lengths of the vectors are modified) along with the x-nullclines $y^2 = |x|$ and $x = 0$ shown in yellow and the y nullcline $y = 0$ in blue.



(i) For the sector on the top left we have $\dot{x} < 0$ and $\dot{y} < 0$.

(ii) For the sector on the top right we have $\dot{x} > 0$ and $\dot{y} < 0$, etc.

(iii) When a solution crosses the yellow curve $y^2 = |x|$ the sign of $\dot{x}$ switches.

(iv) Solutions cannot cross the y-axis (part of the nullcline for $\dot{y}$) nor can solutions cross the x-axis (the nullcline for $\dot{y}$).

(v) This implies that a solution that starts in a quadrant stays in that quadrant.

(vi) The orbits shown (green) and vector field indicate that the equilibrium solution at the origin is asymptotically stable.

From the first-order derivative of the components of the vector field,

$$\frac{\partial f_1}{\partial x} = -3x^2 + y^4 \qquad \frac{\partial f_1}{\partial y} = 4xy^3$$
$$\frac{\partial f_2}{\partial x} = -4x^3 y \qquad \frac{\partial f_2}{\partial y} = -3y^2 - x^4,$$

the matrix for the linearization is

$$A = D\mathbf{f}(0,0) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

The eigenvalues of A are 0 and 0. Hence this equilibrium is not hyperbolic, so that the theory developed above does not apply.

5.3 *Epidemics and Structural Stability

This section explores the dynamics of SIR model and Lotka–Volterra models. The SIR model has equilibrium solutions that seem to be stable, but they are not hyperbolic, which means we can't use linearization and the Asymptotic Stability Theorem to show that they are stable. The Lotka–Volterra models are interesting because they give an example of *structural instability*, which, roughly speaking, means that small change to the model's vector field results in a drastic change to the phase portrait. That's bad for modelers because in real life small measurement errors are inevitable, and those small changes change the vector field slightly, which changes the solutions drastically.

5.3.1 Epidemic modeling and intermediate evolution of dynamical systems

We now return to the SIR model introduced in the first Chapter of this volume in (1.9). For population ratios S, I, and R, representing susceptible, infected, and recovering, the SIR model is

$$\frac{dS}{dt} = -bSI, \quad \frac{dI}{dt} = bSI - kI, \quad \frac{dR}{dt} = kI, \tag{5.7}$$

where $b > 0$ is a constant describing the rate by which susceptible become inflected, and $k > 0$ is the rate by which infected overcome the infection.

The three population ratios S, I, and R all belong to $[0, 1]$ and are assumed to sum to 1 at all times,

$$S(t) + I(t) + R(t) = 1 \text{ for all } t.$$

The solutions we are interested in are contained in the triangle in the first octant of the plane

$$S + I + R = 1, \qquad S, I, R \in [0, 1],$$

in $\mathbb{R}^3$ with coordinates (S, I, R).

When we first introduced this model it made sense to consider the following assumptions:

(i) The total population is fixed.

(ii) Every individual will recover in a reasonable amount of time.

(iii) Once an individual has recovered, they are no longer susceptible to a recurrent infection.

(iv) All of the individuals mix with each other and that everyone is simultaneously exposed.

When we are analyzing the SIR model, the fundamental questions we ask are very different than what we have been looking at, and what we will ask of the Lotka–Volterra model for predator-prey dynamics. Nevertheless we will march forward with the tools that we have so carefully just developed.

Equilibria for SIR

The use of the nullclines of the SIR vector field in (5.7), used to find the equilibrium solutions, produces three surfaces in $\mathbb{R}^3$, which can be difficult to visualize. Instead find the equilibrium solutions by setting the vector field to be zero and solving algebraically:

$$-bSI = 0, \quad bSI - kI = 0, \qquad \text{and} \qquad kI = 0.$$

The last equation implies that $I = 0$, which implies the first two equations are also satisfied, meaning that S is arbitrary. Since R does not appear, then R is arbitrary as well.

This means, ignoring the constraint $S + I + R = 1$, that solutions of (5.7) have a plane of equilibrium solutions of the form

$$\{(S, 0, R) : S, R \in \mathbb{R}\} \subset \mathbb{R}^3.$$

Thus none of these equilibrium solutions are isolated.

Because none of the equilibria of the SIR model is isolated, Proposition 5.2.3 implies that none is hyperbolic, which means we cannot use the Asymptotic Stability Theorem (Theorem 5.2.2).

Eigenvalues of the SIR Linearization

Nevertheless, let us proceed as if we are misguided dynamicists who always consider linearization first no matter the inherent danger in doing so. The linearization of (1.9) leads to the derivative matrix

$$D\mathbf{f}(S, I, R) = \begin{bmatrix} -bI & -bS & 0 \\ bI & bS - k & 0 \\ 0 & k & 0 \end{bmatrix}.$$

Evaluation of this matrix at any $(S_0, 0, R_0)$ gives

$$A = D\mathbf{f}(S_0, 0, R_0) = \begin{bmatrix} 0 & -S_0 & 0 \\ 0 & bS_0 - k & 0 \\ 0 & k & 0 \end{bmatrix}.$$

A little work shows that eigenvalues of this matrix are: $\lambda = 0,\ 0,\ bS_0 - k$. The fact that the zero eigenvalue has multiplicity two reflects that the equilibrium $(S_0, 0, R_0)$ belongs to a 2-dimensional plane of equilibria.

Although we cannot use the Asymptotic Stability Theorem, the sign of the eigenvalue $bS_0 - k$ does hint at something dynamical. The change in the quantity I is determined by

$$\dot{I} = bSI - kI = (bS - k)I,$$

and the sign of $\dot{I}$ depends on $bS - k$. This leads to a change in the intermediate behavior of $I(t)$: as S decreases (it decreases because $\dot{S} = -bSI < 0$) eventually $bS - k$ becomes negative, making $\dot{I}$ negative, meaning that I is decreasing.

Dimension Reduction via the Population Constraint

With the constraint that

$$S(t) + I(t) + R(t) = 1, \text{ for } \ S(t), I(t), R(t) \in [0, 1],$$

the dynamics are really 2-dimensional, and we can reduce the number of differential equations from 3 to 2. Once we know two of the quantities $S(t)$, $I(t)$, and $R(t)$, the constraint $S(t) + I(t) + R(t) = 1$ uniquely determines the value of the remaining quantity. We choose to determine $S(t)$ from $I(t)$ and $R(t)$, so we rely on the triangle given by

$$I(t) + R(t) \leq 1, \text{ for } \ I(t), R(t) \in [0, 1],$$

in the (I, R) coordinate plane of the full (S, I, R)-space.

From $S(t) + I(t) + R(t) = 1$, let

$$S(t) = 1 - I(t) - R(t),$$

so the differential equation for $I(t)$ becomes

$$\dot{I} = (b(1 - I - R) - k)I.$$

The differential equation for $R(t)$ remains the same ($\dot{R} = kI$), since it did not contain $S(t)$. Now we have two differential equations in the two quantities $I(t)$ and $R(t)$. The $\dot{I}$ nullcline is

$$\begin{aligned} (b(1 - I - R) - k)I &= 0, \\ b(1 - I - R) - k = 0 \text{ or } \ I &= 0. \end{aligned}$$

The second of these corresponds to the interval $[0,1]$ on the vertical R axis in the (I,R)-plane, while the first of these (after rewriting) is the line

$$R = -I + 1 - \frac{k}{b}.$$

If $1 - \frac{k}{b} < 0$ then $b < k$ and this line does not appear in the triangle required by the constant population constraint in the (I,R)-plane.

The $\dot{R}$ nullcline is $I = 0$, which is the interval $[0,1]$ on the vertical R axis. The nullclines intersect at the points $(0, R_0)$ for all $R_0 \in [0,1]$; these are the equilibrium solutions and they are not isolated (hence not hyperbolic).

From the vector field

$$\mathbf{f}(I,R) = (f_1(I,R), f_2(I,R)) = (bI - bI^2 - bIR - kI, kI)$$

we have

$$D\mathbf{f}(I,R) = \begin{bmatrix} b - 2bI - bR - k & -bI \\ k & 0 \end{bmatrix}.$$

The linearization at an equilibrium $(0, R_0)$, $R_0 \in [0,1]$, is

$$A = D\mathbf{f}(0, R_0) = \begin{bmatrix} b - bR_0 - k & 0 \\ k & 0 \end{bmatrix}.$$

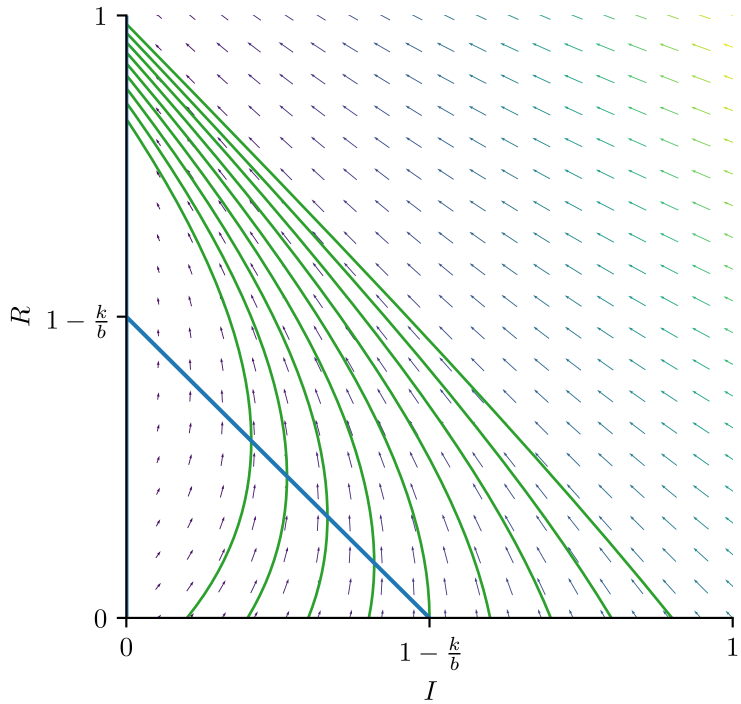
The eigenvalues of the lower-triangular matrix A are $b - bR_0 - k$ and 0. As expected, the equilibrium solutions are not hyperbolic since the equilibria are not isolated. The eigenvalue $b - bR_0 - k$ appears to differ from the eigenvalue $bS_0 - k$ we saw from the 3×3 matrix. But recall that the constraint $1 = S + I + R$ and the equilibrium $(0, R_0)$ imply the value of S_0 satifies $1 = S_0 + 0 + R_0$, i.e., $S_0 = 1 - R_0$, so that

$$b - bR_0 - k = b(1 - R_0) - k = bS_0 - k,$$

hence they are the same eigenvalue.

To get a visualization of the 2-dimensional dynamics fix $b = 0.2$ and $k = 0.1$ and plot the vector field, the nullcline $R + I = 1 - \frac{k}{b}$ for $\dot{I}$ where this derivative changes sign (the blue line), and a sampling of solutions that start with initial conditions of the form $I > 0$ and $R = 0$ (the green curves).

Here is a graph with all of this information plotted in the (I,R) plane.



For orbits that cross the blue nullcline, what does the crossing represent? It is the point where $\dot{I} = 0$, where the $\dot{I}$ changes from positive to negative, that is, the maximum infection percentage.

From the graph it appears that solutions starting near an equilibrium solution $(0, R_0)$ above the blue line stay near that equilibrium solution so it should be stable. Unfortunately the linearization is not sufficient to justify this, and so we rely on the vector field description to understand the qualitative nature of the solution.

5.3.2 Lotka–Volterra and structural stability

The rigorous definition of structural stability for a C^1 system $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ requires concepts beyond this course. But the basic question of structurally stability is this: Are the phase portraits of $\dot{\mathbf{x}} = \mathbf{g}(\mathbf{x})$ qualitatively the same as that of $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ for all vector fields $\mathbf{g}$ close to $\mathbf{f}$?

This is a serious question of modeling: if the vector field $\mathbf{f}$ is perturbed slightly, will the phase portrait stay qualitatively the same or will it change dramatically? A dramatic change indicates that the model defined by the vector field $\mathbf{f}$ is structurally unstable. We illustrate structurally instability in the ever-so-slight perturbation from the Lotka–Volterra first-order model to the Lotka–Volterra second-order model.

Example 5.3.1 (Lotka–Volterra first order). Recall the nondimensional form of the first-order Lotka–Volterra system:

$$\dot{x} = bx - xy,$$
$$\dot{y} = -y + xy.$$

We first compute the nullclines for this system.

(i) $\dot{x} = 0$ leads to nullclines defined by

$$x(b - y) = 0 \Rightarrow x = 0, \quad y = b.$$

Note that this means that all trajectories crossing these two lines must be vertical in the $x - y$ plane so that no solution can cross the y axis. This is comforting because it assures us that if the prey population x is positive initially then it must remain positive for all time (imagining a negative population is a headache inducing exercise).

(ii) The $\dot{y} = 0$ nullclines are similarly given by

$$y = 0, \quad x = 1,$$

implying that solutions can not cross the x axis either. The general picture is illustrated in phase space in Figure 5.11.

The fixed points of this system are at the intersection of the nullclines, at $(0, 0)$ and $(1, b)$ the last one being the coexistence fixed point where both species can survive simultaneously. As shown in the homework, the origin is an unstable saddle. Linearization for the coexistence fixed point indicates that solutions are centers, i.e. purely oscillatory cycles about this fixed point. Interestingly this is exactly what happens in the full nonlinear system although we don't show this explicitly here.

The direction of the vector field shown in Figure 5.11 can be found by simply evaluating the vector field at different points in this region. We know that the sign of each component of the vector field can only change by passing through a nullcline, so we need only consider a single point in each of the four delineated regions. For instance, if we consider a point lying above the line $y = b$ then we see that $\dot{x} < 0$ meaning the flow field must point to the left. In addition for points to the right of the line $x = 1$ then $\dot{y} > 0$.

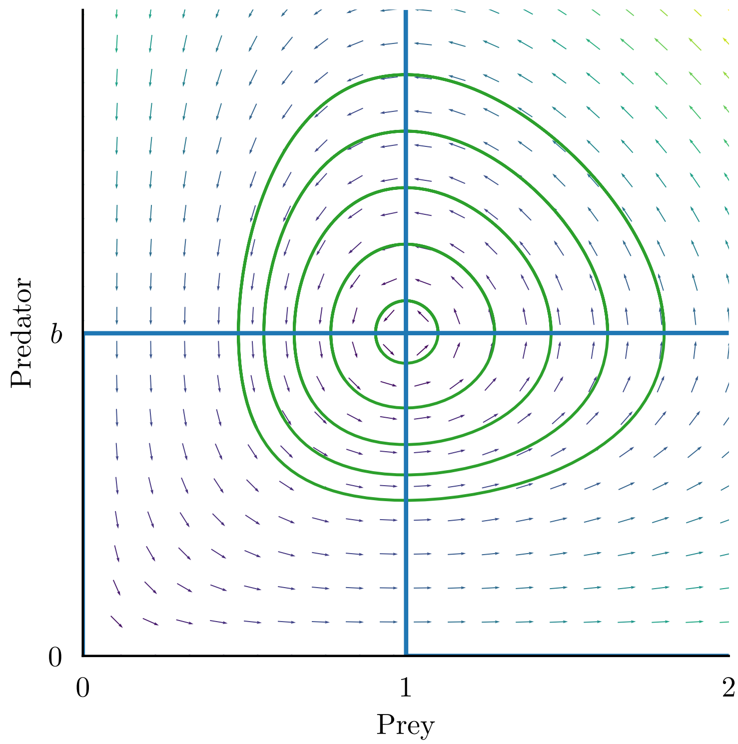


Figure 5.11: Here is a graph with the vector field (rescaled so the vectors don't overlap each other), the nullclines $y = b$ and $x = 1$ in blue, and some periodic solutions (the green closed curves).

If this model were structurally stable, then modifying it ever so slightly should result in the same type of dynamics, that is oscillating solutions that cycle around the coexistence equilibrium. On the other hand if this model is *not* structurally stable then the dynamics will be fundamentally different when the setup changes slightly. As noted in Chapter 1, the dynamics change substantially if we introduce a finite carrying capacity for the prey's resources. This occurs even if the carrying capacity is some very large (but finite) value. This can be seen clearly by analyzing the local stability of each of the equilibria for the model with carrying capacity introduced.

Example 5.3.2 (Lotka–Volterra second order model). Consider the second order Lotka–Volterra model which has the nondimensional form:

$$\dot{x} = bx - gx^2 - xy,$$
$$\dot{y} = -y + xy.$$

Nullclines for $\dot{x} = 0$ are given by:

$$x = 0, \quad y = b - gx,$$

and for $\dot{y} = 0$

$$y = 0, \quad x = 1.$$

This now gives three equilibria, $(0, 0)$, $(\frac{b}{g}, 0)$ and $(1, b - g)$ where the second equilibria is introduced in this iteration of the model as the extinction of the predator population, and when the prey reaches the carrying capacity of the environment.

The stability of the origin is unchanged by the introduction of a carrying capacity, and interestingly the predator extinction equilibrium when the prey are at carrying capacity is a saddle (as you find out in the homework). Of most significant interest, for a reasonable choice of the model parameters b and g, the coexistence fixed point is now asymptotically stable for this model configuration (you will see this in the homework, how exciting!).

This is demonstrated in Figure 5.12 where a specific set of parameter values has led to convergence of all solutions regardless of initial state, to the co-existence equilibrium. As the carrying capacity is decreased this fundamentally changes as illustrated in Figure 5.13.

The key difference between this version of the model and the previous one is that any restriction on the resources available to the prey has changed the coexistence fixed point from a center that has solutions oscillating around it, to a source or sink (depending on exact values of the parameters). This is a significant change that alters the very nature of all the possible solutions, independent of what the initial conditions were.

Such a drastic change in the dynamics even for a minor change (adjusting the carrying capacity to some extremely large yet finite value) to the system indicates that neither of these versions of the Lotka–Volterra system are structurally stable.

Remark 5.3.3. Structural stability is generally one of the most pertinent kinds of stability. Without structural stability we would need to question the validity of our model and possibly even the approach we are taking to model it. On the other hand, if we have a model that is structurally stable, but perhaps gives some less than desirable results compared to reality, then we would only need to reconsider the precise parameters that we are using; for example, perhaps we do not have the exact values of b and g specified and need to look at the data to obtain the correct values. This kind of situation is what we hope to reach, one where we believe our model is structurally correct, and we only need to identify the best values of the relevant parameters that fit the existing data the 'best'.

Unfortunately we aren't there yet for the predator-prey system. Modifying the two examples of this section significantly alters the dynamics of the phase space, which is something that we want to avoid. We have still not 'solved' our structural stability problem for the Lotka–Volterra model, but first we need to introduce some more ideas on stability before we can see what really is happening and how we can 'fix' it.

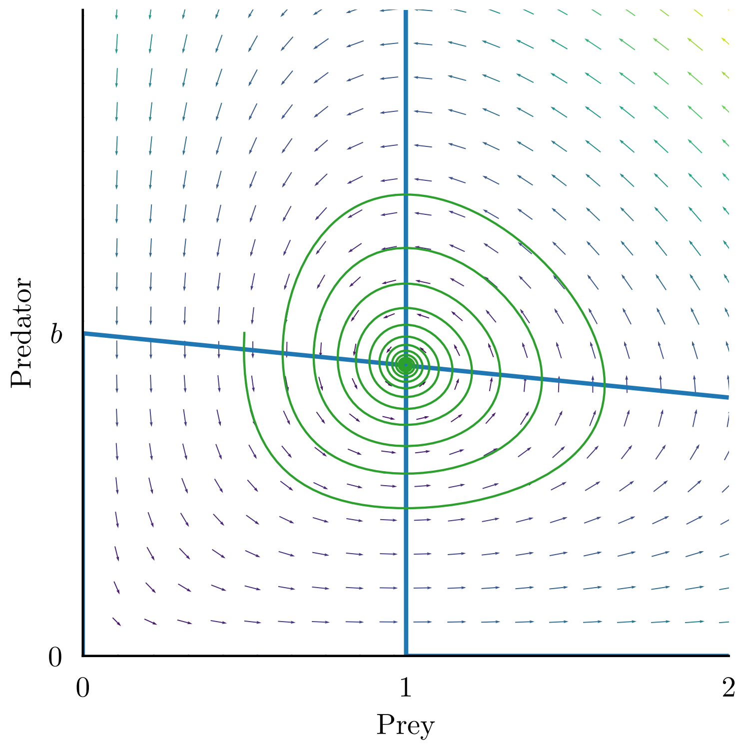


Figure 5.12: Here is a graph with the vector field (scaled so the vectors do not overlap) of the second order Lotka–Volterra model, the nullclines $x = 1$ and $y = b - gx$, and a spiral solution that appears to be converging to the equilibrium at $(1, \frac{b}{g})$.

5.4 Lyapunov's method

Linearization at an equilibrium provides one method by which asymptotical stability may be determined for the nonlinear dynamics near the equilibrium. This is one of the most common methods for investigating asymptotic stability, but as we have seen it is applicable only for hyperbolic equilibria which severely limits the applicability of the technique.

This section introduces another technique known as *Lyapunov's method*.[27] The basic idea is to find a function $V(\mathbf{x}(t))$ where $\mathbf{x}(t)$ is the solution of the differential equation in question. Lyapunov's method treats $V(\mathbf{x}(t))$ like a distance function from the equilibrium and shows that values of $V(\mathbf{x}(t))$ decreases enough along all orbits near the equilibrium to give asymptotic stability.

[27] Pronounced LYAh pooh-NOFF.

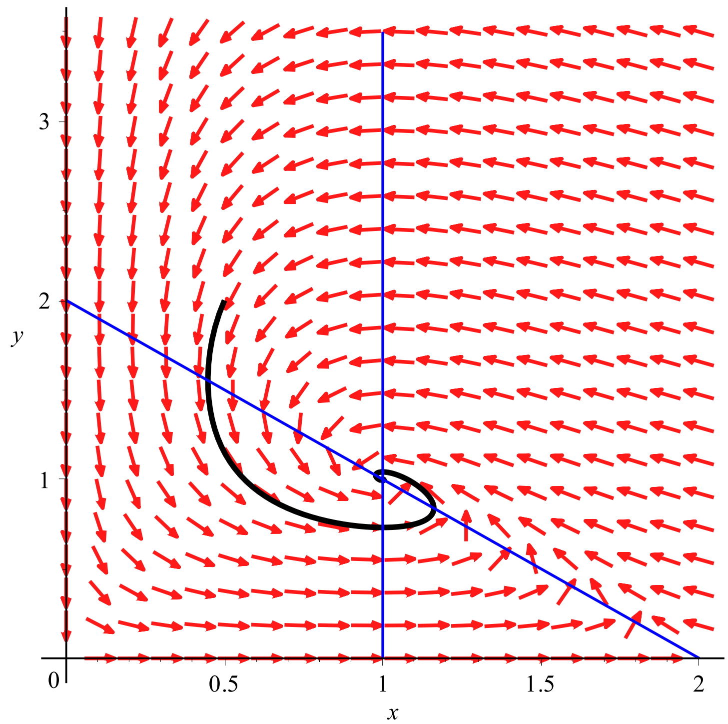


Figure 5.13: To “see” the new equilibrium $(\frac{b}{g}, 0)$ we increase g to $g = 1$. Here is a graph of the vector field (with the vectors scaled so they don’t overlap each other), the nullclines in blue, one solution (the black curve), and the third equilibrium now appearing at $(\frac{b}{g}, 0) = (2, 0)$.

When we suspect from the vector field and/or phase portrait that an equilibrium is asymptotical stable but the linearization method fails to show it is asymptotically stable, we may be able to use Lyapunov’s method to show that it is asymptotically stable.

Remark 5.4.1. The tricky part of Lyapunov’s method is finding the function $V(\mathbf{x}(t))$. The total energy of the system is often a reasonable choice when constructing such a function for physical systems (what does total energy mean for population dynamics though?). Until the last two decades there has been no systematic method for constructing Lyapunov functions, and the choice of such functions has been referred to in the literature as given by ‘divine inspiration’. In the very recent past, an algebraic technique using sums of squares has been used to find Lyapunov functions for dynamical systems with a polynomial basis. This technique relies heavily on recent developments in semi-definite programming that were not available just a few decades ago.

Remark 5.4.2. An added bonus with Lyapunov's method is that we can get stability, which we didn't get with linearization. This occurs when we can find a function $V(\mathbf{x}(t))$ whose value does not increase along orbits near the equilibrium. When we suspect from the vector field and/or phase portrait that an equilibrium is stable but not asymptotically stable we seek for a function $V(\mathbf{x}(t))$ whose value is constant along all orbits near the equilibrium.

Example 5.4.3. The second order scalar equation

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

is called a *conservative equation* because it has a conserved quantity, often called an *energy*, associated with it. To identify the conserved quantity for this system, multiply the equation by dx/dt to get

$$\frac{dx}{dt}\frac{d^2x}{dt^2} = -\frac{dV}{dx}\frac{dx}{dt},$$

and then recognize, by the chain rule, that the term on the left is

$$\frac{1}{2}\frac{d}{dt}\left(\frac{dx}{dt}\right)^2$$

and the term on the right is

$$-\frac{d}{dt}V(x).$$

Thus the second order equation becomes

$$\frac{1}{2}\frac{d}{dt}\left(\frac{dx}{dt}\right)^2 = -\frac{d}{dt}V(x).$$

Integration gives

$$\frac{1}{2}\left(\frac{dx}{dt}\right)^{1/2} = -V(x) + E.$$

for an arbitrary constant E (for energy). Rewriting the integrated relation gives *total energy* or simply *energy* for short,

$$E\left(x(t), \frac{dx}{dt}\right) = \frac{1}{2}\left(\frac{dx}{dt}\right)^{1/2} + V(x).$$

We think of E as a function of x and dx/dt, where the first term $K = (1/2)(dx/dt)^2$ is called the *kinetic energy* and the second term $P = V(x)$ is called the *potential energy*.

In the homework you show that for these types of equations the energy $E = K + P$ is a conserved quantity, that is,

$$\frac{d}{dt}E\left(x, \frac{dx}{dt}\right) = 0.$$

The conserved quantity E is an example of an *invariant* or *integral* of the differential equation.

Definition 5.4.4. *For a first-order system* $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ *where* $\mathbf{f} : \Omega \to \mathbb{R}^n$ *is* C^1 *for* Ω *open in* $\mathbb{R}^n$*, a* C^1 *function* $N : \Omega \to \mathbb{R}$ *is an* invariant *if*

$$\frac{dN}{dt} = DN(\mathbf{x})\frac{d\mathbf{x}}{dt} = DN(\mathbf{x})\mathbf{f}(\mathbf{x}) = 0.$$

The derivative here is called the orbital derivative *since it is being evaluated along solutions of the differential equation.*

Remark 5.4.5. Recall the constraint $S + I + R = 1$ for the SIR model. This constraint is an invariant for the SIR model because the time derivative of the function $N = 1 - S - I - R$ is zero. This means that solutions that start on the plane $N = 0$ stay on the plane $N = 0$.

Why do we care about invariants or integrals? As an example, we saw with the three-dimension SIR model that the invariant in this case allowed us to reduce the number of differential equations from three to two. This is a very significant simplification.

The existence of an integral for a planar systems actually means much more.

Example 5.4.6 (Nonlinear pendulum). For the conservative system

$$\frac{d^2x}{dt^2} = -\frac{dV}{dx} = -\sin(x) \quad \text{where} \quad V(x) = 1 - \cos(x) \geq 0$$

the energy invariant is

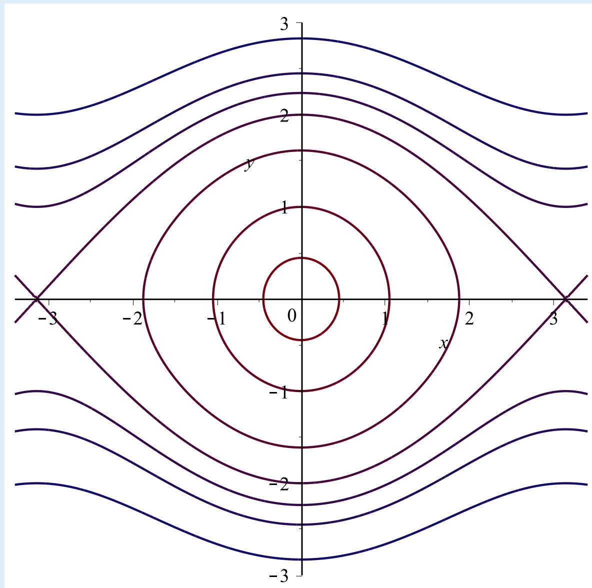
$$E = \frac{1}{2}\left(\frac{dx}{dt}\right)^2 + V(x) = \frac{1}{2}\left(\frac{dx}{dt}\right)^2 + 1 - \cos(x).$$

[Note: we choose $V(x) = 1 - \cos(x)$ instead of $V(x) = -\cos(x)$ so that the potential energy is positive. Notice that the constant 1 makes no difference to the differential equation.]

Setting $y = dx/dt$ we convert this second order equation into a first-order system:

$$\begin{aligned} \frac{dx}{dt} &= y, \\ \frac{dy}{dt} &= \frac{d^2x}{dt^2} = -\sin(x). \end{aligned}$$

Here is a graph showing a sampling of the level curves of the energy invariant E.



The level curves of E are orbits of the phase portrait; we have "solved" this conservative system.

Near the equilibrium at the origin the solutions move clockwise; on the top the solutions move left to right, and on the bottom the solutions move right to left.

What about the solutions that approach the equilibrium solutions on the x-axis near $-\pi$ and π? What do you expect will happen here?

Example 5.4.7. The second order scalar equation

$$\frac{d^2x}{dt^2} + \delta\frac{dx}{dt} + \frac{dV}{dt} = 0,$$

for a given potential energy function $V(x) \geq 0$ and positive $\delta > 0$, is no longer a conservative equation. The energy function

$$E = \frac{1}{2}\left(\frac{dx}{dt}\right)^2 + V(x)$$

is no longer conserved because, as shown in the homework, it satisfies

$$\frac{d}{dt}E\left(x, \frac{dx}{dt}\right) \leq 0.$$

Remark 5.4.8. We have now seen the use of functions from phase space to $\mathbb{R}$ that help in the understanding of behavior of orbits. This is the basic idea of Lyapunov functions and stability theory which we develop next.

Before finalizing the formal definitions below, we require a few standard notational considerations:

(i) For open $\Omega \subset \mathbb{R}^n$ let $\mathbf{f} : \Omega \to \mathbb{R}^n$ be C^1.

(ii) Suppose that $\overline{\mathbf{x}}$ is an equilibrium of $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ and WLOG assume that $\overline{\mathbf{x}} = 0$ (by translation), i.e., $\mathbf{f}(0) = 0$.

Definition 5.4.9. *For an open set Ω in $\mathbb{R}^n$ containing $\mathbf{0}$, a function $V : \Omega \to \mathbb{R}$ is* positive definite *if $V(\mathbf{x}) > 0$ for all $\mathbf{x} \in \Omega \setminus \{0\}$. A function $V : \Omega \to \mathbb{R}$ is* negative definite *if $-V$ is positive definite.*

Definition 5.4.10. *A C^1 function $V : \Omega \to \mathbb{R}$ is a* Lyapunov function *for the equilibrium at the origin of the vector field $\mathbf{f} : \Omega \to \mathbb{R}^n$ if V is positive definite on an open subset U of Ω containing 0, if $V(\mathbf{0}) = 0$, and if its orbital (time) derivative satisfies*

$$\frac{d}{dt}V(\mathbf{x}) = D_{\mathbf{x}}V(\mathbf{x})\frac{d\mathbf{x}}{dt} = DV(\mathbf{x})\mathbf{f}(\mathbf{x}) \leq 0 \quad \textit{for} \quad \textit{all} \quad \mathbf{x} \in \Omega.$$

Theorem 5.4.11 (Lyapunov's Local Stability Theorem). *For a C^1 vector field $\mathbf{f} : \Omega \to \mathbb{R}^n$ where Ω is open in $\mathbb{R}^n$, suppose $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ has an equilibrium $\overline{\mathbf{x}} = 0$. If there exists a Lyapunov function $V : \Omega \to \mathbb{R}$ for the equilibrium $\overline{\mathbf{x}}$, then $\overline{\mathbf{x}} = 0$ is a stable equilibrium.*

Proof. Choose $\varepsilon > 0$ small enough so the $\overline{B_\varepsilon(0)} \subset \Omega$. Set m_ε to be the minimum of the continuous $V(\mathbf{x})$ on the compact set

$$S_\varepsilon = \{\mathbf{x} \in \Omega : \|\mathbf{x}\| = \varepsilon\},$$

i.e. $m_\varepsilon = \min_{\mathbf{x} \in S_\varepsilon} V(\mathbf{x})$. Since $V(\mathbf{x})$ is positive definite, it follows that $m_\varepsilon > 0$.

Because $V(\mathbf{x})$ is continuous and $V(0) = 0$, there exists $\delta > 0$ such that $\|\mathbf{x}\| < \delta$ implies that $V(\mathbf{x}) < m_\varepsilon$. For an initial condition $\mathbf{x}_0$ satisfying $\|\mathbf{x}_0\| < \delta$, let $\mathbf{x}(t)$ be the solution satisfying $\mathbf{x}(0) = \mathbf{x}_0$.

Since $V(\mathbf{x})$ is a Lyapunov function,

$$\frac{d}{dt}V(\mathbf{x}(t)) \leq 0.$$

Integrating this over the interval $[0, t]$ for $t \geq 0$ gives

$$V(\mathbf{x}(t)) \leq V(\mathbf{x}(0)) < m_\varepsilon. \tag{5.8}$$

We want to show that $\|\mathbf{x}_0\| < \delta$ implies the solution $\mathbf{x}(t)$ with initial condition $\mathbf{x}_0$ satisfies $\|\mathbf{x}(t)\| < \varepsilon$ for all $t \geq 0$, which yields stability.

We argue by contradiction: suppose for $\|\mathbf{x}_0\| < \delta$ there exists $t_1 > 0$ such that the solution $\mathbf{x}(t)$ with initial condition $\mathbf{x}_0$ satisfies $\|x(t_1)\| = \varepsilon$, i.e., that $\mathbf{x}(t_1) \in S_\varepsilon$. Since m_ε is the minimum of $V(\mathbf{x})$ on S_ε, the condition $\|\mathbf{x}(t_1)\| = \varepsilon$ implies that

$$V(\mathbf{x}(t_1)) \geq m_\varepsilon.$$

But this contradicts the inequality (5.8), i.e., that the value of $V(\mathbf{x})$ along the solution was bounded above by m_ε.

Thus we obtain for all small enough $\varepsilon > 0$ there exist $\delta > 0$ such that for $\|\mathbf{x}_0\| < \delta$ the solution $\mathbf{x}(t)$ with $\mathbf{x}(0) = \mathbf{x}_0$ satisfies $\|x(t)\| < \varepsilon$ for all $t \geq 0$, i.e., the equilibrium at the origin is stable. □

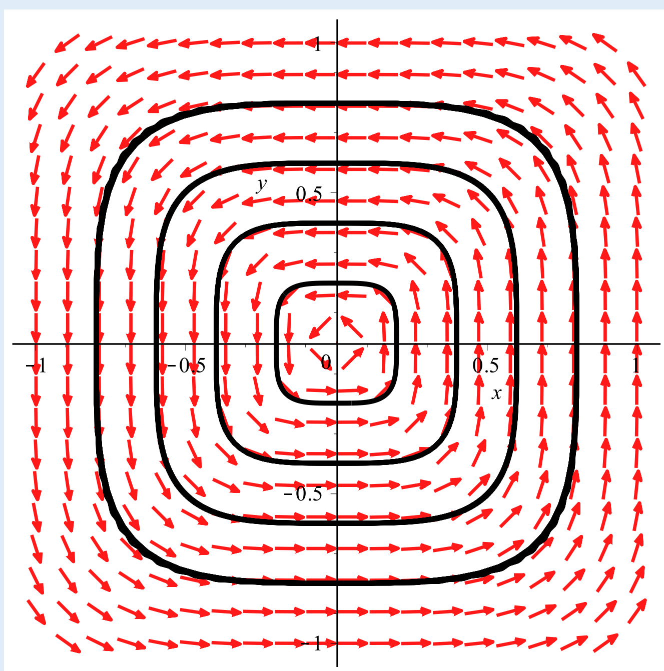
Remark 5.4.12. Lyapunov's Local Stability Theorem is the method to use to establish stability when linearization gives a center and there is evidence, such as the numerically generated phase portrait, which suggests stability but not asymptotic stability.

Example 5.4.13. The origin for the system

$$\dot{x} = -y^3 + x^2y^5,$$
$$\dot{y} = x^3 - x^5y^2,$$

is an equilibrium whose linearization is the zero matrix.

Here is a graph with the vector field (scaled so the vectors don't overlap) and some solutions plotted to get the sense the the origin is stable but not asymptotically stable.



The nullclines in the graph are the x-axis and the y-axis.

A Lyapunov function for the equilibrium at the origin of the vector field is

$$V(x, y) = x^4 + y^4.$$

It is easy to check that this is positive definite, that $V(0, 0) = 0$, and its orbital derivative is

$$\begin{aligned}\frac{dV}{dt} &= DV(x, y)\mathbf{f}(x, y) \\ &= (4x^3, 4y^3) \cdot (-y^3 + x^2y^5, x^3 - x^5y^2) \\ &= 4x^3(-y^3 + x^2y^5) + 4y^3(x^3 - x^5y^2) \\ &= -4x^3y^3 + 4x^5y^5 + 4y^3x^3 - 4x^5y^5 \\ &= 0.\end{aligned}$$

Thus by Lyapunov's Local Stability Theorem the equilibrium at the origin is stable.

If the orbital derivative of a Lyapunov function satisfies a stronger inequality, then we get a stronger stability result.

Theorem 5.4.14 (Lyapunov's Local Asymptotic Stability Theorem). *For a C^1 vector field $\mathbf{f} : \Omega \to \mathbb{R}^n$ where Ω is open in $\mathbb{R}^n$, suppose $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ has an equilibrium $\overline{\mathbf{x}} = 0$. If there exists a Lyapunov function $V : \Omega \to \mathbb{R}$ such that its orbital derivative satisfies $(d/dt)V(\mathbf{x}) < 0$ for all nonzero $\mathbf{x}$ in a neighbourhood of 0, then the equilibrium $\overline{\mathbf{x}} = 0$ is asymptotically stable.*

Proof. The stronger condition on the orbital derivative of $V(\mathbf{x})$ implies that it satisfies $(d/dt)V(\mathbf{x}) \geq 0$ for all $\mathbf{x}$ in a neighborhood of 0. So by the Lyapunov Local Stability Theorem the equilibrium at 0 is stable.

Thus for a small $\varepsilon > 0$ there exists $\delta > 0$ such that for any solution $\mathbf{x}(t)$ with initial condition $\mathbf{x}(0) = \mathbf{x}_0$ satisfying $\|\mathbf{x}_0\| < \delta$ there holds $\|\mathbf{x}(t)\| < \varepsilon$ for all $t \geq 0$.

If $\mathbf{x}_0 = 0$, there is nothing is show. Hence take $\mathbf{x}_0 \neq 0$, and suppose to the contrary that $\mathbf{x}(t)$ is bounded away from 0 for $t \geq 0$, i.e., there exists $a > 0$ such that

$$a \leq \|\mathbf{x}(t)\| \leq \varepsilon \quad \text{for all } t \geq 0.$$

The set

$$K = \{\mathbf{x} \in \Omega : a \leq \|\mathbf{x}\| \leq \varepsilon\}$$

is compact and on this compact set the negative orbital derivative $(d/dt)V(\mathbf{x})$ has a maximum value $-\mu$ for $\mu > 0$, i.e.,

$$\frac{d}{dt}V(\mathbf{x}) \leq -\mu \quad \text{for all } \mathbf{x} \in K.$$

Integration of $(d/dt)V(\mathbf{x}(t))$ over the interval $[0, t]$ gives

$$V(\mathbf{x}(t)) - V(\mathbf{x}(0)) = \int_0^t \frac{d}{ds}V(\mathbf{x}(s))\, ds \leq -\int_0^t \mu\, ds = -\mu t.$$

Rearranging this gives

$$V(\mathbf{x}(t)) \leq V(\mathbf{x}_0) - \mu t. \tag{5.9}$$

The Lyapunov function V is positive definite on Ω which includes K, but as $t \to \infty$ inequality (5.9) implies that $V(\mathbf{x}(t))$ becomes negative on K, a contradiction.

Thus the solution $\mathbf{x}(t)$ can not be bounded away from the origin, i.e., for all $a > 0$ there exists $t \geq 0$ such that $\|\mathbf{x}(t)\| < a$. To show that $\mathbf{x}(t)$ converges to 0 as $t \to \infty$, we again proceed by contradiction, that is suppose there exists $0 < \eta < \varepsilon$ such that for all $T \geq 0$ there exists $t > T$ such that $\|\mathbf{x}(t)\| \geq \eta$. This implies there exists a sequence $\{t_n\}$ with $t_n > n$ such that $\|\mathbf{x}(t_n)\| \geq \eta$.

For $0 \leq s_1 < s_2$ the negative orbital derivative implies that

$$V(\mathbf{x}(s_2)) - V(\mathbf{x}(s_1)) = \int_{s_1}^{s_2} \frac{d}{ds} V(\mathbf{x}(s))\, ds < 0.$$

From this it follows for $0 \leq s_1 < s_2$ that V is strictly decreasing along $\mathbf{x}(t)$,

$$V(\mathbf{x}(s_2)) < V(\mathbf{x}(s_1)).$$

Applying the strictly decreasing property of $V(\mathbf{x}(t))$ to the sequence $\{\mathbf{x}(t_n)\}$ we obtain the strictly decreasing sequence $\{V(\mathbf{x}(t_n))\}$.

The sequence $\{V(\mathbf{x}(t_n))\}$ lies in the compact set $\{\mathbf{x} \in \Omega : \eta \leq \|\mathbf{x}\| \leq \varepsilon\}$ on which the minimum value $V_{\min}$ of the positive definite V is positive, implying that

$$V(\mathbf{x}(t_n)) \geq V_{\min} \quad \text{for all} \quad n.$$

On the other hand, because $\mathbf{x}(t)$ is not bounded away from the origin, there exists $t^* \geq 0$ for which $\|\mathbf{x}(t^*)\|$ is sufficiently small enough to imply by the continuity of $V(\mathbf{x})$ and $V(0) = 0$ that $V(\mathbf{x}(t^*)) < V_{\min}$.

The value of t^* satisfies

$$t_n < t^* < t_{n+1}$$

for some n and gives the inequalities

$$V(\mathbf{x}(t_n)) > V(\mathbf{x}(t^*)) \quad \text{and} \quad V(\mathbf{x}(t^*)) < V(\mathbf{x}(t_{n+1})).$$

But the last inequality is a contradiction to the Lyapunov function being strictly decreasing along $\mathbf{x}(t)$. $\square$

Example 5.4.15. Recall the planar system

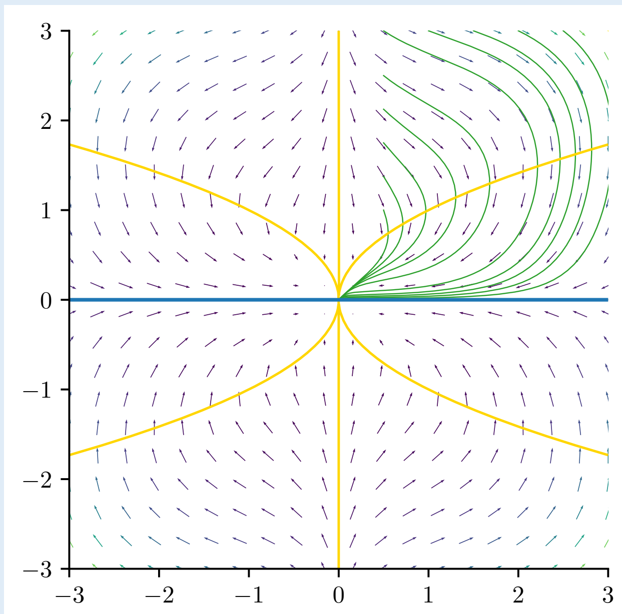
$$\begin{aligned} \dot{x} &= -x^3 + xy^4, \\ \dot{y} &= -y^3 - x^4 y. \end{aligned}$$

The vector field for this system is

$$\mathbf{f}(x, y) = (-x^3 + xy^4, -y^3 - x^4 y).$$

The vector field and phase portrait for the system are shown below. We concluded from this phase portrait the the equilibrium at the origin is asymptotically stable, but the linearization of the vector field at the origin gave the zero matrix which means that the linearization is completely uninformative for the nonlinear dynamics.

We use Lyapunov's Local Asymptotic Stability Theorem to prove the equilibrium at the origin is asymptotically stable.



A Lyapunov function for the equilibrium of the vector field is

$$V(x,y) = x^4 + y^4$$

because it is positive definite, $V(0,0) = 0$, and its orbital derivative is

$$\begin{aligned}
\frac{d}{dt}V(x,y) &= DV(x,y)\mathbf{f}(x,y) \\
&= (4x^3, 4y^3) \cdot (-x^3 + xy^4, -y^3 - x^4 y) \\
&= 4x^3(-x^3 + xy^4) + 4y^3(-y^3 - x^4 y) \\
&= -4x^6 + 4x^4y^4 - 4y^6 - 4x^4y^4 \\
&= -4(x^6 + y^6) \\
&< 0
\end{aligned}$$

for all nonzero (x, y) in a neighborhood of the origin. Thus by Lyapunov's Local Asymptotic Stability Theorem, the origin is asymptotically stable.

Example 5.4.16 (Damped nonlinear pendulum). The damped nonlinear pendulum is modeled by the second order scalar equation

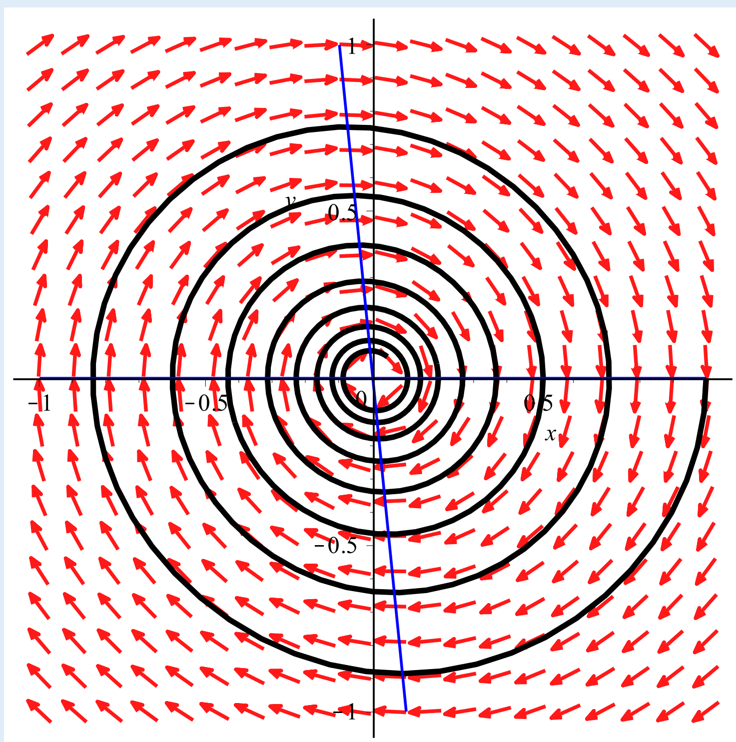
$$\ddot{\theta} + s\dot{\theta} + \sin\theta = 0,$$

where θ is the angle the pendulum makes with the downward direction.

Convert this into a first-order system by seting $x = \theta$ and $y = \dot{\theta}$, so that

$$\begin{aligned}\dot{x} &= \dot{\theta} = y,\\ \dot{y} &= \ddot{\theta} = -s\dot{\theta} - \sin\theta = -sy - \sin x.\end{aligned}$$

Here is a graph of the vector field, the nullclines of the x-axis and a line near the y-axis (in blue), and one solution suggesting the equilibrium at the origin is asymptotically stable.



We saw the undamped pendulum ($s = 0$) before which has an energy integral

$$E = \frac{1}{2}\left(\frac{dx}{dt}\right)^2 + 1 - \cos x = \frac{y^2}{2} + 1 - \cos x.$$

This function E is a Lyapunov function V since in a a neighborhood of the equilibrium at the origin V is positive definite, $V(0,0) = 0$, and its orbital derivative is

$$\begin{aligned}\frac{d}{dt}V(x,y) &= DV(x,y)\cdot \mathbf{f}(x,y)\\ &= (\sin x, y)\cdot(y, -sy-\sin x)\\ &= y\sin x - y\sin x - sy^2\\ &= -sy^2\\ &\le 0.\end{aligned}$$

The orbital derivative is not however strictly negative for all nonzero (x,y) in a neighbourhood of the origin, so we cannot apply the Lyapunov Asymptotic Stability Theorem.

We can only apply Lyapunov's Local Stability Theorem to get stability of the equilibrium at the origin.

Linearization of the vector field at the equilibrium at the origin gives the matrix

$$\begin{bmatrix} 0 & 1\\ -1 & -s\end{bmatrix}$$

whose characteristic polynomial is $\lambda^2 + s\lambda + 1$, hence whose eigenvalues are the complex conjugate with negative real part

$$\frac{-s \pm \sqrt{s^2-4}}{2}.$$

By the Asymptotical Stability Theorem for linearization the equilibrium at the origin is asymptotically stable.

The purpose of this last example was to show us that we need both methods of linearization and Lyapunov functions to establish asymptotic stability results, for sometimes one will fail. It is important to consider both methods when analyzing a new system. Complete dedication and devotion to a single approach will blind you to potential issues that may arise...blinders only work well for horses.

Example 5.4.17 (Lorenz equations). Many decades before most readers of this book were born, Edward Lorenz was a meteorologist (and applied mathematician) trying to come up with a definitive method for predicting weather patterns. Starting with the standard set of partial differential equations (PDEs) that describe convective processes (fluid motion driven by thermal buoyant instabilities), Lorenz derived a very simplified ODE that he numerically integrated (this was much more laborious than using Python's odeint).

At one point Lorenz wanted to reproduce some of his earlier calculations, but decided to restart his simulation midway through where he had left off previously. At the time this required him to enter (by hand) the initial values that he took from halfway through his previous simulation. Surprisingly (to him and the world at the time) almost immediately the solutions diverged from each other. After several agonizing days (weeks/months), Lorenz realized that while the computer he was using kept six significant digits, the printout that he had used to re-enter his initial values kept only three significant digits, and that this minor change (thought to be unimportant at the time) caused the solutions to diverge rapidly.

In his 1963 paper 'Deterministic Nonperiodic Flow' (appearing in *Journal of the Atmospheric Sciences*, Lorenz formalized these ideas, giving birth to *chaos theory*. In part of this article, he explains

> Two states differing by imperceptible amounts may eventually evolve into two considerably different states ... If, then, there is any error whatever in observing the present state—and in any real system such errors seem inevitable—an acceptable prediction of an instantaneous state in the distant future may well be impossible....In view of the inevitable inaccuracy and incompleteness of weather observations, precise very-long-range forecasting would seem to be nonexistent.

The simple three-dimensional model that Lorenz settled on and used in his seminal 1963 paper is given below:

$$\dot{x} = \sigma(y - x) \tag{5.10a}$$

$$\dot{y} = rx - y - xz \tag{5.10b}$$

$$\dot{z} = xy - bz. \tag{5.10c}$$

A rough interpretation of this system is that $x(t)$ and $y(t)$ represent different modes of the flow field, and $z(t)$ represents a component of the temperature. The parameters of the system are: 1) σ which is called the Prandtl number, and represents a material property of the fluid in question, 2) r which is proportional to the thermal driving force of the system, and 3) b which is a geometric parameter that dictates the size of the domain the fluid is constrained to be in.

This system of equations (5.10) is one of the most famous in differential equations. Although these equations formed the foundation of chaos theory, and for certain parameters, can yield extremely complicated (yet bounded) behavior, for other parameters the system remains quite calm.

In fact, if we let $V(x, y, z) = \frac{1}{\sigma}x^2 + y^2 + z^2$ then you can see in the exercises that $V' < 0$ so long as $r < 1$. This implies that the origin is asymptotically stable for the Lorenz equations so long as $r < 1$, which clearly is not a chaotic system. As r increases however, much more exciting things are bound to happen.

5.5 *More Lyapunov Stability

Definition 5.5.1. *A matrix $A \in M_n(\mathbb{F})$ is said to be* Hurwitz *if $\Re(\lambda) < 0$ for all $\lambda \in \sigma(A)$.*

5.5.1 Lyapunov's Equation

Lemma 5.5.2. *A matrix $A \in M_n(\mathbb{F})$ is Hurwitz if and only if given a positive definite $Q \in M_n(\mathbb{F})$ (hereafter denoted $Q > 0$), there exists a unique $P > 0$ such that*

$$A^{\mathsf{H}}P + PA = -Q. \tag{5.11}$$

We call (5.11) Lyapunov's equation.

Proof. ($\Longrightarrow$) If A is Hurwitz, then there exists $\eta > 0$ and $C > 0$ such that $\|e^{At}\| \leq Ce^{-\eta t}$. Given $Q > 0$, let

$$P = \int_0^\infty e^{A^{\mathsf{H}}t}Qe^{At}\,dt.$$

Since A is Hurwitz, we have that

$$\left|\int_0^\infty e^{A^{\mathsf{H}}t}Qe^{At}\,dt\right| \leq \|Q\|\int_0^\infty \|e^{At}\|^2\,dt \leq \|Q\|\int_0^\infty e^{-2\eta t}\,dt \leq \frac{\|Q\|}{2\eta}.$$

Thus, P is well defined and clearly Hermitian. Note that $\|e^{At}\mathbf{x}\| > 0$ for all $\mathbf{x} \neq 0$. Thus, $e^{A^{\mathsf{H}}t}Qe^{At} > 0$ and hence so is P. Finally, note that

$$\begin{aligned}
A^{\mathsf{H}}P + PA &= \int_0^\infty \left(A^{\mathsf{H}}e^{A^{\mathsf{H}}t}Qe^{At} + e^{A^{\mathsf{H}}t}Qe^{At}A\right)\,dt \\
&= \int_0^\infty \frac{d}{dt}\left(e^{A^{\mathsf{H}}t}Qe^{At}\right)\,dt \\
&= e^{A^{\mathsf{H}}t}Qe^{At}\bigg|_0^\infty = -Q.
\end{aligned}$$

To prove uniqueness, assume that $\widehat{P} > 0$ also satisfies (5.11). Thus, we have that

$$A^{\mathsf{T}}(P - \widehat{P}) + (P - \widehat{P})A = 0.$$

Multiplying on the left and right by $e^{A^{\mathsf{H}}t}$ and e^{At}, respectively, we have

$$e^{A^{\mathsf{H}}t}\left(A^{\mathsf{H}}(P - \widehat{P}) + (P - \widehat{P})A\right)e^{At} = 0.$$

This is equivalent to

$$\frac{d}{dt}e^{A^{\mathsf{H}}t}(P - \widehat{P})e^{At} = 0,$$

which integrates to

$$e^{A^{\mathsf{H}}t}(P - \widehat{P})e^{At} = C,$$

where C is a constant matrix. Since the exponentials converge to zero as $t \to \infty$ we have that $C = 0$, which implies that $e^{A^{\mathsf{H}}t}(P - \widehat{P})e^{At} = 0$ Since the exponentials are both invertible, we have that $P = \widehat{P}$.
($\Longleftarrow$) Assuming the existence of $P > 0$, given $Q > 0$, such that (5.11), we let $V(\mathbf{x}) = \mathbf{x}^{\mathsf{H}} P \mathbf{x}$. Thus,

$$\frac{d}{dt}V(\mathbf{x}) = \mathbf{x}^{\mathsf{H}\prime} P\mathbf{x} + \mathbf{x}^{\mathsf{H}} P\dot{\mathbf{x}} = \mathbf{x}^{\mathsf{H}} A^{\mathsf{H}} P\mathbf{x} + \mathbf{x}^{\mathsf{H}} PA\mathbf{x} = \mathbf{x}^{\mathsf{H}}(A^{\mathsf{H}}P + PA)\mathbf{x} = -\mathbf{x}^{\mathsf{H}} Q\mathbf{x}.$$

Thus, $\mathbf{x}(t) \to \mathbf{0}$ as $t \to \infty$ regardless of the initial point x_0. Thus, $\|e^{At}\| \to 0$, which implies that A is Hurwitz. □

Remark 5.5.3. Lyapunov's equation (5.11) is a special case of Sylvester's equation, which is $AX + XB = C$. Here one would be given A, B, and C and solve for X. In our formulation of Lyapunov's equation, we are finding P given A and Q.

Remark 5.5.4. If A is diagonalizable, we can prove the converse directly. Assume $A = SDS^{-1}$, where $D = \operatorname{diag}(\lambda_1, \cdots, \lambda_n)$ are the eigenvalues of A. Note that (5.11) can be written

$$S^{-\mathsf{H}} D^{\mathsf{H}} S^{\mathsf{H}} P + PSDS^{-1} = -Q.$$

Multiplying both sides on the left and right by S^{H} and S, respectively, we have

$$D^{\mathsf{H}}\widehat{P} + \widehat{P}D = -\widehat{Q}. \tag{5.12}$$

where $\widehat{P} = S^{\mathsf{H}} PS > 0$ and $\widehat{Q} = S^{\mathsf{H}} QS > 0$. Since the diagonals of a positive definite matrix are positive, we have that the diagonal elements of (5.12) take the form

$$p_{kk}(\lambda_k + \overline{\lambda}_k) = -q_{kk},$$

where $\widehat{P} = [p_{ij}]$ and $\widehat{Q} = [q_{ij}]$. Hence, we have that $\lambda_k + \overline{\lambda}_k = 2\Re(\lambda_k) < 0$, which implies that A is Hurwitz.

5.5.2 Stability Theorem Revisited

We re-do the stability theorem, but here we use Lyapunov functions.

Theorem 5.5.5. *Let Ω be an open subset of the Banach space X. Let $\overline{\mathbf{x}} \in \Omega$ be an equilibrium solution of the dynamical system $\dot{\mathbf{x}} = f(\mathbf{x})$, where $f : \Omega \to X$ is smooth and $A = Df(\overline{\mathbf{x}})$. If A is Hurwitz, then $\overline{\mathbf{x}}$ is asymptotically stable.*

Proof. Given that A is Hurwitz and that $Q > 0$, there exists $P > 0$ such that $A^{\mathsf{T}}P + PA = -Q$. Writing

$$\dot{\mathbf{x}} = f(\mathbf{x}) = A\mathbf{x} + (f(\mathbf{x}) - A\mathbf{x}) = A\mathbf{x} + g(\mathbf{x}),$$

where $g(\mathbf{x}) = f(\mathbf{x}) - A\mathbf{x}$. Thus,

$$\begin{aligned}\frac{d}{dt}V(\mathbf{x}) &= \mathbf{x}^{\mathsf{H}}P\dot{\mathbf{x}} + \dot{\mathbf{x}}^{\mathsf{H}}P\mathbf{x} \\ &= \mathbf{x}^{\mathsf{H}}P(A\mathbf{x} + g(\mathbf{x})) + (\mathbf{x}^{\mathsf{H}}A^{\mathsf{T}} + g(\mathbf{x})^{\mathsf{H}})P\mathbf{x} \\ &= \mathbf{x}^{\mathsf{H}}PA\mathbf{x} + \mathbf{x}^{\mathsf{H}}Pg(\mathbf{x}) + \mathbf{x}^{\mathsf{H}}\mathbf{x} + g(\mathbf{x})^{\mathsf{H}}P\mathbf{x} \\ &= \mathbf{x}^{\mathsf{H}}(PA + A^{\mathsf{H}}P)\mathbf{x} + 2\mathbf{x}^{\mathsf{H}}Pg(\mathbf{x}) \\ &= -\mathbf{x}^{\mathsf{H}}Q\mathbf{x} + 2\mathbf{x}^{\mathsf{H}}Pg(\mathbf{x}).\end{aligned}$$

Since

$$\frac{\|g(\mathbf{x})\|}{\|\mathbf{x}\|} \to 0 \qquad \text{as} \qquad \|\mathbf{x}\| \to 0,$$

then for all $\varepsilon > 0$ there exists $\delta > 0$ such that $\|g(\mathbf{x})\| \leq \varepsilon\|\mathbf{x}\|$ when $\|\mathbf{x}\| \leq \delta$. Thus,

$$\frac{d}{dt}V(\mathbf{x}) \leq -\mathbf{x}^{\mathsf{T}}Q\mathbf{x} + 2\|\mathbf{x}\|\|P\|\|g(\mathbf{x})\| \leq -(\lambda_{\min}(Q) - 2\varepsilon\|P\|)\|\mathbf{x}\|^2$$

for all $\|\mathbf{x}\| < \delta$. Thus, choosing $\varepsilon < \frac{\lambda_{\min}(Q)}{2\|P\|}$ gives stability. $\square$

Remark 5.5.6. Here $\lambda_{\min}(Q)$ denotes the value of the smallest eigenvalue of Q.

Remark 5.5.7. The theorem is another version of the stability theorem that we already proved.

5.5.3 *The Bartles–Stewart Algorithm

Instead of diagonalizing A, we can solve for P numerically by using Schur's theorem. Assume that $A = UTU^{\mathsf{H}}$, where T is upper triangular and U is orthonormal. Thus, we have that

$$U^{-\mathsf{H}}T^{\mathsf{H}}U^{\mathsf{H}}P + PUTU^{-1} = -Q.$$

Left and right multiplying by U^{H} and U, respectively, gives

$$T^{\mathsf{H}}\widetilde{P} + \widetilde{P}T = -\widetilde{Q},$$

where $\widetilde{P} = U^{\mathsf{H}}PU$ and $\widetilde{Q} = U^{\mathsf{H}}QU$.

By multiplying out components, we can solve for $\widetilde{P}$. This allows us to solve for $\widetilde{P}$ regardless of whether A is diagonalizable. Also, Schur has better stability properties since it's about finding an orthonormal basis instead of an eigenbasis, which can be ill-conditioned.

5.6 Poincaré–Bendixson Theorem and bounding regions via Lyapunov functions

The Lyapunov methods are not really that complicated, other than the fact that one must 'miraculously find' the appropriate Lyapunov function. Finding such a Lyapunov function is far more powerful than just proving that the origin for specific systems is globally attracting. Lyapunov's first theorem shows how we can prove that solutions stay within some bounded region in phase space. This is not very exciting if there is an attracting fixed point in this region, but what happens if we have a region in phase space that we know the solution must stay in, but there is no stable fixed point there? Where does the solution go in this case and what is the long term behavior? In dimension three and higher such a situation breeds chaos (sorry we don't define anything better than that) and very complicated, bounded trajectories may result. In 2D the situation is much calmer, and is described fully by the Poincaŕe Bendixson Theorem stated below.

5.6.1 Green's Theorem and nonexistence of periodic orbits

Consider the planar autonomous systems

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}), \ \mathbf{x} \in \mathbb{R}^2, \tag{5.13}$$

and suppose that $\mathbf{f} = (P(x,y), Q(x,y))$. Recall Green's Theorem.

Theorem 5.6.1 (Green's Theorem). *All of the following are considered in the plane in* $\mathbb{R}^2$*:*

(i) *for a simple closed curve* Γ *oriented counterclockwise,*

(ii) S *the interior of* Γ*, and*

(iii) C^1 *functions* P *and* Q *on an open set containing* $\Gamma \cup S$

then it follows that

$$\oint_\Gamma P dx + Q dy = \iint_S \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) \, dxdy.$$

We will need the divergence form of Green's Theorem which is obtained by reversing the roles of P and Q with a minus sign on on Q as follows:

$$\begin{aligned}
\oint_\Gamma (P dy - Q dx) &= \oint_\Gamma (-Q dx + P dy) \\
&= \iint_S \left(\frac{\partial}{\partial x}(P) - \frac{\partial}{\partial y}(-Q) \right) \, dxdy \\
&= \iint_S \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} \right) \, dxdy \\
&= \iint_S \nabla \cdot (\mathbf{f}) \, dxdy.
\end{aligned}$$

Theorem 5.6.2 (Bendixson's Criteria). *Suppose* $\mathbf{f} : U \to \mathbb{R}^2$ *is a* C^1 *vector field where* U *is an open subset of* $\mathbb{R}^2$*. For* $\mathbf{f} = (P, Q)$*, if the divergence* $\nabla \cdot \mathbf{f} = P_x + Q_y$ *is not identically zero and does not change sign on a simply connected subset* $\Omega \subset U$*, then* $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ *has no periodic orbit in* Ω*.*

Proof. Suppose there exists a periodic orbit $\mathbf{x}(t) = (x(t), y(t))$, $0 \leq t \leq T$, of $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ lying entirely in the simply connected region Ω; that is,

$$\Gamma = \{\mathbf{x}(t) : 0 \leq t \leq T\} \subset \Omega.$$

With S denoting the interior of Γ, and Γ oriented counterclockwise, we have by the divergence version of Green's Theorem, and the differential equations

$$\begin{aligned} \dot{x} &= P(x, y), \\ \dot{y} &= Q(x, y), \end{aligned}$$

that

$$\begin{aligned} \iint_S \nabla \cdot (\mathbf{f}) \; dx dy &= \oint_\Gamma (P dy - Q dx) \\ &= \int_0^T (P\dot{y} - Q\dot{x}) \; dt \\ &= \int_0^T (PQ - QP) \; dt \\ &= 0 \end{aligned}$$

By assumption $\nabla \cdot (\mathbf{f})$ is not identically zero and does not change sign in S. This means that $\nabla \cdot (\mathbf{f}) \geq 0$ or $\nabla \cdot (\mathbf{f}) \leq 0$ and there exists $\tilde{\mathbf{x}} \in S$ such that $\nabla \cdot (\mathbf{f})(\tilde{\mathbf{x}}) \neq 0$.

By the continuity of $\nabla \cdot (\mathbf{f})$ there exists a compact subset K of the open subset S, with $\tilde{\mathbf{x}} \in K$, and $\varepsilon > 0$ for which $|\nabla \cdot (\mathbf{f})(\mathbf{x})| \geq \varepsilon$ for all $\mathbf{x} \in K$. This implies that

$$\iint_S \nabla \cdot (\mathbf{f}) \; dx dy \neq 0.$$

This gives a contradiction, so there is no periodic orbit Γ of $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ in Ω. □

Example 5.6.3. The planar system

$$\begin{aligned} \dot{x} &= P(x, y) = y - x^5 - xy^4, \\ \dot{y} &= Q(x, y) = -x - x^4 y - y^5, \end{aligned}$$

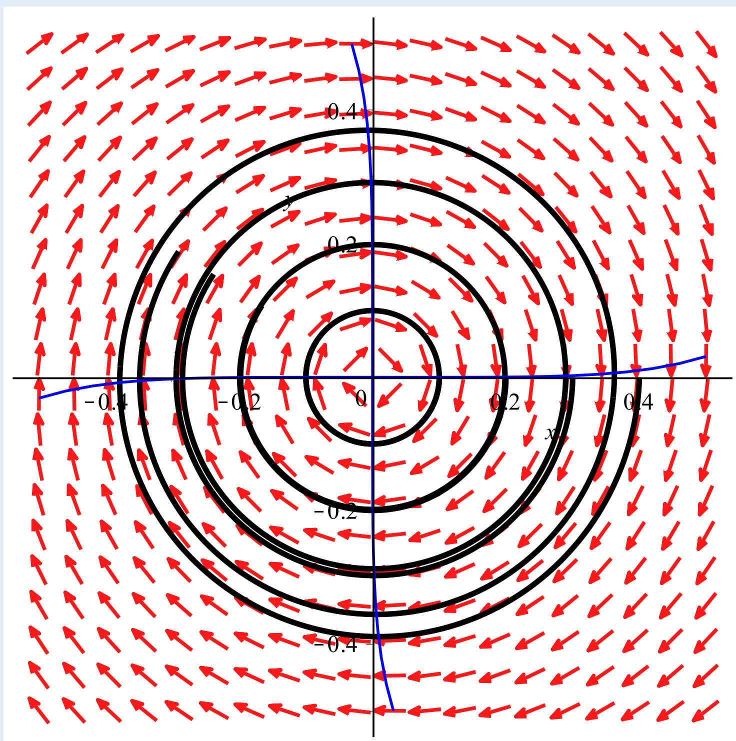
has an equilibrium at the origin.

Linearization of the vector field $\mathbf{f}(x, y) = (P, Q)^T$ at the origin gives the matrix

$$
\begin{aligned}
D\mathbf{f}(0,0) &= \begin{bmatrix} \dfrac{\partial P}{\partial x} & \dfrac{\partial P}{\partial y} \\ \dfrac{\partial Q}{\partial x} & \dfrac{\partial Q}{\partial y} \end{bmatrix}_{(x,y)=(0,0)} \\
&= \begin{bmatrix} -5x^4 - y^4 & 1 - 4xy^3 \\ -1 - 4x^3 y & -x^4 - 5y^4 \end{bmatrix}_{(x,y)=(0,0)} \\
&= \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.
\end{aligned}
$$

The equilibrium is a linear center.

Here is a graph of the vector field (vectors scaled so they don't overlap each other), the nullclines (from $\dot{x} = 0$ and from $\dot{y} = 0$), and some orbits.



You might conclude from this phase portrait that locally the equilibrium at the origin matches its linearization, that it is surrounded by periodic orbits. But as the divergence of the vector field,

$$\begin{aligned}\nabla\cdot(\mathbf{f}) &= P_x + Q_y \\ &= -5x^4 - y^4 - x^4 - 5y^4 \\ &= -6x^4 - 6y^4 \\ &= -6(x^4 + y^4),\end{aligned}$$

is not identically equal to zero and does not change sign on the simply connected (x, y)-plane, the Bendixson Criteria implies there are no periodic orbits of this planar system.

The Lyapunov function $V(x, y) = x^2 + y^2$ has orbital derivative

$$\begin{aligned}\frac{d}{dt}V(x, y) &= DV(x, y)\cdot\mathbf{f}(x, y) \\ &= (2x, 2y)\cdot(y - x^5 - xy^4, -x - x^4y - y^5) \\ &= 2xy - 2x^6 - 2x^2y^4 - 2xy - 2x^4y^2 - 2y^6 \\ &= -2x^6 - 2x^2y^4 - 2x^4y^2 - 2y^6 \\ &= -2(x^6 + x^2y^4 + x^4y^2 + y^6) \\ &< 0,\end{aligned}$$

for all $(x, y) \neq 0$ near the origin.

Thus by Lyapunov's Local Asymptotic Stability Theorem, the equilibrium at the origin is asymptotically stable; it just take solutions a long time to get to the origin.

Remark 5.6.4. As truly exciting as a non-existent result is, we are really keen on identifying conditions wherein a periodic orbit can occur.

5.6.2 Existence of periodic orbits in the plane

The Poincaré-Bendixson Theorem gives sufficient conditions for the existence of a special kind of periodic orbit known as a limit cycle. Before giving the formal definition of a limit cycle and the Poincaré-Bendixson Theorem (we will skip the proof—it takes about three days of class time if we were to present it), we present an example to motivate the notion of a limit cycle and the Poincaré-Bendixson Theorem.

Example 5.6.5. The planar system

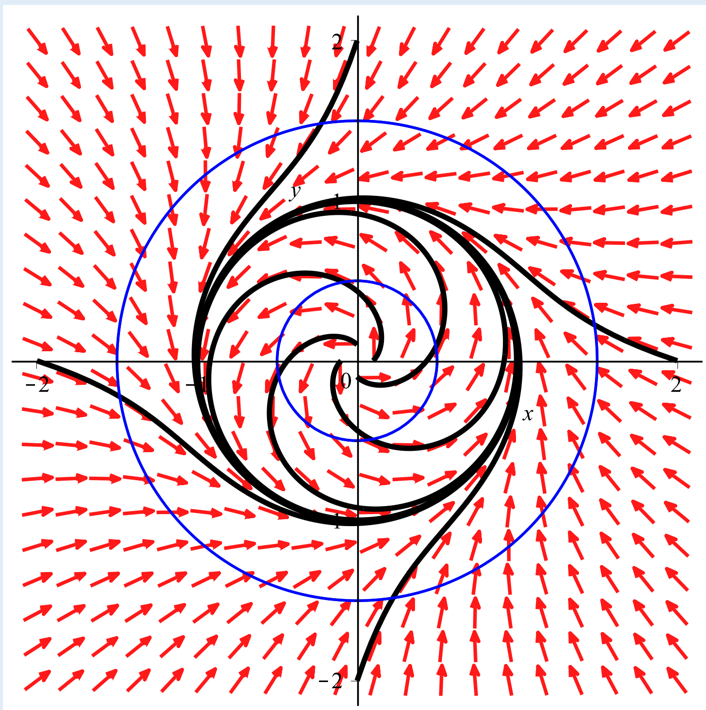
$$\begin{aligned}\dot{x} &= x - y - x^3 - xy^2, \\ \dot{y} &= x + y - x^2y - y^3,\end{aligned}$$

has an equilibrium at the origin whose linearization is given by,

$$D\mathbf{f}(0,0) = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix},$$

resulting in eigenvalues $1 \pm i$, meaning that the origin is an unstable spiral point.

Here is a graph of the vector field, eight solutions, and two circles (one with radius 0.5 and the other with radius 1.5) that bound an annular region.



The eight solutions that start outside the annular region cross into the compact annular region

$$A = \{(x,y) \in \mathbb{R}^2 : (0.5)^2 \leq \sqrt{x^2+y^2} \leq (1.5)^2\},$$

and stay in this annular region, i.e, A "captures" solutions.

(i) In addition we note that any solution starting in A will stay in A.

(ii) Notice that A does not contain any equilibrium solutions of the planar system, and solutions in A appear to be limiting onto the unit circle.

(iii) There appears to be a periodic orbit on the unit circle, which is aptly called a limit cycle because nearby solutions are "converging" to it.

(iv) Notice there are no other periodic orbits near the periodic orbit on the unit circle, which is to say the periodic orbit is isolated.

We can justify the capturing aspect of A and the limiting behavior of solutions in A with the Lyapunov function $V(x, y) = x^2 + y^2$ whose orbital derivative is

$$\begin{aligned}
\frac{d}{dt}V(x, y) &= DV(x, y)\mathbf{f}(x, y) \\
&= (2x, 2y) \cdot (x - y - x^3 - xy^2, x + y - x^2y - y^3) \\
&= 2x^2 - 2xy - 2x^4 - 2x^2y^2 + 2xy + 2y^2 - 2x^2y^2 - 2y^4 \\
&= 2x^2 + 2y^2 - 2x^4 - 4x^2y^2 - 2y^4 \\
&= 2\big[x^2 + y^2 - x^4 - 2x^2y^2 - y^4\big] \\
&= 2\big[x^2 + y^2 - (x^2 + y^2)^2\big] \\
&= 2(x^2 + y^2)\big[1 - (x^2 + y^2)\big].
\end{aligned}$$

(i) The value of $(d/dt)V(x, y)$ on the boundary $x^2 + y^2 = (0.5)^2$ is positive, indicating that V is increasing on a neighborhood of this boundary, so solutions cross this boundary of A with an increasing value of V.

(ii) The value of $(d/dt)V(x, y)$ on the boundary $x^2 + y^2 = (1.5)^2$ is negative, indicating that V is decreasing on a neighborhood of this boundary, so solutions cross this boundary of A with a decreasing value of V.

(iii) Also the Lyapunov function implies solutions that start in the annulus A stay in A.

(iv) Note as well that $(d/dt)V(x, y) = 0$ when $x^2 + y^2 = 1$, which is the unit circle to which the solutions in A are limiting onto, i.e. the limit cycle we are interested in identifying.

Remark 5.6.6. This is where the fun begins! Lyapunov methods, linearization and various other approaches to analyzing dynamical systems in the plane pale in comparison to drawing good pictures. Graphical representations of a dynamical system are often the more insightful than any hard analysis that can be performed. That being said, the analysis certainly plays a vital role and is essential to understanding the system in question.

Now we are ready to formalize this entire process.

Definition 5.6.7. *Let Ω be open in $\mathbb{R}^n$ for $n \geq 2$ and let $\mathbf{f} : \Omega \to \mathbb{R}^n$ be a C^1 vector field. A periodic orbit of $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ is a closed trajectory.*

Remark 5.6.8. We have seen non-isolated periodic orbits in the example of the first-order Lotka–Volterra model. These periodic orbits were not isolated because any nearby solution was also periodic, but not identical.

Definition 5.6.9. *Let Ω be open in $\mathbb{R}^n$ for $n \geq 2$ and let $\mathbf{f} : \Omega \to \mathbb{R}^n$ be a C^1 vector field. A subset A of Ω is* forward invariant *for $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ if for every solution $\mathbf{x}(t)$ with $\mathbf{x}(0) \in A$ there holds $\mathbf{x}(t) \in A$ for all $t \geq 0$.*

Remark 5.6.10. We also refer to a forward invariant set as a *trapping region*. Lyapunov functions are often helpful for constructing trapping regions.

Theorem 5.6.11 (Poincaré-Bendixson Theorem). *Let Ω be an open subset of $\mathbb{R}^2$ and let $\mathbf{f} : \Omega \to \mathbb{R}^2$ be a C^2 vector field. If there exists a compact subset A of Ω that is forward invariant for $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ and A does not contain any equilibrium solutions, then A contains a periodic orbit.*

Remark 5.6.12. In the previous example we used a Lyapunov function to construct a compact forward invariant equilibrium-free set A. The Poincaré-Bendixson Theorem guarantees the existence of an isolated periodic orbit in A. We happen to know this isolated periodic orbit is the unit circle. We do not always get this lucky with Lyapunov functions, i.e. Lyapunov functions are a good place to start when constructing trapping regions, but it is usually not that simple.

Remark 5.6.13. The Poincaré-Bendixson Theorem does not give the stability of the limit cycle, only its existence. The example we used to motivate the existence of limit cycles has an asymptotically stable limit cycle, but this is not what always happens. There is theory on the stability of the limit cycles (it involves determining $\nabla\cdot(\mathbf{f})$).

Example 5.6.14. The Lotka–Volterra third order model,

$$\dot{x} = x - \frac{x^2}{4} - \frac{xy}{1+x},$$
$$\dot{y} = -\frac{3y}{2} + 3\frac{xy}{1+x},$$

accounts for the predator not always searching for the prey.

The exercises investigate the linear stability of the three equilibria $(0,0)$, $(4,0)$, and $(1,3/2)$ of this planar system. We would like to use a simple Lyapunov function to define a trapping region for this problem in the hopes of proving the existence of a limit cycle, but the simplest Lyapunov functions (quadratic in x and y) are not conducive to our desires. Instead we will need to be more devious in finding a trapping region.

(i) We take a piece of an orbit near the equilibrium $(1,3/2)$ that circles about the equilibrium just more than one complete cycle (recall that because this equilibrium is an unstable spiral that solutions will be rotating away from this point in the plane).

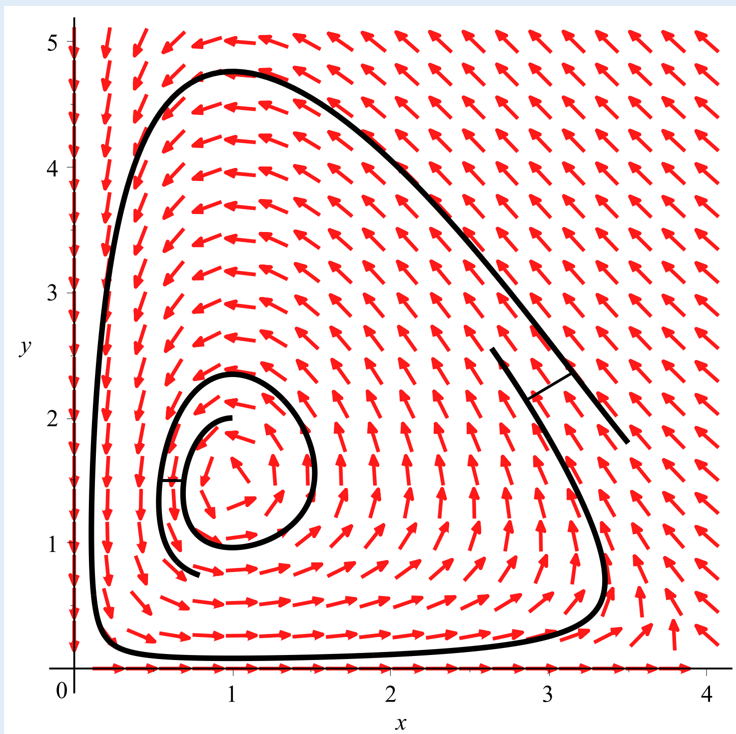
(ii) We then place a line segment across the gap where the orbit segment passes itself (as shown in the graph below).

(iii) We also take a piece of an orbit that starts near the equilibrium $(4, 0)$ that goes around the equilibrium $(1, 3/2)$ a bit more than once (just as we did for the previous step).

(iv) We then place a line segment across the gap where the orbit segment passes itself (as shown in the graph below).

Both line segments are chosen so they are "transverse" to the flow, i.e., the orbits cross the line segments in a non-parallel manner.

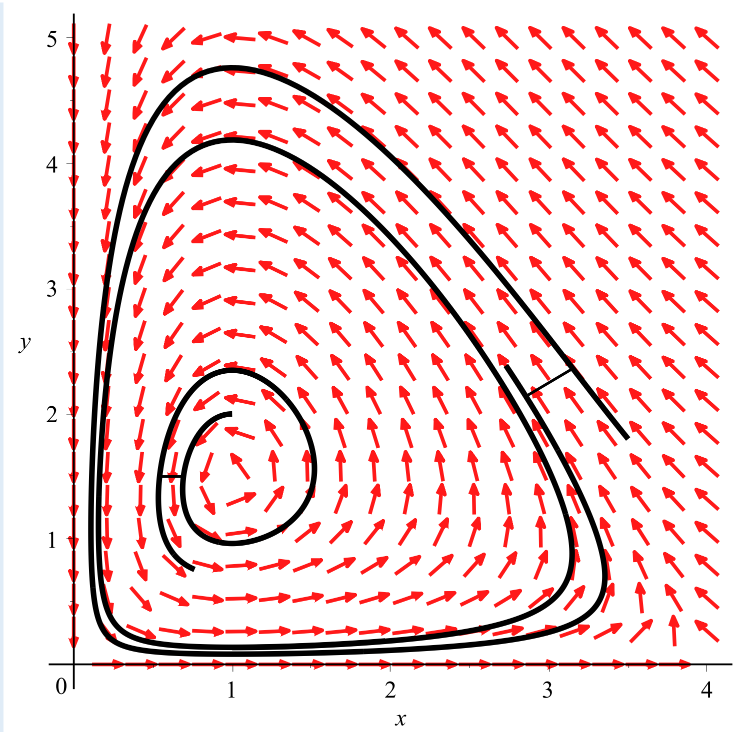
Here is the graph of the vector field (with the vectors scaled so they don't cross each other), the two orbit segments, and the two line segments that are transverse to the flow.



The orbits segments and the transverse line segments are the boundaries of a compact set A that is forward invariant (where we ignore the extra pieces of orbit segments in the graph).

There are no equilibrium solutions in A. Thus by the Poincaré-Bendixson Theorem there exists a limit cycle in A.

Here is a graph with the limit cycle contained in the compact set A.



It would appear from the phase portrait that the limit cycle is asymptotically stable, meaning that all other nearby orbits approach the limit cycle.

Time to celebrate, as we have finally found a predator-prey model for which there is a limit cycle that should imitate the true behavior of predator-prey interactions!

Remark 5.6.15. ...Except there are still clearly questions about this model. The previous example used very specific values of the parameters of the Lotka–Volterra model. Are these parameters physically reasonable? If they are, for what populations are they reasonable for? Basically, we aren't really done, we just answered one question of many.

Example 5.6.16. A model for glycolysis, the process by which living cells break down sugar to get energy, is the planar system

$$\dot{x} = -x + ay + x^2y,$$
$$\dot{y} = b - ay - x^2y,$$

where x and y are concentrations of reactants and a and b are positive parameters.

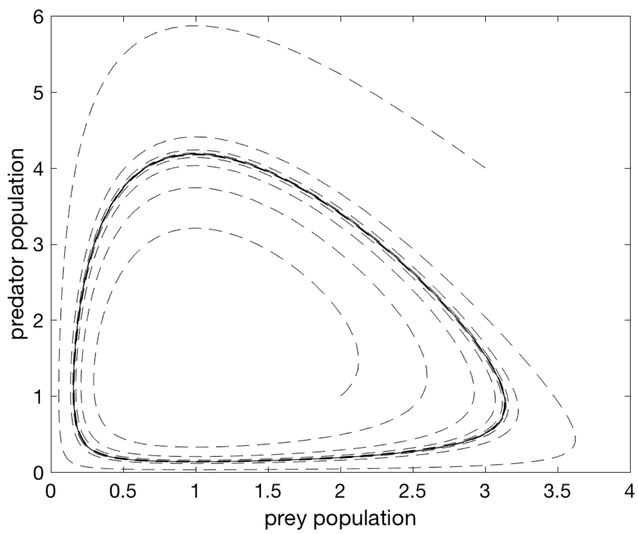


Figure 5.14: The appearance of the limit cycle for the third order Lotka–Volterra system described in Example 5.6.14.

The nullcline for $\dot{x}$ is

$$0 = -x + ay + x^2 y \;\Rightarrow\; x = y(a + x^2) \;\Rightarrow\; y = \frac{x}{a + x^2},$$

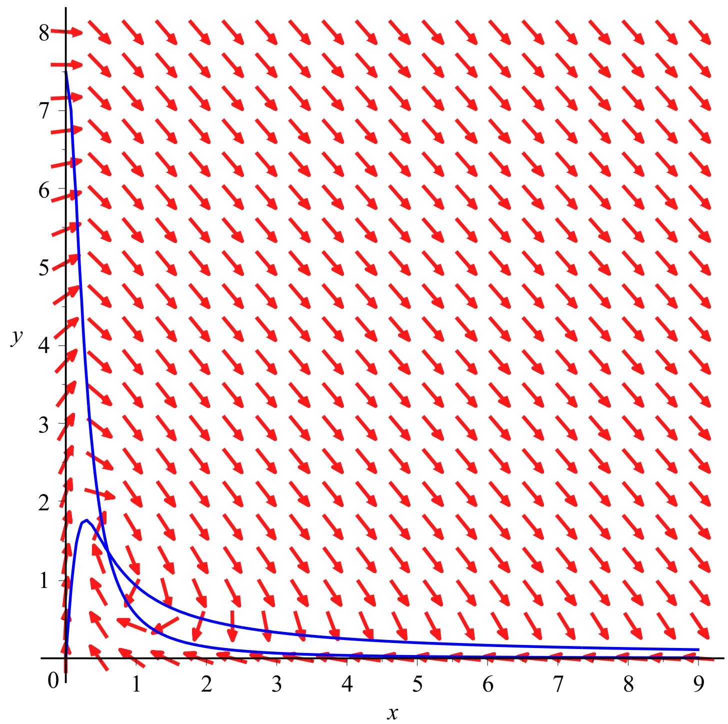
and the nulllcline for $\dot{y}$ is

$$0 = b - ay - x^2 y \;\Rightarrow\; b = y(a + x^2) \;\Rightarrow\; y = \frac{b}{a + x^2}.$$

You will find in the homework, the single equilibrium solution for this planar system, determine its linear stability, and verify that it is unstable for the parameter values

$$a = 0.08 \quad \text{and} \quad b = 0.6.$$

Here is a graph of the vector field for these parameters values (vectors scaled so they don't overlap) and the nullclines in blue.



To construct a trapping region:

(i) First show that the first quadrant is invariant: a solution that starts in the first quadrant stays in the first quadrant. This can be established by showing that solutions that start on either the positive x-axis or the positive y-axis move into the first quadrant.

(a) The level sets of the function $V(x, y) = y$ are the lines parallel to the x-axis, with the 0-level set of V being the x-axis. The orbital derivative of V is

$$\frac{d}{dt}V(x, y) = (0, 1) \cdot (-x + ay + x^2y, b - ay - x^2y) = b - ay - x^2y.$$

On the level set $V(x, y) = 0$, i.e., $y = 0$, the orbital derivative is

$$\frac{d}{dt}V(x, 0) = b > 0 \quad \text{for} \quad x \geq 0.$$

Since the level sets of $V(x, y) = c$ move upward as c increases, solutions that start on $V(x, y) = 0$ with $x > 0$ move up into the first quadrant.

(b) The level sets of the function $V(x, y) = x$ are lines parallel to the y-axis, with the 0-level set of V being the y-axis. The orbital derivative of V is

$$\frac{d}{dt}V(x, y) = (1, 0) \cdot (-x + ay + x^2 y, b - ay - x^2 y) = -x + ay + x^2 y.$$

On the level set $V(x, y) = 0$, i.e., $x = 0$, the orbital derivative is

$$\frac{d}{dt}V(0, y) = ay > 0 \quad \text{for} \quad y > 0.$$

Since the level sets of $V(x, y) = c$ move to the right as c increase, solutions that start on the positive y-axis move into the first quadrant.

(c) This leaves the solution that starts at the origin, but by continuity with respect to initial conditions, this solution also moves into the first quadrant.

Thus the first quadrant is invariant.

(ii) Of course the first quadrant is not compact, so in order to apply the Poincaré-Bendixson Theorem, we must find another "side" to obtain a compact region that bounds the first quadrant, as well as excise an open neighbourhood of the unstable equilibrium.

(a) Visualizing the vector field suggests we try a function of the form $V(x, y) = x + y$ and look for a level of $V(x, y)$ larger than 7.5 (where the top of thenullcline ends). The orbital derivative of $V(x, y) = x + y$ is

$$\begin{aligned}\frac{d}{dt}V(x, y) &= (1, 1) \cdot (-x + ay + x^2 y, b - ay - x^2 y) \\ &= -x + ay + x^2 y + b - ay - x^2 y \\ &= -x + b.\end{aligned}$$

This switches sign at $x = b$ and is negative for $x > b$.

The level sets of $V(x, y) = x + y$ are parallel lines with slope -1 and whose level values decrease from right to left. Solutions starting on the level $x + y = V(x, y) = 8.6$ on the interval $b < x \leq 8.6$ move to the left and down because the orbital derivative of $V(x, y)$ is negative.

(b) The issue is that this doesn't quite bound the first quadrant because of the gap near the vertical axis. To deal with this gap between $0 \leq x \leq b$ when $y = 8$, we return to the function $V(x, y) = y$ and its orbital derivative

$$\frac{d}{dt}V(x, y) = b - ay - x^2 y,$$

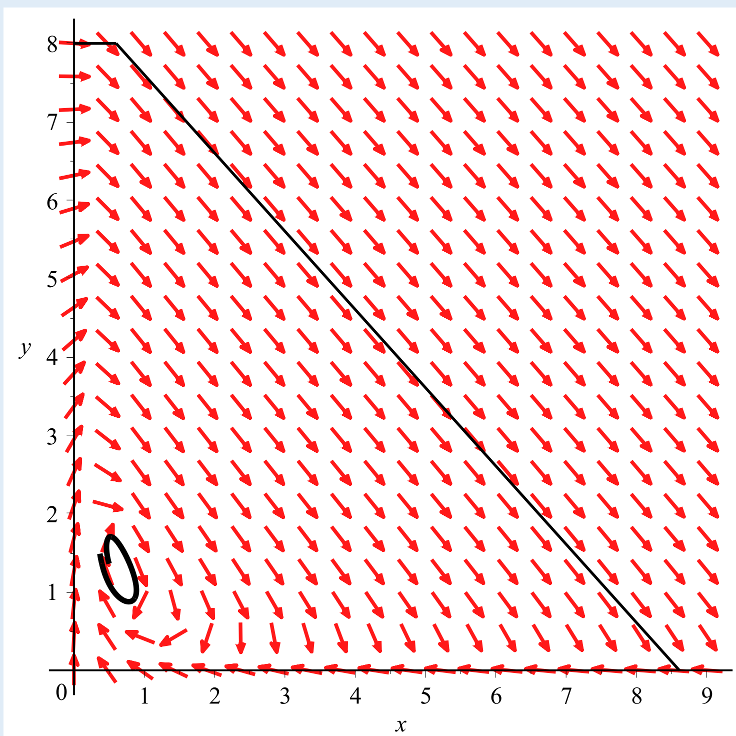
and notice that when $y = 8$ and $a = 0.06$ and $b = 0.6$ we have

$$\begin{aligned}\frac{d}{dt}V(x,8) &= b - 8a - 8x^2 \\ &= 0.6 - 8(0.08) - 8x^2 \\ &= 0.6 - 0.64 - 8x^2 \\ &= -0.04 - 8x^2 < 0.\end{aligned}$$

Thus the remaining "side" is the line segment $y = 8$ on the interval $0 \le x \le 0.6$ combined with the line segment $x + y = 8.6$ on the interval $0.6 < x \le 8.6$.

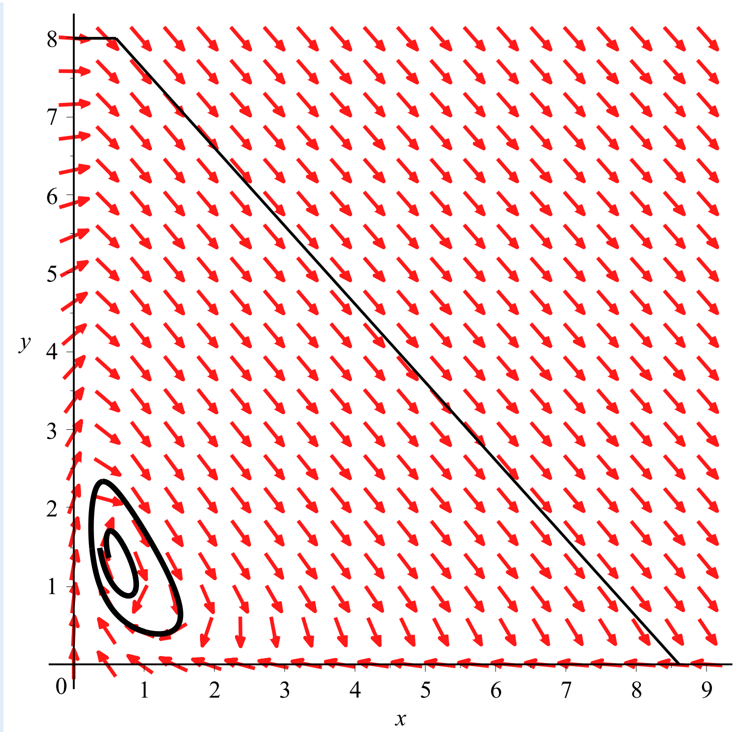
(c) Using the instability of the equilibrium we form an orbit segment that loops once around the equilibrium and place a transverse line segment across the gap, as we did in the previous example for the Lotka–Volterra system.

(iii) This gives a compact set A without equilibria for the planar system; here is a graph of this compact set A.



By the Poincaré-Bendixson Theorem there exists a limit cycle in A.

Here is a graph of the limit cycle with the compact forward invariant set A.



Remark 5.6.17. Bendixson's Criteria and the Poincaré-Bendixson Theorem are strictly two-dimensional phenomena. They do not have higher dimensional analogues. Instead at higher dimensions we actually obtain much more interesting phenomena such as chaos.

5.7 Convergence and stability for Numerical Methods

In case you haven't noticed, the term *stability* is thrown around a lot in this field of mathematics, and it can mean something different in many unique situations. In fact, stability means a lot of different things just in the context of dynamical systems. There are many more forms of stability that appear in a thorough discussion of dynamic modeling, but as has been emphasized previously in this text, this is not a comprehensive review of all these topics.

We are then satisfied to only present a handful of the methods and potential approaches for using stability in its various forms. We take the same approach when discussing the stability of numerical methods, i.e. we are only introducing the concepts of stability and convergence for numerical methods. This is NOT a comprehensive treatment.

Up to this point, we have only been concerned with the truncation error of numerical methods and have pushed any other concerns that each method presents under the proverbial rug. This is of course a horrible idea, and one that if totally ignored would haunt us for years to come (or at least it should).

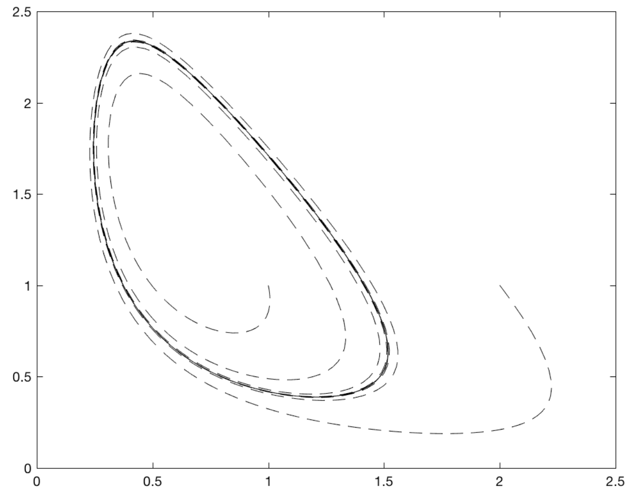


Figure 5.15: The limit cycle for Example 5.6.16 with $a = 0.08$ and $b = 0.6$.

There are many other concerns and issues to bring up for numerical methods then just considering the cost of each, and the truncation error associated with a particular approach. In fact this is such a nuanced subject that there are multiple courses dedicated to the analysis of these numerical schemes, and many a PhD dissertation has been generated from this topic. We certainly don't have the time to worry about all of this, so we will restrict ourselves to a cherry-picked subset of the topic, one that naturally generates some fascinating figures independent of the dynamics that are being investigated.

5.7.1 Convergence

We aren't going to go into detail about convergence here, but the basic idea is that a given numerical approximation to the solution of (**??**) is prescribed e.g. via Euler's method so that

$$\mathrm{U}_n \approx \mathbf{u}(n\Delta t), \tag{5.14}$$

is the approximation of the true solution $\mathbf{u}(t)$ at time $t = n\Delta t$. Convergence of the approximation scheme given by (5.14) simply guarantees that as $\Delta t \to 0$ then (5.14) will converge to the true solution $\mathbf{u}(T)$ for a fixed value of $T = k\Delta t$ (note that $k \to \infty$ as $\Delta t \to 0$ in order for T to remain fixed).

Convergence of a numerical scheme is not always guaranteed, and indeed may not always be intuitive. Essentially we are asking that the the truncation errors in the numerical approximation do not overwhelm the solution as we take more and more time-steps (recall that $k \to \infty$ in this case), and this is only guaranteed because the time step is getting smaller and smaller at the same time. Convergence works for most numerical schemes so long as the truncation error is truly controlled, i.e. taking more steps doesn't allow the truncation error to accumulate and overwhelm the approximation.

To see how this works, we will consider the accumulation of error for the forward Euler method applied to the linear, scalar IVP

$$\dot{u} = \lambda u, \quad u(0) = u_0. \tag{5.15}$$

Extension to linear systems is easily accomplished by analyzing the scalar ODE that arises in each eigenspace of the full system, and extension to nonlinear systems is carried out relying heavily on the fact that $\mathbf{f}(\mathbf{u})$ (the right hand side of the ODE) is Lipschitz continuous.

For $T > 0$, partition the interval $[0, T]$ into $N \in \mathbb{N}$ subintervals $[t_{n-1}, t_n]$ of equal length

$$\Delta t = \frac{T}{N}$$

and endpoints

$$t_n = n\Delta t, \;\; n = 0, 1, 2, \ldots, N.$$

The forward Euler method gives approximations $U_n = U(t_n)$ of the exact values $u(t_n)$ of the solution $u(t)$ of the IVP, where $U(0) = u(0)$.

The time step of the forward Euler method relates U_n to U_{n+1} by

$$U_{n+1} = U_n + (\Delta t)\lambda U_n = \big(1 + (\Delta t)\lambda\big)U_n.$$

In terms of the notation $u(t_n)$ the truncation error of a time step for the forward Euler method is

$$\tau_n = \tau(t_n) = \frac{u(t_{n+1}) - u(t_n)}{\Delta t} - \lambda u(t_n) = \frac{\Delta t}{2}\ddot{u}(t_n) + O((\Delta t)^2).$$

Rewritten this is

$$u(t_{n+1}) = \big(1 + (\Delta t)\lambda\big)u(t_n) + (\Delta t)\tau_n.$$

This implies that

$$\mathscr{L}(t_n) = (\Delta t)\tau_n$$

is the local error of a single time step.

The (accumulated) global error at time step n is the difference

$$E_n = U_n - u(t_n).$$

At the beginning, since $U_0 = u(0)$, we have $E_0 = 0$. For $n = 1$ we have

$$\begin{aligned}
E_1 &= U_1 - u(t_1) \\
&= \big(1 + (\Delta t)\lambda\big)u_0 - \big[(1 + (\Delta t)\lambda)u(t_0) + (\Delta t)\tau_0\big] \\
&= \big(1 + (\Delta t)\lambda\big)\big(U_0 - u(t_0)\big) - (\Delta t)\tau_0 \\
&= -(\Delta t)\tau_0.
\end{aligned}$$

For $n = 2$ we have

$$\begin{aligned}
E_2 &= U_2 - u(t_2)\\
&= \big(1+(\Delta t)\lambda\big)U_1 - \big[(1+(\Delta t)\lambda)u(t_1) + (\Delta t)\tau_1\big]\\
&= \big(1+(\Delta t)\lambda\big)\big(U_1 - u(t_1)\big) - (\Delta t)\tau_1\\
&= \big(1+(\Delta t)\lambda\big)E_1 - (\Delta t)\tau_1\\
&= \big(1+(\Delta t)\lambda\big)(-(\Delta t)\tau_0) - (\Delta t)\tau_1\\
&= -(\Delta t)\big[(1+(\Delta t)\lambda)\tau_0 + \tau_1\big]\\
&= -(\Delta t)\sum_{m=1}^{2}\big(1+(\Delta t)\lambda\big)^{2-m}\tau_{m-1}.
\end{aligned}$$

This summation formula for E_2 matches the error for

$$E_1 = -(\Delta t)\sum_{m=1}^{1}\big(1+(\Delta t)\lambda\big)^{1-m}\tau_{m-1} = -(\Delta t)\tau_0.$$

This suggests that the global error at step n is

$$E_n = -(\Delta t)\sum_{m=1}^{n}\big(1+(\Delta t)\lambda\big)^{n-m}\tau_{m-1}.$$

We verify this by induction.

For $n+1$ we have

$$\begin{aligned}
E_{n+1} &= U_{n+1} - u(t_{n+1})\\
&= \big(1+(\Delta t)\lambda\big)U_n - \big[(1+(\Delta t)\lambda)u(t_n) + (\Delta t)\tau_n\big]\\
&= \big(1+(\Delta t)\lambda\big)\big(U_n - u(t_n)\big) - (\Delta t)\tau_n\\
&= \big(1+(\Delta t)\lambda\big)E_n - (\Delta t)\tau_n\\
&= \big(1+(\Delta t)\lambda\big)\left(-(\Delta t)\sum_{m=1}^{n}(1+(\Delta t)\lambda)^{n-m}\tau_{m-1}\right) - (\Delta t)\tau_n\\
&= -(\Delta t)\sum_{m=1}^{n}(1+(\Delta t)\lambda)^{n+1-m}\tau_{m-1} - (\Delta t)\tau_n\\
&= -(\Delta t)\sum_{m=1}^{n+1}(1+(\Delta t)\lambda)^{n+1-m}\tau_{m-1}.
\end{aligned}$$

With this exact expression for the error we hope to show that as $\Delta t \to 0$ so does E_n.

In pursuit of this we recognize that the expression $1+(\Delta t)\lambda$ in the error consists of the first two terms of the exponential of $\exp((\Delta t)\lambda)$:

$$\exp((\Delta t)\lambda) = 1 + (\Delta t)\lambda + \frac{((\Delta t)\lambda)^2}{2} + \cdots .$$

Since the quantities here are positive for small enough Δt we obtain the inequality

$$|1+(\Delta t)\lambda| = 1 + (\Delta t)\lambda \le \exp((\Delta t)\lambda) \le \exp((\Delta t)|\lambda|).$$

Hence for positive integers n and m with $m \le n \le N$ we obtain the inequalities

$$|1 + (\Delta t)\lambda|^{n-m} \le \exp((n-m)(\Delta t)|\lambda|) \le \exp(n(\Delta t)|\lambda|) \le \exp(|\lambda|\, T).$$

For the truncation error we set

$$\|\tau\|_\infty = \max\{|\tau_m| : 1 \le m \le N-1\}$$

where

$$\max|\tau_m| \approx \frac{(\Delta t)}{2}\|\ddot{u}\|_\infty = O(\Delta t)$$

and

$$\|\ddot{u}\|_\infty = \max\{|\ddot{u}(t)| : 0 \le t \le T\}.$$

With these we get an upper bound on the global error E_N:

$$\begin{aligned}
|E_N| &= \left|(\Delta t)\sum_{m=1}^{N}\big(1 + (\Delta t)\lambda\big)^{N-m}\tau_{m-1}\right| \\
&\le (\Delta t)\sum_{m=1}^{N}|1 + (\Delta t)\lambda|^{N-m}\,|\tau_{m-1}| \\
&\le (\Delta t)\sum_{m=1}^{N}\exp(|\lambda|\,T)\|\tau\|_\infty \\
&= (\Delta t)N\exp(|\lambda|\,T)\|\tau\|_\infty \\
&= T\exp(|\lambda|\,T)\|\tau\|_\infty.
\end{aligned}$$

Since $\|\tau\|_\infty = O(\Delta t)$ we have that $E_N \to 0$ as $\Delta t \to 0$.

Nota Bene 5.7.1. Caution: this convergence for forward Euler is a limit as Δt goes to 0, not for every small Δt you wish to use, as we demonstrate below.

Remark 5.7.2. Convergence for numerical approximations of ODEs is wonderful, but also clearly doesn't tell us quite everything. For instance, just like we did when discussing continuous dependence on initial conditions and parameters, we have ignored the fact that the estimate produced above is exponential in the final time T, and the parameter λ.

If we select Δt sufficiently small then we shouldn't care about this exponential dependence of the error, but if we are waiting long enough, i.e. we are really interested in long time dynamics for very large values of T then we should be a little more concerned with this because not only are we allowing for a very large exponential term to affect the global error, but in order to overcome this large global error, we are forced to take even smaller time steps Δt over a longer time interval, which is not going to be computationally cheap.

Remark 5.7.3. Clearly the derivation for forward Euler given above is not unique to this method, or this particular ODE. It turns out that one can construct a very similar proof for nonlinear ODEs with a Lipschitz constant λ. It is also not that painful to work through a proof for convergence for most other numerical methods such as midpoint, trapezoid, etc. The multistep and/or multistage methods are more difficult, but once again, careful tracking of the local truncation error allows these algorithms to be properly analyzed as well.

As we see below, convergence is nice, but actually not the most useful concept when it comes to analyzing numerical methods. We actually need something much stronger and more descriptive than convergence in the sense described here. This is is best illustrated by the following example.

Example 5.7.4. Consider solving the ODE

$$\dot{u} = \lambda(u - \cos t) - \sin t, \qquad u(0) = 1, \tag{5.16}$$

which has the explicit solution given by $u(t) = \cos t$ for any value of λ. Suppose that we are interested in the error that accumulates at time $T = 2$, that is we want to compare how well our numerical solution approximates cos(2) starting with initial condition $u(0) = 0$. Now, because the exact solution for any value of λ is the same, then forward Euler is guaranteed to converge no matter what λ is. If we pick $\Delta t = 0.01$ as our time step, then the error we get at time $T = 2$ for various values of λ is given in the Table below:

λ	error at $T = 2$
-10	1.6×10^{-4}
-50	4.0×10^{-5}
-200	1.0
-250	1.6×10^{35}

Clearly something goes terribly wrong as λ grows in magnitude, even though the exact solution remains the same. To illustrate this even more, we select the problematic case $\lambda = -250$ and adjust the time step to see if we can get the correct solution. This is seen in the following Table:

Δt (for $\lambda = -250$)	error at $T = 2$
0.01	1.6×10^{35}
0.009	3.2×10^{21}
0.0085	1.0×10^{12}
0.008	1.0
0.0075	2.8×10^{-6}

Thus, if we pick a time step small enough then it looks like we do get an answer that makes sense. Something magical seems to happen right around $\Delta t \approx 0.008$.

This scalar example is a classic one that illustrates the dangers of blindly using numerical methods. While it is used for the forward Euler method, similar issues arise for almost all explicit numerical schemes.

To demonstrate how this happens, we point toward Example 4.1.11 where we noticed that some numerical schemes worked better than others. In particular, we noticed that the implicit backward Euler method demonstrated fairly robust behavior even for decently large values of Δt whereas the explicit forward Euler method did not. Both these examples demonstrate that something more sinister is going on.

5.7.2 Absolute Stability

There are a lot of different kinds of stability even for the study of numerical methods. We will focus on only one type here, but it is important to recognize that this topic is much more nuanced than we are diving into here. Assuming that this discussion is comprehensive is just like the captain of the Titanic assuming that he could see all of the iceberg above the surface of the water (actually I doubt the captain was on deck when the Titanic hit the iceberg...). Nevertheless, it is of course essential to notice that the iceberg is present, and either decide to avoid it at all costs, or to return and explore it with the proper tools and gear (taking an additional course in numerical analysis).

We will only discuss the absolute stability of numerical schemes for scalar differential equations. Extension to higher dimensional systems is relatively straightforward, as discussed further below. We first note that so long as our time steps are sufficiently small then it makes sense to consider analyzing a linearization of the full nonlinear solution, i.e. since $\mathbf{f}(\mathbf{x})$ is Lipschitz continuous, and supposing that our solutions are infinitely continuously differentiable, then we are justified in supposing that small increases in time (denoted by Δt) will result in small changes in $\mathbf{f}(\mathbf{x})$. This is part of the motivation for considering stability of the numerical schemes on linear systems only.

It turns out that the restrictions we obtain for the linear system (5.15) apply to the full nonlinear system when $f'(\tilde{x}) = \lambda$, where $\tilde{x} = x(t)$ is the current state of our system. Hence, we will focus on (5.15). Extension to systems is accomplished by noting that the same restrictions placed below must apply to every eigenvalue of the Jacobian matrix $D\mathbf{f}(\mathbf{x})$.

We begin by analyzing the forward Euler method to demonstrate what absolute stability means, and then we make the formal definitions, and demonstrate how absolute stability is calculated for a variety of methods.

Example 5.7.5. In the induction proof of the exact formula for the error E_n we obtained the recursion formula

$$E_{n+1} = \big(1 + (\Delta t)\lambda\big)E_n + (\Delta t)\tau_n, \text{ for } n \in \{0, 1, 2, \dots, N\}.$$

Since in the forward Euler method the truncation error is $O(\Delta t)$, the term $(\Delta t)\tau_n$ is $O((\Delta t)^2)$.

From the recursion formula for the error and $E_0 = 0$ we obtain

$$E_1 = \big(1 + (\Delta t)\lambda\big)E_0 + (\Delta t)\tau_0 = (\Delta t)\tau_0 = O((\Delta t)^2).$$

From the recursion formula for the error and the value of E_1 we obtain

$$E_2 = \big(1 + (\Delta t)\lambda\big)E_1 + (\Delta t)\tau_1 = \big(1 + (\Delta t)\lambda\big)E_1 + O((\Delta t)^2).$$

From the recursion formula for the error and the value of E_2 we obtain

$$\begin{aligned}
E_3 &= \big(1 + (\Delta t)\lambda\big)E_2 + (\Delta t)\tau_2 \\
&= \big(1 + (\Delta t)\lambda\big)\big[(1 + (\Delta t)\lambda)E_1 + (\Delta t)\tau_1\big] + (\Delta t)\tau_2 \\
&= \big(1 + (\Delta t)\lambda\big)^2 E_1 + (1 + (\Delta t)\lambda)(\Delta t)\tau_1 + (\Delta t)\tau_2 \\
&= \big(1 + (\Delta t)\lambda\big)^2 E_1 + (\Delta t)\tau_1 + \lambda(\Delta t)^2\tau_1 + (\Delta t)\tau_2 \\
&= \big(1 + (\Delta t)\lambda\big)^2 E_1 + O((\Delta t)^2).
\end{aligned}$$

Continuing this we obtain

$$E_N = \big(1 + (\Delta t)\lambda\big)^{N-1} E_1 + O((\Delta t)^2).$$

The error will be amplified or grow exponentially fast in N if

$$|1 + (\Delta t)\lambda| > 1.$$

To avoid this unfavorable situation we choose $z = (\Delta t)\lambda$ to satisfy the complex equation

$$|1 + z| \leq 1.$$

This gives the closed disk in the complex plane with center at -1 and radius 1 as shown in Figure 5.16.

This closed disk is called the region of absolute stability for the forward Euler method for the scalar IVP $u' = \lambda u$, $u(0) = u_0$, $\lambda \neq 0$. You can see why when $\lambda > 0$ the forward Euler method has issues with accurately approximating the solution $u(t) = u(0)\exp(\lambda t)$ which grows, because there no choice of Δt for which $|1 + (\Delta t)\lambda| \leq 1$.

On the other hand, when $\lambda < 0$ the forward Euler method will accurately approximate the solution $u(t) = u(0)\exp(\lambda t)$ because there are choices of Δt for which $|1 + (\Delta t)\lambda| \leq 1$.

For the previous example given above where we were trying to solve $\dot{u} = \lambda(u - \cos t) - \sin t$ we note that the non-autonomous terms will actually balance out when we look at the evolution of the error, and hence we are really just looking at ensuring that $|1 + \lambda\Delta t| \leq 1$ to ensure stability. For the problematic case of $\lambda = -250$ this would indicate that $\Delta t \leq \frac{-2}{\lambda} = 0.008$, which is remarkably consistent with the results we noticed from our numerical calculations above.

Now that we have seen the example above for forward Euler, we are ready to define what we mean by absolute stability for a numerical method.

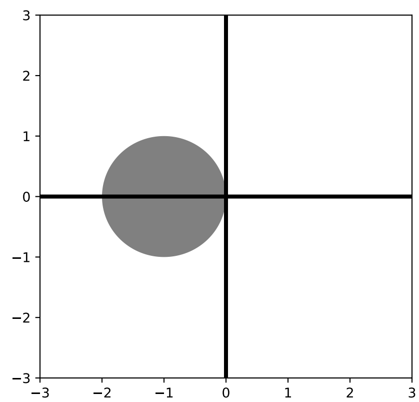


Figure 5.16: The absolute stability region in the complex plane for the forward Euler method.

Definition 5.7.6. *For a numerical scheme that approximates the IVP* (5.15)*, we define the* region of absolute stability *in the complex plane* $\mathbb{C}$ *as the values of* $z = \lambda\Delta t$ *for which the errors for this particular scheme are amplified by a factor with magnitude less than or equal to one.*

Remark 5.7.7. That definition is a mouthful, and strangely for a mathematics textbook, doesn't have an equation in it. Unfortunately the concept of absolute stability is a little difficult to formalize without introducing several pages worth of notation that would be motivated purely by a desire to get the definition correct. While many individuals enjoy working with complicated notation, one of the authors of this text does not, so we have opted for the definition given above instead. Sincerest apologies are due to those who prefer lengthy and messy forms of notation.

Sincere apologies are also due to those who just like everything to be neat and tidy. Eventually these remarks will be cleaned up, sarcastic comments will be removed, and the previous definition will be properly formalized. Until such a date, continue to suffer quietly the whims of a reluctant writer.

The previous example demonstrates how to find the absolute stability region for the forward Euler method. Rather than only considering real values of $z = \lambda\Delta t$ however, we want to also consider complex values, because when we consider systems of differential equations we will be considering λ as an eigenvalue of the corresponding Jacobian matrix, which may very easily be complex valued even for a real valued function $\mathbf{f}(\mathbf{x})$.

Thus, we see that the absolute stability region for forward Euler is the interior of the unit circle centered at $z = -1$ in the complex plane. This means that forward Euler will remain stable so long as we select Δt so that $\lambda\Delta t$ remains in this region.

The stability region depicted in Figure 5.16 indicates the region where $z = \lambda\Delta t$ is stable for forward Euler. As noted above, even for nonlinear ODEs this is remarkably accurate in that selecting Δt to ensure that $z = f'(x)\Delta t$ remains in the shaded region of Figure 5.16, will ensure that the method remains stable. This provides a good way to clarify stability of the scheme and provides a method for determining the most appropriate choice of Δt.

Computing the absolute stability region for a numerical method is not an easy or straightforward task. As there isn't necessarily a simple approach to this, we demonstrate how it works on a few additional examples.

Example 5.7.8. To look at yet one more example of calculating the absolute stability region, we consider the midpoint or Leapfrog method which provides the update formula

$$u(t + \Delta t) = u(t - \Delta t) + 2\Delta t f(u(t)).$$

Applied to (5.15) this becomes

$$u(t + \Delta t) = u(t - \Delta t) + 2\lambda\Delta t u(t).$$

If we let $E(t + \Delta t) = E_1$ be the error at time $t + \Delta t$ and $z = \lambda\Delta t$ then this leads to the recursive equation

$$E_1 = E_{-1} + 2zE_0,$$

or after n steps

$$E_{n+1} = E_{n-1} + 2zE_n.$$

Determining the absolute stability region for the midpoint method then reduces to solving this recursive equation, and identifying for what values of $z \in \mathbb{C}$ the solution $E_n(z)$ does not grow with n. To solve this relationship, we consider the ansatz $E_n = \xi^n$ so that

$$\xi^2 = 1 + 2z\xi.$$

This yields solutions:

$$\xi = z \pm \sqrt{z^2 + 1}.$$

The actual error would be some linear combination of these two solutions. Guaranteeing that the error doesn't grow with n would require that $|\xi| \leq 1$ for both such solutions which becomes:

$$|z \pm \sqrt{z^2 + 1}| \leq 1.$$

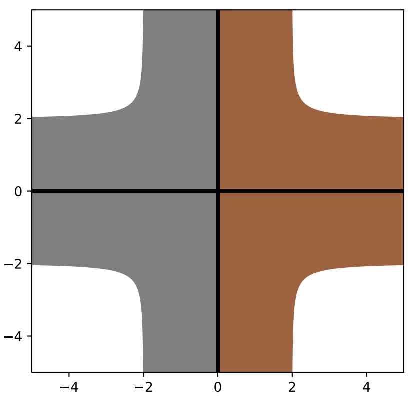


Figure 5.17: The absolute stability region in the complex plane for the midpoint method. The gray region represents the area wherein $z + \sqrt{z^2+1}$ is less than 1, and the copper colored region details where $z - \sqrt{z^2+1}$ is less than 1. The absolute stability region is at the intersection of both regions, i.e. the imaginary axis.

> The region where $|z + \sqrt{z^2+1}| \le 1$ is shown in Figure 5.17 in the shaded gray region. The region where $|z - \sqrt{z^2+1}| \le 1$ is shown in copper in the same Figure. In order for the error to not grow, z must be in the intersection of these two regions. The intersection of these two regions is the imaginary axis, i.e. the midpoint method is only absolutely stable for z along the imaginary axis.

We can perform the same type of analysis for other multistep methods. The absolute stability regions are quite fascinating to visualize, and can create some of the most interesting figures that you can imagine. For instance, the absolute stability region for 2-step Adams–Bashforth is shown in Figure 5.18.

Remark 5.7.9. We have only discussed the absolute stability region for two-step methods. Higher order multistep methods can of course be analyzed in a similar way, but clearly things get more complicated rather quickly as the quadratic formula doesn't apply. Numerical approximations are then necessary to estimate the absolute stability region of the numerical method in question.

Remark 5.7.10. Derivation of the absolute stability region for Runge–Kutta methods is actually substantially simpler than for multistep methods. Simply substituting $f(u) = \lambda u$ into the algorithm, writing out all of the terms on a single line, and then guaranteeing that the error doesn't grow with each iteration, leads to the absolute stability region in $\mathbb{C}$. Luckily you get to do this in the exercises, but just to give you something to work toward, Figure 5.19 shows the absolute stability region for explicit 2-stage Runge–Kutta.

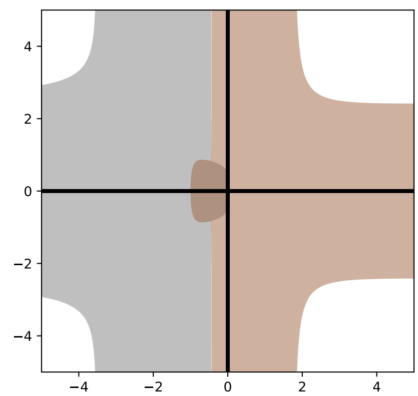


Figure 5.18: The absolute stability region in the complex plane for the 2-step Adams–Bashforth method. Just as for the midpoint method, the two colors represent regions where the two possible roots of the recursion equation are guaranteed to be less than one in magnitude. The absolute stability region is the intersection of these two, which is a narrow region in the left half plane.TODO: either this figure or the following figure are incorrect...can you tell which one it is?

This leads to an interesting observation that the extra degree of freedom in the choice of coefficients for Runge–Kutta methods gives an option for optimizing the method for stability purposes. In other words, rather than choosing the coefficients of the Runge–Kutta method to be 'nice' and easy to remember, we may choose them to extend the absolute stability region as far into the left half plane as possible. This leads to a class of time-stepping methods often referred to as strong stability preserving (SSP) Runge–Kutta methods.

Remark 5.7.11. In the interest of having as many remarks as possible (and besides, they at least make it easier to read through this as opposed to having just paragraphs like a history textbook), we now comment on the extension of absolute stability to systems. Basically everything stays the same except that λ now represents the eigenvalues of the Jacobian of the system at any given point in time, and we must choose Δt so that $z = \lambda \Delta t$ is in the absolute stability region for all eigenvalues λ of the system.

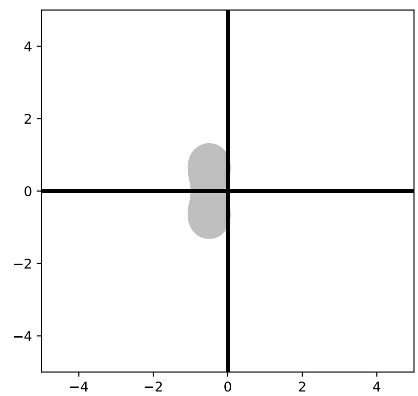


Figure 5.19: The absolute stability region in the complex plane for the 2-stage, second-order, explicit Runge–Kutta method.

5.8 *Stable Manifold Theorem

Now that we have hopelessly beat the predator-prey model to death, and hopefully have a good feeling for the long time behavior of solutions for a stable fixed point (node or equilibrium), we turn our attention to unstable fixed points. In particular, we are interested in the behavior of fixed points such as the one shown in Example 5.2.13, i.e. a saddle. We are interested in the behavior and evolution of solutions near a saddle node, i.e. how do solutions approach such an equilibrium and how do they diverge away from it? This section will illustrate the geometric construction of the answer to this question.

First, we need to establish some notation and definitions. Returning to the formula for a matrix exponential to solve (5.1.17),

$$\mathbf{x}(t) = \sum_{\lambda\in\sigma(A)} e^{\lambda t}\left(P_\lambda + \sum_{k=1}^{m_\lambda-1} \frac{D_\lambda^t}{k!}\right)\mathbf{x}_0.$$

If A is semi-simple this reduces to

$$\mathbf{x}(t) = \sum_{\lambda\in\sigma(A)} e^{\lambda t} P_\lambda \mathbf{x}_0,$$

so if $\mathbf{x}_0$ is in the eigenspace corresponding to λ, then $P_\lambda\mathbf{x}_0 = \mathbf{x}_0$ and $P_{\lambda'}\mathbf{x}_0 = 0$ for all $\lambda' \neq \lambda$. If this is the case, there are three possibilities

(i) If $\Re(\lambda) < 0$, then $\lim_{t\to\infty} |\mathbf{x}(t)| = 0$ (stable)

(ii) If $\Re(\lambda) > 0$, then $\lim_{t\to\infty} |\mathbf{x}(t)| = 0$ (unstable)

(iii) If $\Re(\lambda) = 0$, then $|\mathbf{x}(t)| = |\mathbf{x}(0)|$.

We employ the following definition to distinguish between the different parts of the solution.

Definition 5.8.1. *For an $n \times n$ matrix A, the stable subspace, E_s is the subspace of $\mathbb{R}^n$ that is spanned by the set of eigenvectors of A whose corresponding eigenvalues have negative real part. The unstable subspace E_u is defined similarly for eigenvalues with positive real part, and the center subspace for eigenvalues with zero real part.*

Note that this works for all cases except $\Re(\lambda) = 0$, even when A is not semi-simple ($t^n e^{\lambda t} \to 0$ for $\lambda = 0$). This allows us to break up e^{tA} in three parts,

$$e^{tA} = \sum_{\substack{\lambda\in\sigma(A)\\ \Re(\lambda)<0}} \underbrace{\left(P_\lambda + \sum_{k=1}^{m\lambda-1} \frac{t^k D_\lambda^k}{k!}\right)}_{P_s} + \sum_{\substack{\lambda\in\sigma(A)\\ \Re(\lambda)>0}} e^{\lambda t} \underbrace{\left(P_\lambda + \sum_{k=1}^{m\lambda-1} \frac{t^k D_\lambda^k}{k!}\right)}_{P_u} + \sum_{\substack{\lambda\in\sigma(A)\\ \Re(\lambda)=0}} e^{\lambda t} \underbrace{\left(P_\lambda + \sum_{k=1}^{m\lambda-1} \frac{t^k D_\lambda^k}{k!}\right)}_{P_0}$$

$\uparrow$ stable subspace $\qquad \uparrow$ unstable subspace $\qquad \uparrow$ center subspace

(i) If $P_0\mathbf{x}_0 = P_c\mathbf{x}_0 = 0$, then $\mathbf{x}_0$ is in the stable subspace and $\lim_{t\to\infty} |\mathbf{x}(t)| = 0$.

(ii) If $P_s\mathbf{x}_0 = P_c\mathbf{x}_0 = 0$, then $\mathbf{x}_0$ is in the unstable subspace and $\lim_{t\to-\infty} |\mathbf{x}(t)| = 0$.

(iii) Otherwise, $\mathbf{x}_0$ is in the center subspace, and then if $m_\lambda = 1$ for $\Re(\lambda) = 0$, then $|\mathbf{x}(t)| = |\mathbf{x}_0|$ for all t.

If $m_\lambda > 1$, then there is at most polynomial growth of $\mathbf{x}(t)$ in time.

In essence, for linear systems then we can find linear subspaces E_s, E_u and E_c that are invariant, that is solutions starting on one of these subspaces will remain on it for all time, and in addition we know the asymptotic behavior of all solutions on these subspaces immediately. The goal of the Stable Manifold Theorem and Center Manifold Theorem in the following section are to derive an analog of this for nonlinear systems. When nonlinearity is present however, a linear subspace is no longer sufficient but we must instead return to the concept of a manifold.

Definition 5.8.2. *A manifold of dimension n is a subset of some linear vector space that is locally isomorphic to $\mathbb{R}^n$.*

Remark 5.8.3. There are more careful definitions of manifolds out there, but we will resort to this one for simplicity and because we don't really need the more complicated versions. Typically you can consider a manifold of dimension n something that locally 'looks' like $\mathbb{R}^n$.

Example 5.8.4. Some examples of commonly seen manifolds include

(i) The unit circle in $\mathbb{R}^2$ is a 1 dimensional manifold.

(ii) The surface of a donut or torus is a 2-manifold.

(iii) A pringle chip is a 2-manifold.

(iv) The figure eight is not a manifold because the line crossing itself does not locally look like $\mathbb{R}$.

Now to construct the unstable or stable manifold, we will need to make use of the Contraction Mapping Principle yet again, in addition we recall that

$$\sum_{\lambda \in \sigma(A)} P_\lambda = I$$

so that for a hyperbolic fixed point $P_s + P_u = I$. We also note that Theorem 5.1.2 and Theorem **??** apply to initial data projected onto the stable subspace i.e. $\|P_s e^{tA}\|$ will decay to 0 exponentially fast.

Recall that for a hyperbolic fixed point $\overline{\mathbf{x}}$ of (5.1.17) and $\mathbf{v} = \mathbf{x} - \overline{\mathbf{x}}$ then

$$\mathbf{v}' = A\mathbf{v} + Q(\mathbf{v}), \tag{5.17}$$

where $A = Df(\overline{\mathbf{x}})$, $Q(\mathbf{v}) = P_s Q(\mathbf{v}) + P_u Q(\mathbf{v})$ and

$$Q(\mathbf{v}) = f(\mathbf{v}) - A\mathbf{v} \Rightarrow Q(0) = 0 \quad \text{and} \quad DQ(0) = 0.$$

Theorem 5.8.5 (Stable Manifold Theorem). *Let $\overline{\mathbf{x}}$ be a hyperbolic fixed point of* (5.1.17) *with linearization* (5.17)*. If $\dim(E_s) = k$ (there are k $\lambda \in \sigma(A)$ such that $\Re(\lambda) < 0$), then there is a k-dimensional differentiable manifold W_s^{loc} tangent to E_s such that for any $\mathbf{v}(0) \in W_s^{loc}$, $\mathbf{v}(t) \in W_s^{loc}$ (W_s^{loc} is invariant under the flow induced by* (5.17)*) and $\lim_{t\to\infty} \mathbf{v}(t) = 0$.*

Similarly (not shown here) there is an $n-k$ dimensional manifold W_u^{loc} tangent to E_u such that if $\mathbf{v}_0 \in W_u^{loc}$, then $\mathbf{v}(t) \in W_u^{loc}$ for all $t \leq 0$ and $\lim_{t\to\infty} \mathbf{v}(t) = 0$.

Proof. [Stable Manifold Theorem] The first step is to consider

$$\mathbf{v}(t) = P_s e^{tA}\mathbf{a} + \int_0^t P_s e^{A(t-s)} Q(\mathbf{v}(s))ds - \int_t^\infty P_u e^{A(t-s)} Q(\mathbf{v}(s))ds. \tag{5.18}$$

(i) You show in the homework that this is satisfied only if $\mathbf{v}(t)$ is a solution of (5.17).

(ii) Using this solution we construct the slow manifold.

(iii) The Uniform Contraction Mapping Principle is then used to prove there is a unique, continuous solution to (5.18).

(iv) Finally, we show that $\mathbf{v}(t)$ is exponentially decaying so that if $\mathbf{v}(0)$ is on the stable manifold, then $\mathbf{v}(t) \to 0$ as $t \to \infty$.

We proceed through each of these items below.

(i) Shown in the homework.

(ii) Now suppose that such a solution exists, then

$$\mathbf{v}(0) = \underbrace{P_s\mathbf{a}}_{E_s} - \int_0^\infty P_u e^{-As} Q(\mathbf{v}(s))ds.$$

This implicitly defines a manifold the same dimension as E_s by the graph $(P_s\mathbf{a}, h(P_s\mathbf{a}))$ where

$$h(P_s\mathbf{a}) = -\int_0^\infty P_u e^{-As} Q(\mathbf{v}(s))ds,$$

and $\mathbf{v}(s)$ inside the integral is implicitly dependent on $P_s\mathbf{a}$ (a k-dimensional object). Letting $\tilde{\mathbf{a}} = P_s\mathbf{a}$, then

$$Dh(\tilde{\mathbf{a}}) = -\int_0^\infty P_u e^{-As} DQ(\mathbf{v}(\tilde{\mathbf{a}})ds) \Rightarrow Dh(0) = 0$$

and

$$h(0) = 0(\text{since } \mathbf{v}(\tilde{\mathbf{a}} = 0) = 0 \text{ and } DQ(0) = 0)$$

so that this manifold defined by the graph is tangent to E_s at the origin.

(iii) First note (recall) that for $\xi \le |\Re(\sigma(A))|$,

$$\|(P_s e^{tA})\| \le c_0 e^{-(\xi-\varepsilon)t} \quad \text{and} \quad \|P_u e^{-tA}\| \le c_1 e^{-(\xi-\varepsilon)t} \quad \text{for} \quad t \ge 0.$$

Hence,

$$F(\mathbf{v}, \mathbf{a}) = P_s e^{tA}\mathbf{a} + \int_0^t P_s e^{A(t-s)} Q(\mathbf{v}(s))ds - \int_t^\infty P_u e^{A(t-s)} Q(\mathbf{v}(s))ds$$

$$\Rightarrow \|F(\mathbf{v}_1, \mathbf{a}) - F(\mathbf{v}_2, \mathbf{a})\| \le \int_0^t \|P_s e^{A(t-s)}(Q(\mathbf{v}_1(s)) - Q(\mathbf{v}_2(s)))\|ds$$

$$+ \int_t^\infty \|P_u e^{A(t-s)}(Q(\mathbf{v}_2(s)) - Q(\mathbf{v}_1(s)))\|ds$$

Now from the homework, because $Q(0) = 0$ and $DQ(0) = 0$, then $\exists\, \delta > 0$ such that if $\|\mathbf{v}_1\| < \delta$ and $\|\mathbf{v}_2\| < \delta$, then $\|Q(\mathbf{v}_1) - Q(\mathbf{v}_2))\| < \frac{\xi-\varepsilon}{4c}\|\mathbf{v}_1 - \mathbf{v}_2\|$ where $c = \max\{c_0, c_1\}$. Thus

$$\|F(\mathbf{v}_1, \mathbf{a}) - F(\mathbf{v}_2, \mathbf{a})\| \le \frac{\xi - \varepsilon}{4}\|\mathbf{v}_1 - \mathbf{v}_2\| \left[\int_0^t e^{-(\xi-\varepsilon)(t-s)}ds + \int_t^\infty e^{(t-s)}ds\right]$$

$$\le \frac{1}{4}\|\mathbf{v}_1 - \mathbf{v}_2\|[1 - e^{-(\xi-\varepsilon)t} - 0 + 1]$$
$$= \left(\frac{1}{2} - \frac{1}{4}e^{-(\xi-\varepsilon)t}\right)\|\mathbf{v}_1 - \mathbf{v}_2\|$$

$$\le \frac{1}{2}\|\mathbf{v}_1 - \mathbf{v}_2\| \Rightarrow F(\mathbf{v}(s), \mathbf{a})$$

is a uniform contraction mapping implying there is a unique fixed point $\mathbf{v}_a$ such that

$$\mathbf{v}_a(t) = P_s e^{tA}\mathbf{a} + \int_0^t P_s e^{A(t-s)} Q(\mathbf{v}_1(s))ds - \int_t^\infty P_u e^{A(t-s)} Q(\mathbf{v}_1(s))ds$$

is a solution of (5.17).

(iv) Now we want to show that $\mathbf{v}_a(t) \to 0$ as $t \to +\infty$ so all points on W_s^{loc} decay to the fixed point $\overline{\mathbf{x}}$ of the original system (5.1.17).

$$\|\mathbf{v}_a(t)\| \le e^{-(\xi-\varepsilon)t}\|\mathbf{a}\| + \int_0^t e^{-(\xi-t)(t-s)}\|Q(\mathbf{v}_a(s))\|ds + \int_t^\infty e^{(\xi-\varepsilon)(t-s)}\|Q(\mathbf{v}_a(s))\|ds.$$

Just as before, there is a $\delta > 0$ such that if $\|\mathbf{v}\| < \delta$, then $\|Q(\mathbf{v})\| < \frac{\xi-\varepsilon}{4c}\|\mathbf{v}\|$. If we let $\|\mathbf{a}\| < \frac{\delta}{2}$, then

$$\|\mathbf{v}\| \le e^{-(\xi-\varepsilon)t}\|\mathbf{a}\| + \frac{\xi - \varepsilon}{4}\left[\int_0^t e^{-(\xi-\varepsilon)(t-s)}\|\mathbf{v}\|ds + \int_t^\infty e^{(\xi-\varepsilon)(t-s)}\|\mathbf{v}\|ds\right]$$

$$= e^{-(\xi-\varepsilon)t}\|\mathbf{a}\| + \frac{1}{4}[2 - e^{-(\xi-\varepsilon)t}]\|\mathbf{v}\| \le e^{-(\xi-\varepsilon)t}\|\mathbf{a}\| + \frac{1}{2}\|\mathbf{v}\|$$

$$\Rightarrow \frac{1}{2}\|\mathbf{v}\| \le e^{-(\xi-\varepsilon)t}\|\mathbf{a}\|$$
$$\Rightarrow \|\mathbf{v}\| \le 2e^{-(\xi-\varepsilon)t}\|\mathbf{a}\|$$

so long as $\|\mathbf{v}\| < \delta$ but when $\|\mathbf{a}\| < \frac{\delta}{2} \Rightarrow \|\mathbf{v}\| < \delta$ so $\|(\mathbf{v}(t))\| \leq 2e^{-(\xi-\varepsilon)t}\|\mathbf{a}\|$ and $\mathbf{v}(t) \to 0$ as $t \to +\infty$.

$\square$

Remark 5.8.6. Note that the Stable Manifold at this point is only locally defined. This is easily carried to a global result by taking all points on this locally defined stable manifold and integrating backward in time to find points that will eventually fall onto it (speaking for $t > 0$). This backward integration taken to $t \to -\infty$ will define a global stable manifold with the local properties near $\overline{\mathbf{x}}$ dictated by the Stable Manifold Theorem.

Remark 5.8.7. The unstable manifold can be computed just as in the proof above, but with $t \to -t$ and P_s being replaced by P_u.

Remark 5.8.8. The utility of the Contraction Mapping Principle can be seen here as well, as we are given an algorithm for approximating the stable manifold. This is insightful for the underlying dynamics of the system (5.1.17) because the stable manifold is an invariant set.

Example 5.8.9. Construct the stable manifold at the origin for the nonlinear system:

$$\begin{aligned} \dot{x} &= -x - y^2 \\ \dot{y} &= y + x^2. \end{aligned}$$

First we note that the linearization of this problem at $x = y = 0$ is given by

$$\dot{\mathbf{x}} = A\mathbf{x} + Q(\mathbf{x}), \quad \text{where} \quad A = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad Q(\mathbf{x}) = \begin{bmatrix} -y^2 \\ x^2 \end{bmatrix}.$$

From the proof of the Stable Manifold Theorem, we need to find the fixed point to the map:

$$\mathbf{x}(t) = P_s e^{tA}\mathbf{a} + \int_0^t P_s e^{A(t-s)} Q(\mathbf{x}(s))ds - \int_t^\infty P_u e^{A(t-s)} Q(\mathbf{x}(s))ds,$$

and then $\mathbf{x}(0)$ will yield a parametric form of the slow manifold. For this problem,

$$P_s = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad P_u = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

If we let the initial guess be the fixed point $\mathbf{x}_0 = 0$ then we can easily compute

$$
\begin{aligned}
\mathbf{x}_1(t) &= \begin{bmatrix} a_1 e^{-t} \\ 0 \end{bmatrix} \\
\mathbf{x}_2(t) &= \begin{bmatrix} a_1 e^{-t} \\ 0 \end{bmatrix} + \int_0^t \begin{bmatrix} e^{-(t-s)} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ a_1^2 e^{-2s} \end{bmatrix} ds - \int_t^\infty \begin{bmatrix} 0 & 0 \\ 0 & e^{t-s} \end{bmatrix} \begin{bmatrix} 0 \\ a_1^2 e^{-2s} \end{bmatrix} ds \\
&= \begin{bmatrix} a_1 e^{-t} \\ 0 \end{bmatrix} - \int_t^\infty \begin{bmatrix} 0 \\ a_1^2 e^{t-3s} \end{bmatrix} ds \\
&= \begin{bmatrix} a_1 e^{-t} \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ -\frac{a_1^2}{3} e^{-2t} \end{bmatrix} \\
&= \begin{bmatrix} a_1 e^{-t} \\ -\frac{a_1^2}{3} e^{-2t} \end{bmatrix} \\
\mathbf{x}_3(t) &= \begin{bmatrix} a_1 e^{-t} \\ 0 \end{bmatrix} + \int_0^t \begin{bmatrix} e^{-(t-s)} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -\frac{a_1^4}{9} e^{-4s} \\ a_1^2 e^{-2s} \end{bmatrix} ds - \int_t^\infty \begin{bmatrix} 0 & 0 \\ 0 & e^{t-s} \end{bmatrix} \begin{bmatrix} -\frac{a_1^4}{9} e^{-4s} \\ a_1^2 e^{-2s} \end{bmatrix} ds \\
&= \begin{bmatrix} a_1 e^{-t} \\ 0 \end{bmatrix} + \int_0^t \begin{bmatrix} -\frac{a_1^4}{9} e^{-t-3s} \\ 0 \end{bmatrix} ds - \int_t^\infty \begin{bmatrix} 0 \\ a_1^2 e^{t-3s} \end{bmatrix} ds \\
&= \begin{bmatrix} a_1 e^{-t} \\ 0 \end{bmatrix} + \begin{bmatrix} \frac{a_1^4}{27} \left(e^{-4t} - e^{-t}\right) \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ \frac{a_1^2}{3} e^{-2t} \end{bmatrix} \\
&= \begin{bmatrix} a_1 e^{-t} + \frac{a_1^4}{27} \left(e^{-4t} - e^{-t}\right) \\ -\frac{a_1^2}{3} e^{-2t} \end{bmatrix},
\end{aligned}
$$

implying that

$$\mathbf{x}(0) \approx \begin{bmatrix} a_1 \\ -\frac{a_1^2}{3} \end{bmatrix}.$$

Hence the local stable manifold for this problem is given by

$$y = -\frac{x^2}{3}.$$

Similarly, the unstable manifold can be computed quite easily as

$$x = -\frac{y^2}{3}.$$

Remark 5.8.10. A much stronger result than the Stable Manifold Theorem is the Hartmann-Grobman Theorem which states that near a hyperbolic fixed point, the phase portrait (often referred to as the flow map) is topologically equivalent (a homeomorphism i.e. a continuous function with continuous inverse, maps the trajectories) to the linearized system. We will not attempt to prove the Hartmann-Grobman Theorem here, but do note that the key ingredient is the hyperbolicity of the fixed point $\overline{\mathbf{x}}$.

5.9 *Center Manifold Theory

The Stable Manifold Theorem, and the more generic Hartmann-Grobman Theorem discussed in the last section work only for hyperbolic fixed points. Although they are less frequently encountered as explained in Section **??**, non-hyperbolic fixed points still occur naturally, and are essential to consider in greater detail. This is manifested in the form of the Center Manifold Theorem, which is unfortunately far more complicated than the Stable Manifold Theorem, and to some degree, less satisfying. Before going any further, we require one additional definition to clarify the issues that relate to a non hyperbolic fixed point.

Definition 5.9.1. *For an $n \times n$ matrix A, define the* spectral gap *of A as*

$$\min\{|\Re(\lambda)| : \lambda \textit{ is an eigenvalue with non-zero real part}\}.$$

Note that if we let P_c be the projector onto the center subspace of (5.1.17) with corresponding spectral gap β then the arguments of Theorem 5.1.2 can be easily modified to see that

$$\begin{aligned} \|P_s e^{tA}\| &\leq C_\varepsilon e^{-(\beta-\varepsilon)t} \quad t \geq 0, \\ \|P_u e^{tA}\| &\leq C_\varepsilon e^{(\beta-\varepsilon)t} \quad t \leq 0, \\ \|P_c e^{tA}\| &\leq C_\varepsilon e^{\varepsilon|t|} \quad t \in \mathbb{R}, \end{aligned}$$

for any positive number $\varepsilon < \beta$.

With this in mind, we can state

Theorem 5.9.2 (Center Manifold Theorem). *Consider* (5.17)*, the linearization of* (5.1.17) *about a non hyperbolic fixed point $\overline{\mathbf{x}}$, where $Q(\mathbf{x})$ is in C^{k+1}. Let E_s, E_u, and E_c be the corresponding stable, unstable and center subspaces at this fixed point. Then there exists $\delta > 0$ and a local center manifold $\mathscr{M}$ so that*

(*i*) *There exists a C^k function $\phi : E_c \to \mathbb{R}^n$ with $P_c\phi(\mathbf{a}) = \mathbf{a}$ such that*

$$\mathscr{M} = \{\phi(\mathbf{a}) : \mathbf{a} \in E_c, \quad \|\mathbf{a}\| < \delta\}.$$

(*ii*) *The manifold $\mathscr{M}$ is locally invariant under* (5.17)*, meaning that any solution starting inside $\mathscr{M}$ will remain inside $\mathscr{M}$ for $|t|$ sufficiently small.*

(*iii*) *$\mathscr{M}$ is tangent to E_c at the origin.*

(*iv*) *Every globally bounded orbit remaining in a suitably small neighborhood of $\overline{\mathbf{x}}$ is entirely contained within $\mathscr{M}$.*

(*v*) *For any solution $\mathbf{x}(t)$ such that $\mathbf{x}(t) \to 0$ as $t \to +\infty$, there is an $\eta > 0$ and solution $\mathbf{y}(t) \in \mathscr{M}$ so that*

$$e^{\eta t}\|\mathbf{x}(t) - \mathbf{y}(t)\| \to 0 \quad \textit{as} \quad t \to +\infty.$$

Remark 5.9.3. We will not prove all parts of this Theorem here, in particular leaving the precise amount of differentiability available to the center manifold to further courses in dynamics. The key part of this result, and the reason for its usefulness is that for a non hyperbolic fixed point of an n dimensional system, the dynamics can be effectively reduced to a system of dimension $k < n$ where k is the dimension of V_c, the center subspace.

Proof. [Proof of Center Manifold Theorem] As with the proof of the Stable Manifold Theorem, there are several parts to this proof which we review before proceeding to each one individually.

(i) We show that it is sufficient to consider $Q(\mathbf{x})$ with compact support (vanishing outside of a compact set) and with arbitrarily small C^1 norm.

(ii) Once again in the homework, you will show that there is a particular solution to (5.17).

(iii) Using the contraction mapping principle in the properly defined space, we construct the center manifold.

(iv) The local invariance of the center manifold is established.

(v) We show that $\mathscr{M}$ is tangent to V_c.

(vi) Finally we establish that all solutions approaching the fixed point can be approximated by solutions on the center manifold.

(vii) The precise smoothness of the center manifold is not considered here but is covered in great detail in some notes written by Alberto Bressan (check his website and get a reference for this).

Proceeding as outlined:

(i) To show that $Q(\mathbf{x})$ can be replaced with a function with compact support and arbitrarily small C^1 norm by using a standard argument dependent on the use of a cutoff function. We will also make use of the variable $\mathbf{v} = \mathbf{x} - \overline{\mathbf{x}}$ where the fixed point now occurs at $\mathbf{v} = 0$. Let $\rho : \mathbb{R} \to [0, 1]$ be a sufficiently smooth (at least C^k), even function with compact support such that

$$\rho(z) = \begin{cases} 1 & \text{if } |z| \leq 1, \\ 0 & \text{if } |z| \geq 2. \end{cases}$$

Note here that

$$\rho(z) = \frac{1}{2} - \frac{1}{2}\tanh[(z - 3/2)/\varepsilon],$$

for $z \in [1, 2]$ would be sufficient. This is not a very constructive choice of $\rho(z)$ however. For $\varepsilon > 0$, define

$$Q_\varepsilon(\mathbf{v}) = \rho\left(\frac{\|\mathbf{v}\|}{\varepsilon}\right) Q(\mathbf{v}).$$

Recall that because $\mathbf{v} = 0$ is a fixed point of (5.17) then

$$\|Q(\mathbf{v})\|_\infty = O\left(\|\mathbf{v}\|^2\right) \quad \|DQ(\mathbf{v})\|_\infty = O\left(\|\mathbf{v}\|\right).$$

Using this we are interested in the C^1 norm of $Q_\varepsilon(\mathbf{v})$ which we can bound as follows:

$$\begin{aligned} \|Q_\varepsilon\|_{C^1} &= \|Q_\varepsilon\|_{L^\infty} + \|DQ_\varepsilon\|_{L^\infty} \\ &= \|Q_\varepsilon\|_{L^\infty(K)} + \|DQ_\varepsilon\|_{L^\infty(K)}, \end{aligned}$$

where the domain K is prescribed as all $\mathbf{v}$ such that $\|\mathbf{v}\| < 2\varepsilon$. Expanding the last line, and employing the triangle inequality we find that

$$\begin{aligned} \|Q_\varepsilon\|_{C^1} &\leq \|Q\|_{L^\infty(K)} + \frac{1}{\varepsilon}\left\|\rho'\left(\frac{\|\mathbf{v}\|}{\varepsilon}\right)\right\|_{L^\infty(K)} \|Q(\mathbf{v})\|_{L^\infty(K)} + \|DQ\|_{L^\infty(K)} \\ &= O(\varepsilon). \end{aligned}$$

When considering (5.17) we can then replace Q with Q_ε. In the following we then establish the existence of a global center manifold, using Q_ε to represent the nonlinearity implying the local results for the original problem as $Q = Q\varepsilon$ for $\|\mathbf{v}\| < \varepsilon$.

(ii) You will show in the homework that

$$\mathbf{x}(t) = e^{tA}\mathbf{a} + \int_0^t P_c e^{A(t-s)} Q(\mathbf{x}(s))ds + \int_{-\infty}^t P_s e^{A(t-s)} Q(\mathbf{x}(s))ds - \int_t^\infty P_u e^{A(t-s)} Q(\mathbf{x}(s))ds \tag{5.19}$$

is a solution to (5.17).

(iii) We first need to define the space of functions we are using. Let

$$X_\eta = \{\mathbf{x}(t) : \mathbb{R} \to \mathbb{R}^n : \quad \|x\|_\eta = \sup_t e^{-\eta|t|}\|\mathbf{x}(t)\| < \infty\},$$

be the normed linear space in question with norm provided in the definition. We define the map $\mathbb{F} : V_c \times X_\eta \to X_\eta$ by

$$\mathbb{F}(\mathbf{a},\mathbf{x})(t) = P_c e^{tA}\mathbf{a} + \int_0^t P_c e^{A(t-s)} Q(\mathbf{x}(s))ds + \int_{-\infty}^t P_s e^{A(t-s)} Q(\mathbf{x}(s))ds - \int_t^\infty P_u e^{A(t-s)} Q(\mathbf{x}(s))ds. \tag{5.20}$$

We first need to show that $\mathbb{F}(\mathbf{a},\mathbf{x}) \in X_\eta$, which is done using the bounds on e^{tA} discussed at the beginning of this section where we let $\varepsilon = \eta$ and $\varepsilon = \beta - \eta$ for the stable and unstable components respectively:

$$\begin{aligned}
\|\mathbb{F}(\mathbf{a},\mathbf{x})\|_\infty &\le C_\eta e^{\eta|t|}\|\mathbf{a}\|_\infty + \int_0^t C_\eta e^{\eta|t-s|}\|Q\|_\infty ds + \int_{-\infty}^t C_{\beta-\eta} e^{-\eta(t-s)}\|Q\|_\infty ds \\
&+ \int_t^\infty C_{\beta-\eta} e^{\eta(t-s)}\|Q\|_\infty ds \\
&= C\|Q\|_\infty e^{\eta|t|},
\end{aligned}$$

where C is a constant related to the required integrations, but is finite nonetheless indicating that $\|F\|_\eta \le C\|Q\|_\infty = O(\eta)$.

We next show that $\mathbb{F}$ is a contraction mapping in $\mathbf{x}$.

$$\begin{aligned}
\|\mathbb{F}(\mathbf{a},\mathbf{x}) - \mathbb{F}(\mathbf{a},\mathbf{y})\|_\infty &\le \int_0^t \left\|P_c e^{A(t-s)}\right\|_\infty \|Q(\mathbf{x}(s)) - Q(\mathbf{y}(s))\|_\infty ds \\
&+ \int_{-\infty}^t \left\|P_s e^{A(t-s)}\right\|_\infty \|Q(\mathbf{x}(s)) - Q(\mathbf{y}(s))\|_\infty ds \\
&+ \int_t^\infty \left\|P_u e^{A(t-s)}\right\|_\infty \|Q(\mathbf{x}(s)) - Q(\mathbf{y}(s))\|_\infty ds \\
&\le C_\varepsilon \nu \int_0^t e^{\varepsilon|t-s|} e^{\eta|s|}\|\mathbf{x}-\mathbf{y}\|_\eta ds + C_\varepsilon \nu \int_{-\infty}^t e^{-(\beta-\varepsilon)(t-s)} e^{\eta|s|}\|\mathbf{x}-\mathbf{y}\|_\eta ds \\
&+ C_\varepsilon \nu \int_t^\infty e^{(\beta-\varepsilon)(t-s)} e^{\eta|s|}\|\mathbf{x}-\mathbf{y}\|_\eta ds \\
&\le C'\nu e^{\eta|t|}\|\mathbf{x}-\mathbf{y}\|_\eta,
\end{aligned}$$

where ν is chosen so that $\nu \le \frac{1}{2C'}$, which is possible as noted in the proof of the Stable Manifold Theorem (and shown in the HW) when $\mathbf{x}$ and $\mathbf{y}$ are chosen sufficiently close to 0. Thus we have

$$\|\mathbb{F}(\mathbf{a},\mathbf{x}) - \mathbb{F}(\mathbf{a},\mathbf{y})\|_\eta \le \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_\eta,$$

so that $\mathbb{F}(\mathbf{a},\mathbf{x})$ is a contraction mapping in X_η with a unique fixed point $\mathbf{x}_a$.

In fact, because $\mathbb{F}(\mathbf{a},\mathbf{x})$ is Lipschitz (linear) with respect to $\mathbf{a}$ then the mapping $\mathbf{a} \to \mathbf{x}_a$ is Lipschitz as well. Just as in the Stable Manifold Theorem, we let the center manifold be defined by $\phi(\mathbf{a}) = \mathbf{x}_a(t=0)$, i.e.

$$\mathscr{M} = \{\phi(\mathbf{a}) : \mathbf{a} \in V_c\}.$$

(iv) To see that $\mathscr{M}$ is invariant (at least locally) under the flow, it is enough to note that all solutions $\mathbf{x}(t)$ that start on $\mathscr{M}$ at $t = 0$, will remain in X_η (at least for finite time) implying that they will also remain in $\mathscr{M}$.

(v) We now show that the center manifold is tangent to V_c. Let $\mathbf{a} \in V_c$ and consider $\mathbf{x}(t) = e^{tA}\mathbf{a}$. While this will not be a solution itself, we can use the fact that $\mathbb{F}$ is a contraction to estimate the distance between $\mathbf{x}$ and the fixed point $\mathbf{x}_a$ as

$$\|\mathbf{x} - \mathbf{x}_a\|_\eta \leq 2\|\mathbf{x} - \mathbb{F}(\mathbf{a}, \mathbf{x})\|_\eta.$$

Noting that $Q(\mathbf{x})$ is $O(|\mathbf{x}|^2)$ we see that

$$\|Q(\mathbf{x}(t))\|_\infty \leq C\|\mathbf{x}(t)\|_\infty \leq C\left(C_\varepsilon e^{\varepsilon|t|}\|\mathbf{a}\|_\infty\right)^2.$$

Using the definition of $\mathbb{F}(\mathbf{a}, \mathbf{x})$ we can then estimate

$$\begin{aligned}
\|\mathbf{x}(t) - \mathbb{F}(\mathbf{a}, \mathbf{x}(t))\|_\eta \leq C &\int_0^t C_\varepsilon e^{\varepsilon|t-s}\left(C_\varepsilon e^{\varepsilon|s|}\|\mathbf{a}\|_\infty\right)^2 ds \\
&+ C\int_{-\infty}^t C_\varepsilon e^{-(\beta-\varepsilon)(t-s)}\left(C_\varepsilon e^{\varepsilon|s|}\|\mathbf{a}\|_\infty\right)^2 ds \\
&+ C\int_t^\infty C_\varepsilon e^{(\beta-\varepsilon)(t-s)}\left(C_\varepsilon e^{\varepsilon|s|}\|\mathbf{a}\|_\infty\right)^2 ds \\
&\leq C'\|\mathbf{a}\|_\infty^2 e^{\eta|t|},
\end{aligned}$$

where the constant C' is independent of $\|\mathbf{a}\|_\infty$. It follows then that

$$\|\mathbf{x}(0) - \mathbf{x}_a(0)\|_\infty \leq \|\mathbf{x} - \mathbf{x}_a\|_\eta \leq 2\|\mathbf{x} - \mathbb{F}(\mathbf{a}, \mathbf{x})\|_\eta \leq 2C'\|\mathbf{a}\|_\infty^2.$$

Now, because $\mathbf{x}(0) = \mathbf{a}$ and $\mathbf{x}_a(0)$ defines the center manifold then we can see from this last estimate that the center manifold will be at worst $O\left(\|\mathbf{a}\|_\infty^2\right)$ near $\mathbf{a} = 0$ implying that the center manifold is indeed tangent to V_c (which is defined here by the particular function $\mathbf{x}(t)$).

(vi) Finally we will show the asymptotic approximation property.
Let $\mathbf{x}(t)$ be a solution of (5.17) such that $\mathbf{x}(t) \to 0$, and define the continuation of $\mathbf{x}(t)$ as

$$\overline{\mathbf{x}}(t) = \begin{cases} \mathbf{x}(t) & \text{if } t \geq 0, \\ \mathbf{x}(0) & \text{if } t < 0. \end{cases}$$

Note that $\overline{\mathbf{x}}(t)$ is a solution to the continued system of (5.17):

$$\overline{\mathbf{x}}'(t) = A\overline{\mathbf{x}} + Q(\mathbf{x}) + \phi(t), \qquad \phi(t) = \begin{cases} 0 & \text{if } t > 0, \\ -A\mathbf{x}(0) - Q(\mathbf{x}(0)) & \text{if } t < 0. \end{cases}$$

Thus we see that $\overline{\mathbf{x}}(t)$ can be rewritten using Duhamel's Principle as

$$\begin{aligned}
\tilde{x}(t) = e^{A(t-t_0)}P_s\overline{\mathbf{x}}(t_0) &+ \int_{t_0}^t e^{A(s-t_0)}P_sQ(\overline{\mathbf{x}}(s))ds + \int_{t_0}^t e^{A(s-t_0)}P_s\phi(s)ds \\
+ e^{A(t-t_1)}P_{cu}\tilde{x}(t_1) &+ \int_{t_1}^t e^{A(s-t_1)}P_{cu}Q(\overline{\mathbf{x}}(s))ds + \int_{t_1}^t e^{A(s-t_1)}P_{cu}\phi(s)ds,
\end{aligned}$$

where $P_{cu} = P_c + P_u$ denotes the projection onto the center and unstable subspaces.
We now consider the space of functions defined by

$$Z_\eta = \left\{z : \mathbb{R} \to \mathbb{R}^n, \|z(\cdot)\|_\eta = \sup_t e^{\eta t}\|\mathbf{z}(t)\|_\infty \leq \infty\right\}.$$

The procedure is then to find a function $\mathbf{z}(t) \in Z_\eta$ so that $\mathbf{y}(t) = \overline{\mathbf{x}}(t) + \mathbf{z}(t)$ is a global solution of (5.17) contained entirely on the center manifold $\mathscr{M}$. If this is true, then for any choice of t_0 and t_1 from the definition of $\overline{\mathbf{x}}(t)$ we see that $\mathbf{z}(t)$ must satisfy

$$\begin{aligned}
\mathbf{z}(t) &= -P_s\overline{\mathbf{x}}(t) + e^{A(t-t_0)}P_s(\overline{\mathbf{x}}(t_0) + \mathbf{z}(t_0)) + \int_{t_0}^{t} e^{A(s-t_0)}P_sQ(\overline{\mathbf{x}}(s) + \mathbf{z}(s))ds \\
&\quad - P_{cu}\overline{\mathbf{x}}(t) + e^{A(t-t_1)}P_{cu}(\overline{\mathbf{x}}(t_1) + \mathbf{z}(t_1)) + \int_{t_1}^{t} e^{A(s-t_1)}P_{cu}Q(\overline{\mathbf{x}}(s) + \mathbf{z}(s))ds \\
&= e^{A(t-t_0)}P_s\mathbf{z}(t_0) + \int_{t_0}^{t} e^{A(t-s)}P_s\left[Q(\overline{\mathbf{x}}(s) + \mathbf{z}(s)) - Q(\overline{\mathbf{x}}(s))\right]ds - \int_{t_0}^{t} e^{A(t-s)}P_s\phi(s)ds \\
&\quad + e^{A(t-t_1)}P_{cu}\mathbf{z}(t_1) + \int_{t_1}^{t} e^{A(t-s)}P_{cu}\left[Q(\overline{\mathbf{x}}(s) + \mathbf{z}(s)) - Q(\overline{\mathbf{x}}(s))\right]ds - \int_{t_1}^{t} e^{A(t-s)}P_{cu}\phi(s)ds.
\end{aligned}$$

If we let $t_0 \to -\infty$ and $t_1 \to +\infty$ then this defines a mapping which we can refer to as $\mathbf{G}(\mathbf{z})(t)$, which we can then show (just as was shown above) is a contraction mapping, indicating that there is a unique fixed point $\mathbf{z}(t) \in Z_\eta \subset X_\eta$ so that $\mathbf{y}(t) = \mathbf{z}(t) + \overline{\mathbf{x}}(t) \in X_\eta$ and thus $\mathbf{y}(t) \in \mathscr{M}$ and for all $t > 0$:

$$\|\mathbf{x}(t) - \mathbf{y}(t)\| = \|\mathbf{z}(t)\| \leq e^{-\eta t}\|\mathbf{z}\|_\eta.$$

□

Remark 5.9.4. The problem with this proof is that although one may suppose it is constructive via the use of the contraction mapping principle, the use of the cut-off function $\rho(z)$ in the first step overly complicates matters. In addition, we did not specify this function to be anything in particular (we only gave an example of such a function that may work in most cases). Construction of the center manifold does depend on this function however, indicating that the center manifold is unique only up to this choice, i.e. the center manifold is actually not unique, and there are typically a one-parameter family of such manifolds that can be used.

Remark 5.9.5. The center manifold theorem is useful because it allows us to treat the fixed points where the asymptotic stability is undetermined due to non-hyperbolicity. It in fact reduces the problem to the dimension of the center subspace.

Example 5.9.6. Consider the system

$$\begin{aligned}
\dot{x} &= x^2y - x^5 \\
\dot{y} &= -y + x^2.
\end{aligned}$$

This system's equilibrium at the origin has linearization given by the matrix

$$A = \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}, \quad Q(\mathbf{x}) = \begin{bmatrix} x^2y - x^5 \\ x^2 \end{bmatrix},$$

which has projections defined by

$$P_c = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad P_s = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \quad P_u = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

The matrix exponential is thus given by:

$$e^{tA} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + e^{-t} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix},$$

because $D_0 = D_{-1} = 0$.

Using the iterative technique gained from the proof above, and choosing an initial guess of $\mathbf{x}_0 = 0$,

$$\begin{aligned} \mathbf{x}_1 &= \begin{bmatrix} a_1 \\ 0 \end{bmatrix} \\ \mathbf{x}_2 &= \begin{bmatrix} a_1 \\ 0 \end{bmatrix} + \int_0^t \begin{bmatrix} -a_1^5 \\ 0 \end{bmatrix} ds + e^{-t} \int_{-\infty}^t \begin{bmatrix} 0 \\ a_1^2 e^s \end{bmatrix} ds \\ &= \begin{bmatrix} a_1 - a_1^5 t \\ a_1^2 \end{bmatrix}. \end{aligned}$$

Now this defines an approximate center manifold given by $\mathbf{x}_2(t=0) = (a_1, a_1^2)^T$, i.e. the approximate center manifold is described by $x = a_1$ and $y = a_1^2$ or $y = x^2$.

The real question at this point is what this means for the actual system. Returning to the original system, one might suppose that if we want to find out what happens near $x = y = 0$ we could simply look at the evolution equation for x (as this is the component of the flow in V_c) and insert $y = 0$, yielding the evolution equation

$$\dot{x} = -x^5,$$

which would indicate that the origin is stable (there is no unstable subspace in this case). However, inserting the form of the approximate center manifold instead we arrive at

$$\dot{x} = x^4 + O(x^5),$$

which for x small will be unstable.

Exercises

Note to the student: Each section of this chapter has several corresponding exercises, all collected here at the end of the chapter. The exercises between the first and second line are for Section 1, the exercises between the second and third lines are for Section 2, and so forth.

You should **work every exercise** (your instructor may choose to let you skip some of the advanced exercises marked with *). We have carefully selected them, and each is important for your ability to understand subsequent material. Many of the examples and results proved in the exercises are used again later in the text. Exercises marked with ⚠ are especially important and are likely to be used later in this book and beyond. Those marked with † are harder than average, but should still be done.

Although they are gathered together at the end of the chapter, we strongly recommend you do the exercises for each section as soon as you have completed the section, rather than saving them until you have finished the entire chapter.

5.1. Classify the type of equilibrium solution (for example, saddle point, stable node, etc.) of the origin for the linear system $\dot{\mathbf{x}} = A\mathbf{x}$ where

(i) $A = \begin{bmatrix} 1 & 0 \\ 4 & 2 \end{bmatrix}$.

(ii) $A = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$.

5.2. Perform the same analysis as the previous question but now for the following matrices:

(i) $A = \begin{bmatrix} 1 & -5 \\ 1 & -1 \end{bmatrix}$.

(ii) $A = \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix}$.

5.3. Prove Corollary 5.1.11.

5.4. Prove Proposition 5.1.20.

5.5. Find the equilibrium solutions and linearize about each one for the system given below. Then sketch the phase portrait:

$$\begin{aligned} \dot{x} &= \sin y \\ \dot{y} &= x - x^3. \end{aligned}$$

5.6. Find the equilibrium solutions and linearize about each one for the system given below. Then sketch the phase portrait:

$$\begin{aligned} \dot{x} &= y + x - x^3 \\ \dot{y} &= -y. \end{aligned}$$

5.7. Determine whether the following system is stable at the origin (you may use a computer to calculate the eigenvalues for this system):

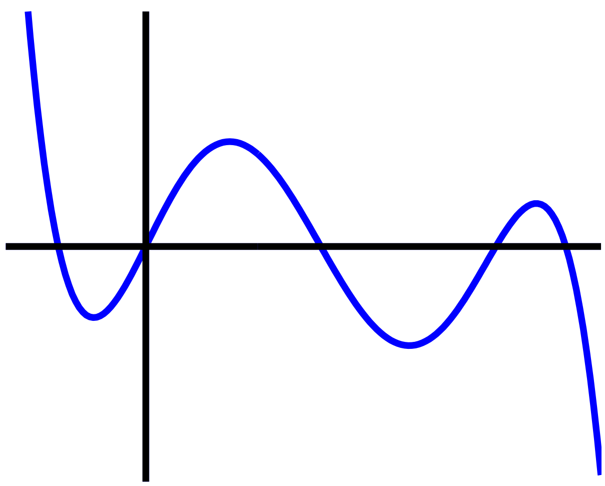
$$\begin{aligned} \dot{x} &= -10x + 10y \\ \dot{y} &= 28x - y - xz \\ \dot{z} &= xy - \frac{8}{3}z. \end{aligned}$$

5.8. For the system considered in the previous problem, find the additional equilibrium points, and evaluate the stability of these equilibria. Once again, feel free to use a computational resource for the calculation of the necessary eigenvalues. *From these analyses of the linearization at the equilibria, do you expect solutions to remain bounded for this system?*

5.9. Sketch the null clines, fixed points and a plausible phase portrait for the 2D system

$$\dot{x} = x(x - y)$$
$$\dot{y} = y(2x - y).$$

5.10. Graphically analyze the stability of the five fixed points $x_1 < x_2 < x_3 < x_4 < x_5$ of $\dot{x} = g(x)$ where $g(x)$ is shown in the figure below. Determine which fixed points are stable, and which are unstable, and what the interval of attraction for each of the stable fixed points is.



5.11.* Show that for Q as defined in (5.2), the claim in (5.5) holds; that is, if $\lim_{\|v\|\to 0} \frac{\|Q(\mathbf{v})\|}{\|\mathbf{v}\|} = 0$ then for any $\eta > 0$ there exists an $\alpha > 0$ so that if $\|\mathbf{v}\| < \alpha$ then $\frac{\|Q(\mathbf{v})\|}{\|\mathbf{v}\|} < \eta$.

5.12.* Examine the linearized stability of the origin and the coexistence fixed point for the nondimensional Lotka–Volterra system:

$$\dot{x} = bx - xy,$$
$$\dot{y} = -y + xy,$$

where $b > 0$.

5.13.* Determine the linear asymptotic stability of the three equilibria of the second order Lotka–Volterra model with $b > g > 0$.

5.14.* Consider the modified Lotka–Volterra system below, with $b > 0$:

$$\dot{x} = bx - \frac{xy}{1+x}$$
$$\dot{y} = -2y + \frac{xy}{1+x}.$$

(i) Examine the linearized stability of the origin of this system

(ii) Examine the linearized stability of the additional fixed point for this system.

(iii) Sketch a phase portrait and explain why this model is also unrealistic for predator-prey dynamics.

5.15. Let $H(x, y)$ be a C^2 function (meaning all derivatives are continuous). A Hamiltonian system is one of the form

$$\dot{x} = \frac{\partial H}{\partial y}$$
$$\dot{y} = -\frac{\partial H}{\partial x}.$$

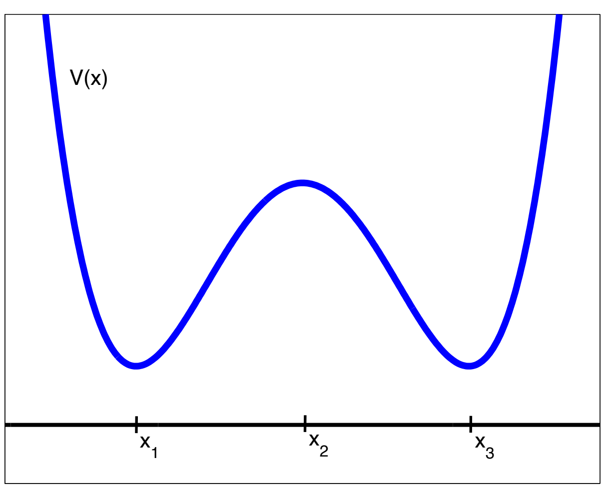
(i) Show that if the Hamiltonian system has an equilibrium (fixed point) then it can not be a sink or source.

(ii) Sketch the phase portrait for the Hamiltonian system with Hamiltonian $H(x, y) = x^2 + 4y^2$.

5.16. For a conservative system described by

$$\ddot{x} = -\frac{\partial V}{\partial x},$$

show that the total energy given by $E = \frac{1}{2}(\dot{x})^2 + V(x)$ remains constant for all time.



5.17. Consider the conservative system described in the last problem with potential $V(x)$ given by the figure above. Show that $x = x_k$ and $\dot{x} = 0$ gives three fixed points of this system. Show that the fixed points at x_1 and x_3 are centers (purely imaginary eigenvalues) and the fixed point at x_2 is a saddle (two real eigenvalues of opposite sign).

5.18. Examine the linear stability of the origin for the Lorenz system:

$$\begin{aligned}\dot{x} &= \sigma(y - x)\\ \dot{y} &= rx - y - xz\\ \dot{z} &= xy - bz,\end{aligned}$$

where b, r, and σ are all assumed to be positive. Hint: The stability of the origin is dependent on *one* of the parameters r, σ, and b.

5.19. Using $V(x) = \frac{1}{\sigma}x^2 + y^2 + z^2$, show that the origin in the Lorenz system with $0 < b$, $0 < r < 1$, and $< 0\sigma$ is nonlinearly (globally) asymptotically stable for the linearized stability. This indicates that the local linear stability result is actually global in this case. Hint: This is not straightforward, you will need to consider a quantity like $(x - \frac{r+1}{2}y)^2$, for example.

5.20.* TODO Homework for Lyapunov 2

5.21. Determine the linear stability of the origin and the fixed point where the predator population is zero for the predator-prey model of Example 5.6.14.

5.22. Determine the linear stability of the coexistence fixed point for the same system as the previous problem.

5.23. Find the value of the single fixed point for Example 5.6.16 and evaluate its linear stability. Verify that for $a = 0.08$ and $b = 0.6$ this point is unstable as indicated in Figure 5.15.

5.24. Show that the positive y axis and the horizontal line segment connecting $\left(0, \frac{b}{a}\right)$ and $\left(b, \frac{b}{a}\right)$ are boundaries on the trapping region as described in Example 5.6.16, that is show that the corresponding vector field points into the region along these curves.

5.25. Show that the origin is an unstable spiral (positive real part) for the system below:

$$\begin{aligned}\dot{x} &= x - y - x^3\\ \dot{y} &= x + y - y^3.\end{aligned}$$

Numerically (or rigorously if you can) convince yourself that there are no other *real* fixed points for this system.

5.26. Use the Lyapunov functional $V(x, y) = x^2+y^2$ and the Poincaré–Bendixson Theorem to show that there is a periodic orbit for the system of the previous problem. Hint: Consider level sets $V = 2 + \varepsilon$ and $V = 1 - \varepsilon$ for any small $\varepsilon > 0$.

5.27.* Following the proof in the text, show that the backward Euler method is convergent for the scalar ODE $\dot{u} = \lambda u$.

5.28. For the nonlinear scalar ODE $\dot{u} = f(u)$ where $f(u)$ has Lipschitz constant λ, show that the forward Euler method converges for a fixed time T as $\Delta t \to 0$.

5.29. Identify the region in the complex plane where the backward Euler method is absolutely stable.

5.30. Find the recursive relation defining the error E_n for the 2-step Adams–Bashforth method, and identify the corresponding roots of the equation.

5.31. Identify the recursive relation for the error E_n of the 2-stage, explicit Runge–Kutta method, and reproduce Figure 5.19.

5.32. *Find the absolute stability region for the explicit 4-stage Runge–Kutta method derived in Chapter 3.

5.33.* Let $f : \mathbb{R}^n \to \mathbb{R}^k$ be at least C^3, and defined such that $f(0) = 0$ and $Df(0) = 0$. Show that for the traditional Euclidian norm, and for every $K > 0$ there is a $\delta > 0$ such that if $\|\mathbf{u}\| < \delta$ and $\|\mathbf{v}\| < \delta$ then $\|f(\mathbf{u}) - f(\mathbf{v})\| \leq K\|\mathbf{u} - \mathbf{v}\|$. Hint: This is closely related to the mean-value theorem.

5.34. Show that (5.18) is a solution of (5.17).

5.35. Find the first three iterates to approximate the stable manifold at the origin for the following system:

$$\dot{x} = -x$$
$$\dot{y} = y + x^2.$$

5.36. Compute the first three iterates for the unstable manifold at the origin of the previous problem.

5.37. Sketch the unstable and stable manifolds as well as the stable and unstable subspaces for the previous problem. Also indicate how this affects the phase portrait of the system.

5.38. Consider the linear systems $\dot{\mathbf{x}} = A\mathbf{x}$ and $\dot{\mathbf{y}} = B\mathbf{y}$ where

$$A = \begin{bmatrix} -1 & -3 \\ -3 & -1 \end{bmatrix} \qquad B = \begin{bmatrix} 2 & 0 \\ 0 & -4 \end{bmatrix}.$$

Show that these two systems are topologically equivalent by constructing a homeomorphism that maps trajectories near the origin of one system onto trajectories near the origin of the other.

5.39. Show that (5.19) is a solution of (5.17).

5.40. Determine the qualitative behavior of solutions near the origin for the following system using the concepts drawn from center manifold theory:

$$\dot{x} = xy$$
$$\dot{y} = -y - x^2$$

5.41. As in the last problem, determine the behavior of solutions near the origin for the following system, relying on center manifold theory:

$$\dot{x} = y$$
$$\dot{y} = -y + xy$$

5.42. Now determine the qualitative behavior of solutions to the following system for $\alpha \neq 0$ which is a slight modification of the previous problem

$$\dot{x} = y$$
$$\dot{y} = -y + \alpha x^2 + xy.$$

5.43. For the linear system $\dot{\mathbf{x}} = A\mathbf{x}$, where A is the same matrix considered in Exercise 2.5 given by

$$A = \begin{bmatrix} -\delta & -\omega & 0 \\ \omega & -\delta & 0 \\ 0 & 0 & \alpha \end{bmatrix}$$

describe the stable, unstable and center subspaces for $\alpha > 0$ and both $\delta = 0$ and $\delta > 0$.

Notes

6 Bifurcation Theory

Math is a game played according to certain simple rules with meaningless marks on paper
—David Hilbert

Bifurcations are some of the most fascinating topics in Mathematics. Not only do they represent some of the most startling and surprising phenomena that we can actually observe, but they are also very fun to visualize. In fact most common bifurcations can best be explained by drawing a carefully designed picture far better than any formal mathematical expression.

What is a bifurcation? Up to this point we have talked a lot about the importance of different types of stability of solutions. We have also discussed the need to consider the stability of the model itself (think structural stability). As we have looked at all of these examples though, there has often been a parameter that appears. For instance we have shown (or you did in the exercises at least) that the famed Lorenz system:

$$\begin{aligned}\dot{x} &= \sigma(y - x),\\ \dot{y} &= rx - y - xz,\\ \dot{z} &= xy - bz,\end{aligned}$$

is globally asymptotically stable to the origin if $r < 1$. If $r > 1$ then we find that this is no longer the case. This means that there is an abrupt change in the phase portrait and hence the qualitative behavior of solutions as r passes through 1. This is called a bifurcation, although a formal definition of a bifurcation requires more mathematical sophistication, the basic idea is quite simple.

Every one of us experiences bifurcations on a daily basis. For instance, if we consider a pot of water on the stove, the temperature knob can be seen as a specific parameter that alters the behavior of the water in the pot (a parameter that we can adjust). At a specific position of the knob (dependent on elevation etc.) the water will start to boil which is a significant change in the water's behavior. This is a bifurcation point.

Another place where bifurcations occur is in political or economic markets. In this case, variations in populations can lead to changes in the stability of political parties or even pervading schools of thought. Naturally such bifurcations are a little harder to quantify, but the general concepts that underly bifurcation theory still apply, and in some cases can explain the swing between two different political or economic extremes.

In this book we will focus only on codimension 1 bifurcations meaning we are concerned with a bifurcation driven by changes in a single parameter of the system. Higher dimensional bifurcations are of course interesting and relevant, but also difficult to graphically demonstrate and analyze. Bifurcations of codimension 1 are straightforward and can all be illustrated with a few simple examples.

Now, let's draw some pictures!

6.1 Bifurcations of codimension 1

We denote a first-order system of differential equations with parameters as

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \eta)$$

where $\mathbf{f}$ is C^1 function from the product $\Omega \times \mathscr{U}$ to $\mathbb{R}^n$ for Ω open in $\mathbb{R}^n$ and $\mathscr{U}$ open in $\mathbb{R}^k$. Since we can translate $\mathbf{x}$ by a fixed vector, we assume without loss of generality that $0 \in \Omega$.

6.1.1 Persistence of Nondegenerate Equilibria

For $\eta_0 \in \mathscr{U}$, suppose that $\overline{\mathbf{x}} = 0$ is an equilibrium of $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \eta_0)$, i.e., $\mathbf{f}(0, \eta_0) = 0$.

Definition 6.1.1. *We call the equilibrium* $\overline{\mathbf{x}} = 0$ nondegenerate *if the linearization matrix* $A(\eta_0) = D\mathbf{f}(\overline{\mathbf{x}}, \eta_0)$ *does not have* 0 *as an eigenvalue; otherwise the equilibrium is called* degenerate*.*

Theorem 6.1.2 (Persistence Theorem). *For* $\eta_0 \in \mathscr{U}$*, if* $\overline{\mathbf{x}} = 0$ *is a nondegenerate equilibrium of* $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \eta_0)$*, then there exists an open* $V \subset \mathscr{U}$ *and a unique* C^1 *function* $g : V \to \Omega$ *such that* $\eta_0 \in V$*,* $g(\eta_0) = \mathbf{0}$*, and*

$$\mathbf{f}(g(\eta), \eta) = 0 \quad \textit{for} \quad \textit{all } \eta \in V.$$

Proof. Apply the Implicit Function Theorem to $\mathbf{f}(\mathbf{x}, \eta)$. □

For a nondegenerate equilibrium $\overline{\mathbf{x}} = 0$ with nondegenerate parameter η_0, changing or perturbing the parameter from η_0 to a nearby value η will simply move the equilibrium from $g(\eta_0) = 0$ to $g(\eta)$, i.e., the equilibrium persists uniquely.

Local Structural Stability of Hyperbolic Equilibria

Now suppose that the equilibrium $\overline{\mathbf{x}} = 0$ is hyperbolic, i.e., no eigenvalue of $A(\eta_0)$ has zero real part.

(i) Then for some $k \in \{0, 1, 2, \dots, n\}$ there are k eigenvalues of $A(\eta_0)$ with negative real part and $n - k$ eigenvalues with positive real part.

(ii) This means that the equilibrium is nondegenerate and so the Persistence Theorem applies.

(iii) Furthermore, since the eigenvalues of $A(\eta) = D\mathbf{f}(g(\eta), \eta)$ depend continuously on η, there are k eigenvalues of $A(\eta)$ that have negative real part and $n - k$ eigenvalues that have positive real part.

(iv) Thus the perturbed equilibrium $g(\eta)$ remains hyperbolic.

(v) As the eigenvalues of $A(\eta)$ determine the local dynamics of the hyperbolic equilibrium, the local dynamics near the perturbed equilibrium $\overline{\mathbf{x}} = g(\eta)$ do not change.

(vi) This says that hyperbolic equilibrium $\overline{\mathbf{x}} = g(\eta_0)$ are locally structurally stable.

6.1.2 Necessary Condition for Bifurcations

For there to be a change or bifurcation in the local dynamics near $\overline{\mathbf{x}} = 0$ the matrix $A = D\mathbf{f}(0, \eta_0)$ would have to have an eigenvalue with zero real part (including the case of having zero as an eigenvalue), i.e., the equilibrium can not be hyperbolic for a bifurcation to occur.

Hence a a necessary condition for a bifurcation is when the equilibrium $\overline{\mathbf{x}} = 0$ is not hyberbolic.

Remark 6.1.3. Note that nonhyperbolicity of the equilibrium is not a sufficient condition for the occurrence of a bifurcation. For example, consider the equation

$$\dot{x} = \eta^2 x + x^3 = x(\eta^2 + x^2),$$

which has a degenerate equilibrium $x = 0$ when $\eta = 0$ however for every $\eta \in \mathbb{R}$, the phase portrait consists of one unstable equilibrium at $x = 0$.

Definition 6.1.4 (Pseudo-definition of Bifurcation). *A bifurcation occurs for*

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \eta),$$

at $\eta = \eta_0$ when the local dynamics at an equilibrium $\overline{\mathbf{x}}$ for η_0 differ from the local dynamics near $\overline{\mathbf{x}}$ for $\eta \neq \eta_0$.

6.1.3 Bifurcations of codimension 1

The number of parameters required to describe the bifurcation is called the codimension of the bifurcation. We will illustrate "generic" codimension-one bifurcations.

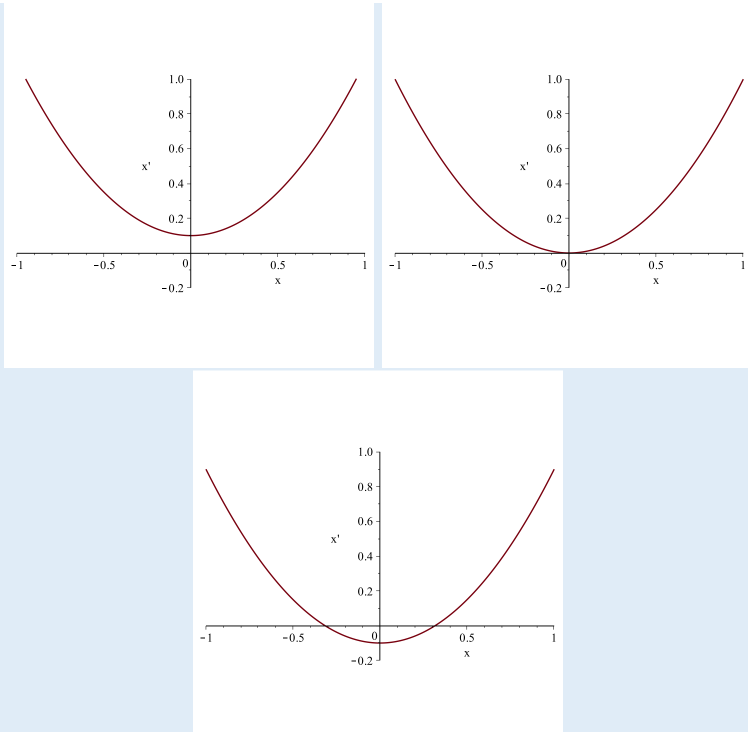
Let's consider the case where $x \in \mathbb{R}$. For this setting an equilibrium being not hyperbolic is equivalent to it being degenerate, i.e., the 1×1 matrix $A(\eta) = D\mathbf{f}(0, \eta)$ has an eigenvalue with zero real part if and only if the eigenvalue is zero.

Example 6.1.5 (Saddle Node Bifurcation). The one-parameter family of ODEs

$$\dot{x} = x^2 + \eta$$

has no equilibrium solutions when $\eta > 0$, exactly one equilibrium at $x = 0$ when $\eta = 0$, and two equilibria $x = \pm\sqrt{\eta}$ when $\eta < 0$.

Here are graphs of $f(x) = x^2 + \eta$ for $\eta = 0.1, 0, -0.1$.



(i) The phase portrait for $\eta > 0$ (upper left plot) has no equilibrium.

(ii) (a) The phase portrait for $\eta = 0$ (upper right plot) has one equilibrium at the origin with orbits to the left converging to the equilibrium and orbits to the right moving away from the equilibrium at the origin; this makes the equilibrium at the origin half-stable or semistable, meaning it is attracting on one side, and unstable on the other side.

(b) The linearization of the vector field at the equilibrium at the origin when $\eta = 0$, is $Df(0) = 2(0) = 0$, i.e. it is degenerate (and nonhyperbolic).

(iii) The phase portrait for $\eta < 0$ has two equilibria; the equilibrium at $-\sqrt{\eta}$ is asymptotically stable, and the equilibrium at $\sqrt{\eta}$ is unstable.

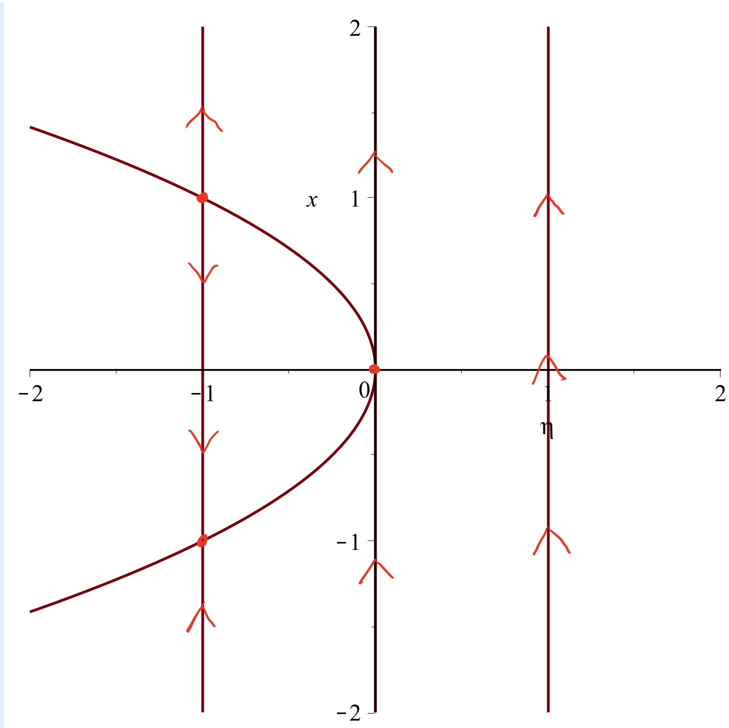
Since the phase portraits for $\eta = 0$ and for $\eta \neq 0$ are different, the equation $\dot{x} = x^2 + \eta$ has a bifurcation at $\eta = 0$. We call $\eta = 0$ the bifurcation value.

We sketch what is called the bifurcation diagram for $\dot{x} = x^2 + \eta$ below. The bifurcation diagram:

(i) places the parameter on the horizontal axis,

(ii) the variable x is on the vertical axis,

(iii) and on some values of η (vertical lines in the bifurcation diagram), we plot the phase line for $\dot{x} = x^2 + \eta$. In this particular bifurcation diagram we do this for $\eta = -1, 0, 1$.



This example describes what is called a saddle-node bifurcation which occurs when two equilibria with opposite stability coalesce into one degenerate equilibrium at the bifurcation value, then disappear altogether.

Example 6.1.6 (Transcritical Bifurcation). The one-parameter family

$$\dot{x} = \eta x - x^2$$

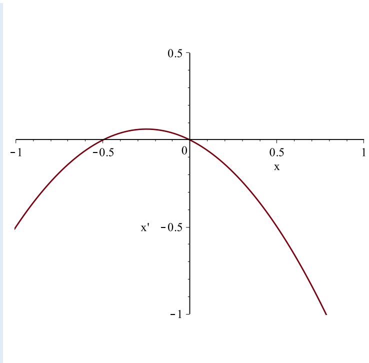
has:

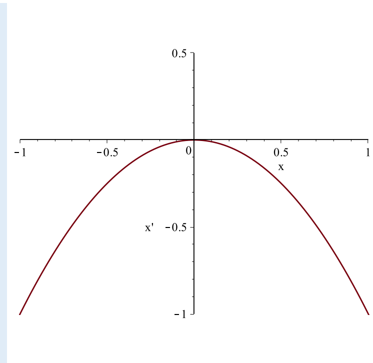
(i) two equilibria at $x = \eta$ and $x = 0$ when $\eta < 0$,

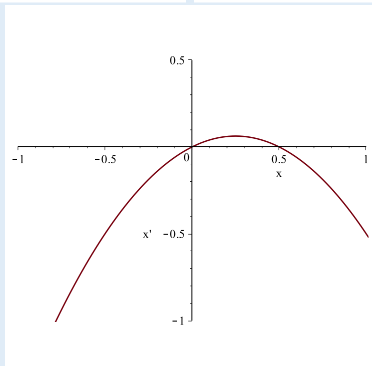
(ii) one equilibrium $x = 0$ when $\eta = 0$,

(iii) and two equilibria 0 and η when $\eta > 0$.

Here are graphs of $f(x) = \eta x - x^2$ for $\eta = -0.5, 0, 0.5$.







(i) The phase portrait when $\eta < 0$ has an unstable equilibrium at $x = \eta$ and an asymptotically stable equilibrium at $x = 0$.

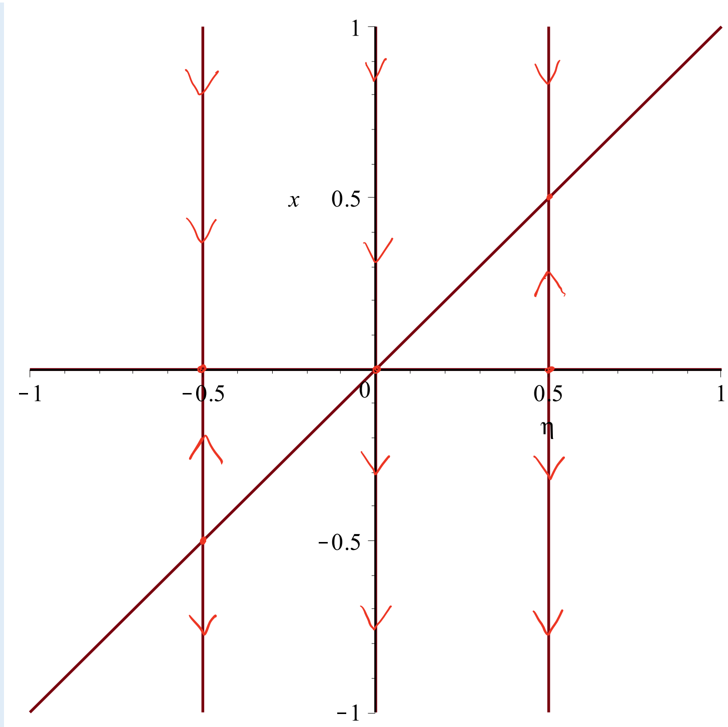
(ii) (a) The phase portrait when $\eta = 0$ has a half-stable equilibrium at the origin (unstable on the negative side, asymptotically stable on the positive side).

(b) The linearization of the vector field at the equilibrium at the origin is $Df(0) = -2(0) = 0$, i.e., it is degenerate (and nonhyperbolic).

(iii) The phase portrait when $\eta > 0$ has an unstable equilibrium at the origin and an asymptotically stable equilibrium at η.

Since the phase portraits for $\eta = 0$ and for $\eta \neq 0$ are different, the equation $\dot{x} = \eta x - x^2$ has a bifurcation at $\eta = 0$.

Here is the bifurcation diagram for this bifurcation.



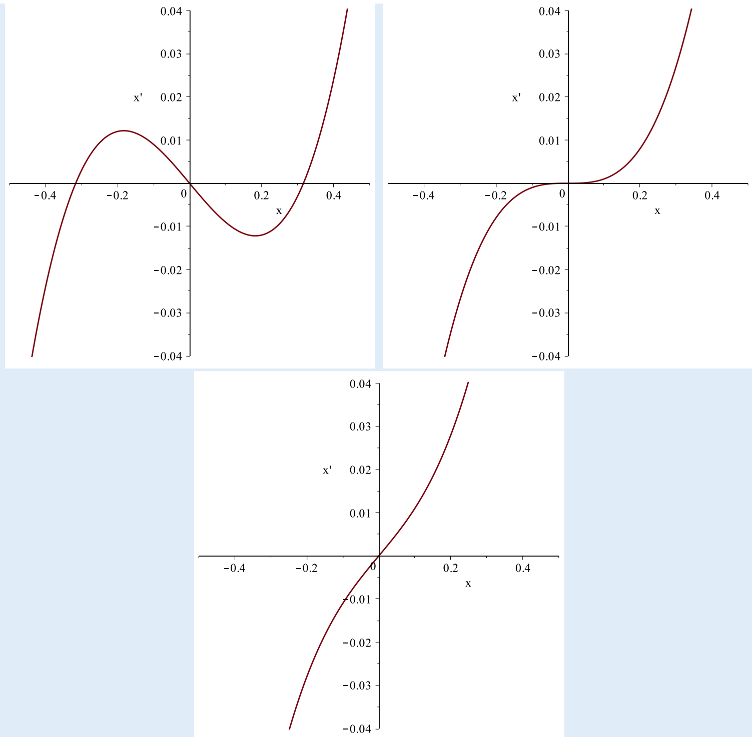
The bifurcation is called *transcritical* when two equilibria with opposite forms of stability pass through each other at the bifurcation value and exchange the type of stability; at the bifurcation value the two equilibria coalesce into one degenerate equilibrium.

Example 6.1.7 (Subcritical Pitchfork Bifurcation). The one-parameter family

$$\dot{x} = \eta x + x^3$$

has three equilibria near the origin when $\eta < 0$ and is close to 0, has one equilibrium near the origin when $\eta = 0$, and has one equilibrium near the origin when $\eta > 0$ and η is small.

Here are the graphs of $f(x) = \eta x + x^3$ for $\eta = -0.1, 0, 0.1$.

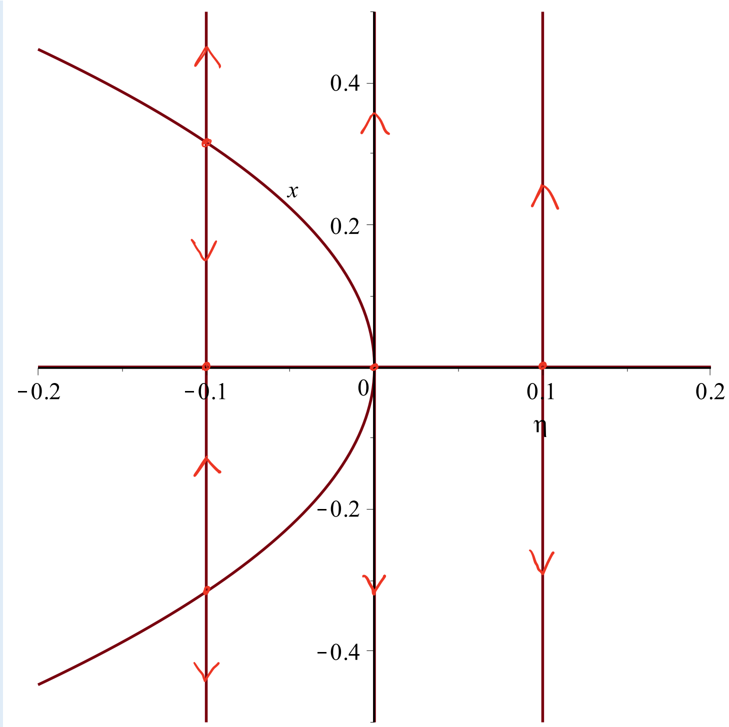


(i) The phase portrait when $\eta < 0$ is that of an unstable equilibrium to the left of the origin, an asymptotically stable equilibrium at the origin, and an unstable equilibrium to the right of the origin.

(ii) The local phase portrait when $\eta = 0$ is that of unstable equilibrium at the origin. The linearization of the vector field at the equilibrium at the origin is $Df(0) = \eta + 3x^2|_{x=0,\eta=0} = 0$, so the equilibrium is degenerate (and nonhyperbolic).

(iii) The phase portrait when $\eta > 0$ is that of an unstable equilibrium.

Here is the bifurcation diagram for this subcritical pitchfork bifurcation.



The subcritical pitchfork bifurcation is when three equilibria with alternating stability types coalesce into one equilibrium at the bifurcation value, with the "middle" equilibrium surviving but switching stability type. Do you see the pitchfork of the equilibria in the bifurcation diagram?

Remark 6.1.8. Clearly we have not (nor will we) demonstrated all of the possible bifurcations of codimension 1. Even so, we note that all of the examples in this Chapter are sufficient to demonstrate all of the qualitative features of potential codimension 1 bifurcations. This means that the pictures/figures we have sketched here will adequately describe all codimension 1 bifurcations at least locally or for values of the parameter η that are near the bifurcation value.

Example 6.1.9 (Hysteresis). The one-parameter family

$$\dot{x} = (\eta - 0.3)x + x^3 - x^5$$

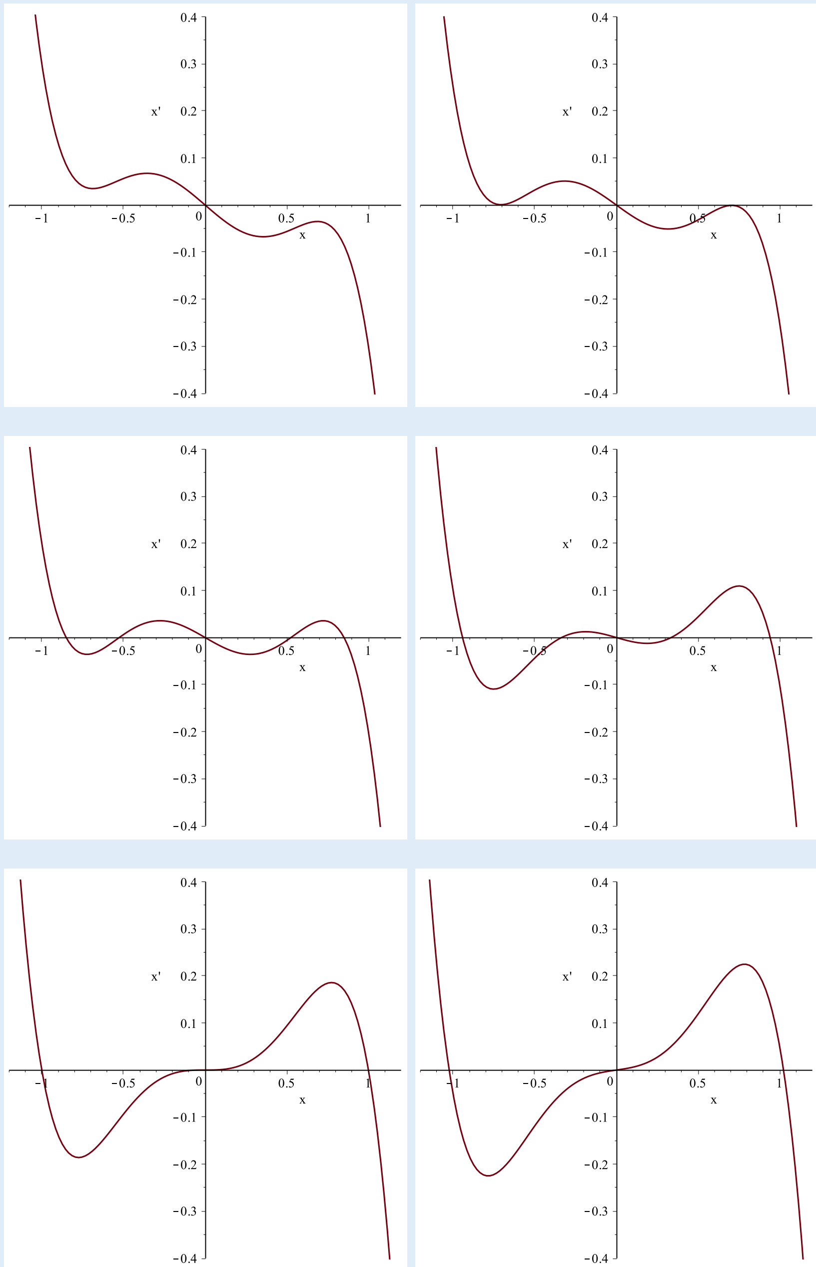
has several local bifurcations in it which demonstrate a qualitative phenomenon referred to as *hysteresis*.

To find the equilibrium we find the roots of the fifth degree polynomial

$$f(x) = (\eta - 0.3)x + x^3 - x^5 = x((\eta - 0.3) + x^2 + x^4).$$

The quartic factor is a quadratic in x^2 and so can be solved for x^2 by the quadratic formula, i.e. these equilibria can be identified exactly (up to a root of the quadratic formula).

Here are graphs of $f(x) = \eta x + x^3 - x^5$ for $\eta = 0, 0.049, 0.1, 0.2, 0.3, 0.305$.

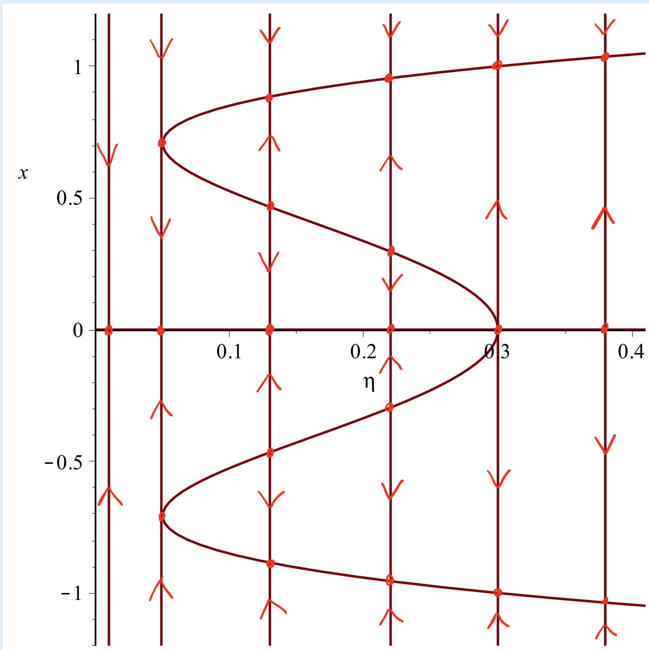


(i) The parameter $\eta = 0.049$ is a bifurcation value where two local saddle-node bifurcations occur simultaneously, with one equilibrium appearing near $x = -0.7$ and the other appearing near $x = 0.7$.

(ii) The parameter $\eta = 0.3$ is a bifurcation value where a local subcritical pitchfork bifurcation occurs at the origin.

The phase portrait of $\dot{x} = (\eta - 0.3)x + x^3 - x^5$ for $\eta \neq 0.049$ and $\eta \neq 0.3$ is best explained by considering the bifurcation diagram.

Here is the bifurcation diagram for $\dot{x} = (\eta - 0.3)x + x^3 - x^5$.



To interpret this bifurcation diagram, suppose that the equation

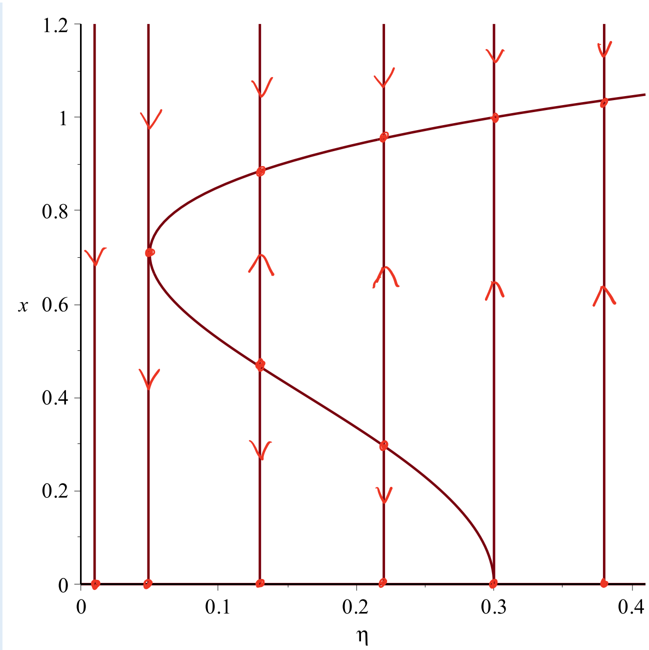
$$\dot{x} = (\eta - 0.3)x + x^3 - x^5 = x((\eta - 0.3) + x^2 - x^4)$$

models a percent of a population, such as fruit flies where η is a parameter related to the food supply, then:

(i) $0 \leq x \leq 1$ is a natural constraint to use.

(ii) The equilibrium at the origin is there for all η, and hence the set $x \geq 0$ is invariant.

Here is the relative part of the bifurcation diagram, i.e. $x \geq 0$.



To explain the notion of hysteresis:

(i) Suppose that $\eta = 0.1$ initially (very little food supply, i.e., fruit, for fruit flies is available) and $x = 0$ (no fruit flies).

(ii) Then suddenly η changes to $\eta = 0.4$ (a plentiful amount of fruit is introduced) and accompanying that fruit are a few fruit flies, meaning $x > 0$ and small.

(iii) The equilibrium $x = 0$ is unstable when $\eta = 0.4$ and so x moves away from 0 towards the asymptotically stable equilibrium near 1 (meaning we are nearing the maximum amount of fruit flies the available fruit can sustain).

(iv) Removing most of the fruit (the food supply for the fruit flies) drops the value of η back to $\eta = 0.1$, but the fruit fly percentage population x doesn't decay to 0. Why?

(v) Because x decays to the asymptotically stable equilibrium near 0.8.

This is hysteresis: returning the parameter η to its original value does not mean the value of x returns to its original value.

To return x to its original value of 0 we would have to remove all the fruit, i.e., move η from 0.1 to 0, in which the equilibrium at the origin is globally asymptotically stable and wait for x to decay to 0, then increase η to 0.1, and which point we have $x = 0$.

Remark 6.1.10. The previous example illustrates the point that while we do not plan on walking through every single bifurcation of codimension 1, we have effectively demonstrated all of the qualitative features for any such bifurcation. For instance, this hystersis example displays two *local* saddle-node bifurcations. If we had focused on just a single one of these bifurcations, they would have looked very much like the saddle-node example we have already examined. This occurs similarly when the pitchfork bifurcation occurs at $\eta = 0.3$.

Remark 6.1.11. If we think of $x(t)$ as the unemployment rate in the UK between 1980 and 1986, and η as the parameter collectively describing the state of economy (smaller η means better economy, higher η means worse economy), then with the recession that hit the UK in 1981, i.e., η changed from slightly below 0.3 to slightly above 0.3, so that the unemployment rate grew dramatically (from 1.5 million unemployed to 2.0 million unemployed). After the recession, i.e., when η returned to below 0.3, the unemployment rate continued to grow between 1984 and 1986 (with more than 3.0 million unemployed), i.e., x was above the unstable equilibrium and moving toward the asymptotically stable equilibrium that is near 1. This again is hysteresis. [Online source for UK recession: Investopedia—Hysteresis.]

Remark 6.1.12. Hysteresis is far more common than most people realize, and it can explain some of the most confusing phenomena observed in all of the different scientific fields. For instance, there are many economic policies that have hysteretic effects with drastic results. Specifically, recessions and depressions are most certainly caused by a myriad of complicated factors, however the basic principle of hysteresis explains why a simple reversal of policies and procedures can not reverse a recession or depression, but far more drastic measures are necessary. The same can be said for public health guidelines etc.. High level government decisions that are based on the premise that the results can be reversed, are questionable at best, and this can be seen rather easily by as simple a picture as Figure **??**.

Remark 6.1.13. Thus far we have considered what happens to the system when the eigenvalues of a fixed point pass through the origin on the real axis as the parameter is adjusted. It is also of interest to consider what happens as a pair of complex valued eigenvalues passes through the imaginary axis. This is referred to as a Hopf bifurcation (Hopf-Andreev bifurcation as well), and just as with the pitchfork bifurcation there is a subcritical and a supercritical case.

You will explore the supercritical case in the homework, but the basic idea is that the stability of the fixed point changes from being a stable spiral, to unstable but the solutions remain bounded (this is usually shown via some Lyapunov type argument). The instability of the fixed point then leads to a limit cycle appearing in the solution space. See the youtube video: 'Vector field: what is a Hopf bifurcation'.

Example 6.1.14 (Subcritical Hopf Bifurcation). Consider the two-dimensional system

$$\begin{aligned} r' &= \eta r + r^3 - r^5 \\ \theta' &= \omega + br^2. \end{aligned}$$

Fixed points of the $r' = 0$ equation in these cylindrical coordinates correspond to either a fixed point, or a limit cycle (perfect circle in this case). These correspond to what we would expect for a subcritical pitchfork bifurcation in the radius r, i.e.

$$r = 0,\ \pm\sqrt{\frac{1}{2} \pm \frac{1}{2}\sqrt{1+4\eta}}.$$

The only difference here is that $r < 0$ is not physically relevant, so we really only need worry about the positive root. If $\eta < 0$ there are 3 real solutions, i.e. the fixed point at the origin which it turns out is stable (as seen for the subcritical pitchfork bifurcation above), another unstable limit cycle of relatively small amplitude, and another large amplitude stable limit cycle. As η passes through zero, the unstable limit cycle coalesces at the origin, vanishing for $\eta > 0$ but making the origin unstable in the process and the stable, large amplitude limit cycle remains.

Remark 6.1.15. The subcritical Hopf bifurcation is another example of hysteresis. For $\eta < 0$ solutions near the origin will remain near the origin, but if the parameters of the system change ever so slightly so that $\eta > 0$ then solutions near the origin will suddenly and abruptly jump to the large amplitude oscillations represented by the stable limit cycle of large amplitude. Simply decreasing η back through zero will not return these solutions to the origin, but there is a gap wherein the large amplitude oscillations remain stable and away from the origin.

Exercises

Note to the student: Each section of this chapter has several corresponding exercises, all collected here at the end of the chapter. The exercises between the first and second line are for Section 1, the exercises between the second and third lines are for Section 2, and so forth.

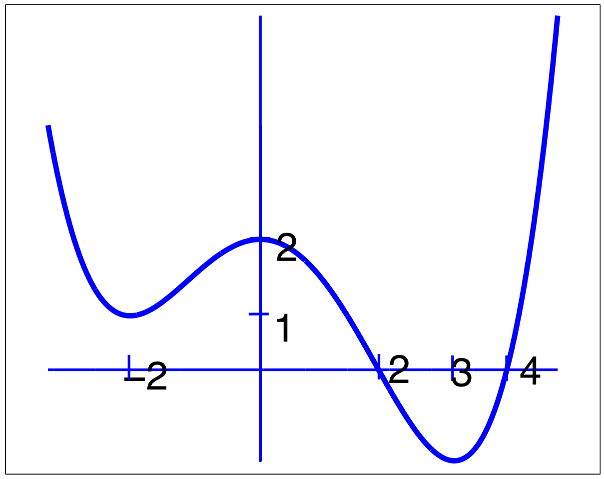
You should **work every exercise** (your instructor may choose to let you skip some of the advanced exercises marked with *). We have carefully selected them, and each is important for your ability to understand subsequent material. Many of the examples and results proved in the exercises are used again later in the text. Exercises marked with ⚠ are especially important and are likely to be used later in this book and beyond. Those marked with † are harder than average, but should still be done.

Although they are gathered together at the end of the chapter, we strongly recommend you do the exercises for each section as soon as you have completed the section, rather than saving them until you have finished the entire chapter.

6.1. Sketch and justify the bifurcation diagram for the *supercritical pitchfork* bifurcation diagram $\dot{x} = \eta x - x^3$.

6.2. For the ODE $\dot{x} = \sin(x) + \eta$, sketch a bifurcation diagram and the appropriate phase portraits as $\eta \in \mathbb{R}$ varies.

6.3. The Figure below is a graph of the function $y = f(x)$. Sketch the bifurcation diagram of $\dot{x} = f(x) - a$ for $-\infty < a < +\infty$.



6.4. Analyze the bifurcations by sketching a bifurcation diagram and relevant phase portraits for:

$$\dot{x} = x - \eta x^2.$$

6.5. Determine the linear stability of the origin for the system:

$$\begin{aligned} \dot{x} &= -y + x(\eta - x^2 - y^2) \\ \dot{y} &= x + y(\eta - x^2 - y^2). \end{aligned}$$

How does the linear stability of the origin depend on η?

6.6.* For the system of the previous problem, show that in polar coordinates $x = r\cos\theta$ and $y = r\sin\theta$ that the ODE is equivalent to:

$$\begin{aligned} r' &= r(\eta - r^2) \\ \theta &= 1. \end{aligned}$$

6.7.* Show that for $\eta > 0$,

$$\gamma_\eta(t) = \sqrt{\eta}\begin{bmatrix}\cos t \\ \sin t\end{bmatrix},$$

is a periodic solution of the system of the previous two problems.

Notes