One-dimensional dynamical systems

Consider the following initial value problem for one (autonomous[1]) ODE

$$\begin{cases} \dfrac{dx}{dt} = f(x) \\ x(0) = x_0 \end{cases} \tag{1}$$

where $f : D \mapsto \mathbb{R}$ and $D \subseteq \mathbb{R}$ is a subset of $\mathbb{R}$. In order for the initial value problem (1) to be well-posed (a problem is well-posed if the solution exist and is unique), it is necessary and sufficient for $f(x)$ to be Lipschitz continuous in D.

Definition 1 (Lipschitz continuity)**.** Let $D \subseteq \mathbb{R}$ be a subset of $\mathbb{R}$. We say that $f : D \times [0, T] \to \mathbb{R}$ is Lipschitz continuous in D if there exists a positive constant $0 \leq L < \infty$ (Lipschitz constant) such that

$$|f(x_1) - f(x_2)| \leq L\,|x_1 - x_2| \qquad \text{for all } x_1, x_2 \in D. \tag{2}$$

The smallest number L^* such that the inequality above is satisfied is called "best" Lipschitz constant.

Lipschitz continuity is stronger than continuity, which requires only that[2]

$$\lim_{y_1 \to y_2^{\pm}} |f(y_1) - f(y_2)| = 0 \quad \text{for all } t \in [0, T] \text{ and for all } y_2 \in D \text{ (excluding boundary).} \tag{3}$$

In fact, Lipschitz continuity implies that the rate at which $f(x_1)$ approaches $f(x_2)$ as $x_1 \to x_2$ cannot be larger than L. In other words, a Lipschitz continuous function $f(x)$ has a growth rate that is bounded by L for all x in D.

Example: Let $D = [-1, 1]$ be a closed interval, i.e., an interval including the endpoints -1 and 1. The function $f(x) = x^{1/3}$ is continuous in D for all $t \in \mathbb{R}$ (see Figure 1). However, $f(x)$ is not Lipschitz continuous in D. The problem here is that $f(x)$ has infinite "slope" at the point $x = 0$. In other words, there is no constant $0 \leq L < \infty$ such that

$$|f(x) - f(0)| \leq L\,|x - 0| \quad \text{for all} \quad x \in D. \tag{4}$$

This can be seen by substituting $f(x) = y^{1/3}$ in (4)

$$|f(x)| \leq L\,|x| \quad \Rightarrow \quad \left|\frac{x^{1/3}}{x}\right| = \left|\frac{1}{x^{2/3}}\right| \leq L \quad \text{for all} \quad x \in D. \tag{5}$$

Clearly, if we send x to zero we have that $\left|x^{-2/3}\right| \to \infty$. Hence, it cannot be bounded from above by any finite constant L. In other words, $f(x)$ is not Lipschitz continuous in D because its growth rate at $x = 0$ is too large. However, if we remove $x = 0$ and consider, e.g., the domain

$$D = \left[\frac{1}{10}, 1\right] \tag{6}$$

then $f(x)$ is Lipschitz continuous (actually infinitely-differentiable with continuous derivatives) in D. Finally we notice that $f(x)$ is not Lipschitz continuous in the open interval $D =]0, 1]$. In fact the growth rate of $f(x)$ cannot be bounded by a finite constant L as $x \to 0^+$.

[1]The ODE (1) is called "autonomous" if the right hand side f does not depend explicitly on t.

[2]The notation $x_1 \to x_2^{\pm}$ means that x_1 is approaching x_2 either from the left ("$-$") or from the right ("$+$"). Note that we can equivalently write (3) as

$$\lim_{x_1 \to x_2^+} f(x_1) = \lim_{x_1 \to x_2^-} f(x_1) = f(y_2).$$

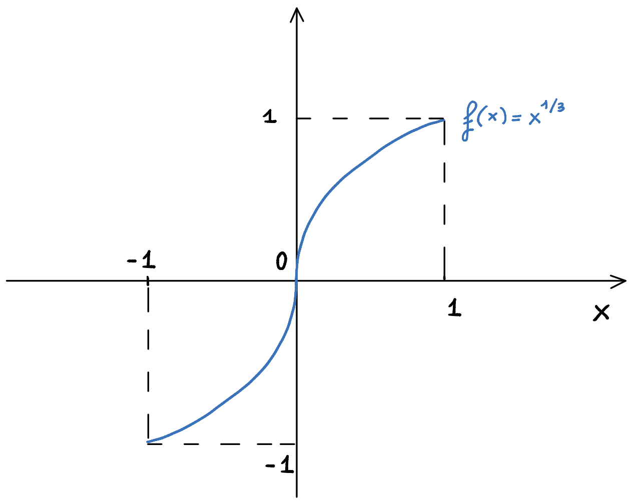


Figure 1: Sketch of the function $f(x) = x^{1/3}$ in $D = [-1, 1]$. The function is continuous in D, but it has an infinite slope at $x = 0$ and therefore it is not Lipschitz continuous in D.

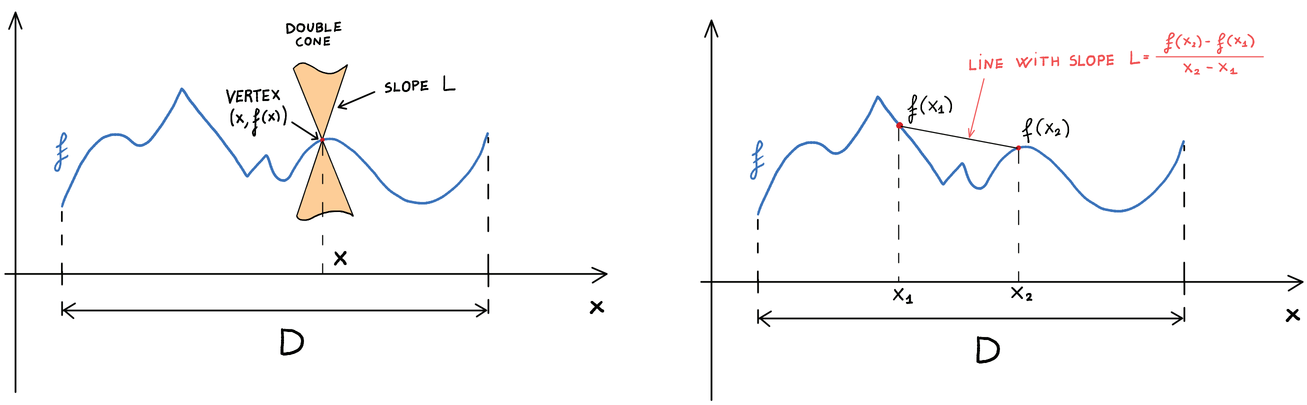


Figure 2: Geometric meaning of Lipschitz continuity.

Geometric meaning of the Lipschitz continuity condition. The Lipschitz continuity condition (2) has a nice geometric interpretation. In practice it says that the function $f(x)$ cannot enter a double cone with slope L and vertex placed on any point of the graph $(x, f(x))$ with $x \in D$. In other words, if we can slide the vertex of the double cone over the graph of the function $f(x)$ for all $x \in D$ and the function never enters the cone then $f(x)$ is Lipschitz continuous in D. To explain this, let us divide the inequality (2) by $|x_1 - x_2|$ (for $x_1 \neq x_2$). This yields

$$\underbrace{\left|\frac{f(x_1) - f(x_2)}{x_1 - x_2}\right|}_{|K|} \leq L \quad \text{for all} \quad x_1, x_2 \in D. \tag{7}$$

As shown in Figure 2, K represents the slope of the line connecting the points $(x_1, f(x_1))$ and $(x_2, f(x_2))$. The "best" Lipshitz constant is obtained as

$$L^* = \max_{x_1, x_2 \in D} \left|\frac{f(x_1) - f(x_2)}{x_1 - x_2}\right|. \tag{8}$$

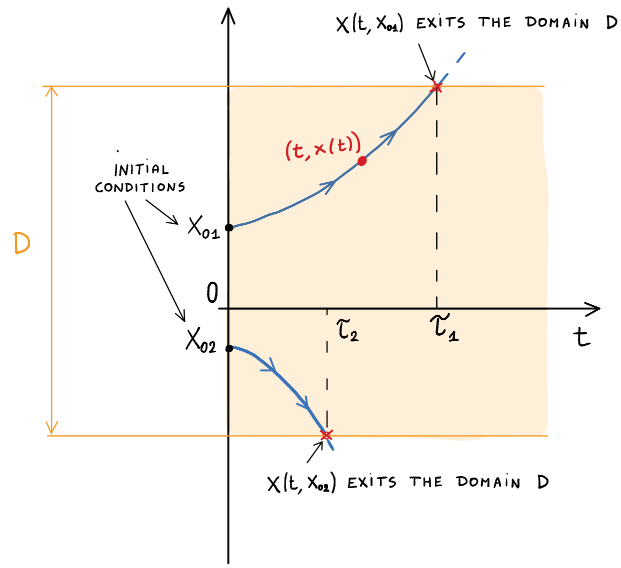


Figure 3: Geometric meaning of the existence and uniqueness theorem for the solution of one ODE.

Any finite number $L \geq L^*$ is still a Lipschitz constant, though not the best one. If the function $f(x)$ is continuously differentiable on a closed set $D \subset \mathbb{R}$ then

$$L^* = \max_{x \in D} \left| \frac{df(x)}{dx} \right| < \infty. \tag{9}$$

Lemma 1. If $f(x)$ is continuously differentiable on a closed set $D \subseteq \mathbb{R}$ then $f(x)$ is Lipschitz continuous in D.

Proof. By assumption the derivative of $df(x)/dx$ is continuous in the closed set $D \subseteq \mathbb{R}$. This implies that the minimum and the maximum of $df(x)/dx$ are attained at some points in D (Extreme Value Theorem). By using the mean value theorem we immediately see that

$$|f(x_1) - f(x_2)| = \left| \frac{df(x^*)}{dx} \right| |x_1 - x_2| \,. \tag{10}$$

where x^* is some point within the interval $[x_1, x_2] \subset D$. The point x^* depends on f, x_1 and x_2. The right hand side of (10) can be bounded as

$$|f(x_1) - f(x_2)| \leq \underbrace{\max_{x \in D} \left| \frac{df(x)}{dx} \right|}_{L^*} |x_1 - x_2| \qquad \text{for all} \quad y_1, y_2 \in D. \tag{11}$$

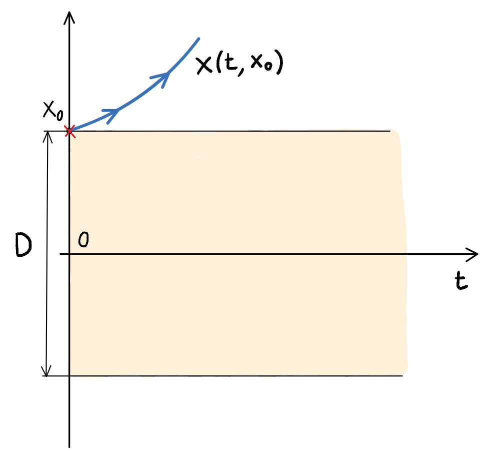
□

Example: The function $f(x) = x^2$ is of class C^∞ (infinitely differentiable with continuous derivatives) in any closed set $D \subset \mathbb{R}$. However, the function $f(x) = x^2$ is not Lipschitz continuous at $x = \pm\infty$, since the slope of the first-order derivative $f'(x) = 2x$ grows unboundedly as $x \to \pm\infty$.

Well-posedness of the initial value problem. Next, we formulate the existence and uniqueness theorem for the solution of the first-order ODE (1).

Theorem 1 (Existence and uniqueness of the solution to (1))**.** Let $D \subset \mathbb{R}$ be an open set, $x_0 \in D$. If $f : D \to \mathbb{R}$ is Lipschitz continuous in D then there exists a unique solution to the initial value problem (1) within the time interval $[0, \tau[$, where τ is the instant at which $x(t)$ exits[3] the domain D (see Figure 3). The solution $x(t)$ is continuously differentiable in $[0, \tau[$.

Remark: In Theorem 1, we required that D is an open set so that we can have solutions in D at least for some $t \in [0, \tau[$. On the other hand, if D is closed then we can pick x_0 right at the boundary of D so that the solution[4] $x(t) = X(t, x_0)$ never enters D, which is the region in which f is assumed to be Lipschitz continuous. In this case, the "exit time" τ may be zero, and Theorem 1 does not provide any information on the existence and uniqueness of the solution.



Global solutions. If $f(x)$ is Lipschitz continuous on the entire real line $\mathbb{R}$ then the solution to the initial value problem (1) is *global*. This means that the solution exists and is unique for all $t \geq 0$. In fact, $x(t)$ never exits the domain in which $f(x)$ is Lipschitz continuous, and therefore we can extend τ in Theorem 1 to infinity. It is important to emphasize that existence and uniqueness of the solution to (1) has nothing to do with the smoothness of $f(x)$ but rather with the rate at which $f(x)$ grows or decays.

Computing the solution of one-dimensional autonomous ODEs. The initial value problem (1) is separable, i.e., it can be equivalently written in an integral form as

$$\boxed{\int_{x_0}^{x(t)} \frac{dx}{f(x)} = t} \tag{12}$$

Hence, if we know how to compute the primitive of $1/f(x)$, i.e., the integral at the left hand side of (12), then we have an algebraic equation that relates $x(t)$, x_0 and t. This does not mean that we can always easily write $x(t)$ explicitly in terms of x_0 and t. This is demonstrated in the following simple example.

Example: Consider the initial value problem (1) and set

$$f(x) = \frac{1}{x^4 - x^2 + 1} \quad \text{and} \quad x_0 = 0. \tag{13}$$

As it is seen in Figure 4, $f(x)$ continuously differentiable in $\mathbb{R}$ with bounded derivative.

[3]As shown in Figure 3, the "exit time" τ depends on D $f(x)$ and x_0.

[4]The nonlinear map $X(t, x_0)$ represents the solution of (1) corresponding to the initial condition x_0, where x_0 is left unspecified. As we shall see hereafter $X(t, x_0)$ is called *flow* generated by the dynamical system (1).

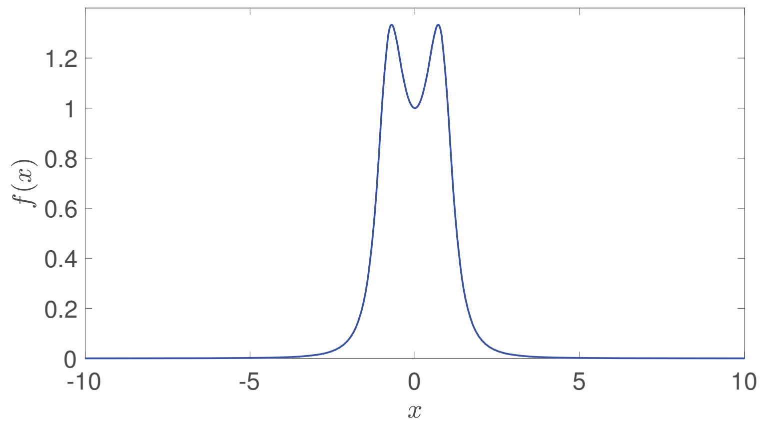


Figure 4: Plot of the function defined in equation (13).

Therefore, the solution of the initial value problem (1), with f and x_0 as in (13) is global, meaning that it exists and it is unique for all $t \geq 0$. A substitution of (13) into the integral equation (12) yields

$$\frac{x(t)^5}{5} - \frac{x(t)^3}{3} + x(t) = t. \tag{14}$$

Hence, to express $x(t)$ as a function of t we need to compute the roots of the fifth-order polynomial (14) as a function of t and among them select the one that passes through $x(0) = 0$.

Example: Consider the initial value problem

$$\begin{cases} \dfrac{dx}{dt} = \sin(x) \\ x(0) = x_0 \end{cases} \tag{15}$$

where x_0 is any number in the interval $D = [0, \pi]$. The solution to (15) can be obtained by computing the integral[5]

$$\int_{x_0}^{x(t)} \frac{dx}{\sin(x)} = t \quad \Rightarrow \quad \left[\log\left(\left|\tan\left(\frac{x}{2}\right)\right|\right)\right]_{x_0}^{x(t)} = t \tag{16}$$

By using the properties of the logarithm we obtain

$$\log\left|\frac{\tan\left(\dfrac{x(t)}{2}\right)}{\tan\left(\dfrac{x_0}{2}\right)}\right| = t \qquad \Rightarrow \qquad x(t) = 2\arctan\left(e^t \tan\left(\frac{x_0}{2}\right)\right). \tag{17}$$

Note that

$$\lim_{t\to\infty} x(t) = \pi \tag{18}$$

The trajectories of the system (15) are shown in Figure 8.

Hereafter we provide an example of an initial value problem the solution of which blows-up in a finite time, and an example of an initial value problem that has an infinite number of solutions.

[5]Recall that the primitive of $1/\sin(x)$ is

$$\log\left(\left|\tan\left(\frac{x}{2}\right)\right|\right).$$

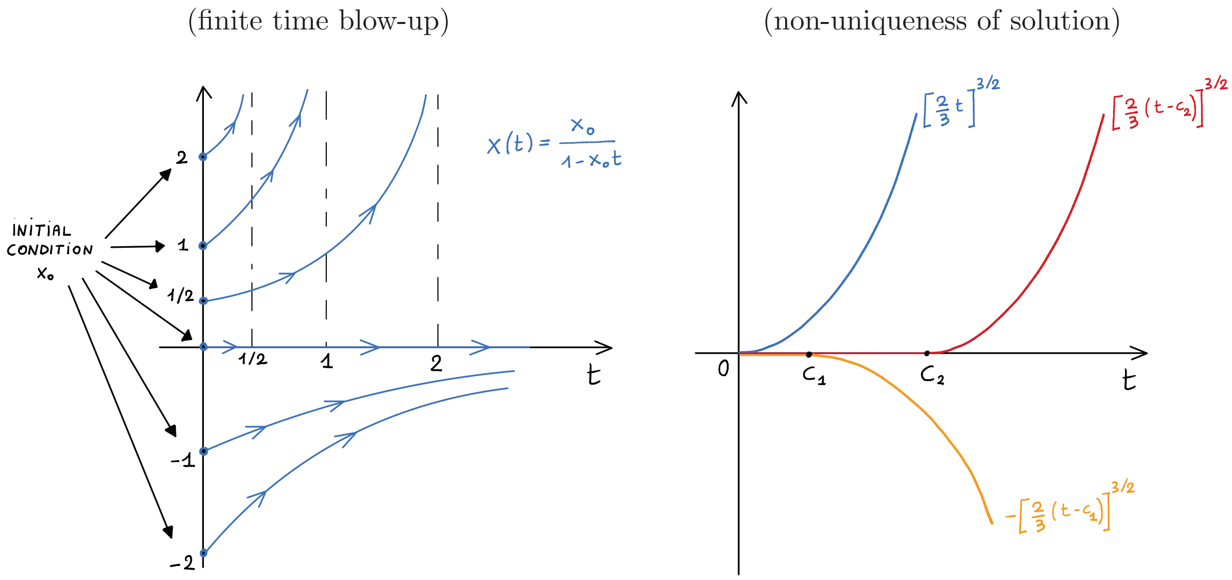


Figure 5: (left) Solutions of the initial value problem (19) for different initial conditions x_0. It is seen that for $x_0 > 0$ the solution blows up at the fine time $t^* = 1/x_0$. On the other hand, if $x_0 \leq 0$ the solution exists and is unique for all $t \geq 0$. (b) Solutions of the initial value problem (21) corresponding to the initial condition $x_0 = 0$. This problem has an infinite number of solutions.

- *Finite-time blow-up*: Consider the initial value problem

$$\frac{dx}{dt} = x^2 \qquad x(0) = x_0. \tag{19}$$

We know that $f(x) = x^2$ is not Lipschitz continuous at infinity. By using separation of variables, i.e., equation (12), it is straightforward to show that

$$\int_{x_0}^{x(t)} \frac{dx}{x^2} = -\frac{1}{x(t)} + \frac{1}{x_0} = t \quad \Rightarrow \quad x(t) = \frac{x_0}{1 - x_0 t}. \tag{20}$$

The function $x(t)$ clearly blows up to infinity as t approaches $1/x_0$ (from the left) for positive initial conditions x_0. On the other hand, if $x_0 \leq 0$ the solution exists and is unique for all $t \geq 0$.

- *Non-uniqueness of solutions:* Consider the initial value problem

$$\frac{dx}{dt} = x^{1/3} \qquad x(0) = 0. \tag{21}$$

We have seen that $f(x) = x^{1/3}$ is not Lipschitz continuous in any domain D that includes the point the point $x = 0$. Note that we are setting the initial condition exactly at the point in which the slope of $f(x)$ is infinity (see Figure 1). By using separation of variables it straightforward to show that a solution to (21) is

$$x(t) = \left(\frac{2}{3}t\right)^{3/2}. \tag{22}$$

However, note that the functions

$$x(t) = \begin{cases} 0 & \text{for } 0 \leq t < c \\ \pm\left(\frac{2}{3}(t-c)\right)^{3/2} & \text{for } t \geq c \end{cases} \tag{23}$$

are also solutions to (21) for all $c \geq 0$.

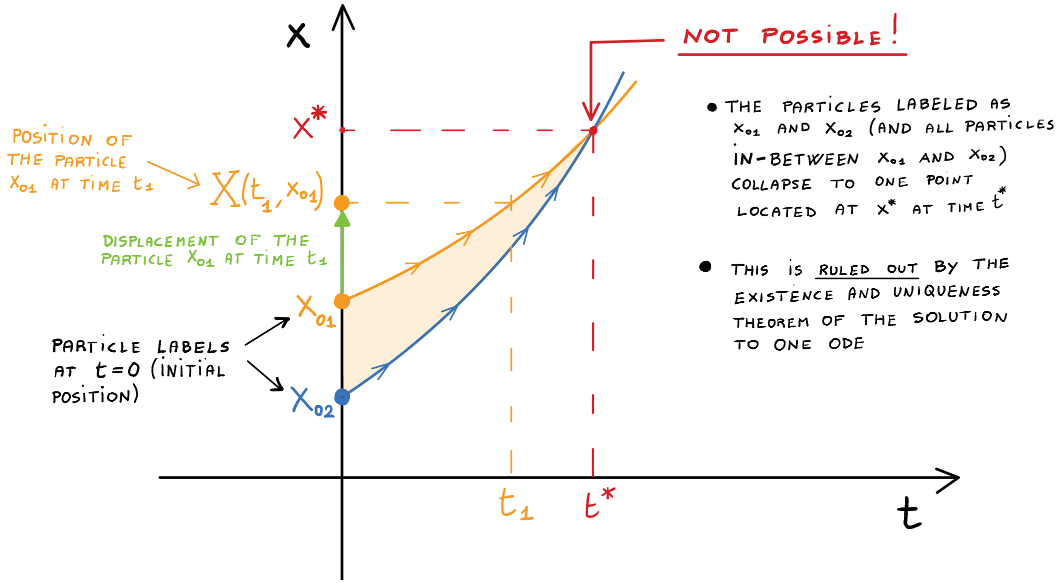


Figure 6: Trajectories corresponding to different initial conditions cannot intersect.

One-dimensional flows. We have seen that the initial value problem (1) admits a unique solution $x(t)$ (continuously differentiable in t) if $f(x)$ is Lipschitz continuous on an open subset $D \subset \mathbb{R}$ (Theorem 1), and if x_0 is chosen in D. This means that the solution $x(t)$ depends on $f(x)$ and x_0. We will denote the dependence of $x(t)$ on x_0 as $X(t, x_0)$, i.e.,

$$x(t) = X(t, x_0). \tag{24}$$

Let us first notice that because of the existence and uniqueness Theorem 1, it is not possible for two solutions corresponding to two different initial conditions to intersect at any finite time t (see Figure 6). This implies that $X(t, x_0)$ is invertible at each finite time[6] (see below), i.e., we can always identify which "particle" x_0 sits at location $x(t) = X(t, x_0)$ at time t. Moreover, it is impossible for two "particles" x_{01} and x_{02} to collide at any finite time, or for one particle to split into two or more particles (Figure 6). Next, we characterize how the flow $X(t, x_0)$ depends on the initial condition x_0 at each fixed time t.

Theorem 2 (Regularity of the ODE solution with respect to x_0)**.** Let $D \subset \mathbb{R}$ be an open set, $x_0 \in D$. If $f : D \to \mathbb{R}$ is Lipschitz continuous in D then the solution of the initial value problem (1), i.e., $X(t, x_0)$ (i.e., the flow generated by the ODE) is continuous in x_0. Moreover, if $f(x)$ is of class C^k in D (continuously differentiable k-times in D with continuous derivative), then $X(t, x_0)$ is of class C^k in D.

In summary, Theorem 2 states that the smoother $f(x)$, the smoother the dependency of $X(t, x_0)$ on x_0. The two-dimensional function $X(t, x_0)$ is called *flow generated by the dynamical system* (1), and it represents the full set of solutions to (1) for every initial condition x_0.

Theorem 3 (Regularity of the ODE solution in time t)**.** Let $D \subset \mathbb{R}$ be an open set, $x_0 \in D$. if $f(x)$ is of class C^k in D (continuously differentiable k-times in D with continuous derivative), then $X(t, x_0)$ is of class C^{k+1} in time.

The continuity of higher-order derivatives, and its link to the the regularity of $f(x)$ can be established by differentiating the ODE

$$\frac{dX(t, x_0)}{dt} = f(X(t, x_0)) \tag{25}$$

[6] Solutions corresponding to different initial conditions can, however, intersect at $t = \infty$, e.g., when there exist an attracting set such as a stable equilibrium point (proof below).

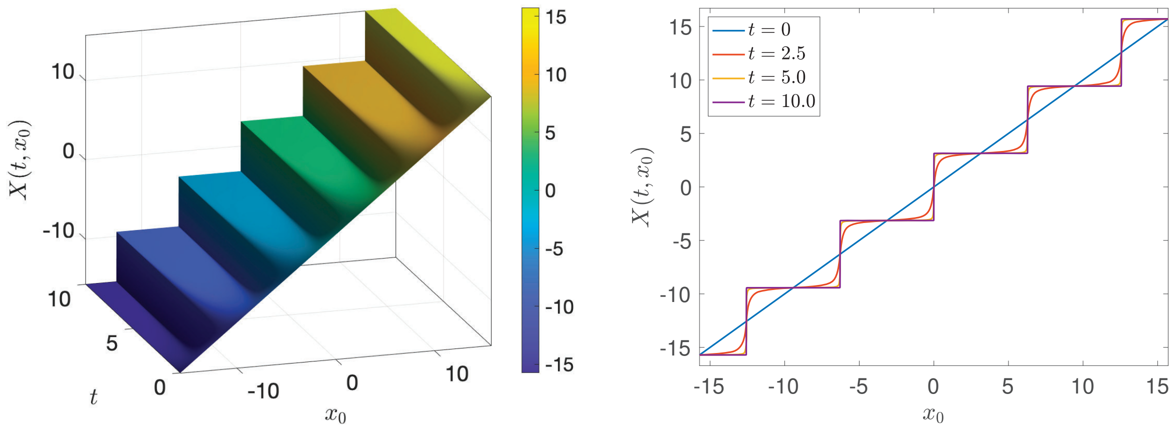


Figure 7: Visualization of the flow generated by the ODE (28).

with respect to time. For instance, we have

$$\frac{d^2X(t,x_0)}{dt^2} = f'(X(t,x_0))f(X(t,x_0)), \tag{26}$$

$$\frac{d^3X(t,x_0)}{dt^3} = f''(X(t,x_0))f^2(X(t,x_0)) + [f'(X(t,x_0))]^2 f(X(t,x_0)). \tag{27}$$

At this point we can use the existence and uniqueness theorem for the solution of higher-dimensional dynamical systems to conclude that the derivatives $d^nX(t,x_0)/dt^n$ are continuous if $d^{n-1}f(x)/dx^{n-1}$ is continuous.

Example: In Figure 7 we visualize the flow generated by the ODE

$$\frac{dx}{dt} = \sin(x) \tag{28}$$

for all $x_0 \in [-5\pi, 5\pi]$ and $t \in [0, 10]$. Such flow was computed by solving the ODE (28) numerically (see Appendix A) for a large number of initial conditions x_0. Similarly, in Figure 8 we plot the trajectories of the system (28) corresponding to an evenly-spaced grid of 100 initial conditions in $[-5\pi, 5\pi]$.

Properties of the flow. The flow generated by one dimensional dynamical systems of the form (1) satisfies the following properties:

- $X(0,x_0) = x_0$. This means that at $t = 0$ the mapping $X(t,x_0)$ is the identity.
- $X(t,x_0)$ is monotonic in x_0 for each fixed t, i.e.,

$$X(t,x_{02}) > X(t,x_{01}) \quad \text{for all } x_{02} > x_{01}. \tag{29}$$

  This property can be proved easily by substituting $x(t) = X(t,x_0)$ into the ODE $dx/dt = f(x)$ and differentiating it with respect to x_0. This yields

$$\frac{\partial}{\partial x_0}\left(\frac{dX(t,x_0)}{dt}\right) = \frac{\partial f(X(t,x_0))}{\partial x_0} \quad \Rightarrow \quad \frac{d}{dt}\left(\frac{\partial X(t,x_0)}{\partial x_0}\right) = f'(X(t,x_0))\frac{\partial X(t,x_0)}{\partial x_0}. \tag{30}$$

  The last one ODE is linear and that can be easily integrated in time from the initial condition

$$\frac{\partial X(0,x_0)}{\partial x_0} = 1 \tag{31}$$

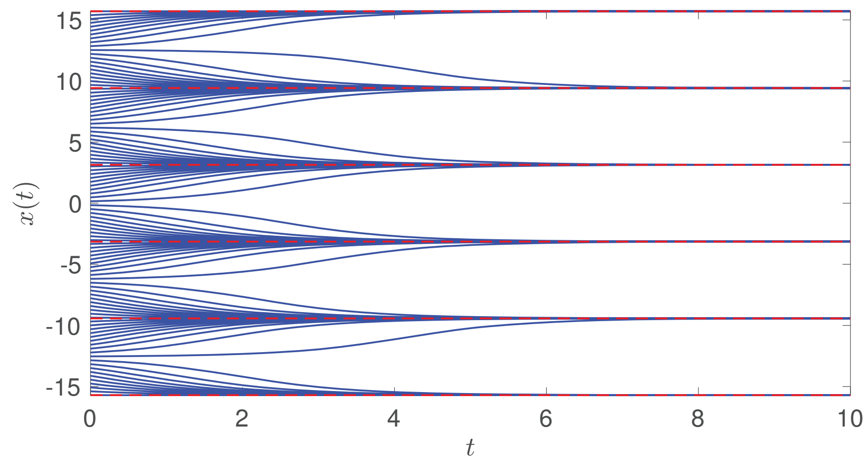


Figure 8: Trajectories of the dynamical system (28) corresponding to 100 evenly spaced initial conditions in $[-5\pi, 5\pi]$. All trajectories are computed numerically. The red dashed lines represent the stable fixed points (equilibria) of the system.

to obtain[7]

$$\frac{\partial X(t, x_0)}{\partial x_0} = \exp\left[\int_0^t f'(X(\tau, x_0))d\tau\right]. \tag{34}$$

The right hand side of (34) is strictly positive for each $t \geq 0$, which implies

$$\frac{\partial X(t, x_0)}{\partial x_0} > 0 \qquad \text{for each finite } t \geq 0. \tag{35}$$

This proves that the flow map $X(t, x_0)$ is monotonic in x_0 and therefore invertible for each finite t.

- $X(t, x_0)$ satisfies the composition rule $X(t+s, x_0) = X(t, X(s, x_0)) = X(s, x(t, x_0))$. This property is called "semi-group property" of the flow and it follows from the fact that we can restart integration of the ODE (1) at time t (or time s) from the new initial condition $X(t, x_0)$ (or $X(s, x_0)$) to get to the final integration time $s + t$. Again, this property holds because of the existence and uniqueness theorem 1.

Inverse flows. The monotonicity property (29) guarantees that the flow map is *invertible* for each finite $t \geq 0$. In other words, it is always possible to determine which x_0 sits at a certain location x at time t. As we mentioned above, this also means that it is impossible to have simultaneous occupation of one location x by more than one "particle" x_0, i.e., the trajectories of the (1) corresponding to two different initial conditions cannot intersect (see Figure 6). The invertibility of $X(t, x_0)$ in x_0 for each fixed t allows us to define the inverse flow, which gives the label x_0 of the particle that sits at x at time t. In practice, the inverse flow can be computed by integrating the (1) from the initial condition x (at time t) backwards

[7]Equation (34) characterizes the dynamics of an infinitesimal "line element" with length dx_0 as it is "transported" by the flow $X(t, x_0)$. In fact, from (34) it follows that

$$dX(t, x_0) = dx_0 \exp\left[\int_0^t f'(x(\tau, x_0))d\tau\right]. \tag{32}$$

Moreover, if x_0 is a fixed-point, i.e. if $X(\tau, x_0) = x_0$, then

$$dX(t, x_0) = dx_0 e^{tf'(x_0)}. \tag{33}$$

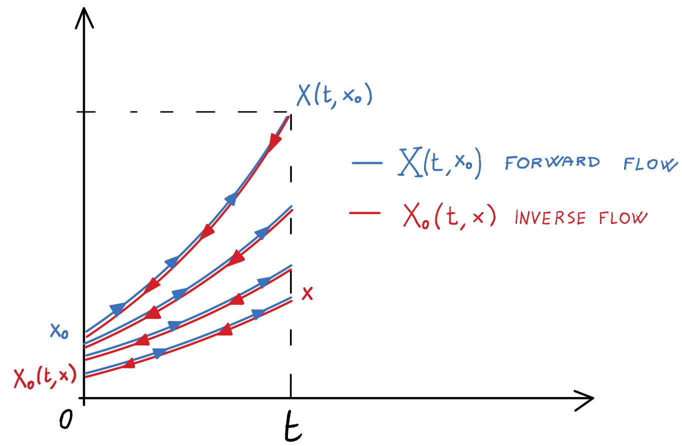


Figure 9: Illustration of forward and inverse flows.

in time to $t = 0$. Integrating (1) backwards in time is equivalent to integrating forward in time the ODE system with reversed velocity vector (see Figure 9)

$$\begin{cases} \frac{dx}{dt} = -f(x) \\ x(0) = x \end{cases} \tag{36}$$

The flow associated with this system will be denoted as $X_0(t, x)$. Clearly, for each fixed t the inverse flow $X_0(t, x)$ is the inverse of the forward flow $X(t, x_0)$, i.e.,

$$X(t, X_0(t, x)) = x, \qquad X_0(t, X(t, x_0)) = x_0. \tag{37}$$

Flow map equation. It can be shown that the flow $X(t, x_0)$ generated by the initial value problem (1) is governed by the first-order partial differential equation (PDE)

$$\begin{cases} \frac{\partial X(t,x_0)}{\partial t} - f(x_0)\frac{\partial X(t,x_0)}{\partial x_0} = 0 \\ X(0, x_0) = x_0 \end{cases} \tag{38}$$

This can be verified, e.g., by substituting the flow

$$X(t, x_0) = \frac{x_0}{1 - x_0 t} \tag{39}$$

generated by (19) into (38). Indeed, computing the derivatives

$$\frac{\partial X(t,x_0)}{\partial t} = \frac{x_0^2}{(1-x_0t)^2}, \qquad \frac{\partial X(t,x_0)}{\partial x_0} = \frac{1}{(1-x_0t)^2}. \tag{40}$$

and recalling that $f(x_0) = x_0^2$ in this case, we see that (38) is identically satisfied. Equation (9) is a hyperbolic PDE that can be solved numerically, e.g., with the method of characteristics, finite differences, or spectral methods, to obtain the flow map. The solution to (38) can be formally expressed in terms of an exponential operator known as *Koopman operator*. To this end, we first define the linear (differential) operator

$$K(x_0) = f(x_0)\frac{\partial}{\partial x_0}, \tag{41}$$

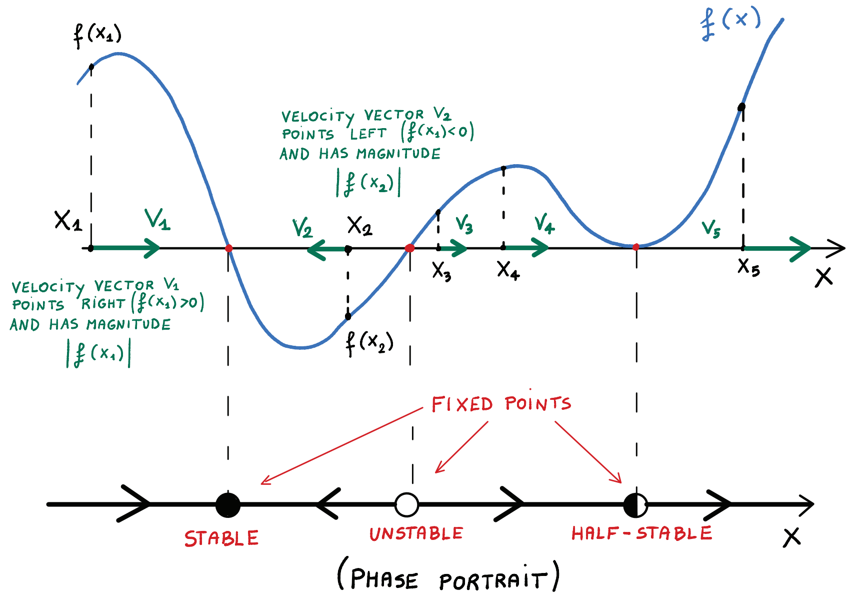


Figure 10: Vector field associated with $f(x)$, fixed points, and phase portrait.

which is known as "generator" of the Koopman operator. This allows us to write (38) as

$$\frac{\partial X(t,x_0)}{\partial t} = K(x_0)X(t,x_0), \tag{42}$$

and therefore obtain the formal solution

$$X(t,x_0) = e^{tK(x_0)}x_0, \tag{43}$$

where $e^{tK(x_0)}$ is the Koopman operator. In this form, it is immediate to prove the semi-group property of the flow discussed previously. In fact,

$$X(t+s,x_0) = e^{(t+s)K(x_0)}x_0 = e^{tK(x_0)}e^{sK(x_0)}x_0 = e^{tK(x_0)}X(s,x_0) = X(t,X(s,x_0)). \tag{44}$$

Similarly, the inverse flow $X_0(t,x)$ defined by the dynamical system (36) is governed by the PDE

$$\begin{cases} \dfrac{\partial X_0(t,x)}{\partial t} + f(x)\dfrac{\partial X_0(t,x)}{\partial x} = 0 \\ X_0(0,x) = x \end{cases} \tag{45}$$

The solution to this PDE is

$$X_0(t,x) = e^{-tK(x_0)}x. \tag{46}$$

Geometric approach. We have seen that dynamical systems of the form (1) generate a flow $X(t,x_0)$ that maps every initial condition x_0 to the solution of the ODE at time t. If we think as x_0 as the initial position of a particle sitting on a line (phase space), then from elementary mechanics we know that $dX(0,x_0)/dt = f(x_0)$ represents the velocity of such particle. Hence, given $f(x)$ we can immediately plot the *vector field* [8] associated with the dynamical system, which represents how fast a particle at any particular location x moves left or right. Clearly, if the velocity vector $f(x)$ is equal to zero at some

[8] A vector field is a vector that is continuously indexed by one or more variables. For one-dimensional dynamical systems the vector field $f(x)$ is indexed by coordinate x.

locations x^* then any particle placed at that location won't move at all as time increases. These points are called *fixed points* (or *equilibria*) of the dynamical system (1). Fixed points can be rigorously defined as the points $x^* \in \mathbb{R}$ such that for all $t \geq 0$

$$X(t, x^*) = x^*. \tag{47}$$

By differentiating the previous equation with respect to time yields

$$0 = \frac{\partial X(t, x^*)}{\partial t} = f(X(t, x^*)) = f(x^*). \tag{48}$$

Therefore, the fixed points of the system (1) are *zeros* of the nonlinear function $f(x)$, i.e.

$$f(x^*) = 0, \tag{49}$$

(see Figure 10). The calculation of the fixed points can be done analytically for prototype dynamical systems. In general, computing the fixed points requires a root-finding numerical algorithm such as the Newton's method.

Distribution of fixed points. The (Lipschitz) continuity condition on $f(x)$ in Theorem 1 imposes topological constraints on the distribution of fixed points. Specifically, fixed points facing each other cannot be both stable or unstable, but rather they must have opposite stability properties (Figure 10).

Stability analysis of fixed points. A fixed point x^* of the dynamical system (1) is said to be *asymptotically stable* if

$$\lim_{t \to \infty} |X(t, x_0) - x^*| = 0 \tag{50}$$

for all x_0 in some neighborhood of x^*. In other words, stable fixed points attract trajectories of the dynamical system from both left and right (see Figure 10). Of course, by plotting $f(x)$ we can immediately infer the stability properties of all fixed points. This can be also done analytically by a technique known as *linearization*. The basic idea is very simple. If $f(x)$ is smooth (at least continuously differentiable) then the more we "zoom-in" at a fixed point x^* the more $f(x)$ looks linear in a neighborhood of x^*, and therefore it can be replaced by its fist-order term in a Taylor series expansion. In other words, by "zooming-in" we are studying the local dynamics of the system nearby the fixed point. To this end, let us pick an initial condition x_0 that is sufficiently close to to the fixed point x^*, say $x_0 - x^* = 10^{-10}$. By continuity, the flow $X(t, x_0)$ will map x_0 to a position that is still close to x^* at least for some time (see Figure 11). The distance between $X(t, x_0)$ and the fixed point x^* can be expressed as function[9]

$$\eta(t, x_0) = X(t, x_0) - x^* \qquad \Leftrightarrow \qquad X(t, x_0) = \eta(t, x_0) + x^*. \tag{52}$$

A substitution of $X(t, x_0) = \eta(t, x_0) + x^*$ into (1) yields

$$\begin{cases} \dfrac{d\eta}{dt} = f(\eta + x^*) \\ \eta(0, x_0) = x_0 - x^* \end{cases} \tag{53}$$

If $\eta(0, x_0)$ is very small then $\eta(t, x_0)$ is very small too (at least for some time). This allows us expand $f(\eta + x^*)$ in a Taylor series as

$$f(\eta(t, x_0) + x^*) = \underbrace{f(x^*)}_{=0} + f'(x^*)\eta(x, t) + \frac{1}{2} f''(x^*)\eta(x, t)^2 + \cdots \tag{54}$$

[9]Note that

$$\eta(0, x_0) = X(0, x_0) - x^* = x_0 - x^*. \tag{51}$$

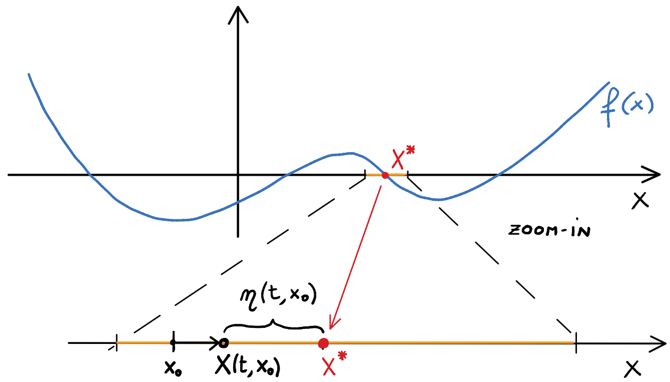


Figure 11: Linearization nearby the fixed point x^*.

Hence, to first-order in η we obtain the linear initial value problem

$$\begin{cases} \dfrac{d\eta}{dt} = f'(x)\eta \\ \eta(0, x_0) = x_0 - x^* \end{cases} \tag{55}$$

The solution of (55) is

$$\eta(t, x_0) = (x - x_0)e^{f'(x^*)t}. \tag{56}$$

The last equation allows us to conclude that

- $f'(x^*) < 0 \quad \Rightarrow \quad x^*$ is asymptotically stable
- $f'(x^*) > 0 \quad \Rightarrow \quad x^*$ is unstable
- $f'(x^*) = 0 \quad \Rightarrow \quad$ results of linear stability analysis are inconclusive.

If $f'(x_0) = 0$ then need to expand f to higher order in η, and solve a nonlinear ODE to classify the stability of the fixed point x^*.

Example: The dynamical system

$$\frac{dx}{dt} = \underbrace{x^2 - 1}_{f(x)} \tag{57}$$

has two fixed points located at $x^*_{1,2} = \pm 1$. Of course, $f'(x) = 2x$. By evaluating $f'(x)$ at the fixed points we see that $f'(1) = 2 > 0$ and $f'(-1) = -2 < 0$. Hence $x_1^* = 1$ is unstable, and $x_2^* = -1$ is asymptotically stable.

Example: The dynamical system

$$\frac{dx}{dt} = 1 + \sin(x) \tag{58}$$

has a global solution for all initial conditions x_0, and an infinite number of fixed points located at (see Figure 12)

$$x_k^* = \frac{3\pi}{2} + 2k\pi. \tag{59}$$

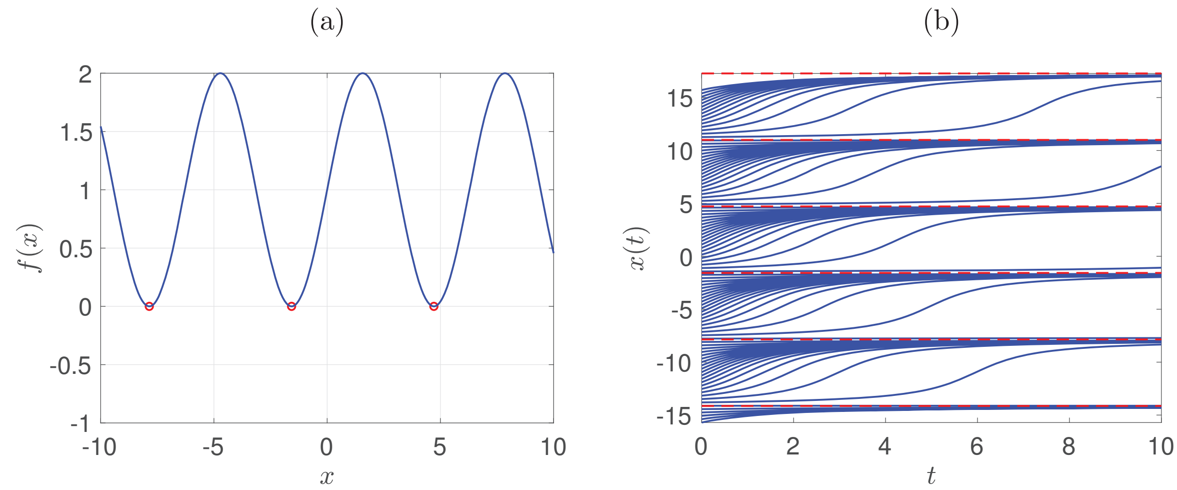


Figure 12: (a) Graph of function the $f(x) = 1 + \sin(x)$ and some of its fixed points (red circles). (b) Trajectories of the dynamical system (58).

By expanding $f(x) = 1 + \sin(x)$ in a Taylor series at $x_0^* = 3\pi/2$ we obtain

$$\begin{aligned}\sin\left(\eta + \frac{3\pi}{2}\right) &= \sin\left(\frac{3\pi}{2}\right) + \cos\left(\frac{3\pi}{2}\right)\eta - \frac{1}{2}\sin\left(\frac{3\pi}{2}\right)\eta^2 + \cdots \\ &= -1 + \frac{\eta^2}{2} + \cdots,\end{aligned} \tag{60}$$

i.e.,

$$1 + \sin\left(\eta + \frac{3\pi}{2}\right) = \frac{\eta^2}{2} + \cdots. \tag{61}$$

Substituting this back into (53) yields the nonlinear system

$$\begin{cases} \dfrac{d\eta}{dt} = \dfrac{\eta^2}{2} \\ \eta(0, x_0) = x_0 - \dfrac{3\pi}{2} \end{cases} \tag{62}$$

We computed the analytical solution to this system before (see Eq. (20)),

$$\eta(t, x_0) = \frac{\left(x_0 - \dfrac{3\pi}{2}\right)}{1 - \left(x_0 - \dfrac{3\pi}{2}\right)\dfrac{t}{2}}. \tag{63}$$

Clearly, if $x_0 > 3\pi/2$ then trajectory tends to go further away from the fixed point $x_0^* = 3\pi/2$. On the other hand, if $x_0 < 3\pi/2$ then the trajectories are attracted to $x_0^* = 3\pi/2$. Note that the second-order polynomial approximation of the system (58) at the fixed point $x_0^* = 3\pi/2$ we just considered seems to blow-up in a finite time for $x_0 > 3\pi/2$, while the trajectories plotted in Figure 12 exist and are unique for all times. This is due to the fact that we did not include a sufficient number of terms in the Taylor expansion, some of which become increasingly important in stabilizing the polynomial approximation of the dynamical system as η becomes larger.

Example: The dynamical system

$$\frac{dx}{dt} = \underbrace{-x^3}_{f(x)} \tag{64}$$

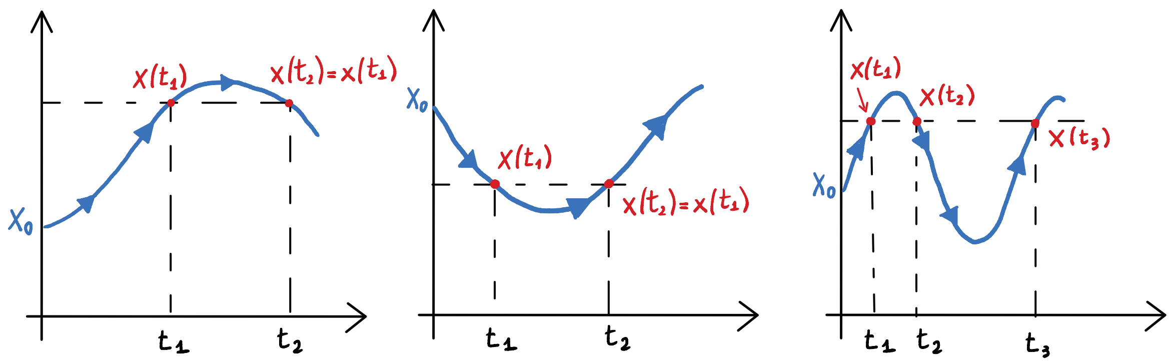


Figure 13: These trajectories are impossible for one-dimensional dynamical systems of the form (1).

has a fixed point at $x^* = 0$. Linear stability analysis in this case is ineffective at inferring stability. In fact $f'(x) = -3x^2$, which is equal to zero at $x^* = 0$. The analytical solution to (64) is obtained as

$$\int_{x_0}^{x(t)} \frac{dx}{x^3} = -t \quad \Rightarrow \quad \frac{1}{2}\left(\frac{1}{x(t)^2} - \frac{1}{x_0^2}\right) = t. \tag{65}$$

Therefore,

$$X(t, x_0) = \text{sign}(x_0)\sqrt{\frac{x_0^2}{1 + 2x_0^2 t}}, \tag{66}$$

which shows that $x^* = 0$ is a globally attracting fixed point. This means that $x^* = 0$ attracts all trajectories generated by the ODE (64) independently of the initial condition x_0.

Lyapunov functions (potentials). Lyapunov functions are used to make conclusions about trajectories of a system (1) without finding the trajectories (i.e., solving the differential equation). A typical Lyapunov theorem has the form: "if there exists a function $V(x)$ that satisfies some conditions on $V(x)$ and $dV(x(t))/dt$, then the trajectories of the system satisfy some property". A Lyapunov function V can be thought of as generalized potential for a system.

- *Asymptotic stability of fixed points:* If there exists a smooth function $V(x)$ in a neighborhood of the fixed point x^* satisfying

  a) $V(x)$ has a local minimum at x^*,

  b) $V(x)$ does not increase along trajectories of (1), i.e., $dV(x(t))/dt < 0$, in a neighborhood of x^*, (excluding x^*).

  Then x^* is is an asymptotically stable fixed point. The proof is very simple. Suppose that $x(t_1)$ is in a neighborhood of x^* then

  $$V(x(t_2)) = V(x(t_1)) + \int_{t_1}^{t_2} \frac{dV(x(\tau))}{d\tau} d\tau < V(x(t_1)) \quad \text{for all } t_2 \geq t_1 \tag{67}$$

  Hence $x(t)$ converges monotonically to the local minimum of V located at x^* as time increases, implying that x^* is asymptotically stable.

- *Impossibility of trajectory reversals:* If there exists a smooth function $V(x)$ satisfying $dV(x(t))/dt < 0$ then there cannot be any maxima or minima of $x(t)$ at any finite time t. In particular, this rules out trajectories of the form shown in Figure 13. The proof follows immediately from (67). In fact, for any trajectory reversal there exist t_1 and t_2 such that $x(t_2) = x(t_1)$ (see Figure 13). Hence,

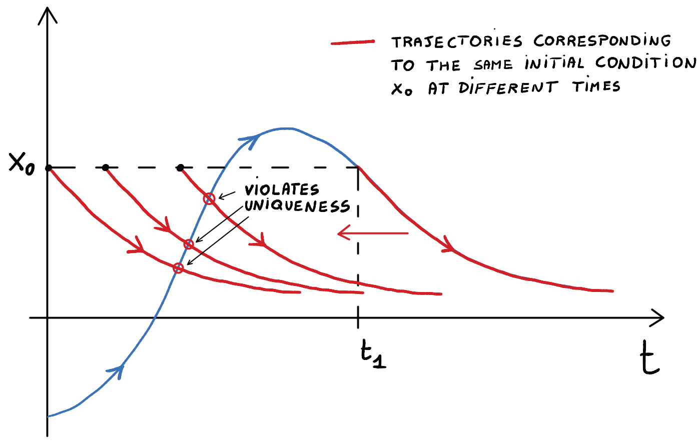


Figure 14: An autonomous dynamical generates trajectories that depend only on x_0 and f. This means that we can translate a trajectory left and right to obtain other solutions of the same system. This translational symmetry, together with the existence and uniqueness theorem 1, rules out the possibility of trajectory reversals, e.g., the blue trajectory.

$V(x(t_2)) = V(x(t_1))$ which immediately contradicts (67). Note in fact, that $dV(x(t))/dt$ is not zero and does not change sign in $[t_1, t_2]$).

How do we construct a function $V(x)$ with the properties stated above? For one-dimensional systems it is sufficient to consider primitive of $-f(x)$, i.e.,

$$\frac{dV(x)}{dx} = -f(x). \tag{68}$$

In fact,

$$\frac{dV(x(t))}{dt} = -\frac{dV(x(t))}{dx}\frac{dx(t)}{dt} = -f(x(t))^2 \leq 0. \tag{69}$$

The equality sign holds only at fixed points, which are indeed the only points where $dx(t)/dt = 0$. Note that this rules out the possibility of trajectories of the form shown in Figure 13. If we interpret $f(x)$ as a vector field in the sense described in Figure 10, then $V(x)$ defined in (68) is called *potential* for $f(x)$. The potential is defined up to an additive constant.

An alternative method to rule out the possibility trajectories reversals such those in Figure 13 relies on the existence and uniqueness Theorem 1. In fact, since the dynamical system (1) is autonomous, it doesn't really matter the time at which we set the initial condition. This implies that we are free to translate the trajectories left and right in the plane $(x(t), t)$, to obtain all possible solutions to the system. However, as show in Figure 14, it is not possible to do so without violating the existence and uniqueness theorem if there exists a trajectory reversal.

Note that this also means that to compute flow of 1D systems we just need a few trajectories which can be then translated left or right as shown in Figure 15 for the system $dx/dt = 1 - x^2$.

Example: Consider the dynamical system (58). A potential for such system is

$$V(x) = V(x_0) - \int_{x_0}^{x} (1 + \sin(y))\, dy = \cos(x) - x + C, \tag{70}$$

where C is a constant. This function is plotted in Figure 16 for $C = 0$. It is seen that $V(x)$ has inflection points at the fixed points suggesting that such fixed points are half-stable.

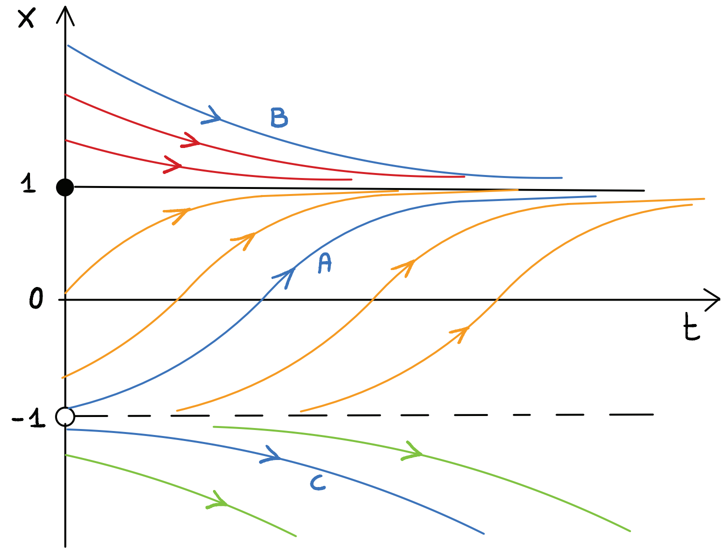


Figure 15: We can use the translational symmetry of the solutions to the autonomous system (1) to construct the entire flow. Specifically, the yellow trajectories are all obtained by translating the trajectory labeled by "A" to the left and to the right. Similarly, the red trajectories are obtained by translating the trajectory labeled as "B" to the left, while the green trajectories are obtained by translating the trajectory labeled by "C" to the left and to the right.

Dynamics of one-dimensional dynamical systems. In summary, the trajectories of a one-dimensional dynamical system

- Can get to a stable (or half-stable) fixed-point in an infinite time,
- Can blow-up to infinity in a finite or an infinite time,
- Cannot have maxima or minima at any finite time (no overshoot/undershoot, no periodic orbits).

The only attracting sets of one-dimensional dynamical systems are fixed points. In higher dimensions we can have attracting sets that are more complicated, e.g., limit cycles, saddle nodes connected by heteroclinic orbits, strange attractors, etc.

Appendix A: Elementary numerical methods for ODEs

The initial value problem (1) can be equivalently written in an integral form as

$$x(t) = x_0 + \int_0^t \frac{dx(s)}{ds} ds = x_0 + \int_0^t f(x(s)) ds \tag{71}$$

i.e., as an integral equation for $x(s)$. This formulation is quite convenient for developing numerical methods for ODEs based on *numerical quadrature formulas*, i.e., numerical approximations of the temporal integral appearing at the right hand side of (71). For example, consider a discretization of the time interval $[0, T]$ in terms of $N + 1$ evenly-spaced time instants

$$t_i = i\Delta t \qquad i = 0, 1, \ldots, N \qquad \text{where} \quad \Delta t = \frac{T}{N}. \tag{72}$$

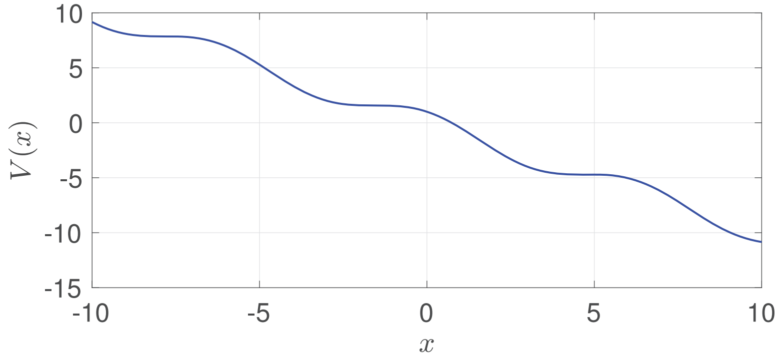
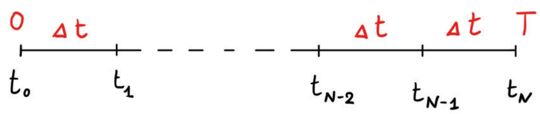


Figure 16: A potential function for the dynamical system (58).



By applying (71) within each time interval $[t_i, t_{i+1}]$ we obtain

$$x(t_{i+1}) = x(t_i) + \int_{t_i}^{t_{i+1}} f(x(s))ds. \tag{73}$$

At this point we can approximate the integral at the right hand side if (73), e.g., by using the simple rectangle rule

$$\int_{t_i}^{t_{i+1}} f(x(s))ds \simeq \Delta t f(x(t_i)) \tag{74}$$

This yields the *Euler forward scheme*

$$u_{i+1} = u_i + \Delta t f(u_i), \tag{75}$$

where u_i is an approximation of $x(t_i)$. The Euler forward scheme is an explicit one-step scheme. The adjective "explicit" emphasizes the fact that u_{i+1} can be computed explicitly based on the knowledge of f and u_i using (75). On the other hand, if we approximate the integral at the right hand side of (71) with the trapezoidal rule

$$\int_{t_i}^{t_{i+1}} f(x(s))ds \simeq \frac{\Delta t}{2}\left[f(x(t_{i+1})) + f(x(t_i))\right] \tag{76}$$

we obtain the *Crank-Nicolson scheme*

$$u_{i+1} = u_i + \frac{\Delta t}{2}\left[f(u_i) + f(u_{i+1})\right]. \tag{77}$$

The Crank-Nicolson scheme is "implicit" because the approximate solution at time t_{i+1}, i.e., u_{i+1}, cannot be computed explicitly based on u_i, but requires the solution of a nonlinear equation. Such a solution can be computed numerically by using any method to solve nonlinear equations. These methods are usually iterative, e.g., the bisection method, or the Newton method if f is continuously differentiable. Iterative methods for nonlinear equations can be formulated as fixed point iteration problems. In the specific case of (77) we have

$$u_{i+1} = G(u_{i+1}) \qquad \text{where} \qquad G(u_{i+1}) = u_i + \frac{\Delta t}{2}\left[f(u_i) + f(u_{i+1})\right]. \tag{78}$$

If Δt is small then u_i is close to u_{i+1}. Moreover, if Δt is sufficiently small we have that the Lipschitz constant of G is smaller than 1, which implies that the fixed point iterations will convergence globally to a unique solution u_{i+1} .

Bifurcations of equilibria in one-dimensional dynamical systems

In previous lecture note we studied one-dimensional dynamical systems of the form

$$\begin{cases} \dfrac{dx}{dt} = f(x) \\ x(0) = x_0 \end{cases} \tag{1}$$

and fully characterized their properties. In particular, we proved that trajectory reversals or oscillations are impossible, and that the dynamics is essentially determined by the location of the fixed points and their stability properties. In this course note we study what happens to the locations of the fixed points if $f(x)$ in (1) depends on a parameter $\mu \in \mathbb{R}$, and we are allowed to vary such parameter μ. To this end, we consider the dynamical system[1]

$$\begin{cases} \dfrac{dx}{dt} = f(x,\mu) \\ x(0) = x_0 \end{cases} \tag{2}$$

For each fixed value of μ we can plot $f(x,\mu)$ versus x, and see if there are any fixed points. Equivalently, we can think of $f(x,\mu)$ as a real-valued function in 2 variables, i.e., a surface in the three-dimensional Euclidean space.

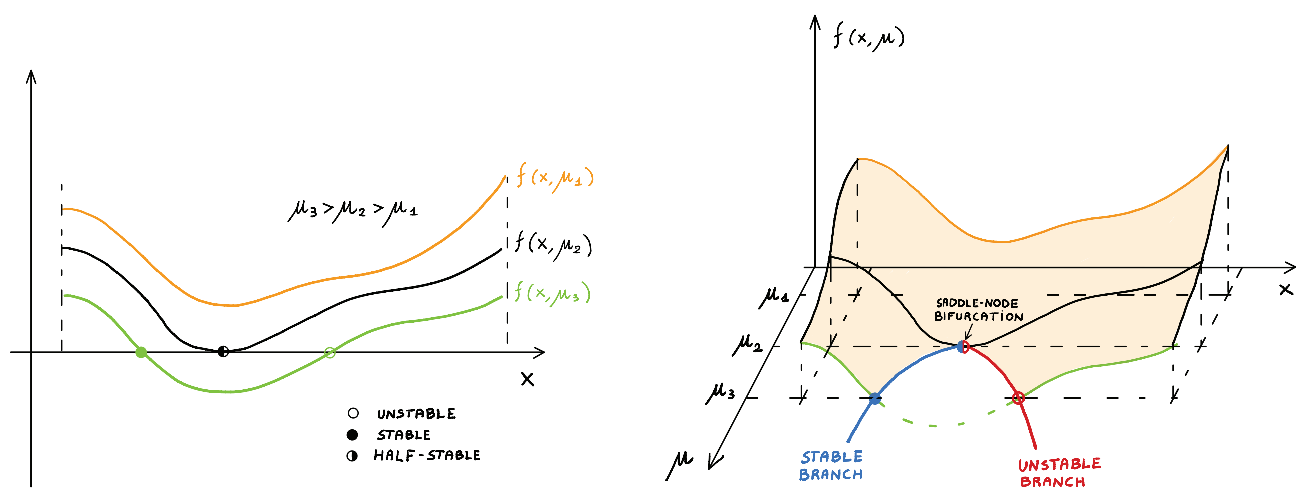


Figure 1: Sketch of a saddle-node bifurcation.

Clearly, the fixed points of the system (2) are in the zero level set[2] of $f(x,\mu)$, i.e., they are solutions of the equation

$$f(x,\mu) = 0. \tag{4}$$

In Figure 1, the zero level set of $f(x,\mu)$ is represented by the stable and unstable branches of fixed points that originate from a saddle-node bifurcation at $\mu = \mu_2$. Of course, there are many other ways the function $f(x,\mu)$ can intersect the plane (x,μ). For instance, we can have the zero level set corresponding to the so-called sub-critical pitchfork bifurcation. This is sketched in Figure 2.

[1]More generally, $f(x)$ can depend on multiple parameters, i.e., we can have $f(x,\mu_1,\ldots,\mu_M)$ in equation (2).

[2]The zero level set of a function $f(x,\mu)$ is the set of points $(x,\mu) \in \mathbb{R}^2$ such that the function is equal to zero, i.e.,

$$\{(x,\mu) \in \mathbb{R}^2 : \ f(x,\mu) = 0\} \qquad \text{(zero level set of } f\text{)}. \tag{3}$$

If the function $f(x,\mu)$ does not intersect the (x,μ) plane, then the zero level set is empty.

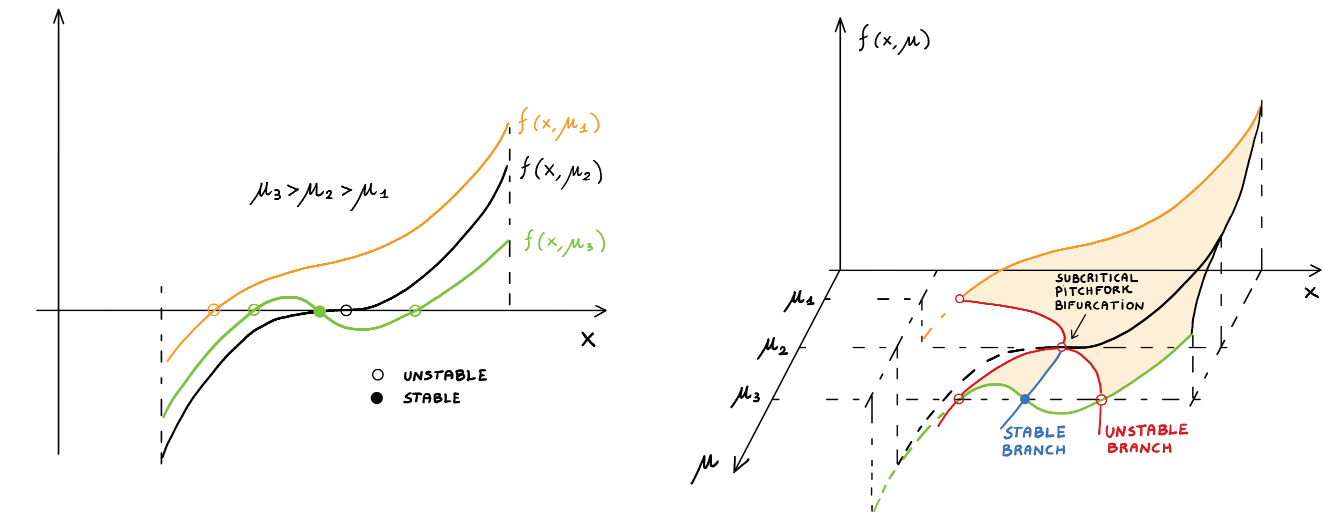


Figure 2: Sketch of a subcritical pitchfork bifurcation.

Bifurcation diagram. The zero level set of $f(x, \mu)$, i.e., the set of points $(x, \mu) \in \mathbb{R}^2$ satisfying (4) is called *bifurcation diagram* of fixed points. In practice, the bifurcation diagram provides the location of all fixed points of the system as a function of the parameter μ. Hereafter we sketch the bifurcation diagrams corresponding to the saddle-node bifurcation sketched in Figure 1, and the subcritical pitchfork bifurcation sketched Figure 2.

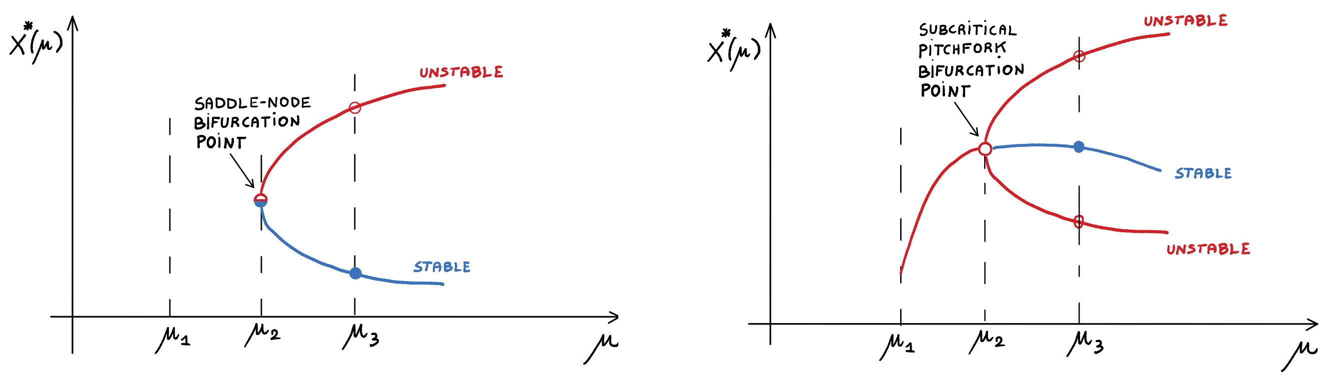


Figure 3: Bifurcation diagrams corresponding to the saddle-node bifurcation sketched in Figure 1, and the subcritical pitchfork bifurcation sketched in Figure 2.

In the bifurcation diagram we usually plot the location of the fixed points $x^*(\mu)$ as a function of the bifurcation parameter μ, but it is also possible to plot the bifurcation parameter $\mu(x^*)$ versus the location of the fixed points x^*. In this setting, the saddle-node bifurcation diagram shown in Figure 3(left) becomes a parabolic function $\mu(x^*)$ with upward concavity.

What is the relation between the coordinates of the fixed points x^* and the parameter μ? In particular, is it possible to express the zero level set of $f(x, \mu)$ as a graph of a smooth function? The answer is provided by the implicit function theorem.

Theorem 1 (Implicit function theorem)**.** Let $f(x,\mu)$ be a function of class $\mathcal{C}^1$ (i.e., continuously differentiable) in x and μ in a neighborhood of a point (x^*,μ^*) such that $f(x^*,\mu^*)=0$. If

$$\frac{\partial f(x^*,\mu^*)}{\partial x} \neq 0 \tag{5}$$

then there exists a neighborhood B of μ^* in which the zero level set of $f(x,\mu)$ can be represented as a graph of a smooth function $x^*(\mu)$, i.e.,

$$f(x^*(\mu),\mu)=0 \qquad \text{for all } \mu\in B. \tag{6}$$

The function $x^*(\mu)$ is of class $\mathcal{C}^1(B)$ (continuously differentiable in B) and it satisfies the additional properties:

$$x^*(\mu^*)=x^*, \qquad\qquad \frac{dx^*(\mu^*)}{d\mu} = -\frac{\dfrac{\partial f(x^*,\mu^*)}{\partial \mu}}{\dfrac{\partial f(x^*,\mu^*)}{\partial x}}. \tag{7}$$

Comments on the implicit function theorem:

- Theorem 1 indicates that the bifurcation diagram is composed of smooth curves $x^*(\mu)$ (continuously differentiable in μ), except at points (x^*,μ^*) where
$$\frac{\partial f(x^*,\mu^*)}{\partial x} = 0. \tag{8}$$
- Property (7) follows immediately by differentiating (6) with respect to μ and evaluating the derivative at at $\mu=\mu^*$. In fact,
$$f(x^*(\mu),\mu)=0 \quad \Rightarrow \quad \frac{df(x^*(\mu),\mu)}{d\mu}=0 \quad \Rightarrow \quad \frac{\partial f(x^*(\mu),\mu)}{\partial \mu} + \frac{\partial f(x^*(\mu),\mu)}{\partial x}\frac{dx^*(\mu)}{d\mu} = 0. \tag{9}$$
By evaluating the last equation at $\mu=\mu^*$ and recalling that $x^*(\mu^*)=x^*$ we obtain (7). Note that we can divide by $\partial f(x^*,\mu^*)/\partial x$ because it is nonzero by assumption (5).
- The role of x and μ can be reversed in Theorem 1. In other words, it is possible to formulate the implicit function theorem by choosing x as independent variable and μ as dependent variable. In this formulation, if (x^*,μ^*) is in the zero level set of f and $\partial f(x^*,\mu^*)/\partial\mu\neq 0$ then there exists a smooth function $\mu^*(x)$ in a neighborhood of x^* that represents the zero level set of $f(x,\mu)$ for all x in such a neighborhood, i.e., $f(x,\mu^*(x))=0$. With reference to the saddle-node bifurcation sketched in Figure 3, it is seen that at the saddle-node bifurcation point we have $dx^*(\mu^*)/d\mu=\infty$, suggesting that $\partial f(x^*,\mu^*)/\partial x=0$ and $\partial f(x^*,\mu^*)/\partial \mu\neq 0$ (see Eq. (7)). On the other hand, $d\mu^*(x^*)/dx=0$, implying again that $\partial f(x^*,\mu^*)/\partial x=0$.

From the discussion above, it appears that the conditions

$$f(x^*,\mu^*)=0, \qquad \frac{\partial f(x^*,\mu^*)}{\partial x}=0 \tag{10}$$

may suggest that a bifurcation is taking place at (x^*,μ^*). While these condition are indeed necessary (otherwise the implicit function theorem applies), they are not sufficient to guarantee the existence of a bifurcation as the following example clearly demonstrates.

Example: Consider the function

$$f(x,\mu)=x^3+\mu. \tag{11}$$

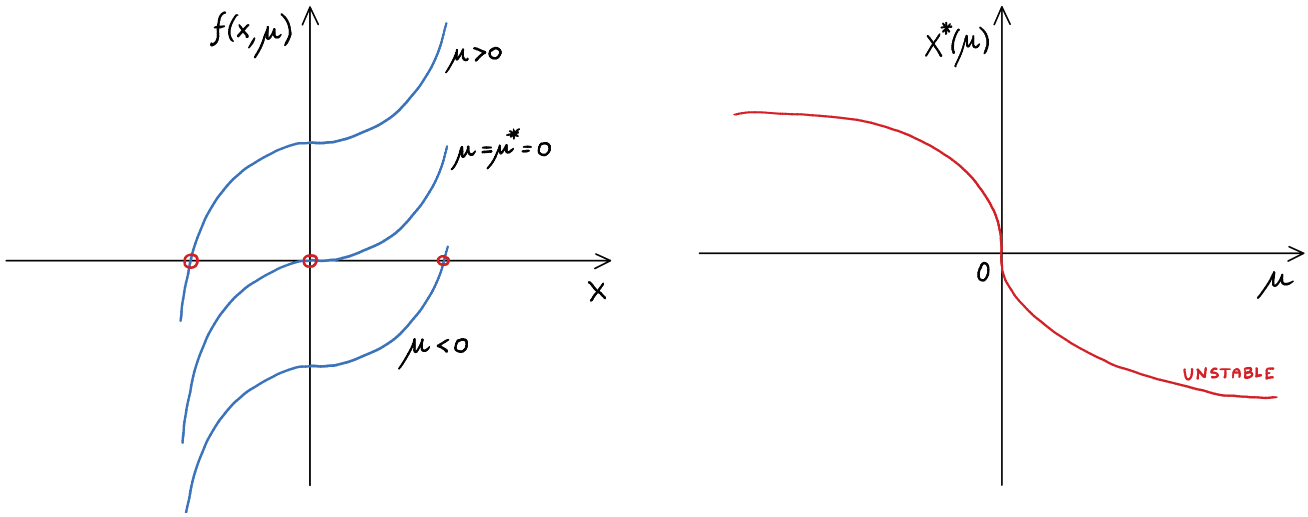


Figure 4: Sketch of the function (11) for different μ (left) and bifurcation diagram (right). In this case there is only one unstable fixed point that moves around as we very the parameter μ, and no bifurcation whatsoever. Yet, the (necessary) conditions for a bifurcation summarized in equation (10) are both satisfied at $(x^*, \mu^*) = (0, 0)$.

Clearly

$$f(0,0) = 0, \qquad \frac{\partial f(0,0)}{\partial x} = 2x^2\Big|_{(x,\mu)=(0,0)} = 0. \tag{12}$$

Hence, $(x^*, \mu^*) = (0, 0)$ may be a bifurcation point. However, the zero level set in this case can be expressed analytically as (see Figure 4)

$$f(x,\mu) = x^3 + \mu = 0 \qquad \Rightarrow \qquad x^*(\mu) = -\sqrt[3]{\mu}. \tag{13}$$

This shows that there is indeed no bifurcation at $(x^*, \mu^*) = (0, 0)$. In Figure 4 we sketch the function (11) for different values of μ, and the corresponding bifurcation diagram.

In general, a bifurcation is characterized by two or more branches of fixed point intersecting at some location for some value of the bifurcation parameter μ (see Figure 5). Such multiplicity of branches emanating from the bifurcation point (x^*, μ^*) is usually associated with non-invertibility of the zero level set of $f(x, \mu)$ at (x^*, μ^*), which can be studied by using Taylor series (next section). Of course, it is possible to have functions $f(x, \mu)$ with rather complicated zero level sets, and multiple bifurcations of different types. For example, in Figure 5 we sketch a bifurcation diagram with five different types of bifurcations. In particular,

- Saddle-node bifurcation,
- Transcritical bifurcation,
- Pitchfork bifurcation (supercritical and subcritical).

The "exotic" bifurcation looks like a saddle-node but it involves four branches instead of two.

It is important to always keep in mind that the bifurcation diagram represents the location of the fixed points of the systems as a function of μ. The continuity requirement we imposed on $f(x, \mu)$ prohibits fixed points of the same type, e.g., two stable nodes, to face each other. Consequently, two stable branches or two unstable branches cannot be facing each other in the bifurcation diagram.

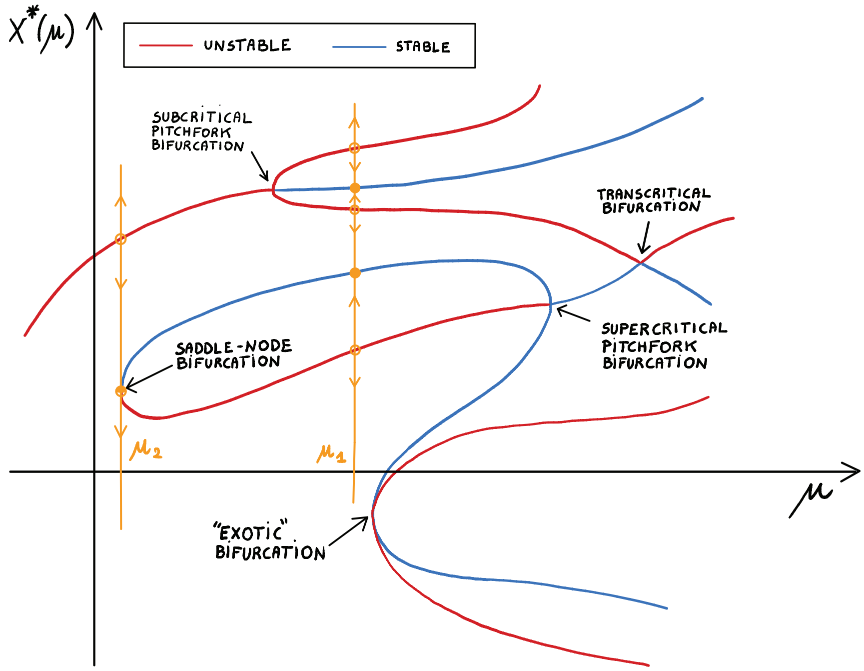


Figure 5: Sketch of a bifurcation diagram with 5 different bifurcations.

Polynomial approximation at bifurcation points. To study invertibility of the zero level set of $f(x,\mu)$ in a neighborhood of a point (x^*,μ^*) that belongs to such zero level set we expand $f(x,\mu)$ in a Taylor series

$$f(x,\mu) = \sum_{k,m=0}^{\infty} \frac{1}{k!m!}\frac{\partial^{k+m} f(x^*,\mu^*)}{\partial x^k \partial \mu^m}(x-x^*)^k(\mu-\mu^*)^m, \tag{14}$$

i.e.,

$$\begin{aligned} f(x,\mu) =& f(x^*,\mu^*) + \frac{\partial f(x^*,\mu^*)}{\partial x}(x-x^*) + \frac{\partial f(x^*,\mu^*)}{\partial \mu}(\mu-\mu^*) + \frac{1}{2}\frac{\partial^2 f(x^*,\mu^*)}{\partial x^2}(x-x^*)^2 + \\ & \frac{1}{2}\frac{\partial^2 f(x^*,\mu^*)}{\partial \mu^2}(\mu-\mu^*)^2 + \frac{\partial^2 f(x^*,\mu^*)}{\partial x \partial \mu}(x-x^*)(\mu-\mu^*) + \frac{1}{6}\frac{\partial^3 f(x^*,\mu^*)}{\partial x^3}(x-x^*)^3 + \\ & \frac{1}{6}\frac{\partial^3 f(x^*,\mu^*)}{\partial \mu^3}(\mu-\mu^*)^3 + \frac{1}{2}\frac{\partial^3 f(x^*,\mu^*)}{\partial x \partial \mu^2}(x-x^*)(\mu-\mu^*)^2 + \frac{1}{2}\frac{\partial^3 f(x^*,\mu^*)}{\partial x^2 \partial \mu}(x-x^*)^2(\mu-\mu^*) + \cdots \end{aligned} \tag{15}$$

We know that if (x^*,μ^*) is a bifurcation point then

$$f(x^*,\mu^*) = 0, \qquad \text{and} \qquad \frac{\partial f(x^*,\mu^*)}{\partial x} = 0. \tag{16}$$

A substitution of (16) into (15) yields

$$\begin{aligned} f(x,\mu) =& \frac{\partial f(x^*,\mu^*)}{\partial \mu}R + \frac{1}{2}\frac{\partial^2 f(x^*,\mu^*)}{\partial x^2}X^2 + \frac{1}{2}\frac{\partial^2 f(x^*,\mu^*)}{\partial \mu^2}R^2 + \frac{\partial^2 f(x^*,\mu^*)}{\partial x \partial \mu}XR + \\ & \frac{1}{6}\frac{\partial^3 f(x^*,\mu^*)}{\partial x^3}X^3 + \frac{1}{6}\frac{\partial^3 f(x^*,\mu^*)}{\partial \mu^3}R^3 + \frac{1}{2}\frac{\partial^3 f(x^*,\mu^*)}{\partial x \partial \mu^2}XR^2 + \frac{1}{2}\frac{\partial^3 f(x^*,\mu^*)}{\partial x^2 \partial \mu}X^2 R + \cdots \end{aligned} \tag{17}$$

where we defined the "centered" variables

$$X = x - x^*, \qquad R = \mu - \mu^*. \tag{18}$$

If (x, μ) is also in the zero level set then $f(x, \mu) = 0$ in (17). This yields polynomial equation in X and R that characterizes the locations of the fixed points in a neighborhood of a fixed point (x^*, μ^*) satisfying (16). The multiplicity of possible solutions to such polynomial equation is what eventually yields multiple branches of equilibria emanating from the fixed point (x^*, μ^*).

Depending on the leading order of the polynomial expansion of the system at the bifurcation point, we can have a different number of branches of equilibria involved in the bifurcation process. As we shall see hereafter, both saddle-node and transcritical bifurcations are represented locally by polynomials of degree 2. Pitchfork bifurcations by polynomials of degree 3. The "exotic" bifurcation shown in Figure 5 is generated by system that is locally equivalent to a polynomial of degree 4.

Saddle-node bifurcation. We have now all element to characterize the saddle-node bifurcation sketched in Figure 1 and Figure 3(left).

Theorem 2 (Saddle-node bifurcation)**.** Let (x^*, μ^*) be a fixed point of the dynamical system (2), i.e., $f(x^*, \mu^*) = 0$. If the following conditions are satisfied

$$\frac{\partial f(x^*, \mu^*)}{\partial x} = 0, \qquad \frac{\partial f(x^*, \mu^*)}{\partial \mu} \neq 0, \qquad \frac{\partial^2 f(x^*, \mu^*)}{\partial x^2} \neq 0, \tag{19}$$

then the system undergoes a saddle-node bifurcation at (x^*, μ^*).

To characterize the saddle-node bifurcation *quantitatively*, we choose R in (17) to be of the same order of magnitude as X^2. For example, if $X \simeq 10^{-5}$ then $R \simeq 10^{-10}$. In this way, the leading terms in the Taylor series (17) have the same order of magnitude, and we can neglect higher-order terms in a straightforward way. In this assumption, the Taylor series (17) can be written as

$$f(x, \mu) = AX^2 + BR + \cdots . \tag{20}$$

where we set

$$A = \frac{\partial f(x^*, \mu^*)}{\partial \mu} \neq 0, \qquad \text{and} \qquad B = \frac{1}{2}\frac{\partial^2 f(x^*, \mu^*)}{\partial x^2} \neq 0. \tag{21}$$

Equation (20) allows us to write the following polynomial approximation[3] of the dynamical system (2) in a neighborhood of the bifurcation point (x^*, μ^*)

$$\frac{dX}{dt} = AX^2 + BR. \tag{23}$$

Diving by A and rescaling the time variable t as $\tau = At$ yields[4]

$$\frac{dX}{d\tau} = X^2 + H \qquad \text{(normal form)} \tag{25}$$

where $\tau = At$, and $H = RB/A$ is a rescaled version of the bifurcation parameter μ. Any dynamical system that undergoes a saddle-node bifurcation can be written as (25) in a neighborhood of the bifurcation point, i.e., for (x, μ) very close to (x^*, μ^*). This is the reason why (25) is called the *normal form* of a dynamical system that undergoes a saddle-node bifurcation.

In Figure 6 we clarify the meaning of the normal form of a saddle-node bifurcation. In Figure 7 we sketch the corresponding bifurcation diagram.

[3]Note that since x^* is a constant we have

$$\frac{dX}{dt} = \frac{d(x - x^*)}{dt} = \frac{dx}{dt}. \tag{22}$$

[4]In equation (25) we assumed that $A > 0$. If $A <$ then we divide by the modulus of A, i.e., $|A|$, which leaves a minus sign in front of X^2 in (25), i.e.,

$$\frac{dX}{d\tau} = -X^2 + H \qquad , \qquad \tau = |A|t, \qquad , H = RB/|A|. \tag{24}$$

.

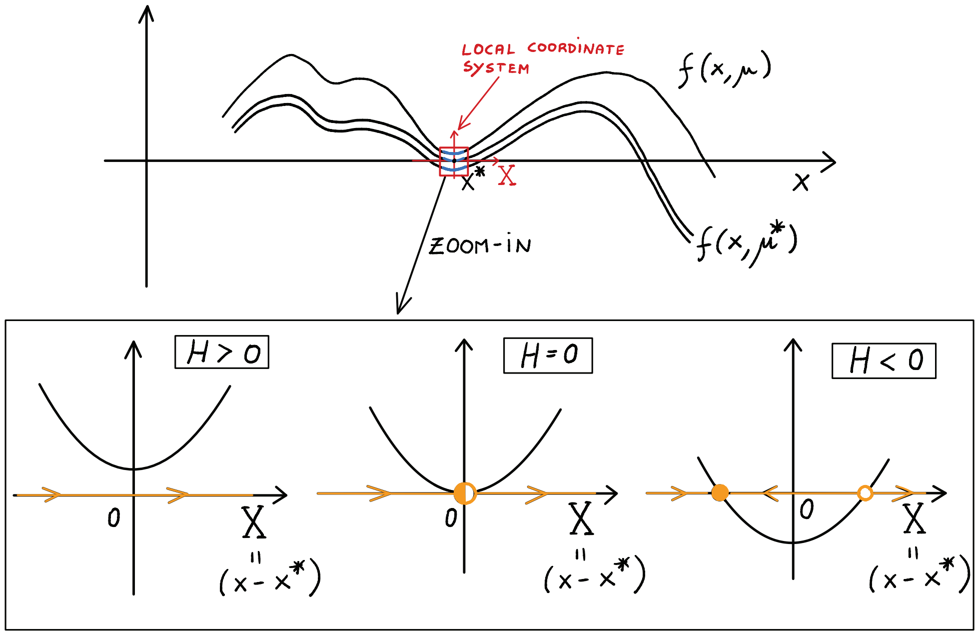


Figure 6: Saddle-node bifurcation. Shown is the function $f(x, \mu)$ for three values of μ, and a zoom-in of the bifurcation process. The nonlinear dynamical system (2) in such a small region is approximated by the normal form (25) (after appropriate rescaling).

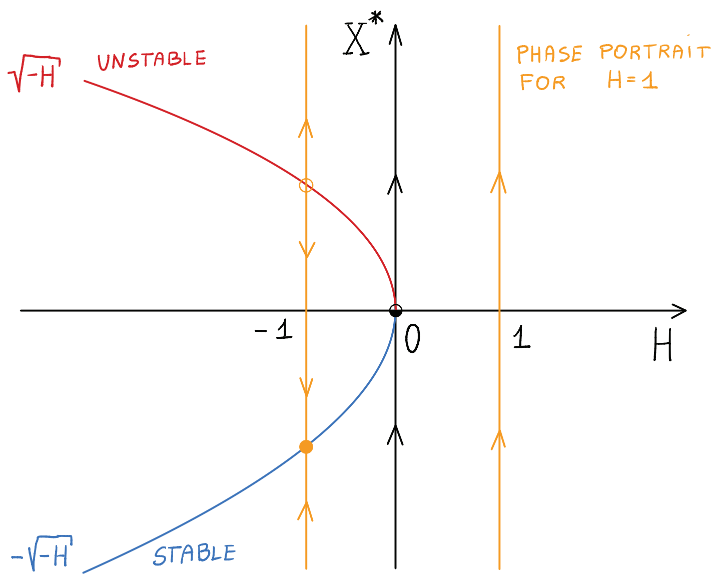


Figure 7: Bifurcation diagram for the normal form of a saddle-node bifurcation. Similarly to Figure 6, this bifurcation diagram describes what happens in an extremely small region that includes the bifurcation point (x^*, μ^*), i.e., the region in red in Figure 6.

Example: Consider the nonlinear system

$$\frac{dx}{dt} = \sin(x) + \mu \tag{26}$$

In Figure 8 we plot $f(x, \mu)$ together with its zero level set, i.e., the bifurcation diagram. Note that if $\mu^* = 1$

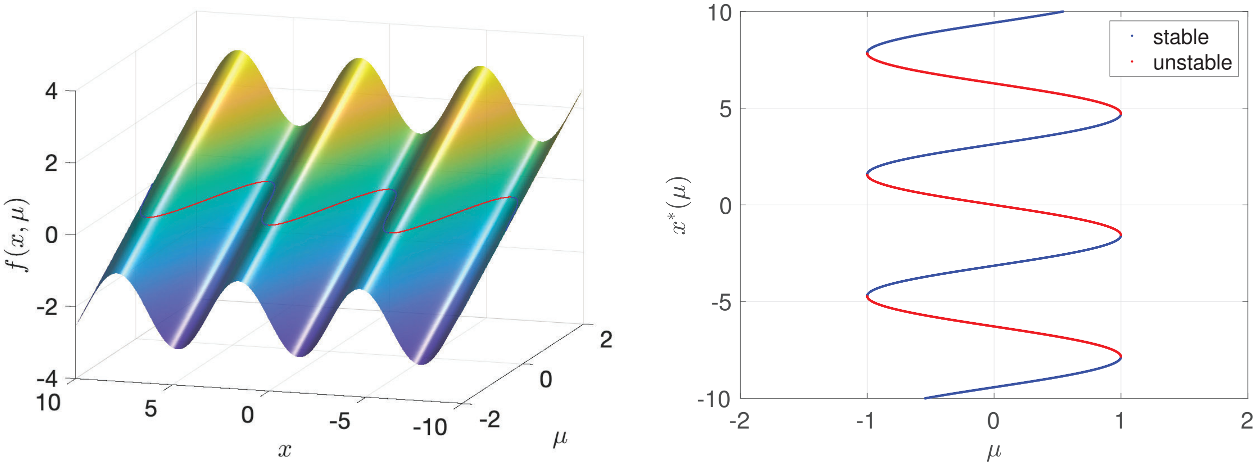


Figure 8: Plot of the right hand side of equation 26, i.e., $f(x,\mu) = \sin(x) + \mu$ together with its zero level set (left), and bifurcation diagram (right).

we have that $f(x,\mu) = \sin(x) + \mu$ is tangent to the x axis at the points

$$x_k^* = -\frac{\pi}{2} + 2k\pi, \qquad k \in \mathbb{Z}. \tag{27}$$

At such points we have

$$\frac{\partial f(x_k^*,\mu^*)}{\partial x} = \cos(x_k^*) = 0 \qquad \frac{\partial f(x_k^*,\mu^*)}{\partial \mu} = 1 \neq 0 \qquad \frac{\partial^2 f(x_k^*,\mu^*)}{\partial x^2} = -\sin(x_k^*) = -1 \neq 0. \tag{28}$$

Hence, the conditions of Theorem 2 are satisfied, implying that (x_k^*, μ^*) ($k \in \mathbb{Z}$) are all saddle-node bifurcations. It is straightforward to show that when $\mu^* = -1$ there is another infinite number of of saddle-node bifurcations at

$$x_k^* = \frac{\pi}{2} + 2k\pi, \qquad k \in \mathbb{Z}. \tag{29}$$

Example: Consider the nonlinear system

$$\frac{dx}{dt} = e^{-x^2/\mu} - \frac{\sin(x\mu)}{(x^2+1)}. \tag{30}$$

In this case, it is not possible to determine the fixed points of the system analytically. In fact, the fixed points are solutions to the transcendental equation

$$e^{-x^2/\mu} = \frac{\sin(x\mu)}{(x^2+1)}, \tag{31}$$

which cannot be solved analytically. However, it is rather straightforward to compute the fixed points numerically, e.g., as zero level sets of the two dimensional function (30) or using any root finding solver. The result is shown in Figure 9, where we see that there is an infinite number of saddle node bifurcations that tend to cluster as μ increases

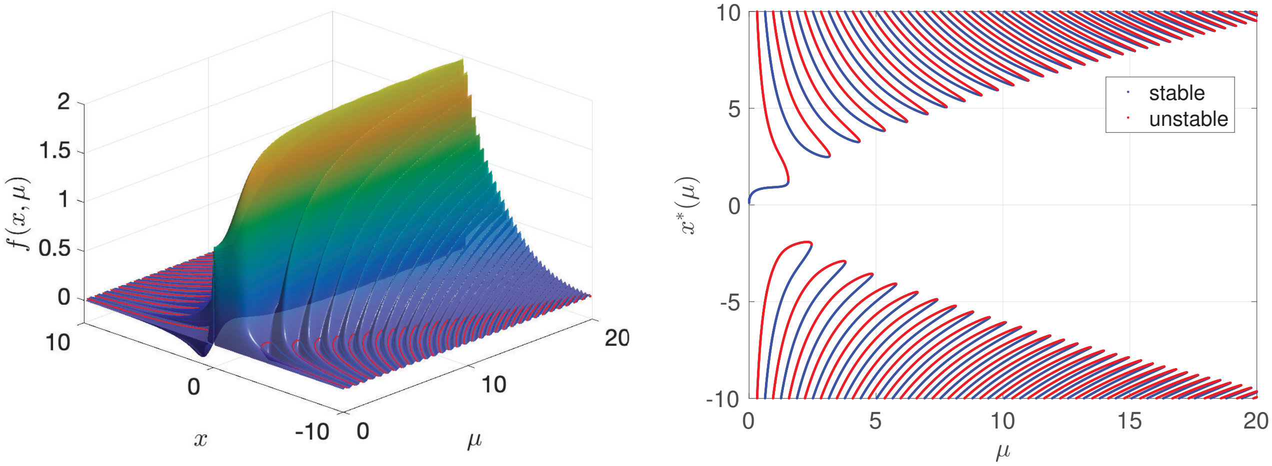


Figure 9: Plot of $f(x, \mu)$ defined in equation 30 (right hand side) together with its zero level set (left), and bifurcation diagram (right).

Transcritical bifurcation. Transcritical bifurcations are rather common bifurcations of equilibria, which can be characterized by the following theorem.

Theorem 3 (Transcritical bifurcation)**.** Let (x^*, μ^*) be a fixed point of the dynamical system (2), i.e., $f(x^*, \mu^*) = 0$. If the following conditions are satisfied

$$\frac{\partial f(x^*, \mu^*)}{\partial x} = 0, \qquad \frac{\partial f(x^*, \mu^*)}{\partial \mu} = 0, \qquad \frac{\partial^2 f(x^*, \mu^*)}{\partial x^2} \neq 0, \qquad \frac{\partial^2 f(x^*, \mu^*)}{\partial x \partial \mu} \neq 0 \tag{32}$$

then the system undergoes a transcritical bifurcation at (x^*, μ^*).

A substitution of (32) into (17) yields (to leading order in $X = x - x^*$ and $R = \mu - \mu^*$)

$$f(x, \mu) = BX^2 + CXR + \cdots . \tag{33}$$

In equation (33) we set

$$B = \frac{1}{2}\frac{\partial^2 f(x^*, \mu^*)}{\partial x^2} \neq 0, \qquad \text{and} \qquad C = \frac{\partial^2 f(x^*, \mu^*)}{\partial x \partial \mu} \neq 0. \tag{34}$$

Hence, the dynamics nearby a transcritical bifurcation point is characterized by the following polynomial approximation of the dynamical system (2)

$$\frac{dX}{dt} = BX^2 + CXR, . \tag{35}$$

which can be normalized (divide by B) as[5]

$$\frac{dX}{d\tau} = X^2 + XH \qquad \text{(normal form)}, \tag{37}$$

where $\tau = Bt$, and $H = CR/B$. The fixed points of (37) are $X^* = 0$ and $X^* = -H$. In Figure 10 plot the velocity vector that defines the normal form of a transcritical bifurcation and sketch the bifurcation diagram.

[5] As in the case of the saddle-node bifurcation, if $B < 0$ then we divide by the modulus of B, which yields a minus in front of X^2 in (35), i.e.,

$$\frac{dX}{d\tau} = -X^2 + XH. \tag{36}$$

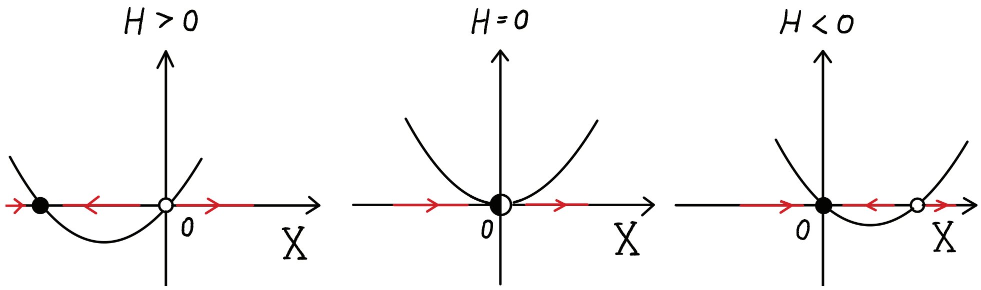


Figure 10: Transcritical bifurcation in local coordinates. Shown is the function $X^2 + XH$ appearing at the right hand side of the normal form (37) for three values of H. The nonlinear dynamical system (2) is approximated by the normal form (37) in a neighborhood of the bifurcation point (after appropriate rescaling).

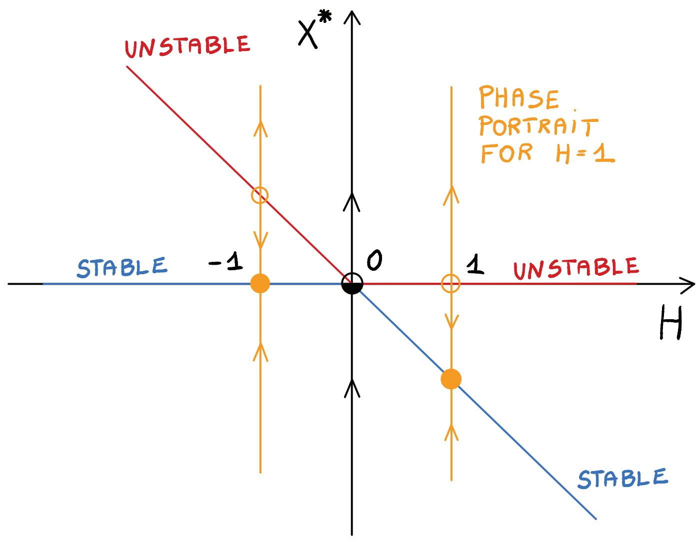


Figure 11: Bifurcation diagram for the normal form of a transcritical bifurcation.

Example: Consider the following dynamical system

$$\frac{dx}{dt} = \underbrace{\mu \log(x) + x - 1}_{f(x,\mu)}. \tag{38}$$

The fixed points are obtained by setting $f(x, \mu) = 0$. This yields,

$$\mu \log(x) = 1 - x. \tag{39}$$

Clearly, for $x = 1$ the equation above reads $0 = 0$, which means that $x^* = 1$ is a fixed point for all values of μ. Next, we compute the derivative of $f(x, \mu)$ with respect to x

$$\frac{\partial f(x, \mu)}{\partial x} = \frac{\mu}{x} + 1. \tag{40}$$

A necessary condition for (x^*, μ^*) to be a bifurcation point is

$$\frac{\partial f(x^*, \mu^*)}{\partial x} = 0 \qquad \Rightarrow \qquad \mu^* = -x^*. \tag{41}$$

Recalling that $x^* = 1$ is always a fixed point, we find that $(x^*, \mu^*) = (1, -1)$ could be a bifurcation point. Let us verify that $(1, -1)$ is indeed a transcritical bifurcation point. To this end, we just need to verify the conditions in Theorem 3. We have,

$$\frac{\partial f(1,-1)}{\partial \mu} = 0, \qquad \frac{\partial^2 f(1,-1)}{\partial x^2} = 1 \neq 0, \qquad \frac{\partial^2 f(1,-1)}{\partial x \partial \mu} = 1 \neq 0. \tag{42}$$

Therefore $(x^*, \mu^*) = (1, -1)$ is a transcritical bifurcation point. Let us compute the normal form of the system (38) at the bifurcation point. Recalling (34)-(35) and using (42) we have

$$\frac{dX}{dt} = \frac{X^2}{2} + XR, \tag{43}$$

where $X = x - 1$ and $R = \mu + 1$. Divide (43) by 1/2 to obtain the normal form

$$\frac{dX}{d\tau} = X^2 + XH, \tag{44}$$

where

$$\tau = \frac{t}{2}, \qquad \text{and} \qquad H = 2(\mu + 1). \tag{45}$$

Pitchfork bifurcation. Another type of rather common bifurcation of equilibria is the pitchfork bifurcation, which can be *supercritical* or *subcrititical* (see Figure 2 and Figure 5). The following Theorem characterizes pitchfork bifurcations.

Theorem 4 (Pitchfork bifurcation)**.** Let (x^*, μ^*) be a fixed point of the dynamical system (2), i.e., $f(x^*, \mu^*) = 0$. If the following conditions are satisfied

$$\frac{\partial f(x^*,\mu^*)}{\partial x} = 0, \qquad \frac{\partial f(x^*,\mu^*)}{\partial \mu} = 0, \qquad \frac{\partial^2 f(x^*,\mu^*)}{\partial x^2} = 0, \qquad \frac{\partial^3 f(x^*,\mu^*)}{\partial x^3} \neq 0, \qquad \frac{\partial^2 f(x^*,\mu^*)}{\partial x \partial \mu} \neq 0 \tag{46}$$

then the system undergoes a pitchfork bifurcation at (x^*, μ^*).

As mentioned above, pitchfork bifurcations can be supercritical or subcritical, depending on the sign of $\partial^3 f/\partial x^3$ and $\partial^2 f/\partial x \partial \mu$. A substitution of (46) into (17) yields (to leading order in $X = x - x^*$ and $R = \mu - \mu^*$

$$f(x, \mu) = DX^3 + CXR + \cdots . \tag{47}$$

where we set

$$D = \frac{1}{6}\frac{\partial^3 f(x^*,\mu^*)}{\partial x^3} \neq 0, \qquad \text{and} \qquad C = \frac{\partial^2 f(x^*,\mu^*)}{\partial x \partial \mu} \neq 0. \tag{48}$$

Hence, to leading order, we obtain the following polynomial approximation of (2) at a pitchfork bifurcation point

$$\frac{dX}{dt} = DX^3 + CXR. \tag{49}$$

Dividing by the modulus of D yields

$$\frac{dX}{d\tau} = X^3 + XH \qquad (D > 0) \qquad \text{subcritical pitchfork,} \tag{50}$$

$$\frac{dX}{d\tau} = -X^3 + XH \qquad (D < 0) \qquad \text{supercritical pitchfork,} \tag{51}$$

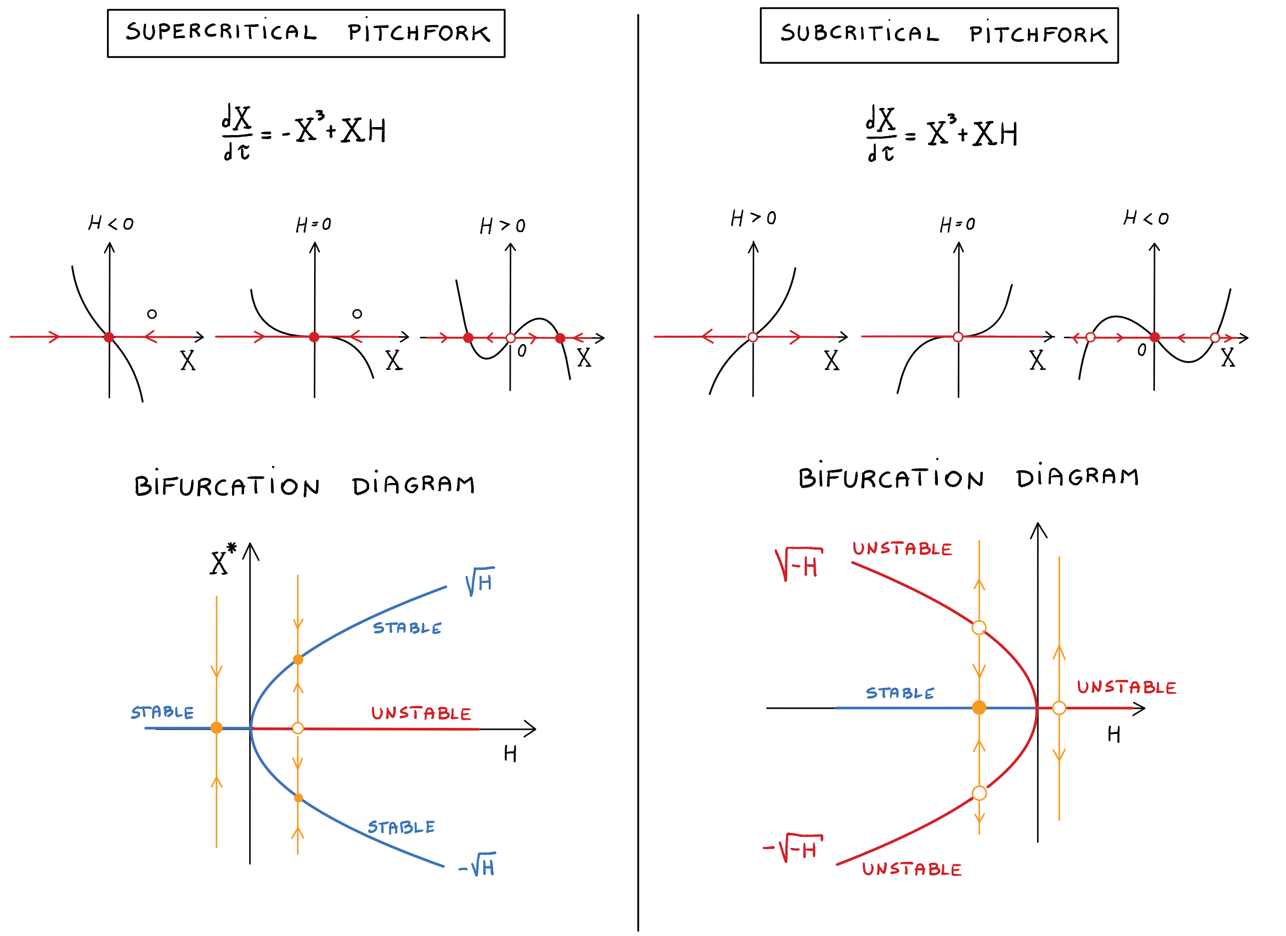


Figure 12: Bifurcation diagrams for supercritical and subcritical pitchfork bifurcations.

where we set $\tau = |D|t$ and $H = CR/|D|$. Any dynamical system of the form (2) that undergoes a pitchfork bifurcation at (x^*, μ^*) can be written (to leading order in X and R after appropriate rescaling) either as (50) or (51) in a neighborhood of (x^*, μ^*). For this reason, (50) and (51) are referred to as the em normal forms of the supercritical and subcritical pitchfork bifurcations, respectively. In Figure 12 plot the velocity vectors associated with the normal forms (50) or (51) and sketch the bifurcation diagrams.

Example: Consider the system

$$\frac{dx}{dt} = \underbrace{\sin(x) + \mu x}_{f(x,\mu)}. \tag{52}$$

The fixed points are solutions to the transcendental equation

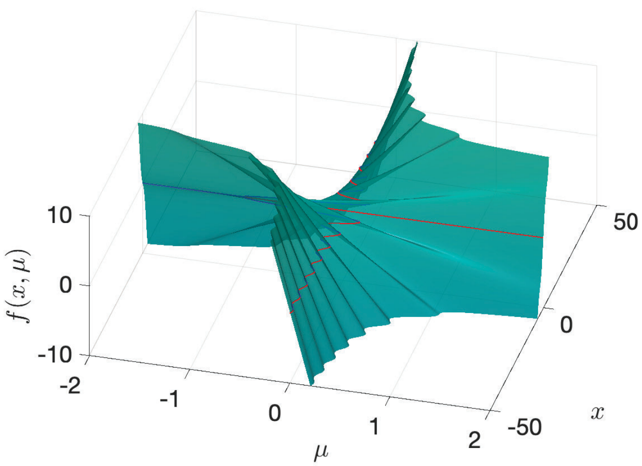
$$\sin(x) + \mu x = 0. \tag{53}$$

Clearly $x^* = 0$ is a fixed point for all μ. Note also that for $x \neq 0$ the bifurcation diagram is completely defined by the equation

$$\mu(x^*) = \frac{\sin(x^*)}{x^*}, \tag{54}$$

which explicitly expresses the bifurcation parameter as a function of the location of the fixed points. The derivative of the right hand side of (52) is

$$\frac{\partial f}{\partial x} = \cos(x) + \mu. \tag{55}$$



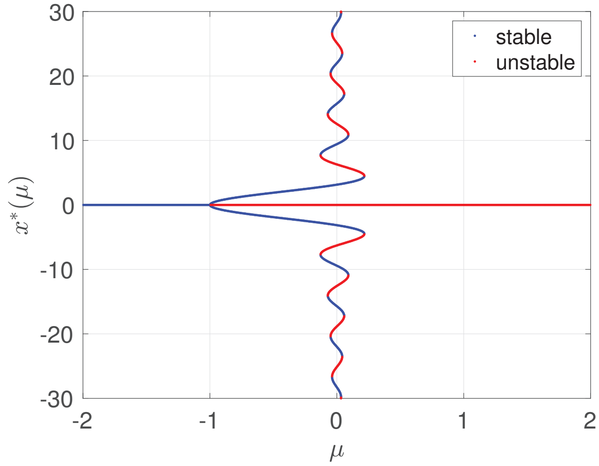


Figure 13: Bifurcation diagram for the system (52).

Hence, for $x^* = 0$ we have that $\mu^* = -1$ makes (55) equal to zero. This means that $(x^*, \mu^*) = (0, -1)$ could be a bifurcation point as it satisfies the necessary conditions (10). Let us verify that $(x^*, \mu^*) = (0, -1)$ is indeed a pitchfork bifurcation point. To this end, it is necessary and sufficient to verify the conditions in Theorem 4. We have

$$\frac{\partial^2 f}{\partial x^2} = -\sin(x) \quad \Rightarrow \frac{\partial^2 f(x^*, \mu^*)}{\partial x^2} = 0, \tag{56}$$

$$\frac{\partial^3 f}{\partial x^3} = -\cos(x) \quad \Rightarrow \frac{\partial^3 f(x^*, \mu^*)}{\partial x^3} = -1, \tag{57}$$

$$\frac{\partial^2 f}{\partial x \partial \mu} = 1 \quad \Rightarrow \frac{\partial^2 f(x^*, \mu^*)}{\partial x \partial \mu} = 1. \tag{58}$$

Hence, there is a *supercritical pitchfork bifurcation* point at $(x^*, \mu^*) = (0, -1)$. Note, in fact, that (57) implies that the coefficient D in (48) is negative, and therefore the normal form representing the bifurcation in this case is (51). As shown in Figure 13, the system exhibits also an infinite number of saddle-node bifurcations as the parameter μ is varied. Such saddle-node bifurcations are defined analytically by equation (54).

Other bifurcations of equilibria. The Taylor series (17) can be (to leading order) a rather arbitrary polynomial in X and R. This opens the possibility to have more "exotic" bifurcations of equilibria in which one point splits into four points or more (see Figure 5), or bifurcation in which multiple stable and unstable branches intersect at one point.

Introduction to n-dimensional dynamical systems

Consider the following n-dimensional system of nonlinear ODEs

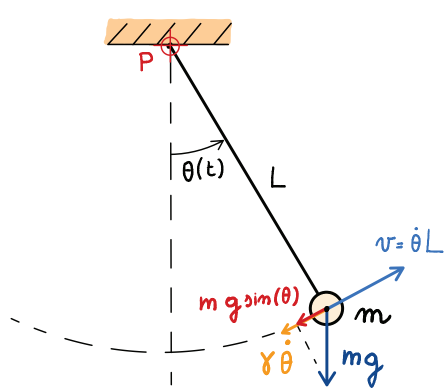
$$\begin{cases} \dfrac{d\boldsymbol{x}}{dt} = \boldsymbol{f}(\boldsymbol{x}) \\ \boldsymbol{x}(0) = \boldsymbol{x}_0 \end{cases} \tag{1}$$

where $\boldsymbol{x}(t) = [x_1(t) \cdots x_n(t)]^T$ is a vector of phase variables, $\boldsymbol{f} : D \to \mathbb{R}^n$, and D is a subset of $\mathbb{R}^n$. In an expanded notation the system (1) can be written as

$$\begin{cases} \dfrac{dx_1}{dt} = f_1(x_1, \ldots, x_n) \\ \dfrac{dx_2}{dt} = f_2(x_1, \ldots, x_n) \\ \quad\vdots \\ \dfrac{dx_n}{dt} = f_n(x_1, \ldots, x_n) \\ x_1(0) = x_{10} \\ x_2(0) = x_{20} \\ \quad\vdots \\ x_n(0) = x_{n0} \end{cases} \tag{2}$$

Dynamical systems of the form (1) can model phenomena in classical mechanics (e.g., the pendulum equations), health and medicine (e.g., cancer models), weather patterns, material science, and quantum physics. They can also be used to approximate the dynamics of partial differential equations (PDEs). Let us provide two simple examples.

Example: Consider the following sketch of a pendulum (point mass), subject to gravity and friction with friction



As is well-known from physics, the rate of change (time derivative) of the angular momentum of the point mass m with respect to the point P equals the momentum of the external forces acting on the point mass. The external forces in this case are gravity and friction. Setting up the balance of momenta yields

$$mL^2 \frac{d^2\theta}{dt^2} = -mgL\sin(\theta) - \gamma L \frac{d\theta}{dt}, \tag{3}$$

i.e.

$$\frac{d^2\theta}{dt^2} = -\frac{g}{L}\sin(\theta) - \frac{\gamma}{Lm}\frac{d\theta}{dt}. \tag{4}$$

This is a second-order ordinary nonlinear differential equation in $\theta(t)$. The equation can be easily transformed into a system two first-order nonlinear ODEs by defining the new variables

$$x_1(t) = \theta(t), \qquad x_2(t) = \frac{d\theta(t)}{dt}. \tag{5}$$

Clearly, based on the definition of $x_1(t)$ and $x_2(t)$ we have

$$\frac{dx_1}{dt} = x_2. \tag{6}$$

Moreover, by differentiating $x_2(t)$ with respect to time and using equation (4) we obtain

$$\frac{dx_2}{dt} = -\frac{g}{L}\sin(x_1) - \frac{\gamma}{Lm}x_2. \tag{7}$$

Hence, equation (4) is equivalent to the two-dimensional nonlinear dynamical system

$$\begin{cases} \dfrac{dx_1}{dt} = x_2, \\ \dfrac{dx_2}{dt} = -\dfrac{g}{L}\sin(x_1) - \dfrac{\gamma}{Lm}x_2. \end{cases} \tag{8}$$

Of course, in order to solve the system of ODEs (8), we need an initial condition for the position of the pendulum $x_1(0)$, and an initial condition for the velocity of the pendulum $x_2(0)$.

Example: Consider the following initial-boundary value problem for the heat equation in the periodic spatial domain $[0, 2\pi]$

$$\begin{cases} \dfrac{\partial u(t,y)}{\partial t} = \alpha \dfrac{\partial^2 u(y,t)}{\partial y^2} & \text{heat equation} \\ u(0,y) = u_0(y) & \text{initial condition} \\ u(t,0) = u(t,2\pi) & \text{periodic boundary conditions} \end{cases} \tag{9}$$

A finite-difference approximation of the PDE (9) on the evenly-spaced grid with n points

$$y_k = (k-1)\Delta y \qquad k = 1,\ldots,n, \qquad \Delta y = \frac{2\pi}{n} \tag{10}$$

yields the n-dimensional linear dynamical system

$$\begin{cases} \dfrac{dx_1(t)}{dt} = \dfrac{\alpha}{\Delta y^2}\left[x_2(t) - 2x_1(t) + x_n(t)\right] \\ \dfrac{dx_2(t)}{dt} = \dfrac{\alpha}{\Delta y^2}\left[x_3(t) - 2x_2(t) + x_1(t)\right] \\ \qquad \vdots \\ \dfrac{dx_n(t)}{dt} = \dfrac{\alpha}{\Delta y^2}\left[x_1(t) - 2x_n(t) + x_{n-1}(t)\right] \end{cases} \tag{11}$$

where we defined $x_k(t) = u(y_k, t)$. Note that $x_k(t)$ represents an approximation of the solution to the partial differential equation (9) at the grid point $y = y_k$. Hence, by computing the solution to (11), we are computing an approximation of the solution to the PDE (9) at the grid points $(y_1, \ldots, y_n)$.

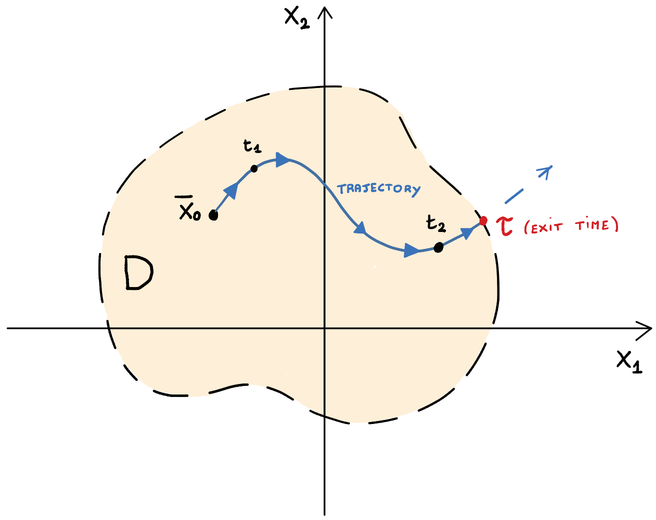


Figure 1: Illustration of the meaning of Theorem 1 in two-dimensions. Shown are the open set $D \subset \mathbb{R}^2$ in which $\boldsymbol{f}(\boldsymbol{x})$ is Lipschitz continuous, the trajectory corresponding to a particular $\boldsymbol{x}_0 \in D$ and the exit time τ for such trajectory.

Well-posedness of the initial value problem. Let us recall theorem that guarantees existence and uniqueness of the solution to the system of first order ODEs (1).

Theorem 1 (Existence and uniqueness of the solution to (1))**.** Let $D \subset \mathbb{R}^n$ be an open set, $\boldsymbol{x}_0 \in D$. If $\boldsymbol{f} : D \to \mathbb{R}^n$ is Lipschitz continuous in D then there exists a unique solution to the initial value problem (1) within the time interval $[0, \tau[$, where τ is the time instant at which the trajectory $\boldsymbol{x}(t)$ exits[1] the domain D. The solution $\boldsymbol{x}(t)$ is continuously differentiable in $[0, \tau[$.

How do we define Lipschitz continuity for a vector-valued function $\boldsymbol{f}(\boldsymbol{y})$ defined on subset of $\mathbb{R}^n$? By a simple generalization of the definition we gave for one-dimensional functions.

Definition 1. Let D be a subset of $\mathbb{R}^n$, $\boldsymbol{f} : D \to \mathbb{R}^n$. We say that $\boldsymbol{f}$ is Lipschitz continuous in D if there exists a constant $0 \leq L < \infty$ such that

$$\|\boldsymbol{f}(\boldsymbol{x}_1) - \boldsymbol{f}(\boldsymbol{x}_2)\| \leq L \|\boldsymbol{x}_1 - \boldsymbol{x}_2\| \quad \text{for all} \quad \boldsymbol{x}_1, \boldsymbol{x}_2 \in D, \tag{12}$$

where $\|\cdot\|$ is any norm defined in $\mathbb{R}^n$ (see Appendix B). Recall, in fact that all norms defined in a finite-dimensional space such as $\mathbb{R}^n$ are *equivalent*.

Similarly to what we have seen for one-dimensional dynamical systems, there conditions that simpler to verify than Lipschitz continuity.

Lemma 1. If $\boldsymbol{f}(\boldsymbol{x})$ is of class C^1 in a compact convex domain $D \subset \mathbb{R}^n$, then $\boldsymbol{f}(\boldsymbol{x})$ is Lipschitz continuous in D.

The proof of this lemma is provided in Appendix B.

Lemma 2. Let $\boldsymbol{f}(\boldsymbol{y}, t)$ be of class C^1 (continuously differentiable) in $D \subseteq \mathbb{R}^n$. If $\boldsymbol{f}(\boldsymbol{y}, t)$ has bounded derivatives $\partial f_i / \partial y_j$ then $\boldsymbol{f}(\boldsymbol{y}, t)$ is Lipschitz continuous in D.

[1] As shown in Figure 1, the "exit time" τ depends on D, $\boldsymbol{f}(x)$ and $\boldsymbol{x}_0$.

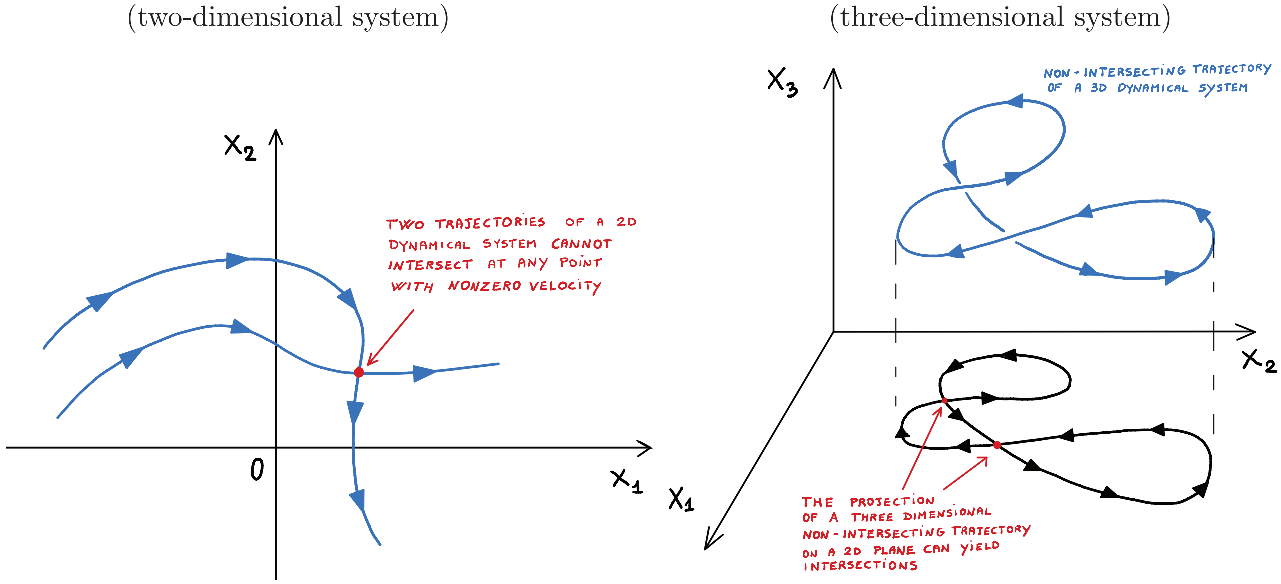


Figure 2: It is impossible for two trajectories to intersect with nonzero velocity any point in phase space. This would make $\boldsymbol{f}(\boldsymbol{x})$ non-unique at such point and also violate the existence and uniqueness Theorem 1.

Global solutions. If $\boldsymbol{f}(\boldsymbol{x})$ is Lipschitz continuous on the entire space $\mathbb{R}^n$ then the solution to the initial value problem (1) is *global*. This means that the solution exists and is unique for all $t \geq 0$. In fact, $x(t)$ never exits the domain in which $\boldsymbol{f}(\boldsymbol{x})$ is Lipschitz continuous, and therefore we can extend τ in Theorem 1 to infinity. It is important to emphasize that existence and uniqueness of the solution to (1) has nothing to do with the smoothness of $\boldsymbol{f}(\boldsymbol{x})$ but rather with the rate at which $\boldsymbol{f}(\boldsymbol{x})$ grows or decays.

Example: The solution to both dynamical systems (8) and (11) is global in time, meaning that exists and is unique for all $t \geq 0$. In fact the right hand side of such systems if globally Lipschitz in $\mathbb{R}^2$ and $\mathbb{R}^n$, respectively.

Flow generated by nonlinear dynamical systems. The solution of initial value problem (1) depends on $\boldsymbol{f}(\boldsymbol{x})$ and $\boldsymbol{x}_0$. As before, we will denote the dependence of the solution $\boldsymbol{x}(t)$ on $\boldsymbol{x}_0$ as $\boldsymbol{X}(t, \boldsymbol{x}_0)$, i.e.,

$$\boldsymbol{x}(t) = \boldsymbol{X}(t, \boldsymbol{x}_0). \tag{13}$$

Similarly to what we have seen for one-dimensional systems, it is not possible for two solutions corresponding to two different initial conditions to intersect at any finite time t. Otherwise we could use such intersection point as initial condition for (1) and conclude that there are two orbits emanating from such point (see Figure 2), hence violating the existence and uniqueness Theorem 1. This implies that $\boldsymbol{X}(t, \boldsymbol{x}_0)$ is invertible at each finite time[2] (see below), i.e., we can always identify which "particle" $\boldsymbol{x}_0$ sits at location $\boldsymbol{x}(t) = \boldsymbol{X}(t, \boldsymbol{x}_0)$ at time t. Moreover, it is impossible for two "particles" to collide at any finite time, or for one particle to split into two or more particles (Figure 2).

Theorem 2 (Regularity of the flow with respect to $\boldsymbol{x}_0$)**.** Let $D \subset \mathbb{R}^n$ be an open set, $\boldsymbol{x}_0 \in D$. If $\boldsymbol{f} : D \to \mathbb{R}^n$ is Lipschitz continuous in D then the flow $\boldsymbol{X}(t, \boldsymbol{x}_0)$ generated by the initial value problem (1), is continuous in $\boldsymbol{x}_0$. Moreover, If $\boldsymbol{f}(\boldsymbol{x})$ is of class $C^k(D)$ (continuously differentiable k-times in D) in D then $\boldsymbol{X}(t, \boldsymbol{x}_0)$ is of class $C^k(D)$ relative to $\boldsymbol{x}_0$ (continuously differentiable k-times with respect to $\boldsymbol{x}_0$).

[2] Solutions corresponding to different initial conditions can, however, intersect at $t = \infty$, e.g., when there exist an attracting set.

In summary, Theorem 2 states that the smoother $\boldsymbol{f}(\boldsymbol{x})$, the smoother the dependency of the flow $\boldsymbol{X}(t, \boldsymbol{x}_0)$ on $\boldsymbol{x}_0$. The n-dimensional function $\boldsymbol{X}(t, \boldsymbol{x}_0)$ is called *flow generated by the dynamical system* (1), and it represents the full set of solutions to (1) for each initial condition $\boldsymbol{x}_0$.

Theorem 3 (Regularity of the flow in time)**.** Let $D \subset \mathbb{R}^n$ be an open set, $\boldsymbol{x}_0 \in D$. if $\boldsymbol{f}(\boldsymbol{x})$ is of class C^k in D (continuously differentiable k-times in D with continuous derivative), then $\boldsymbol{X}(t, \boldsymbol{x}_0)$ is of class C^{k+1} in time for all $t \in [0, \tau[$, where τ is the time at which $X(t, \boldsymbol{x}_0)$ exits the domain D.

Properties of the flow: The flow $\boldsymbol{X}(t, \boldsymbol{x}_0)$ satisfies the following properties

- $\boldsymbol{X}(0, \boldsymbol{x}_0) = \boldsymbol{x}_0$. This means that at $t = 0$ the mapping $\boldsymbol{X}(0, \boldsymbol{x}_0) = \boldsymbol{x}_0$ is the identity.
- $\boldsymbol{X}(t, \boldsymbol{x}_0)$ is invertible for all t for which the solution to the initial value problem (1) exists and is unique.
- $\boldsymbol{X}(t, \boldsymbol{x}_0)$ satisfies the composition rule $\boldsymbol{X}(t+s, \boldsymbol{x}_0) = \boldsymbol{X}(t, \boldsymbol{X}(s, \boldsymbol{x}_0)) = \boldsymbol{X}(s, \boldsymbol{X}(t, \boldsymbol{x}_0))$. This property is called "semi-group property" of the flow and it follows from the fact that we can restart integration of the ODE (1) at time t (or time s) from the new initial condition $\boldsymbol{X}(t, \boldsymbol{x}_0)$ (or $\boldsymbol{X}(s, \boldsymbol{x}_0)$) to get to the final integration time $s + t$. Again, this property holds because of the existence and uniqueness theorem 1.
- The flow $\boldsymbol{X}(t, \boldsymbol{x}_0)$ satisfies the system of first-order PDEs

$$\begin{cases} \dfrac{\partial \boldsymbol{X}(t, \boldsymbol{x}_0)}{\partial t} - \boldsymbol{f}(\boldsymbol{x}_0) \cdot \nabla_{\boldsymbol{x}_0} \boldsymbol{X}(t, \boldsymbol{x}_0) = 0 \\ \boldsymbol{X}(0, x_0) = \boldsymbol{x}_0 \end{cases} \tag{14}$$

- The inverse flow $\boldsymbol{X}_0(t, \boldsymbol{x})$ satisfies the system of first-order PDEs

$$\begin{cases} \dfrac{\partial \boldsymbol{X}_0(t, \boldsymbol{x})}{\partial t} + \boldsymbol{f}(\boldsymbol{x}) \cdot \nabla_{\boldsymbol{x}} \boldsymbol{X}_0(t, \boldsymbol{x}) = 0 \\ \boldsymbol{X}_0(0, x) = \boldsymbol{x} \end{cases} \tag{15}$$

Geometric approach. The flow $\boldsymbol{X}(t, \boldsymbol{x}_0)$ generated by the ODE system (1) maps any initial condition $\boldsymbol{x}_0$ to the solution of the ODE at time t. If we think as $\boldsymbol{x}_0$ as the initial position of a particle in $\mathbb{R}^n$, then from elementary mechanics we know that $d\boldsymbol{X}(0, \boldsymbol{x}_0)/dt = \boldsymbol{f}(\boldsymbol{x}_0)$ represents the velocity of such particle. Hence, given $\boldsymbol{f}(\boldsymbol{x})$ we can immediately sketch the *vector field*[3] associated with the dynamical system, which represents where a particle sitting at any particular location in phase space is heading to To clarify this idea, consider the following two-dimensional dynamical system

$$\begin{cases} \dfrac{dx_1}{dt} = f_1(x_1, x_2) \\ \dfrac{dx_2}{dt} = f_2(x_1, x_2) \end{cases} \tag{16}$$

In Figure 3 we plot the velocity vector

$$\boldsymbol{f}(\boldsymbol{x}) = (f_1(\boldsymbol{x}), f_2(\boldsymbol{x})) \tag{17}$$

[3] A vector field is a vector that is continuously indexed by one or more variables. For one-dimensional dynamical systems the vector field $f(x)$ is indexed by coordinate x, and it is represented by a vector sitting on a line. For two-dimensional system the vector $\boldsymbol{f}(\boldsymbol{x}) = (f_1(\boldsymbol{x}), \boldsymbol{f}_2(\boldsymbol{x}))$ is a vector with two components: one along x_1 and the other along x_2. For three-dimensional systems the vector field has three components, and so on so forth.

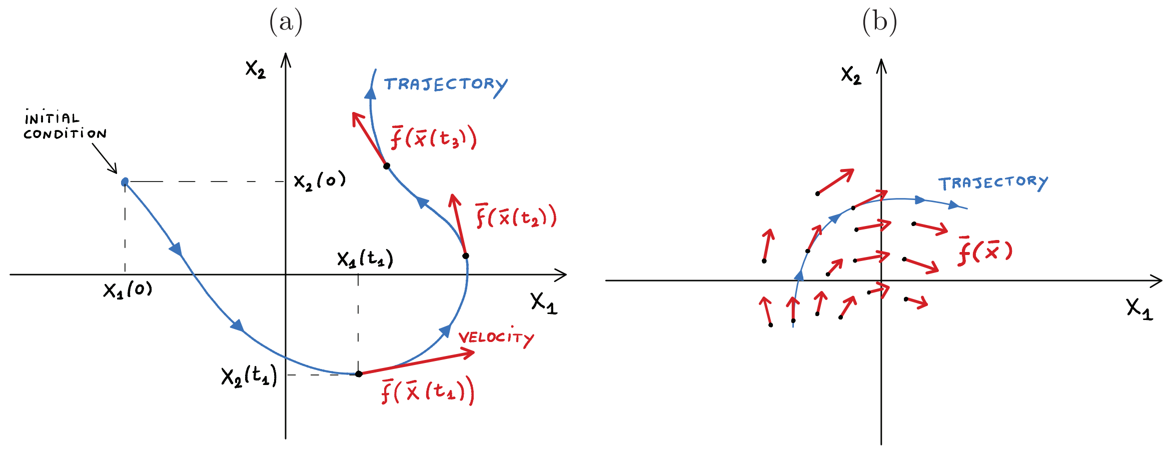


Figure 3: Geometric approach in 2D. (a) Sketch of a trajectory in the phase plane and associated velocity vectors $\boldsymbol{f}(\boldsymbol{x}(t))$. This process can be reversed in the sense that we can also guess how the trajectory $\boldsymbol{x}(t)$ looks like by plotting a bunch of velocity vectors $\boldsymbol{f}(\boldsymbol{x})$ evaluated at different points $\boldsymbol{x}$ in the plane (b).

along a trajectory $\boldsymbol{x}(t)$. This process can be reversed in the sense that we can guess how the trajectory $\boldsymbol{x}(t)$ looks like by plotting a sufficiently large number of velocity vectors $\boldsymbol{f}(\boldsymbol{x})$ evaluated at points nearby a point of interest in the phase space (see Figure 4(b)).

Example: In Figure 4 we plot the vector fields and corresponding trajectories defined by following two-dimensional dynamical systems

$$\begin{cases} \dot{x}_1 = 2x_1x_2 - 1 \\ \dot{x}_2 = -x_1^2 - x_2^2 + 10 \end{cases} \tag{18}$$

and

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = -\sin(x_1) - \dfrac{1}{10}x_2 \end{cases} \qquad \text{(pendulum with friction)} \tag{19}$$

As we shall see hereafter, the curves $\dot{x}_1 = 0$ and $\dot{x}_2 = 0$ are called *nullclines*. Fixed points are at the intersection of nullclines.

Fixed points. If the velocity vector $\boldsymbol{f}(\boldsymbol{x})$ is equal to zero at some point $\boldsymbol{x}^* \in \mathbb{R}^n$ then any particle placed at that point won't move at all as time increases. These points are called *fixed points* (or *equilibria*) of the dynamical system (1). Mathematically, we can define a fixed point $\boldsymbol{x}^* \in \mathbb{R}^n$ as

$$\boldsymbol{X}(t, \boldsymbol{x}^*) = \boldsymbol{x}^* \qquad \text{for all } t \geq 0. \tag{20}$$

By differentiating this previous equation with respect to time we obtain

$$\frac{\partial \boldsymbol{X}(t, \boldsymbol{x}^*)}{\partial t} = \boldsymbol{f}(\boldsymbol{X}(t, \boldsymbol{x}^*)) = \boldsymbol{f}(\boldsymbol{x}^*) = 0. \tag{21}$$

Therefore, the fixed points of the system (1) are solutions to the nonlinear system of equations

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{0}. \tag{22}$$

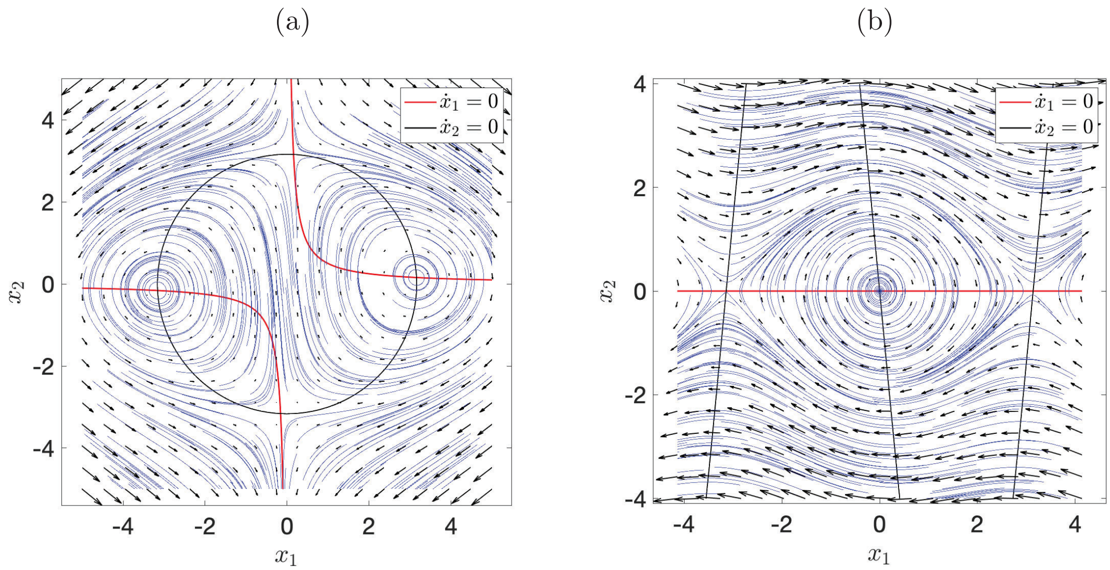


Figure 4: Vector field and trajectories generated by the two-dimensional nonlinear dynamical system (18) (Figure (a)), and (19) (Figure (b)). Shown are also the nullclines for both systems.

This system can be written as

$$\begin{cases} f_1(x_1, \ldots, x_n) = 0 \\ f_2(x_1, \ldots, x_n) = 0 \\ \quad \vdots \\ f_n(x_1, \ldots, x_n) = 0 \end{cases} \tag{23}$$

In this form, it is clear that the fixed points (if any) of the system (1) lie at the intersection of zero level sets[4] of n functions f_j each one of which is n-dimensional. In two-dimensions such zero level sets are identified by the intersection of two surfaces $f_1(x_1, x_2)$ and $f_1(x_1, x_2)$ with the (x_1, x_2) plane

$$\begin{cases} f_1(x_1, x_2) = 0 \\ f_2(x_1, x_2) = 0 \end{cases} \tag{24}$$

In two dimensions, the zero level sets of $f_1(x_1, x_2)$ and $f_1(x_1, x_2)$ are called *nullclines*. The reason for the definition is that the vector field $\boldsymbol{f}(x_1, x_2)$ is vertical at all points sitting on the nullcline $f_1(x_1, x_2) = 0$, and horizontal at all point sitting on the nullcline $f_2(x_1, x_2) = 0$. Correspondingly, the trajectories intersect the nullclines $f_1(x_1, x_2) = 0$ and $f_2(x_1, x_2) = 0$ vertically and horizontally, respectively (see Figure 4).

Example: Let us calculate the nullclines of the dynamical system (18). The first nullcline is

$$f_1(x_1, x_2) = 2x_1x_2 - 1 = 0 \qquad \Leftrightarrow \qquad x_2 = \frac{1}{2x_1} \qquad (\text{nullcline } \dot{x}_1 = 0), \tag{25}$$

i.e., the hyperbola depicted in red in Figure 4(a). The second nullcline is a circle with radius $\sqrt{10}$ centered at the origin

$$f_2(x_1, x_2) = -x_1^2 - x_2^2 + 10 \qquad \Leftrightarrow \qquad x_1^2 + x_2^2 = 10 \qquad (\text{nullcline } \dot{x}_2 = 0). \tag{26}$$

[4]The calculation of the fixed points can be done analytically only for prototype dynamical systems. In general, computing the fixed points requires a root-finding numerical algorithm for nonlinear systems of algebraic equations, e.g., the Newton's method.

This is depicted in black in Figure 4(a). We also see that the trajectories of the system intersect the nullcline $\dot{x}_2 = 0$ (black curve) horizontally. In fact, such nullcline represents the set of points in the phase plane where the velocity has zero vertical component. Similarly the trajectories intersect the nullcline $\dot{x}_1 = 0$ (red curve) vertically. The four fixed points of the system are at the intersection of the nullclines and can be computed analytically (left as exercise).

Example: Let us calculate the nullclines and the fixed points of the dynamical system (19). The first nullcline is

$$f_1(x_1, x_2) = x_2 = 0 \qquad \Leftrightarrow \qquad x_2 = 0 \qquad (\text{nullcline } \dot{x}_1 = 0), \tag{27}$$

Such nullcline is plotted in red in Figure 4(b). The second nullcline is

$$f_2(x_1, x_2) = \sin(x_1) - x_2/10 \qquad \Leftrightarrow \qquad x_2 = 10\sin(x_1) \qquad (\text{nullcline } \dot{x}_2 = 0), \tag{28}$$

and it is plotted in black in Figure 4(b).. The fixed points are at the intersection of the nullclines. In this case we obtain two physically different fixed points:

$$x_1^* = (0, 0) \qquad x_2^* = (\pi, 0) \tag{29}$$

corresponding to a pendulum in a vertical position, i.e., $x_1 = 0$ or $x_1 = \pi$ with zero velocity $x_2 = 0$.

Analysis of fixed points. A quick look at the phase portraits in Figure 4 suggests that the dynamics in a neighborhood of a fixed point can be quite different. Such dynamics can often be computed via a linearization process that is similar to the process we used in one-dimensional dynamical systems. The idea is "zoom-in" on a fixed point $\boldsymbol{x}^*$ and compute the orbits of the dynamical systems in a small neighborhood of $\boldsymbol{x}^*$ by solving a linearized version of the system (1). To this end, consider an initial condition $\boldsymbol{x}_0$ that is very close $\boldsymbol{x}^*$, and define the perturbation

$$\boldsymbol{\eta}(t, \boldsymbol{x}_0) = \boldsymbol{X}(t, \boldsymbol{x}_0) - \boldsymbol{x}^*. \tag{30}$$

By expanding $\boldsymbol{f}(\boldsymbol{X}(t, \boldsymbol{x}_0)) = \boldsymbol{f}(\boldsymbol{x}^* + \boldsymbol{\eta}(t, \boldsymbol{x}_0))$ in a neighborhood of $\boldsymbol{x}^*$, i.e., for small $\boldsymbol{\eta}(t, \boldsymbol{x}_0)$ we obtain

$$\boldsymbol{f}(\boldsymbol{x}^* + \boldsymbol{\eta}(t, \boldsymbol{x}_0)) = \underbrace{\boldsymbol{f}(\boldsymbol{x}^*)}_{=\boldsymbol{0}} + \boldsymbol{J}_{\boldsymbol{f}}(\boldsymbol{x}^*)\boldsymbol{\eta}(t, \boldsymbol{x}_0) + \cdots, \tag{31}$$

where

$$\boldsymbol{J}_{\boldsymbol{f}}(\boldsymbol{x}^*) = \begin{bmatrix} \dfrac{\partial f_1(\boldsymbol{x}^*)}{\partial x_1} & \cdots & \dfrac{\partial f_1(\boldsymbol{x}^*)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_n(\boldsymbol{x}^*)}{\partial x_1} & \cdots & \dfrac{\partial f_n(\boldsymbol{x}^*)}{\partial x_n} \end{bmatrix} \tag{32}$$

is the Jacobian[5] of $\boldsymbol{f}(\boldsymbol{x})$ evaluated at the fixed point $\boldsymbol{x}^*$. Hence, the first-order approximation of the nonlinear dynamical system (1) at $\boldsymbol{x}^*$ can be written as

$$\begin{cases} \dfrac{d\boldsymbol{\eta}}{dt} = \boldsymbol{J}_{\boldsymbol{f}}(\boldsymbol{x}^*)\boldsymbol{\eta} \\ \boldsymbol{\eta}(0, \boldsymbol{x}_0) = \boldsymbol{x}_0 - \boldsymbol{x}^* \end{cases} \tag{33}$$

[5] The Jacobian of $\boldsymbol{f}(\boldsymbol{x})$ is a matrix-valued function that takes in a function $\boldsymbol{f}(\boldsymbol{x})$ and it returns a $n \times n$ matrix-valued function. The entries of such Jacobian matrix are functions. Of course, if we evaluate the Jacobian of $\boldsymbol{f}(\boldsymbol{x})$ at a specific point $\boldsymbol{x}^*$ then we obtain a matrix with real entries (provided $\boldsymbol{f}$ is real).

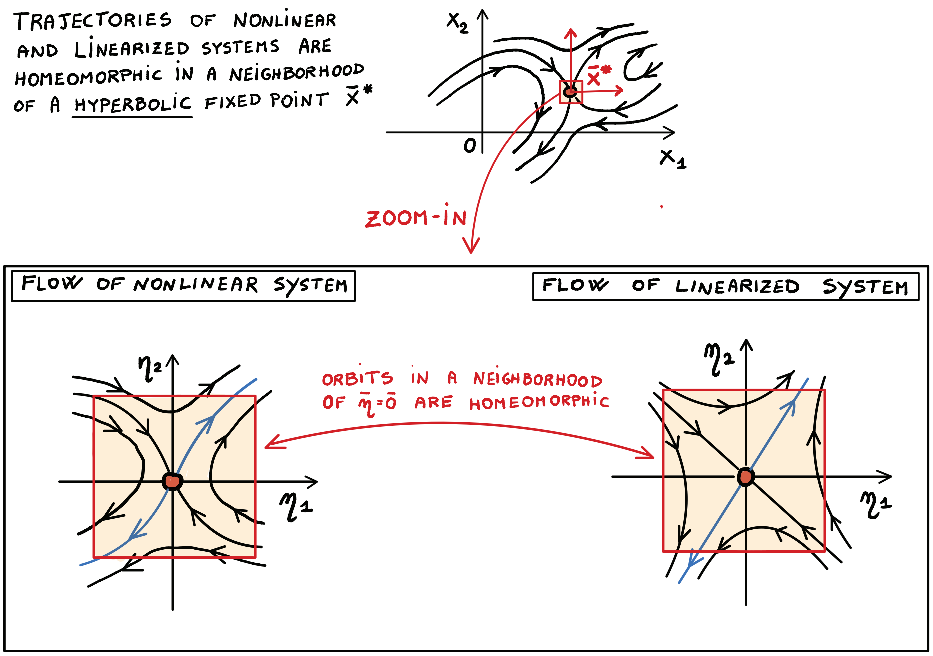


Figure 5: Geometric meaning of the Hartmman-Grobman Theorem 4. The trajectories of a nonlinear system in a neighborhood of any hyperbolic fixed point are homeomorphic to the trajectories of the linearized system at $\boldsymbol{x}^*$. This means that the trajectories of the nonlinear and linearized system are not exactly the same in the aforementioned neighborhood of $\boldsymbol{x}^*$, but they can be mapped to each other by a continuous transformation that has a continuous inverse.

Theorem 4 (Hartman-Grobman)**.** Let $\boldsymbol{x}^* \in \mathbb{R}^n$ be a fixed point of the dynamical system (1). If the Jacobian (32) has no eigenvalue with zero real part then there exists a homeomorphism (i.e., continuous invertible mapping with continuous inverse) defined on some neighborhood of $\boldsymbol{x}^*$ that takes orbits of the system (1) and maps them into orbits of the linearized system (30)-(33). The mapping preserves the orientation of the orbits.

Remark: Theorem (4) is saying that if $\boldsymbol{x}^*$ is a hyperbolic[6] fixed point then the flow of the nonlinear dynamical system (1) nearby $\boldsymbol{x}^*$ is "homemorphic" (i.e., it can be mapped back and forth by a continuous transformation) to the flow of the linearized system (30)-(33).

At this point it is natural to ask the following questions:

1. To study the flow nearby a hyperbolic fixed point of nonlinear system we need to compute the flow of the linear system (33). Is there a general method to compute such flow? Note that flows of linear systems are very important on their own as there are many system that are actually are linear (e.g., the discretized PDE system (11)).

2. What happens if the fixed point is non-hyperbolic? As we will see, in this case we need to use a generalization of the Hartman-Grobman theorem known as *center manifold* theorem.

[6]A fixed point $\boldsymbol{x}^*$ is called *hyperbolic* if the Jacobian of $\boldsymbol{J_f}(\boldsymbol{x}^*)$ has no eigenvalue with zero real part.

Appendix A: Elementary numerical methods for systems of ODEs

As before, we can re-write the Cauchy problem (1) as an integral equation

$$\boldsymbol{x}(t) = \boldsymbol{x}(0) + \int_0^t \boldsymbol{f}(\boldsymbol{x}(s))ds. \tag{34}$$

This form is quite handy to derive numerical methods to solve (1) based on quadrature rules applied to the one-dimensional integral at the right hand side. For instance, consider a partition of the $[0, T]$ into an evenly-spaced grid points such that $t_{i+1} = t_i + \Delta t$, and write (34) within each time interval

$$\boldsymbol{x}(t_{i+1}) = \boldsymbol{x}(t_i) + \int_{t_i}^{t_{i+1}} \boldsymbol{f}(\boldsymbol{x}(s))ds. \tag{35}$$

Explicit midpoint method. By approximating the integral at the right hand side of (35), e.g., using the midpoint rule yields

$$\int_{t_i}^{t_{i+1}} \boldsymbol{f}(\boldsymbol{x}(s))ds \simeq \Delta t \boldsymbol{f}\left(\boldsymbol{x}\left(t_i + \frac{\Delta t}{2}\right)\right) \tag{36}$$

At this point, we can approximate $\boldsymbol{x}(t_i + \Delta t/2)$ using the Euler forward method

$$\boldsymbol{x}\left(t_i + \frac{\Delta t}{2}\right) \simeq \boldsymbol{x}(t_i) + \frac{\Delta t}{2}\boldsymbol{f}(\boldsymbol{x}(t_i)) \tag{37}$$

to obtain the *explicit midpoint method*

$$\boldsymbol{x}(t_{i+1}) = \boldsymbol{x}(t_i) + \Delta t \boldsymbol{f}\left(\boldsymbol{x}(t_i) + \frac{\Delta t}{2}\boldsymbol{f}(\boldsymbol{x}(t_i))\right). \tag{38}$$

The explicit midpoint method is a one-step method that belongs to the larger class of Runge-Kutta methods.

Appendix B: Equivalent norms in $\mathbb{R}^n$

As is well known, all norms defined in a finite-dimensional vector space such as $\mathbb{R}^n$ are *equivalent*. This means that if we pick two arbitrary norms in $\mathbb{R}^n$, say $\|\cdot\|_a$ and $\|\cdot\|_b$, then there exist two numbers C_1 and C_2 such that

$$C_1 \|\boldsymbol{x}\|_a \leq \|\boldsymbol{x}\|_b \leq C_2 \|\boldsymbol{x}\|_a \quad \text{for all} \quad \boldsymbol{x} \in \mathbb{R}^n. \tag{39}$$

The most common norms in $\mathbb{R}^n$ are

$$\|\boldsymbol{x}\|_\infty = \max_{k=1,..,n} |x_k|, \tag{40}$$

$$\|\boldsymbol{x}\|_1 = \sum_{k=1}^n |x_k|, \tag{41}$$

$$\|\boldsymbol{x}\|_2 = \left(\sum_{k=1}^n |x_k|^2\right)^{1/2}, \tag{42}$$

$$\vdots \tag{43}$$

$$\|\boldsymbol{x}\|_p = \left(\sum_{k=1}^n |y_k|^p\right)^{1/p} \qquad p \in \mathbb{N} \setminus \{\infty\}. \tag{44}$$

Based on these definitions it can be shown that, e.g., that

$$\|\boldsymbol{x}\|_\infty \le \|\boldsymbol{x}\|_1 \le n\,\|\boldsymbol{x}\|_\infty\,, \tag{45}$$
$$\|\boldsymbol{x}\|_2 \le \|\boldsymbol{x}\|_1 \le \sqrt{n}\,\|\boldsymbol{x}\|_2\,, \tag{46}$$
$$\|\boldsymbol{x}\|_\infty \le \|\boldsymbol{x}\|_2 \le \sqrt{n}\,\|\boldsymbol{x}\|_\infty\,. \tag{47}$$

Therefore if the $\boldsymbol{f}(\boldsymbol{x})$ is Lipschitz continuous in D with respect to the 1-norm, i.e.,

$$\|\boldsymbol{f}(\boldsymbol{x}_1)-\boldsymbol{f}(\boldsymbol{x}_2)\|_1 \le L_1\,\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_1 \quad \text{for all} \quad \boldsymbol{x}_1,\boldsymbol{x}_2\in D \tag{48}$$

then it is also Lipschitz continuous in D with respect to the uniform norm. In fact, by using (45) we have

$$\|\boldsymbol{f}(\boldsymbol{x}_1)-\boldsymbol{f}(\boldsymbol{x}_2)\|_\infty \le \underbrace{L_1 n}_{L_\infty}\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_\infty\,. \tag{49}$$

Similarly, by using (46), we see that if $\boldsymbol{f}$ is Lipschitz in D with respect to the 1-norm then $\boldsymbol{f}$ is Lipschitz in D with respect to the 2-norm.

Appendix B: Proof of Lemma 1

Let $D\subseteq\mathbb{R}^n$ be a compact convex domain and let

$$M=\max_{\boldsymbol{x}\in D}\left|\frac{\partial f_j(\boldsymbol{x})}{\partial x_i}\right|. \tag{50}$$

Clearly M exists and is finite because we assumed that D is compact and that $\boldsymbol{f}$ is of class C^1 in D[7]. Consider two points $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ in D, and the line that connects $\boldsymbol{x}_1$ to $\boldsymbol{x}_2$, i.e.,

$$\boldsymbol{z}(s)=(1-s)\boldsymbol{x}_1+s\boldsymbol{x}_2 \qquad s\in[0,1]. \tag{51}$$

Since D is convex, we have that the line $\boldsymbol{z}(s)$ lies entirely within D. Therefore we can use the mean value theorem applied to the one-dimensional function $f_i(\boldsymbol{z}(s))$ ($s\in[0,1]$) to obtain

$$f_i(\boldsymbol{x}_2)-f_i(\boldsymbol{x}_1)=\nabla f_i(\boldsymbol{z}(s^*),t)\cdot(\boldsymbol{x}_2-\boldsymbol{x}_1) \qquad \text{for some } s^*\in[0,1]. \tag{52}$$

By taking the absolute value and using the Cauchy-Schwartz inequality we obtain

$$\begin{aligned}|f_i(\boldsymbol{x}_1)-f_i(\boldsymbol{x}_1)|^2 &=\left|\sum_{j=1}^n\frac{\partial f_i(\boldsymbol{z}(s^*))}{\partial x_j}(x_{2j}-x_{1j})\right|^2\\ &\le\left|\sum_{j=1}^n\frac{\partial f_i(\boldsymbol{z}(s^*))}{\partial x_j}\right|^2\left|\sum_{j=1}^n(x_{2j}-x_{1j})\right|^2\\ &\le nM^2\,\|\boldsymbol{x}_2-\boldsymbol{x}_1\|_2^2\,.\end{aligned} \tag{53}$$

This implies that

$$\|\boldsymbol{f}(\boldsymbol{x}_2)-\boldsymbol{f}(\boldsymbol{x}_1)\|_2 \le \underbrace{nM}_{L_2}\|\boldsymbol{x}_2-\boldsymbol{x}_1\|_2\,. \tag{54}$$

i.e., $\boldsymbol{f}(\boldsymbol{y},t)$ is Lipschitz continuous in the 2-norm, or any other norm that is equivalent to the 2-norm. In particular, by using the inequalities (45)-(47) we have that $\boldsymbol{f}(\boldsymbol{x})$ is Lipschitz continuous relative to the 1-norm.

[7] A compact domain is by definition bounded and closed. The minimum and maximum of a continuous function in defined on a compact domain is attained at some points within the domain or on its boundary. Note that this is not true if the domain is not compact. For example, the function $f(y)=1/y$ is continuously differentiable on $]0,1]$ (bounded domain by not compact), but the function is unbounded on $]0,1]$.

Linear dynamical systems

Consider the following n-dimensional linear dynamical system

$$\begin{cases} \dfrac{d\boldsymbol{x}}{dt} = \boldsymbol{A}\boldsymbol{x} \\ \boldsymbol{x}(0) = \boldsymbol{x}_0 \end{cases} \tag{1}$$

where $\boldsymbol{x}(t) = [x_1(t) \cdots x_n(t)]^T$ is a column vector of phase variables, and $\boldsymbol{A} \in M_{n\times n}(\mathbb{R})$ is a $n \times n$ matrix with real coefficients.

It is immediate to show that the linear function $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$ is Lipschitz continuous on $\mathbb{R}^n$. In fact,

$$\|\boldsymbol{A}\boldsymbol{x}_1 - \boldsymbol{A}\boldsymbol{x}_2\| = \|\boldsymbol{A}(\boldsymbol{x}_1 - \boldsymbol{x}_2)\| \leq \|\boldsymbol{A}\| \, \|\boldsymbol{x}_1 - \boldsymbol{x}_2\| \tag{2}$$

for any matrix norm $\|\boldsymbol{A}\|$ that is compatible with the vector norm $\|\boldsymbol{x}\|$ (see Appendix B). Alternatively, note that the function $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$ has bounded derivatives for all $\boldsymbol{x} \in \mathbb{R}^n$ (provided the entries of the matrix $\boldsymbol{A}$ are finite), i.e.,

$$\frac{\partial f_i(\boldsymbol{x})}{\partial x_j} = A_{ij} < \infty \qquad \text{for all } i, j = 1, \ldots, n. \tag{3}$$

Therefore by Lemma 2 in the course note 2, we immediately conclude that the solution of (1) is global, i.e., it exists and is unique for all $t \geq 0$. Moreover, since $\boldsymbol{A}\boldsymbol{x}$ is continuously differentiable an infinite number of times on $\mathbb{R}^n$, then by Theorem 2 and Theorem 3 in the course note 3 we have that the flow $\boldsymbol{X}(t, \boldsymbol{x}_0)$ generated by (1) is of class C^∞ in t and $\boldsymbol{x}_0$.

Fixed points. The fixed points of the linear dynamical system (1) are solutions of the linear equation

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}_{\mathbb{R}^n}, \tag{4}$$

i.e., they lie at the intersection of n hyper-planes passing through the origin in $\mathbb{R}^n$. Such hyper-planes are defined by the linear equations

$$A_{j1}x_1 + A_{j2}x_2 + \cdots + A_{jn}x_n = 0, \qquad j = 1, \ldots, n. \tag{5}$$

Clearly, if the matrix $\boldsymbol{A}$ is invertbile then we have a unique fixed point at

$$\boldsymbol{x}^* = \boldsymbol{0}_{\mathbb{R}^n}. \tag{6}$$

On the other hand, if the matrix $\boldsymbol{A}$ is not invertible then we have an infinite number of fixed points, i.e., all points in the nullspace[1] of $\boldsymbol{A}$ are fixed points. For example, the fixed points of the 2D linear dynamical system defined by the rank 1 matrix

$$\boldsymbol{A} = \begin{bmatrix} 5 & 1 \\ 10 & 2 \end{bmatrix} \tag{7}$$

are obtained by solving

$$\begin{bmatrix} 5 & 1 \\ 10 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \Rightarrow \qquad x_2 = 2x_1 \tag{8}$$

Hence, in this case we have an infinite number of fixed points sitting on a line with slope 2 passing through the origin of the phase plane (x_1, x_2).

[1] Recall that the nullspace of a matrix $\boldsymbol{A}$ is the set of vectors that are sent to the zero vector by applying $\boldsymbol{A}$. The nullspace of an $n \times n$ matrix is a vector subspace of $\mathbb{R}^n$.

Flow generated by linear dynamical systems. As shown in Appendix C, the analytical solution of the initial value problem (1) can be formally expressed in terms of a matrix exponential[2], i.e.,

$$\boldsymbol{X}(t, \boldsymbol{x}_0) = e^{t\boldsymbol{A}}\boldsymbol{x}_0. \tag{10}$$

This expression shows that the flow map is indeed of class C^∞ in both x_0 and t, as anticipated above. Hereafter we take a linear algebraic approach to the problem of solving the linear system of ODEs (1), i.e., we focus on linear algebraic techniques to compute the matrix exponential $e^{t\boldsymbol{A}}$ explicitly in terms of the spectral properties (eigenvalues, eigenvectors and generalized eigenvectors) of the matrix $\boldsymbol{A}$.

Computation of the matrix exponential. The matrix exponential in appearing in (10) can be written explicitly in terms of the eigenvalues and the eigenvectors (or generalized eigenvectors) of the matrix $\boldsymbol{A}$. In Appendix A we provide a thorough review of the matrix eigenvalue problem, including calculation of the eigenvalues, eigenvectors and generalized eigenvectors of a matrix. Please read through Appendix A very carefully, as everything that is discussed hereafter assumes that you are familiar with eigenvalues, eigenspaces, generalized eigenvectors, and similarity transformations. The computation of the matrix the matrix exponential $e^{t\boldsymbol{A}}$, and therefore the solution (10) of the linear system (1), differs depending on whether or not

- the matrix $\boldsymbol{A}$ is diagonalizable,
- the matrix $\boldsymbol{A}$ is not diagonalizable.

For a definition diagonalizable and non-diagonalizable matrices see Appendix A. As we will see, the non-diagonalizable case includes the diagonalizable one. Therefore, in principle, we could just develop the formula for the matrix exponential in the case where $\boldsymbol{A}$ is not-diagonalizable. However, for clarity of exposition, here we present the two cases separately.

Matrix exponential for diagonalizable matrices. If $\boldsymbol{A}$ is diagonalizable then there exists a set of n distinct eigenvectors $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ and a similarity transformation $\boldsymbol{P}$ such that (see Appendix A)

$$\boldsymbol{A}\boldsymbol{P} = \boldsymbol{P}\boldsymbol{\Lambda}, \tag{11}$$

where

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix} \tag{12}$$

is a diagonal matrix that contains all eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$ of $\boldsymbol{A}$ and

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix} \tag{13}$$

is a matrix that contains all eigenvectors of $\boldsymbol{A}$. Each vector $\boldsymbol{v}_i$ in (13) is a column vector. Since the matrix $\boldsymbol{P}$ is invertible we have

$$\boldsymbol{A} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{-1}. \tag{14}$$

[2]Recall that the matrix exponential is formally defined by the power series

$$e^{t\boldsymbol{A}} = I + t\boldsymbol{A} + \frac{t^2}{2}\boldsymbol{A}^2 + \cdots = \sum_{k=0}^{\infty} \frac{t^k \boldsymbol{A}^k}{k!}, \tag{9}$$

which converges uniformly for all $t \geq 0$.

This matrix factorization is very effective when computing the matrix powers appearing in the definition of the matrix exponential (9). In fact,

$$\boldsymbol{A}^2 = \boldsymbol{P}\boldsymbol{\Lambda}\underbrace{\boldsymbol{P}^{-1}\boldsymbol{P}}_{\boldsymbol{I}}\boldsymbol{\Lambda}\boldsymbol{P}^{-1} = \boldsymbol{P}\boldsymbol{\Lambda}^2\boldsymbol{P}^{-1}. \tag{15}$$

Similarly,

$$\boldsymbol{A}^3 = \boldsymbol{P}\boldsymbol{\Lambda}^3\boldsymbol{P}^{-1}, \qquad \cdots, \boldsymbol{A}^k = \boldsymbol{P}\boldsymbol{\Lambda}^k\boldsymbol{P}^{-1}. \tag{16}$$

This implies that

$$e^{t\boldsymbol{A}} = \boldsymbol{P}\left(I + t\boldsymbol{\Lambda} + \frac{t^2}{2}\boldsymbol{\Lambda}^2 + \cdots\right)\boldsymbol{P}^{-1} = \boldsymbol{P}e^{t\boldsymbol{\Lambda}}\boldsymbol{P}^{-1}. \tag{17}$$

The exponential the diagonal matrix $\boldsymbol{\Lambda}$ in (12) is easily obtained as

$$e^{t\boldsymbol{\Lambda}} = \begin{bmatrix} e^{t\lambda_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & e^{t\lambda_n} \end{bmatrix}. \tag{18}$$

Hence, steps to compute the analytical solution (1) in the case where $\boldsymbol{A}$ is diagonalizable are:

1. Compute the eigenvalues and the eigenvectors of $\boldsymbol{A}$;
2. Construct the matrix $\boldsymbol{P}$ in (13) and the matrix exponential (18);
3. Compute the analytical solution of (1) using matrix-vector products

$$\boxed{\boldsymbol{X}(t, \boldsymbol{x}_0) = \boldsymbol{P}e^{t\boldsymbol{\Lambda}}\boldsymbol{P}^{-1}\boldsymbol{x}_0.} \tag{19}$$

Matrix exponential for non-diagonalizable matrices. If the matrix $\boldsymbol{A}$ is not diagonalizable then there exist a similarity transformation $\boldsymbol{P}$ such that

$$\boldsymbol{A}\boldsymbol{P} = \boldsymbol{P}\boldsymbol{J}, \tag{20}$$

where (assuming that $\boldsymbol{A}$ has p distinct eigenvalues[3])

$$\boldsymbol{J} = \begin{bmatrix} \boldsymbol{J}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{J}_p \end{bmatrix} \tag{21}$$

is a block-diagonal matrix called the Jordan form of $\boldsymbol{A}$ (see Appendix A and Table 1). The matrix

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix} \tag{22}$$

is the matrix that contains the eigenvectors and the generalized eigenvectors of $\boldsymbol{A}$ columnwise.

Since the matrix $\boldsymbol{P}$ is invertible (eigenvectors and generalized eigenvectors are linearly independent) we have the matrix factorization

$$\boldsymbol{A} = \boldsymbol{P}\boldsymbol{J}\boldsymbol{P}^{-1}, \tag{23}$$

By following exactly the same steps as in (15)-(17) we obtain the following expression for the matrix exponential of $\boldsymbol{A}$ in the case where $\boldsymbol{A}$ is non-diagonalizable

$$e^{t\boldsymbol{A}} = \boldsymbol{P}e^{t\boldsymbol{J}}\boldsymbol{P}^{-1}. \tag{24}$$

[3] The sum of the algebraic multiplicities of the eigenvalues $\{\lambda_1, \ldots, \lambda_p\}$ must be equal to n.

Properties of the eigenvalue	Jordan block	Exponential of Jordan block
λ_i has algebraic multiplicity one	$\boldsymbol{J}_i = \begin{bmatrix}\lambda_i\end{bmatrix}$	$e^{t\boldsymbol{J}_i} = \begin{bmatrix}e^{t\lambda_i}\end{bmatrix}$
λ_i has algebraic multiplicity two and geometric multiplicity two	$\boldsymbol{J}_i = \begin{bmatrix}\lambda_i & 0\\ 0 & \lambda_i\end{bmatrix}$	$e^{t\boldsymbol{J}_i} = \begin{bmatrix}e^{t\lambda_i} & 0\\ 0 & e^{t\lambda_i}\end{bmatrix}$
λ_i has algebraic multiplicity two and geometric multiplicity one	$\boldsymbol{J}_i = \begin{bmatrix}\lambda_i & 1\\ 0 & \lambda_i\end{bmatrix}$	$e^{t\boldsymbol{J}_i} = \begin{bmatrix}e^{t\lambda_i} & te^{t\lambda_i}\\ 0 & e^{t\lambda_i}\end{bmatrix}$
λ_i has algebraic multiplicity three and geometric multiplicity three	$\boldsymbol{J}_i = \begin{bmatrix}\lambda_i & 0 & 0\\ 0 & \lambda_i & 0\\ 0 & 0 & \lambda_i\end{bmatrix}$	$e^{t\boldsymbol{J}_i} = \begin{bmatrix}e^{t\lambda_i} & 0 & 0\\ 0 & e^{t\lambda_i} & 0\\ 0 & 0 & e^{t\lambda_i}\end{bmatrix}$
λ_i has algebraic multiplicity three and geometric multiplicity two	$\boldsymbol{J}_i = \begin{bmatrix}\lambda_i & 0 & 0\\ 0 & \lambda_i & 1\\ 0 & 0 & \lambda_i\end{bmatrix}$	$e^{t\boldsymbol{J}_i} = \begin{bmatrix}e^{t\lambda_i} & 0 & 0\\ 0 & e^{t\lambda_i} & te^{t\lambda_i}\\ 0 & 0 & e^{t\lambda_i}\end{bmatrix}$
λ_i has algebraic multiplicity three and geometric multiplicity one	$\boldsymbol{J}_i = \begin{bmatrix}\lambda_i & 1 & 0\\ 0 & \lambda_i & 1\\ 0 & 0 & \lambda_i\end{bmatrix}$	$e^{t\boldsymbol{J}_i} = \begin{bmatrix}e^{t\lambda_i} & te^{t\lambda_i} & t^2e^{t\lambda_i}/2\\ 0 & e^{t\lambda_i} & te^{t\lambda_i}\\ 0 & 0 & e^{t\lambda_i}\end{bmatrix}$

Table 1: Jordan blocks and matrix exponentials of Jordan blocks (see Appendix A) corresponding to eigenvalues λ_i with different algebraic and geometric multiplicities.

The Jordan canonical form of $\boldsymbol{A}$ is a block-diagonal matrix (see equation (153)), with blocks given in Table 1. The matrix exponential of a block-diagonal matrix is a matrix that has the exponential of each block in the diagonal

$$e^{t\boldsymbol{J}} = \begin{bmatrix} e^{t\boldsymbol{J}_1} & & & \\ & e^{t\boldsymbol{J}_2} & & \\ & & \ddots & \\ & & & e^{t\boldsymbol{J}_p} \end{bmatrix}. \tag{25}$$

In Table 1 we summarize the Jordan blocks corresponding to different types of eigenvalues. The mathematical proof of each Jordan block is given in Appendix A. Hence, steps to compute the analytical solution (1) in the case where $\boldsymbol{A}$ is not diagonalizable are:

1. Compute the eigenvalues, the eigenvectors, and the generalized eigenvectors of $\boldsymbol{A}$;
2. Construct the the matrix $\boldsymbol{J}$ using the Jordan blocks in Table 1;
3. Construct the matrix $\boldsymbol{P}$ in (22) and the matrix exponential (25) by exponentiating each Jordan block as in Table 1;

4. Compute the analytical solution of (1) using matrix-vector products

$$\boxed{\boldsymbol{X}(t,\boldsymbol{x}_0) = \boldsymbol{P}e^{t\boldsymbol{J}}\boldsymbol{P}^{-1}\boldsymbol{x}_0.} \tag{26}$$

Fundamental matrix. In the theory of autonomous linear ODEs the general solution of the system (1) is often expressed in terms of a *fundamental matrix* $\boldsymbol{\Phi}(t)$ as

$$\boldsymbol{x}_g(t) = \boldsymbol{\Phi}(t)\boldsymbol{c}, \tag{27}$$

where $\boldsymbol{c}$ is an arbitrary vector. Enforcing the initial condition $\boldsymbol{x}_g(0) = \boldsymbol{x}_0$ we find that

$$\boldsymbol{c} = \boldsymbol{\Phi}^{-1}(0)\boldsymbol{x}_0. \tag{28}$$

Substituting this expression for $\boldsymbol{c}$ back into (27) gives

$$\boldsymbol{X}(t,\boldsymbol{x}_0) = \boldsymbol{\Phi}(t)\boldsymbol{\Phi}^{-1}(0)\boldsymbol{x}_0. \tag{29}$$

Comparing this expression to (10) suggests that we can equivalently write the matrix exponential of $\boldsymbol{A}$ as

$$e^{t\boldsymbol{A}} = \boldsymbol{\Phi}(t)\boldsymbol{\Phi}^{-1}(0). \tag{30}$$

Regarding the analytical expression of the fundamental matrix $\boldsymbol{\Phi}(t)$, it can be obtained immediately by comparing (30) with (24). This yields

$$\boldsymbol{\Phi}(t) = \boldsymbol{P}e^{t\boldsymbol{J}} \tag{31}$$

where $\boldsymbol{P}$ is the matrix (22) that has the eigenvectors and generalized eigenvectors of $\boldsymbol{A}$ as columns. As before, the exponential of the Jordan canonical form of $\boldsymbol{A}$, i.e., $e^{t\boldsymbol{J}}$, can be computed by using (25) and exponentiating each Jordan block as in Table 1.

Two-dimensional linear dynamical systems. In this section we compute the analytical solution/flow of several prototype two-dimensional dynamical systems using the mathematical techniques we just discussed. Specifically, we study the flow corresponding to the *saddle node*, *spiral*, *center*, and *degenerate node*.

Saddle node. Consider the linear dynamical system

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\end{bmatrix} = \underbrace{\begin{bmatrix}2 & 3\\ 3 & -6\end{bmatrix}}_{\boldsymbol{A}}\begin{bmatrix}x_1\\ x_2\end{bmatrix}. \tag{32}$$

We have seen in Appendix A (Example 4) that the eigenvalues of $\boldsymbol{A}$ are

$$\lambda_1 = 3, \qquad \lambda_2 = -7. \tag{33}$$

Since the eigenvalues are simple, the matrix $\boldsymbol{A}$ is diagonalizable. A basis for the eigenspace corresponding to each eigenvalue is

$$\boldsymbol{v}_1 = \begin{bmatrix}3\\ 1\end{bmatrix}, \qquad \boldsymbol{v}_2 = \begin{bmatrix}1\\ -3\end{bmatrix}. \tag{34}$$

The matrix of eigenvectors that defines the similarity transformation (11) is

$$\boldsymbol{P} = \begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2\end{bmatrix} = \begin{bmatrix}3 & 1\\ 1 & -3\end{bmatrix}. \tag{35}$$

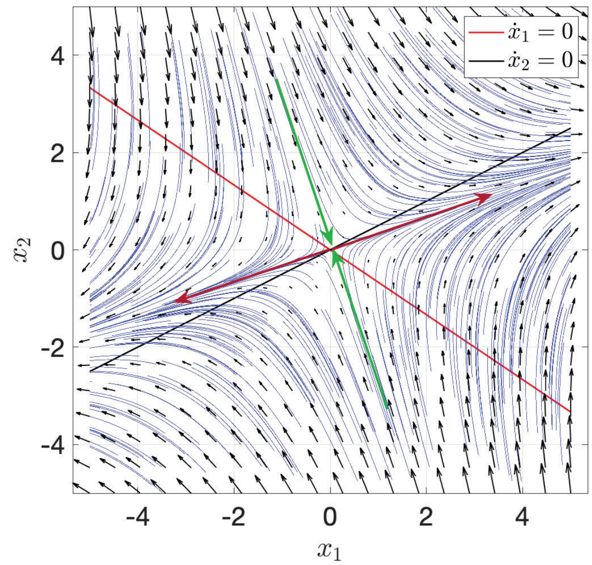


Figure 1: Saddle node. Shown are the nullclines, and the unstable (red arrows)/stable (green arrows) manifolds of the saddle identified by the eigenvectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$, respectively.

The inverse of $\boldsymbol{P}$ is

$$\boldsymbol{P}^{-1} = \frac{1}{10} = \begin{bmatrix} 3 & 1 \\ 1 & -3 \end{bmatrix}. \tag{36}$$

This yields the analytical solution

$$\begin{bmatrix} X_1(t, \boldsymbol{x}_0) \\ X_2(t, \boldsymbol{x}_0) \end{bmatrix} = \underbrace{\begin{bmatrix} 3 & 1 \\ 1 & -3 \end{bmatrix}}_{\boldsymbol{P}} \underbrace{\begin{bmatrix} e^{3t} & 0 \\ 0 & e^{-7t} \end{bmatrix}}_{e^{t\boldsymbol{\Lambda}}} \underbrace{\frac{1}{10}\begin{bmatrix} 3 & 1 \\ 1 & -3 \end{bmatrix}}_{\boldsymbol{P}^{-1}} \underbrace{\begin{bmatrix} x_{01} \\ x_{02} \end{bmatrix}}_{\boldsymbol{x}_0} \tag{37}$$

Developing the matrix products yields the desired flow

$$\begin{cases} X_1(t, \boldsymbol{x}_0) = \dfrac{x_{01}}{10}\left(9e^{3t} + e^{-7t}\right) + \dfrac{x_{02}}{10}\left(e^{3t} + 9e^{-7t}\right) \\ X_2(t, \boldsymbol{x}_0) = \dfrac{x_{01}}{10}\left(3e^{3t} - 3e^{-7t}\right) + \dfrac{x_{02}}{10}\left(3e^{3t} - 3e^{-7t}\right) \end{cases} \tag{38}$$

The phase portrait of this flow is shown in Figure 1.

Stable spiral. Consider the linear dynamical system

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \underbrace{\begin{bmatrix} -1 & -1 \\ 1 & -1 \end{bmatrix}}_{\boldsymbol{A}} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}. \tag{39}$$

The eigenvalues of the matrix $\boldsymbol{A}$ are

$$\lambda_1 = -1 + i, \qquad \lambda_2 = -1 - i. \tag{40}$$

These eigenvalues are complex conjugates and both have algebraic multiplicity one (simple eigenvalues), which implies that they have geometric multiplicity one. Therefore the matrix $\boldsymbol{A}$ is diagonalizable, and there exits a one-dimensional eigenspace (spanned by a complex vector) for each λ_i. To compute such eigenspaces/eigenvectors we proceed as usual

$$(\boldsymbol{A} - \lambda_1 \boldsymbol{I})\boldsymbol{v}_1 = \boldsymbol{0}_{\mathbb{R}^2} \quad \Leftrightarrow \quad \begin{bmatrix} -i & -1 \\ 1 & -i \end{bmatrix} \begin{bmatrix} v_{11} \\ v_{12} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Leftrightarrow \quad \begin{cases} -iv_{11} = v_{12} \\ v_{12} \text{ or } v_{11} \text{ free} \end{cases} \tag{41}$$

$$(\boldsymbol{A} - \lambda_2 \boldsymbol{I})\boldsymbol{v}_2 = \boldsymbol{0}_{\mathbb{R}^2} \quad \Leftrightarrow \quad \begin{bmatrix} i & -1 \\ 1 & i \end{bmatrix} \begin{bmatrix} v_{21} \\ v_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Leftrightarrow \quad \begin{cases} iv_{21} = v_{22} \\ v_{21} \text{ or } v_{22} \text{ free} \end{cases} \tag{42}$$

We choose $v_1 = v_{21} = i$, which yields the following basis for the (complex) eigenspaces corresponding to λ_1 and λ_2, respectively

$$\boldsymbol{v}_1 = \begin{bmatrix} i \\ 1 \end{bmatrix}, \qquad \boldsymbol{v}_2 = \begin{bmatrix} i \\ -1 \end{bmatrix}. \tag{43}$$

The similarity matrix $\boldsymbol{P}$ and its inverse are

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 \end{bmatrix} = \begin{bmatrix} i & i \\ 1 & -1 \end{bmatrix}, \qquad \boldsymbol{P}^{-1} = \frac{1}{2} \begin{bmatrix} -i & 1 \\ -i & 1 \end{bmatrix}. \tag{44}$$

The matrix exponential (17) is easily obtained as

$$\begin{aligned} e^{t\boldsymbol{A}} &= \underbrace{\begin{bmatrix} i & i \\ 1 & -1 \end{bmatrix}}_{\boldsymbol{P}} \underbrace{\begin{bmatrix} e^{t(-1+i)} & 0 \\ 0 & e^{t(-1-i)} \end{bmatrix}}_{e^{t\boldsymbol{\Lambda}}} \underbrace{\frac{1}{2} \begin{bmatrix} -i & 1 \\ -i & 1 \end{bmatrix}}_{\boldsymbol{P}^{-1}} \\ &= \frac{e^{-t}}{2} \begin{bmatrix} i & i \\ 1 & -1 \end{bmatrix} \begin{bmatrix} -ie^{it} & e^{it} \\ -ie^{-it} & -e^{-it} \end{bmatrix} \\ &= \frac{e^{-t}}{2} \begin{bmatrix} e^{it} + e^{-it} & ie^{it} - ie^{-it} \\ -ie^{it} + ie^{-it} & e^{it} + e^{-it} \end{bmatrix}. \end{aligned} \tag{45}$$

At this point we use the Euler formulas

$$\cos(t) = \frac{e^{it} + e^{-it}}{2}, \qquad \sin(t) = \frac{e^{it} - e^{-it}}{2i}, \tag{46}$$

to obtain

$$e^{t\boldsymbol{A}} = e^{-t} \begin{bmatrix} \cos(t) & -\sin(t) \\ \sin(t) & \cos(t) \end{bmatrix}. \tag{47}$$

Applying $e^{t\boldsymbol{A}}$ to the initial condition $\boldsymbol{x}_0$ gives us the analytical solution

$$\begin{cases} X_1(t, \boldsymbol{x}_0) = e^{-t} \left[\cos(t) x_{01} - \sin(t) x_{02}\right] \\ X_2(t, \boldsymbol{x}_0) = e^{-t} \left[\sin(t) x_{01} + \cos(t) x_{02}\right] \end{cases}. \tag{48}$$

The phase portrait of this flow is shown in Figure 2.

- **Center.** Consider the linear dynamical system

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}}_{\boldsymbol{A}} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}. \tag{49}$$

The eigenvalues of $\boldsymbol{A}$ are

$$\lambda_1 = i, \qquad \lambda_2 = -i. \tag{50}$$

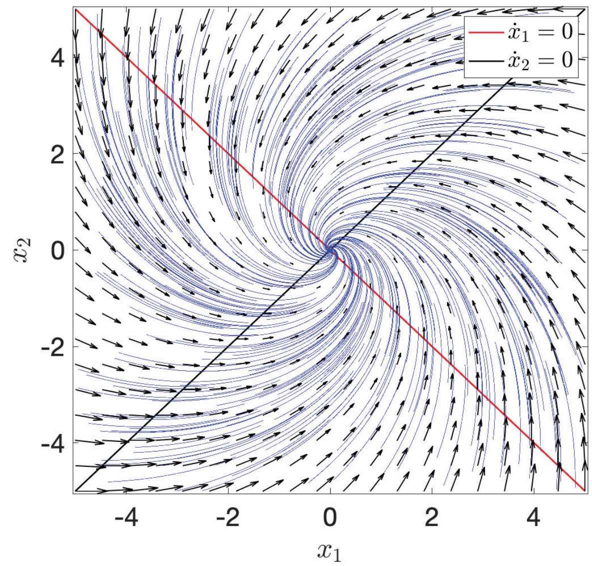


Figure 2: Stable spiral.

The eigenspaces associated with λ_1 and λ_2 are both one-dimensional (both eigenvalues are simple). Let us compute a basis for the eigenspace associated with λ_1

$$(\boldsymbol{A} - \lambda_1 \boldsymbol{I})\boldsymbol{v}_1 = \boldsymbol{0}_{\mathbb{R}^2} \quad \Leftrightarrow \quad \begin{bmatrix} -i & 1 \\ -1 & -i \end{bmatrix} \begin{bmatrix} v_{11} \\ v_{12} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Leftrightarrow \quad \begin{cases} iv_{11} = v_{12} \\ v_{11} \text{ or } v_{12} \text{ free} \end{cases} . \tag{51}$$

We choose $v_{11} = 1$, which yields

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ i \end{bmatrix}. \tag{52}$$

Similarly, for the eigenspace associated with λ_2 we have

$$(\boldsymbol{A} - \lambda_2 \boldsymbol{I})\boldsymbol{v}_2 = \boldsymbol{0}_{\mathbb{R}^2} \quad \Leftrightarrow \quad \begin{bmatrix} i & 1 \\ -1 & i \end{bmatrix} \begin{bmatrix} v_{21} \\ v_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Leftrightarrow \quad \begin{cases} iv_{21} = -v_{22} \\ v_{21} \text{ or } v_{22} \text{ free} \end{cases} . \tag{53}$$

We choose $v_{21} = 1$, which yields

$$\boldsymbol{v}_2 = \begin{bmatrix} 1 \\ -i \end{bmatrix}. \tag{54}$$

The similarity matrix $\boldsymbol{P}$ and its inverse are

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}, \qquad \boldsymbol{P}^{-1} = \frac{1}{2} \begin{bmatrix} 1 & -i \\ 1 & i \end{bmatrix}. \tag{55}$$

The matrix exponential $e^{t\boldsymbol{A}}$ can be computed using equation (17)

$$\begin{aligned} e^{t\boldsymbol{A}} &= \underbrace{\begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}}_{\boldsymbol{P}} \underbrace{\begin{bmatrix} e^{it} & 0 \\ 0 & e^{-it} \end{bmatrix}}_{e^{t\boldsymbol{\Lambda}}} \underbrace{\frac{1}{2} \begin{bmatrix} 1 & -i \\ 1 & i \end{bmatrix}}_{\boldsymbol{P}^{-1}} \\ &= \frac{1}{2} \begin{bmatrix} e^{it} + e^{-it} & -i\left(e^{it} - e^{-it}\right) \\ i\left(e^{it} - e^{-it}\right) & e^{it} + e^{-it} \end{bmatrix} && (56) \\ &= \begin{bmatrix} \cos(t) & \sin(t) \\ -\sin(t) & \cos(t) \end{bmatrix}, && (57) \end{aligned}$$

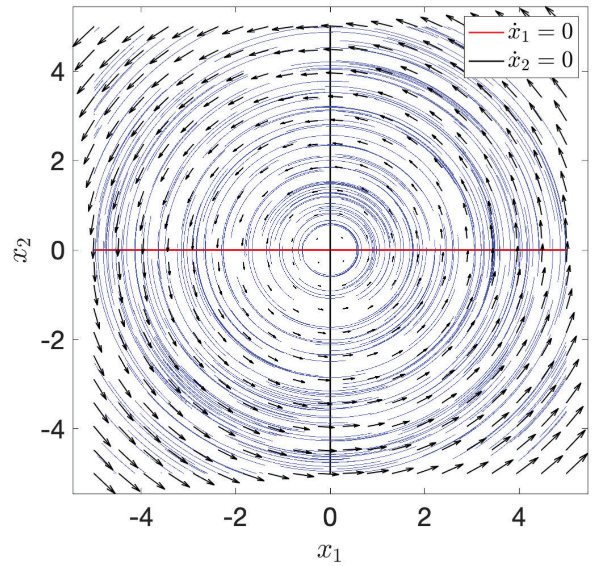


Figure 3: Center.

where we used again the Euler formulas (46). A substitution of the matrix exponential into (10) yields the analytical solution

$$\begin{cases} X_1(t, \boldsymbol{x}_0) = x_{01}\cos(t) + x_{02}\sin(t) \\ X_2(t, \boldsymbol{x}_0) = -x_{01}\sin(t) + x_{02}\cos(t) \end{cases} . \tag{58}$$

The phase portrait of this flow is shown in Figure 3.

- **Degenerate node.** Consider the linear dynamical system

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 1 \\ -1 & 3 \end{bmatrix}}_{\boldsymbol{A}} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}. \tag{59}$$

The matrix $\boldsymbol{A}$ has only one eigenvalue $\lambda = 2$ with algebraic multiplicity 2. The dimension of the corresponding eigenspace, i.e., the dimension of the nullspace of $(\boldsymbol{A} - \lambda \boldsymbol{I})$ (geometric multiplicity of λ), can be calculated using the matrix rank theorem

$$\dim\left(N(\boldsymbol{A} - \lambda\boldsymbol{I})\right) = 2 - \text{rank}(\boldsymbol{A} - \lambda\boldsymbol{I}) = 2 - \underbrace{\text{rank}\left(\begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix}\right)}_{=1} = 1 \tag{60}$$

Hence the dimension of the eigenspace associated with $\lambda = 2$, is equal to one. This implies that the matrix $\boldsymbol{A}$ is *not* diagonalizable. Let us compute a basis for the one-dimensional eigenspace. We have

$$(\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{v}_1 = \boldsymbol{0}_{\mathbb{R}^2} \quad \Leftrightarrow \quad \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} v_{11} \\ v_{12} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Leftrightarrow \quad \begin{cases} v_{11} = v_{12} \\ v_{11} \text{ or } v_{12} \text{ free} \end{cases} \tag{61}$$

We choose $v_{12} = 1$, which yields

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \tag{62}$$

At this point we need to complement $\boldsymbol{v}_1$ to a basis of $\mathbb{R}^2$ by adding one linearly independent vector. To this end, we compute the so-called generalized eigenvector[4] by solving the linear equation

$$(\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{v}_2 = \boldsymbol{v}_1 \tag{64}$$

We obtain

$$\begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} v_{21} \\ v_{22} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \Rightarrow \qquad \begin{cases} -v_{21} + v_{22} = 1 \\ v_{21} \text{ or } v_{22} \text{ free} \end{cases} \tag{65}$$

We choose $v_{22} = 1$ which gives the generalized eigenvector

$$\boldsymbol{v}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{66}$$

The similarity matrix in this case has the eigenvector $\boldsymbol{v}_1$ and the generalized eigenvector $\boldsymbol{v}_2$ as columns

$$\boldsymbol{P} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \qquad \Leftrightarrow \qquad \boldsymbol{P}^{-1} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}. \tag{67}$$

The matrix exponential of the Jordan block that corresponds to the eigenvalue $\lambda = 2$ with algebraic multiplicity two and geometric multiplicity one is (see Table 1)

$$e^{t\boldsymbol{J}} = \begin{bmatrix} e^{2t} & te^{2t} \\ 0 & e^{2t} \end{bmatrix}. \tag{68}$$

The the matrix exponential $e^{t\boldsymbol{A}}$ can now be computed explicitly via the formula (24)

$$\begin{aligned} e^{t\boldsymbol{A}} &= \underbrace{\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}}_{\boldsymbol{P}} \underbrace{\begin{bmatrix} e^{2t} & te^{2t} \\ 0 & e^{2t} \end{bmatrix}}_{e^{t\boldsymbol{J}}} \underbrace{\begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}}_{\boldsymbol{P}^{-1}} \\ &= \begin{bmatrix} e^{2t} - te^{2t} & te^{2t} \\ -te^{2t} & e^{2t} + te^{2t} \end{bmatrix}. \end{aligned} \tag{69}$$

This gives the analytical solution

$$\begin{cases} X_1(t, \boldsymbol{x}_0) = \left(e^{2t} - te^{2t}\right) x_{01} + te^{2t} x_{02} \\ X_2(t, \boldsymbol{x}_0) = -te^{2t} x_{01} + \left(e^{2t} + te^{2t}\right) x_{02} \end{cases}. \tag{70}$$

The phase portrait of this flow is shown in Figure 4.

Classification of two-dimensional flows generated linear dynamical systems. In Figure 5 and Figure 6 we provide a classification of all possible flows generated by two-dimensional dynamical systems in terms of the eigenvalues of the matrix $\boldsymbol{A}$. Of course, changing the sign of the eigenvalues of $\boldsymbol{A}$ is equivalent to transforming the matrix from $\boldsymbol{A}$ to $-\boldsymbol{A}$. This yields an inversion in the orientation of all trajectories, which implies, e.g., that stable nodes become unstable, centers spin the other way around, etc.

Three-dimensional linear dynamical systems. In this section we calculate analytically the flow generated by three dimensional linear systems. Higher-dimensional system can be dealt with using similar

[4]Note that the generalized eigenvector $\boldsymbol{v}_2$ defined in (64) is in the nullspace of the matrix $(\boldsymbol{A} - \lambda\boldsymbol{I})^2$. In fact,

$$(\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{v}_2 = \boldsymbol{v}_1 \quad \Rightarrow (\boldsymbol{A} - \lambda\boldsymbol{I})^2 \boldsymbol{v}_2 = \boldsymbol{0}_{\mathbb{R}^2}. \tag{63}$$

It can be shown that eigenvectors and generalized eigenvectors are linearly independent.

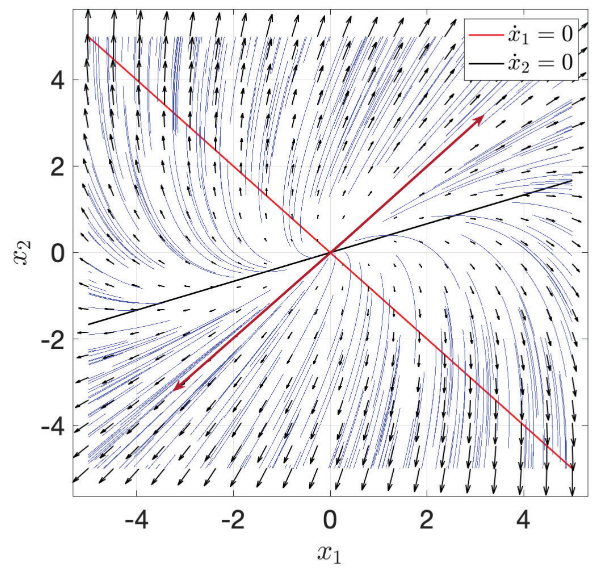


Figure 4: Degenerate node. Shown is the unstable manifold of the node (red arrows), which is defined by the eigendirection $\boldsymbol{v}_1$ corresponding to the eigenvalue $\lambda = 2$.

techniques. The general approach to compute the analytical solution (10) of a general linear system (1) is described beginning at pages 2-5 of this course note.

Example: Consider the three dimensional linear system Consider the linear dynamical system

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & -1 \end{bmatrix}}_{\boldsymbol{A}} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \tag{71}$$

The matrix $\boldsymbol{A}$ has eigenvalues $\lambda_1 = 1$ (with algebraic multiplicity two) and $\lambda_2 = -1$ (with algebraic multiplicity one). The dimension of the eigenspace corresponding to λ_1, i.e., the geometric multiplicity of λ_1 is

$$\dim(N(\boldsymbol{A} - \lambda_1 \boldsymbol{I})) = 3 - \operatorname{rank}(\boldsymbol{A} - \lambda_1 \boldsymbol{I}) = 3 - \operatorname{rank}\left(\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & -2 \end{bmatrix}\right) = 3 - 1 = 2. \tag{72}$$

Therefore the matrix is diagonalizable. The eigenvectors corresponding to λ_1 are solution to the linear system $N(\boldsymbol{A} - \lambda_1 \boldsymbol{I})\boldsymbol{v} = \boldsymbol{0}_{\mathbb{R}^3}$, i.e.,

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & -2 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \qquad \Rightarrow \qquad \begin{cases} v_1 + v_2 - 2v_3 = 0 \\ (v_1, v_2) \text{ or } (v_1, v_3) \text{ or } (v_2, v_3) \text{ are arbitrary} \end{cases} \tag{73}$$

We pick $(v_2, v_3) = (1, 1)$ and $(v_2, v_3) = (2, 1)$ which yields the following eigenvectors

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \qquad \boldsymbol{v}_2 = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}. \tag{74}$$

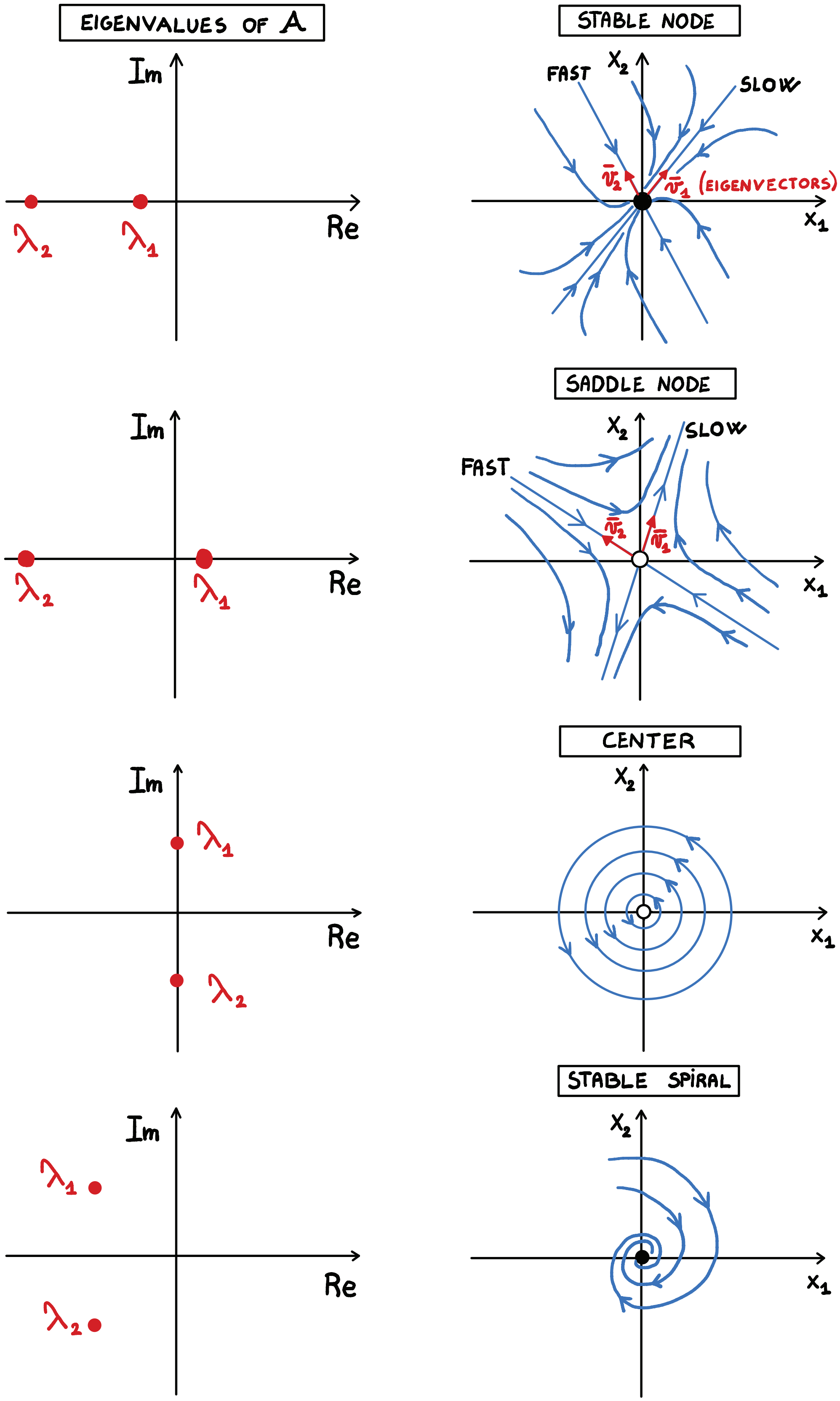


Figure 5: Classification of flows generated by two-dimensional dynamical systems in terms of the eigenvalues of the matrix $\boldsymbol{A}$.

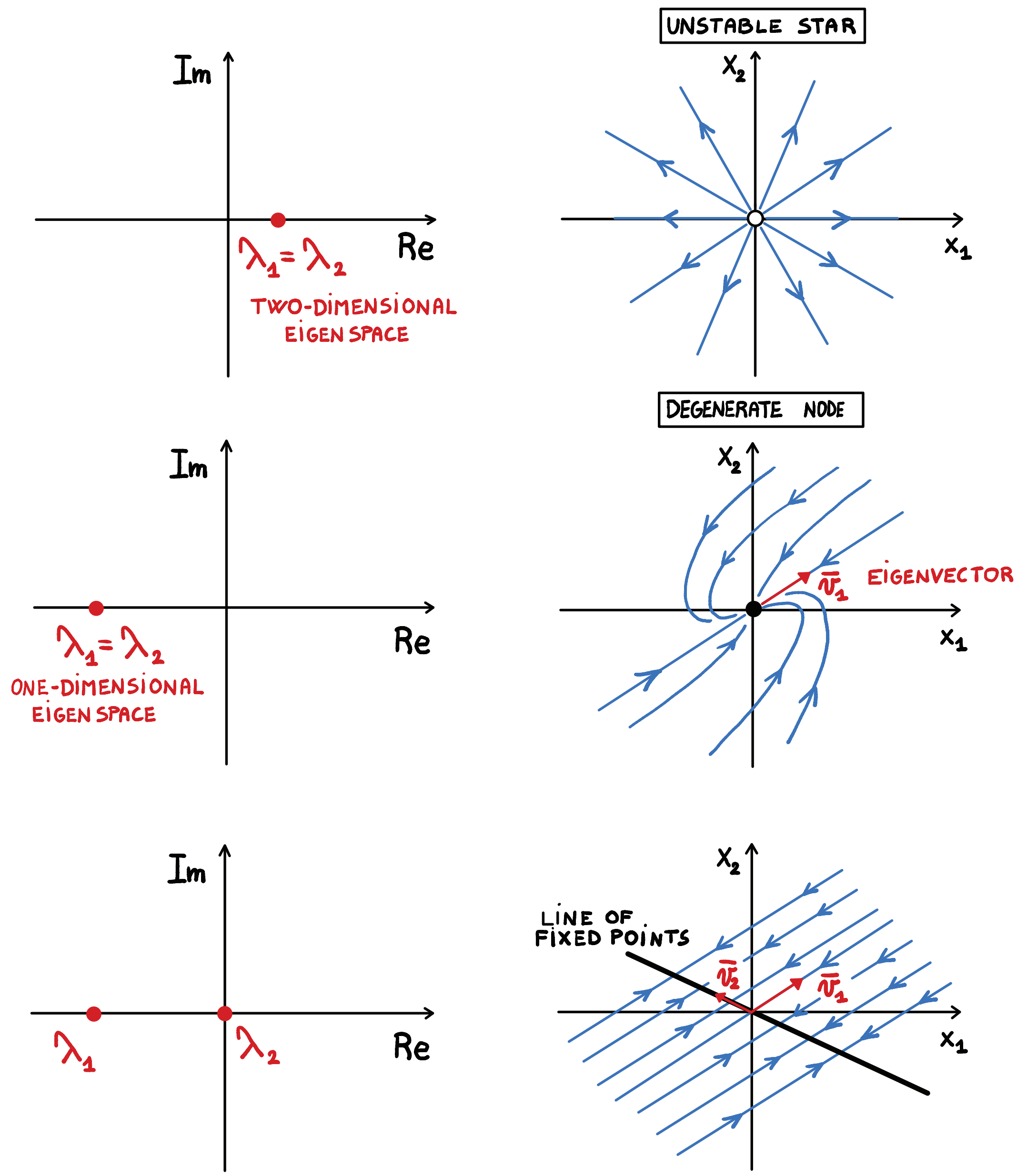


Figure 6: Classification of flows generated by two-dimensional dynamical systems in terms of the eigenvalues of the matrix $\boldsymbol{A}$.

Any linear combination of $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ is still an eigenvector. The eigenvectors corresponding to $\lambda_2 = -1$ are solutions to the linear system $N(\boldsymbol{A} - \lambda_2 \boldsymbol{I})\boldsymbol{v} = \boldsymbol{0}_{\mathbb{R}^3}$, i.e.,

$$\begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \qquad \Rightarrow \qquad \begin{cases} v_1 = 0 \\ v_2 = 0 \\ v_3 \text{ is arbitrary} \end{cases} \tag{75}$$

We choose

$$\boldsymbol{v}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{76}$$

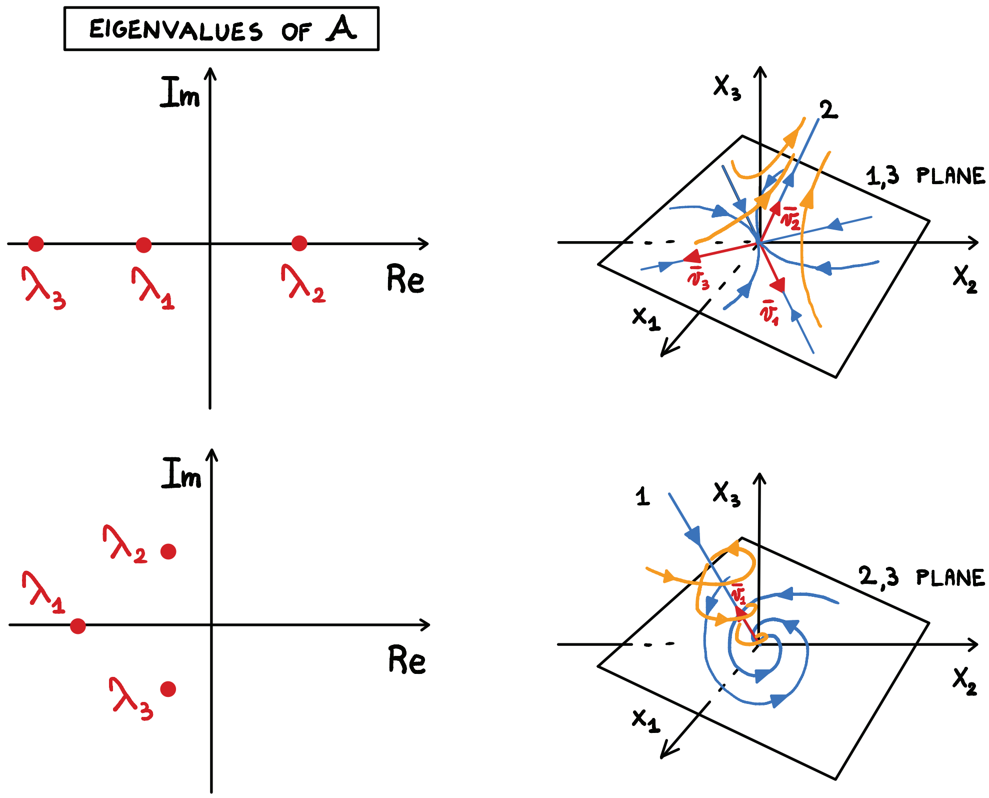


Figure 7: Examples of flows generated by three-dimensional dynamical systems in terms of the eigenvalues of the matrix $\boldsymbol{A}$.

The similarity matrix and its inverse are

$$\boldsymbol{P} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ 1 & 1 & 1 \end{bmatrix}, \qquad \boldsymbol{P}^{-1} = \frac{1}{2}\begin{bmatrix} 2 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & -1 & 2 \end{bmatrix} \tag{77}$$

Therefore, the analytical solution of the 3D linear system (71) is

$$\begin{bmatrix} X_1(t,\boldsymbol{x}_0) \\ X_2(t,\boldsymbol{x}_0) \\ X_3(t,\boldsymbol{x}_0) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} e^t & 0 & 0 \\ 0 & e^t & 0 \\ 0 & 0 & e^{-t} \end{bmatrix} \frac{1}{2} \begin{bmatrix} 2 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & -1 & 2 \end{bmatrix} \begin{bmatrix} x_{01} \\ x_{01} \\ x_{03} \end{bmatrix} \tag{78}$$

i,e,

$$\begin{cases} X_1(t,\boldsymbol{x}_0) = e^t x_{01} \\ X_2(t,\boldsymbol{x}_0) = e^t x_{02} \\ X_3(t,\boldsymbol{x}_0) = \dfrac{(e^t - e^{-t})}{2}(x_{01} + x_{02}) + e^{-t} x_{03} \end{cases} \tag{79}$$

Classification of three-dimensional flows generated linear dynamical systems In Figure 7 we provide a few sketches of three-dimensional flows corresponding to matrices $\boldsymbol{A}$ with various eigenvalues. As easily seen, the classification of these flows is not as straightforward as in the 2D case. In fact, we can have spiraling directions, saddle node planes, etc.

Appendix A: The matrix eigenvalue problem

In this Appendix we briefly review the eigenvalue problem for a $n \times n$ matrix $\boldsymbol{A}$ with real coefficients. The eigenvalue problem is essentially the problem of finding all real (or complex) numbers λ (eigenvalues) and all nonzero real (or complex) vectors $\boldsymbol{v}$ (eigenvectors) satisfying the equation

$$\boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{v}. \tag{80}$$

Computation of eigenvalues. From equation (80) it follows that

$$(\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{v} = \boldsymbol{0}_{\mathbb{R}^n}, \tag{81}$$

Hence, the eigenvector v (which is non-zero by definition) is in the nullspace of the matrix $(\boldsymbol{A} - \lambda\boldsymbol{I})$. This implies that the matrix $(\boldsymbol{A} - \lambda\boldsymbol{I})$ is not invertible[5]. A necessary and sufficient condition for $(\boldsymbol{A} - \lambda\boldsymbol{I})$ to be not invertible is

$$\det(\boldsymbol{A} - \lambda\boldsymbol{I}) = 0 \qquad \text{(characterististic equation)}. \tag{82}$$

The polynomial

$$p(\lambda) = \det(\boldsymbol{A} - \lambda\boldsymbol{I}) \tag{83}$$

is known as *characteristic polynomial* associated with the matrix $\boldsymbol{A}$. The characteristic equation (82) implies that the eigenvalues of the matrix $\boldsymbol{A}$ are roots of the characteristic polynomial $p(\lambda)$.

How many eigenvalues do we have for a given $n \times n$ matrix $\boldsymbol{A}$? The characteristic polynomial $p(\lambda)$ associated with the matrix $\boldsymbol{A}$ is a polynomial of degree n with real coefficients. Hence, by using the fundamental theorem of algebra we conclude $p(\lambda)$ has exactly n roots which may be real or complex conjugates. In other words, every $n \times n$ matrix has exactly n eigenvalues. Such eigenvalues may be repeated, in which case we say that they have "algebraic multiplicity" greater than one. In other words, the multiplicity of an eigenvalue as a root of the characteristic polynomial is called *algebraic multiplicity* the eigenvalue.

Example 1: Compute the eigenvalues of the matrix

$$\boldsymbol{A} = \begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix}. \tag{84}$$

The characteristic polynomial is

$$p(\lambda) = \det(\boldsymbol{A} - \lambda\boldsymbol{I}) = \det \begin{bmatrix} 2-\lambda & 3 \\ 3 & -6-\lambda \end{bmatrix} = -(2-\lambda)(6+\lambda) - 9, \tag{85}$$

i.e.,

$$p(\lambda) = \lambda^2 + 4\lambda - 21. \tag{86}$$

The eigenvalues of $\boldsymbol{A}$ are roots of $p(\lambda)$. Setting $p(\lambda) = 0$ yields

$$\lambda_{1,2} = -2 \pm \sqrt{4+21} = -2 \pm 5 \qquad \Rightarrow \qquad \lambda_1 = 3, \quad \lambda_2 = -7. \tag{87}$$

In this case, both eigenvalues have algebraic multiplicity one, i.e., they are simple roots of $p(\lambda)$. The characteristic polynomial can be factored as

$$p(\lambda) = (\lambda - 3)(\lambda + 7), \tag{88}$$

[5] The matrix $(\boldsymbol{A} - \lambda\boldsymbol{I})$ in (81) maps a non-zero vector $\boldsymbol{v}$ into $\boldsymbol{0}_{\mathbb{R}^n}$. Hence the the nullspace of $(\boldsymbol{A} - \lambda\boldsymbol{I})$ has a nonzero vector in it, which implies that the matrix $(\boldsymbol{A} - \lambda\boldsymbol{I})$ is not invertible.

suggesting once again that $\lambda = 3$ and $\lambda = -7$ are simple roots.

Example 2: Compute the eigenvalues of the matrix

$$\boldsymbol{A} = \begin{bmatrix} 2 & 5 & 1 & -5 \\ 0 & 4 & 3 & 0 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{89}$$

In this case we have

$$\boldsymbol{A} - \lambda \boldsymbol{I} = \begin{bmatrix} 2-\lambda & 5 & 1 & -5 \\ 0 & 4-\lambda & 3 & 0 \\ 0 & 0 & 2-\lambda & 4 \\ 0 & 0 & 0 & 1-\lambda \end{bmatrix} \tag{90}$$

and

$$p(\lambda) = \det(\boldsymbol{A} - \lambda \boldsymbol{I}) = (2-\lambda)^2(4-\lambda)(1-\lambda). \tag{91}$$

Hence, the matrix $\boldsymbol{A}$ has three eigenvalues:

$$\begin{aligned} \lambda_1 &= 2 \qquad \text{with algebraic multiplicity 2,} \\ \lambda_2 &= 4 \qquad \text{with algebraic multiplicity 1,} \\ \lambda_3 &= 1 \qquad \text{with algebraic multiplicity 1.} \end{aligned}$$

Note that the eigenvalues coincides with the diagonal entries of the matrix $\boldsymbol{A}$. This is a general fact about upper or or lower triangular matrices, i.e., the eigenvalues of such matrices coincides with the diagonal entries of the matrix. For example, the following matrix

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{92}$$

has two eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 0$, both with algebraic multiplicity 2.

Example 3: Compute the eigenvalues of the following matrix

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix}. \tag{93}$$

The characteristic polynomial is

$$p(\lambda) = \det(\boldsymbol{A} - \lambda \boldsymbol{I}) = \det \begin{bmatrix} 1-\lambda & 2 \\ -1 & 1-\lambda \end{bmatrix} = -(1-\lambda)^2 + 2, \tag{94}$$

i.e.,

$$p(\lambda) = \lambda^2 - 2\lambda + 3. \tag{95}$$

Hence, the eigenvalues are

$$\lambda_1 = 1 + i\sqrt{2} \qquad \lambda_2 = 1 - i\sqrt{2} \tag{96}$$

Note that λ_1 and λ_2 are complex conjugates eigenvalues. Clearly, for 2×2 matrices with real entries the fundamental theorem of algebra tells us that the eigenvalues are either both real or complex conjugates.

Eigenvectors and eigenspaces. By definition, an eigenvector of a $n \times n$ matrix $\boldsymbol{A}$ is a nonzero vector $\boldsymbol{v} \in \mathbb{R}^n$ such that

$$\boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{v}. \tag{97}$$

This means that $\boldsymbol{v}$ is a nonzero vector in the nullspace of the matrix $(\boldsymbol{A} - \lambda \boldsymbol{I})$. In fact, $\boldsymbol{v}$ is mapped onto the zero of $\mathbb{R}^n$ by $(\boldsymbol{A} - \lambda \boldsymbol{I})$ (see equation (81)). We know that the nullspace of a $n \times n$ matrix is a vector subspace of $\mathbb{R}^n$.

We denote by $N(\boldsymbol{A} - \lambda \boldsymbol{I})$ the nullspace of the matrix $(\boldsymbol{A} - \lambda \boldsymbol{I})$, and call $N(\boldsymbol{A} - \lambda \boldsymbol{I})$ the *eigenspace* of $\boldsymbol{A}$ corresponding to the eigenvalue λ. The dimension of the eigenspace $N(\boldsymbol{A} - \lambda \boldsymbol{I})$ is called *geometric multiplicity* of the eigenvalue λ. By definition, an eigenvector cannot be zero and therefore the eigenspace corresponding to each eigenvalue has dimension at least equal to one. The dimension of the eigenspace corresponding to some eigenvalue can be computed by using the matrix rank theorem.

Example 4: Compute the eigenspaces of the matrix

$$\boldsymbol{A} = \begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix} \tag{98}$$

We have seen in Example 1 that the eigenvalues of $\boldsymbol{A}$ are $\lambda_1 = 3$ and $\lambda_2 = -7$. Let us compute the eigenspace corresponding to λ_1. To this end, we first compute the dimension of such eigenspace by using the matrix rank theorem

$$\dim(N(\boldsymbol{A} - \lambda_1 \boldsymbol{I}) = 2 - \operatorname{rank}(\boldsymbol{A} - \lambda_1 \boldsymbol{I}) = 2 - \operatorname{rank}\left(\begin{bmatrix} -1 & 3 \\ 3 & -9 \end{bmatrix}\right) = 2 - 1 = 1 \tag{99}$$

Hence, the eigenspace corresponding to λ_1 has dimension one. Any vector of such an eigenspace is an eigenvector of A corresponding to λ_1. To compute a basis for the eigenspace $N(\boldsymbol{A} - \lambda_1 \boldsymbol{I})$ consider

$$(\boldsymbol{A} - \lambda_1 \boldsymbol{I})\boldsymbol{v} = \boldsymbol{0}_{\mathbb{R}^2} \quad \Leftrightarrow \quad \begin{bmatrix} -1 & 3 \\ 3 & -9 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Leftrightarrow \quad -v_1 + 3v_2 = 0 \tag{100}$$

Hence,

$$\boldsymbol{v} = \begin{bmatrix} 3 \\ 1 \end{bmatrix} \tag{101}$$

is a basis for $N(\boldsymbol{A} - \lambda_1 \boldsymbol{I})$, and an eigenvector of $\boldsymbol{A}$ corresponding to λ_1. All eigenvectors of $\boldsymbol{A}$ corresponding to λ_1 are in the form

$$c \begin{bmatrix} 3 \\ 1 \end{bmatrix} \quad \text{with } c \neq 0. \tag{102}$$

Similarly, the eigenspace corresponding to λ_2 has dimension 1 and can be determined by solving the linear system

$$(\boldsymbol{A} - \lambda_2 \boldsymbol{I})\boldsymbol{v} = \boldsymbol{0}_{\mathbb{R}^2} \quad \Leftrightarrow \quad \begin{bmatrix} 9 & 3 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Leftrightarrow \quad 3v_1 + v_2 = 0. \tag{103}$$

Hence,

$$\boldsymbol{v} = \begin{bmatrix} 1 \\ -3 \end{bmatrix} \tag{104}$$

is a basis for $N(\boldsymbol{A} - \lambda_2 \boldsymbol{I})$ and an eigenvector of $\boldsymbol{A}$ corresponding to λ_2. In summary, λ_1 and λ_2 are eigenvalues with algebraic multiplicity one and geometric multiplicity one. Geometric multiplicity one means that the eigenspaces $N(\boldsymbol{A} - \lambda_1 \boldsymbol{I})$ and $N(\boldsymbol{A} - \lambda_2 \boldsymbol{I})$ are both one-dimensional. A basis for $N(\boldsymbol{A} - \lambda_1 \boldsymbol{I})$ and $N(\boldsymbol{A} - \lambda_2 \boldsymbol{I})$ is given by (101) and (104), respectively.

The following theorem establishes a relationship between the algebraic multiplicity and the geometric multiplicity of an eigenvalue λ.

Theorem 1. Let λ be an eigenvalue of a $n \times n$ matrix $\boldsymbol{A}$. Denote by s the algebraic multiplicity of λ. Then

$$\dim(N(\boldsymbol{A} - \lambda \boldsymbol{I})) \leq s. \tag{105}$$

In other words the geometric multiplicity of the eigenvalue λ (i.e., the dimension of the associated eigenspace) is always smaller or equal than the algebraic multiplicity).

Of course, if λ is a simple eigenvalue ($s = 1$) then $\dim(N(\boldsymbol{A} - \lambda \boldsymbol{I})) = 1$, i.e., the eigenspace corresponding to simple eigenvalues is always one-dimensional. If λ has algebraic multiplicity 2, i.e., it is a repeated eigenvalue, then it is possible to have geometric multiplicity equal to one or equal to two. In the latter case the eigenspace is two-dimensional and any vector in such eigenspace (including linear combinations of multiple eigenvectors) is an eigenvector. Let us provide a simple example of a 2×2 matrix with one eigenvalue of algebraic multiplicity two and geometric multiplicity one.

Example 5: Consider the following matrix

$$\boldsymbol{A} = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}. \tag{106}$$

We know that $\lambda = 2$ is the only eigenvalue and it has algebraic multiplicity two. In fact, the characteristic polynomial is $p(\lambda) = (2 - \lambda)^2$. The geometric multiplicity of $\lambda = 2$ can be calculated by using the matrix rank theorem

$$\dim(N(\boldsymbol{A} - \lambda \boldsymbol{I})) = 2 - \operatorname{rank}(\boldsymbol{A} - \lambda \boldsymbol{I}) = 2 - \underbrace{\operatorname{rank}\left(\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\right)}_{=1} = 2 - 1 = 1. \tag{107}$$

Hence, the eigenspace associated with $\lambda = 2$ is one-dimensional. A basis for such an eigenspace is obtained as follows:

$$(\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{v} = \boldsymbol{0}_{\mathbb{R}^2} \quad \Leftrightarrow \quad \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Leftrightarrow \quad v_2 = 0. \tag{108}$$

We choose

$$\boldsymbol{v} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \tag{109}$$

Example 6: Compute the eigenvalues and the eigenvectors of the following matrix

$$\boldsymbol{A} = \begin{bmatrix} 2 & 1 & 3 \\ 0 & 1 & 5 \\ 0 & 0 & 2 \end{bmatrix}. \tag{110}$$

This is an upper triangular matrix and therefore the eigenvalues coincide with the diagonal entries. Hence we have $\lambda_1 = 2$ with algebraic multiplicity two and $\lambda_2 = 1$ with algebraic multiplicity one.

$$\boldsymbol{A} - \lambda_1 \boldsymbol{I} = \begin{bmatrix} 0 & 1 & 3 \\ 0 & -1 & 5 \\ 0 & 0 & 0 \end{bmatrix} \quad \Leftrightarrow \quad \dim(N(\boldsymbol{A} - \lambda_1 \boldsymbol{I})) = 3 - \underbrace{\operatorname{rank}(\boldsymbol{A} - \lambda_1 \boldsymbol{I})}_{=2} = 1, \tag{111}$$

$$\boldsymbol{A} - \lambda_2 \boldsymbol{I} = \begin{bmatrix} 1 & 1 & 3 \\ 0 & 0 & 5 \\ 0 & 0 & 1 \end{bmatrix} \qquad \Leftrightarrow \qquad \dim(N(\boldsymbol{A} - \lambda_2 \boldsymbol{I})) = 3 - \underbrace{\text{rank}(\boldsymbol{A} - \lambda_2 \boldsymbol{I})}_{=2} = 1. \tag{112}$$

Therefore, the dimension of the eigenspaces associated with λ_1 and λ_2 is one. Let us find a basis for such eigenspaces.

$$(\boldsymbol{A} - \lambda_1 \boldsymbol{I})\boldsymbol{v} = \boldsymbol{0}_{\mathbb{R}^3} \quad \Rightarrow \quad \begin{bmatrix} 0 & 1 & 3 \\ 0 & -1 & 5 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \qquad \Leftrightarrow \qquad \begin{cases} v_1 \text{ arbitrary} \\ v_2 + 3v_3 = 0 \\ -v_2 + 5v_3 = 0 \end{cases} \tag{113}$$

Hence, an eigenvector that spans $N(\boldsymbol{A} - \lambda_1 \boldsymbol{I})$ is

$$\boldsymbol{v} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}. \tag{114}$$

Similarly,

$$(\boldsymbol{A} - \lambda_2 \boldsymbol{I})\boldsymbol{v} = 0_{\mathbb{R}^3} \quad \Rightarrow \quad \begin{bmatrix} 1 & 1 & 3 \\ 0 & 0 & 5 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \qquad \Leftrightarrow \qquad \begin{cases} v_1 + v_2 + 3v_3 = 0 \\ v_3 = 0 \end{cases} \tag{115}$$

Hence, an eigenvector that spans $N(\boldsymbol{A} - \lambda_2 \boldsymbol{I})$ is

$$\boldsymbol{v} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}. \tag{116}$$

Hereafter, we recall an important theorem on eigenvectors corresponding to different eigenvalues.

Theorem 2. Eigenvectors corresponding to different eigenvalues are linearly independent.

Of course if an eigenvalue λ has geometric multiplicity larger than one, then we can construct a basis for $N(\boldsymbol{A} - \lambda \boldsymbol{I})$. In any case, such basis will be linearly independent on any other eigenvector corresponding to a different eigenvalue.

Similarity transformations. Let $\boldsymbol{A}, \boldsymbol{B} \in M_{n\times n}(\mathbb{R}^n)$. We say that $\boldsymbol{A}$ is *similar* to $\boldsymbol{B}$ is there exists an invertible matrix $\boldsymbol{P} \in M_{n\times n}(\mathbb{R}^n)$ such that

$$\boldsymbol{AP} = \boldsymbol{PB} \qquad \Leftrightarrow \qquad \boldsymbol{A} = \boldsymbol{PBP}^{-1} \tag{117}$$

The transformation $\boldsymbol{B} \to \boldsymbol{PBP}^{-1}$ is called *similarity transformation*. An example of similarity transformation is the change of basis transformation.

Theorem 3. Similar matrices have the same eigenvalues.

Proof. Let $A, B \in M_{n\times n}$ be two similar matrices, i.e., $P \in M_{n\times n}$ such that

$$A = PBP^{-1}. \tag{118}$$

Then

$$\det(A - \lambda I) = \det(PBP^{-1} - \lambda PP^{-1}) = \det(P)\det(B - \lambda I)\det(P^{-1}) = \det(B - \lambda I) \tag{119}$$

□

Diagonalization. Consider a $n\times n$ matrix $\boldsymbol{A}$. We have seen in Theorem 2 that eigenvectors corresponding to different eigenvalues are linearly independent. Hence, if the algebraic multiplicity of each eigenvalue is equal to the geometric multiplicity then it is possible to construct a basis for $\mathbb{R}^n$ made of eigenvectors of A. Let us organize such n eigenvectors as columns of a matrix P

$$\boldsymbol{P} = \begin{bmatrix}\boldsymbol{v}_1 & \cdots & \boldsymbol{v}_n\end{bmatrix}. \tag{120}$$

Clearly,

$$\boldsymbol{A}\boldsymbol{P} = \begin{bmatrix}\boldsymbol{A}\boldsymbol{v}_1 & \cdots & \boldsymbol{A}\boldsymbol{v}_n\end{bmatrix} = \begin{bmatrix}\boldsymbol{v}_1 & \cdots & \boldsymbol{v}_n\end{bmatrix}\underbrace{\begin{bmatrix}\lambda_1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \lambda_n\end{bmatrix}}_{\Lambda} = \boldsymbol{P}\boldsymbol{\Lambda}, \tag{121}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix with the eigenvalues of A (counted with their multiplicity) sitting along the diagonal. Equation (121) shows that if $\boldsymbol{A}$ has n linearly independent eigenvectors then $\boldsymbol{A}$ is similar to a diagonal matrix[6] $\boldsymbol{\Lambda}$. The similarity transformation is defined by the matrix $\boldsymbol{P}$ in (120), i.e., the matrix that has the eigenvectors of $\boldsymbol{A}$ as columns.

A simple corollary of this statement is that matrices with simple eigenvalues are always diagonalizable, since they have n linearly independent eigenvectors.

Theorem 4. Let $\boldsymbol{A}$ be a $n\times n$ matrix with eigenvalues $\{\lambda_1,\dots,\lambda_p\}$ with algebraic multiplicities $\{s_1,\dots,s_p\}$, respectively. Then $\boldsymbol{A}$ is diagonalizable if and only if

$$\dim(N(\boldsymbol{A}-\lambda_i\boldsymbol{I})) = s_i \quad \text{for all} \quad i=1,\dots,p. \tag{122}$$

This theorem is saying that if each eigenvalue of a matrix $\boldsymbol{A}$ has algebraic multiplicity equal to its geometric multiplicity then the matrix $\boldsymbol{A}$ is similar to a diagonal matrix. Conversely, if a matrix $\boldsymbol{A}$ is similar to a diagonal matrix then each eigenvalue of $\boldsymbol{A}$ has algebraic multiplicity equal to its geometric multiplicity.

Example 7: The matrix

$$\boldsymbol{A} = \begin{bmatrix}2 & 3\\ 3 & -6\end{bmatrix} \tag{123}$$

is diagonalizable. In fact, we have seen that the eigenvalues are $\lambda_1 = 3$ and $\lambda_2 = -7$ (simple eigenvalues). This implies that the dimension of the associated eigenspace is one for both eigenvalues. The eigenvectors of $\boldsymbol{A}$ are

$$\boldsymbol{v}_1 = \begin{bmatrix}3\\ 1\end{bmatrix} \qquad \text{and} \qquad \boldsymbol{v}_2 = \begin{bmatrix}1\\ -3\end{bmatrix}. \tag{124}$$

Define

$$\boldsymbol{P} = \begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2\end{bmatrix} = \begin{bmatrix}3 & 1\\ 1 & -3\end{bmatrix}, \qquad \boldsymbol{\Lambda} = \begin{bmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{bmatrix} = \begin{bmatrix}3 & 0\\ 0 & -7\end{bmatrix}. \tag{125}$$

It is straightforward to verify that

$$\boldsymbol{P}^{-1} = \frac{1}{10}\begin{bmatrix}3 & 1\\ 1 & -3\end{bmatrix} \tag{126}$$

and

$$\boldsymbol{A} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{-1} \qquad \text{or} \qquad \boldsymbol{\Lambda} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}. \tag{127}$$

[6]In general, we say that a matrix $\boldsymbol{A}$ is *diagonalizable* if there exists an invertible matrix $\boldsymbol{P}$ such that $\boldsymbol{A}$ is similar to a diagonal matrix.

Example 8: The matrix

$$\boldsymbol{A} = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix} \tag{128}$$

is *not* diagonalizable. In fact the algebraic multiplicity of the eigenvalue $\lambda = 2$ is two, while its geometric multiplicity is one. We will see hereafter that it is possible to complement the eigenvector that spans the eigenspace with another linearly independent vector called "generalized eigenvector" to form a basis of $\mathbb{R}^2$. Such generalized eigenvector of $\boldsymbol{A}$, makes $\boldsymbol{A}$ similar to a matrix $\boldsymbol{J}$ called *Jordan form* of $\boldsymbol{A}$. In this particular example, the Jordan form of $\boldsymbol{A}$ coincides with $\boldsymbol{A}$, i.e., $\boldsymbol{A}$ is already in a Jordan form.

Example 9: Verify that the matrix

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 2 \end{bmatrix} \tag{129}$$

is diagonalizable. The matrix is lower-triangular with eigenvalues $\lambda_1 = 1$ (algebraic multiplicity two) and $\lambda_2 = 2$ (algebraic multiplicity one). To verify that A is diagonalizable we just need to check that the geometric multiplicity of $\lambda_1 = 1$ is equal to two. To this end, we use the matrix rank theorem:

$$\dim(N(\boldsymbol{A} - \lambda_1 \boldsymbol{I})) = 3 - \mathrm{rank}(\boldsymbol{A} - \lambda_1 \boldsymbol{I}) = 3 - \mathrm{rank}\left(\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \right) = 3 - 1 = 2 \tag{130}$$

This shows that the dimension of the nullspace of $N(\boldsymbol{A} - \lambda_1 \boldsymbol{I})$, i.e., the dimension of the eigenspace associated with $\lambda_1 = 1$ is two. Let us compute a basis for such an eigenspace. To this end,

$$(\boldsymbol{A} - \lambda_1 \boldsymbol{I})\boldsymbol{v} = \boldsymbol{0}_{\mathbb{R}^3} \quad \Rightarrow \quad \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad \Leftrightarrow \quad \begin{cases} v_1 \text{ arbitrary} \\ v_2 \text{ arbitrary} \\ v_3 = -v_2 \end{cases} \tag{131}$$

Hence, a basis for the eigenspace corresponding to λ_1 is

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} \right\}. \tag{132}$$

On the other hand, the eigenspace $N(\boldsymbol{A} - \lambda_2 \boldsymbol{I})$ is spanned by a vector that can be computed as

$$(\boldsymbol{A} - \lambda_2 \boldsymbol{I})\boldsymbol{v} = \boldsymbol{0}_{\mathbb{R}^3} \quad \Rightarrow \quad \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad \Leftrightarrow \quad \begin{cases} v_1 = 0 \\ v_2 = 0 \\ v_3 \text{ arbitrary} \end{cases} \tag{133}$$

Therefore a matrix $\boldsymbol{P}$ that diagonalizes $\boldsymbol{A}$ is

$$\boldsymbol{P} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}. \tag{134}$$

Indeed, it can be verified by a direct calculation that

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}}_{\boldsymbol{\Lambda}} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}}_{\boldsymbol{P}^{-1}} \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 2 \end{bmatrix}}_{\boldsymbol{A}} \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}}_{\boldsymbol{P}}. \tag{135}$$

Generalized eigenvectors and Jordan canonical form. The set of eigenvectors of any $n \times n$ matrix $\boldsymbol{A}$ can be complemented to a basis of $\mathbb{R}^n$. To this end, we can add a certain number of so-called *generalized eigenvectors*, to each "defective" eigenspace of $\boldsymbol{A}$. A defective eigenspace of $\boldsymbol{A}$ is an eigenspace with dimension $\dim(N(\boldsymbol{A} - \lambda_i \boldsymbol{I}))$ smaller than the algebraic multiplicity s_i of the associated eigenvalue λ_i (see Theorem 1). For such defective eigenspaces we compute

$$s_i - \dim(N(\boldsymbol{A} - \lambda_i \boldsymbol{I})) \tag{136}$$

additional generalized eigenvectors. This yields a basis of $\mathbb{R}^n$ made of eigenvectors and generalized eigenvectors of $\boldsymbol{A}$. Such basis, also induces a similarity transformation between $\boldsymbol{A}$ and a matrix called *Jordan canonical form* of $\boldsymbol{A}$. Let us describe the procedure to compute the Jordan form of a matrix $\boldsymbol{A}$. To this end, let us first consider the simple 2 matrix

$$\boldsymbol{A} = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}. \tag{137}$$

We know that the eigenspace corresponding to the eigenvalue λ is one-dimensional with basis

$$\boldsymbol{v} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \tag{138}$$

To complement $\boldsymbol{v}$ with another vector and form a basis of $\mathbb{R}^2$ we choose $\boldsymbol{w}$ as follows

$$(\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{w} = \boldsymbol{v}. \tag{139}$$

Clearly, $\boldsymbol{w}$ is in the nullspace of the matrix $(\boldsymbol{A} - \lambda \boldsymbol{I})^2$. In fact, by applying $(\boldsymbol{A} - \lambda \boldsymbol{I})$ to both sides of (139) we obtain

$$(\boldsymbol{A} - \lambda \boldsymbol{I})^2 \boldsymbol{w} = (\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{v} = \boldsymbol{0}_{\mathbb{R}^2}. \tag{140}$$

It can be shown that $\boldsymbol{w}$ and $\boldsymbol{v}$ are linearly independent. To compute the generalized eigenvector $\boldsymbol{w}$ we solve the linear system (139)

$$(\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{w} = \boldsymbol{v} \quad \Leftrightarrow \quad \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \Leftrightarrow \quad \begin{cases} w_1 \text{ arbitrary} \\ w_2 = 1 \end{cases}. \tag{141}$$

Hence a generalized eigenvector for the eigenspace $N(\boldsymbol{A} - \lambda \boldsymbol{I})$ is

$$\boldsymbol{w} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{142}$$

At this point we define the similarity transformation

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{v} & \boldsymbol{w} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{143}$$

and apply $\boldsymbol{A}$ to $\boldsymbol{P}$ to obtain

$$\boldsymbol{AP} = \begin{bmatrix} \boldsymbol{Av} & \boldsymbol{Aw} \end{bmatrix} = \begin{bmatrix} \boldsymbol{v} & \boldsymbol{w} \end{bmatrix} \underbrace{\begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}}_{\boldsymbol{J}} = \boldsymbol{PJ}. \tag{144}$$

Hence, $\boldsymbol{A}$ is similar to a matrix $\boldsymbol{J}$ in a particular form (not diagonal but almost diagonal), known as *Jordan canonical form* of $\boldsymbol{A}$. In this particular example, $\boldsymbol{A}$ is already in a Jordan form so the similarity transformation defined by $\boldsymbol{P}$ turns out to be the identity transformation.

Next, let us consider a 3×3 matrix $\boldsymbol{A}$ with only one eigenvalue λ of algebraic multiplicity three and geometric multiplicity two.

$$\boldsymbol{A} = \begin{bmatrix} \lambda & 1 & 1 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix}. \tag{145}$$

The eigenspace of $\boldsymbol{A}$ corresponding to the eigenvalue λ is

$$(\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{v} = \boldsymbol{0}_{\mathbb{R}^3} \quad \Leftrightarrow \quad \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad \Leftrightarrow \quad \begin{cases} v_1 \text{ arbitrary} \\ v_2 \text{ arbitrary} \\ v_3 = -v_2 \end{cases}. \tag{146}$$

Hence a basis for $N(\boldsymbol{A} - \lambda \boldsymbol{I})$ is

$$\boldsymbol{v}_1 = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} \qquad \boldsymbol{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}. \tag{147}$$

To complement $\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ to a basis of $\mathbb{R}^3$ we add a generalized eigenvector $\boldsymbol{v}_3$ that solves the following linear system[7]

$$(\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{v}_3 = \boldsymbol{v}_2. \tag{148}$$

We obtain

$$\begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} v_{31} \\ v_{32} \\ v_{33} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \Leftrightarrow \quad \begin{cases} v_{31} \text{ arbitrary} \\ v_{32}, \text{arbitrary} \\ v_{32} + v_{33} = 1 \end{cases}. \tag{149}$$

Hence a generalized eigenvector for the eigenspace $N(\boldsymbol{A} - \lambda \boldsymbol{I})$ is

$$\boldsymbol{v}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}. \tag{150}$$

We define the similarity trasformation $\boldsymbol{P}$ by using the eigenvectors $\begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 \end{bmatrix}$ and the generalized eigenvector $\boldsymbol{v}_3$ of $\boldsymbol{A}$

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \boldsymbol{v}_3 \end{bmatrix}. \tag{151}$$

Since $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$ are linearly independent we have that P is invertible. Clearly,

$$\boldsymbol{A}\boldsymbol{P} = \begin{bmatrix} \boldsymbol{A}\boldsymbol{v}_1 & \boldsymbol{A}\boldsymbol{v}_2 & \boldsymbol{A}\boldsymbol{v}_3 \end{bmatrix} = \begin{bmatrix} \lambda\boldsymbol{v}_1 & \lambda\boldsymbol{v}_2 & \boldsymbol{v}_2 + \lambda\boldsymbol{v}_3 \end{bmatrix} = \underbrace{\begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \boldsymbol{v}_3 \end{bmatrix}}_{\boldsymbol{P}} \underbrace{\begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix}}_{\boldsymbol{J}} = \boldsymbol{P}\boldsymbol{J}. \tag{152}$$

Jordan blocks. At this point it is clear that by computing the generalized eigenvectors it is always possible to construct a similarity transformation $\boldsymbol{P}$ that takes any matrix $\boldsymbol{A}$ into its Jordan canonical form

$$\boldsymbol{J} = \begin{bmatrix} \boldsymbol{J}_1 & & & \\ & \boldsymbol{J}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{J}_p \end{bmatrix}, \tag{153}$$

[7] Note there is really no reason why we should choose $\boldsymbol{v}_1$ instead of $\boldsymbol{v}_2$ at the right hand side of (148). In fact, the choice of both eigenvectors and generalized eigenvectors is not really unique.

where p is the total number of distinct eigenvalues of $\boldsymbol{A}$. The Jordan canonical form is a block-diagonal matrix in which each block $\boldsymbol{J}_i$ can be of the form summarized in Table 1.

Matrix exponentials of Jordan blocks. The matrix exponential of the Jordan form of (153) is a block-diagonal matrix that has the matrix exponential of each Jordan block along the diagonal.

$$e^{t\boldsymbol{J}} = \begin{bmatrix} e^{t\boldsymbol{J}_1} & & & \\ & e^{t\boldsymbol{J}_2} & & \\ & & \ddots & \\ & & & e^{t\boldsymbol{J}_p} \end{bmatrix}. \tag{154}$$

Hence, to compute the matrix exponential of the Jordan form of $\boldsymbol{A}$, we just need a formula for the matrix exponential of each Jordan block in Table 1. The case in which the Jordan block is diagonal is trivial, since the matrix exponential is just the exponential of the diagonal elements. For instance,

$$\boldsymbol{J}_i = \begin{bmatrix} \lambda_i & 0 \\ 0 & \lambda_i \end{bmatrix} \qquad \Rightarrow \qquad e^{t\boldsymbol{J}_i} = \begin{bmatrix} e^{t\lambda_i} & 0 \\ 0 & e^{t\lambda_i} \end{bmatrix}. \tag{155}$$

Let us now show how to compute the matrix exponential of the following Jordan blocks

$$\text{a)} \quad \boldsymbol{J}_i = \begin{bmatrix} \lambda_i & 1 \\ 0 & \lambda_i \end{bmatrix}, \qquad \text{b)} \quad \boldsymbol{J}_i = \begin{bmatrix} \lambda_i & 0 & 0 \\ 0 & \lambda_i & 1 \\ 0 & 0 & \lambda_i \end{bmatrix}, \qquad \text{c)} \quad \boldsymbol{J}_i = \begin{bmatrix} \lambda_i & 1 & 0 \\ 0 & \lambda_i & 1 \\ 0 & 0 & \lambda_i \end{bmatrix}. \tag{156}$$

a) Let us write the 2D Jordan block as

$$\boldsymbol{J}_i = \begin{bmatrix} \lambda_i & 1 \\ 0 & \lambda_i \end{bmatrix} = \underbrace{\begin{bmatrix} \lambda_i & 0 \\ 0 & \lambda_i \end{bmatrix}}_{\boldsymbol{B}_i} + \underbrace{\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}}_{\boldsymbol{C}}. \tag{157}$$

The matrix commutator of $\boldsymbol{B}_i$ and $\boldsymbol{C}$ equals zero. In fact,

$$[\boldsymbol{B}_i, \boldsymbol{C}] = \boldsymbol{B}_i\boldsymbol{C} - \boldsymbol{C}\boldsymbol{B}_i = \begin{bmatrix} \lambda_i & 0 \\ 0 & \lambda_i \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \lambda_i & 0 \\ 0 & \lambda_i \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}. \tag{158}$$

This implies that[8]

$$e^{t\boldsymbol{J}_i} = e^{t(\boldsymbol{B}_i + \boldsymbol{C})} = e^{t\boldsymbol{B}_i} e^{t\boldsymbol{C}}. \tag{162}$$

Since $\boldsymbol{B}_i$ is a diagonal matrix

$$e^{t\boldsymbol{B}_i} = \begin{bmatrix} e^{t\lambda_i} & 0 \\ 0 & e^{t\lambda_i} \end{bmatrix}. \tag{163}$$

[8] In general, given two square matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ we have

$$e^{\boldsymbol{A}} e^{\boldsymbol{B}} = e^{\boldsymbol{Z}}, \tag{159}$$

where

$$\boldsymbol{Z} = \boldsymbol{A} + \boldsymbol{B} + \frac{1}{2}[\boldsymbol{A}, \boldsymbol{B}] + \frac{1}{12}[\boldsymbol{A}, [\boldsymbol{A}, \boldsymbol{B}]] + \cdots. \tag{160}$$

This formula is known as Baker-Campbell-Hausdorff formula. If $\boldsymbol{A}$ and $\boldsymbol{B}$ commute, i.e., if $[\boldsymbol{A}, \boldsymbol{B}] = \boldsymbol{0}_{M_{n\times n}}$ then by (159) and (160) we have

$$e^{\boldsymbol{A}} e^{\boldsymbol{B}} = e^{\boldsymbol{A}+\boldsymbol{B}}. \tag{161}$$

Regarding the exponential of $\boldsymbol{C}$ we have the exact formula[9]

$$e^{t\boldsymbol{C}} = \boldsymbol{I} + t\boldsymbol{C} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + t \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}. \tag{165}$$

Finally, a substitution of (165) and (163) into (162) yields the desired expression

$$e^{t\boldsymbol{J}_i} = \begin{bmatrix} e^{t\lambda_i} & 0 \\ 0 & e^{t\lambda_i} \end{bmatrix} \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} e^{t\lambda_i} & te^{\lambda_i t} \\ 0 & e^{t\lambda_i} \end{bmatrix}. \tag{166}$$

b) The exponential of the 3D Jordan block

$$\boldsymbol{J}_i = \begin{bmatrix} \lambda_i & 0 & 0 \\ 0 & \lambda_i & 1 \\ 0 & 0 & \lambda_i \end{bmatrix} \tag{167}$$

can be computed using the formula (166) we just proved. In fact,

$$e^{t\boldsymbol{J}_i} = \begin{bmatrix} e^{t\lambda_i} & 0 & 0 \\ 0 & e^{t\lambda_i} & te^{\lambda_i t} \\ 0 & 0 & e^{t\lambda_i} \end{bmatrix}. \tag{168}$$

c) The exponential of the 3D Jordan block

$$\boldsymbol{J}_i = \begin{bmatrix} \lambda_i & 1 & 0 \\ 0 & \lambda_i & 1 \\ 0 & 0 & \lambda_i \end{bmatrix} \tag{169}$$

requires more work. We begin by splitting $\boldsymbol{J}_i$ as the sum of a diagonal matrix and and a non-diagonal matrix

$$\boldsymbol{J}_i = \begin{bmatrix} \lambda_i & 1 & 0 \\ 0 & \lambda_i & 1 \\ 0 & 0 & \lambda_i \end{bmatrix} = \underbrace{\begin{bmatrix} \lambda_i & 0 & 0 \\ 0 & \lambda_i & 0 \\ 0 & 0 & \lambda_i \end{bmatrix}}_{\boldsymbol{B}_i} + \underbrace{\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}}_{\boldsymbol{C}}. \tag{170}$$

As before, it is straightforward to show that $\boldsymbol{B}_i$ and $\boldsymbol{C}$ commute

$$[\boldsymbol{B}_i, \boldsymbol{C}] = \boldsymbol{B}_i\boldsymbol{C} - \boldsymbol{C}\boldsymbol{B}_i = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{171}$$

Moreover, by a direct calculation, we have

$$\boldsymbol{C}^2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad \boldsymbol{C}^3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad \dots, \qquad \boldsymbol{C}^k = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{172}$$

Therefore, the matrix exponential of the Jordan block (169) is

$$e^{t\boldsymbol{J}_i} = e^{t\boldsymbol{B}_i} e^{t\boldsymbol{C}} = e^{t\boldsymbol{B}_i} \left(\boldsymbol{I} + \boldsymbol{C} + \frac{\boldsymbol{C}^2}{2} \right). \tag{173}$$

[9] In fact,

$$\boldsymbol{C}^2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}. \tag{164}$$

Of course all matrix powers $\boldsymbol{C}^k$ are all zero for $k \geq 2$ since $\boldsymbol{C}^2 = \boldsymbol{0}$, and we can write $\boldsymbol{C}^k = \boldsymbol{C}^2\boldsymbol{C}^{k-2}$.

Substituting (172) into (173) finally yields

$$e^{t\boldsymbol{J}_i} = \begin{bmatrix} e^{t\lambda_i} & 0 & 0 \\ 0 & e^{t\lambda_i} & 0 \\ 0 & 0 & e^{t\lambda_i} \end{bmatrix} \underbrace{\begin{bmatrix} 1 & t & t^2/2 \\ 0 & 1 & t \\ 0 & 0 & 1 \end{bmatrix}}_{\boldsymbol{I}+\boldsymbol{C}+\boldsymbol{C}^2/2}. \tag{174}$$

Developing the product finally yields

$$e^{t\boldsymbol{J}_i} = \begin{bmatrix} e^{t\lambda_i} & te^{t\lambda_i} & t^2e^{t\lambda_i}/2 \\ 0 & e^{t\lambda_i} & te^{t\lambda_i} \\ 0 & 0 & e^{t\lambda_i} \end{bmatrix}. \tag{175}$$

The matrix exponential of all Jordan blocks we discussed in this section are summarized in Table 1. Formulas for matrix exponentials of higher-dimensional Jordan blocks can be computed by using the techniques we discussed in this section.

Appendix B: Matrix norms compatible with vector norms

Let us define the following matrix norm

$$\|\boldsymbol{A}\| = \sup_{\boldsymbol{y}\neq \boldsymbol{0}_{\mathbb{R}^n}} \frac{\|\boldsymbol{A}\boldsymbol{y}\|}{\|\boldsymbol{y}\|} = \sup_{\|\boldsymbol{y}\|=1} \|\boldsymbol{A}\boldsymbol{y}\|. \tag{176}$$

Clearly, $\|\boldsymbol{A}\|$ is matrix norm (prove it as exercise), which satisfies, by definition, the following inequality

$$\|\boldsymbol{A}\| \geq \frac{\|\boldsymbol{A}\boldsymbol{y}\|}{\|\boldsymbol{y}\|} \qquad \text{i.e.} \qquad \|\boldsymbol{A}\boldsymbol{y}\| \leq \|\boldsymbol{A}\| \, \|\boldsymbol{y}\|. \tag{177}$$

It is straightforward to show that

$$\|\boldsymbol{A}\|_\infty = \max_{i=1,..,n} \left(\sum_{j=1}^n |A_{ij}| \right), \tag{178}$$

$$\|\boldsymbol{A}\|_1 = \max_{j=1,..,n} \left(\sum_{i=1}^n |A_{ij}| \right), \tag{179}$$

$$\|\boldsymbol{A}\|_2 = \sqrt{\lambda_{\max}\left(\boldsymbol{A}^T\boldsymbol{A}\right)} = \sigma_{\max}(\boldsymbol{A}), \tag{180}$$

where $\sigma_{\max}(\boldsymbol{A})$ is the largest singular value of the matrix $\boldsymbol{A}$. For example,

$$\|\boldsymbol{A}\boldsymbol{y}\|_\infty = \max_{i=1,\dots,n} \left| \sum_{j=1}^n A_{ij}y_j \right| \leq \max_{i=1,\dots,n} \left(\sum_{j=1}^n |A_{ij}|\,|y_j| \right) \leq \|\boldsymbol{y}\|_\infty \max_{i=1,\dots,n} \left(\sum_{j=1}^n |A_{ij}| \right) \tag{181}$$

which implies that

$$\frac{\|\boldsymbol{A}\boldsymbol{y}\|_\infty}{\|\boldsymbol{y}\|_\infty} \leq \max_{i=1,\dots,n} \left(\sum_{j=1}^n |A_{ij}| \right) \qquad \text{for all } \boldsymbol{y} \neq \boldsymbol{0}_{\mathbb{R}^n}, \tag{182}$$

i.e.,

$$\sup_{\boldsymbol{y}\neq \boldsymbol{0}_{\mathbb{R}^n}} \frac{\|\boldsymbol{A}\boldsymbol{y}\|_\infty}{\|\boldsymbol{y}\|_\infty} = \max_{i=1,\dots,n} \left(\sum_{j=1}^n |A_{ij}| \right) = \|\boldsymbol{A}\|_\infty. \tag{183}$$

With any compatible matrix norm available we immediately see that the function $\boldsymbol{f}(\boldsymbol{y}) = \boldsymbol{A}\boldsymbol{y}$ is Lipschitz continuous in $\mathbb{R}^n$. In fact, we have

$$\|\boldsymbol{A}\boldsymbol{y}_1 - \boldsymbol{A}\boldsymbol{y}_2\| \leq \|\boldsymbol{A}\| \, \|\boldsymbol{y}_1 - \boldsymbol{y}_2\| \qquad \text{for all} \qquad \boldsymbol{y}_1, \boldsymbol{y}_2 \in \mathbb{R}^n, \tag{184}$$

where $L = \|\boldsymbol{A}\|$ is the Lipschitz constant.

Appendix C: Solution of a linear system in terms of the matrix exponential

We first write the ODE (1) as a linear integral equation

$$\boldsymbol{X}(t, \boldsymbol{x}_0) = \boldsymbol{x}_0 + \int_0^t \boldsymbol{A}\boldsymbol{X}(s, \boldsymbol{x}_0) ds.$$

To solve this equation we use the Picard iteration method, which is a fixed point iteration method. To this end, we define the iterative sequence

$$\boldsymbol{X}^{(n)}(t, \boldsymbol{x}_0) = \boldsymbol{x}_0 + \int_0^t \boldsymbol{A}\boldsymbol{X}^{(n-1)}(s, \boldsymbol{x}_0) ds \qquad \boldsymbol{X}^{(0)}(t, \boldsymbol{x}_0) = \boldsymbol{x}_0 \tag{185}$$

For nonlinear systems Picard's iterations usually converge only within a small time interval. On the other hand, for linear systems Picard's iterations are globally convergent. Let us we start with $n = 1$

$$\boldsymbol{X}^{(1)}(t, \boldsymbol{x}_0) = \boldsymbol{x}_0 + \int_0^t \boldsymbol{A}\boldsymbol{x}_0 ds = \boldsymbol{x}_0 + \boldsymbol{A}\boldsymbol{x}_0 t = (\boldsymbol{I} + \boldsymbol{A}t)\boldsymbol{x}_0.$$

We can use this to compute $n = 2$ which gives

$$\boldsymbol{X}^{(2)}(t, \boldsymbol{x}_0) = \boldsymbol{x}_0 + \int_0^t \boldsymbol{A}\boldsymbol{X}^{(1)}(s, \boldsymbol{x}_0) ds = \boldsymbol{x}_0 + \int_0^t \boldsymbol{A}(\boldsymbol{I} + \boldsymbol{A}t)\boldsymbol{x}_0 ds = \left(\boldsymbol{I} + \boldsymbol{A}t + \frac{t^2}{2}\boldsymbol{A}^2\right)\boldsymbol{x}_0.$$

By induction it is straightforward to show that

$$\boldsymbol{X}^{(n)}(t, \boldsymbol{x}_0) = \left(\sum_{k=0}^{n} \frac{\boldsymbol{A}^k t^k}{k!}\right)\boldsymbol{x}_0.$$

Clearly,

$$\boldsymbol{X}(t, \boldsymbol{x}_0) = \lim_{n\to\infty} \boldsymbol{X}^{(n)}(t, \boldsymbol{x}_0) = \left(\lim_{n\to\infty} \sum_{k=0}^{n} \frac{\boldsymbol{A}^k t^k}{k!}\right)\boldsymbol{x}_0 = e^{t\boldsymbol{A}}\boldsymbol{x}_0. \tag{186}$$