Ordinary Differential Equations

Lecture Notes for MATH 4051

FREDERICK TSZ-HO FONG

Hong Kong University of Science and Technology
(Version: January 10, 2020)

Contents

Preface

This is the lecture note used by the author at Brown University in 2013-14, and at HKUST in Spring 2020. The purpose of the course and the note is to give an introduction to Theory of Ordinary Differential Equations (ODEs). There is not much emphasis on solving ODEs, finding explicit solutions or applying numerical methods, but several theoretical issues of ODEs including existence, uniqueness, stability and periodicity are examined.

The major purpose of the course and the lecture notes is to strength students' knowledge on basic mathematical analysis by illustrating the use of analysis on the qualitative studies on differential equations.

The pre-requisite of the course is workable background mathematical analysis. Students should already have a good sense of concepts such as limit, continuity, differentiability of functions of both one and several variables. Basic linear algebra such as eigenvalues and eigenvectors are also needed. The course will essentially start from Section 1.3. Students should read Sections 1.2 by their own as a review of MATH 2351/2352.

Frederick Tsz-Ho Fong
January 10, 2020
HKUST, Clear Water Bay, Hong Kong

Chapter 1

Linear Systems

1.1. Systems of ODEs

Ordinary Differential Equations (ODEs) are equations whose unknowns functions have only one independent variable. You should have already encountered some of them in single-variable calculus, physics or engineering classes. Here are some examples:

- Exponential decay equation:
$$\frac{d}{dt}y(t) = -ky(t).$$
- Newton's Second Law:
$$m\frac{d^2}{dt^2}x(t) = F(x(t)).$$

The above examples are ODEs with a single unknown function. This course will mainly investigate **systems of ODEs**. A system of ODEs is a number of simultaneous ODEs with one or more unknown functions. In order for these functions to form a solution to the system, all ODEs in the system have to hold simultaneously. The **order** of the system refers to the highest derivative order involved in the equations. In this course, we will mostly focus on **first-order** systems, meaning that the system involves only first derivatives with respect to the independent variable.

Here is an example of a first-order system of ODEs (which governs the *glycolysis* inside human bodies):

$$\begin{aligned}\frac{dx}{dt} &= -x + ay + x^2y \\ \frac{dy}{dt} &= b - ay - x^2y.\end{aligned}$$

The variable t, often regarded as the *time*, is the **independent variable**. The functions $x(t)$ and $y(t)$ are the unknowns of the system, whereas a and b are constants.

Remark 1.1. Unless otherwise is stated, we always use t to denote the independent variable in this course. □

While there are many systems of ODEs of scientific interest, we will study systems of ODEs *regardless* of their roles in sciences. We will investigate ODE systems from *linear-algebraic*, *geometric* and *analytic* viewpoints and will focus on general ODEs which may not be related to any current scientific fields.

The general form of a first-order system of ODEs is:

$$\begin{aligned} x_1' &= F_1(x_1, \dots, x_d) \\ x_2' &= F_2(x_1, \dots, x_d) \\ &\vdots \\ x_d' &= F_d(x_1, \dots, x_d). \end{aligned}$$

It is a simultaneous system of d equations. In this system, $x_1(t), \dots, x_d(t)$ are the unknown functions of the system, and $F_1, \dots, F_d$ are given functions of d variables. We often put a system of ODEs in a **vector form**. The reasons for doing so are many-folded. One obvious reason, as we will investigate soon in a more detail, is that it allows the interaction between the ODE system and its *geometry*.

To represent an ODE system in vector form, we let $\mathbf{x}(t)$ be the **unknown vector** whose components are the unknown functions $x_1, \dots, x_d$, i.e.

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_d(t) \end{bmatrix}.$$

If the independent variable t is clear from the context, we may simply write the unknown vector as

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix}.$$

We also put the right-hand side of an ODE system into a vector by letting:

$$\mathbf{F}(x_1, \dots, x_d) = \begin{bmatrix} F_1(x_1, \dots, x_d) \\ \vdots \\ F_n(x_1, \dots, x_d) \end{bmatrix}.$$

Since the components $(x_1, \dots, x_d)$ can be represented[1] by $\mathbf{x}$, we can further abbreviate $\mathbf{F}(x_1, \dots, x_d)$ by $\mathbf{F}(\mathbf{x})$. Therefore, a general system of ODEs can be written in the form of:

$$\mathbf{x}' = \mathbf{F}(\mathbf{x}).$$

The vector form of an ODE system links the theory of ODEs with geometry. We think of a solution $\mathbf{x}(t)$ as a **parametrized curve**, $\mathbf{x}(t) = (x_1(t), \dots, x_d(t))$ in $\mathbb{R}^d$. The components $x_i(t)$'s give the coordinates of the trajectory at time t. The t-derivative, $\mathbf{x}'(t) = (x_1'(t), \dots, x_d'(t))$, represents the **tangent**, or **velocity**, vector of the curve at time t. Therefore, we will often call a solution $\mathbf{x}(t)$ as a **solution curve**, or a **trajectory**, etc.

The right-hand side of the equation, written as $\mathbf{F}(\mathbf{x})$, defines a **vector field** on $\mathbb{R}^d$. A vector field on $\mathbb{R}^d$ is a function that associates each point $\mathbf{x} \in \mathbb{R}^d$ to a vector. The vector field corresponding to the glycolysis system is shown in Figure 1.1

In order for $\mathbf{x}(t)$ to be a solution curve, it has to satisfy $\mathbf{x}' = \mathbf{F}(\mathbf{x})$. Geometrically, it means that the tangent vector $\mathbf{x}'(t)$ of the curve is at any time equal to the vector field $\mathbf{F}$ at the point $\mathbf{x}(t)$. To put it in an even simpler terms, the solution curve $\mathbf{x}(t)$ *flows* along the vector field $\mathbf{F}$ at any time. Figure 1.2 shows the relation between the family of solution curves (in red) and the vector field (in blue) of the glycolysis example.

---

[1] In this course, we make very little distinction between $(x_1, \dots, x_d)$ and the column vector $\begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix}$

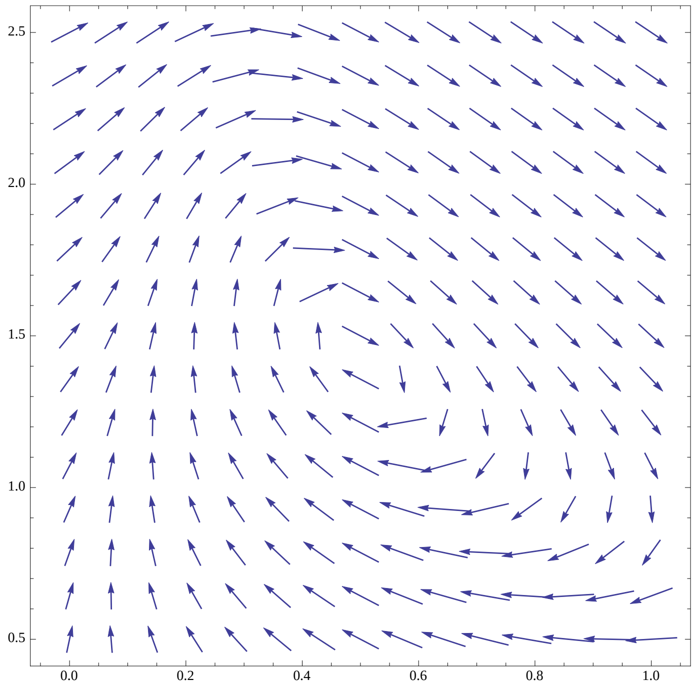


Figure 1.1. The vector field plot of the system: $\frac{dx}{dt} = -x + \frac{1}{10}y + x^2y$, $\frac{dy}{dt} = \frac{1}{2} - \frac{1}{10}y - x^2y$.

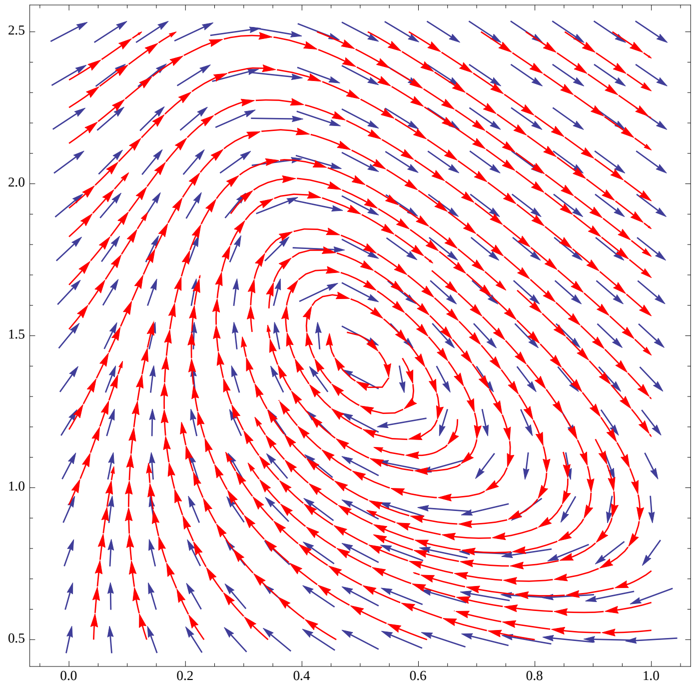


Figure 1.2. The vector field plot with solution curves with of the system: $\frac{dx}{dt} = -x + \frac{1}{10}y + x^2y$, $\frac{dy}{dt} = \frac{1}{2} - \frac{1}{10}y - x^2y$

A plot consisting of only solution curves of an ODE system is called the **phase portrait** of the system.

1.1.1. Second and higher order ODEs. The reason why we talk mostly about first-order ODE systems because any higher order ODEs can be reduced to a first-order ODE system, at the expense of having more unknown functions. Take the second-order ODE $x''+bx'+kx=0$ as an example. One can let $v=x'$, then $v'=x''=-bx'-kx=-bv-kx$. The second-order ODE is therefore equivalent to the following first-order system:

$$\begin{aligned} x' &= v \\ v' &= -kx - bv, \end{aligned}$$

or in matrix form:

$$\begin{bmatrix} x \\ v \end{bmatrix}' = \begin{bmatrix} 0 & 1 \\ -k & -b \end{bmatrix} \begin{bmatrix} x \\ v \end{bmatrix}.$$

Similarly, for a higher-order ODE $x^{(m)} + c_{m-1}x^{(m-1)} + \ldots + c_1 x' + c_0 x = 0$, we can let:

$$x_1 = x, \quad x_2 = x', \quad x_3 = x'', \ldots, \; x_m = x^{(m-1)}$$

or in short, $x_i = x^{(i-1)}$ for any $i = 1, 2, \ldots, m$. Then the higher-order ODE can be written as:

$$\begin{aligned} x_1' &= x_2 \\ x_2' &= x_3 \\ &\vdots \\ x_{m-1}' &= x_m \\ x_m' &= -c_0 x_1 - c_1 x_2 - \ldots - c_{m-1} x_m, \end{aligned}$$

or in matrix form:

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}' = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -c_0 & -c_1 & -c_2 & \cdots & -c_{m-1} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}.$$

1.2. Planar Linear Systems

This section examines **planar linear systems with constant coefficients** (for simplicity we will just call them **planar linear systems**) which can always be solved using Linear Algebra. We will derive the general solution of these systems and investigate their phase portraits. Theory of nonlinear ODEs, as you will see later, is heavily built upon linear theory.

We first give several important definitions:

Definition 1.2 (Linear Systems)**.** A system of ODE $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ is said to be **linear** if $\mathbf{F} : \mathbb{R}^d \to \mathbb{R}^d$ is a linear transformation[2], or equivalently, $\mathbf{F}(\mathbf{x})$ is of the form $\mathbf{F}(\mathbf{x}) = A\mathbf{x}$ where A is an $d \times d$ matrix of real numbers.

Definition 1.3 (Planar Linear Systems)**.** A **planar linear system** is a linear system defined on $\mathbb{R}^2$. That is, an ODE system of the form $\mathbf{x}' = A\mathbf{x}$ where A is a 2×2 matrix of real numbers.

Example 1.1. The system

$$x' = x + 3y$$
$$y' = x - y$$

is a planar linear system as it can be written in the following matrix-vector form:

$$\underbrace{\begin{bmatrix} x' \\ y' \end{bmatrix}}_{\mathbf{x}'} = \underbrace{\begin{bmatrix} 1 & 3 \\ 1 & -1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x \\ y \end{bmatrix}}_{\mathbf{x}}$$

□

The phase portrait of the system is shown in Figure 1.3.

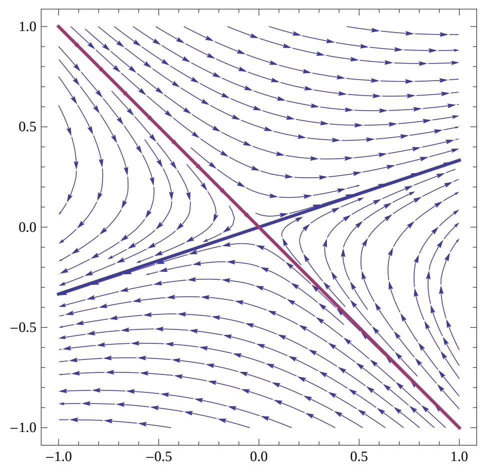


Figure 1.3. Phase portrait of the system $x' = x + 3y$, $\quad y' = x - y$. The curves represent solutions of the system.

It is not difficult to notice from the phase portrait that there are two *straight-line* solutions through the origin. We are going to show that if a linear system $\mathbf{x}' = A\mathbf{x}$ has a straight-line solution, it must be parallel to an **eigenvector** direction:

Proposition 1.4. *Suppose $\mathbf{x}(t)$ is a straight-line solution to the linear system $\mathbf{x}' = A\mathbf{x}$, then $\mathbf{x}(t) = Ce^{\lambda t}\mathbf{v}_0$ where $\mathbf{v}_0$ is an eigenvector of A with eigenvalue λ, and C is an arbitrary constant.*

Proof. Suppose $\mathbf{x}(t)$ is a straight-line solution, then $\mathbf{x}(t)$ is parallel to a fixed non-zero vector $\mathbf{v}_0$ at any time t, i.e. there exists a function $c(t)$ of t such that:

$$\mathbf{x}(t) = c(t)\mathbf{v}_0 \quad \text{for any } t \in \mathbb{R}.$$

Note that we do not set $\mathbf{x}(t) = c\mathbf{v}_0$ or $\mathbf{x}(t) = t\mathbf{v}_0$ because the former is a stationary solution and the latter travels at constant speed, neither of them may be true.

Substitute $\mathbf{x}(t) = c(t)\mathbf{v}_0$ into the system $\mathbf{x}' = A\mathbf{x}$, we have:

$$(c(t)\mathbf{v}_0)' = A(c(t)\mathbf{v}_0)$$
$$c'(t)\mathbf{v}_0 = c(t)A\mathbf{v}_0. \tag{1.1}$$

If $c(t_0) \neq 0$ for some $t_0 \in \mathbb{R}$, then we have $A\mathbf{v}_0 = \frac{c'(t_0)}{c(t_0)}\mathbf{v}_0$, which is equivalent to saying $\mathbf{v}_0$ is a eigenvector of A with eigenvalue $\frac{c'(t_0)}{c(t_0)} =: \lambda$. Now we have $A\mathbf{v}_0 = \lambda\mathbf{v}_0$, so by (1.1), we have

$$c'(t)\mathbf{v}_0 = \lambda c(t)\mathbf{v}_0$$
$$c'(t) = \lambda c(t) \qquad \text{(since } \mathbf{v}_0 \neq \mathbf{0})$$

It can be easily shown (see Exercise 1.1) using *separation of variables* that $c(t) = Ce^{\lambda t}$ for some constant C. Therefore, we have $\mathbf{x}(t) = Ce^{\lambda t}\mathbf{v}_0$. □

Exercise 1.1. Show that the general solution of the first-order ODE $c'(t) = \lambda c(t)$ is given by:

$$c(t) = Ce^{\lambda t}$$

where C is a constant.

As shown in Proposition 1.4, to determine the straight-line solutions (if there are any) of a linear system $\mathbf{x}' = A\mathbf{x}$, it amounts to finding the eigenvalues and eigenvectors of the matrix A. In the next subsection, we will first review the matrix algebra required.

1.2.1. Review of Linear Algebra.

We first recall the definition of eigenvectors and eigenvalues of a matrix.

Definition 1.5 (Eigenvectors and Eigenvalues)**.** Let A be a square matrix of real numbers. We say $\mathbf{v}$ is an **eigenvector** of A if $\mathbf{v} \neq 0$ and $A\mathbf{v} = \lambda\mathbf{v}$ for some scalar λ. The scalar λ is called an **eigenvalue** of the matrix A.

Remark 1.6. While an eigenvector $\mathbf{v}$ must be non-zero, an eigenvalue λ can be zero. □

To determine the eigenvalues of a matrix A, one may use:

Proposition 1.7. *A scalar λ is an eigenvalue of A if and only if $\det(A - \lambda I) = 0$.*

Proof. See any standard Linear Algebra textbook. □

The determinant equation $\det(A - \lambda I) = 0$ is called the **characteristic equation** of A, and $\det(A - \lambda I)$ is called the **characteristic polynomial** of A. Proposition 1.7 tells us that to find the eigenvalues of A, one can solve the characteristic polynomial. After the eigenvalues are all determined, we can find their corresponding eigenvectors by solving systems of linear equations. Here is an example:

Example 1.2. Consider the matrix

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

The characteristic equation is given by:

$$\begin{aligned}
0 &= \det(A - \lambda I) \\
&= \det\left(\begin{bmatrix} 1 & 3 \\ 1 & -1 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) \\
&= \begin{vmatrix} 1-\lambda & 3 \\ 1 & -1-\lambda \end{vmatrix} \\
&= (1-\lambda)(-1-\lambda) - 3 \\
&= \lambda^2 - 4 \\
&= (\lambda - 2)(\lambda + 2).
\end{aligned}$$

Therefore, $\lambda = 2$ or $\lambda = -2$. The matrix A has two eigenvalues: 2 and -2.

To find the corresponding eigenvectors of each eigenvalue, we simply solve the system $A\mathbf{v} = \lambda\mathbf{v}$ for $\mathbf{v}$.

For $\lambda = 2$, $A\mathbf{v} = 2\mathbf{v}$ can be written as:

$$\begin{bmatrix} 1 & 3 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 2 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

which can be rewritten as:

$$\begin{aligned}
x_1 + 3x_2 &= 2x_1 \\
x_1 - x_2 &= 2x_2
\end{aligned}$$

Both equations can be rearranged as: $x_1 = 3x_2$. Therefore, one of the equations is considered to be redundant, and any vector $\mathbf{v} = (3x_2, x_2)$ with $x_2 \neq 0$ is an eigenvector of A with eigenvalue 2.

The eigenvectors corresponding to eigenvalue -2 can be found by the same way. It can be verified that any vector $\mathbf{w} = (x_1, -x_1)$ with $x_1 \neq 0$ is an eigenvector with eigenvalue -2. It is left as an exercise for readers. See Exercise 1.2 below. □

Exercise 1.2. Find all eigenvectors corresponding to eigenvalue -2 of the matrix:

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

For each eigenvalue of $A = \begin{bmatrix} 1 & 3 \\ 1 & -1 \end{bmatrix}$ we have found, let's pick an eigenvector with that eigenvalue:

$$\lambda_1 = 2 \qquad \mathbf{v}_1 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

$$\lambda_2 = -2 \qquad \mathbf{v}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Then by Proposition 1.4, both

$$\begin{aligned}
\mathbf{x}_1(t) &= e^{\lambda_1 t}\mathbf{v}_1 = e^{2t}\begin{bmatrix} 3 \\ 1 \end{bmatrix} \\
\mathbf{x}_2(t) &= e^{\lambda_2 t}\mathbf{v}_2 = e^{-2t}\begin{bmatrix} 1 \\ -1 \end{bmatrix}
\end{aligned}$$

are straight-line solutions of the system $\mathbf{x}' = A\mathbf{x}$.

By linearity, it can be easily verified (see Exercise 1.3) that any linear combination of $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$ is also a solution to the system.

Exercise 1.3. Show that if $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$ are both solutions to the linear system $\mathbf{x}' = A\mathbf{x}$, then any $c_1\mathbf{x}_1(t) + c_2\mathbf{x}_2(t)$, where c_1 and c_2 are constants, is also a solution to the linear system $\mathbf{x}' = A\mathbf{x}$.

In the next subsection, we will show in fact $c_1e^{2t}\begin{bmatrix}3\\1\end{bmatrix} + c_2e^{-2t}\begin{bmatrix}1\\-1\end{bmatrix}$ form **all** possible solutions of the system $\mathbf{x}' = A\mathbf{x}$ where A is defined as above.

Exercise 1.4. Find all eigenvalues and their corresponding eigenvectors of the following matrices:

(i) $\begin{bmatrix}1 & 2\\0 & 3\end{bmatrix}$ (ii) $\begin{bmatrix}1 & 2\\3 & 6\end{bmatrix}$ (iii) $\begin{bmatrix}1 & 2\\1 & 0\end{bmatrix}$ (iv) $\begin{bmatrix}1 & 2\\3 & -3\end{bmatrix}$.

Exercise 1.5. Rewrite the second-order ODE $x'' + bx' + kx = 0$, where b and k are constants, into a first-order planar linear system. Determine the range of values of b and k for which the matrix of the planar system has real, distinct eigenvalues.

1.2.2. Distinct Eigenvalues. Given a 2×2 matrix A with characteristic polynomial given by

$$\det(A - \lambda I) = (\lambda - \lambda_1)(\lambda - \lambda_2)$$

where λ_1 and λ_2 are two distinct real numbers, then the matrix A has distinct eigenvalues λ_1 and λ_2. If for each $i = 1, 2$, we let $\mathbf{v}_i$ be an eigenvector with eigenvalue λ_i, then we have seen that

$$\mathbf{x}(t) = c_1e^{\lambda_1 t}\mathbf{v}_1 + c_2e^{\lambda_2 t}\mathbf{v}_2$$

are solutions to the system $\mathbf{x}' = A\mathbf{x}$ for any constants c_1 and c_2. We will prove in fact they form **all** possible solutions of the system. To establish this, we need the following fundamental fact about eigenvectors:

Theorem 1.8. *Eigenvectors of a square matrix with distinct eigenvalues must be linearly independent. Precisely, if $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are eigenvectors of a matrix A with distinct eigenvalues $\lambda_1, \ldots, \lambda_k$ respectively, then $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are linearly independent, meaning that whenever*

$$c_1\mathbf{v}_1 + \ldots + c_k\mathbf{v}_k = \mathbf{0}$$

we must have $c_1 = \ldots = c_k = 0$.

Proof. The proof with an arbitrary integer k can be found in any standard Linear Algebra textbook. Here we give the proof for the case $k = 2$, i.e. two vectors involved.

Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be two eigenvectors of A with distinct eigenvalues λ_1 and λ_2 respectively, i.e. $A\mathbf{v}_1 = \lambda_1\mathbf{v}_1$ and $A\mathbf{v}_2 = \lambda_2\mathbf{v}_2$.

Suppose c_1 and c_2 are two scalars such that

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = \mathbf{0}. \tag{1.2}$$

We need to show $c_1 = c_2 = 0$.

Multiply A from the left on both sides of the above vector equation, we get:

$$\begin{aligned} A(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) &= \mathbf{0}\\ c_1A\mathbf{v}_1 + c_2A\mathbf{v}_2 &= \mathbf{0}\\ c_1\lambda_1\mathbf{v}_1 + c_2\lambda_2\mathbf{v}_2 &= \mathbf{0} \qquad (1.3) \end{aligned}$$

Subtract (1.3) by $\lambda_1 \times$ (1.2), we get:

$$c_2(\lambda_2 - \lambda_1)\mathbf{v}_2 = \mathbf{0}.$$

Since $\lambda_1 \neq \lambda_2$ and $\mathbf{v}_2 \neq 0$ (being an eigenvector), therefore we must have $c_2 = 0$. Substitute $c_2 = 0$ into (1.2), we get $c_1\mathbf{v}_1 = \mathbf{0}$, and hence $c_1 = 0$ too as $\mathbf{v}_1 \neq \mathbf{0}$ being an eigenvector. □

Back to the planar linear system we are considering, if the matrix A has two distinct eigenvalues λ_1 and λ_2, with corresponding eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$ respectively, then these two vectors are **linearly independent**. Recall that any $d \times d$ matrix whose columns are linearly independent vectors in $\mathbb{R}^d$ must be invertible. This observation will be crucial to establish the following:

Theorem 1.9. *Given a planar linear system $\mathbf{x}' = A\mathbf{x}$ whose matrix A has two distinct real eigenvalues λ_1 and λ_2. Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be two eigenvectors of A with eigenvalues λ_1 and λ_2 respectively, then any solution $\mathbf{x}(t)$ to the planar linear system $\mathbf{x}' = A\mathbf{x}$ can be expressed as:*

$$\mathbf{x}(t) = c_1 e^{\lambda_1 t}\mathbf{v}_1 + c_2 e^{\lambda_2 t}\mathbf{v}_2 \tag{1.4}$$

where c_1 and c_2 are any real constants.

In order to establish Theorem 1.9, we review how to *diagonalize* matrix A using the eigen-data.

Lemma 1.10. *Given a 2×2 matrix A with two distinct real eigenvalues λ_1 and λ_2. Suppose $\mathbf{v}_1$ and $\mathbf{v}_2$ are corresponding eigenvectors with eigenvalues λ_1 and λ_2 respectively. Then the matrix A can be decomposed into the following diagonal form:*

$$A = \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2\end{bmatrix} \begin{bmatrix}\lambda_1 & 0 \\ 0 & \lambda_2\end{bmatrix} \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2\end{bmatrix}^{-1}. \tag{1.5}$$

Here $\begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2\end{bmatrix}$ is a 2×2 matrix whose columns are $\mathbf{v}_1$ and $\mathbf{v}_2$.

Remark 1.11. Note that the matrix $\begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2\end{bmatrix}$ is invertible since $\mathbf{v}_1$ and $\mathbf{v}_2$ are linearly independent. □

Proof of Lemma 1.10. Since $\mathbf{v}_i$'s are eigenvectors of A, we have:

$$\begin{aligned} A\begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2\end{bmatrix} &= \begin{bmatrix}A\mathbf{v}_1 & A\mathbf{v}_2\end{bmatrix} \\ &= \begin{bmatrix}\lambda_1\mathbf{v}_1 & \lambda_2\mathbf{v}_2\end{bmatrix} \\ &= \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2\end{bmatrix}\begin{bmatrix}\lambda_1 & 0 \\ 0 & \lambda_2\end{bmatrix}. \end{aligned}$$

As $\mathbf{v}_1$ and $\mathbf{v}_2$ are eigenvectors with distinct eigenvalues, they are linearly independent and so the matrix $\begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2\end{bmatrix}$ is invertible. By multiplying $\begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2\end{bmatrix}^{-1}$ on both sides from the right, we have

$$A = \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2\end{bmatrix} \begin{bmatrix}\lambda_1 & 0 \\ 0 & \lambda_2\end{bmatrix} \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2\end{bmatrix}^{-1}$$

which is desired. □

Now we are ready to prove the theorem using the diagonal decomposition of the matrix.

Proof of Theorem 1.9. For simplicity, we denote $P = \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2\end{bmatrix}$ and $D = \begin{bmatrix}\lambda_1 & 0 \\ 0 & \lambda_2\end{bmatrix}$. Should P be invertible, we have $A = PDP^{-1}$. The diagonal decomposition of A has a nice consequence toward solving linear ODE systems: if we let $\mathbf{y} = P^{-1}\mathbf{x}$, then $\mathbf{x} = P\mathbf{y}$ and the ODE system $\mathbf{x}' = A\mathbf{x}$ can be expressed as:

$$\begin{aligned}(P\mathbf{y})' &= A(P\mathbf{y}) \\ P\mathbf{y}' &= AP\mathbf{y} \quad \text{(since } P \text{ is a constant matrix)} \\ P\mathbf{y}' &= PDP^{-1}P\mathbf{y} \\ P\mathbf{y}' &= PD\mathbf{y} \\ \mathbf{y}' &= D\mathbf{y}.\end{aligned}$$

Therefore under this change of variables $\mathbf{y} = P^{-1}\mathbf{x}$, the ODE system $\mathbf{x}' = A\mathbf{x}$ can be transformed into $\mathbf{y}' = D\mathbf{y}$ where D is a diagonal matrix. If we write $\mathbf{y}(t) = \begin{bmatrix}y_1(t) \\ y_2(t)\end{bmatrix}$, then $\mathbf{y}' = D\mathbf{y}$ in equation form is:

$$\begin{aligned}y_1'(t) &= \lambda_1 y_1(t) \\ y_2'(t) &= \lambda_2 y_2(t).\end{aligned}$$

The system for $y_i(t)$'s is a *decoupled* one, and the solution for $y_i(t)$'s can be found by separation of variables. Precisely,

$$y_1(t) = c_1 e^{\lambda_1 t}, \qquad y_2(t) = c_2 e^{\lambda_2 t}$$

where c_1 and c_2 are constants. In other words, the general solution for $\mathbf{y}' = D\mathbf{y}$ is:

$$\mathbf{y}(t) = \begin{bmatrix}c_1 e^{\lambda_1 t} \\ c_2 e^{\lambda_2 t}\end{bmatrix} = c_1 e^{\lambda_1 t}\begin{bmatrix}1 \\ 0\end{bmatrix} + c_2 e^{\lambda_2 t}\begin{bmatrix}0 \\ 1\end{bmatrix}.$$

Since $\mathbf{x}$ and $\mathbf{y}$ are related via the matrix P, i.e. $\mathbf{x} = P\mathbf{y}$, the general solution for the original system $\mathbf{x}' = A\mathbf{x}$ is:

$$\mathbf{x}(t) = P\mathbf{y}(t) = \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2\end{bmatrix}\left(c_1 e^{\lambda_1 t}\begin{bmatrix}1 \\ 0\end{bmatrix} + c_2 e^{\lambda_2 t}\begin{bmatrix}0 \\ 1\end{bmatrix}\right).$$

It is easy to verify that $\begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2\end{bmatrix}\begin{bmatrix}1 \\ 0\end{bmatrix} = \mathbf{v}_1$ and $\begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2\end{bmatrix}\begin{bmatrix}0 \\ 1\end{bmatrix} = \mathbf{v}_2$. Therefore,

$$\mathbf{x}(t) = c_1 e^{\lambda_1 t}\mathbf{v}_1 + c_2 e^{\lambda_2 t}\mathbf{v}_2$$

which is exactly what we need to prove. . □

Remark 1.12. Since (1.4) gives all possible solutions of the system $\mathbf{x}' = A\mathbf{x}$ with real distinct eigenvalues, it is called the **general solution** of the system. Note that (1.4) applies only to the case where A has two distinct real eigenvalues. We have not accounted for the cases where A has *complex* or *repeated* eigenvalues. □

1.2.2.1. Phase portrait of planar linear systems with distinct real eigenvalues. The general solution formula (1.4) allows us to sketch the phase portrait of the planar system $\mathbf{x}' = A\mathbf{x}$ where A has distinct real eigenvalues.

The fate of each solution curve $\mathbf{x}(t) = c_1 e^{\lambda_1 t}\mathbf{v}_1 + c_2 e^{\lambda_2 t}\mathbf{v}_2$ as $t \to \pm\infty$ depends on the signs of the eigenvalues λ_1 and λ_2: the limit of $e^{\lambda t}$ as $t \to \pm\infty$ is either 0, 1 or ∞ when λ is negative, zero or positive respectively.

Let's first assume that both λ_1 and λ_2 are non-zero, and without loss of generality, assume $\lambda_1 < \lambda_2$.

If $\lambda_1 > 0$ and $\lambda_2 > 0$, then $\mathbf{x}(t) = c_1 e^{\lambda_1 t}\mathbf{v}_1 + c_2 e^{\lambda_2 t}\mathbf{v}_2$ approaches $\begin{bmatrix} \pm\infty \\ \pm\infty \end{bmatrix}$ as $t \to \infty$ (except the case where $c_1 = c_2 = 0$), while $\mathbf{x}(t) \to \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ as $t \to -\infty$. Therefore, the solution curves in the phase portrait are *moving way* from the origin. We call this type of phase portrait a **source**. See Figure 1.4.

If $\lambda_1 < 0$ and $\lambda_2 < 0$, then $\mathbf{x}(t) = c_1 e^{\lambda_1 t}\mathbf{v}_1 + c_2 e^{\lambda_2 t}\mathbf{v}_2$ approaches $\begin{bmatrix} \pm\infty \\ \pm\infty \end{bmatrix}$ as $t \to -\infty$ (except the case where $c_1 = c_2 = 0$), while $\mathbf{x}(t) \to \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ as $t \to +\infty$. Therefore, the solution curves in the phase portrait are *attracted* to the origin. We call this type of phase portrait a **sink**. See Figure 1.5.

Now suppose $\lambda_1 < 0 < \lambda_2$, or in other words, the eigenvalues of A are of different signs. Recall that $\mathbf{x}(t) = c_1 e^{\lambda_1 t}\mathbf{v}_1 + c_2 e^{\lambda_2 t}\mathbf{v}_2$. As $t \to \infty$, the coefficient $c_1 e^{\lambda_1 t}$ of $\mathbf{v}_1$ is vanishing while the other coefficient $c_2 e^{\lambda_2 t}$ of $\mathbf{v}_2$ tends to $\pm\infty$ (here again assume c_1 and c_2 are non-zero). Therefore, as time increases, the $\mathbf{v}_1$-part becomes more dominant than the $\mathbf{v}_2$-part, so the solution curves will approach asymptotically to the line spanned by $\mathbf{v}_1$. The reverse scenario happens as $t \to -\infty$. We call this type of phase portrait a **saddle**. See Figure 1.6.

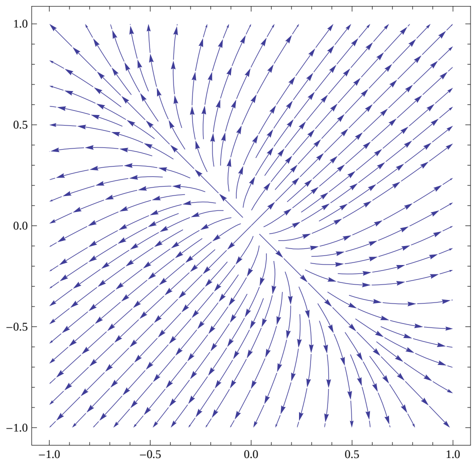


Figure 1.4. The diagram shows a **source** phase portrait. Solution curves are tending away from the origin. Also, it can be noticed that the solution curves are tangent to one of the straight-line solutions as $t \to -\infty$.

In all three cases (source, sink or saddle), there are two straight-line solutions corresponding to the two linearly independent eigenvectors of A. For the source and sink cases, you can observe from Figures 1.4 and 1.5 that the curves are approaching tangentially to one of the straight-line solutions as t tends to either $+\infty$ or $-\infty$. To which eigen-direction the curves approach depends on the how close λ_1 and λ_2 are from 0:

Proposition 1.13. *Suppose $\mathbf{x}' = A\mathbf{x}$ is a planar linear system with distinct, non-zero, real eigenvalues λ_1 and λ_2 of the same sign. Denote $\mathbf{v}_1$ and $\mathbf{v}_2$ the eigenvectors with eigenvalues λ_1 and λ_2 respectively. Assume $|\lambda_1| < |\lambda_2|$, then the solution curves in the phase portrait are tangent to the eigenvector $\mathbf{v}_1$ as they approach to the origin.*

Proof. We will only prove the source case, i.e. the solution curves tend to the origin as $t \to -\infty$. The sink case can be proved in a similar way, *mutatis mutandis*. While it

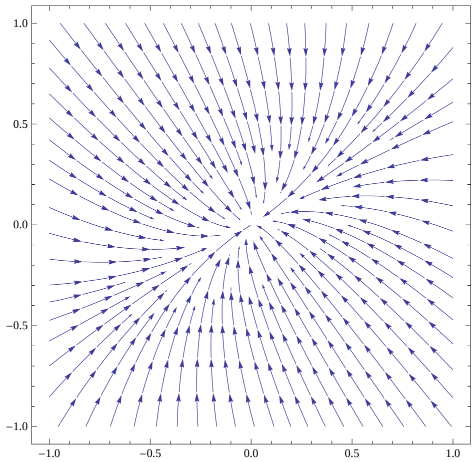


Figure 1.5. The diagram shows a **sink** phase portrait. Solution curves approach to the origin tangentially to one of the straight-line solutions as time increases.

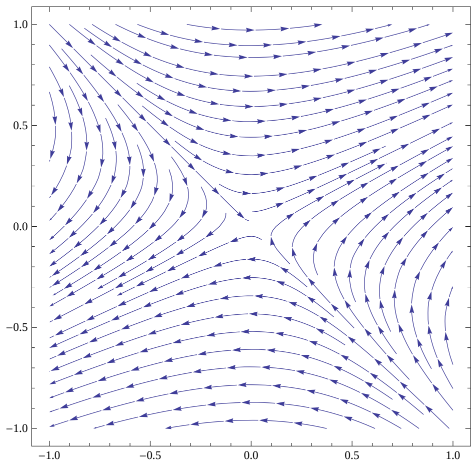


Figure 1.6. The diagram shows a **saddle** phase portrait. Solution curves are asymptotic to the straight-line solutions as $t \to \pm\infty$.

is possible to prove this proposition by computing the asymptotic slope of the position vector and show that it is the same as the slope of the line spanned by $\mathbf{v}_1$, there is a smarter way to establish this using a little bit linear algebra.

Recall from the proof of Theorem 1.9 that a solution $\mathbf{x}(t)$ to the system $\mathbf{x}' = A\mathbf{x}$ can be expressed as $\mathbf{x}(t) = P\mathbf{y}(t)$ where $P = \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2\end{bmatrix}$ and $\mathbf{y}(t)$ is a solution to the diagonalized system:

$$\mathbf{y}' = \begin{bmatrix}\lambda_1 & 0 \\ 0 & \lambda_2\end{bmatrix}\mathbf{y}.$$

Therefore, the phase portrait of the $\mathbf{x}$-system can be obtained by transforming the phase portrait of the $\mathbf{y}$-system via the linear map associated to the matrix P. This transformation takes the standard basis vectors $(1, 0)$ and $(0, 1)$ in the $\mathbf{y}$-portrait to the eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$ respectively in the $\mathbf{x}$-portrait.

As such, to establish this proposition it suffices to show the solution curves in the $\mathbf{y}$-system approach to the y_1-axis as $t \to -\infty$. The general solution to the $\mathbf{y}$-system is:

$$\mathbf{y}(t) = \begin{bmatrix} c_1 e^{\lambda_1 t} \\ c_2 e^{\lambda_2 t} \end{bmatrix}.$$

Letting $\mathbf{y} = \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix}$, one can compute:

$$\frac{dy_2}{dy_1} = \frac{y_2'(t)}{y_1'(t)} = \frac{c_1 e^{\lambda_2 t}}{c_2 e^{\lambda_1 t}} = \frac{c_1}{c_2} e^{(\lambda_2 - \lambda_1)t}.$$

As we assume that $0 < \lambda_1 < \lambda_2$, we have $\lambda_2 - \lambda_1 > 0$ and so $\frac{dy_2}{dy_1} \to 0$ as $t \to -\infty$. Therefore, the solution curves (with non-zero c_1 and c_2) in the $\mathbf{y}$-portrait approach to the origin tangentially to the y_1-axis, and hence the solution curves in the $\mathbf{x}$-portrait approach to the origin tangentially to the line spanned by $\mathbf{v}_1$. □

Exercise 1.6. Complete the proof of the Proposition 1.13 for the sink case.

Next we account for the case where one of the eigenvalues λ_1 and λ_2 is zero. Without loss of generality, let's assume $\lambda_1 = \lambda \neq 0$ and $\lambda_2 = 0$. Then the general solution to the planar system $\mathbf{x}' = A\mathbf{x}$ can be expressed as:

$$\mathbf{x}(t) = c_1 e^{\lambda t}\mathbf{v}_1 + c_2\mathbf{v}_2.$$

The phase portrait of this case is drastically different from the previous three cases we have investigated. First take $c_2 = 0$, then the solution $\mathbf{x}(t) = c_1 e^{\lambda t}\mathbf{v}_1$ is a straight-line solution through the origin and parallel to $\mathbf{v}_1$. Whether this straight trajectory tends to or away from the origin depends on the sign of λ. Adding $c_2\mathbf{v}_2$, which is a constant vector, will parallel translate the solution line by $c_2\mathbf{v}_2$. How much it translates depends on the value of c_2. Therefore, the phase portrait is a parallel family of straight-line solutions as shown in Figure 1.7.

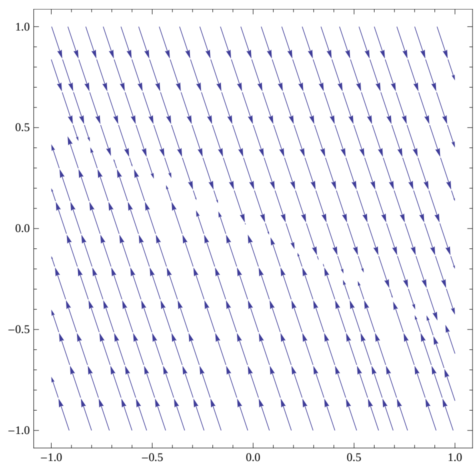


Figure 1.7. The phase portrait of the system $x' = x + 2y, \quad y' = -3x - 6y$ The eigenvalues are 5 with an eigenvector $(1, 3)$, and 0 with an eigenvector $(-1, 2)$.

Note that the line spanned by the null-eigenvector[3] $\mathbf{v}_2$ is not a solution line of the system, but instead each individual *point* is a stationary solution of the system.

[3] A null-eigenvector is an eigenvector with eigenvalue 0.

Exercise 1.7. For each of the following matrix A: (i) find the general solution of the system $\mathbf{x}' = A\mathbf{x}$, (ii) state the phase portrait type and (iii) sketch the phase portrait.

(a) $A = \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}$ (b) $A = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix}$ (c) $A = \begin{bmatrix} -5 & 1 \\ 4 & -2 \end{bmatrix}$.

1.2.3. Complex Eigenvalues. We have settled all possible cases with distinct real eigenvalues in the previous subsection. However, it does happen often in *real* life that complex eigenvalues can occur even for a real 2×2 matrix. A quick example is the matrix $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ whose eigenvalues are $\pm i$. We are going to deal with the complex case in this subsection.

Throughout this subsection (and for most part of the course), A always denote a **real matrix** although its eigenvalues may be complex. Since the characteristic polynomial $\det(A - \lambda I)$ is real, if λ is a complex root then so does $\overline{\lambda}$. In other words, the complex eigenvalues of a real matrix A must come in conjugate pair. If A is 2×2 with complex eigenvalue $\lambda = \alpha + \beta i$ (where $\beta \neq 0$), then the other eigenvalue must be $\alpha - \beta i$ and so A has distinct eigenvalues.

Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be complex eigenvectors corresponding to eigenvalues λ_1 and λ_2 respectively. Although one can mimic what was done for the distinct real case to find the general solution of $\mathbf{x}' = A\mathbf{x}$ with complex eigenvalues, it would only give us a general **complex** solution. Let illustrate this by letting $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. Readers should verify that the eigen-data are:

$$\lambda_1 = i \qquad \mathbf{v}_1 = \begin{bmatrix} i \\ 1 \end{bmatrix}$$
$$\lambda_2 = -i \qquad \mathbf{v}_2 = \begin{bmatrix} -i \\ 1 \end{bmatrix}.$$

The **complex** general solution is then given by:

$$\mathbf{x}(t) = c_1 e^{it} \begin{bmatrix} i \\ 1 \end{bmatrix} + c_2 e^{-it} \begin{bmatrix} -i \\ 1 \end{bmatrix}, \qquad c_1, c_2 \in \mathbb{C}.$$

It left much to be desired because we began with a *real* problem but end up having a *complex* answer! It is not trivial to figure out the geometry of the phase portrait.

Our approach of deriving the **real** general solution formula will be as follows:

(1) We begin by considering matrices of the form $Q = \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}$ where α and β are real numbers and $\beta \neq 0$. These matrices have complex eigenvalues $\alpha \pm \beta i$. We will derive the **real** general solution of the system $\mathbf{y}' = Q\mathbf{y}$. See Theorem 1.14.

(2) Next, we will use a bit Linear Algebra to show that **any** real 2×2 matrix A with complex eigenvalues can be decomposed into the form of:
$$A = PQP^{-1}$$
where P is a *suitably chosen* invertible matrix and Q is a matrix of the form $\begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}$.

(3) Finally, as in the diagonal case, we make a change of variables by letting $\mathbf{y}(t) = P^{-1}\mathbf{x}(t)$. Then $\mathbf{x}(t)$ solves $\mathbf{x}' = A\mathbf{x}$ if and only if $\mathbf{y}(t)$ solves $\mathbf{y}' = Q\mathbf{y}$. Since the real general solution of the $\mathbf{y}$-system can be found, the real general solution of the $\mathbf{x}$-system can be derived via the relation $\mathbf{x}(t) = P\mathbf{y}(t)$.

1.2.3.1. General solution of systems in complex canonical form. We call a 2×2 matrix of the form $Q = \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}$ a **complex canonical form**. The real general solution with matrix in this form is stated in the following theorem:

Theorem 1.14. *The real general solution of the system* $\mathbf{y}' = Q\mathbf{y}$ *where* $Q = \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}$, *with* α, $\beta \in \mathbb{R}$ *and* $\beta \neq 0$, *is given by:*

$$\mathbf{y}(t) = c_1 e^{\alpha t} \begin{bmatrix} \cos \beta t \\ -\sin \beta t \end{bmatrix} + c_2 e^{\alpha t} \begin{bmatrix} \sin \beta t \\ \cos \beta t \end{bmatrix}, \qquad c_1, c_2 \in \mathbb{R}. \tag{1.6}$$

Outline of Proof. We let $\mathbf{y}(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix}$ be a real solution of the system $\mathbf{y}' = Q\mathbf{y}$, which can be written as equations:

$$\begin{aligned} y_1' &= \alpha y_1 + \beta y_2 \\ y_2' &= -\beta y_1 + \alpha y_2. \end{aligned}$$

Define a function $f : \mathbb{R} \to \mathbb{C}$ by:

$$f(t) = (y_1(t) + iy_2(t))e^{(-\alpha + i\beta)t}.$$

One can verify by straight-forward computation that $f'(t) = 0$ provided that $\mathbf{y}(t)$ is a solution to the system $\mathbf{y}' = Q\mathbf{y}$. Let $f(t) = c_1 + ic_2$ where c_1, $c_2 \in \mathbb{R}$. It can be verified that this implies:

$$y_1(t) + iy_2(t) = e^{\alpha t}(c_1 \cos \beta t + c_2 \sin \beta t) + ie^{\alpha t}(-c_1 \sin \beta t + c_2 \cos \beta t)$$

and that $\mathbf{y}(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix}$ equals to the expression given by (1.6). □

Exercise 1.8. Verify all the detail left-out in the proof of Theorem 1.14.

Example 1.3. Let $Q = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ whose eigenvalues are $\pm i$, i.e $\alpha = 0$ and $\beta = -1$. By Theorem 1.14, the real general solution of the system $\mathbf{y}' = Q\mathbf{y}$ is given by:

$$\mathbf{y}(t) = c_1 \begin{bmatrix} \cos t \\ \sin t \end{bmatrix} + c_2 \begin{bmatrix} -\sin t \\ \cos t \end{bmatrix}, \qquad c_1, c_2 \in \mathbb{R}.$$

□

1.2.3.2. Phase portrait of planar linear systems in complex canonical form. The $\sin$ and $\cos$ terms of the real general solution formula (1.6) suggests that the solution curves in the phase portrait should be *circular* or *spiral* depending whether α is zero or not.

When $\alpha > 0$, $e^{\alpha t} \to 0$ as $t \to -\infty$ whereas $e^{\alpha t} \to +\infty$ as $t \to +\infty$. Since $|\cos \beta t| \leq 1$ and $|\sin \beta t| \leq 1$, a simple squeezing argument shows that $\mathbf{y}(t)$ tends *away* from the origin. In contrast with the real distinct case, the solution curves are spiraling around the origin (see Figure 1.8). We call this type of phase portrait a **spiral source**.

Similarly, when $\alpha < 0$, $e^{\alpha t} \to 0$ as $t \to +\infty$. The solution curves $\mathbf{y}(t)$ tend *towards* the origin as $t \to +\infty$ in a spirally manner (see Figure 1.9). We call this type of phase portrait a **spiral sink**.

If $\alpha = 0$, the phase portrait consists of concentric circles centered at the origin (see Figure 1.10). This type of phase portrait is called a **center**.

The sign of α, i.e. the real part of the complex eigenvalue, determines the phase portrait type of the planar linear system. The sign of β determines the **orientation** of the solution curves. From the expression of the general solution (1.6), the solution curves $\begin{bmatrix} \cos\beta t \\ -\sin\beta t \end{bmatrix}$ and $\begin{bmatrix} \sin\beta t \\ \cos\beta t \end{bmatrix}$ determine the clockwise/counterclockwise orientation of the phase portrait.

If $\beta > 0$**::** then the solution curves travel in the clockwise direction as t increases; whereas

If $\beta < 0$**::** then the solution curves travel in the counterclockwise direction as t increases.

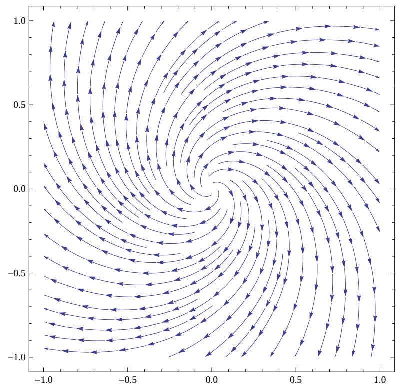


Figure 1.8. A spiral source: phase portrait of a system with $\alpha > 0$ and $\beta > 0$

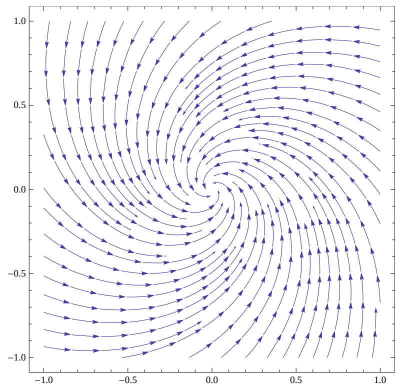


Figure 1.9. A spiral sink: phase portrait of a system with $\alpha < 0$ and $\beta < 0$

1.2.3.3. Complex canonical decomposition of 2×2 matrices. In the distinct real case, we decompose the matrix A into the form of $A = PDP^{-1}$ where D is a diagonal matrix. By the change of variable $\mathbf{y} = P^{-1}\mathbf{x}$, it was shown in the proof of Theorem 1.9 that $\mathbf{x}(t)$ is a solution to the system $\mathbf{x}' = A\mathbf{x}$ if and only if $\mathbf{y}(t)$ is a solution to the system $\mathbf{y}' = D\mathbf{y}$. The $\mathbf{y}$-system can be easily solved.

We will do an analogous change of variables for the complex eigenvalues case, but instead of having a diagonal matrix in the middle, we have a complex canonical form. The next lemma shows that *any* 2×2 matrix A with complex eigenvalues can be decomposed as $A = PQP^{-1}$ where Q is a complex canonical form. We will call this the **complex canonical decomposition** of the matrix A.

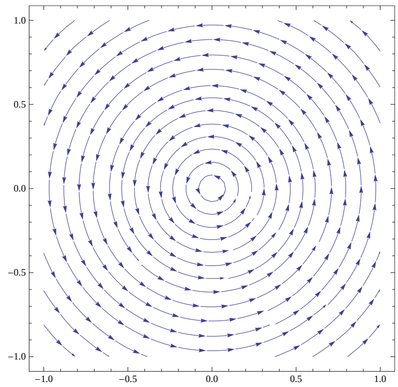


Figure 1.10. A center: phase portrait of a system with $\alpha = 0$ and $\beta < 0$

Lemma 1.15. *Let A be a 2×2 real matrix with complex eigenvalues $\alpha \pm \beta i$. Suppose $\mathbf{v}$ is a complex eigenvector of A with eigenvalue $\alpha + \beta i$. Denote the real and imaginary parts of $\mathbf{v}$ by $\mathbf{v}_{\text{Re}}$ and $\mathbf{v}_{\text{Im}}$ respectively, i.e. $\mathbf{v} = \mathbf{v}_{\text{Re}} + i\mathbf{v}_{\text{Im}}$. Then:*

$$A = \begin{bmatrix} \mathbf{v}_{\text{Re}} & \mathbf{v}_{\text{Im}} \end{bmatrix} \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix} \begin{bmatrix} \mathbf{v}_{\text{Re}} & \mathbf{v}_{\text{Im}} \end{bmatrix}^{-1}.$$

Proof. The given matrix A has complex eigenvalues $\alpha \pm \beta i$ which are in conjugate pair. It is also crucial to observe that their corresponding eigenvectors are also in conjugate pair. To show this, let $\mathbf{v}$ be a complex eigenvector of A with eigenvalue $\alpha + \beta i$. Then $A\mathbf{v} = (\alpha + \beta i)\mathbf{v}$ and we have:

$$\begin{aligned} \overline{A\mathbf{v}} &= \overline{(\alpha + \beta i)\mathbf{v}} && \text{(take conjugate on both sides)} \\ \overline{A}\overline{\mathbf{v}} &= \overline{(\alpha + \beta i)}\overline{\mathbf{v}} && \text{(multiplicative property of conjugate)} \\ A\overline{\mathbf{v}} &= (\alpha - \beta i)\overline{\mathbf{v}} && \text{(since } A \text{ is real)} \end{aligned}$$

Therefore $\overline{\mathbf{v}}$ is an eigenvector of A with eigenvalue $\alpha - \beta i$.

Now that $\mathbf{v}$ and $\overline{\mathbf{v}}$ are eigenvectors with distinct eigenvalues. By Theorem 1.8, which also applies to the complex case, they are linearly independent. It is not difficult to verify that Lemma 1.10 also holds for complex eigenvectors with distinct complex eigenvalues. Therefore, we have:

$$A = \begin{bmatrix} \mathbf{v} & \overline{\mathbf{v}} \end{bmatrix} \begin{bmatrix} \alpha + \beta i & 0 \\ 0 & \alpha - \beta i \end{bmatrix} \begin{bmatrix} \mathbf{v} & \overline{\mathbf{v}} \end{bmatrix}^{-1}. \tag{1.7}$$

Note that $\begin{bmatrix} \mathbf{v} & \overline{\mathbf{v}} \end{bmatrix}$ is invertible because the columns are linearly independent.

Finally, the last step and also the key part of the proof is to relate $\begin{bmatrix} \mathbf{v} & \overline{\mathbf{v}} \end{bmatrix}$ with $\begin{bmatrix} \mathbf{v}_{\text{Re}} & \mathbf{v}_{\text{Im}} \end{bmatrix}$. It can be easily seen that:

$$\begin{bmatrix} \mathbf{v} & \overline{\mathbf{v}} \end{bmatrix} = \begin{bmatrix} \mathbf{v}_{\text{Re}} + i\mathbf{v}_{\text{Im}} & \mathbf{v}_{\text{Re}} - i\mathbf{v}_{\text{Im}} \end{bmatrix} = \begin{bmatrix} \mathbf{v}_{\text{Re}} & \mathbf{v}_{\text{Im}} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}.$$

Taking the determinant on both sides, we get:

$$\det \begin{bmatrix} \mathbf{v} & \overline{\mathbf{v}} \end{bmatrix} = \det \begin{bmatrix} \mathbf{v}_{\text{Re}} & \mathbf{v}_{\text{Im}} \end{bmatrix} \cdot \det \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}.$$

Since $\mathbf{v}$ and $\overline{\mathbf{v}}$ are linearly independent, we have $\det \begin{bmatrix} \mathbf{v} & \overline{\mathbf{v}} \end{bmatrix} \neq 0$ and so $\det \begin{bmatrix} \mathbf{v}_{\text{Re}} & \mathbf{v}_{\text{Im}} \end{bmatrix} \neq 0$ as well. The matrix $\begin{bmatrix} \mathbf{v}_{\text{Re}} & \mathbf{v}_{\text{Im}} \end{bmatrix}$ is invertible. By substituting $\begin{bmatrix} \mathbf{v} & \overline{\mathbf{v}} \end{bmatrix} = \begin{bmatrix} \mathbf{v}_{\text{Re}} & \mathbf{v}_{\text{Im}} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}$

into (1.7), we arrive at:

$$\begin{aligned} A &= \left(\begin{bmatrix} \mathbf{v}_{\mathrm{Re}} & \mathbf{v}_{\mathrm{Im}} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}\right) \begin{bmatrix} \alpha+\beta i & 0 \\ 0 & \alpha-\beta i \end{bmatrix} \left(\begin{bmatrix} \mathbf{v}_{\mathrm{Re}} & \mathbf{v}_{\mathrm{Im}} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}\right)^{-1} \\ &= \begin{bmatrix} \mathbf{v}_{\mathrm{Re}} & \mathbf{v}_{\mathrm{Im}} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix} \begin{bmatrix} \alpha+\beta i & 0 \\ 0 & \alpha-\beta i \end{bmatrix} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{v}_{\mathrm{Re}} & \mathbf{v}_{\mathrm{Im}} \end{bmatrix}^{-1}. \end{aligned}$$

The proof can be completed by verifying

$$\begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix} \begin{bmatrix} \alpha+\beta i & 0 \\ 0 & \alpha-\beta i \end{bmatrix} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}^{-1} = \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}.$$

This is left as an exercise. □

Exercise 1.9. Verify that $\begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix} \begin{bmatrix} \alpha+\beta i & 0 \\ 0 & \alpha-\beta i \end{bmatrix} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}^{-1} = \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}$.

With Lemma 1.15, every 2×2 matrix A with complex eigenvalues admits a complex canonical form $A = PQP^{-1}$. Since the real general solution of $\mathbf{y}' = Q\mathbf{y}$ was already derived in Theorem 1.14, we can now solve for any planar system $\mathbf{x}' = A\mathbf{x}$ with complex eigenvalues via the relation $\mathbf{x} = P\mathbf{y}$. Precisely, we have the following:

Theorem 1.16. *Let A be a 2×2 matrix with complex eigenvalues $\alpha \pm \beta i$, where $\alpha, \beta \in \mathbb{R}$ and $\beta \neq 0$, and $\mathbf{v} = \mathbf{v}_{\mathrm{Re}} + i\mathbf{v}_{\mathrm{Im}}$ be a complex eigenvector (with real and imaginary parts $\mathbf{v}_{\mathrm{Re}}$ and $\mathbf{v}_{\mathrm{Im}}$ respectively) of A with eigenvalue $\alpha + \beta i$. Then, the real general solution of the system $\mathbf{x}' = A\mathbf{x}$ is given by:*

$$\mathbf{x}(t) = P\left(c_1 e^{\alpha t} \begin{bmatrix} \cos\beta t \\ -\sin\beta t \end{bmatrix} + c_2 e^{\alpha t} \begin{bmatrix} \sin\beta t \\ \cos\beta t \end{bmatrix}\right), \qquad c_1, c_2 \in \mathbb{R}, \tag{1.8}$$

where $P = \begin{bmatrix} \mathbf{v}_{\mathrm{Re}} & \mathbf{v}_{\mathrm{Im}} \end{bmatrix}$.

The proof is omitted here since it is in principle the same as in the distinct real case, i.e. Theorem 1.9. The major difference is that the diagonal matrix D now becomes a complex canonical form Q, and the transformation matrix P has a different form. The solving of $\mathbf{y}' = D\mathbf{y}$ in Theorem 1.9 is now replaced by the use of Theorem 1.14.

Exercise 1.10. Write down the whole proof of Theorem 1.16.

1.2.3.4. Phase portrait of planar linear systems with complex eigenvalues. Theorem 1.16 asserts that the real general solution of any planar linear system $\mathbf{x}' = A\mathbf{x}$ with complex eigenvalues can be obtained by multiplying an invertible matrix P by the general solution of the corresponding system $\mathbf{y}' = Q\mathbf{y}$ in complex canonical form, so the phase portrait of the $\mathbf{x}$-system is the transformed image of the $\mathbf{y}$-portrait by the linear map associated to P. The type of the phase portrait (spiral source, spiral sink or center) is preserved since A and Q have the same complex eigenvalues. The solution curves, however, may be distorted, rotated, and more elliptic than those in a complex canonical form system. See Figure 1.11 as an example.

However, one should note the orientation of the solution curves in the $\mathbf{x}$-portrait may not be the same as in the $\mathbf{y}$-portrait. It is because the linear transformation associated to P may change the orientation! From Linear Algebra, a linear transformation associated to a matrix P *preserves* the orientation if $\det(P) > 0$, and *reverses* the orientation if $\det(P) < 0$. Therefore, to determine the orientation of the solution curves, one should take both the sign of β and the sign of $\det(P)$ into account.

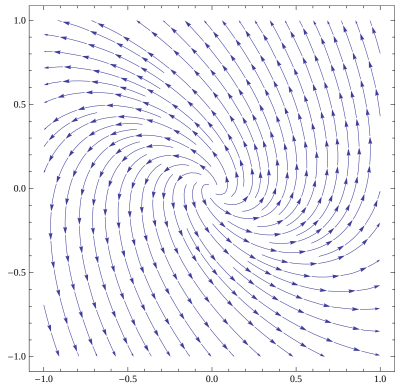


Figure 1.11. A spiral source: phase portrait of the system: $x' = x - 4y, \quad y' = 4x + 5y$

To summarize the above discussion, we let $\mathbf{x}' = A\mathbf{x}$ be a planar system with complex eigenvalue $\alpha \pm \beta i$, and $\mathbf{v} = \mathbf{v}_{\text{Re}} + i\mathbf{v}_{\text{Im}}$ be a complex eigenvector of A with eigenvalue $\alpha + \beta i$. Let $P = \begin{bmatrix} \mathbf{v}_{\text{Re}} & \mathbf{v}_{\text{Im}} \end{bmatrix}$. The phase portrait is a:

- spiral source if $\alpha > 0$;
- spiral sink if $\alpha < 0$;
- center if $\alpha = 0$.

The orientation can be determined by Table 1 below.

Table 1. Orientation of solution curves with complex eigenvalues

	$\det(P) > 0$	$\det(P) < 0$
$\beta > 0$	clockwise	counterclockwise
$\beta < 0$	counterclockwise	clockwise

1.2.4. Repeated Eigenvalues. The last case for deriving the general solution of a planar linear system $\mathbf{x}' = A\mathbf{x}$ is that the matrix A has a repeated real eigenvalue λ. It happens when the characteristic polynomial $\det(A - zI)$ can be factorized as $(z - \lambda)^2$. Note that in the planar case, it is not possible to get a repeated complex eigenvalue since complex roots of a real polynomial must appear in conjugate pairs. However, readers should note that repeated complex eigenvalues may appear in higher dimensional linear systems.

Back to planar systems, suppose from now on A is a 2×2 real matrix with a repeated eigenvalue λ. In order to find the general solution of $\mathbf{x}' = A\mathbf{x}$, we introduce the **Jordan canonical form** which, in 2×2 case, is a matrix of the form:

$$J = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}.$$

We are going to show that *any* 2×2 matrix A with a repeated eigenvalue λ can be decomposed as $A = PJP^{-1}$ for a suitable invertible matrix P (unless A is simply λI). We call this a **Jordan decomposition**. Then, in order find the general solution of $\mathbf{x}' = A\mathbf{x}$, we do the same '$\mathbf{y}$-trick' again: let $\mathbf{y}(t) = P^{-1}\mathbf{x}(t)$, then $\mathbf{x}(t)$ solves $\mathbf{x}' = A\mathbf{x}$ if and only if $\mathbf{y}(t)$ solves $\mathbf{y}' = J\mathbf{y}$. The general solution to the system with a Jordan canonical form J as the matrix can always be found. As in the diagonal and complex case, the general solution of the $\mathbf{x}$-system can be obtained by multiplying the matrix P to the $\mathbf{y}$-system.

Walking on a similar path as in the complex case, we first find the general solution to the planar system $\mathbf{y}' = J\mathbf{y}$ where J is a Jordan canonical form.

Theorem 1.17. *The general solution to the system* $\mathbf{y}' = J\mathbf{y}$ *where* $J = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}$ *is:*

$$\mathbf{y}(t) = c_1 e^{\lambda t} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + c_2 e^{\lambda t} \begin{bmatrix} t \\ 1 \end{bmatrix}, \qquad c_1,\ c_2 \in \mathbb{R}. \tag{1.9}$$

Proof. Let $\mathbf{y}(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix}$ and rewrite the system $\mathbf{y}' = J\mathbf{y}$ in equation form:

$$\begin{aligned} y_1' &= \lambda y_1 + y_2 \\ y_2' &= \lambda y_2. \end{aligned}$$

Clearly, the second equation gives $y_2 = c_2 e^{\lambda t}$ for some $c_2 \in \mathbb{R}$. Substitute this into the first equation, we get:

$$y_1' = \lambda y_1 + c_2 e^{\lambda t}$$

where is an ODE of one unknown function y_1. Some trial of separation of variables should convince you that this equation is not separable. Fortunately, it can be solved by so-called the method of *integration factor*. Multiply $e^{-\lambda t}$ on both sides of the equation. After rearranging, we get:

$$e^{-\lambda t} y_1' - \lambda e^{-\lambda t} y_1 = c_2.$$

It is worthwhile to note that the left-hand side can be rewritten as a total differential. Precisely, we have:

$$\frac{d}{dt}(e^{-\lambda t} y_1) = e^{-\lambda t} y_1' - \lambda e^{-\lambda t} y_1 = c_2.$$

Reader should verify this using the product rule. It then implies

$$e^{-\lambda t} y_1 = c_2 t + c_1, \quad c_1 \in \mathbb{R}.$$

Finally, multiplying both sides by $e^{\lambda t}$ yields:

$$y_1 = c_1 e^{\lambda t} + c_2 t e^{\lambda t},$$

and so

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} c_1 e^{\lambda t} + c_2 t e^{\lambda t} \\ c_2 e^{\lambda t} \end{bmatrix} = c_1 e^{\lambda t} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + c_2 e^{\lambda t} \begin{bmatrix} t \\ 1 \end{bmatrix},$$

as desired. □

1.2.4.1. Phase portrait of planar linear systems in Jordan canonical form. As can be seen from (1.9), the fate of the solution curves $\mathbf{y}(t)$ as $t \to \pm\infty$ is completely determined by the sign of λ.

When $\lambda > 0$, $\mathbf{y}(t)$ tends to 0 as $t \to -\infty$ while it blows up as $t \to +\infty$. The phase portrait (see Figure 1.12) is a **source**.

When $\lambda < 0$, $\mathbf{y}(t)$ tends to 0 as $t \to +\infty$ while it blows up as $t \to -\infty$. The phase portrait (see Figure 1.13) is a **sink**.

These sources and sinks, however, look a bit different from the case of distinct real eigenvalues. There is only one straight-line solution, which happens when $c_2 = 0$, in contrast to the distinct real case.

It can be easily verified that when $c_2 \neq 0$, the slope is given by:

$$\frac{dy_2}{dy_1} = \frac{y_2'}{y_1'} = \frac{c_2 \lambda e^{\lambda t}}{c_1 \lambda e^{\lambda t} + c_2(e^{\lambda t} + \lambda t e^{\lambda t})} = \frac{c_2 \lambda}{c_1 \lambda + c_2(1 + \lambda t)}$$

which tends to 0 as $t \to \pm\infty$. Therefore, the solution curves is tangent to the y_1-axis as it approaches to the origin and also become more and more horizontal as they blow up.

The degenerate case again happens when $\lambda = 0$. In this case, the general solution becomes:

$$\mathbf{y}(t) = c_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} t \\ 1 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + t \begin{bmatrix} c_2 \\ 0 \end{bmatrix}.$$

The phase portrait of the degenerate case is a family of parallel lines parallel to the y_1-axis. When $c_2 > 0$, the solution line is traveling in the positive y_1-direction, whereas when $c_2 < 0$, the solution line is traveling in the negative y_1-direction. However, when $c_2 = 0$, the solution is a stationary point $(c_1, 0)$. Therefore, the y_1-axis is not a solution line, but each *point* on the y_1-axis by itself is a stationary solution to the system (see Figure 1.14)

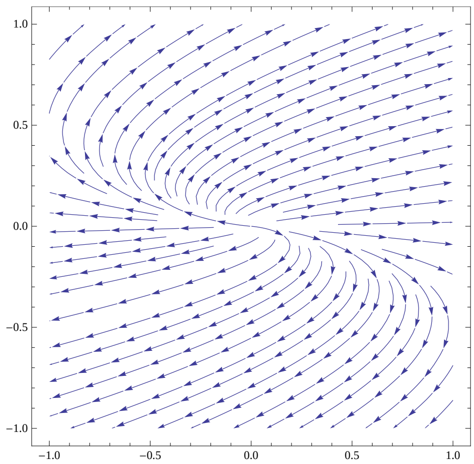


Figure 1.12. A source where the matrix of the system with a positive repeated eigenvalue.

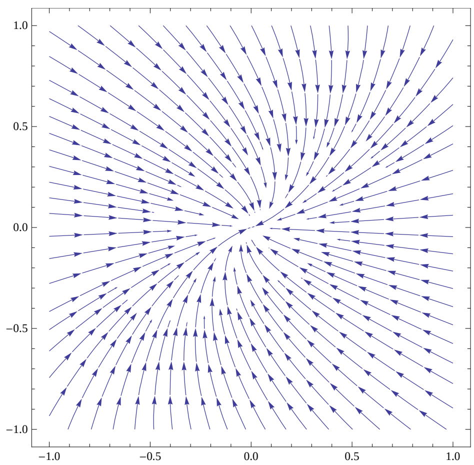


Figure 1.13. A sink where the matrix of the system with a negative repeated eigenvalue.

1.2.4.2. Jordan decomposition of 2×2 matrices with repeated eigenvalues. We are going to show that every 2×2 matrix A with repeated eigenvalues must admit a Jordan decomposition $A = PJP^{-1}$. After this is established, one can find the general solution of all planar linear systems with such matrices using the '$\mathbf{y}$-trick' as in the

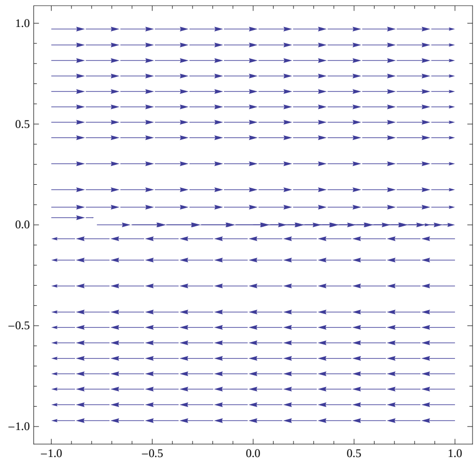


Figure 1.14. A degenerate type of phase portrait of the system with a zero repeated eigenvalue.

diagonal and complex cases. In order to establish this, we need the following celebrated theorem in Linear Algebra:

Theorem 1.18 (Cayley-Hamilton's Theorem)**.** *Any square matrix A satisfies its own characteristic polynomial. Precisely, suppose A is $d \times d$ and its characteristic polynomial is given by:*

$$\det(A - xI) = c_0 + c_1 x + c_2 x^2 + \ldots + c_d x^d.$$

Then, the following holds:

$$c_0 I + c_1 A + c_2 A^2 + \ldots + c_d A^d = 0.$$

The proof of this theorem can be found, for instance, in Friedberg–Insel–Spence's Linear Algebra book. The special case where A is a 2×2 matrix can be verified by direct computations:

Exercise 1.11. Let A be an arbitrary 2×2 matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$. First verify that the characteristic polynomial $\det(A - xI)$ is given by:

$$x^2 - (a + d)x + (ad - bc).$$

Then, verify by direct computation that A satisfies:

$$A^2 - (a + d)A + (ad - bc)I = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

In particular, if A has a repeated eigenvalue λ, what can you say about the matrix $(A - \lambda I)^2$?

Theorem 1.18 will be crucial to prove that any 2×2 matrix with a repeated eigenvalue must have a Jordan decomposition.

Lemma 1.19. *Let A be a 2×2 matrix with a repeated eigenvalue λ. Then, either $A = \lambda I$, or there exists two linearly independent vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ satisfying*

$$(A - \lambda I)\mathbf{v}_1 = 0 \qquad \textit{(i.e. } \mathbf{v}_1 \textit{ is an eigenvector of } A\textit{)}$$
$$(A - \lambda I)\mathbf{v}_2 = \mathbf{v}_1 \qquad \textit{(note that } \mathbf{v}_2 \textit{ is not an eigenvector of } A\textit{)}$$

such that:

$$A = \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2\end{bmatrix} \begin{bmatrix}\lambda & 1 \\ 0 & \lambda\end{bmatrix} \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2\end{bmatrix}^{-1}. \tag{1.10}$$

Proof. Given that λ is an eigenvalue and A is 2×2, there are two cases to consider:

- $\dim \text{null}(A - \lambda I) = 2$, or
- $\dim \text{null}(A - \lambda I) = 1$.

Here null$(A - \lambda I)$ denotes the null space of the matrix $A - \lambda I$. Any non-zero vector in null$(A - \lambda I)$ is an eigenvector of A with eigenvalue λ.

The first case implies $A = \lambda I$ (see Exercise below). For the second case $\dim \text{null}(A - \lambda I) = 1$, suppose null$(A - \lambda I) = $ span$(\mathbf{v}_1)$, then $\mathbf{v}_1$ is an eigenvector of A and any other eigenvector of A must be a constant multiple of $\mathbf{v}_1$. Pick any vector $\mathbf{w}$ in $\mathbb{R}^2$ such that $\mathbf{v}_1$ and $\mathbf{w}$ are linearly independent, then we consider the vector $(A - \lambda I)\mathbf{w}$.

Since A has a repeated eigenvalue λ, its characteristic polynomial must be given by:

$$\det(A - xI) = (x - \lambda)^2.$$

By Theorem 1.18, the matrix A satisfies its own characteristic polynomial, and so

$$(A - \lambda I)^2 = 0.$$

Therefore, we have $(A - \lambda I)[(A - \lambda I)\mathbf{w}] = (A - \lambda I)^2\mathbf{w} = 0$. This shows $(A - \lambda I)\mathbf{w}$ is an eigenvector of A with eigenvalue λ, and by the assumption of this case that null$(A - \lambda I) = $ span$(\mathbf{v}_1)$, we must have

$$(A - \lambda I)\mathbf{w} = c\mathbf{v}_1 \quad \text{for some } c \in \mathbb{R}.$$

This c cannot be zero, otherwise $(A - \lambda I)\mathbf{w} = \mathbf{0}$ then $\mathbf{w}$ is also an eigenvector of A with eigenvalue λ and hence $\mathbf{w}$ would a constant multiple of $\mathbf{v}_1$. It contradicts to our choice of $\mathbf{w}$ that $\mathbf{v}_1$ and $\mathbf{w}$ have to be linearly independent.

Now $c \neq 0$, we have:

$$(A - \lambda I)\frac{\mathbf{w}}{c} = \mathbf{v}_1.$$

Define $\mathbf{v}_2 := \frac{\mathbf{w}}{c}$, then $\mathbf{v}_1$ and $\mathbf{v}_2$ are a pair of linearly independent vectors that satisfy

$$(A - \lambda I)\mathbf{v}_1 = 0$$
$$(A - \lambda I)\mathbf{v}_2 = \mathbf{v}_1.$$

Finally, these relations imply $A\mathbf{v}_1 = \lambda\mathbf{v}_1$ and $A\mathbf{v}_2 = \mathbf{v}_1 + \lambda\mathbf{v}_2$ and we are left to verify (1.10):

$$\begin{aligned} A\begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2\end{bmatrix} &= \begin{bmatrix}A\mathbf{v}_1 & A\mathbf{v}_2\end{bmatrix} \\ &= \begin{bmatrix}\lambda\mathbf{v}_1 & \mathbf{v}_2 + \lambda\mathbf{v}_1\end{bmatrix} \\ &= \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2\end{bmatrix}\begin{bmatrix}\lambda & 1 \\ 0 & \lambda\end{bmatrix} \end{aligned}$$

Since $\mathbf{v}_1$ and $\mathbf{v}_2$ are linearly independent, the matrix $\begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2\end{bmatrix}$ is invertible and therefore (1.10) can be proved by multiplying $\begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2\end{bmatrix}^{-1}$ on both sides from the right.

$\square$

Exercise 1.12. Complete the first case in the proof of Lemma 1.19, i.e. show that if $\dim \text{null}(A - \lambda I) = 2$ for a 2×2 matrix A, then $A = \lambda I$.

Using the '$\mathbf{y}$-trick' which we have been using, we can now derive the general solution of planar linear systems $\mathbf{x}' = A\mathbf{x}$ where A has a repeated eigenvalue λ.

Theorem 1.20. *Given a 2×2 matrix A with a repeated eigenvalue λ. If $A = \lambda I$, the general solution to the system $\mathbf{x}' = A\mathbf{x}$ is:*

$$\mathbf{x}(t) = e^{\lambda t} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}.$$

Otherwise if $A \neq \lambda I$, then A admits a Jordan decomposition $A = P \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix} P^{-1}$ for some invertible matrix P, as shown in Lemma 1.19, and hence the general solution of the system $\mathbf{x}' = A\mathbf{x}$ is:

$$\mathbf{x}(t) = P \left(c_1 e^{\lambda t} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + c_2 e^{\lambda t} \begin{bmatrix} t \\ 1 \end{bmatrix} \right), \qquad c_1\ c_2 \in \mathbb{R}. \tag{1.11}$$

Proof. For the first case if $A = \lambda I$, the system $\mathbf{x}' = A\mathbf{x}$ is simply two decoupled ODEs:

$$x_1' = \lambda x_1$$
$$x_2' = \lambda x_2.$$

Solving each equation separately yields $x_1(t) = c_1 e^{\lambda t}$ and $x_2 = c_2 e^{\lambda t}$. Hence,

$$\mathbf{x}(t) = e^{\lambda t} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}.$$

For the second case $A \neq \lambda I$ and so A admits a Jordan decomposition $A = P \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix} P^{-1}$. Let $\mathbf{y} = P^{-1}\mathbf{x}$, then $\mathbf{x}(t)$ solves $\mathbf{x}' = A\mathbf{x}$ if and only if $\mathbf{y}(t)$ solves $\mathbf{y}' = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix} \mathbf{y}$. By Theorem 1.17,

$$\mathbf{y}(t) = c_1 e^{\lambda t} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + c_2 e^{\lambda t} \begin{bmatrix} t \\ 1 \end{bmatrix}.$$

Since $\mathbf{x}(t) = P\mathbf{y}(t)$, we have (1.11) as desired. □

1.2.4.3. Phase portrait of planar linear systems with a repeated eigenvalue. For a planar linear system $\mathbf{x}' = A\mathbf{x}$ where A has a repeated eigenvalue λ, if $A = \lambda I$ then by its general solution formula the phase portrait is a family of straight-lines passing through the origin. Whether it tends to or away from the origin depends on the sign of λ.

For the *generic* case where A has a Jordan decomposition $A = PJP^{-1}$. The general solution formula (1.11) tells us that the phase portrait of the system $\mathbf{x}' = A\mathbf{x}$ can be obtained by transforming that of the system $\mathbf{y}' = J\mathbf{y}$ by the linear map associated to P. Since P transforms the vector $(1, 0)$ on the $\mathbf{y}$-portrait to an eigenvector $\mathbf{v}_1$ of A in the $\mathbf{x}$-portrait, the $\mathbf{x}$-portrait consists of solution curves which are tangent at the origin to the line spanned $\mathbf{v}_1$. An example of such a phase portrait is shown in Figure 1.15.

Exercise 1.13. Find a Jordan decomposition of each matrix A below, i.e. express the matrix A in the form of PJP^{-1} where P is an invertible matrix and J is a Jordan canonical form. Then, write down the general solution of the system $\mathbf{x}' = A\mathbf{x}$, and find the solution $\mathbf{x}(t)$ with initial condition $\mathbf{x}(0) = (1, -2)$.

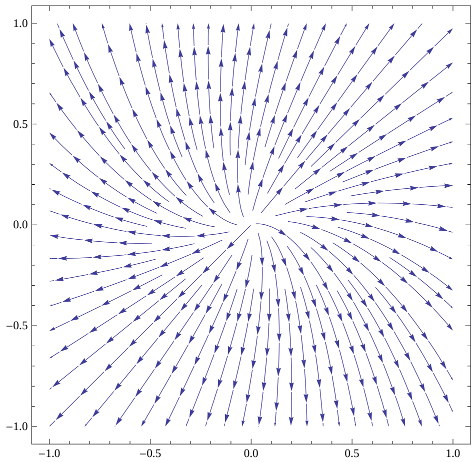


Figure 1.15. A source phase portrait of the following system with repeated eigenvalues: $x' = 4x + y, \quad y' = -x + 6y$.

λ_1	λ_2	type of phase portrait
real, positive	real, positive	source
real, negative	real, negative	sink
real, negative	real, positive	saddle
complex, positive real part	complex, positive real-part	spiral source
complex, negative real part	complex, negative real-part	spiral sink
complex, negative real part	complex, positive real-part	not possible

Table 2. Phase portrait types of linear systems $\mathbf{x}' = A\mathbf{x}$ when all eigenvalues of A has non-zero real part. Denote λ_1 and λ_2 the two eigenvalues of A. Without loss of generality, assume $\lambda_1 \leq \lambda_2$.

(i) $A = \begin{bmatrix} 7 & 1 \\ 0 & 7 \end{bmatrix}$ (ii) $A = \begin{bmatrix} 2 & -3 \\ 3 & -4 \end{bmatrix}$ (iii) $A = \begin{bmatrix} 4 & 1 \\ -1 & 6 \end{bmatrix}$.

Exercise 1.14. Let $A = \begin{bmatrix} 3 & 1 \\ a & 3 \end{bmatrix}$ where a is real. In the table below, fill in the range of value(s) of a for which the system $\mathbf{x}'(t) = A\mathbf{x}(t)$ has the type of phase portraits indicated on the left column. Fill in "$\emptyset$" (i.e. the empty set) for those phase portrait type(s) that cannot exist under this system.

Type of phase portrait	Range of value(s) of a
saddle	
sink	
source	
spiral sink	
spiral source	
center	

Exercise 1.15. Consider the second-order equation $x'' + bx' + kx = 0$. By rewriting this ODE into a planar linear system, determine all possible phase portrait types and indicate the range of values of b and k in which each portrait type occurs.

1.3. Matrix Exponentials

This section presents an elegant and *unified* approach to linear ODE systems with constant coefficients. This approach, known as *matrix exponentials*, not only unifies all three cases (distinct real, complex, repeated eigenvalues) of planar linear systems and extend to , but also gives an explicit form of general solutions to linear systems of any dimension without regarding to the eigendata, which may be difficult to find in higher dimensions.

In one dimension, the solution to the initial-value problem $x'(t) = ax(t),\ x(0) = x_0$ is given by $x(t) = x_0 e^{at}$. The exponential term e^{at} can be written as an infinite series:

$$e^{at} = \sum_{k=0}^{\infty} \frac{(at)^k}{k!} = \sum_{k=0}^{\infty} \frac{a^k t^k}{k!}.$$

Now take a square matrix A of any dimension, we are going to define the *matrix exponential* of A, denoted by e^A or $\exp(A)$, such that the solution to the initial-value problem $\mathbf{x}' = A\mathbf{x},\ \mathbf{x}(0) = \mathbf{x}_0$ can be written as $\mathbf{x}(t) = e^{tA}\mathbf{x}_0$, which is analogous to the one dimension case (with the constant matrix A replaces the role of a in one dimension).

Definition 1.21 (Matrix Exponentials)**.** Given a square matrix A of any dimension, we define e^A (or alternatively denoted by $\exp(A)$) by:

$$e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!},$$

where A^0 is defined to be I, the identity matrix with the same dimension as A.

Remark 1.22. While it makes perfect sense to add up finitely many matrices, convergence has to be justified when summing up infinitely many matrices. It will be done later after computing a few examples. □

Example 1.4. Let $A = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$, then $A^k = \begin{bmatrix} \lambda_1^k & 0 \\ 0 & \lambda_2^k \end{bmatrix}$. Therefore

$$e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!} = \sum_{k=0}^{\infty} \frac{1}{k!} \begin{bmatrix} \lambda_1^k & 0 \\ 0 & \lambda_2^k \end{bmatrix} = \begin{bmatrix} \sum_{k=0}^{\infty} \frac{\lambda_1^k}{k!} & 0 \\ 0 & \sum_{k=0}^{\infty} \frac{\lambda_2^k}{k!} \end{bmatrix} = \begin{bmatrix} e^{\lambda_1} & 0 \\ 0 & e^{\lambda_2} \end{bmatrix}.$$

□

Example 1.5. Let $A = \begin{bmatrix} 0 & -\theta \\ \theta & 0 \end{bmatrix}$. By computing the first few powers $A^2,\ A^3,\ A^4, \ldots$, one should observe that

$$A^{2k} = (-1)^k \theta^{2k} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad A^{2k+1} = (-1)^k \theta^{2k+1} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

for any $k \geq 0$. Therefore, by splitting up $\sum_{k=0}^{\infty} \frac{A^k}{k!}$ into even and odd terms, we get:

$$e^A = \sum_{k=0}^{\infty} \frac{A^{2k}}{(2k)!} + \sum_{k=0}^{\infty} \frac{A^{2k+1}}{(2k+1)!} = \sum_{k=0}^{\infty} \frac{(-1)^k \theta^{2k}}{(2k)!} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \sum_{k=0}^{\infty} \frac{(-1)^k \theta^{2k+1}}{(2k+1)!} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Recall that the Taylor's series of $\sin\theta$ and $\cos\theta$ are:

$$\sin\theta = \sum_{k=0}^{\infty} \frac{(-1)^k \theta^{2k+1}}{(2k+1)!}, \quad \cos\theta = \sum_{k=0}^{\infty} \frac{(-1)^k \theta^{2k}}{(2k)!}.$$

Therefore, $e^A = (\cos\theta)\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} + (\sin\theta)\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix} = \begin{bmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{bmatrix}$. □

Exercise 1.16. Let $A = \begin{bmatrix}\lambda & 1\\ 0 & \lambda\end{bmatrix}$, express e^A as a single matrix.

1.3.0.1. Norm of a matrix. Although e^A is defined in every example we have seen so far, you may wonder whether it is so in any square matrix. The answer is yes! In order to justify this, we first introduce the concept of *norm* of a matrix, which will also appear frequently throughout the course.

Definition 1.23 (Matrix Norm). Given a matrix A (not necessarily square), the norm of A is defined as:

$$\|A\| = \sup\{|A\mathbf{x}| : |\mathbf{x}| = 1\}.$$

Remark 1.24. From now on, we will use *double lines* $\|\cdot\|$ for the norm of a matrix defined as above, and *single lines* $|\cdot|$ for the Euclidean norm of a vector. □

Remark 1.25. In case of $\mathbb{R}^2$, think of a matrix A as a linear map, taking $\mathbf{x}$ to $\mathbf{A}\mathbf{x}$. The set $\{|\mathbf{x}| = 1\}$ represents the unit circle centered at the origin. The linear map associated to A transforms the unit circle to a ellipse centered at the origin. The geometric interpretation of $\|A\|$ is the furthest distance of points on this ellipse from the origin. □

The following are some useful properties of the norm of a matrix.

Lemma 1.26 (Properties of Matrix Norm). *Given any matrices A, B and C such that AB and $B+C$ are defined, we have*

(1) $\|A\| \geq 0$, *and* $\|A\| = 0$ *if and only if* $A = 0$.

(2) $\|cA\| = |c|\,\|A\|$ *for any* $c \in \mathbb{R}$.

(3) $\|B + C\| \leq \|B\| + \|C\|$.

(4) $|A\mathbf{x}| \leq \|A\|\,|\mathbf{x}|$ *for any vector* $\mathbf{x}$ *such that* $A\mathbf{x}$ *is defined.*

(5) $\|AB\| \leq \|A\|\,\|B\|$.

Proof. We leave (1), (2) and (3) as exercises for readers. To prove (4), take an arbitrary vector $\mathbf{x}$. If $\mathbf{x} = \mathbf{0}$, (4) is trivially true. Now assume $\mathbf{x} \neq \mathbf{0}$, then the vector $\frac{\mathbf{x}}{\|\mathbf{x}\|}$ is a unit vector. By the definition of matrix norm, we have

$$\left|A\left(\frac{\mathbf{x}}{\|\mathbf{x}\|}\right)\right| \leq \sup\{|A\mathbf{y}| : |\mathbf{y}| = 1\} =: \|A\|\,.$$

Therefore,

$$\frac{1}{\|\mathbf{x}\|}\,|A\mathbf{x}| \leq \|A\|\,,$$

which implies $|A\mathbf{x}| \leq \|A\|\,|\mathbf{x}|$.

To prove (5), we consider an arbitrary vector $\mathbf{y}$ such that $|\mathbf{y}| = 1$. Then using (4), we have:

$$|AB\mathbf{y}| \leq \|A\|\,|B\mathbf{y}| \leq \|A\|\,\|B\|\,|\mathbf{y}| = \|A\|\,\|B\|\,.$$

Note that $|\mathbf{y}| = 1$. Therefore, taking $\max$ over all unit vectors $\mathbf{y}$, we have:

$$\|AB\| := \sup\{|AB\mathbf{y}| : |\mathbf{y}| = 1\} \leq \|A\|\,\|B\|$$

as desired. □

Exercise 1.17. Prove (1), (2) and (3) of Lemma 1.26.

Using these properties, we can now prove that matrix exponential of any square matrix A must be well-defined. For any matrix P, we denote $[P]_{ij}$ the (i, j)-th entry of P.

Proposition 1.27 (Well-definedness of e^A)**.** *Let A be a square matrix. The (i, j)-th component of e^A, which is the following infinite series,*

$$\sum_{k=0}^{\infty} \frac{[A^k]_{ij}}{k!}$$

must converge. Therefore, e^A is a well-defined matrix for any square matrix A. Furthermore, we have $\left\|e^A\right\| \leq e^{\|A\|}$.

Proof. We first show that for any square matrix P, we must have $|[P]_{ij}| \leq \|P\|$. The argument goes as follows: the j-th column of the matrix P is the components of the vector $P(\mathbf{e}_j)$ where $\mathbf{e}_j$ is the j-th standard basis vector, i.e. $P(\mathbf{e}_j) = \sum_i [P]_{ij}\mathbf{e}_i$. By the definition of matrix norm, we have

$$\begin{aligned}
|P(\mathbf{e}_j)| &\leq \|P\| && \text{(since } \mathbf{e}_j \text{ is unit)} \\
\left|\sum_i [P]_{ij}\mathbf{e}_i\right| &\leq \|P\| \\
\sqrt{\sum_i |[P]_{ij}|^2} &\leq \|P\|
\end{aligned}$$

Since for each i, we have $|[P]_{ij}| \leq \sqrt{|[P]_{1j}|^2 + \ldots + |[P]_{dj}|^2} = \sqrt{\sum_i |[P]_{ij}|^2}$, and so $|[P]_{ij}| \leq \|P\|$.

Now we use the absolute convergence test to show the following infinite series $\sum_{k=0}^{\infty} \frac{[A^k]_{ij}}{k!}$ converges. Consider:

$$\left|\frac{[A^k]_{ij}}{k!}\right| \leq \frac{\|A^k\|}{k!} \leq \frac{\|A\|^k}{k!}.$$

The last inequality follows from (5) of Lemma 1.26. Since $\sum_{k=0}^{\infty} \frac{\|A\|^k}{k!}$ converges to $e^{\|A\|}$, by comparison test and absolute convergence test, $\sum_{k=0}^{\infty} \frac{[A^k]_{ij}}{k!}$ converges absolutely. It shows e^A is defined.

To show the last part of the proposition, we use (3) and (5) of Lemma 1.26:

$$\left\|e^A\right\| = \left\|\sum_{k=0}^{\infty} \frac{A^k}{k!}\right\| \leq \sum_{k=0}^{\infty} \left\|\frac{A^k}{k!}\right\| \leq \sum_{k=0}^{\infty} \frac{\|A\|^k}{k!} = e^{\|A\|}$$

as desired. □

Exercise 1.18. Given a sequence of $d \times d$ matrices $\{A_n\}_{n=1}^{\infty}$, show that A_n converges to a $d \times d$ matrix A_∞ if and only if $\lim_{n\to\infty} \|A_n - A_\infty\| = 0$.

1.3.0.2. Solution to linear systems with initial conditions. Having established that e^A is well-defined, we are about to prove that the initial-value problem $\mathbf{x}' = A\mathbf{x},\ \mathbf{x}(0) = \mathbf{x}_0$ has a solution $\mathbf{x}(t) = e^{tA}\mathbf{x}_0$, analogous to the one dimension case.

Theorem 1.28. *Let A be a $d \times d$ matrix and $\mathbf{x}_0$ is any given vector in $\mathbb{R}^d$. Then the initial-value problem*

$$\mathbf{x}' = A\mathbf{x}, \quad \mathbf{x}(0) = \mathbf{x}_0$$

has a solution $\mathbf{x}(t) = e^{tA}\mathbf{x}_0$.

Proof. Let $\mathbf{x}(t) = e^{tA}\mathbf{x}_0$, then

$$\mathbf{x}'(t) = \frac{d}{dt}\sum_{n=0}^{\infty}\frac{t^n A^n}{n!}\mathbf{x}_0$$

We want to differentiate the infinite series term-by-term. In order to do so, recall that we have to justify that as $N \to \infty$:

(1) $\displaystyle\frac{d}{dt}\sum_{n=0}^{N}\frac{t^n A^n}{n!}\mathbf{x}_0$ converges uniformly on any open interval (a, b), and

(2) $\displaystyle\sum_{n=0}^{N}\frac{t^n A^n}{n!}\mathbf{x}_0$ converges pointwise on (a, b).

Note that (2) follows directly from Proposition 1.27. For (1), we compute that:

$$\frac{d}{dt}\sum_{n=0}^{N}\frac{t^n A^n}{n!}\mathbf{x}_0 = \sum_{n=1}^{N}\frac{t^{n-1}A^n}{(n-1)!}\mathbf{x}_0.$$

For any $t \in (a, b)$, we have for each term that:

$$\left\|\frac{t^{n-1}A^n}{(n-1)!}\mathbf{x}_0\right\| \leq \frac{|t|^{n-1}}{(n-1)!}\|A\|^n|\mathbf{x}_0| \leq \frac{(|a|+|b|)^{n-1}\|A\|^n|\mathbf{x}_0|}{(n-1)!}.$$

Since $\displaystyle\sum_{n=1}^{\infty}\frac{(|a|+|b|)^{n-1}\|A\|^n|\mathbf{x}_0|}{(n-1)!}$ converges as a series of real numbers – can be easily checked by ratio test, Weiestrass' M-test shows the series of functions $\displaystyle\sum_{n=1}^{\infty}\frac{t^{n-1}A^n}{(n-1)!}\mathbf{x}_0$ converges uniformly on (a, b). It proves (1).

By term-by-term differentiation, we get

$$\begin{aligned}\mathbf{x}'(t) &= \sum_{n=1}^{\infty}\frac{t^{n-1}A^n}{(n-1)!}\mathbf{x}_0 = \sum_{n=0}^{\infty}\frac{t^n A^{n+1}}{n!}\mathbf{x}_0 \\ &= A\left(\sum_{n=0}^{\infty}\frac{t^n A^n}{n!}\right)\mathbf{x}_0 \\ &= A\mathbf{x}(t).\end{aligned}$$

Clearly $\mathbf{x}(0) = e^0\mathbf{x}_0 = I\mathbf{x}_0 = \mathbf{x}_0$. Therefore, $e^{tA}\mathbf{x}_0$ is a solution to the given initial-value problem. □

We say in the statement of Theorem 1.28 that the initial-value problem has *a* solution given by $e^{tA}\mathbf{x}_0$ because we haven't shown the solution is unique. Now we are about to prove the uniqueness of this solution. In fact, we have a stronger result:

Theorem 1.29 (Continuous Dependence Inequality for Linear Systems)**.** *Let A be $d \times d$ matrix, and $\mathbf{x}_0$, $\mathbf{y}_0$ be two given vectors in $\mathbb{R}^d$. Suppose $\mathbf{x}(t)$ and $\mathbf{y}(t)$ are solutions to the following initial-value problems respectively (with the same matrix but different initial conditions):*

$$\begin{aligned} \mathbf{x}' &= A\mathbf{x}, \quad \mathbf{x}(0) = \mathbf{x}_0 \\ \mathbf{y}' &= A\mathbf{y}, \quad \mathbf{y}(0) = \mathbf{y}_0, \end{aligned}$$

then we have

$$|\mathbf{x}(t) - \mathbf{y}(t)| \leq |\mathbf{x}_0 - \mathbf{y}_0| \, e^{\|A\||t|} \quad \text{for any } t \in \mathbb{R}. \tag{1.12}$$

Remark 1.30. We call Theorem 1.29 the *Continuity Dependence on Initial Data* because (1.12) shows that in short-time, a slight change of the initial data will only change the solution a little. We will see later that this kind of continuity dependence also holds for nonlinear systems assuming the vector field $\mathbf{F}(\mathbf{x})$ is sufficiently regular. □

As an easy corollary to Theorem 1.29, if $\mathbf{x}(t)$ and $\mathbf{y}(t)$ are solutions with the same initial data, i.e. $\mathbf{x}_0 = \mathbf{y}_0$, then it is necessary that $\mathbf{x}(t) \equiv \mathbf{y}(t)$:

Corollary 1.31 (Uniqueness of Solutions to Linear Systems)**.** *The solution to the initial-value problem*

$$\mathbf{x}' = A\mathbf{x}, \quad \mathbf{x}(0) = \mathbf{x}_0$$

is unique. Hence, $\mathbf{x}(t) = e^{tA}\mathbf{x}_0$ is the only solution.

Proof of Theorem 1.29. The inequality (1.12) trivially holds when $t = 0$. We first assume $t > 0$. We consider:

$$\begin{aligned} |\mathbf{x}(t) - \mathbf{y}(t)|^2 &= (\mathbf{x}(t) - \mathbf{y}(t)) \cdot (\mathbf{x}(t) - \mathbf{y}(t)) \\ \frac{d}{dt} |\mathbf{x}(t) - \mathbf{y}(t)|^2 &= 2(\mathbf{x}(t) - \mathbf{y}(t)) \cdot \frac{d}{dt}(\mathbf{x}(t) - \mathbf{y}(t)) && \text{(product rule)} \\ &= 2(\mathbf{x}(t) - \mathbf{y}(t)) \cdot A(\mathbf{x}(t) - \mathbf{y}(t)) && \text{(since } \mathbf{x}' = A\mathbf{x} \text{ and } \mathbf{y}' = A\mathbf{y}) \\ &\leq 2\,|\mathbf{x}(t) - \mathbf{y}(t)| \cdot |A(\mathbf{x}(t) - \mathbf{y}(t))| && \text{(Cauchy-Schwarz's Inequality)} \\ &\leq 2\,|\mathbf{x}(t) - \mathbf{y}(t)| \cdot \|A\| \, |\mathbf{x}(t) - \mathbf{y}(t)| && \text{(by (4) of Lemma 1.26)} \\ &\leq 2\,\|A\| \cdot |\mathbf{x}(t) - \mathbf{y}(t)|^2. \end{aligned}$$

Therefore, by the "integrating-factor" trick:

$$\begin{aligned} \frac{d}{dt}\left(e^{-2\|A\|t} \cdot |\mathbf{x}(t) - \mathbf{y}(t)|^2\right) &= e^{-2\|A\|t}\left(\frac{d}{dt}|\mathbf{x}(t) - \mathbf{y}(t)|^2 - 2\,\|A\| \cdot |\mathbf{x}(t) - \mathbf{y}(t)|^2\right) \\ &\leq 0 \end{aligned}$$

Hence, for any $t > 0$, we have:

$$e^{-2\|A\|t} \cdot |\mathbf{x}(t) - \mathbf{y}(t)|^2 \leq e^{-2\|A\| \cdot 0} \cdot |\mathbf{x}(0) - \mathbf{y}(0)|^2 = |\mathbf{x}_0 - \mathbf{y}_0|^2 .$$

By rearrangement, one can prove (1.12) for $t > 0$.

The case $t < 0$ is almost the same, except for the Cauchy-Schwarz's Inequality one should consider

$$2(\mathbf{x}(t) - \mathbf{y}(t)) \cdot A(\mathbf{x}(t) - \mathbf{y}(t)) \geq -2\,|\mathbf{x}(t) - \mathbf{y}(t)| \cdot |A(\mathbf{x}(t) - \mathbf{y}(t))|$$

instead. The detail of this case is left as an exercise to readers. □

Exercise 1.19. Complete the proof of Theorem 1.29 for the case $t < 0$.

Proposition 1.32. *Given two real numbers t and s, and a square matrix A, we have*

$$e^{tA}e^{sA} = e^{(t+s)A}. \tag{1.13}$$

Corollary 1.33. *For any square matrix A, $e^{-A} = \left(e^A\right)^{-1}$. Therefore, e^A must be invertible.*

Proof (not rigorous).

$$\begin{aligned}
e^{tA}e^{sA} &= \left(\sum_{k=0}^{\infty}\frac{t^kA^k}{k!}\right)\left(\sum_{l=0}^{\infty}\frac{s^lA^l}{l!}\right)\\
&= \sum_{k,\,l=0}^{\infty}\frac{t^ks^l}{k!\,l!}A^{k+l}\\
&= \sum_{j=0}^{\infty}\sum_{k+l=j}\frac{t^ks^l}{k!\,l!}A^{k+l} && \text{(sum up along diagonals of the (k,l)-array)}\\
&= \sum_{j=0}^{\infty}\sum_{k=0}^{j}\frac{t^ks^{j-k}}{k!\,(j-k)!}A^j && \text{(re-label indices: $l = j-k$)}\\
&= \sum_{j=0}^{\infty}\sum_{k=0}^{j}\frac{j!\,t^ks^{j-k}}{k!\,(j-k)!}\frac{A^j}{j!}.
\end{aligned}$$

The binomial theorem asserts that $(t+s)^j = \sum_{k=0}^{j}\frac{j!}{k!\,(j-k)!}t^ks^{j-k}$, so from above, we deduce:

$$e^{tA}e^{sA} = \sum_{j=0}^{\infty}\frac{(t+s)^jA^j}{j!} = \sum_{j=0}^{\infty}\frac{((t+s)A)^j}{j!} = e^{(t+s)A}$$

as desired. The corollary follows directly from $e^0 = I$. □

We say the proof is not rigorous because it involves rearrangement of a double-indexed sum (over k, l) into a diagonal sum. It has to be justified rigorously by estimating the remainder terms and proving that they go to zero (see my lecture notes on MATH 4023, in which we proved $e^ze^w = e^{z+w}$ for any complex numbers z and w). As a course on theory of ODE, we provide an alternative proof using the uniqueness theorem (Corollary 1.31) that we have just proved.

Proof of Proposition 1.32. Consider two curves

$$\mathbf{X}_s(t) := e^{tA}e^{sA}\mathbf{x}_0 \quad \text{and} \quad \mathbf{Y}_s(t) := e^{(t+s)A}\mathbf{x}_0,$$

where $\mathbf{x}_0$ is any vector in $\mathbb{R}^d$. Here we regard t as the variable and s as fixed. We claim that for any fixed $s \in \mathbb{R}$, both $\mathbf{X}_s(t)$ and $\mathbf{Y}_s(t)$ solve the ODE system $\mathbf{x}'(t) = A\mathbf{x}(t)$ with the same initial data. It would then follow by Corollary (1.31) that $\mathbf{X}_s(t) = \mathbf{Y}_s(t)$ for any $t \in \mathbb{R}$.

$$\begin{aligned}
\frac{d}{dt}\mathbf{X}_s(t) &= \frac{d}{dt}e^{tA}e^{sA}\mathbf{x}_0 \\
&= \left(\frac{d}{dt}e^{tA}\right)e^{sA}\mathbf{x}_0 \\
&= Ae^{tA}e^{sA}\mathbf{x}_0 && \text{(from Theorem 1.28)} \\
&= A\mathbf{X}_s(t) \\
\frac{d}{dt}\mathbf{Y}_s(t) &= \frac{d}{dt}e^{(t+s)A}\mathbf{x}_0 \\
&= \frac{d}{d(t+s)}e^{(t+s)A}\cdot\frac{d(t+s)}{dt}\cdot\mathbf{x}_0 \\
&= Ae^{(t+s)}\mathbf{x}_0 \\
&= A\mathbf{Y}_s(t)
\end{aligned}$$

Hence, both $\mathbf{X}_s(t)$ and $\mathbf{Y}_s(t)$ satisfy the ODE system $\mathbf{x}'(t) = A\mathbf{x}(t)$ for any fixed $s \in \mathbb{R}$. For their initial values at $t = 0$, we can check that they are also the same:

$$\begin{aligned}
\mathbf{X}_s(0) &= Ie^{sA}\mathbf{x}_0 = e^{sA}\mathbf{x}_0 \\
\mathbf{Y}_s(0) &= e^{(0+s)A}\mathbf{x}_0 = e^{sA}\mathbf{x}_0.
\end{aligned}$$

By Corollary 1.31, we have $e^{tA}e^{sA}\mathbf{x}_0 = e^{(t+s)A}\mathbf{x}_0$ for any $t, s \in \mathbb{R}$ and any $\mathbf{x}_0 \in \mathbb{R}^d$. This show $e^{tA}e^{sA} = e^{(t+s)A}$ for any $t, s \in \mathbb{R}$. □

Exercise 1.20. Let A, B and P be square matrices of the same dimension, and further assume that P is invertible. Show that:

(1) $e^{PAP^{-1}} = Pe^AP^{-1}$

(2) $Ae^A = e^AA$

(3) If $AB = BA$, then $e^Ae^B = e^Be^A = e^{A+B}$. Using this, express $\exp\begin{bmatrix}\alpha & -\beta \\ \beta & \alpha\end{bmatrix}$ as a single matrix.

1.3.0.3. Flow of a linear system. Theorem 1.28 and Corollary 1.31 assert that $\mathbf{x}(t) = e^{tA}\mathbf{x}_0$ is the only solution to the initial-value problem $\mathbf{x}' = A\mathbf{x}, \quad \mathbf{x}(0) = \mathbf{x}_0$. Therefore, starting at any point $\mathbf{x}_0$ on the phase portrait, there is one and only one trajectory passing through it, and $\mathbf{x}(t) = e^{tA}\mathbf{x}_0$ can be thought as **flowing along the trajectory through $\mathbf{x}_0$ for t unit time**. This concept motivates the introduction of the flow map of a linear system:

Definition 1.34 (Flow of a Linear System)**.** Given a system of ODEs $\mathbf{x}' = A\mathbf{x}$ on $\mathbb{R}^d$, the flow of the system, denoted by $\Phi(\cdot, t) : \mathbb{R}^d \times (-\infty, \infty) \to \mathbb{R}^d$ such that for any $\mathbf{x}_0 \in \mathbb{R}^d$, $\Phi(\mathbf{x}_0, t)$ is the point on $\mathbb{R}^d$ reached by flowing along the trajectory from $\mathbf{x}_0$ for t unit time. Equivalently, the flow $\Phi(\cdot, t)$ is sometimes denoted as $\Phi_t(\cdot) : \mathbb{R}^d \to \mathbb{R}^d$.

Remark 1.35. The flow can be defined similarly for nonlinear systems, but one should first justify uniqueness and the time interval has to be restricted so that the solution is defined. We will do these in the next chapter. □

Remark 1.36. For linear systems $\mathbf{x}' = A\mathbf{x}$, the flow is explicitly given by

$$\Phi_t(\mathbf{x}_0) = e^{tA}\mathbf{x}_0.$$

However, an explicit expression of the flow for most nonlinear systems are usually extremely difficult to find. □

The flows, for both linear and nonlinear systems, will be used frequently throughout the course. One advantage of using flows $\Phi_t(\mathbf{x}_0)$ instead of using $\mathbf{x}(t)$ to represent a solution to an initial-value problem is that one can read off the initial condition immediately from the $\mathbf{x}_0$-slot. If one uses $\mathbf{x}(t)$, one has to declare aside that its initial condition is given by $\mathbf{x}(0) = \mathbf{x}_0$, which can be cumbersome and confusing if there are many initial conditions being considered simultaneously.

Another feature of using flows concerns about taking compositions of two flows. Since for any given $t,\ s \in \mathbb{R}$ and $\mathbf{x}_0 \in \mathbb{R}^d$, we have:

$$\begin{aligned} \Phi_s\left(\Phi_t(\mathbf{x}_0)\right) &= \Phi_s\left(e^{tA}\mathbf{x}_0\right) \\ &= e^{sA}e^{tA}\mathbf{x}_0 \\ &= e^{(s+t)A}\mathbf{x}_0 && \text{(from (1.13))} \\ &= \Phi_{s+t}(\mathbf{x}_0). \end{aligned}$$

Therefore, $\Phi_s \circ \Phi_t = \Phi_{s+t}$, meaning that flowing along the trajectory starting from $\mathbf{x}_0$ for t unit time, followed by flowing along for s unit time, will reach the same point as flowing along the trajectory starting from $\mathbf{x}_0$ for $s+t$ unit time. It is consistent with our intuition. Evidently, this fact can easily be seen using flows but it is not quite clear why it is true if one labels the trajectories by $\mathbf{x}(t)$ and $\mathbf{y}(t)$.

By Theorem 1.29, the flow $\Phi_t(\cdot)$ can be shown to be continuous:

Proposition 1.37 (Continuity of Flow for Linear Systems)**.** *Let $\Phi_t(\cdot) : \mathbb{R}^d \to \mathbb{R}^d$ be the flow of a linear system $\mathbf{x}' = A\mathbf{x}$. Then $\Phi_t(\cdot)$ is continuous function from $\mathbb{R}^d$ to $\mathbb{R}^d$.*

Proof. Using the flow, (1.12) can be rewritten as:

$$|\Phi_t(\mathbf{x}_0) - \Phi_t(\mathbf{y}_0)| \leq |\mathbf{x}_0 - \mathbf{y}_0|\, e^{\|A\||t|} \tag{1.14}$$

Fix $t \in \mathbb{R}$, regarding $\mathbf{y}_0$ is fixed and $\mathbf{x}_0$ is a variable. When $\mathbf{x}_0 \to \mathbf{y}_0$, meaning that $|\mathbf{x}_0 - \mathbf{y}_0| \to 0$, the right-hand side of (1.14) tends to 0. The squeezing principle shows we have $|\Phi_t(\mathbf{x}_0) - \Phi_t(\mathbf{y}_0)| \to 0$, or equivalently, $\Phi_t(\mathbf{x}_0) \to \Phi(\mathbf{y}_0)$ as $\mathbf{x}_0 \to \mathbf{y}_0$. Therefore, $\Phi_t(\cdot)$ is continuous. □

Remark 1.38. The flow must be continuous, even differentiable, in the t-slot, since $\Phi_t(\mathbf{x}_0)$ is defined to be the solution $\mathbf{x}(t)$ that solves the system $\mathbf{x}' = A\mathbf{x}$ with initial condition $\mathbf{x}(0) = \mathbf{x}_0$. Therefore,

$$\frac{d}{dt}\Phi_t(\mathbf{x}_0) = \mathbf{x}'(t) = A\mathbf{x}(t) = A\Phi_t(\mathbf{x}_0).$$

□

Remark 1.39. Since for each fixed $t \in \mathbb{R}$ the flow $\Phi_t : \mathbb{R}^d \to \mathbb{R}^d$ is invertible with inverse Φ_{-t} being also continuous, in topological terminology we call the flow Φ_t a *homeomorphism* of $\mathbb{R}^d$ to itself. □

The expression $\mathbf{x}(t) = e^{tA}\mathbf{x}_0$ as the solution to the initial-value problem $\mathbf{x}' = A\mathbf{x}, \mathbf{x}(0) = \mathbf{x}_0$, although works for any square matrix A, is not as *explicit* as it may seem. The matrix exponential e^{tA} is defined by an infinite series of matrices. Each term of the series involves the power A^k, which cannot be computed easily. In fact, we need to find the eigen-data and a canonical decomposition in order to find A^k. In higher dimensions, finding solutions to linear systems with matrix A will require a good canonical decomposition of the matrix. We will not go into that because it will deviate from the main theme of this course. Interested readers may consult Chapters 5 and 6 of Hirsch–Smale–Devaney's book for the Linear Algebra treatment of ODE systems in higher dimensions.

Chapter 2

Existence and Uniqueness

This chapter is about the existence and uniqueness theorems of ODEs. The purpose of establishing the existence and uniqueness theorems is that nonlinear ODEs are extremely difficult to solve, but in many applications in other areas such as differential geometry, it is not necessary to know the explicit solution yet we need to make sure there *is* a solution.

If you have taken some elementary ODE courses, you probably have learned various methods to solve *some types* of ODEs. These methods might include *separation of variables*, *integration factor*, *characteristics equations*, *variation of parameters*, *Laplace's Transforms*, etc. However, you probably have realized that these toolkits are far from being complete for solving all ODEs. One can easily write down a nonlinear ODE such as:

$$y'(t) = \log \sin(y(t)^2 + t^2), \quad y(0) = 1.$$

It is almost impossible to find the explicit solution even using computer softwares.

For nonlinear ODE systems, solving for explicit solutions is much harder even for a simple nonlinear ODE system like this:

$$\begin{aligned} x' &= x^2 - y \\ y' &= x + y^2 + 1 \end{aligned}$$

Another goal of working through the existence theorem is to give you a taste of how *differential equations* interact with *analysis*. While most undergraduate ODE/PDE courses focus on *solving* differential equations, the studies of PDE at graduate or research level require a strong background in analysis, especially *functional analysis*.

2.1. Contraction Maps and Iterations

Let's begin by a discussion on the following problem.

Problem 2.1. Suppose you put a map of the HKUST campus on a flat table inside a building at HKUST. Assume the map has size smaller than the HKUST campus (a fairly sensible assumption). Prove that there must be a point on the map which is directly above the actual point it represents, no matter how you orient the map.

You may think this problem is completely off-topic, but you will realize later the key idea of the existence theorems actually stem from this problem. To formulate Problem 2.1 in a mathematical way, we denote U the set representing the whole HKUST University including its boundary. Define $f : U \to U$ to be the function that takes an actual point $\mathbf{x}$ at HKUST, to the point $f(\mathbf{x})$ on the map that represents this point $\mathbf{x}$.

$f(\mathbf{x})$ is in U since the map is put inside HKUST. To solve Problem 2.1, we need to find a point $\mathbf{y}_0 \in U$ such that

$$f(\mathbf{y}_0) = \mathbf{y}_0$$

so that the point $f(\mathbf{y}_0)$ on the map is equal to the point $\mathbf{y}_0$ which it represents.

We will show that such a $\mathbf{y}_0$ always exists. The key ingredient is that such a function f must be a **contraction**. Precisely, there exists a positive $\alpha < 1$, such that given any $\mathbf{x}, \mathbf{y} \in U$, we must have

$$|f(\mathbf{x}) - f(\mathbf{y})| \leq \alpha \, |\mathbf{x} - \mathbf{y}| \, . \tag{2.1}$$

Obviously, this assumption makes sense because the distance between any two points on the map is significantly smaller than that between the two actual points at HKUST they represent. The argument for the existence of such a $\mathbf{y}_0$, which we call it a fixed-point of f, is through an iteration argument.

We take any point $\mathbf{x}_1 \in U$, and consider $\mathbf{x}_2 := f(\mathbf{x}_1)$. If $\mathbf{x}_2 = \mathbf{x_1}$, then $\mathbf{x}_1$ is the desired fixed-point, and we are done. Otherwise, look for $\mathbf{x}_3 := f(\mathbf{x}_2)$, $\mathbf{x}_4 := f(\mathbf{x}_3)$ and continue indefinitely, i.e. $\mathbf{x}_{n+1} := f(\mathbf{x}_n)$ for any integer $n > 0$. Intuitively, the distance between two adjacent points should be decreasing as the process continues. It can be verified by (2.1) that for any integer $n > 0$,

$$|\mathbf{x}_{n+1} - \mathbf{x}_n| = |f(\mathbf{x}_n) - f(\mathbf{x}_{n-1})| \leq \alpha \, |\mathbf{x}_n - \mathbf{x}_{n-1}| \, .$$

Inductively, one can show:

$$\begin{aligned} |\mathbf{x}_{n+1} - \mathbf{x}_n| &\leq \alpha \, |\mathbf{x}_n - \mathbf{x}_{n-1}| \\ &\leq \alpha^2 \, |\mathbf{x}_{n-1} - \mathbf{x}_{n-2}| \\ &\quad \vdots \\ &\leq \alpha^{n-1} \, |\mathbf{x}_2 - \mathbf{x}_1| \, . \end{aligned}$$

Next, we consider the infinite series $\sum_{n=1}^{\infty} |\mathbf{x}_{n+1} - \mathbf{x}_n|$, which is bounded above by the geometric series $\sum_{n=1}^{\infty} \alpha^{n-1} |\mathbf{x}_2 - \mathbf{x}_1|$. Note that $|\mathbf{x}_2 - \mathbf{x}_1|$ is regarded as a constant. As the common ratio α of the geometric series is strictly less than 1, the series $\sum_{n=1}^{\infty} \alpha^{n-1} |\mathbf{x}_2 - \mathbf{x}_1|$ converges, so by comparison test, the series

$$\sum_{n=1}^{\infty} |\mathbf{x}_{n+1} - \mathbf{x}_n|$$

converges too.

Finally, the absolute convergence test shows the telescoping series $\sum_{n=1}^{\infty} (\mathbf{x}_{n+1} - \mathbf{x}_n)$ converges. Then for any integer $N > 1$, we have:

$$\begin{aligned} \mathbf{x}_N &= \mathbf{x}_1 + (\mathbf{x}_2 - \mathbf{x}_1) + (\mathbf{x}_3 - \mathbf{x}_2) + \ldots + (\mathbf{x}_N - \mathbf{x}_{N-1}) \\ &= \underbrace{\mathbf{x}_1}_{\text{constant vector}} + \underbrace{\sum_{n=1}^{N-1} (\mathbf{x}_{n+1} - \mathbf{x}_n)}_{\text{converges as } N \to \infty}, \end{aligned}$$

Letting $N \to \infty$, this shows $\mathbf{x}_N$, as a sequence, must converge too. If we denote $\mathbf{y}_0 = \lim_{n\to\infty} \mathbf{x}_n$, then letting $n \to \infty$ on both sides of $\mathbf{x}_{n+1} = f(\mathbf{x}_n)$, we get $\mathbf{y}_0 = f(\mathbf{y}_0)$, where we used the fact that f is continuous as guaranteed by the contraction inequality (2.1). Therefore, such a fixed point $\mathbf{y}_0$ always exists and it is exactly the point on the map which is directly above the actual point it represents.

The sequence $\mathbf{x}_n$ in this map problem is defined via the recurrence relation $\mathbf{x}_{n+1} = f(\mathbf{x}_n)$ with a 'seed' $\mathbf{x}_1$. We call this type of sequence an **iteration sequence**. As you will see, the existence of solutions for ODE systems is proved by rewriting the ODE problem to a problem of finding a fixed-point. The iterative sequence involved, commonly called the *Picard's iteration sequence*, will be proved to converge to a solution to the ODE system. Proving existence of a solution is analogous to locating the fixed-point of the map problem.

2.1.0.1. More iteration examples. You may have tried the following "experiment" with your scientific calculator: start with any value x_0 in $[0, 1]$, say 0.1; then press the $\boxed{\text{cos}}$ button repeatedly. After pressing it for around 20 times you will see the displayed value will soon "stabilize" and converge to a particular value (around 0.739085133). This value is in fact an approximate solution to the equation $x = \cos x$. To see why is that, let's formulate this "pressing-the-button" experiment in a mathematical way. We let:

$$\begin{aligned} x_0 &= 0.1 \\ x_n &= \cos(x_{n-1}) \quad \text{for any } n \geq 1. \end{aligned}$$

Then the sequence $x_0, x_1, x_2, \ldots$ will be equal to:

$$0.1,\ \cos(0.1),\ \cos(\cos(0.1)),\ \cos(\cos(\cos(0.1))), \ldots$$

which is exactly the sequence of numbers you get by pressing the $\boxed{\text{cos}}$ button repeatedly. If we assume (and we will justify) that $\lim_{n\to\infty} x_n$ exists and converges to a limit L, then taking limit on both sides of $x_n = \cos(x_{n-1})$ will yield:

$$\begin{aligned} \lim_{n\to\infty} x_n &= \lim_{n\to\infty} \cos(x_{n-1}) \\ L &= \cos\left(\lim_{n\to\infty} x_{n-1}\right) && \text{(since cos is continuous)} \\ L &= \cos(L). \end{aligned}$$

It shows the limit L is a solution to the equation $x = \cos x$. That's why when you press the $\boxed{\text{cos}}$ button for many many times, you will get a value which is very close to this L.

In order to complete the above argument, we need to prove the sequence x_n converges. Let's complete the detail here. Define

$$f(x) = \cos x : [0, 1] \to [0, 1].$$

It makes sense to define the codomain of f to be $[0, 1]$ because $\cos x$ is decreasing on $[0, 1]$ and so $f(x) \in [\cos 1, \cos 0] = [\cos 1, 1] \subset [0, 1]$. Now x_n is an iteration sequence by the function f, i.e. $x_n = f(x_{n-1})$ for any $n \geq 1$.

By the mean-value theorem, one can prove a contraction inequality similar to (2.1): for any $x, y \in [0, 1]$

$$\begin{aligned} |f(x) - f(y)| &= |f'(\xi)||x - y| && \text{for some } \xi \text{ between } x \text{ and } y \\ &= |\sin \xi||x - y| \\ &\leq (\sin 1)|x - y| && \text{since } x,\ y,\ \xi \in [0, 1] \end{aligned}$$

Denote $\alpha := \sin 1$ for simplicity, we have the following contraction inequality:

$$|f(x) - f(y)| \leq \alpha|x - y| \quad \text{for any } x, y \in [0, 1]. \tag{2.2}$$

Note that $\alpha < 1$. Therefore, one can mimic the steps after (2.1) for the map problem (now the real number sequence $\{x_n\}$ replaces the role of the vector sequence $\{\mathbf{x}_n\}$) to show that x_n converges to some number $L \in [0, 1]$, which is a fixed-point of f and a root to the equation $x = f(x)$.

Exercise 2.1. Complete the detail of showing the convergence of x_n in the above iteration using the contraction inequality (2.2).

Exercise 2.2. This is an exercise of using iterations to show the existence of a root of another trigonometric equation. Define $g(x) = \cos x - \frac{1}{3}\cos^3 x$ where $x \in [0, \frac{\pi}{3}]$. Consider the following iteration sequence:

$$x_0 = 0, \qquad x_n = g(x_{n-1}) \quad \text{for } n \geq 1.$$

Show that $g(x)$ maps $[0, \frac{\pi}{3}]$ into $[0, \frac{\pi}{3}]$ and satisfies a contraction inequality: there exists $\alpha < 1$ such that $|g(x) - g(y)| \leq \alpha|x - y| \quad$ for any $x,\, y \in [0, \frac{\pi}{3}]$. Hence, prove that the iteration sequence x_n converges to a limit L which is a root of the equation $x = \cos x - \frac{1}{3}\cos^3 x$.

2.2. Picard's Iteration

The proof of the existence theorem of ODEs is also based on an iteration argument. We will first reformulate an ODE problem as an iteration problem like the examples in the previous section. Then, showing the existence of solutions will be equivalent to proving an iteration sequence converges. That is why *analysis* comes into play.

As in Section 1.1, we can represent a system of ODEs (which may not be linear) in vector form:

$$\mathbf{x}' = \mathbf{F}(\mathbf{x}).$$

Here $\mathbf{F}(\mathbf{x})$ is regarded as a vector field on $\mathbb{R}^d$, and the vector equation $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ asserts that the solution curve $\mathbf{x}(t)$ travels with velocity (or tangent) vector $\mathbf{x}'$ equals to the vector field direction for all time t.

The vector fields $\mathbf{F}(\mathbf{x})$ we have considered in Chapter 1 are in the form $A\mathbf{x}$ where A is a constant matrix. From now on, we not only consider these linear systems, but also nonlinear ones which are very challenging to solve. In this chapter in particular, we also allow the vector field $\mathbf{F}$ to be time-dependent, i.e. changing over time. Mathematically speaking, we allow $\mathbf{F}$ to depend on both time t and the space $\mathbf{x}$. Therefore, a general form of an ODE system in this chapter is in the form of:

$$\mathbf{x}' = \mathbf{F}(\mathbf{x}, t)$$

where $\mathbf{F} : \Omega \times I \to \mathbb{R}^d$ is a time-dependent vector field defined on some open region Ω in $\mathbb{R}^d$ and some time interval I.

Definition 2.2 (Autonomous and Non-autonomous Systems)**.** An *autonomous* ODE system is one that the vector field is *time independent*, i.e. $\mathbf{x}' = \mathbf{F}(\mathbf{x})$, whereas a *non-autonomous* ODE system is one that the vector field is *time dependent*, i.e. $\mathbf{x}' = \mathbf{F}(\mathbf{x}, t)$.

2.2.0.1. Integral equation of an initial-value problem. Given an ODE system $\mathbf{x}' = \mathbf{F}(\mathbf{x}, t)$ with an initial condition $\mathbf{x}(0) = \mathbf{x}_0$, we will first rewrite the initial-value problem (from now on will be called IVP) as an equivalent *integral equation*. Then, we will formulate the iteration procedure for the integral equation.

Proposition 2.3 (Integral Equations for an IVP)**.** *Given any fixed vector $\mathbf{x}_0 \in \mathbb{R}^d$ and a continuous vector field $\mathbf{F}$, the trajectory $\mathbf{x}(t)$ is a solution to the IVP*

$$\mathbf{x}' = \mathbf{F}(\mathbf{x}, t), \quad \mathbf{x}(0) = \mathbf{x}_0 \tag{2.3}$$

if and only if $\mathbf{x}(t)$ is a continuous solution to the integral equation

$$\mathbf{x}(t) = \mathbf{x}_0 + \int_0^t \mathbf{F}(\mathbf{x}(s),\, s)\, ds. \tag{2.4}$$

Remark 2.4. Inside the integral we use $\mathbf{x}(s)$ instead of $\mathbf{x}(t)$ so as to avoid confusion with the upper limit t of the integral. The letter s is dummy and can be replaced by your most favorite variable (other than t). □

Proof of Proposition 2.3. $(\Rightarrow)$-part: Given $\mathbf{x}(t)$ solves the IVP (2.3), then:

$$\begin{aligned}
\text{RHS of (2.4)} &= \mathbf{x}_0 + \int_0^t \mathbf{F}(\mathbf{x}(s),\, s)\,ds \\
&= \mathbf{x}_0 + \int_0^t \mathbf{x}'(s)\,ds \\
&= \mathbf{x}_0 + [\mathbf{x}(s)]_{s=0}^{s=t} && \text{(Fundamental Theorem of Calculus)} \\
&= \mathbf{x}_0 + \mathbf{x}(t) - \mathbf{x}(0) \\
&= \mathbf{x}(t) && (\text{since } \mathbf{x}(0) = \mathbf{x}_0) \\
&= \text{LHS of (2.4)}.
\end{aligned}$$

$(\Leftarrow)$-part: Given $\mathbf{x}(t)$ solves the integral equation (2.4), we have:

$$\begin{aligned}
\mathbf{x}'(t) &= \frac{d}{dt}\left(\mathbf{x}_0 + \int_0^t \mathbf{F}(\mathbf{x}(s),\, s)\,ds\right) \\
&= \frac{d}{dt}\int_0^t \mathbf{F}(\mathbf{x}(s),\, s)\,ds \\
&= \mathbf{F}(\mathbf{x}(t),\, t) && \text{(Fundamental Theorem of Calculus)} \\
\mathbf{x}(0) &= \mathbf{x}_0 + \int_0^0 \mathbf{F}(\mathbf{x}(s),\, s)\,ds \\
&= \mathbf{x}_0.
\end{aligned}$$

Therefore, $\mathbf{x}(t)$ is a solution to the IVP (2.3). □

Exercise 2.3. Consider the following IVP (below $\mu > 0$ and $\lambda > 0$ are constants, $h_i(x, y)$'s are some continuous functions of x and y, and x_0, y_0 are two fixed initial conditions):

$$\begin{aligned}
x' &= -\mu x + h_1(x, y) & x(0) &= x_0 \\
y' &= \lambda y + h_2(x, y) & y(0) &= y_0
\end{aligned}$$

Show that the above IVP is equivalent to the following integral system.

$$\begin{aligned}
x(t) &= e^{-\mu t}\left(x_0 + \int_0^t e^{\mu s} h_1(x(s),\, y(s))\,ds\right) \\
y(t) &= e^{\lambda t}\left(y_0 + \int_0^t e^{-\lambda s} h_2(x(s),\, y(s))\,ds\right),
\end{aligned}$$

i.e. $(x(t), y(t))$ solves the IVP if and only if it is a continuous solution to the integral system.

2.2.0.2. Picard's iteration sequence of an initial-value problem. By the equivalence of IVPs and integral equations as illustrated in Proposition 2.3, we can now reformulate the integral equation as an iteration problem. The iteration sequence involved is known as the Picard's iteration sequence.

From now on, the vector field $\mathbf{F}(\mathbf{x}, t)$ is always assumed to be continuous in both $\mathbf{x}$- and t-slots, so that the IVP (2.3) and the integral equation (2.4) are equivalent. This assumption is crucial since the Fundamental Theorem of Calculus applies only to continuous functions.

Definition 2.5 (Picard's Iteration Sequence). Given an IVP

$$\mathbf{x}' = \mathbf{F}(\mathbf{x}, t), \quad \mathbf{x}(0) = \mathbf{x}_0,$$

its Picard's Iteration Sequence is a *sequence of functions* $\{\mathbf{x}_n(t)\}_{n=0}^{\infty}$ defined by:

$$\mathbf{x}_0(t) = \mathbf{x}_0 \qquad \text{(an abuse of notation)}$$

$$\mathbf{x}_n(t) = \mathbf{x}_0 + \int_0^t \mathbf{F}(\mathbf{x}_{n-1}(s),\, s)\, ds \qquad \text{for any } n \geq 1 \tag{2.5}$$

Remark 2.6. We used $\mathbf{x}_0$ to denote both the initial condition and the first term of the iteration sequence. Because we set the two to be equal, there should not be any confusion on this use of notations. □

The Picard's iteration sequence is related to the existence theorem of ODE in the following way. If one is able to show that the sequence $\mathbf{x}_n(t)$ converges a limit function $\mathbf{x}_\infty(t)$ as $n \to \infty$, then *heuristically*, letting $n \to \infty$ on both sides of (2.5) will yield:

$$\begin{aligned}
\lim_{n\to\infty} \mathbf{x}_n(t) &= \lim_{n\to\infty} \left(\mathbf{x}_0 + \int_0^t \mathbf{F}(\mathbf{x}_{n-1}(s),\, s)\, ds \right) \\
\mathbf{x}_\infty(t) &= \mathbf{x}_0 + \lim_{n\to\infty} \int_0^t \mathbf{F}(\mathbf{x}_{n-1}(s),\, s)\, ds \\
&= \mathbf{x}_0 + \int_0^t \lim_{n\to\infty} \mathbf{F}(\mathbf{x}_{n-1}(s),\, s)\, ds && \text{(cheating!)} \\
&= \mathbf{x}_0 + \int_0^t \mathbf{F}\left(\lim_{n\to\infty} \mathbf{x}_{n-1}(s),\, s \right) ds && \text{(true if } \mathbf{F} \text{ is continuous)} \\
\mathbf{x}_\infty(t) &= \mathbf{x}_0 + \int_0^t \mathbf{F}(\mathbf{x}_\infty(s),\, s)\, ds.
\end{aligned}$$

Therefore, $\mathbf{x}_\infty(t)$ solves the integral equation (2.4) which is equivalent to the given IVP. Therefore, pending all justifications of the steps that we 'cheated', we have shown the limit $\mathbf{x}_\infty(t)$ is a solution to the IVP. Although this limit function $\mathbf{x}_\infty(t)$ may not be explicit, it at least shows that a solution to the IVP exists and we can study its qualitative behaviors such as stability.

However, we still have to justify the steps that we 'cheated', namely why we can switch the limit and the integral signs. Even in one dimension, there are examples of sequence of continuous functions $\{x_n(t)\}_{n=0}^{\infty}$ that converges to a limit function $x_\infty(t)$ as $n \to \infty$, but

$$\lim_{n\to\infty} \int_0^t x_n(s) ds \neq \int_0^t \lim_{n\to\infty} x_n(s) ds.$$

A substantial part of the existence theorem is to show that it is legitimate to switch the limit and integral signs for the Picard's iteration sequence. It involves a concept in analysis known as *uniform convergence* (See Appendix A.1).

Let's look at a few examples of Picard's iteration before we go into the analysis of the general case. Some of the examples can possibly be solved in an more elementary way. However, in order to make better sense of the Picard's iteration, let's pretend not knowing how to solve them. These examples will convince us that the Picard's iteration sequence will indeed converge to the solution of the IVP.

Let's start with one dimension:

Example 2.1. Consider the IVP: $x' = x$, $x(0) = 1$. Obviously, the solution should be $x(t) = e^t$, but as said in the previous paragraph, let's pretend we don't know this answer and try to use Picard's iteration to solve this IVP.

By Proposition 2.3, the equivalent integral equation is given by:

$$x(t) = 1 + \int_0^t x(s)ds,$$

and therefore its Picard's iteration sequence is defined as:

$$\begin{aligned} x_0(t) &= 1 \\ x_n(t) &= 1 + \int_0^t x_{n-1}(s)\,ds \quad \text{for } n \geq 1. \end{aligned}$$

Let's compute the first few terms of the iteration:

$$\begin{aligned} x_0(t) &= 1; \\ x_1(t) &= 1 + \int_0^t x_0(s)ds \\ &= 1 + \int_0^t 1ds \\ &= 1 + [s]_0^t = 1 + t; \\ x_2(t) &= 1 + \int_0^t x_1(s)ds \\ &= 1 + \int_0^t (1+s)ds \\ &= 1 + \left[s + \frac{s^2}{2}\right]_0^t \\ &= 1 + t + \frac{t^2}{2}; \end{aligned}$$

$$\begin{aligned} x_3(t) &= 1 + \int_0^t \left(1 + s + \frac{s^2}{2}\right) ds \\ &= 1 + \left[s + \frac{s^2}{2} + \frac{s^3}{3 \cdot 2}\right]_0^t \\ &= 1 + t + \frac{t^2}{2!} + \frac{t^3}{3!}. \end{aligned}$$

Keep iterating, we should see from the pattern that

$$x_n(t) = 1 + t + \frac{t^2}{2!} + \frac{t^3}{3!} + \ldots + \frac{t^n}{n!} = \sum_{k=0}^{n} \frac{t^k}{k!}.$$

Let's prove it by induction. Suppose $x_i(t) = \sum_{k=0}^{i} \frac{t^k}{k!}$, and consider $x_{i+1}(t)$:

$$\begin{aligned}
x_{i+1}(t) = 1 + \int_0^t x_i(s)ds = 1 + \int_0^t \sum_{k=0}^{i} \frac{s^k}{k!} ds \\
&= 1 + \sum_{k=0}^{i} \left[\frac{s^{k+1}}{k!(k+1)} \right]_{s=0}^{s=t} = 1 + \sum_{k=0}^{i} \frac{t^{k+1}}{(k+1)!} \\
&= \frac{t^0}{0!} + \sum_{k=1}^{i+1} \frac{t^k}{k!} && \text{(shifting indices)} \\
&= \sum_{k=0}^{i+1} \frac{t^k}{k!} && \text{(absorbing the zero-th term)}
\end{aligned}$$

as desired. By induction, the Picard's iteration sequence is given as an infinite series $x_n(t) = \sum_{k=0}^{n} \frac{t^k}{k!}$. As $n \to \infty$, it converges to

$$x_\infty(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!}$$

which is the Taylor's series of the function e^t, our expected solution to the IVP. □

Example 2.2. Consider the IVP

$$x' = 2t(1 + x), \quad x(0) = 0.$$

The equivalent integral equation is:

$$x(t) = \int_0^t 2s(1 + x(s))ds$$

and the Picard's iteration sequence is defined as :

$$\begin{aligned}
x_0(t) &= 0 \\
x_n(t) &= \int_0^t 2s(1 + x_{n-1}(s))ds \quad \text{for } n \geq 1.
\end{aligned}$$

By direct computations (exercise), one should get:

$$\begin{aligned}
x_1(t) &= t^2 \\
x_2(t) &= t^2 + \frac{t^4}{2} \\
x_3(t) &= t^2 + \frac{t^4}{2!} + \frac{t^6}{3!}
\end{aligned}$$

and from the pattern we conjecture that

$$x_n(t) = t^2 + \frac{t^4}{2!} + \ldots + \frac{t^{2n}}{n!} = \sum_{k=1}^{n} \frac{t^{2k}}{k!}.$$

Let's verify this is true by induction. Suppose $x_i(t) = \sum_{k=1}^{i} \frac{t^{2k}}{k!}$ and consider $x_{i+1}(t)$:

$$\begin{aligned}
x_{i+1}(t) &= \int_0^t 2s(1 + x_i(s))ds = \int_0^t 2s\left(1 + \sum_{k=1}^{i} \frac{s^{2k}}{k!}\right) ds \\
&= 2\int_0^t \left(s + \sum_{k=1}^{i} \frac{s^{2k+1}}{k!}\right) ds = 2\left(\left[\frac{s^2}{2} + \sum_{k=1}^{i} \frac{s^{2k+2}}{k!(2k+2)}\right]_{s=0}^{s=t}\right) \\
&= t^2 + \sum_{k=1}^{i} \frac{t^{2k+2}}{k!(k+1)} = t^2 + \sum_{k=1}^{i} \frac{t^{2k+2}}{(k+1)!} \\
&= \frac{t^2}{1!} + \sum_{k=2}^{i+1} \frac{t^{2k}}{k!} = \sum_{k=1}^{i+1} \frac{t^{2k}}{k!}
\end{aligned}$$

as desired. By induction, $x_n(t) = \sum_{k=1}^{n} \frac{t^{2k}}{k!}$ and converges as $n \to \infty$ to

$$x_\infty(t) = \sum_{k=1}^{\infty} \frac{t^{2k}}{k!} = -1 + \sum_{k=0}^{\infty} \frac{(t^2)^k}{k!} = -1 + e^{t^2}.$$

One can verify that it is indeed a solution to the IVP. □

Exercise 2.4. For each of the following IVPs, write down the first few terms of its Picard's iteration sequence, deduce the general term of the sequence followed by a proof by induction.

(1) $x' = tx, \quad x(0) = 1$

(2) $x' = t + x, \quad x(0) = 0$

Next we look at a two dimensional example:

Example 2.3. Consider the second-order IVP: $x'' = -x,\ x(0) = 0,\ x'(0) = 1$, which can be written as a first-order two dimensional system of ODE with initial condition

$$\begin{bmatrix} x \\ v \end{bmatrix}' = \begin{bmatrix} v \\ -x \end{bmatrix}, \qquad \begin{bmatrix} x(0) \\ v(0) \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

The system is equivalent to the integral equation:

$$\begin{bmatrix} x(t) \\ v(t) \end{bmatrix}' = \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \int_0^t \begin{bmatrix} v(s) \\ -x(s) \end{bmatrix} ds.$$

Integrating a vector-valued function simply means integrating each of its component, i.e.

$$\int_0^t \begin{bmatrix} v(s) \\ -x(s) \end{bmatrix} ds = \begin{bmatrix} \int_0^t v(s)ds \\ -\int_0^t x(s)ds \end{bmatrix}.$$

The Picard's iteration sequence $\mathbf{x}_n =: \begin{bmatrix} x_n(t) \\ v_n(t) \end{bmatrix}$ is defined as follows:

$$\begin{aligned}
\begin{bmatrix} x_0(t) \\ v_0(t) \end{bmatrix} &= \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\
\underbrace{\begin{bmatrix} x_n(t) \\ v_n(t) \end{bmatrix}}_{\mathbf{x}_n(t)} &= \underbrace{\begin{bmatrix} 0 \\ 1 \end{bmatrix}}_{\mathbf{x}_0} + \underbrace{\begin{bmatrix} \int_0^t v_{n-1}(s)ds \\ -\int_0^t x_{n-1}(s)ds \end{bmatrix}}_{\int_0^t \mathbf{F}(\mathbf{x}_{n-1}(s))ds}
\end{aligned}$$

By direct computations (exercise), we get:

$$\mathbf{x}_1(t) = \begin{bmatrix} t \\ 1 \end{bmatrix} \qquad \mathbf{x}_3(t) = \begin{bmatrix} t - \frac{t^3}{3!} \\ 1 - \frac{t^2}{2} \end{bmatrix}$$

$$\mathbf{x}_2(t) = \begin{bmatrix} t \\ 1 - \frac{t^2}{2} \end{bmatrix} \qquad \mathbf{x}_4(t) = \begin{bmatrix} t - \frac{t^3}{3!} \\ 1 - \frac{t^2}{2!} + \frac{t^4}{4!} \end{bmatrix}$$

Exercise 2.5. Verify the above calculations.

Based on these patterns, we conjectured that

$$\mathbf{x}_{2n-1}(t) = \begin{bmatrix} x_{2n-1}(t) \\ v_{2n-1}(t) \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^{n} \frac{(-1)^{k-1}t^{2k-1}}{(2k-1)!} \\ \sum_{k=0}^{n-1} \frac{(-1)^k t^{2k}}{(2k)!} \end{bmatrix} \tag{2.6}$$

$$\mathbf{x}_{2n}(t) = \begin{bmatrix} x_{2n}(t) \\ v_{2n}(t) \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^{n} \frac{(-1)^{k-1}t^{2k-1}}{(2k-1)!} \\ \sum_{k=0}^{n} \frac{(-1)^k t^{2k}}{(2k)!} \end{bmatrix}. \tag{2.7}$$

Again, to prove this claim, we use induction. Assume the above is true for some n, then we consider both $\mathbf{x}_{2n+1}$ and $\mathbf{x}_{2n+2}$:

$$\begin{aligned}
\mathbf{x}_{2n+1}(t) &= \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \begin{bmatrix} \int_0^1 v_{2n}(s)ds \\ -\int_0^1 x_{2n}(s)ds \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \begin{bmatrix} \int_0^1 \sum_{k=0}^{n} \frac{(-1)^k s^{2k}}{(2k)!} ds \\ -\int_0^t \sum_{k=1}^{n} \frac{(-1)^{k-1} s^{2k-1}}{(2k-1)!} ds \end{bmatrix} \\
&= \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \begin{bmatrix} \sum_{k=0}^{n} \left[\frac{(-1)^k s^{2k+1}}{(2k+1)!}\right]_{s=0}^{s=t} \\ -\sum_{k=1}^{n} \left[\frac{(-1)^{k-1} s^{2k}}{(2k)!}\right]_{s=0}^{s=t} \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^{n+1} \frac{(-1)^{k-1}t^{2k-1}}{(2k-1)!} \\ 1 - \sum_{k=1}^{n} \frac{(-1)^{k-1}t^{2k}}{(2k)!} \end{bmatrix} \\
&= \begin{bmatrix} \sum_{k=1}^{n+1} \frac{(-1)^{k-1}t^{2k-1}}{(2k-1)!} \\ \sum_{k=0}^{n} \frac{(-1)^k t^{2k}}{(2k)!} \end{bmatrix}, \\
\mathbf{x}_{2n+2}(t) &= \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \begin{bmatrix} \int_0^t v_{2n+1}(s)ds \\ -\int_0^t x_{2n+1}(s)ds \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \begin{bmatrix} \int_0^t \sum_{k=0}^{n} \frac{(-1)^k s^{2k}}{(2k)!} ds \\ -\int_0^t \sum_{k=1}^{n+1} \frac{(-1)^{k-1} s^{2k-1}}{(2k-1)!} ds \end{bmatrix} \\
&= \begin{bmatrix} \sum_{k=0}^{n} \frac{(-1)^k t^{2k+1}}{(2k+1)!} \\ 1 + \sum_{k=1}^{n+1} \frac{(-1)^k t^{2k}}{(2k)!} \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^{n+1} \frac{(-1)^{k-1}t^{2k-1}}{(2k-1)!} \\ \sum_{k=0}^{n+1} \frac{(-1)^k t^{2k}}{(2k)!} \end{bmatrix}.
\end{aligned}$$

Therefore, the claim is also true for $\mathbf{x}_{2n+1}(t)$ and $\mathbf{x}_{2n+2}(t)$. Hence (2.6) and (2.7) holds for all $n \geq 1$. Let $n \to \infty$ in either one of (2.6) and (2.7), we have:

$$\mathbf{x}_\infty(t) = \begin{bmatrix} \sum_{k=1}^{\infty} \frac{(-1)^{k-1}t^{2k-1}}{(2k-1)!} \\ \sum_{k=0}^{\infty} \frac{(-1)^k t^{2k}}{(2k)!} \end{bmatrix} = \begin{bmatrix} \sin t \\ \cos t \end{bmatrix}.$$

Therefore $x_\infty(t) = \sin t$ and $v_\infty(t) = \cos t$, which clearly form a solution to the first-order system, and $x_\infty(t) = \sin t$ is also a solution to the second-order IVP. □

Exercise 2.6. Deduce the general term of the Picard's iteration sequence of the first-order system corresponding to the second-order IVP:

$$x'' = -4x, \quad x(0) = 0, \quad x'(0) = 2.$$

Does the sequence converges? If so, does it converge to a solution to this IVP?

Example 2.4. In this last example about Picard's iteration, we consider a linear IVP

$$\mathbf{x}' = A\mathbf{x}, \quad \mathbf{x}(0) = \mathbf{x}_0.$$

which was discussed in detail in Chapter 1. We will show that the Picard's iteration sequence of these linear systems will converge to the solution in terms of matrix exponential, i.e. $e^{tA}\mathbf{x}_0$, as we have seen in Theorem 1.28.

First we rewrite the IVP as an integral equation:

$$\mathbf{x}(t) = \mathbf{x}_0 + \int_0^t A\mathbf{x}(s)ds.$$

and its Picard's iteration sequence is defined by:

$$\begin{aligned}\mathbf{x}_0(t) &= \mathbf{x}_0 \\ \mathbf{x}_n(t) &= \mathbf{x}_0 + \int_0^t A\mathbf{x}_{n-1}(s)ds.\end{aligned}$$

By direction computation, one can show

$$\begin{aligned}\mathbf{x}_1(t) &= \mathbf{x}_0 + \int_0^t A\mathbf{x}_0 ds \\ &= \mathbf{x}_0 + [sA\mathbf{x}_0]_{s=0}^{s=t} && \text{(Regard } A\mathbf{x}_0 \text{ as a constant.)} \\ &= \mathbf{x}_0 + tA\mathbf{x}_0 \\ \mathbf{x}_2(t) &= \mathbf{x}_0 + \int_0^t A\mathbf{x}_1(s)ds = \mathbf{x}_0 + \int_0^t A\mathbf{x}_0 + sA^2\mathbf{x}_0 ds \\ &= \mathbf{x}_0 + tA\mathbf{x}_0 + \frac{t^2}{2}A^2\mathbf{x}_0 = \mathbf{x}_0 + tA\mathbf{x}_0 + \frac{(tA)^2}{2}\mathbf{x}_0.\end{aligned}$$

Based on this pattern, we conjecture:

$$\mathbf{x}_n(t) = \mathbf{x}_0 + tA\mathbf{x}_0 + \frac{(tA)^2}{2!}\mathbf{x}_0 + \ldots + \frac{(tA)^n}{n!}\mathbf{x}_0 = \left(\sum_{k=0}^n \frac{(tA)^k}{k!}\right)\mathbf{x}_0.$$

Exercise 2.7. Prove, by induction, that

$$\mathbf{x}_n(t) = \left(\sum_{k=0}^n \frac{(tA)^k}{k!}\right)\mathbf{x}_0.$$

Let $n \to \infty$, we have $\mathbf{x}_n(t) \to \mathbf{x}_\infty(t)$ which is given by:

$$\mathbf{x}_\infty(t) = \sum_{k=0}^\infty \frac{(tA)^k}{k!}\mathbf{x}_0 = e^{tA}\mathbf{x}_0$$

as desired. □

In all examples we have seen so far, the Picard's iteration sequence converges to the solutions we expect, but it is important to keep in mind that the general terms of most Picard's iteration sequences are difficult to deduce.

We will ultimately show that the Picard's iteration sequence $\{\mathbf{x}_n(t)\}_{n=0}^\infty$ associated to the IVP $\mathbf{x}' = \mathbf{F}(\mathbf{x}, t), \ \mathbf{x}(0) = \mathbf{x}_0$ must converge provided the vector field $\mathbf{F}(\mathbf{x}, t)$ satisfies certain continuity assumption, known as *Lipschitz continuity* to be introduced in the next section. Under this assumption, one can also prove that $\mathbf{x}_n(t)$ converges *uniformly* as $n \to \infty$. Combining the Lipschitz continuity of $\mathbf{F}$, one can justify why it is true that:

$$\lim_{n\to\infty} \int_0^t \mathbf{F}(\mathbf{x}_n(s),\ s)\, ds = \int_0^t \lim_{n\to\infty} \mathbf{F}(x_n(s),\ s)\, ds,$$

which is a step we 'cheated' in page 41. After all these analytic issues have been resolved, one can prove the existence theorem by showing that the Picard's iteration sequence converges to a solution to the IVP.

2.3. Lipschitz Continuity

Lipschitz continuity is a crucial concept for studying ODE systems. As you will see in the proof of the existence theorem, the Lipschitz continuity of the vector field $\mathbf{F}(\mathbf{x}, t)$ is used to show that the Picard's iteration sequence converges uniformly. Furthermore, in Section 2.7, we will show that Lipschitz continuity of the vector field $\mathbf{F}$ always implies the solution to any IVP is unique.

Recall from multivariable calculus that a function $\mathbf{F} : \mathbb{R}^d \to \mathbb{R}^m$ is said to be continuity at $\mathbf{x}_0$ if $|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{x}_0)| \to 0$ as $|\mathbf{x} - \mathbf{x}_0| \to 0$. One can say a function is continuous at a *point*, or on a region. A function $\mathbf{F}$ is continuous on a region Ω if it is continuous at every point in Ω. However, one can only talk about Lipschitz continuity on a region Ω, as demonstrated in its definition:

Definition 2.7. Let Ω be a domain in $\mathbb{R}^d$ and I be any time interval (which can be infinite, closed, open, or half-open). A time-dependent function $\mathbf{F}(\mathbf{x}, t) : \Omega \times I \to \mathbb{R}^m$ is said to be *Lipschitz continuous* on $\Omega \times I$ if there exists a constant $L > 0$, called a Lipschitz constant, such that

$$|\mathbf{F}(\mathbf{x}, t) - \mathbf{F}(\mathbf{y}, t)| \le L|\mathbf{x} - \mathbf{y}| \quad \text{for any } x, y \in \Omega \text{ and } t \in I.$$

A time-independent function $\mathbf{G} : \Omega \to \mathbb{R}^m$ is said to be *Lipschitz continuous* on Ω if there exists a constant $L' > 0$ such that

$$|\mathbf{G}(\mathbf{x}) - \mathbf{G}(\mathbf{y})| \le L'|\mathbf{x} - \mathbf{y}| \quad \text{for any } x, y \in \Omega.$$

Remark 2.8. By the squeezing principle, a function $\mathbf{F}(\mathbf{x}, t)$ being Lipschitz continuous on $\Omega \times I$ implies it is continuous in the $\mathbf{x}$-slot at every point on Ω. □

Remark 2.9. Note that if $\Omega' \subset \Omega$, then $\mathbf{F}(\mathbf{x}, t)$ being Lipschitz continuous on $\Omega \times I$ implies it is Lipschitz continuous on $\Omega' \times I$. However, it is not vice versa. □

Example 2.5. Some easy examples of Lipschitz continuous functions include:

(1) For any constant $a \in \mathbb{R}$, the function $F(x) = ax$ is Lipschitz continuous on $\mathbb{R}$: for any $x, y \in \mathbb{R}$, we have

$$|F(x) - F(y)| = |ax - ay| = |a||x - y|.$$

(2) $F(x) = |x|$ is Lipschitz continuous on $\mathbb{R}$: for any $x, y \in \mathbb{R}$, we have

$$|F(x) - F(y)| = ||x| - |y|| \le |x - y|.$$

(3) $F(x) = \sin x$ is Lipschitz continuous on $\mathbb{R}$: for any $x, y \in \mathbb{R}$, we have

$$\begin{aligned}|F(x) - F(y)| &= |\sin x - \sin y| \\ &= |\cos \xi||x - y| \quad \text{for some } \xi \text{ between } x \text{ and } y \\ &\le 1 \cdot |x - y|.\end{aligned}$$

(4) Given an $d \times d$ matrix A, the vector field $\mathbf{F}(\mathbf{x}) = A\mathbf{x}$ is Lipschitz continuous on $\mathbb{R}^d$: for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, we have

$$|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{y})| = |A(\mathbf{x} - \mathbf{y})| \le \|A\| \, |\mathbf{x} - \mathbf{y}|.$$

Here we have used Lemma 1.26.

(5) $F(x, t) = xt$ is Lipschitz continuous on $\mathbb{R} \times [-T, T]$ where T is a fixed positive number: for any $x, y \in \mathbb{R}$ and $t \in [-T, T]$, we have

$$|F(x, t) - F(y, t)| = |xt - yt| = |t||x - y| \le T \cdot |x - y|.$$

However, it is *not* Lipschitz continuous on $(x,t) \in \mathbb{R} \times (-\infty, \infty)$ since for any $x \neq y$ in $\mathbb{R}$, we have

$$\frac{|F(x,y) - F(y,t)|}{|x-y|} = |t| \to \infty \text{ as } t \to \infty.$$

If F were Lipschitz continuous on $\mathbb{R} \times (-\infty, \infty)$, the fraction $\frac{|F(x,t)-F(y,t)|}{|x-y|}$ (when $x \neq y$) must be bounded.

□

The last example demonstrated that a function can be Lipschitz continuous on a one space-time domain but not on a larger one. Therefore, whenever we talk about Lipschitz continuity, one should indicate the domain. It does not make much sense for just saying that a function is Lipschitz continuous.

2.3.0.1. Bounded derivative test for Lipschitz continuity. Before we give more examples of non-Lipschitz continuous functions, we next talk about a very effective test to determine whether or not a function is Lipschitz continuous on a given domain. As demonstrated in the $F(x) = \sin x$ example, the Lipschitz continuity follows easily from the mean-value theorem since the first derivative F' is bounded from above and below. This technique can actually be stated as a general theorem.

Theorem 2.10 (Bounded Derivative Test: one dimension)**.** *Let $F(x,t) : \Omega \times I \to \mathbb{R}$ be a differentiable function on an interval domain $\Omega \subset \mathbb{R}$ and time interval I, then F is Lipschitz continuous on $\Omega \times I$ if and only if the partial derivative $\frac{\partial F}{\partial x}$ is bounded on $\Omega \times I$, i.e. there exists a constant $C > 0$ such that $\left|\frac{\partial F}{\partial x}(x,t)\right| \leq C$ for all $(x,\, t) \in \Omega \times I$.*

Similar results hold for time-independent functions: let $G(x) : \Omega \to \mathbb{R}$ be a differentiable function on an interval domain $\Omega \subset \mathbb{R}$, then G is Lipschitz continuous on Ω if and only if the first derivative $G'(x)$ is bounded on Ω, i.e. there exists a constant $C' > 0$ such that $|G'(x)| \leq C'$ for all $x \in \Omega$.

Proof. We only give the proof for the time-dependent functions since the proof of time-independent functions is exactly the same.

Suppose $F(x,t) : \Omega \times I \to \mathbb{R}$ is Lipschitz continuous on $\Omega \times I$, then there exists a constant $L > 0$ such that

$$|F(x,t) - F(x_0,t)| \leq L|x - x_0| \quad \text{for any } x,\, x_0 \in \Omega \text{ and } t \in I.$$

Then for any $x \neq x_0$, one has:

$$\left|\frac{F(x,t) - F(x_0,t)}{x - x_0}\right| \leq L.$$

Since $F(x,t)$ is differentiable on $\Omega \times I$, the partial derivative

$$\frac{\partial F}{\partial x}(x_0,t) := \lim_{x \to x_0} \frac{F(x,t) - F(x_0,t)}{x - x_0}$$

exists. It follows from the above inequality that one must have $\left|\frac{\partial F}{\partial x}(x_0,t)\right| \leq L$. Since $(x_0,t) \in \Omega \times I$ is arbitrarily chosen, we have $\left|\frac{\partial F}{\partial x}\right| \leq L$ on every $(x,t) \in \Omega \times I$.

Conversely, suppose $\left|\frac{\partial F}{\partial x}\right| \leq C$ on every $(x,t) \in \Omega \times I$. Then given any $x, y \in \Omega$ and $t \in I$, we have

$$\begin{aligned}
F(x,t) - F(y,t) &= \frac{\partial F}{\partial x}(\xi,t) \cdot (x-y) && \text{(mean-value theorem in the } x\text{-slot)} \\
|F(x,t) - F(y,t)| &= \left|\frac{\partial F}{\partial x}(\xi,t)\right| |x-y| \leq C|x-y| && \text{(by the given hypothesis)}
\end{aligned}$$

which shows $F(x,t)$ is Lipschitz continuous on $\Omega \times I$. $\square$

Example 2.6. Let's look at a few examples on how to apply Theorem 2.10.

(1) $F(x) = x^5$ is Lipschitz continuous on $[0,1]$ because $F'(x) = 5x^4$ and for $x \in [0,1]$, we have $|F'(x)| = |5x^4| \leq 5$ on $x \in [0,1]$. However, it is not Lipschitz continuous on $\mathbb{R}$ because $|F'(x)| = 5x^4 \to \infty$ as $x \to \infty$.

(2) $F(x) = x^{1/2}$ is *not* Lipschitz continuous on $[0,\infty)$, since F is differentiable on $(0,\infty)$ and $F'(x) = \frac{1}{2\sqrt{x}} \to \infty$ as $x \to 0$. Therefore, $|F'(x)|$ is not bounded on $(0,\infty)$ and so Theorem 2.10 asserts it is not Lipschitz continuous on $(0,\infty)$. By the definition of Lipschitz continuity, it is not Lipschitz continuous on the larger domain $[0,\infty)$.

However, it *is* Lipschitz continuous on $[\varepsilon,\infty)$ where $\varepsilon > 0$ is any fixed constant. To see this, consider

$$|F'(x)| = \frac{1}{2\sqrt{x}} \leq \frac{1}{2\sqrt{\varepsilon}} \quad \text{for any } x \in [\varepsilon,\infty).$$

As ε is a fixed non-zero constant, Theorem 2.10 asserts that F is Lipschitz continuous on $[\varepsilon,\infty)$.

(3) Consider the time-dependent function $F(x,t) = tx$ which was discussed before using the definition of Lipschitz continuity. One can show $F(x,t)$ is not Lipschitz continuous on $\mathbb{R} \times (-\infty,\infty)$ because

$$\left|\frac{\partial F}{\partial x}\right| = |t| \to \infty \quad \text{as } t \to \infty.$$

However, it *is* Lipschitz continuous on $\mathbb{R} \times [-T,T]$ for any fixed constant $T > 0$, since $\left|\frac{\partial F}{\partial x}\right| = |t| \leq T$ for any $(x,t) \in \mathbb{R} \times [-T,T]$.

$\square$

Remark 2.11. Although the bounded derivative test is very straight-forward, it only applies to differentiable functions. One cannot use this test for non-differentiable functions such as $F(x) = |x|$. $\square$

Exercise 2.8. Determine whether or not each of the following functions is Lipschitz continuous on the specified domain:

(a) $F(x) = x^{1/3}$ on $x \in [-1,1]$.

(b) $F(x) = x^{1/3}$ on $x \in [-\frac{1}{2},1]$.

(c) $F(x) = x^3$ on $x \in [-M,M]$ where $M > 0$ is a fixed number.

(d) $F(x) = x^3$ on $x \in \mathbb{R}$.

(e) $F(x) = 1/x$ on $x \in [1,\infty)$.

(f) $F(x) = 1/x$ on $x \in (0,1]$.

(g) $F(x) = \sin(\cos x)$ on $x \in \mathbb{R}$.

(h) $F(x) = \sin(\cos x)$ on $x \in [0,\, 2\pi]$.

(i) $F(x,t) = t^{1/3}$ on $(x,t) \in \mathbb{R} \times (-\infty,\infty)$.

(j) $F(x,t) = \sin(\cos(x+t^2))$ on $(x,t) \in \mathbb{R} \times (-\infty,\infty)$.

(k) $F(x,t) = \cos(tx)$ on $(x,t) \in \mathbb{R} \times (-\infty,\infty)$.

Exercise 2.9. Consider the function
$$F(x,t) = \frac{t(x^2+1)}{x}.$$
Determine whether or not $F(x,t)$ is Lipschitz continuous on each of the following domain $\Omega \times I$:

(a) $\Omega \times I = [1,2] \times [0,1]$
(b) $\Omega \times I = (1,2) \times [0,1]$
(c) $\Omega \times I = [1,2] \times [0,\infty)$
(d) $\Omega \times I = [1,\infty) \times [0,T]$ where $T > 0$ is a fixed number.
(e) $\Omega \times I = (0,1) \times [0,1]$.

Exercise 2.10. Determine all possible values of a such that $F(x) = x^a$ is Lipschitz continuous on $[0,1]$.

Exercise 2.11. Given a function $F(x) : \Omega \to \mathbb{R}$ and a function $G(y) : F(\Omega) \to \mathbb{R}$, where $F(\Omega)$ denotes the image of F. Suppose F is Lipschitz continuous on Ω and G is Lipschitz continuous on $F(\Omega)$, show that the composition $G \circ F$ is Lipschitz continuous on Ω.

The bounded derivative test can be generalized to higher dimensions. However, one should note that there is a convexity condition for the domain Ω.

Definition 2.12 (Convex Domain). A subset $\Omega \subset \mathbb{R}^d$ is said to be *convex* if for any pair of points $\mathbf{x}$, $\mathbf{y} \in \Omega$, the line segment joining $\mathbf{x}$ and $\mathbf{y}$ lies completely inside Ω. Precisely, the straight-line
$$\mathbf{r}(s) := s\mathbf{x} + (1-s)\mathbf{y}$$
is contained in Ω for any $s \in [0,1]$.

Theorem 2.13 (Bounded Derivative Test: higher dimensions). *Let $\mathbf{F}(\mathbf{x},t) : \Omega \times I \to \mathbb{R}^m$ be a differentiable function on a convex domain $\Omega \subset \mathbb{R}^d$ and time interval I. Denote x_j, where $1 \le j \le d$, be the j-th component of $\mathbf{x}$, i.e. $\mathbf{x} = (x_1, \ldots, x_d)$, and F_i, where $1 \le i \le m$, be the i-th component of $\mathbf{F}$. Precisely:*
$$\mathbf{F}(\mathbf{x},t) = \begin{bmatrix} F_1(x_1,\ldots,x_d,t) \\ \vdots \\ F_m(x_1,\ldots,x_d,t) \end{bmatrix}.$$
Then, $\mathbf{F}$ is Lipschitz continuous on $\Omega \times I$ if and only if all first partial derivatives $\frac{\partial F_i}{\partial x_j}$, where $1 \le i \le m$ and $1 \le j \le d$, are bounded above and below on $(\mathbf{x},t) \in \Omega \times I$.

Remark 2.14. An equivalent way (without involving components of $\mathbf{x}$ and $\mathbf{F}$) to state the boundedness of all $\frac{\partial F_i}{\partial x_j}$'s is that there exists a constant $C > 0$ such that: $\|D\mathbf{F}(\mathbf{x},t)\| \le C$ for all $(\mathbf{x},t) \in \Omega \times I$, where $D\mathbf{F}(\mathbf{x},t)$ denotes the Jacobian matrix of $\mathbf{F}$ at $(\mathbf{x},t)$ to be defined in Chapter 3. □

Proof of Theorem 2.13. Suppose there exists a constant $C > 0$ such that $\left|\frac{\partial F_i}{\partial x_j}\right| \le C$ on $\Omega \times I$ for any i, j. Take any pair of point $\mathbf{y}$, $\mathbf{z} \in \Omega$ and $t \in I$, the line segment
$$\mathbf{r}(s) = s\mathbf{y} + (1-s)\mathbf{z}, \ s \in [0,1]$$

lies inside Ω by the convexity assumption. Consider the composition of $\mathbf{F}$ and $\mathbf{r}$:

$$\mathbf{G}(s,t) := \mathbf{F}(\mathbf{r}(s),t) : [0,1] \times I \to \mathbb{R}^m$$

which is also differentiable on $[0,1] \times I$. Let $G_i(s,t)$ be the i-th component of $\mathbf{G}(s,t)$. For each i, by the mean-value theorem applied to G_i in the s-slot, we have:

$$\begin{aligned} G_i(1,t) - G_i(0,t) &= \frac{\partial G_i}{\partial s}(\xi_i,t)\cdot(1-0) && \text{for some } \xi_i \in [0,1] \\ &= \sum_{k=1}^{d} \frac{\partial F_i}{\partial x_k}\cdot\frac{dx_k}{ds} && \text{(chain rule)} \end{aligned}$$

Along the straight-line path, the k-th component of $\mathbf{r}(s)$ is $x_k := sy_k + (1-s)z_k$ where x_k and y_k denotes the k-th components of $\mathbf{y}$ and $\mathbf{z}$ respectively. Therefore,

$$\begin{aligned} |G_i(1,t) - G_i(0,t)| &= \left|\sum_{k=1}^{d} \frac{\partial F_i}{\partial x_k}\cdot\frac{d}{ds}(sy_k + (1-s)z_k)\right| \\ &= \left|\sum_{k=1}^{d} \frac{\partial F_i}{\partial x_k}\cdot(y_k - z_k)\right| \le \sum_{k=1}^{d}\left|\frac{\partial F_i}{\partial x_k}\right| |y_k - z_k| && \text{(triangle inequality)} \\ &\le \left(\sum_{k=1}^{d}\left|\frac{\partial F_i}{\partial x_k}\right|^2\right)^{1/2}\left(\sum_{k=1}^{n}|y_k - z_k|^2\right)^{1/2} && \text{(Cauchy-Schwarz)} \\ &\le \underbrace{\left(\sum_{k=1}^{d} C^2\right)^{1/2}}_{\text{given}} \cdot \underbrace{|\mathbf{y}-\mathbf{z}|}_{\text{definition}} = \sqrt{d}C|\mathbf{y}-\mathbf{z}|. \end{aligned}$$

Finally,

$$\begin{aligned} |\mathbf{G}(1,t) - \mathbf{G}(0,t)| &= \left(\sum_{k=1}^{m} |G_i(1,t) - G_i(0,t)|^2\right)^{1/2} \\ &\le \left(\sum_{k=1}^{m} dC^2|\mathbf{y}-\mathbf{z}|^2\right)^{1/2} \\ &= \sqrt{md}C|\mathbf{y}-\mathbf{z}|. \end{aligned}$$

By definitions of $\mathbf{G}$ and $\mathbf{r}$, we have $\mathbf{G}(1,t) = \mathbf{F}(\mathbf{r}(1),s) = \mathbf{F}(\mathbf{y},t)$ and similarly $\mathbf{G}(0,t) = \mathbf{F}(\mathbf{z},t)$. Therefore,

$$|\mathbf{F}(\mathbf{y},t) - \mathbf{F}(\mathbf{z},t)| \le \sqrt{md}C|\mathbf{y}-\mathbf{z}|.$$

Since $\mathbf{y}$, $\mathbf{z}$ and t are arbitrary, $\mathbf{F}(\mathbf{x},t)$ is Lipschitz continuous on $\Omega \times I$.

Conversely, assume $\mathbf{F}$ is Lipschitz continuous on $\Omega \times I$, then there exists $L > 0$ such that

$$|\mathbf{F}(\mathbf{y},t) - \mathbf{F}(\mathbf{z},t)| \le L\,|\mathbf{y}-\mathbf{z}| \quad \text{for any } \mathbf{y},\, \mathbf{z} \in \Omega,\ t \in I.$$

Then this implies for any i, we have

$$|F_i(\mathbf{x}_0 + s\mathbf{e}_j, t) - F_i(\mathbf{x}_0,t)| \le |\mathbf{F}(\mathbf{x}_0 + s\mathbf{e}_j, t) - \mathbf{F}(\mathbf{x}_0,t)| \le L|\mathbf{x}_0 + s\mathbf{e}_j - \mathbf{x}_0| = L|s|$$

for any $\mathbf{x}_0 \in \Omega$, $t \in I$ and s in some small interval $(-\varepsilon,\varepsilon)$. Here $\mathbf{e}_j$ denotes the j-th standard basis vector of $\mathbb{R}^d$, and ε is small enough such that $\mathbf{x}_0 + s\mathbf{e}_j \in \Omega$ for any $s \in (-\varepsilon,\varepsilon)$. Then by the definition of partial derivatives, we have

$$\left|\frac{\partial F_i}{\partial x_j}(\mathbf{x}_0,t)\right| = \lim_{s\to 0}\left|\frac{F_i(\mathbf{x}_0 + s\mathbf{e}_j,t) - F_i(\mathbf{x}_0,t)}{s}\right| \le L.$$

It completes our proof. □

From now on, we will use F_i to denote the i-th component of $\mathbf{F}$, and x_j to denote the j-th component of $\mathbf{x}$, etc.

Example 2.7. This following examples demonstrate the use of the higher dimension bounded derivative test

(1) The linear map $\mathbf{F}(\mathbf{x}) = A\mathbf{x}$, where $\mathbf{x} \in \mathbb{R}^d$ and A is an $m \times d$ matrix, is Lipschitz continuous on $\mathbb{R}^d$. Denote the (i,j)-th entry of A by $[A]_{ij}$. It can be easily verified that $\frac{\partial F_i}{\partial x_j} = [A]_{ij}$. Therefore, the first partials $\left|\frac{\partial F_i}{\partial x_j}\right|$ are all bounded by $M := \max\{|[A]_{ij}| : 1 \leq i \leq m,\ 1 \leq j \leq d.\}$.

(2) Let $\mathbf{F}(x_1, x_2) = \begin{bmatrix} x_2 \\ 1 - x_1^2 \end{bmatrix}$. Then:

$$\frac{\partial F_1}{\partial x_1} = 0 \qquad \frac{\partial F_1}{\partial x_2} = 1$$
$$\frac{\partial F_2}{\partial x_1} = -2x_1 \qquad \frac{\partial F_2}{\partial x_2} = 0$$

which are all bounded by 2 on the infinite strip $(x_1, x_2) \in [-1, 1] \times \mathbb{R}$. Therefore $\mathbf{F}$ is Lipschitz continuous on $[-1, 1] \times \mathbb{R}$ by Theorem 2.13.

However, it is not Lipschitz continuous on the infinite strip $\mathbb{R} \times [-1, 1]$ since $\left|\frac{\partial F_2}{\partial x_1}\right| = 2|x_1| \to \infty$ as $x_1 \to \infty$.

(3) Let $\mathbf{F}(x_1, x_2, t) = \begin{bmatrix} x_1 + t \\ tx_2 \end{bmatrix}$. Then:

$$\frac{\partial F_1}{\partial x_1} = 1 \qquad \frac{\partial F_1}{\partial x_2} = 0$$
$$\frac{\partial F_2}{\partial x_1} = 0 \qquad \frac{\partial F_2}{\partial x_2} = t$$

Therefore, $\mathbf{F}$ is Lipschitz continuous on $\mathbb{R}^2 \times [-T, T]$ for any fixed number $T > 0$, since all first partials are bounded by T in this space-time domain. However, it is not Lipschitz continuous on $\Omega \times (-\infty, \infty)$ for any region $\Omega \subset \mathbb{R}^2$, as $\left|\frac{\partial F_2}{\partial x_2}\right| = |t| \to \infty$ as $t \to \infty$.

□

Exercise 2.12. Show that $\mathbf{F}(\mathbf{x}) = \begin{bmatrix} \ln x_1 \\ x_2^2 + x_1 \end{bmatrix}$ is Lipschitz continuous on the domain $(x_1, x_2) \in [\varepsilon, \infty) \times [-M, M]$ for any fixed $\varepsilon > 0$ and $M > 0$. However, it is not Lipschitz continuous on either $(0, \infty) \times [-M, M]$ or $[\varepsilon, \infty) \times \mathbb{R}$.

The bounded derivative test allows us to prove Lipschitz continuity by simply showing the boundedness of first partial derivatives. However, there are many examples of functions whose first derivatives are not bounded on the whole domain but only on part of it. It prompts us to define a weaker notion, *locally Lipschitz continuous*, that is less restrictive.

There are several topological concepts such as open sets and closed sets before we can define local Lipschitz continuity.

From now on, we denote:

$$\begin{aligned} B_r^d(\mathbf{x}_0) &= \{\mathbf{x} \in \mathbb{R}^d : |\mathbf{x} - \mathbf{x}_0| < r\} \\ &= \text{the open ball in } \mathbb{R}^d \text{ centered at } \mathbf{x}_0 \text{ with radius } r \end{aligned}$$

Both are convex sets in $\mathbb{R}^d$. If the dimension d is clear from the context, we will omit the superscript d and simply write $B_r(\mathbf{x}_0)$. In one dimension, 'balls' are open intervals. For instance, $B_r(x) = (x-r, x+r)$.

Definition 2.15 (Open Sets and Closed Sets)**.** A subset U of $\mathbb{R}^d$ is said to be **an open set in** $\mathbb{R}^d$ (or alternatively, **open in** $\mathbb{R}^d$) if for every $\mathbf{x} \in U$, there exists $\varepsilon > 0$ such that $B_\varepsilon^d(\mathbf{x}) \subset U$.

A subset C of $\mathbb{R}^d$ is said to be **a closed set in** $\mathbb{R}^d$ (or alternatively, **closed in** $\mathbb{R}^d$) if its complement $\mathbb{R}^d \backslash C$ is an open set in $\mathbb{R}^d$. □

Example 2.8. Any open ball $B_r^d(\mathbf{y})$ is open in $\mathbb{R}^d$: given any $\mathbf{x} \in B_r^d(\mathbf{y})$, one can take $\varepsilon = r - |\mathbf{x}-\mathbf{y}|$ then one can verify by triangle inequality that $B_\varepsilon^d(\mathbf{x}) \subset B_r^d(\mathbf{y})$: pick any $\mathbf{z} \in B_\varepsilon^d(\mathbf{x})$ and we need to show $\mathbf{z} \in B_r^d(\mathbf{y})$. In order to show this, we bound:

$$\begin{aligned} |\mathbf{z}-\mathbf{y}| &= |\mathbf{z}-\mathbf{x}+\mathbf{x}-\mathbf{y}| \\ &\le |\mathbf{z}-\mathbf{x}| + |\mathbf{x}-\mathbf{y}| && \text{(triangle inequality)} \\ &< \varepsilon + |\mathbf{x}-\mathbf{y}| && (\text{since } \mathbf{z} \in B_\varepsilon^d(\mathbf{x})) \\ &= r - |\mathbf{x}-\mathbf{y}| + |\mathbf{x}-\mathbf{y}| = r. \end{aligned}$$

Therefore, $|\mathbf{z}-\mathbf{y}| < r$ and equivalently we have $\mathbf{z} \in B_r^d(\mathbf{y})$. □

Exercise 2.13. Draw a diagram to illustrate the above proof that $B_r^d(\mathbf{y})$ is open in $\mathbb{R}^d$.

Exercise 2.14. Let C be the closed ball with radius r centered at $\mathbf{y}$, i.e. $C = \{\mathbf{x} \in \mathbb{R}^d : |\mathbf{x}-\mathbf{y}| \le r\}$. Show that $\mathbb{R}^d \backslash C$ is open in $\mathbb{R}^d$, and hence C is closed in $\mathbb{R}^d$.

In this course, it suffices to verify whether a set is open or closed by visual inspection. In a point-set topology course, you will learn more tools and techniques to prove a certain set is open and/or closed.

Remark 2.16. The null set $\emptyset$ is both open and closed in $\mathbb{R}^d$. Check up a youtube video titled "Hitler learns topology" (Profanity warning). □

Definition 2.17 (Bounded Sets)**.** A subset B of $\mathbb{R}^d$ is **bounded** if there exists a ball $B_r^d(\mathbf{0})$ such that $B \subset B_r^d(\mathbf{0})$. Therefore, $\mathbb{R}^d$ is both open and closed in $\mathbb{R}^d$ as well. □

Definition 2.18 (Local Lipschitz Continuity)**.** Let $\Omega \subset \mathbb{R}^d$ be an open domain and I is a time interval. A function $\mathbf{F}(\mathbf{x}, t) : \Omega \times I \to \mathbb{R}^m$ is said to be *locally Lipschitz continuous* on $\Omega \times I$ if for every point $\mathbf{x}_0 \in \Omega$, there exists a ball $B_r(\mathbf{x}_0) \subset \Omega$ such that $\mathbf{F}$ is Lipschitz continuous on $B_r(\mathbf{x}_0) \times I$.

Similarly, a time-independent function $\mathbf{G} : \Omega \to \mathbb{R}^m$ is *locally Lipschitz continuous* on Ω if for every point $\mathbf{x}_0 \in \Omega$, there exists a ball $B_r(\mathbf{x}_0) \subset \Omega$ such that $\mathbf{G}$ is Lipschitz continuous on $B_r(\mathbf{x}_0)$.

Recall that the motivations for introducing Lipschitz continuity is that $\mathbf{F}$ being Lipschitz continuous, as we will see later, allows us to justify some of the steps we 'cheated' in page 41. However, it is a very strong assumption to require a function being Lipschitz continuous on the whole $\mathbb{R}^d$ as we have seen in some of the examples. Even a simple function like $F(x) = x^2$ is not Lipschitz continuous on $\mathbb{R}$. Yet, it is locally Lipschitz continuous on $\mathbb{R}$ because for any $x_0 \in \mathbb{R}$, the function $F(x) = x^2$ is Lipschitz continuous on any ball centered at x_0 with any finite radius, say $(x_0 - 1, x_0 + 1)$.

In order to establish the convergence of the Picard's iteration sequence, it is in fact good enough to have local Lipschitz continuity. Locally Lipschitz continuous functions are very common as all C^1 functions, to be defined below, satisfy this condition.

Definition 2.19 (C^1-Functions)**.** Let Ω be an open domain in $\mathbb{R}^d$ and I is time interval. A function $\mathbf{F}(\mathbf{x}, t) : \Omega \times I \to \mathbb{R}^m$ is said to be C^1 on $\Omega \times I$ if all its first partial derivatives $\frac{\partial F_i}{\partial x_j}$ and $\frac{\partial F_i}{\partial t}$, for any $1 \leq i \leq m$ and $1 \leq j \leq d$, exist and are continuous on $\Omega \times I$.

Similarly, a time-independent function $\mathbf{G} : \Omega \to \mathbb{R}^m$ is said to be C^1 on Ω if all its first partial derivatives $\frac{\partial G_i}{\partial x_j}$, $1 \leq i \leq m$ and $1 \leq j \leq d$, exist and are continuous on Ω. □

Remark 2.20. A function can be C^1 on a smaller domain but not on a larger one. However, if the domain Ω of the function is clearly indicated, one may simply say the function is C^1 to mean that it is C^1 on Ω. □

Theorem 2.21 (C^1 implies Local Lipschitz Continuity)**.** *Let Ω be an open convex domain in $\mathbb{R}^d$ and I is a closed and bounded time interval. If a function $\mathbf{F}(\mathbf{x}, t) : \Omega \times I \to \mathbb{R}^m$ is C^1 on $\Omega \times I$, then $\mathbf{F}$ must be locally Lipschitz continuous on $\Omega \times I$.*

Similarly, if a time-independent function $\mathbf{G} : \Omega \to \mathbb{R}^m$ is C^1 on Ω, then $\mathbf{G}$ must be locally Lipschitz continuous on Ω.

Proof of Theorem 2.21. We will only prove the time-dependent case since the other case is similar. Suppose $\mathbf{F}(\mathbf{x}, t)$ is C^1 on $\Omega \times I$, then its first partial derivatives are all continuous on $\Omega \times I$. At any point $\mathbf{x}_0 \in \Omega$, since Ω is an open domain, one can choose a small ball $B_r(\mathbf{x}_0) \subset \Omega$.

Denote $\overline{B_{r/2}(\mathbf{x}_0)}$ the closed ball with radius $r/2$ centered at $\mathbf{x}_0$. Clearly

$$B_{r/2}(\mathbf{x}_0) \subset \overline{B_{r/2}(\mathbf{x}_0)} \subset B_r(\mathbf{x}_0) \subset \Omega.$$

The set $\overline{B_{r/2}(\mathbf{x}_0)} \times I$ is a closed and bounded (that's why we need I to be closed and bounded!) and therefore the extreme value theorem asserts that all continuous functions defined on this set must be bounded. In particular, the first partial derivatives $\frac{\partial F_i}{\partial x_j}$ are all bounded on $\overline{B_{r/2}(\mathbf{x}_0)} \times I$. Therefore, by Theorem 2.13, $\mathbf{F}$ is Lipschitz continuous on $\overline{B_{r/2}(\mathbf{x}_0)} \times I$, and on the smaller set $B_{r/2}(\mathbf{x}_0) \times I$ as well. Since the argument holds for every $\mathbf{x}_0 \in \Omega$, the function $\mathbf{F}$ is locally Lipschitz continuous on $\Omega \times I$. □

Example 2.9. The function $\mathbf{F}(\mathbf{x}) = \begin{bmatrix} \ln x_1 \\ x_2^2 + x_1 \end{bmatrix}$ is C^1 on $(0, \infty) \times \mathbb{R}$ because all the first partials derivatives (listed below) are continuous on this domain:

$$\frac{\partial F_1}{\partial x_1} = \frac{1}{x_1} \qquad \frac{\partial F_1}{\partial x_2} = 0$$
$$\frac{\partial F_2}{\partial x_2} = 1 \qquad \frac{\partial F_2}{\partial x_2} = 2x_2$$

Therefore it is also locally Lipschitz continuous on $(0, \infty) \times \mathbb{R}$. However, it is not Lipschitz continuous on $(0, \infty) \times \mathbb{R}$. □

Exercise 2.15. Let $F(\mathbf{x}) : B_1(\mathbf{0}) \to \mathbb{R}$ be the function defined by:

$$F(x_1, x_2) = \frac{1}{1 - x_1^2 - x_2^2}.$$

Is F Lipschitz continuous on $B_1(\mathbf{0})$? Is F locally Lipschitz continuous on $B_1(\mathbf{0})$?

Exercise 2.16. Let $F(x) := \sqrt{x}$.

(a) Is F differentiable at $x = 0$?

(b) Is F C^1 on $[0, 1]$?

(c) Is F C^1 on $(0, 1]$?

(d) Is F Lipschitz continuous on $[0, \infty)$?

(e) Is F locally Lipschitz continuous on $[0, \infty)$?

(f) Is F Lipschitz continuous on $(0, \infty)$?

(g) Is F locally Lipschitz continuous on $(0, \infty)$?

2.4. Picard-Lindelöf's Existence Theorem

After imposing the (local) Lipschitz continuity condition on $\mathbf{F}(x, y)$, we are now ready to state and give a complete proof of the existence theorem due to Picard and Lindelöf. Let's first recap the ingredients of proving the existence theorems. We begin with a domain $\Omega \subset \mathbb{R}^d$, a point $\mathbf{x}_0 \in \Omega$ and an IVP:

$$\mathbf{x}' = \mathbf{F}(\mathbf{x}, t), \quad \mathbf{x}(0) = \mathbf{x}_0.$$

We assume, as always, that the vector field $\mathbf{F}$ is continuous, then by Proposition 2.3 the IVP is equivalent to the integral equation (2.4):

$$\mathbf{x}(t) = \mathbf{x}_0 + \int_0^t \mathbf{F}(\mathbf{x}(s),\, s)\, ds.$$

In order to show the integral equation has a solution, we mimic the iteration examples discussed in Section 2.1 and defined the Picard's iteration sequence:

$$\begin{aligned} \mathbf{x}_0(t) &= \mathbf{x}_0 \\ \mathbf{x}_n(t) &= \mathbf{x}_0 + \int_0^t \mathbf{F}(\mathbf{x}_{n-1}(s),\, s)\, ds \qquad\qquad \text{for any } n \geq 1. \end{aligned}$$

We will resolve the following two analytic issues:

(1) show that $\mathbf{x}_n$ converges uniformly on some interval $(-\varepsilon, \varepsilon)$ as $n \to \infty$; and

(2) show that $\mathbf{F}(\mathbf{x}_n, \cdot)$ converges uniformly on the interval as well, so that we can perform the following interchanging of the limit and integral signs:

$$\lim_{n\to\infty} \int_0^t \mathbf{F}(\mathbf{x}_{n-1}(s),\, s)\, ds = \int_0^t \lim_{n\to\infty} \mathbf{F}(\mathbf{x}_{n-1}(s),\, s)\, ds.$$

After they are resolved, the limit function $\mathbf{x}_\infty$ will become a solution to the integral equation, and by continuity of the limit function $\mathbf{x}_\infty$, it is also a solution to the IVP. Hence, the existence of a solution, at least for a short time, is established.

Now we state the existence theorem:

Theorem 2.22 (Picard-Lindelöf's Existence Theorem)**.** *Let Ω be an open domain in $\mathbb{R}^d$, $\mathbf{x}_0$ be a point in Ω, and $I = [-T, T]$ be a closed and bounded time interval. Suppose $\mathbf{F}(\mathbf{x}, t) : \Omega \times I \to \mathbb{R}^d$ is a vector field which is locally Lipschitz continuous on $\Omega \times I$ (see Definition 2.18), then the initial-value problem*

$$\mathbf{x}' = \mathbf{F}(\mathbf{x}, t), \quad \mathbf{x}(0) = \mathbf{x}_0$$

has a solution $\mathbf{x}(t)$ defined on an interval $[-\varepsilon, \varepsilon] \subset I$ for some $\varepsilon > 0$.

Similarly, for an autonomous system with a vector field $\mathbf{G} : \Omega \to \mathbb{R}^d$ which is locally Lipschitz continuous on Ω, the IVP

$$\mathbf{x}' = \mathbf{G}(\mathbf{x}), \quad \mathbf{x}(0) = \mathbf{x}_0$$

has a solution defined on an interval $[-\varepsilon', \varepsilon']$ for some $\varepsilon' > 0$.

Remark 2.23. By Theorem 2.21, any C^1 vector field must be locally Lipschitz continuous. Therefore, Theorem 2.22 applies to all C^1 vector fields. Many examples we have seen so far are C^1 on their domain. □

Like in the previous chapter, we only present the proof of the non-autonomous case, since the proof for the autonomous case is the same, *mutatis mutandis*. Here is the outline of the whole proof:

(1) Since $\mathbf{F}$ is locally Lipschitz continuous, there exists a ball $B_r(\mathbf{x}_0)$ such that $\mathbf{F}$ is Lipschitz continuous on $B_r(\mathbf{x}_0) \times I$.

(2) Then, we show that when restricted to a small time interval $[-\varepsilon, \varepsilon]$, the Picard's iteration sequence $\mathbf{x}_n(t)$ will all lie inside this ball $B_r(\mathbf{x}_0)$. See Lemma 2.24. It will allow us to apply the Lipschitz continuity of $\mathbf{F}$ on the term $|\mathbf{F}(\mathbf{x}_n(s), s) - \mathbf{F}(\mathbf{x}_{n-1}(s), s)|$ that will appear in the next step.

(3) Next we show that the adjacent terms $\mathbf{x}_n$ and $\mathbf{x}_{n-1}$ of the Picard's iteration sequence will get closer and closer to each other as n increases. See Lemma 2.25.

(4) As in Problem 2.1, when the adjacent terms $\mathbf{x}_n$ and $\mathbf{x}_{n-1}$ get closer at suitable rate, one can use absolute convergence and telescoping method on $\sum_n |\mathbf{x}_n - \mathbf{x}_{n-1}|$ to show the convergence of $\mathbf{x}_n$. See Lemma 2.25.

(5) Finally, we resolve the two analytic issues mentioned before, and complete the proof of the theorem.

From now on until Theorem 2.22 is proved, the domain Ω, the point $\mathbf{x}_0$, the $I = [-T, T]$ and the vector field $\mathbf{F}$ are all defined as in the statement of Theorem 2.22. The sequence $\mathbf{x}_n$ denotes the Picard's iteration sequence associated to the IVP.

Furthermore, we denote $B_r(\mathbf{x}_0)$ to be the ball in Ω such that $\mathbf{F}$ is Lipschitz continuous on $B_r(\mathbf{x}_0) \times I$. By shrinking the radius of the ball if necessary, we also assume that the closed ball $\overline{B_r(\mathbf{x}_0)} \subset \Omega$.

Lemma 2.24. *Assume the hypotheses of Theorem 2.22. Then, there exists $\varepsilon > 0$ such that $[-\varepsilon, \varepsilon] \in I$ and $\mathbf{x}_n(t) \in B_r(\mathbf{x}_0)$ for all $n \geq 0$ and $t \in [-\varepsilon, \varepsilon]$. In other words, the trajectory of the Picard's iteration sequence must stay inside the ball $B_r(\mathbf{x}_0)$ during the time $t \in [-\varepsilon, \varepsilon]$.*

Proof. Since $\mathbf{F}$ is locally Lipschitz continuous on Ω, it is continuous on Ω *afortiori*. By the extreme-value theorem, $|\mathbf{F}|$ is bounded on the closed and bounded set $\overline{B_r(\mathbf{x}_0)} \times [-T, T]$.

Denote $M := \max\{|\mathbf{F}(\mathbf{x}, t) : (\mathbf{x}, t)| \in \overline{B_r(\mathbf{x}_0)} \times [-T, T]\}$ which is finite. To show $\mathbf{x}_n(t)$ all lie in the ball $B_r(\mathbf{x}_0)$, we first investigate the first few terms:

Obviously, $\mathbf{x}_0(t) = \mathbf{x}_0 \in B_r(\mathbf{x}_0)$ at all time. Since

$$\mathbf{x}_1(t) = \mathbf{x}_0 + \int_0^t \mathbf{F}(\mathbf{x}_0(s),\, s)\, ds,$$

we have:

$$\begin{aligned}|\mathbf{x}_1(t) - \mathbf{x}_0| &= \left|\int_0^t \mathbf{F}(\mathbf{x}_0(s),\, s)\, ds\right| \\ &\leq \int_0^t |\mathbf{F}(\mathbf{x}_0(s),\, s)|\, ds \\ &\leq \int_0^t M ds = M|t|.\end{aligned}$$

Therefore, if one chooses ε such that $M\varepsilon < r$ (and $[-\varepsilon, \varepsilon] \subset I$), then

$$|\mathbf{x}_1(t) - \mathbf{x}_0| \leq M|t| < r \text{ for any } t \in [-\varepsilon, \varepsilon].$$

In other words, $\mathbf{x}_1(t) \in B_r(\mathbf{x}_0)$ for any $t \in [-\varepsilon, \varepsilon]$. We claim that for this choice of ε, i.e. $\varepsilon < \min\{\frac{r}{M}, T\}$, we have for any $n \geq 1$,

$$\mathbf{x}_n(t) \in B_r(\mathbf{x}_0) \quad \text{for any } t \in [-\varepsilon, \varepsilon].$$

We have already proved this is true for $n = 1$. Assume it is true when $n = k - 1$, i.e.

$$|\mathbf{x}_{k-1}(t) - \mathbf{x}_0| < r \quad \text{for any } t \in [-\varepsilon, \varepsilon].$$

Then when $n = k$, we consider:

$$\begin{aligned}
\mathbf{x}_k(t) &= \mathbf{x}_0 + \int_0^t \mathbf{F}(\mathbf{x}_{k-1}(s),\, s)\, ds \\
|\mathbf{x}_k(t) - \mathbf{x}_0| &= \left| \int_0^t \mathbf{F}(\mathbf{x}_{k-1}(s),\, s)\, ds \right| \\
&\leq \int_0^t |\mathbf{F}(\mathbf{x}_{k-1}(s),\, s)|\; ds \\
&\leq \int_0^t M ds = M|t| \\
&\leq M\varepsilon && \text{(since } t \in [-\varepsilon, \varepsilon]\text{)} \\
&< M \cdot \frac{r}{M} = r && \text{(since } \varepsilon < r/M\text{)}
\end{aligned}$$

Therefore, $\mathbf{x}_k(t) \in B_r(\mathbf{x}_0)$ when $t \in [-\varepsilon, \varepsilon]$. By induction, $\mathbf{x}_n(t) \in B_r(\mathbf{x}_0)$ for any $t \in [-\varepsilon, \varepsilon]$ and $n \geq 1$. □

The next lemma shows that the sequence $\mathbf{x}_n(t)$ are getting closer and closer to each other as n becomes large.

Lemma 2.25. *Assume the hypotheses of Theorem 2.22. As a consequence of Lemma 2.24, there exists $\varepsilon > 0$ such that for all $n \geq 1$, the Picard's iteration sequence $\mathbf{x}_n(t) \in B_r(\mathbf{x}_0)$ when $t \in [-\varepsilon, \varepsilon]$. Then, for all $t \in [-\varepsilon, \varepsilon]$ and $n \geq 1$, we have:*

$$|\mathbf{x}_n(t) - \mathbf{x}_{n-1}(t)| \leq \frac{KL^{n-1}|t|^n}{n!} \tag{2.8}$$

for some constants $K > 0$ and $L > 0$.

Proof. We prove only the case $t > 0$, since the proof is similar for $t < 0$. First we estimate:

$$\begin{aligned}
|\mathbf{x}_1(t) - \mathbf{x}_0(t)| &= \left| \mathbf{x}_0 + \int_0^t \mathbf{F}(\mathbf{x}_0(s),\, s)\, ds - \mathbf{x}_0 \right| \\
&= \left| \int_0^t \mathbf{F}(\mathbf{x}_0(s),\, s)\, ds \right| \\
&\leq \int_0^t |\mathbf{F}(\mathbf{x}_0(s)\, s)|\, ds.
\end{aligned}$$

Define $K := \max\{|\mathbf{F}(\mathbf{x}_0, t)| : t \in [-\varepsilon, \varepsilon]\}$ which is finite by extreme-value theorem. Then we have $|\mathbf{x}_1(t) - \mathbf{x}_0(t)| \leq K|t|$. It verifies that (2.8) holds for $n = 1$. We let L be a Lipschitz constant of $\mathbf{F}$ on $B_r(\mathbf{x}_0) \times [-\varepsilon, \varepsilon]$, i.e. L is a constant such that for any $\mathbf{y}$, $\mathbf{z} \in B_r(\mathbf{x}_0)$ and $t \in [-\varepsilon, \varepsilon]$, we have:

$$|\mathbf{F}(\mathbf{y}, t) - \mathbf{F}(\mathbf{y}, t)| \leq L\, |\mathbf{y} - \mathbf{z}|\,.$$

Next we see if (2.8) holds when $n = 2$. Consider:

$$\begin{aligned}
|\mathbf{x}_2(t) - \mathbf{x}_1(t)| &= \left| \int_0^t \mathbf{F}(\mathbf{x}_1(s), s) - \mathbf{F}(\mathbf{x}_0(s), s) ds \right| && \text{(by definitions of } \mathbf{x}_2 \text{ and } \mathbf{x}_1\text{)} \\
&\leq \int_0^t |\mathbf{F}(\mathbf{x}_1(s), s) - \mathbf{F}(\mathbf{x}_0(s), s)| ds \\
&\leq \int_0^t L|\mathbf{x}_1(s) - \mathbf{x}_0(s)| ds && \text{(Lipschitz continuity)} \\
&\leq \int_0^t KLs ds && \text{(using (2.8) for } n = 0\text{)} \\
&= \frac{KLt^2}{2},
\end{aligned}$$

which verifies that (2.8) holds for $n = 2$. Note that we have implicitly used Lemma 2.24 in the third inequality above so as to guarantee that $\mathbf{x}_1(s)$ lies in the ball $B_r(\mathbf{x}_0)$ on which $\mathbf{F}$ is Lipschitz continuous.

We are going to prove (2.8) by induction. Suppose it holds when $n = k$, i.e.

$$|\mathbf{x}_k(t) - \mathbf{x}_{k-1}(t)| \leq \frac{KL^{k-1}t^k}{k!} \quad \text{for any } t \in [-\varepsilon, \varepsilon].$$

Then for $n = k + 1$, we have:

$$\begin{aligned}
|\mathbf{x}_{k+1}(t) - \mathbf{x}_k(t)| &= \left| \int_0^t \mathbf{F}(\mathbf{x}_k(s), s) - \mathbf{F}(\mathbf{x}_{k-1}(s), s) \, ds \right| \\
&\leq \int_0^t |\mathbf{F}(\mathbf{x}_k(s), s) - \mathbf{F}(\mathbf{x}_{k-1}(s), s)| \, ds \\
&\leq \int_0^t L \, |\mathbf{x}_k(s) - \mathbf{x}_{k-1}(s)| \, ds \\
&\leq \int_0^t L \cdot \frac{KL^{k-1}s^k}{k!} \, ds \\
&= \frac{KL^k t^{k+1}}{(k+1)!},
\end{aligned}$$

and so (2.8) holds when $n = k + 1$. By induction, (2.8) holds for any $n \geq 1$. □

The above lemma proves that the terms in the Picard's iteration sequence are getting closer and closer as $n \to \infty$. By mimicking the argument in the 'map' problem in Section 2.1, one can show $\mathbf{x}_n(t)$ converges. In fact, the estimates proved in Lemma 2.25 assert that this convergence is uniform on $[-\varepsilon, \varepsilon]$. Precisely, we have:

Lemma 2.26. *Assume all hypothese of Theorem 2.22. The Picard's iteration sequence $\mathbf{x}_n$ converges uniformly on $[-\varepsilon, \varepsilon]$ to a limit function as $n \to \infty$. Here $[-\varepsilon, \varepsilon]$ is the interval obtained in Lemma 2.24.*

Proof. By Lemma 2.25, we proved (2.8) for any $n \geq 1$:

$$|\mathbf{x}_n(t) - \mathbf{x}_{n-1}(t)| \leq \frac{KL^{n-1}|t|^n}{n!} \leq \frac{KL^{n-1}\varepsilon^n}{n!}, \quad t \in [-\varepsilon, \varepsilon],$$

and so $\|\mathbf{x}_n - \mathbf{x}_{n-1}\|_\infty \leq \frac{KL^{n-1}\varepsilon^n}{n!}$. Here the L^∞-norm is taken over the interval $[-\varepsilon, \varepsilon]$.

By ratio test, $\sum_{n=1}^\infty \frac{KL^{n-1}\varepsilon^n}{n!}$ converges, so the Weierstrass's M-test (applied with $\mathbf{a}_n = \mathbf{x}_n - \mathbf{x}_{n-1}$) shows $\sum_{n=1}^\infty (\mathbf{x}_n - \mathbf{x}_{n-1})$ converges uniformly on $[-\varepsilon, \varepsilon]$.

Note that:

$$\begin{aligned}\mathbf{x}_N &= \mathbf{x}_0 + (\mathbf{x}_1 - \mathbf{x}_0) + \ldots + (\mathbf{x}_N - \mathbf{x}_{N-1}) \\ &= \mathbf{x}_0 + \sum_{n=1}^{N} (\mathbf{x}_n - \mathbf{x}_{n-1}).\end{aligned}$$

Since $\sum_{n=1}^{N}(\mathbf{x}_n - \mathbf{x}_{n-1}) \to \sum_{n=1}^{\infty}(\mathbf{x}_n - \mathbf{x}_{n-1})$ uniformly on $[-\varepsilon, \varepsilon]$ as $N \to \infty$, the sequence $\mathbf{x}_N$ converges uniformly on $[-\varepsilon, \varepsilon]$ to $\mathbf{x}_0 + \sum_{n=1}^{\infty}(\mathbf{x}_n - \mathbf{x}_{n-1})$ as $N \to \infty$. □

Lemma 2.27. *Assume all hypotheses of Theorem 2.22. Denote $\mathbf{F}(\mathbf{x}_n, \cdot) : [-\varepsilon, \varepsilon] \to \mathbb{R}^d$ be a function that takes $t \in [-\varepsilon, \varepsilon]$ to $\mathbf{F}(\mathbf{x}_n(t), t)$. Then, $\mathbf{F}(\mathbf{x}_n, \cdot)$ converges uniformly on $[-\varepsilon, \varepsilon]$ to $\mathbf{F}(\mathbf{x}_\infty, \cdot)$, where $\mathbf{x}_\infty$ is the uniform convergence limit of $\mathbf{x}_n$ as guaranteed by Lemma 2.26.*

Proof. As before, let L be a Lipschitz constant of $\mathbf{F}$ is on $B_r(\mathbf{x}_0) \times [-\varepsilon, \varepsilon]$, i.e. L is a constant such that

$$|\mathbf{F}(\mathbf{y}, t) - \mathbf{F}(\mathbf{z}, t)| \leq L\,|\mathbf{y} - \mathbf{z}|\,, \quad \text{for any } \mathbf{y},\, \mathbf{z} \in B_r(\mathbf{x}_0) \text{ and } t \in [-\varepsilon, \varepsilon].$$

Take an arbitrary $t \in [-\varepsilon, \varepsilon]$, and subsitute $\mathbf{y} = \mathbf{x}_n(t)$ and $\mathbf{z} = \mathbf{x}_\infty(t)$, we have:

$$|\mathbf{F}(\mathbf{x}_n(t), t) - \mathbf{F}(\mathbf{x}_\infty(t), t)| \leq L\,|\mathbf{x}_n(t) - \mathbf{x}_\infty(t)| \quad \text{for any } t \in [-\varepsilon, \varepsilon].$$

Taking the maximum over $t \in [-\varepsilon, \varepsilon]$ on both sides of the inequality, we have:

$$\|\mathbf{F}(\mathbf{x}_n, \cdot) - \mathbf{F}(\mathbf{x}_\infty, \cdot)\|_\infty \leq L\,\|\mathbf{x}_n - \mathbf{x}_\infty\|_\infty\,.$$

Since $\mathbf{x} \to \mathbf{x}_\infty$ uniformly on $[-\varepsilon, \varepsilon]$ by Lemma 2.26, we have $\|\mathbf{x}_n - \mathbf{x}_\infty\|_\infty$ as $n \to \infty$. By squeezing principle, it implies $\|\mathbf{F}(\mathbf{x}_n, \cdot) - \mathbf{F}(\mathbf{x}_\infty, \cdot)\|_\infty \to 0$ and therefore $\mathbf{F}(\mathbf{x}_n, \cdot)$ converges uniformly on $[-\varepsilon, \varepsilon]$ to $\mathbf{F}(\mathbf{x}_\infty, \cdot)$ as $n \to \infty$. □

Lemma 2.27 allows us to apply Theorem A.11 for switching the limit and integral signs. Finally, we can make use of all these lemmas proven to establish the existence theorem:

Proof of Theorem 2.22. Consider the definition of the Picard's iteration sequence:

$$\mathbf{x}_n(t) = \mathbf{x}_0 + \int_0^t \mathbf{F}(\mathbf{x}_{n-1}(s),\, s)\, ds.$$

Restrict $t \in [-\varepsilon, \varepsilon]$. Let $n \to \infty$ on both sides:

$$\begin{aligned}\lim_{n\to\infty} \mathbf{x}_n(t) &= \mathbf{x}_0 + \lim_{n\to\infty} \int_0^t \mathbf{F}(\mathbf{x}_{n-1}(s),\, s)\, ds \\ \mathbf{x}_\infty(t) &= \mathbf{x}_0 + \int_0^t \lim_{n\to\infty} \mathbf{F}(\mathbf{x}_{n-1}(s),\, s)\, ds \quad \text{(by Lemma 2.27 and Theorem A.11)} \\ \mathbf{x}_\infty(t) &= \mathbf{x}_0 + \int_0^t \mathbf{F}(\mathbf{x}_\infty(s),\, s)\, ds \qquad \text{(by continuity of } \mathbf{F})\end{aligned}$$

Therefore, $\mathbf{x}_\infty(t)$ solves the integral equation (2.4). As $\mathbf{x}_n \to \mathbf{x}_\infty$ uniformly on $[-\varepsilon, \varepsilon]$, the limit function $\mathbf{x}_\infty$ must be continuous by Theorem A.11. Therefore, by Proposition 2.3, $\mathbf{x}_\infty(t)$ solves the IVP:

$$\mathbf{x}' = \mathbf{F}(\mathbf{x}, t), \quad \mathbf{x}(0) = \mathbf{x}_0$$

on $t \in [-\varepsilon, \varepsilon]$, completing the proof. □

Remark 2.28. We will prove uniqueness of solution of this IVP in Section 2.7 using the Grönwall's Inequality. □

Remark 2.29. It is also possible to prove existence of the IVP by just assuming $\mathbf{F}$ is continuous (while Theorem 2.22 requires local Lipschitz continuity). The proof, which is more complicated in terms of analysis, will be presented in Section 2.8 using Arzela-Ascoli's Theorem. This existence result, commonly called the Peano's Existence Theorem, do not require Lipschitz continuity but uniqueness is not guaranteed. □

2.4.0.1. Existence time interval. From the proof of Theorem 2.22, it is possible to estimate the width ε of the time interval on which existence of solutions is guaranteed. Precisely, if $|\mathbf{F}|$ is bounded by M on $\mathbf{B}_r(\mathbf{x}_0) \times [-T, T]$, then ε can be taken to be $\min\{\frac{r}{M}, T\}$. For an autonomous IVP, the ε' can be simply taken to be $\frac{r}{M}$. However, readers should be caution that the time interval $[-\varepsilon, \varepsilon]$ is one that we can *guarantee* the IVP has a solution. It is possible that the solution can be defined beyond this time interval. For instance, consider the one-dimensional IVP:

$$x' = \underbrace{x^2}_{F(x)} \qquad x(0) = \underbrace{1}_{x_0}.$$

Then on the ball $B_r(x_0) = (1-r, 1+r)$ with $|F(x)| = x^2$ is bounded by $(1+r)^2 =: M$, and so one can take $\varepsilon' = \frac{r}{M} = \frac{r}{(1+r)^2}$.

Solving this IVP, one should get $x(t) = \frac{1}{1-t}$. Therefore, the solution exists for $t \in (-\infty, 1)$. However, no matter what $r > 0$ we pick, it is impossible for $\varepsilon' = \frac{r}{(1+r)^2}$ to be 1 since it is at most $\frac{1}{4}$. To summarize, Theorem 2.22 gives us a time interval on which a solution must exist, but it fails to tell us what is the maximal time interval for a solution to be defined. In fact, it is usually impossible to find the exact maximal existence time since most solutions of ODE systems cannot be written down explicitly. However, we will see in the Section 2.5 that if $T_{\max} > 0$ is the maximal existence time, i.e. the solution $\mathbf{x}(t)$ to an IVP exists on $t \in [0, T_{\max})$ but does not exist pass $T_{\max}$, then one can assert what could happen on the *spatial position* of the trajectory $\mathbf{x}(t)$ when $t \to T_{\max}^-$.

Theorem 2.22 requires the vector field $\mathbf{F}(\mathbf{x}, t)$ to be locally Lipschitz continuous. However, if the vector field is *globally* Lipschitz continuous, i.e. $\mathbf{F}$ is Lipschitz continuous on the whole $\mathbb{R}^d \times (-\infty, \infty)$, or for autonomous case Lipschitz continuous on the whole $\mathbb{R}^d$, then one can actually prove long-time existence, i.e. the solution $\mathbf{x}(t)$ obtained in the proof is defined for all $t \in (-\infty, \infty)$.

Exercise 2.17. Prove the following global existence theorem: Given $\mathbf{x}_0 \in \mathbb{R}^d$ and suppose $\mathbf{F}(\mathbf{x}, t) : \mathbb{R}^d \times (-\infty, \infty)$ is Lipschitz continuous on the whole $\mathbb{R}^d \times (-\infty, \infty)$, then the IVP:

$$\mathbf{x}' = \mathbf{F}(\mathbf{x}, t), \quad \mathbf{x}(0) = \mathbf{x}_0$$

has a *global* solution $\mathbf{x}(t)$ defined on $t \in (-\infty, \infty)$.

You should modify the proof of Theorem 2.22. First go through and understand the whole local existence proof, then write up a complete coherent proof for the global existence theorem. Some part(s) in the local existence proof can be omitted, while some part(s) need to be modified. As a hint, we are now dealing with an infinite time interval. You will see that the K defined in the proof of Lemma 2.25 may be infinite if one replace $[-\varepsilon, \varepsilon]$ by $(-\infty, \infty)$. To overcome this issue, you may first fix an arbitrary $T > 0$, and try to first show that the Picard's iteration sequence converges uniformly on $[-T, T]$ to a solution defined on $[-T, T]$. Since T can be taken arbitrarily large, the solution extends to $(-\infty, \infty)$.

Example 2.10. Let's take a look at several examples about finding the existence time interval $[-\varepsilon, \varepsilon]$ as best as we can. Readers should be reminded that the intervals found below may not be the maximal time interval.

(1) Consider the IVP:

$$x_1' = \sin x_2 + t \qquad\qquad x_1(0) = 1$$
$$x_2' = \cos x_1 + t^2 \qquad\qquad x_2(0) = 2$$

In this case, $\mathbf{F}(\mathbf{x}, t) = \begin{bmatrix} \sin x_2 + t \\ \cos x_1 + t^2 \end{bmatrix}$. Its first partial derivatives are:

$$\frac{\partial F_1}{\partial x_1} = 0 \qquad\qquad \frac{\partial F_1}{\partial x_2} = \cos x_2$$
$$\frac{\partial F_2}{\partial x_1} = -\sin x_1 \qquad\qquad \frac{\partial F_2}{\partial x_2} = 0$$

which are all bounded on $\mathbb{R}^2 \times (-\infty, \infty)$. Therefore, $\mathbf{F}$ is Lipschitz continuous on $\mathbb{R}^2 \times (-\infty, \infty)$. By Exercise 2.17, the IVP has a solution $\mathbf{x}(t)$ defined for all $t \in (-\infty, \infty)$.

(2) Consider the one dimensional autonomous IVP:

$$x' = e^x, \quad x(0) = 1.$$

Take $F(x) = e^x$, then $\frac{\partial F}{\partial x} = e^x$ which is bounded on every bounded $[a, b] \subset \mathbb{R}$. Take an arbitrary $r > 0$, and consider the ball $B_r(1) = (1 - r, 1 + r)$. Then for $x \in B_r(1)$ we have:

$$|F(x)| = e^x \leq e^{1+r} =: M.$$

The local existence theorem asserts that a solution exists on $t \in [-\frac{r}{e^{1+r}}, \frac{r}{e^{1+r}}]$, which is the largest when $r = 1$ by elementary calculus. Therefore, the IVP has a solution defined on $[-e^{-2}, e^{-2}]$, but keep in mind that the solution can extend beyond this time interval. In fact, this IVP can be solved explicitly by separation of variables. The solution is given by:

$$x(t) = 1 - \ln(1 - et)$$

which is defined on $t \in (-\infty, e^{-1})$.

(3) Consider the one dimensional non-autonomous IVP:

$$x' = t^2 + x^2, \quad x(0) = 0.$$

In this IVP, $F(x, t) = t^2 + x^2$ which is defined on $\mathbb{R} \times (-\infty, \infty)$. Also, $\frac{\partial F}{\partial x} = 2x$ which is bounded on $[a, b] \times [-T, T]$ for any finite a, b and T. Therefore, $F(x, t)$ is Lipschitz continuous on any such subset of $\mathbb{R} \times (-\infty, \infty)$, but is not globally Lipschitz continuous.

Take an arbitrary $r > 0$ and consider the ball $B_r(0) = (-r, r)$. Then for $(x, t) \in B_r(0) \times [-T, T]$, we have

$$|F(x, t)| = t^2 + x^2 \leq T^2 + r^2 =: M.$$

Theorem 2.22 asserts that solution $x(t)$ exists on $t \in [-\frac{r}{T^2+r^2}, \frac{r}{T^2+r^2}] \cap [-T, T]$. From calculus, $\min\{T, \frac{r}{T^2+r^2}\}$ is the largest possible when $r = T = \sqrt{2}/2$. Therefore, one can guarantee that a solution exists for $t \in [-\sqrt{2}/2, \sqrt{2}/2]$.

□

Exercise 2.18. For each of the following IVPs, show that a solution exists using Theorem 2.22, and determine the largest existence time τ as guaranteed by the theorem (which is not necessarily the actual maximal existence time $T_{\max}$ of the solution). If it is possible to find the solution of the IVP, compare τ and $T_{\max}$ of the solution.

(a) $x' = \sin(\sin x), \quad x(0) = 1$

(b) $x' = 1 + \tan x, \quad x(0) = 0$

(c) $x' = 1 + x^2, \quad x(0) = 0$

(d) $x' = t + x^3, \quad x(0) = 1$

2.5. Finite-Time Singularity

Although it is not always possible to tell exactly what *time* a solution to an IVP ceases to exist, there is something we can say about *where* the trajectory is at the time the solution ceases to exist.

Definition 2.30 (Finite-Time Singularity). A solution $\mathbf{x}(t)$ to an ODE system $\mathbf{x}' = \mathbf{F}(\mathbf{x}, t)$ is said to have a forward **finite-time singularity** at $T_{\max} < \infty$ if the solution $\mathbf{x}(t)$ exists on $[0, T_{\max})$ but not on $[0, T_{\max} + \delta)$ for any $\delta > 0$. Similarly, $\mathbf{x}(t)$ is said to have a backward finite-time singularity at $T_{\min} > -\infty$ if the solution $\mathbf{x}(t)$ exists on $(T_{\min}, 0]$ but not on $(T_{\min} - \delta, 0]$ for any $\delta > 0$. The description *backward* or *forward* can be omitted when it is clear from the context.

The following discussion will be mostly about forward finite-time singularity, i.e. as $t \to T_{\max}^-$. Similarly results hold for backward finite-time singularity. Furthermore, to avoid some complications with the time interval I, let's focus only on autonomous systems which will be sufficient for the rest of the course.

Theorem 2.31. *Let $\Omega \subset \mathbb{R}^d$ be an open domain. Suppose the vector field $\mathbf{F}(\mathbf{x}) : \Omega \to \mathbb{R}^d$ is locally Lipschitz continuous on Ω, and $\mathbf{x}(t)$ is a solution to $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ with (forward) finite-time singularity at $T_{\max} < \infty$. Then, if $\mathbf{x}(t) \to \mathbf{y}$ as $t \to T_{\max}^-$ for some $\mathbf{y} \in \mathbb{R}^d$, we must have $\mathbf{y} \in \mathbb{R}^d \backslash \Omega$.*

Remark 2.32. Heuristically, Theorem 2.31 asserts that if a finite-time singularity occurs at some $\mathbf{y}$, then $\mathbf{y}$ must be outside of Ω (most likely on $\partial\Omega$). In other words, as long as the solution $\mathbf{x}(t)$ stays in the interior of Ω, then solution must continue for a while before it becomes singular. $\square$

Example 2.11. An easy example to illustrate this scenario is the one-dimensional ODE: $x' = -\frac{1}{x}$ on $x \in (0, \infty)$. Using separation of variables, the solution with initial data $x(0) = x_0 > 0$ is given by: $x(t) = \sqrt{x_0^2 - 2t}$, which cannot be continued when t approach $\frac{x_0^2}{2} =: T_{\max}$. We can see that $x(t) \to 0$ as $t \to T_{\max}$, and 0 is a boundary point of the domain $(0, \infty)$. $\square$

In order to establish this result, we need two lemmas, one about the solution (Lemma 2.33) and another about the *topology of open sets*. The first lemma shows that for any closed and bounded set K in Ω, one can always find a time $\tau < T_{\max}$ such that $\mathbf{x}(\tau)$ is outside K. Since the lemma holds for any closed and bounded set K inside Ω, one may take larger and larger closed and bounded set K to cover up the domain Ω, then the second lemma will 'push' the limit $\mathbf{y}$ out of the interior of Ω.

Lemma 2.33. *Assume Ω, $\mathbf{F}$ and $\mathbf{x}(t) : [0, T_{\max}) \to \mathbb{R}^d$ as in Theorem 2.31. Let K be a closed and bounded subset of Ω, then for any $\varepsilon > 0$ there exists a time $\tau \in [T_{\max} - \varepsilon, T_{\max})$ such that $\mathbf{x}(\tau) \notin K$.*

Proof. We prove by contradiction: assume $\mathbf{x}(t) \in K$ for all $t \in [T_{\max} - \varepsilon, T_{\max})$. By the Bolzano-Weierstrass's Theorem, one can find a sequence $t_n \to T_{\max}$ as $n \to \infty$ such that $\mathbf{x}(t_n) \in K$ converges to a limit $\mathbf{z}_0$ as $n \to \infty$. By the closedness of K, the limit $\mathbf{z}_0$ must be in K.

Since K is closed and bounded and $\mathbf{F}$ is *afortiori* continuous by the local Lipschitz continuity assumption, the extreme value theorem shows $|\mathbf{F}|$ must be bounded by some constant M on K. The solution $\mathbf{x}(t)$ satisfies the integral equation:

$$\mathbf{x}(t) = \mathbf{x}(\xi) + \int_{\xi}^{t} \mathbf{F}(\mathbf{x}(s))ds.$$

for any ξ, $t \in [0, T_{\max})$. By our assumption, $\mathbf{x}(s) \in K$ for any $s \in [T_{\max} - \varepsilon, T_{\max})$ and therefore, we have:

$$\begin{aligned}
|\mathbf{x}(t) - \mathbf{x}(\xi)| &\leq \left| \int_{\xi}^{t} \mathbf{F}(\mathbf{x}(s))ds \right| \\
&\leq \int_{\xi}^{t} |\mathbf{F}(\mathbf{x}(s))| \, ds \\
&\leq \int_{\xi}^{t} M ds = M|t - \xi|.
\end{aligned}$$

Substitute $t = t_n$, we get: $|\mathbf{x}(t_n) - \mathbf{x}(\xi)| \leq M|t_n - \xi|$ for any $n \geq 1$. Letting $n \to \infty$, we have $|\mathbf{z}_0 - \mathbf{x}(\xi)| \leq M|T_{\max} - \xi|$ since $\mathbf{x}(t_n) \to \mathbf{z}_0$.

Let $\xi \to T_{\max}^-$ we have $M\,|T_{\max} - \xi| \to 0$, and by applying the squeezing principle, we have $|\mathbf{z}_0 - \mathbf{x}(\xi)| \to 0$ as $\xi \to T_{\max}^-$, or in other words, $\mathbf{x}(\xi) \to \mathbf{z}_0$. Since the limit exists, one can decree $\mathbf{x}(T_{\max}) := \mathbf{z}_0$.

Since $\mathbf{z}_0 \in K \subset \Omega$, by the existence theorem, there exists a solution $\mathbf{z}(t) : [0, \delta) \to \mathbb{R}^d$ of the IVP:

$$\mathbf{z}' = \mathbf{F}(\mathbf{z}), \quad \mathbf{z}(0) = \mathbf{x}(T_{\max}).$$

Then, by 'gluing' the solution $\mathbf{z}(t)$ to $\mathbf{x}(t)$, one can extend the solution $\mathbf{x}(t)$ beyond $T_{\max}$. Precisely, define:

$$\mathbf{x}(t) := \mathbf{z}(t - T_{\max}) \quad \text{for } t \in [T_{\max}, T_{\max} + \delta),$$

then $\mathbf{x}' = \frac{d}{dt}\mathbf{z}(t - T_{\max}) = \mathbf{F}(\mathbf{z}(t - T_{\max})) = \mathbf{F}(\mathbf{x})$ for $t \in [T_{\max}, T_{\max} + \delta)$, and $\mathbf{x}(T_{\max}) = \mathbf{z}(0) = \mathbf{x}(T_{\max})$. Therefore, the solution $\mathbf{x}(t)$ is defined on $[0, T_{\max} + \delta)$, which contradicts the maximality of $T_{\max}$. Hence, our assumption that $\mathbf{x}(t) \in K$ for all $t \in [T_{\max} - \varepsilon, T_{\max})$ does not hold, and so there exists $\tau \in [T_{\max} - \varepsilon, T_{\max})$ such that $\mathbf{x}(\tau) \notin K$. □

Example 2.12. To illustrate the use of Lemma 2.33 and to give some motivations of the next lemma, let's look at the following system again: $x' = -\frac{1}{x}$. The function $F(x) = -\frac{1}{x}$ is defined on $\Omega := (0, \infty)$. Suppose $x(t)$ is a solution with a finite-time singularity at $T_{\max} < \infty$. For each n, consider the closed and bounded set $K_n = [\frac{1}{n}, n] \subset \Omega$. Lemma 2.33 asserts that there exists $\tau_n \in [T_{\max} - \frac{1}{n}, T_{\max})$ such that $x(\tau_n) \notin K_n$, i.e. $x(\tau_n) \in (-\infty, \frac{1}{n}) \cup (n, \infty)$. If $x(\tau_n)$ converges to a limit $y \in \mathbb{R}$ as $n \to \infty$, then by $x(\tau_n) < \frac{1}{n}$ or $x(\tau_n) > n$, one must have $y \leq \lim_{n\to\infty} \frac{1}{n} = 0$, which is not in Ω. □

In the above example, we showed that if we can find a 'good' sequence of closed and bounded set K_n to cover the open set Ω, then using Lemma 2.33 one can find τ_n such that $\mathbf{x}(\tau_n) \notin K_n$ for each n, then one can show that the limit of $\mathbf{x}(\tau_n)$ as $n \to \infty$, must be beyond any of the $K_n' s$, and hence must be outside Ω.

Lemma 2.34 (Exhaustion Lemma of Open Sets)**.** *Let Ω be an open set in $\mathbb{R}^d$. Then, there exists a sequence of closed and bounded sets $K_i \subset \Omega$ with the following properties:*

(1) For any i, we have $K_i \subset K_{i+1}^\circ$ where K_{i+i}° denotes the interior of K_{i+1}. We call such a sequence of sets $\{K_i\}$ strictly increasing; and

(2) these closed and bounded sets K_i cover the open set Ω, i.e. $\Omega = \bigcup_{i=1}^{\infty} K_i$.

Furthermore, these two properties implies

$$\bigcap_{i=1}^{\infty} \overline{\mathbb{R}^d \backslash K_i} = \mathbb{R}^d \backslash \Omega.$$

This sequence of closed and bounded sets K_i is called an exhaustion of Ω.

Proof. Define

$$K_i := \overline{B_i(\mathbf{0})} \bigcap \left(\mathbb{R}^d \backslash \bigcup_{\mathbf{x} \in \mathbb{R}^d \backslash \Omega} B_{1/i}(\mathbf{x}) \right)$$

and we claim that it is the sequence $\{K_i\}$ we need. Clearly each K_i is closed and bounded.

Recall from basic point-set topology that $(A \cap B)^\circ = A^\circ \cap B^\circ$, so we have

$$K_{i+1}^\circ = B_{i+1}(\mathbf{0}) \cap \left(\mathbb{R}^d \backslash \bigcup_{\mathbf{x} \in \mathbb{R}^d \backslash \Omega} B_{1/i}(\mathbf{x}) \right)^\circ = B_{i+1}(\mathbf{0}) \cap \left(\mathbb{R}^d \backslash \bigcup_{\mathbf{x} \in \mathbb{R}^d \backslash \Omega} \overline{B_{1/(i+1)}(\mathbf{x})} \right).$$

Hence, $K_i \subset K_{i+1}^\circ$ follows directly from $\overline{B_i(\mathbf{0})} \subset B_{i+1}(\mathbf{0})$ and $\overline{B_{1/(i+1)}(\mathbf{x})} \subset B_{1/i}(\mathbf{x})$. This proves (1).

For (2), we first rewrite using $\cup_i (A_i \cap B_i) \subset (\cup_i A_i) \cap (\cup_i B_i)$ and $\cup_i (A \backslash B_i) = A \backslash \cap_i B_i$:

$$\begin{aligned}
\bigcup_{i=1}^{\infty} K_i &\subset \bigcup_{i=1}^{\infty} \overline{B_i(\mathbf{0})} \bigcap \bigcup_{i=1}^{\infty} \left(\mathbb{R}^d \backslash \bigcup_{\mathbf{x} \in \mathbb{R}^d \backslash \Omega} B_{1/i}(\mathbf{x}) \right) \\
&= \mathbb{R}^d \bigcap \left(\mathbb{R}^d \backslash \bigcap_{i=1}^{\infty} \bigcup_{\mathbf{x} \in \mathbb{R}^d \backslash \Omega} B_{1/i}(\mathbf{x}) \right) \\
&\subset \mathbb{R}^d \backslash \bigcap_{i=1}^{\infty} \bigcup_{\mathbf{x} \in \mathbb{R}^d \backslash \Omega} \{\mathbf{x}\} && (\text{as } \mathbf{x} \in B_{1/i}(\mathbf{x})) \\
&= \mathbb{R}^d \backslash (\mathbb{R}^d \backslash \Omega) = \Omega.
\end{aligned}$$

To prove $\Omega \subset \bigcup_{i=1}^{\infty} K_i$, we need to use the condition that Ω is open. For any $\mathbf{y} \in \Omega$, by open-ness there exists a large i_0 such that $B_{1/i_0}(\mathbf{y}) \subset \Omega$ and $|\mathbf{y}| \leq i_0$. We claim that $\mathbf{y} \in K_{i_0}$. To see this, clearly we have $\mathbf{y} \in \overline{B_{i_0}(\mathbf{0})}$. Suppose on the contrary that $\mathbf{y} \notin K_{i_0}$, then one must have

$$\mathbf{y} \notin \mathbb{R}^d \backslash \bigcup_{\mathbf{x} \in \mathbb{R}^d \backslash \Omega} B_{1/i_0}(\mathbf{x}), \text{ or equivalently } \mathbf{y} \in B_{1/i_0}(\mathbf{x}) \text{ for some } \mathbf{x} \notin \Omega.$$

However, this means $|\mathbf{x} - \mathbf{y}| < \frac{1}{i_0}$ for some $\mathbf{x} \notin \Omega$, but that would contradict to the fact that

$$\mathbf{x} \in B_{1/i_0}(\mathbf{y}) \subset \Omega.$$

Therefore, $\mathbf{y} \in K_{i_0} \subset \bigcup_{i=1}^{\infty} K_i$, proving $\Omega \subset \bigcup_{i=1}^{\infty} K_i$ and hence (2).

To prove the last statement, we consider

$$\mathbb{R}^d\backslash\Omega = \mathbb{R}^d\backslash\bigcup_{i=1}^{\infty} K_i = \bigcap_{i=1}^{\infty}\mathbb{R}^d\backslash K_i \subset \bigcap_{i=1}^{\infty}\overline{\mathbb{R}^d\backslash K_i}.$$

Moreover, we also have

$$\overline{\mathbb{R}^d\backslash K_i} \subset \overline{\mathbb{R}^d\backslash K_i^\circ} = \mathbb{R}^d\backslash K_i^\circ \subset \mathbb{R}^d\backslash K_{i-1},$$

so we get

$$\bigcap_{i=1}^{\infty}\overline{\mathbb{R}^d\backslash K_i} = \bigcap_{i=2}^{\infty}\overline{\mathbb{R}^d\backslash K_i} \subset \bigcap_{i=2}^{\infty}(\mathbb{R}^d\backslash K_{i-1}) = \mathbb{R}^d\backslash\bigcup_{i=2}^{\infty} K_{i-1} = \mathbb{R}^d\backslash\Omega.$$

In the first equality we have used the fact that $K_1 \subset K_2$. □

Using these two lemmas, we are ready to give the proof of theorem:

Proof of Theorem 2.31. Using Lemma 2.34, there exists a sequence of closed and bounded sets $K_i \subset \Omega$ with the properties stated in the lemma.

For each i, Lemma 2.33 shows there exists $\tau_i \in [T_{\max} - \frac{1}{i}, T_{\max})$ such that $\mathbf{x}(\tau_i) \notin K_i$. In other words $\mathbf{x}(\tau_i) \in \mathbb{R}^d\backslash K_i$.

Since K_i is a strictly increasing sequence of sets, $K_i \subset K_{i+1}^\circ \subset K_{i+1}$, we have $\mathbb{R}^d\backslash K_{i+1} \subset \mathbb{R}^d\backslash K_i$ for any i. Therefore, for any $n \geq i$, we have

$$\mathbf{x}(\tau_n) \in \mathbb{R}^d\backslash K_n \subset \mathbb{R}^d\backslash K_i.$$

Given that $\mathbf{x}(t) \to \mathbf{y}$ as $t \to T_{\max}^-$, we know $\mathbf{x}(\tau_n) \to \mathbf{y}$ as $n \to \infty$. Therefore,

$$\mathbf{y} = \lim_{n\to\infty}\mathbf{x}(\tau_n) \in \overline{\mathbb{R}^d\backslash K_i}$$

for each i, i.e. $\mathbf{y} \in \bigcap_{i=1}^{\infty}\overline{\mathbb{R}^d\backslash K_i}$. By the property of K_i's, we have $\mathbf{y} \in \mathbb{R}^d\backslash\Omega$. □

Corollary 2.35. *Suppose $\mathbf{F}$ is locally Lipschitz continuous on $\mathbb{R}^d$, and $\mathbf{x}(t)$ is a solution of the system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ with a finite-time singularity at $T_{\max}$. Then, $\mathbf{x}(t)$ must blow-up at $T_{\max}$, i.e. $\mathbf{x}(t) \to \infty$ as $t \to T_{\max}^-$.*

Moreover, if $\mathbf{y}(t)$ is a solution to the same system and $\mathbf{y}(t)$ is bounded for $t > 0$, then $\mathbf{y}(t)$ must be defined for all $t \in [0, \infty)$.

Exercise 2.19. By modifying Lemma 2.33, and using Lemma 2.34, prove a 'sequential' version of Theorem 2.31: Assume Ω, $\mathbf{F}$ as in Theorem 2.31. Now suppose $\mathbf{x}(t)$ is a solution to $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ with (forward) finite-time singularity at $T_{\max} < \infty$, and there exists a sequence of times $t_n \to T_{\max}$ as $n \to \infty$ such that $\mathbf{x}(t_n) \to \mathbf{y}$ as $n \to \infty$, then we must have $\mathbf{y} \in \mathbb{R}^d\backslash\Omega$.

To summarize, if a vector field $\mathbf{F}$ is locally Lipschitz continuous on an open domain Ω, and $\mathbf{x}(t)$ is a solution to $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ with a finite-time singularity, then one of the following cases will happen:

(1) As $t \to T_{\max}^-$, the solution $\mathbf{x}(t)$ becomes unbounded.

(2) If $\mathbf{x}(t)$ is bounded on $t \in [0, T_{\max})$, then by the Bolzano-Weierstrass's Theorem, there is a sequence $t_n \to T_{\max}$ as $n \to \infty$ such that $\mathbf{x}(t_n)$ converges to some limit $\mathbf{y} \in \mathbb{R}^d$. By Exercise 2.19 whose proof is a slight modification of Theorem 2.31, this limit $\mathbf{y}$ must be outside Ω.

Furthermore, if $\Omega = \mathbb{R}^d$, only the first case can happen.

2.6. Grönwall's Inequality

In this section, we introduce a fundamental inequality in the theory of ODEs, the Grönwall's Inequality. It will be used to prove the uniqueness theorem in the next section.

The one dimensional IVP: $x' = Lx, \quad x(t_0) = C$, where $L > 0$ is a constant, can be written as an integral equation

$$x(t) = C + \int_{t_0}^{t} Lx(s)ds.$$

Remark 2.36. We have been using the term IVP to refer an ODE system with initial condition at $t = 0$. However, for the purpose of future discussion, we need to allow a condition given at a time t_0 other than 0. We will still call it an 'initial condition' even though t_0 may not be 0, and a problem of this sort will still be called an IVP. □

Now given a continuous function $y(t)$ that satisfies the following integral inequality,

$$y(t) \leq C + \int_{t_0}^{t} Ly(s)ds.$$

We ask what one can say about the relation between $x(t)$, a solution to the integral equation, and $y(t)$, a function satisfying the integral inequality? The Grönwall's Inequality tells us that one of them will act as a *barrier* of the other.

Theorem 2.37 (Grönwall's Inequality). *Let $L > 0$ be a positive constant, and C be any real constant. Suppose $y(t)$ is a continuous functions defined on a time interval I containing t_0 and it satisfies the following integral inequality:*

$$y(t) \leq C + \int_{t_0}^{t} Ly(s)ds \tag{2.9}$$

for all $t \in I$. Then, we have $y(t) \leq Ce^{L|t-t_0|}$ for all $t \in I$.

Proof. We prove only the case when $t > t_0$. The case $t < t_0$ is similar and left as an exercise for readers. The proof of the inequality is by the *barrier* method, which is very common in the studies of PDEs too. When $t > t_0$, $x(t) := Ce^{L(t-t_0)}$ is a solution to the integral equation

$$x(t) = C + \int_{t_0}^{t} Lx(s)ds. \tag{2.10}$$

Graphically, the desired result $y(t) \leq x(t)$ for all $t \geq t_0$ (and $t \in I$) means that the graph of $y(t)$ always stays below that of $x(t)$. We will prove that it is true by contradiction. Heuristically, we assume there is a first time $t_1 > t_0$ at which $y(t)$ overtakes $x(t)$, we will try that show that it will contradict (2.9) and (2.10). However, there is a subtle issue for this idea but it can be resolved by a common ODE/PDE technique so-called the 'ε-trick'. We will explain why it is needed in Remark 2.38 after the proof.

Given any $\varepsilon > 0$, (2.9) implies that for all $t \in I$, the following holds:

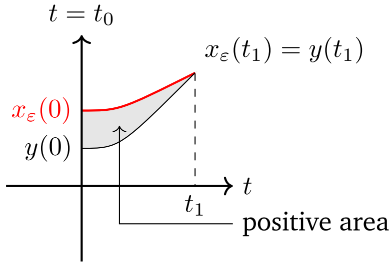
$$y(t) < (C + \varepsilon) + \int_{t_0}^{t} Ly(s)ds. \tag{2.11}$$

Let $x_\varepsilon(t) = (C + \varepsilon)e^{L(t-t_0)}$, which is clearly a solution to the integral equation:

$$x_\varepsilon(t) = (C + \varepsilon) + \int_{t_0}^{t} Lx_\varepsilon(s)ds. \tag{2.12}$$

Initially at $t = t_0$, we see $y(t_0) < C + \varepsilon$ and $x_\varepsilon(t_0) = C + \varepsilon$, and so y is strictly below x_ε. We claim that $y(t)$ stays below $x_\varepsilon(t)$ at all time $t \in I$.

We prove by contradiction: assume that there is a time $t_1 > t_0$ such that $y(t_1) = x_\varepsilon(t_1)$, and that t_1 is the first such time, meaning that $y(t) < x_\varepsilon(t)$ for $t \in [t_0, t_1)$, while $y(t_1) = x_\varepsilon(t_1)$. Then, the area bounded by $y(t)$ and $x_\varepsilon(t)$ for $t \in [t_0, t_1]$ must be positive, i.e.



$$\int_{t_0}^{t_1} (x_\varepsilon(s) - y(s))\, ds > 0. \tag{2.13}$$

However, by substituting $t = t_1$ into (2.11) and (2.12), followed by a subtraction (2.11) - (2.12), we get:

$$y(t_1) - x_\varepsilon(t_1) < \int_{t_0}^{t_1} L(y(s) - x_\varepsilon(s))ds.$$

Since $L > 0$ and $y(t_1) = x_\varepsilon(t_1)$, we have

$$0 < \int_{t_0}^{t_1} (y(s) - x_\varepsilon(s))ds.$$

which contradicts (2.13).

This proves $y(t) < x_\varepsilon(t) = (C + \varepsilon)e^{L(t-t_0)}$ for any $t \in [t_0, \infty) \cap I$. Since $\varepsilon > 0$ is arbitrarily small, letting $\varepsilon \to 0^+$ shows

$$y(t) \le \lim_{\varepsilon \to 0^+} (C + \varepsilon)e^{L(t-t_0)} = Ce^{L(t-t_0)}$$

for any $t \in [t_0, \infty) \cap I$. It completes the proof of the case $t > t_0$, the other case is left for readers. $\square$

Exercise 2.20. Complete the proof of Theorem 2.37 for the case $t < t_0$. As a hint, compare $y(t)$ with $z_\varepsilon(t) = (C + \varepsilon)e^{-L(t-t_0)}$ which solves the integral equation:

$$z_\varepsilon(t) = (C + \varepsilon) - \int_{t_0}^{t} Lz_\varepsilon(s)ds.$$

$\square$

Remark 2.38. We need to invoke the 'ε-trick' in the proof but not directly compare $y(t)$ and $x(t) = Ce^{L(t-t_0)}$ because we can guarantee only $y(t_0) \le x(t_0)$ but not $y(t_0) < x(t_0)$, so it may not be possible to produce a region of positive area that gives a result similar to (2.13). $\square$

Exercise 2.21. Let C be any real constant and $v(t) : (-\infty, \infty) \to \mathbb{R}$ is a continuous and **nonnegative** function. Suppose $u : [0, \alpha] \to \mathbb{R}$ is a continuous function that satisfies:

$$u(t) \le C + \int_0^t v(s)u(s)ds \quad \text{for all } t \in [0, \alpha].$$

Prove that:

$$u(t) \le C \exp\left(\int_0^t v(s)ds\right) \quad \text{for all } t \in [0, \alpha].$$

2.7. Uniqueness of Solutions

One important consequence of the Grönwall's Inequalty (Theorem 2.37) is the uniqueness theorem of ODEs with an initial condition. Recall in the linear case, uniqueness was established as a Corollary 1.31 to Theorem 1.29. We showed that if both $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$ solve a linear system $\mathbf{x}' = A\mathbf{x}$, then the following inequality holds for all $t \in (-\infty, \infty)$:

$$|\mathbf{x}_1(t) - \mathbf{x}_2(t)| \leq |\mathbf{x}_1(0) - \mathbf{x}_2(0)| \, e^{\|A\||t|}.$$

Consequently, if they have the same initial conditions, i.e. $\mathbf{x}_1(0) = \mathbf{x}_2(0)$, then this inequality implies $\mathbf{x}_1(t) = \mathbf{x}_2(t)$ for all time $t \in (-\infty, \infty)$. Furthermore, this continuous dependence inequality also gives an estimate of the two solutions to the same system but with different initial conditions.

Now we move on to the nonlinear (possibly non-autonomous) systems. We will establish a similar continuous dependence inequality for a system $\mathbf{x}' = \mathbf{F}(\mathbf{x}, t)$ where $\mathbf{F}$ is a vector field with sufficiently regularity, such as C^1, or more generally, locally Lipschitz continuous.

We first start with the Lipschitz continuous case:

Theorem 2.39 (Continuous Dependence Inequality for Nonlinear Systems)**.** *Let $\Omega \subset \mathbb{R}^d$ be an open domain and I be a time interval. Suppose $\mathbf{F}(\mathbf{x}, t) : \Omega \times I \to \mathbb{R}^d$ is a vector field which is Lipschitz continuous on $\Omega \times I$ with a Lipschitz constant L. If $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$, are both solutions to the system $\mathbf{x}' = \mathbf{F}(\mathbf{x}, t)$, and $\mathbf{x}_1(t),\ \mathbf{x}_2(t) \in \Omega$ for t in some interval $I' \subset I$, then we have:*

$$|\mathbf{x}_1(t) - \mathbf{x}_2(t)| \leq |\mathbf{x}_1(t_0) - \mathbf{x}_2(t_0)| \, e^{L|t-t_0|} \tag{2.14}$$

for any $t_0,\ t \in I'$.

Remark 2.40. In simpler terms, the inequality (2.14) holds as long as both solutions stay inside Ω. □

Remark 2.41. The linear case (Theorem 1.29) is a special case of Theorem 2.39 since for any square matrix A, the map $\mathbf{x} \mapsto A\mathbf{x}$ is Lipschitz continuous on $\mathbb{R}^d$ with a Lipschitz constant $\|A\|$. □

Remark 2.42. Similar result holds for autonomous systems $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ provided that $\mathbf{F}$ is Lipschitz continuous on Ω. The proof is the same. □

Proof of Theorem 2.39. Define $y(t) := |\mathbf{x}_1(t) - \mathbf{x}_2(t)|$ for $t \in I'$. We are going to show $y(t)$ satisfies the integral inequality (2.9) for some suitable constant C. Since both $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$ solve the system $\mathbf{x}' = \mathbf{F}(\mathbf{x}, t)$, they solve the following integral equations:

$$\mathbf{x}_1(t) = \mathbf{x}_1(t_0) + \int_{t_0}^{t} \mathbf{F}(\mathbf{x}_1(s),\ s)\, ds$$
$$\mathbf{x}_2(t) = \mathbf{x}_2(t_0) + \int_{t_0}^{t} \mathbf{F}(\mathbf{x}_2(s),\ s)\, ds$$

Subtracting the two integral equations, we have:

$$\mathbf{x}_1(t) - \mathbf{x}_2(t) = \mathbf{x}_1(t_0) - \mathbf{x}_2(t_0) + \int_{t_0}^{t} (\mathbf{F}(\mathbf{x}_1(s),\ s) - \mathbf{F}(\mathbf{x}_2(s),\ s))\, ds.$$

Therefore,

$$\begin{aligned}|\mathbf{x}_1(t) - \mathbf{x}_2(t)| &\leq |\mathbf{x}_1(t_0) - \mathbf{x}_2(t_0)| + \left|\int_{t_0}^{t} (\mathbf{F}(\mathbf{x}_1(s),\, s) - \mathbf{F}(\mathbf{x}_2(s),\, s))\, ds\right| \\ &\leq |\mathbf{x}_1(t_0) - \mathbf{x}_2(t_0)| + \int_{t_0}^{t} |\mathbf{F}(\mathbf{x}_1(s),\, s) - \mathbf{F}(\mathbf{x}_2(s),\, s)|\, ds \\ &\leq |\mathbf{x}_1(t_0) - \mathbf{x}_2(t_0)| + \int_{t_0}^{t} L\, |\mathbf{x}_1(s) - \mathbf{x}_2(s)|\, ds.\end{aligned}$$

Here we have used the fact that $\mathbf{F}$ is Lipschitz continuous with a Lipschitz constant L, and both $\mathbf{x}_1(t),\ \mathbf{x}_2(t) \in \Omega$ for $t \in I'$. Therefore, for any $t \in I'$, we have:

$$y(t) \leq \underbrace{|\mathbf{x}_1(t_0) - \mathbf{x}_2(t_0)|}_{=:C} + \int_{t_0}^{t} Ly(s)\, ds$$

for all $t \in I'$. By the Grönwall's Inequality (Theorem 2.37), we have:

$$y(t) \leq |\mathbf{x}_1(t_0) - \mathbf{x}_2(t_0)|\, e^{L|t-t_0|}$$

for all $t \in I'$, as desired. □

As in the linear case, if $\mathbf{x}_1(t_0) = \mathbf{x}_2(t_0)$, then we must have $\mathbf{x}_1(t) = \mathbf{x}_2(t)$ as long as they are in Ω on which $\mathbf{F}$ is Lipschitz continuous. Thus we have the following corollary:

Corollary 2.43 (Uniqueness Theorem: Lipschitz). *Suppose $\mathbf{F}(\mathbf{x}, t) : \Omega \times I \to \mathbb{R}^d$ be a vector field which is Lipschitz continuous on $\Omega \times I$ and $\mathbf{x}_0$ is a point in Ω. If $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$, defined on $t \in I' \subset I$ such that $\mathbf{x}_1(t),\ \mathbf{x}_2(t) \in \Omega$ for $t \in I'$, are both solutions to the IVP:*

$$\mathbf{x}' = \mathbf{F}(\mathbf{x}, t), \quad \mathbf{x}(t_0) = \mathbf{x}_0,$$

then we have $\mathbf{x}_1(t) = \mathbf{x}_2(t)$ for all $t \in I'$.

Remark 2.44. In simpler terms, the corollary asserts that the solution to an IVP is unique as long as the solution lies in Ω. □

Note that both Theorem 2.39 and Corollary 2.43 require $\mathbf{F}$ to be Lipschitz continuous on $\Omega \times I$ (or on Ω for autonomous systems). This condition may be quite restrictive because $\mathbf{F}$ is usually not Lipschitz continuous if Ω is taken to be $\mathbb{R}^d$. However, with a slightly extended argument, one can prove the IVP still has uniqueness if $\mathbf{F}$ is just assumed to be locally Lipschitz continuous.

The key idea is as follows: suppose $\mathbf{F}$ is locally Lipschitz continuous on $\Omega \times I$, and we have two solutions $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$ to an IVP:

$$\mathbf{x}' = \mathbf{F}(\mathbf{x}, t), \quad \mathbf{x}(0) = \mathbf{x}_0.$$

One can choose a small ball $B_r(\mathbf{x}_0)$ such that $\mathbf{F}$ is Lipschitz continuous on $B_r(\mathbf{x}_0) \times I$. Then apply Corollary 2.43 with $B_r(\mathbf{x}_0)$ in place of Ω, then the two solutions $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$ agree as long as they are inside the ball. At the time they leave the ball, we draw another ball on which the vector field is Lipschitz continuous, and then one can extend the uniqueness result for a little while. Heuristically, the uniqueness result can be extended to the whole Ω by successively covering the trajectory by these balls. We will give a proof for this in a more rigorous way:

Corollary 2.45 (Uniqueness Theorem: Locally Lipschitz). *Suppose $\mathbf{F}(\mathbf{x},t):\Omega\times I\to\mathbb{R}^d$ be a vector field which is locally Lipschitz continuous on $\Omega\times I$. Let $\mathbf{x}_0$ be a point in Ω. If $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$, defined on $t\in I'\subset I$ such that $\mathbf{x}_1(t),\ \mathbf{x}_2(t)\in\Omega$ for $t\in I'$, are both solutions to the IVP:*

$$\mathbf{x}'=\mathbf{F}(\mathbf{x},t),\quad \mathbf{x}(t_0)=\mathbf{x}_0,$$

then we have $\mathbf{x}_1(t)=\mathbf{x}_2(t)$ for all $t\in I'$.

Proof. We prove by contradiction. Initially, $\mathbf{x}_1(t_0)=\mathbf{x}_2(t_0)$. Assume $\mathbf{x}_1$ and $\mathbf{x}_2$ start branching out at a time $T\in I'$. Then, $\mathbf{x}_1(t)=\mathbf{x}_2(t)$ for any $t\in[t_0,T]$. Let $\mathbf{p}:=\mathbf{x}_1(T)$, there exists a ball $B_r(\mathbf{p})\subset\Omega$ such that $\mathbf{F}$ is Lipschitz continuous on $B_r(\mathbf{p})\times I$. However, since $\mathbf{x}_1(T)=\mathbf{x}_2(T)$, Corollary 2.43 asserts that $\mathbf{x}_1(t)=\mathbf{x}_2(t)$ for $t\in[T-\delta,T+\delta]$ for some small $\delta>0$ as long as the solutions $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$ stay inside the ball $B_r(\mathbf{p})$ for $t\in[T-\delta,T+\delta]$.

This shows $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$ agree at least on the time interval $[t_0,T+\delta]$, contradicting to the fact that T is the time they start branching out.

Therefore, there is no such $T\in I'$ and so $\mathbf{x}_1(t)=\mathbf{x}_2(t)$ for all $t\in I'$. Similar argument applies to show uniqueness for backward time. □

Recall in Theorem 2.21 asserts that a vector field $\mathbf{F}\in C^1(\Omega\times I)$, or $\mathbf{F}\in C^1(\Omega)$ on an open domain Ω must also be locally Lipschitz continuous. Therefore, Corollary 2.45 applies to C^1 vector fields.

The (local) Lipschitz continuous condition was used to establish the uniqueness of solutions. Without this condition, the solution may not be unique as demonstrated by the following counter-example:

Consider the IVP:

$$x'=x^{2/3},\quad x(0)=0.$$

While $x(0)\equiv 0$ is an obvious solution to the IVP, there is an infinite family of non-zero solutions $\{x_\alpha(t)\}_{\alpha>0}$ given by:

$$x_\alpha(t)=\begin{cases}0 & \text{if } t\in(-\infty,\alpha];\\ \frac{(t-\alpha)^3}{27} & \text{if } t\in(\alpha,\infty).\end{cases}$$

Exercise 2.22. Verify that for each $\alpha>0$, $x_\alpha(t)$ is a solution to the IVP

$$x'=x^{2/3},\quad x(0)=0.$$

The uniqueness theorem does not apply to this IVP because the initial condition is $x_0=0$. Take any open interval Ω in $\mathbb{R}$ that contains 0, the function $F(x)=x^{2/3}$ is not locally Lipschitz continuous on Ω because $F'(x)=\frac{2}{3}x^{-1/3}$ is not bounded as $x\to 0$.

However, the uniqueness theorem applies if we take $\Omega=(0,\infty)$, for instance. $F(x)=x^{2/3}$ is locally Lipschitz continuous on $(0,\infty)$. Take any $x_0\in(0,\infty)$ and then the IVP

$$x'=x^{2/3},\quad x(0)=x_0$$

has a unique solution as far as $x(t)\in(0,\infty)$. However, it is possible for $x(t)$ to branch out when it approaches 0 either in forward or backward time.

Exercise 2.23. Fix $\beta\in(0,1)$, show that the IVP:

$$x'=x^{\beta},\quad x(0)=0$$

has infinitely many solutions by explicitly constructing them.

The continuity dependence inequality (2.14) can be generalized to allow the two solutions which not only have different initial conditions but also satisfy two different systems. See the exercise below:

Exercise 2.24. Let Ω be an open domain in $\mathbb{R}^d$ and I be an interval. Suppose $\mathbf{F} : \Omega \times I \to \mathbb{R}^d$ and $\mathbf{G} : \Omega \times I \to \mathbb{R}^d$ are two continuous vector fields defined on Ω such that

$$|\mathbf{F}(\mathbf{z}, t) - \mathbf{G}(\mathbf{z}, t)| < \varepsilon$$

for any $(\mathbf{z}, t) \in \Omega \times I$. Suppose further that $\mathbf{F}$ is Lipschitz continuous on $\Omega \times I$ with a Lipschitz constant $L > 0$ (while $\mathbf{G}$ is not assumed to be Lipschitz continuous). Suppose $\mathbf{x}(t)$ and $\mathbf{y}(t)$ be solutions to the systems $\mathbf{x}' = \mathbf{F}(\mathbf{x}, t)$ and $\mathbf{y}' = \mathbf{G}(\mathbf{y}, t)$ respectively, show that:

$$|\mathbf{x}(t) - \mathbf{y}(t)| \leq |\mathbf{x}(t_0) - \mathbf{y}(t_0)|e^{L|t-t_0|} + \frac{\varepsilon}{L}(e^{L|t-t_0|} - 1) \tag{2.15}$$

for any t_0, $t \in I$ as long as solutions exist. [Hint: Write the systems as two integral equations, and use Grönwall's Inequality at some point.]

2.7.0.1. Flow of autonomous systems. For linear systems $\mathbf{x}' = A\mathbf{x}$, the flow $\Phi_t(\cdot) : \mathbb{R}^d \to \mathbb{R}^d$ is defined as a map whose input is a point $\mathbf{x}_0$ in $\mathbb{R}^d$ and output is a point reached by flowing along the phase portrait from $\mathbf{x}_0$ for t unit time (forward if $t > 0$ and backward if $t < 0$). In Section 1.3, we know that for linear systems the flow is given by $\Phi_t(\mathbf{x}_0) = e^{tA}\mathbf{x}_0$ where t is defined on $(-\infty, \infty)$ for every $\mathbf{x}_0$.

We are going to define a similar flow for nonlinear systems. As an introductory course, we focus on autonomous systems $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ only because the flow, as we will prove, will satisfies a nice property analogous to $\Phi_t \circ \Phi_s = \Phi_{t+s}$ for the linear case. However, a similar results does not hold if we define the flow for non-autonomous systems.

Let Ω be an open domain of $\mathbb{R}^d$, and $\mathbf{F}$ is locally Lipschitz continuous on Ω. The existence and uniqueness theorems we established earlier shows that the IVP

$$\mathbf{x}' = \mathbf{F}(\mathbf{x}), \quad \mathbf{x}(0) = \mathbf{x}_0$$

has a unique solution $\mathbf{x}(t)$ as long as it is inside Ω. This allows us to define the flow $\varphi_t(\mathbf{x}_0)$ to be the solution $\mathbf{x}(t)$. However, readers should be caution that the flow for a nonlinear system may not define for all t. In fact, different initial conditions may give different maximal existence times!

For example, consider the one dimensional system $x' = x^2$. If the initial condition is $x(0) = 0$, then $x(t) = 0$ is the solution which is defined for all $t \in (-\infty, \infty)$. However, if the initial condition is $x(0) = 1$, then $x(t) = \frac{1}{1-t}$ which is defined for $t \in (-\infty, 1)$ only.

As per the above discussion, the flow of a nonlinear system may not be defined on all of $\Omega \times (-\infty, \infty)$ but only on a subset of it. Here we denote the set for which a flow map of a nonlinear system can be defined by:

$$\Sigma(\mathbf{F}) := \bigcup_{\mathbf{x}_0 \in \Omega} \{\mathbf{x}_0\} \times I_{\mathbf{x}_0}$$

where $I_{\mathbf{x}_0}$ is the maximal time interval of the IVP $\mathbf{x}' = \mathbf{F}(\mathbf{x}), \quad \mathbf{x}(0) = \mathbf{x}_0$.

Remark 2.46. Note that $\Sigma(\mathbf{F})$ depends on the vector field $\mathbf{F} : \Omega \to \mathbb{R}^d$. If the vector field can be understood in the context, we can simply write Σ for the domain of flow. $\square$

Example 2.13. Consider the one-dimensional system $x' = x^2$, i.e. $F(x) = x^2$ which is locally Lipschitz continuous on $\mathbb{R}$. If the initial condition is $x(0) = 0$, then the solution to the IVP is $x(t) \equiv 0$, and so $I_0 = (-\infty, \infty)$. If $x_0 \neq 0$ and the initial condition is $x(0) = x_0$, then a simple separation of variables shows that the solution to the IVP is $x(t) = \frac{x_0}{1-x_0 t}$. Therefore,

$$I_{x_0} = \begin{cases} (x_0^{-1}, \infty) & \text{if } x_0 < 0; \\ (-\infty, x_0^{-1}) & \text{if } x_0 > 0. \end{cases}$$

To summarize, the domain of flow for this system is the open region bounded by the hyperbolas $xt = 1$ in the (t, x)-plane. □

Exercise 2.25. Find the domain of flow Σ of the one-dimensional system $x' = 1+x^2$. Sketch the domain of flow on the (t, x)-plane.

Definition 2.47 (Flow of Nonlinear Autonomous System)**.** Let Ω be an open domain of $\mathbb{R}^d$, and $\mathbf{F}(\mathbf{x}) : \Omega \to \mathbb{R}^d$ be a vector field which is locally Lipschitz continuous on Ω. The flow of the system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ is a map $\varphi^{\mathbf{F}}(\mathbf{x}, t) : \Sigma(\mathbf{F}) \to \Omega$ such that for each $\mathbf{x}_0 \in \Omega$, the curve $\varphi^{\mathbf{F}}(\mathbf{x}_0, t)$ is a solution to the IVP:

$$\mathbf{x}' = \mathbf{F}(\mathbf{x}), \quad \mathbf{x}(0) = \mathbf{x}_0.$$

Alternatively, the flow $\varphi^{\mathbf{F}}(\mathbf{x}_0, t)$ can be denoted by $\varphi_t^{\mathbf{F}}(\mathbf{x}_0)$. Furthermore, if the vector field $\mathbf{F}$ of a flow can be understood from the context, we can omit the superscript $\mathbf{F}$ and simply write $\varphi(\mathbf{x}_0, t)$ or $\varphi_t(\mathbf{x}_0)$.

Remark 2.48. Throughout the course, we will use Φ_t to denote the flow of a linear system, and φ_t for nonlinear systems. □

Remark 2.49. The flow φ_t is well-defined on Σ by the existence and uniqueness theorems (Theorem 2.22 and Corollary 2.45). □

Example 2.14. For ODE systems of more than one equations, the solution to an IVP is often difficult (if not impossible) to find. Therefore, it is almost impossible to write down an explicit domain of flow Σ and the flow φ_t of the system. However, it is sometimes possible to do so in one dimension:

(1) Consider the system $x' = x^2$ which C^1 on $\mathbb{R}$. With an initial condition $x(0) = x_0$, the solution is given by $x(t) = \frac{x_0}{1-x_0 t}$. The flow, on its domain of flow, is given by $\varphi_t(x_0) = \frac{x_0}{1-x_0 t}$.

(2) The system $x' = 1 + x^2$ is C^1 on $\mathbb{R}$, and the solution with initial condition $x(0) = x_0$ is given by $x(t) = \tan(t + \arctan(x_0))$. The flow is therefore $\varphi_t(x_0) = \tan(t + \arctan(x_0))$ wherever it is defined.

□

Exercise 2.26. Find the domain of flow Σ and the flow φ_t of the system $x' = x^3$.

2.7.0.2. Continuity of flow. The flow Φ_t of a linear system is a continuous function on $\mathbb{R}^d$ for any fixed $t \in (-\infty, \infty)$. It is a consequence of the continuous dependence inequality (Theorem 1.29). For nonlinear systems, since we also have a similar continuous dependence inequality (Theorem 2.39), one should expect φ_t is also continuous for each fixed t.

However, there is a subtle issue we need to resolve for nonlinear systems on an open set Ω, namely the domain of flow Σ may not be all of $\Omega \times (-\infty, \infty)$. Therefore, for each fixed t, the flow φ_t cannot be always regarded as a map from Ω to Ω. For instance, consider the ODE $x' = x^2$, with $\Omega = \mathbb{R}$, where the domain of flow is the open region bounded hyperbolas $tx = 1$. Therefore, if we fix $t = 1$, the flow φ_1 is only defined on $x_0 \in (-\infty, 1)$.

As per discussion of the previous paragraph, we will denote

$$\Sigma_t(\mathbf{F}) := \{\mathbf{x}_0 \in \Omega : (\mathbf{x}_0, t) \in \Sigma(\mathbf{F})\}$$

and call $\Sigma_t(\mathbf{F})$ the t-slice of $\Sigma(\mathbf{F})$. If the vector field $\mathbf{F}$ is clear from the context, it can be omitted in the notation and we can simply write Σ_t.

Remark 2.50. Note that the domain of flow $\Sigma(\mathbf{F})$ is a subset of $\Omega \times (-\infty, \infty)$ while the t-slice $\Sigma_t(\mathbf{F})$ is a subset of Ω. □

Example 2.15. Here is an examples of finding Σ_t of one-dimensional systems. In general, the t-slice of the domain of flow of a nonlinear system cannot be easily found.

The system $x' = x^2$ has domain of flow Σ equal to the open region bounded by hyperbolas $tx = 1$. In other words

$$\Sigma = \{(t, x) : tx < 1\}.$$

For each fixed t, the t-slice is given by:

$$\Sigma_t = \begin{cases} (t^{-1}, \infty) & \text{if } t < 0; \\ \mathbb{R} & \text{if } t = 0; \\ (-\infty, t^{-1}) & \text{if } t > 0. \end{cases}$$

□

Exercise 2.27. Find Σ_t for each fixed t of the system $x' = 1 + x^2$.

We are about to prove that the flow φ_t at each fixed time t is a continuous map. The key ingredient is to apply the continuous dependence inequality (Theorem 2.39) and rewrite it using the flow notations. However, the continuous dependence inequality requires the vector field $\mathbf{F}$ to be Lipschitz continuous on the whole domain, but we only assume local Lipschitz continuity here. Thanks to the Heine-Borel's Theorem, one can show local Lipschitz continuity on a **closed and bounded** set K implies Lipschitz continuity on K (proof left as an exercise for readers). We will invoke this result in the proof below.

Proposition 2.51 (Continuity of Flow)**.** *Suppose $\mathbf{F} : \Omega \to \mathbb{R}^d$ is a locally Lipschitz vector field on an open set $\Omega \subset \mathbb{R}^d$. Then, for each $t \in (-\infty, \infty)$, the t-slice $\Sigma_t(\mathbf{F})$ is an open set of Ω, and the flow $\varphi_t : \Sigma_t(\mathbf{F}) \to \mathbb{R}^d$ of the system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ is a continuous map.*

Proof. For each fixed $t > 0$ (similar for $t < 0$) and $\mathbf{x}_0 \in \Sigma_t(\mathbf{F})$, the trajectory of the flow $\{\varphi_s(\mathbf{x}_0)\}_{s\in[0,t]}$, as long as it stays in Ω, is a bounded since $t \mapsto \varphi_t(\mathbf{x}_0)$ is a continuous function and $[0, t]$ is closed and bounded. One can then find a bounded open set $\mathcal{O}$ containing the trajectory $\{\varphi_s(\mathbf{x}_0)\}_{s\in[0,t]}$, and that $\overline{\mathcal{O}} \subset \Omega$. Since $\mathbf{F}$ is locally Lipschitz continuous on Ω (and hence on $\overline{\mathcal{O}}$ as well), by the standard Heine-Borel's argument there exists a Lipschitz constant L such that $|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{y})| \leq L\,|\mathbf{x} - \mathbf{y}|$ for any $\mathbf{x}, \mathbf{y} \in \overline{\mathcal{O}}$. Using the continuous dependence inequality (Theorem 2.39) rewritten in flow notations, we have:

$$|\varphi_t(\mathbf{x}) - \varphi_t(\mathbf{x}_0)| \leq |\mathbf{x} - \mathbf{x}_0|\, e^{L|t|}$$

provided that $\{\varphi_s(\mathbf{x})\}_{s\in[0,t]}$ is in $\overline{\mathcal{O}}$.

The continuity of φ_t follows directly from the continuous dependence inequality (Theorem 2.39) which, using the notation of flows, is stated as $|\varphi_t(\mathbf{x}) - \varphi_t(\mathbf{x}_0)| \leq$

$|\mathbf{x}-\mathbf{x}_0|\, e^{L|t|}$. By letting $\mathbf{x} \to \mathbf{x}_0$ (with t fixed), the squeezing principle implies $\varphi_t(\mathbf{x}) \to \varphi_t(\mathbf{x}_0)$, which is exactly what is meant by φ_t being continuous at $\mathbf{x}_0$ for each fixed t.

The open-ness of $\Sigma_t(\mathbf{F})$ is also a consequence of the continuous dependence inequality, but in a slightly non-trivial way. Here we need to show that when $\mathbf{x}$ is sufficiently close to $\mathbf{x}_0$, the flow $\varphi_t(\mathbf{x})$ is defined at t. Since the trajectory from $\{\varphi_s(\mathbf{x}_0)\}_{s\in[0,t]}$ (keep in mind t is fixed) is bounded in $\mathcal{O}$, there exists $d > 0$ such that the distance from any point on the boundary $\partial\mathcal{O}$ is at least distance d from the trajectory. We claim if $|\mathbf{x}-\mathbf{x}_0| < \frac{d}{2}e^{-Lt}$, then $\varphi_t(\mathbf{x})$ must be defined at t. Suppose otherwise $\varphi_s(\mathbf{x})$ is only defined for $s \in [0, t')$ where $t' < t$. Then the solution $\varphi_s(\mathbf{x})$ encounter finite-time singularity and by Theorem 2.31, $\varphi_s(\mathbf{x})$ must go outside of Ω as $s \to t'$. There must be a time $\tau \in [0, t')$ at which $\varphi_\tau(\mathbf{x})$ leaves $\mathcal{O}$ the first time. Then $\varphi_\tau(\mathbf{x})$ is on the boundary of Ω and so is at least distance d from the trajectory $\{\varphi_s(\mathbf{x}_0)\}_{s\in[0,t]}$. Therefore, the continuous dependence inequality shows:

$$d \le |\varphi_\tau(\mathbf{x}) - \varphi_\tau(\mathbf{x}_0)| \le |\mathbf{x}-\mathbf{x}_0|\, e^{Lt} < \frac{d}{2}e^{-Lt}\cdot e^{Lt} = \frac{d}{2}$$

which is clearly a contradiction. Therefore, $\varphi_t(\mathbf{x})$ must be defined at t for $\mathbf{x} \in B_{\frac{d}{2}e^{-Lt}}(\mathbf{x}_0)$, proving that Σ_t is an open set. □

The continuity of φ_t, as we will see, will be crucial for studying stability in Chapter 3 and proving the Poincaré-Bendixson's Theorem in Chapter 4.

One can also prove that φ_t is C^k on Ω whenever $\mathbf{F}$ is C^k on Ω. The proof is more technical and hence omitted in the course. Interested reader may consult the lecture notes by Brendle for the case $k = 1$.

2.7.0.3. Autonomous versus non-autonomous systems. For linear systems $\mathbf{x}' = A\mathbf{x}$, the flow Φ_t has an intuitive property that $\Phi_t(\Phi_s(\mathbf{x}_0)) = \Phi_{t+s}(\mathbf{x}_0)$ for any $t,\ s \in (-\infty, \infty)$ and $\mathbf{x}_0 \in \mathbb{R}^d$. This turns out to be true for nonlinear **autonomous** systems too, i.e. $\varphi_t(\varphi_s(\mathbf{x}_0)) = \varphi_{t+s}(\mathbf{x}_0)$ provided that $(\mathbf{x}_0, t)$, $(\mathbf{x}_0, s)$ and $(\mathbf{x}_0, s+t)$ are all in Σ. However, it is important to keep in mind that it is in general not true if one defines the flow for non-autonomous systems in a similar fashion.

Proposition 2.52. *Let $\Omega \subset \mathbb{R}^d$ be an open domain. Suppose $\mathbf{F} : \Omega \to \mathbb{R}^d$ is locally Lipschitz continuous on Ω. Denote $\varphi(\mathbf{x}_0, t) : \Sigma(\mathbf{F}) \to \Omega$ the flow of the autonomous system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$. Then, for any t, s and $\mathbf{x}_0$ such that $(\mathbf{x}_0, t)$, $(\mathbf{x}_0,\ s)$, $(\mathbf{x}_0,\ t+s) \in \Sigma(\mathbf{F})$, we have:*

$$\varphi_t(\varphi_s(\mathbf{x}_0)) = \varphi_{t+s}(\mathbf{x}_0).$$

Proof. The proof follows mostly from the uniqueness theorem (Corollary 2.45). Regarding s as a constant and t as the parameter, we will show both $\varphi_t(\varphi_s(\mathbf{x}_0))$ and $\varphi_{t+s}(\mathbf{x}_0)$ are solutions to the same system with the same initial data at $t = 0$. As they are given to be inside Ω, it will then follow from the uniqueness theorem that they must be equal.

For any $\mathbf{y}_0 \in \Omega$, the $\varphi(\mathbf{y}_0, t)$ for $t \in I_{\mathbf{y}_0}$ is a solution to $\mathbf{x}' = \mathbf{F}(\mathbf{x})$. In other words, we have:

$$\frac{\partial \varphi_t(\mathbf{y}_0)}{\partial t} = \mathbf{F}(\varphi_t(\mathbf{y}_0)).$$

Using this, we can deduce:

$$\frac{\partial}{\partial t}\varphi_t(\varphi_s(\mathbf{x}_0)) = \mathbf{F}(\varphi_t(\varphi_s(\mathbf{x}_0))) \qquad \text{(substitute } \mathbf{y}_0 = \varphi_s(\mathbf{x}_0))$$

On the other hand, we have:

$$\begin{aligned}\frac{\partial}{\partial t}\varphi_{t+s}(\mathbf{x}_0) &= \frac{\partial\varphi_{t+s}(\mathbf{x}_0)}{\partial(t+s)}\cdot\frac{\partial(t+s)}{\partial t} && \text{(chain rule)}\\ &= \mathbf{F}(\varphi_{t+s}(\mathbf{x}_0)) && \text{(substitute } \mathbf{y}_0 = \varphi_{t+s}(\mathbf{x}_0))\end{aligned}$$

These show both $\varphi_t(\varphi_s(\mathbf{x}_0))$ and $\varphi_{t+s}(\mathbf{x}_0)$ satisfies the system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$. Initially at $t = 0$, we have:

$$\begin{aligned}[\varphi_t(\varphi_s(\mathbf{x}_0))]_{t=0} &= \varphi_s(\mathbf{x}_0) && \text{(using } \varphi_0 = \mathrm{id})\\ [\varphi_{t+s}(\mathbf{x}_0)]_{t=0} &= \varphi_s(\mathbf{x}_0)\end{aligned}$$

Therefore, they have the same initial data! By the uniqueness theorem, we have

$$\varphi_t(\varphi_s(\mathbf{x}_0)) = \varphi_{t+s}(\mathbf{x}_0),$$

as desired. □

Remark 2.53. Note that the same argument does not apply to non-autonomous systems $\mathbf{x}' = \mathbf{F}(\mathbf{x}, t)$ if φ_t is defined in a similar way. Since then

$$\frac{\partial}{\partial t}\varphi_t(\varphi_s(\mathbf{x}_0)) = \mathbf{F}(\varphi_t(\varphi_s(\mathbf{x}_0)), t)$$

but,

$$\frac{\partial}{\partial t}\varphi_{t+s}(\mathbf{x}_0) = \frac{\partial\varphi_{t+s}(\mathbf{x}_0)}{\partial(t+s)}\cdot\frac{\partial(t+s)}{\partial t} = \mathbf{F}(\varphi_{t+s}(\mathbf{x}_0), \underbrace{t+s}_{\text{not the same!}}).$$

Even for a simple non-autonomous system like $x' = 2t$, the solution with initial condition $x(0) = x_0$ is $x(t) = t^2 + x_0$. The "flow" is given as $\varphi_t(x_0) = t^2 + x_0$. It can be easily verified that:

$$\begin{aligned}\varphi_t(\varphi_s(x_0)) &= \varphi_t(s^2 + x_0)\\ &= t^2 + (s^2 + x_0) = t^2 + s^2 + x_0\\ \varphi_{t+s}(x_0) &= (t+s)^2 + x_0\\ &= t^2 + 2ts + s^2 + x_0.\end{aligned}$$

So $\varphi_t \circ \varphi_s \neq \varphi_{t+s}$! □

Exercise 2.28. Let φ_t be the flow of a C^1 vector field $\mathbf{F} : \mathbb{R}^d \to \mathbb{R}^d$, i.e. autonomous. Suppose $\varphi_t(\mathbf{x}_0)$ is defined for all $t \in [0, \infty)$ and as $t \to \infty$, we have $\varphi_t(\mathbf{x}_0) \to \mathbf{y} \in \mathbb{R}^d$. Show that $\mathbf{y}_0$ is an equilibrium point, i.e. $\varphi_s(\mathbf{y}_0) = \mathbf{y}_0$ for any $s \in (-\infty, \infty)$.

2.8. Peano's Existence Theorem

Lipschitz continuity is a fundamental assumption in all results we have discussed in this chapter. If the vector field $\mathbf{F}$ is not (local) Lipschitz continuous, we have seen that uniqueness is not guaranteed. A quick counterexample is

$$x' = x^{2/3}, \quad x(0) = 0.$$

The function $F(x) = x^{2/3}$ is not Lipschitz continuous on any domain containing 0, and we have seen there are infinitely many solutions to this IVP.

It is natural to ask whether we still have existence of solutions if $\mathbf{F}$ is not (local) Lipschitz continuous? The answer is positive, provided that $\mathbf{F}$ is continuous (which is not necessarily Lipschitz continuous). In this section, we will give a proof of existence of solutions by assuming $\mathbf{F}$ is continuous only. This proof is a bit more technical than the Picard-Lindelöf's Existence proof. It uses a famous theorem in analysis, the Arzelá-Ascoli's Theorem.

In order to state the theorem, we need to give a new definition.

Definition 2.54 (Equicontinuity). Let $I = [a, b]$ be a closed and bounded interval. Suppose $\mathbf{x}_n(t) : I \to \mathbb{R}^d$ is sequence of continuous functions defined on I. This sequence is said to be **equicontinuous** on I if for any given $\varepsilon > 0$, there exists a $\delta > 0$ which depends *only* on ε, such that whenever $|t - s| < \delta$ and $t,\ s \in I$, we have

$$|\mathbf{x}_n(t) - \mathbf{x}_n(s)| < \varepsilon$$

for every n.

Remark 2.55. The key difference between "all $\mathbf{x}_n(t)$'s are continuous on I" and "the sequence $\mathbf{x}_n(t)$ is equicontinuous on I" is about how δ depends on other quantities. We say that all $\mathbf{x}_n(t)$'s are continuous on I if for any $\varepsilon > 0$, any $t_0 \in I$ and any $n \in \mathbb{N}$, there exists a $\delta > 0$ which may **depend on** ε, t_0 **and** n, such that whenever $|t - t_0| < \delta$, we have $|\mathbf{x}_n(t) - \mathbf{x}_n(t_0)| < \varepsilon$. However, when we say the sequence $\mathbf{x}_n(t)$ is equicontinuous on I, this δ can only depend on ε. □

Example 2.16. Let $x_n(t) = \frac{t}{n}$ on $I = [0, 1]$. Then $\{x_n(t)\}$ is an equicontinuous sequence on I because for any $\varepsilon > 0$, one can find $\delta = \varepsilon > 0$ (of course depends only on ε), such that whenever $t,\ s \in [0, 1]$ and $|t - s| < \delta$, we have:

$$|x_n(t) - x_n(s)| = \left|\frac{t}{n} - \frac{s}{n}\right| = \frac{1}{n}|t - s| \le |t - s| < \delta = \varepsilon.$$

However, $y_n(t) = nt$ on $I = [0, 1]$ is not an equicontinuous sequence on I. Take $\varepsilon_0 = 1$, for any $\delta > 0$, one can take $t = 0$ and $s = \frac{\delta}{2}$, and we have $t,\ s \in I$ and $|t - s| < \delta$. However, then

$$|y_n(t) - y_n(s)| = n|t - s| = \frac{n\delta}{2} \ge \varepsilon$$

when $n \ge \frac{2\varepsilon}{\delta}$.

Nonetheless, for each n, the function $y_n(t)$ is continuous on I. □

Exercise 2.29. Let $I = [0, 2\pi]$. Show that the sequence of functions $x_n(t) := \sin \frac{t}{n}$ is equicontinuous on I, but $y_n(t) := \sin(nt)$ is not.

Exercise 2.30. Suppose $\{\mathbf{x}_n(t)\}_{n=1}^{\infty}$ is a sequence of differentiable functions on $t \in [a, b]$ such that $\mathbf{x}_n'(t)$ is uniformly bounded on $[a, b]$. Show that $\{\mathbf{x}_n(t)\}$ is a equicontinuous family on $[a, b]$.

Theorem 2.56 (Arzelà-Ascoli's Theorem)**.** *Let $I = [a, b]$ be a closed and bounded interval. Let $\mathbf{x}_n(t) : I \to \mathbb{R}^d$ be a sequence of functions such that:*

(1) there exists a constant $M > 0$ (independent of t and n) such that $|\mathbf{x}_n(t)| \leq M$ for any n and $t \in I$ (in other words, we say the sequence $\mathbf{x}_n(t)$ is uniformly bounded); and

(2) the sequence $\{\mathbf{x}_n(t)\}$ is equicontinuous on I.

Then, there exists a subsequence $\mathbf{x}_{n_k}(t)$ that converges uniformly on I to a limit function $\mathbf{y}(t)$ as $k \to \infty$.

We omit the proof here. The proof is essentially by a diagonalization argument. Using the Arzelà-Ascoli's Theorem, Peano gave the following ingenious proof of an existence theorem in 1890.

Theorem 2.57 (Peano's Existence Theorem)**.** *Let Ω be an open domain in $\mathbb{R}^d$ and $I = [-T, T]$ be a closed and bounded time interval. Suppose $\mathbf{F}(\mathbf{x}, t) : \Omega \times I \to \mathbb{R}^d$ is a continuous vector field on $\Omega \times I$. Then for any $\mathbf{x}_0 \in \Omega$, the IVP*

$$\mathbf{x}' = \mathbf{F}(\mathbf{x}, t), \quad \mathbf{x}(0) = \mathbf{x}_0$$

has a solution $\mathbf{x}(t)$ defined on $t \in [-\varepsilon, \varepsilon]$ for some small $\varepsilon > 0$.

Proof. We will only prove that there is a solution defined on $t \in [0, \varepsilon]$ for some $\varepsilon > 0$, since to show there is a solution on $[-\varepsilon, 0]$ is similar.

Let $B_r(\mathbf{x}_0)$ be an open ball in Ω such that $\overline{B_r(\mathbf{x}_0)} \subset \Omega$. By continuity of the vector field $\mathbf{F}$, there exists $M > 0$ such that $|\mathbf{F}(\mathbf{x}, t)| \leq M$ for any $(\mathbf{x}, t) \in B_r(\mathbf{x}_0) \times I$. Let $\varepsilon < \min\{\frac{r}{M}, T\} > 0$. We define a sequence of functions $\mathbf{x}_n(t) : [0, \varepsilon] \to \mathbb{R}^d$ in the following way:

Denote $J = [0, \varepsilon]$. For each n, divide J into n-subintervals (each has width $\frac{\varepsilon}{n}$):

$$\begin{aligned} J_n^1 &= [0,\ \varepsilon/n] \\ J_n^2 &= [\varepsilon/n,\ 2(\varepsilon/n)] \\ &\vdots \\ J_n^n &= [(n-1)(\varepsilon/n),\ \varepsilon], \end{aligned}$$

or in short, $J_n^k = [(k-1)(\varepsilon/n),\ k(\varepsilon/n)]$ for each $1 \leq k \leq n$.

Unlike the Picard's iteration sequence whose n-th term is defined by the previous $(n-1)$-th term, we define $\mathbf{x}_n(t)$ on each subinterval J_n^k successively by its previous values on J_n^{k-1}. We first define:

$$\mathbf{x}_n(t) := \mathbf{x}_0 \quad \text{for } t \in J_n^1.$$

Then on the next subinterval J_n^2, we define:

$$\mathbf{x}_n(t) := \mathbf{x}_0 + \int_0^{t-\frac{\varepsilon}{n}} \mathbf{F}(\underbrace{\mathbf{x}_n(s)}_{\text{not a typo!}}, s)\, ds \quad \text{for } t \in J_n^2.$$

When $t \in J_n^2$ which is an interval of width $\frac{\varepsilon}{n}$, then $t - \frac{\varepsilon}{n}$ is in the previous subinterval J_n^1. Therefore, the integral

$$\int_0^{t-\frac{\varepsilon}{n}} \mathbf{F}(\mathbf{x}_n(s),\, s)ds$$

is well-defined since $s \in [0,\, t - \frac{\varepsilon}{n}] \subset J_n^1$ on which we have already defined $\mathbf{x}_n$.

Now that we have already defined $\mathbf{x}_n(t)$ for $t \in J_n^1 \cup J_n^2$, next we move on to J_n^3 in a same fashion. Define:

$$\mathbf{x}_n(t) := \mathbf{x}_0 + \int_0^{t-\frac{\varepsilon}{n}} \mathbf{F}(\mathbf{x}_n(s),\, s)ds \quad \text{for } t \in J_n^3.$$

It is again well-defined since $t \in J_n^3$ implies $t - \frac{\varepsilon}{n} \in J_n^2$, and $\mathbf{x}_n(s)$ is already defined for $s \in J_n^1 \cup J_n^2$.

By successive definition of $\mathbf{x}_n(t)$ on each J_n^k via the relation:

$$\mathbf{x}_n(t) = \mathbf{x}_0 + \int_0^{t-\frac{\varepsilon}{n}} \mathbf{F}(\mathbf{x}_n(s),\, s)ds \quad \text{for } t \in J_n^2 \cup \ldots \cup J_n^n, \tag{2.16}$$

we get a sequence of functions $\{\mathbf{x}_n(t)\}_{n=1}^{\infty}$ defined on $J := [0, \varepsilon]$.

Next we show that $\mathbf{x}_n(t) \in B_r(\mathbf{x}_0)$ for $t \in J$. Obviously, this is true when $t \in J_n^1$. If this is true for $t \in J_n^1 \cup \ldots \cup J_n^{k-1}$, then for $t \in J_n^k$, the relation (2.16) implies:

$$|\mathbf{x}_n(t) - \mathbf{x}_0| \leq \int_0^{t-\frac{\varepsilon}{n}} |\mathbf{F}(\underbrace{\mathbf{x}_n(s)}_{\in B_r(\mathbf{x}_0)},\, s)|ds \leq M\left|t - \frac{\varepsilon}{n}\right| \leq M \cdot \varepsilon < M \cdot \frac{r}{M} = r.$$

By induction, we have $\mathbf{x}_n(t) \in B_r(\mathbf{x}_0)$ for any $t \in J$. In particular, the sequence of functions $\mathbf{x}_n(t)$ is uniformly bounded on J since by triangle inequality, we have:

$$|\mathbf{x}_n(t)| = |\mathbf{x}_n(t) - \mathbf{x}_0 + \mathbf{x}_0| \leq |\mathbf{x}_n(t) - \mathbf{x}_0| + |\mathbf{x}_0| \leq \underbrace{r + |\mathbf{x}_0|}_{\text{finite constant}}$$

for any $t \in J$ and $n \in \mathbb{N}$. It verifies the first condition of the Arzelà-Ascoli's Theorem.

Next we argue that the sequence $\mathbf{x}_n(t)$ is equicontinuous on J:

We first show that for any $t,\, s \in J$, we have $|\mathbf{x}_n(t) - \mathbf{x}_n(s)| \leq M|t-s|$, then the equicontinuity will follow easily.

When $t,\, s \in J_n^2 \cup \ldots \cup J_n^n$, by (2.16), we have (without loss of generality assuming $s < t$):

$$\begin{aligned}
|\mathbf{x}_n(t) - \mathbf{x}_n(s)| &= \left|\int_0^{t-\frac{\varepsilon}{n}} \mathbf{F}(\mathbf{x}_n(\tau),\, \tau)d\tau - \int_0^{s-\frac{\varepsilon}{n}} \mathbf{F}(\mathbf{x}_n(\tau),\, \tau)d\tau\right| \\
&= \left|\int_{s-\frac{\varepsilon}{n}}^{t-\frac{\varepsilon}{n}} \mathbf{F}(\mathbf{x}_n(\tau),\, \tau)d\tau\right| \\
&\leq \int_{s-\frac{\varepsilon}{n}}^{t-\frac{\varepsilon}{n}} |\mathbf{F}(\mathbf{x}_n(\tau),\, \tau)|\, d\tau \\
&\leq M\left|\left(t - \frac{\varepsilon}{n}\right) - \left(s - \frac{\varepsilon}{n}\right)\right| = M|t-s|.
\end{aligned}$$

We leave it as an exercise for readers to verify this is also true if at least one of $t,\, s$ is in J_n^1.

For any $\varepsilon' > 0$ (we use ε' here because ε was already used to denote the width of the interval), let $\delta < \frac{\varepsilon'}{M}$, then given any $t,\, s \in J$ and $|t-s| < \delta$, we have:

$$|\mathbf{x}_n(t) - \mathbf{x}_n(s)| \leq M|t-s| \leq M\delta < \varepsilon'.$$

Therefore, the sequence $\mathbf{x}_n(t)$ is equicontinuous on J.

Now the sequence $\mathbf{x}_n(t)$ is both uniformly bounded and equicontinuous on J. By Arzelà-Ascoli's Theorem, there exists a subsequence $\mathbf{x}_{n_i}(t)$ which converges uniformly on J to a limit function $\mathbf{x}_\infty(t)$ as $n_i \to \infty$.

Finally, we are going to show that this $\mathbf{x}_\infty(t)$ is a solution to the given IVP. Clearly, $\mathbf{x}_n(0) = \mathbf{x}_0$ for all $n \in \mathbb{N}$, and so $\mathbf{x}_\infty(0) = \mathbf{x}_0$, which verifies the initial condition. For any $t > 0$, we and choose n_i sufficiently large so that $t \notin J^1_{n_i}$ which has width $\frac{\varepsilon}{n_i} \to 0$ as $n_i \to \infty$. By (2.16), we have:

$$\begin{aligned}\mathbf{x}_{n_i}(t) &= \mathbf{x}_0 + \int_0^{t-\frac{\varepsilon}{n_i}} \mathbf{F}(\mathbf{x}_{n_i}(s),\, s)ds \\ &= \mathbf{x}_0 + \int_0^t \mathbf{F}(\mathbf{x}_{n_i}(s),\, s)ds - \underbrace{\int_{t-\frac{\varepsilon}{n_i}}^t \mathbf{F}(\mathbf{x}_{n_i}(s),\, s)ds}_{\leq M\cdot\frac{\varepsilon}{n_i} \to 0 \text{ as } n_i\to\infty}.\end{aligned}$$

As $n_i \to \infty$, we have:

$$\mathbf{x}_\infty(t) = \mathbf{x}_0 + \lim_{n_i\to\infty} \int_0^t \mathbf{F}(\mathbf{x}_{n_i}(s),\, s)ds.$$

Since $\mathbf{F}$ is continuous on the compact set $\overline{B_r(\mathbf{x}_0)} \times J$, it is also uniformly continuous on $\overline{B_r(\mathbf{x}_0)} \times J$. For any $\varepsilon' > 0$, there exists $\delta > 0$ depending only on ε', such that whenever $\mathbf{y}, \mathbf{z} \in B_r(\mathbf{x}_0)$ and $|\mathbf{y} - \mathbf{z}| < \delta$, we have

$$|\mathbf{F}(\mathbf{y}, t) - \mathbf{F}(\mathbf{z}, t)| < \varepsilon'$$

for any $t \in J$.

As $\mathbf{x}_{n_i}$ converges uniformly on J to $\mathbf{x}_\infty$ as $n_i \to \infty$, there exists $N > 0$ such that whenever $n_i > N$, we have $\|\mathbf{x}_{n_i} - \mathbf{x}_\infty\|_{\infty,J} < \delta$, which implies

$$|\mathbf{F}(\mathbf{x}_{n_i}(t)) - \mathbf{F}(\mathbf{x}_\infty(t))| < \varepsilon'$$

for any $t \in J$. Therefore, $\mathbf{F}(\mathbf{x}_{n_i}, \cdot)$ converges uniformly on J to $\mathbf{F}(\mathbf{x}_\infty, \cdot)$, thus allow switching between limit and integral signs, i.e.

$$\lim_{n_i\to\infty} \int_0^t \mathbf{F}(\mathbf{x}_{n_i}(s),\, s)ds = \int_0^t \lim_{n_i\to\infty} \mathbf{F}(\mathbf{x}_{n_i}(s),\, s)ds = \int_0^t \mathbf{F}(\mathbf{x}_\infty(s),\, s)ds.$$

Therefore, the continuous function $\mathbf{x}_\infty(t)$ on J satisfies the integral equation:

$$\mathbf{x}_\infty(t) = \mathbf{x}_0 + \int_0^t \mathbf{F}(\mathbf{x}_\infty(s),\, s)ds$$

and so it is a solution to the given IVP. $\square$

To summarize, given an IVP $\mathbf{x}' = \mathbf{F}(\mathbf{x}, t), \quad \mathbf{x}(0) = \mathbf{x}_0$:

- if $\mathbf{F}$ is (globally) Lipschitz continuous on $\mathbb{R}^d \times (-\infty, \infty)$, then the IVP has a unique global solution $\mathbf{x}(t)$ defined on $t \in (-\infty, \infty)$;
- if $\mathbf{F}$ is Lipschitz continuous on $\Omega \times [-T, T]$ where Ω is an open set in $\mathbb{R}^d$ containing $\mathbf{x}_0$, then the IVP has a unique solution defined at least for short-time $t \in [-\varepsilon, \varepsilon]$.
- if $\mathbf{F}$ is only continuous near $\mathbf{x}_0$, the IVP still has a solution defined at least for short-time $t \in [-\varepsilon, \varepsilon]$, but it may not be unique.

Chapter 3

Stability

An equilibrium solution of an ODE system is a special solution which is stationary in the phase portrait. Precisely, it is a solution of the form $\mathbf{x}(t) = \mathbf{x}^*$ for all t where $\mathbf{x}^* \in \mathbb{R}^d$ is fixed. For an autonomous system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$, an equilibrium solution $\mathbf{x}(t) = \mathbf{x}^*$ can only happen when $\mathbf{F}(\mathbf{x}^*) = \mathbf{0}$.

In the rest of the course, we will focus exclusively on autonomous systems. The central theme of this chapter is about whether an equilibrium solution is stable or not, which is an important concept to be defined. Heuristically, a stable equilibrium is one that if you move slightly away from the equilibrium point, it will stay close or even flow towards to the equilibrium point as time goes. An unstable equilibrium, as the name implies, will move away from the equilibrium point.

For linear systems $\mathbf{x}' = A\mathbf{x}$, the general solution suggested that if the eigenvalues of A are all negative, then the origin, which is an equilibrium solution, will tend to be stable, while if A has a positive eigenvalue, the solution will tend away from the origin if one moves slightly away from it along the eigenvector direction.

In this chapter, we will first give the rigorous definition of stability, and verify that the stability of a planar linear system is determined by the eigenvalues of the matrix. Then, we will work with nonlinear systems using linear approximations (i.e., compare them with their linear counterparts). One neat and elegant fact is that phase portrait of the a nonlinear system effectively resembles its approximated linear systems near the equilibrium point[1]. Since the stability of the approximated linear system can be studied by solving for its eigenvalue, one can then determine the stability of the nonlinear system by looking at its linear approximation.

3.1. Definitions of Stability

We begin by defining several notions of stability.

Definition 3.1 (Equilibrium Point). A point $\mathbf{x}^* \in \mathbb{R}^d$ is an **equilibrium point**, or an **equilibrium solution**, of an autonomous system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ if $\mathbf{F}(\mathbf{x}^*) = \mathbf{0}$.

[1]There are some exceptions though. We will discuss that later.

Example 3.1. To find equilibrium point(s), we can set $\mathbf{F}(\mathbf{x}) = \mathbf{0}$ and solve for $\mathbf{x}$. For instance, equilibrium points of the system

$$\begin{bmatrix} x_1' \\ x_2' \end{bmatrix} = \begin{bmatrix} x_1^2 + x_2 \\ x_1 - x_2^2 \end{bmatrix}$$

can be found by solving:

$$\begin{aligned} x_1^2 + x_2 &= 0 \\ x_1 - x_2^2 &= 0. \end{aligned}$$

The second equation implies $x_1 = x_2^2$ and substitute it into the first, we have $x_2^4 + x_2 = 0$, or equivalently $x_2(x_2^3 + 1) = 0$. Therefore, $x_2 = 0$ or $x_2 = -1$. The former case gives $(x_1, x_2) = (0, 0)$, the second case gives $(x_1, x_2) = (1, -1)$.

Therefore, there are two equilibrium points $(0, 0)$ and $(1, -1)$. □

Definition 3.2 (Stable and Unstable Equilibria)**.** Let Ω be an open domain in $\mathbb{R}^d$ and $\mathbf{F} : \Omega \to \mathbb{R}^d$ is a C^1 vector field. Consider an autonomous system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$. Suppose $\mathbf{x}^* \in \Omega$ is an equilibrium point of the system, i.e. $\mathbf{F}(\mathbf{x}^*) = \mathbf{0}$, then

(1) We say $\mathbf{x}^*$ is **stable** if for any open ball $B_\varepsilon(\mathbf{x}^*) \subset \Omega$, there exists an open ball $B_\delta(\mathbf{x}^*)$ such that whenever $\mathbf{x}_0 \in B_\delta(\mathbf{x}^*)$, we have $\varphi_t(\mathbf{x}_0) \in B_\varepsilon(\mathbf{x}^*)$ for all $t \in [0, \infty)$.

(2) We say $\mathbf{x}^*$ is **asymptotically stable** if for any open ball $B_\varepsilon(\mathbf{x}^*) \subset \Omega$, there exists an open ball $B_\delta(\mathbf{x}^*)$ such that whenever $\mathbf{x}_0 \in B_\delta(\mathbf{x}^*)$, we have $\varphi_t(\mathbf{x}_0) \in B_\varepsilon(\mathbf{x}^*)$ for all $t \in [0, \infty)$ and $\varphi_t(\mathbf{x}_0) \to \mathbf{x}^*$ as $t \to \infty$. [Therefore, *asymptotically stable* implies *stable*.]

(3) We say $\mathbf{x}^*$ is **unstable** if it is not stable. Precisely, $\mathbf{x}^*$ is an unstable equilibrium if there exists an open ball $B_\varepsilon(\mathbf{x}^*) \subset \Omega$, such that for any open ball $B_\delta(\mathbf{x}^*)$, there exist a point $\mathbf{x}_0 \in B_\delta(\mathbf{x}^*)$ and a time $\tau \in [0, \infty)$ with $\varphi_\tau(\mathbf{x}_0) \notin B_\varepsilon(\mathbf{x}^*)$.

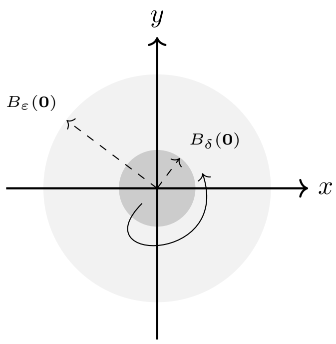


Figure 3.1. For a stable equilibrium point $\mathbf{0}$: one can find a small δ such that for any $\mathbf{x}_0 \in B_\delta(\mathbf{0})$, the forward flow $\varphi_t(\mathbf{0})$, $t \geq 0$, always stays inside $B_\varepsilon(\mathbf{0})$.

Example 3.2. Suppose the flow φ_t of a nonlinear system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ with an equilibrium point $\mathbf{x}^*$ satisfies:

$$(*) \qquad |\varphi_t(\mathbf{x}_0) - \mathbf{x}^*| \leq C\,|\mathbf{x}_0 - \mathbf{x}^*|\,e^{-t}$$

for all $t \in [0, \infty)$ and any $\mathbf{x}_0 \in \mathbb{R}^d$, where $C > 0$ is a constant. We are going to verify from the definition that $\mathbf{x}^*$ is asymptotically stable.

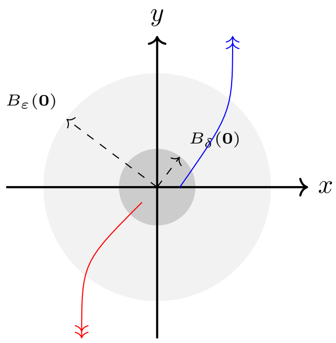


Figure 3.2. For an unstable equilibrium point $\mathbf{0}$: no matter how small δ is, one can always find $\mathbf{x}_0 \in B_\delta(\mathbf{0})$ such that it flows outside the ε-ball after some finite time.

Given any $\varepsilon > 0$, we need to find $\delta =$ ____ (which is left blank for a while), such that whenever $|\mathbf{x}_0 - \mathbf{x}^*| < \delta$, we have $|\varphi_t(\mathbf{x}_0) - \mathbf{x}^*| < \varepsilon$ for any $t \in [0, \infty)$. In order to achieve this, we consider:

$$\begin{aligned} |\varphi_t(\mathbf{x}_0) - \mathbf{x}^*| &\leq C\,|\mathbf{x}_0 - \mathbf{x}^*|\,e^{-t} && \text{(by the given condition)} \\ &\leq C\,|\mathbf{x}_0 - \mathbf{x}^*| && \text{(since } e^{-t} \leq 1 \text{ for } t \geq 0\text{)} \\ &< C\delta && \text{(whenever } |\mathbf{x}_0 - \mathbf{x}^*| < \delta\text{)} \end{aligned}$$

To ensure $|\varphi_t(\mathbf{x}_0) - \mathbf{x}^*| < \varepsilon$, one can choose δ equal to anything smaller than εC^{-1}.

To recap the whole argument: given any $\varepsilon > 0$, pick $\delta = \frac{1}{2}\varepsilon C^{-1}$, then whenever $|\mathbf{x}_0 - \mathbf{x}^*| < \delta$, we have

$$|\varphi_t(\mathbf{x}_0) - \mathbf{x}^*| \leq C\,|\mathbf{x}_0 - \mathbf{x}^*| < C\delta = C \cdot \frac{1}{2}\varepsilon C^{-1} = \frac{\varepsilon}{2} < \varepsilon$$

for any $t \in [0, \infty)$. In other words, whenever $\mathbf{x}_0 \in B_\delta(\mathbf{x}^*)$, we have $\varphi_t(\mathbf{x}_0) \in B_\varepsilon(\mathbf{x}^*)$ for all $t \in [0, \infty)$. From the definition, $\mathbf{x}^*$ is stable.

To show $\mathbf{x}^*$ is in fact asymptotically stable, we apply the squeezing principle on (*). Let $t \to +\infty$, the right-hand side $C\,|\mathbf{x}_0 - \mathbf{x}^*|\,e^{-t} \to 0$. Therefore $|\varphi_t(\mathbf{x}_0) - \mathbf{x}^*| \to 0$ as well. In other words, $\varphi_t(\mathbf{x}_0) \to \mathbf{x}^*$ for any $\mathbf{x}_0 \in \mathbb{R}^d$. Therefore $\mathbf{x}^*$ is asymptotically stable. □

Exercise 3.1. Suppose φ_t is the flow of a nonlinear system which satisfies:

$$|\varphi_t(\mathbf{x}_0)| \leq f(t)\,|\mathbf{x}_0|$$

for any $t \in [0, \infty)$ and any $\mathbf{x}_0 \in \mathbb{R}^d$, where $f : [0, \infty) \to \mathbb{R}$ is some positive-valued function. First, show that $\mathbf{0}$ is an equilibrium point.

Show, from the definitions, that:

- if there exists a constant C such that $f(t) < C$ for all $t \geq 0$, then $\mathbf{0}$ is stable.
- furthermore, if $f(t) \to 0$ as $t \to +\infty$, then $\mathbf{0}$ is asymptotically stable.

Exercise 3.2. Let φ_t be the flow of a nonlinear system with an equilibrium point $\mathbf{x}^*$. Suppose for any $\mathbf{x}_0 \in \mathbb{R}^d$, the magnitude $|\varphi_t(\mathbf{x}_0) - \mathbf{x}^*|$ is always decreasing as t increases. Show that $\mathbf{x}^*$ is a stable equilibrium.

3.1.0.1. Stability of Planar Linear Systems. We next discuss the stability of planar linear systems whose phase portraits and general solutions are well understood.

As in the previous example and exercise, in order to prove stability from the definition, it is crucial to establish some inequalities on the flow $\varphi_t(\mathbf{x}_0)$ so that its magnitude can be controlled by the magnitude of $\mathbf{x}_0 - \mathbf{x}^*$. The flow of a linear system is given by $\Phi_t(\mathbf{x}_0) = e^{tA}\mathbf{x}_0$. In view of that, we will first derive several inequalities concerning e^{tA} which will be helpful for us later to pick the right δ for each given ε.

Lemma 3.3. *Let D, Q and J be the following canonical matrices:*

$$D = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}, \quad Q = \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}, \quad J = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}$$

where λ_1, λ_2, α, β and λ are all real and $\beta \neq 0$. Then, for all $t \geq 0$:

$$\left\|e^{tD}\right\| \leq e^{\max\{\lambda_1,\, \lambda_2\}t}, \quad \left\|e^{tQ}\right\| \leq e^{\alpha t}, \quad \left\|e^{tJ}\right\| \leq (1+t)e^{\lambda t}.$$

Proof. Given any $\mathbf{x} = (x_1,\, x_2) \in \mathbb{R}^2$ such that $|\mathbf{x}| = 1$, we have:

$$\begin{aligned} \left|e^{tD}\mathbf{x}\right|^2 &= \left|\begin{bmatrix} e^{\lambda_1 t} & 0 \\ 0 & e^{\lambda_2 t} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right|^2 = \left|\begin{bmatrix} e^{\lambda_1 t}x_1 \\ e^{\lambda_2 t}x_2 \end{bmatrix}\right|^2 \\ &= e^{2\lambda_1 t}x_1^2 + e^{2\lambda_2 t}x_2^2 \\ &\leq \max\{e^{2\lambda_1 t}, e^{2\lambda_2 t}\}(x_1^2 + x_2^2) \\ &= \max\{e^{2\lambda_1 t}, e^{2\lambda_2 t}\} = e^{2\max\{\lambda_1,\lambda_2\}t}. \end{aligned}$$

Therefore, $\left|e^{tD}\mathbf{x}\right| \leq e^{\max\{\lambda_1,\lambda_2\}t}$ whenever $|\mathbf{x}| = 1$. In other words, we have

$$\left\|e^{tD}\right\| \leq e^{\max\{\lambda_1,\lambda_2\}t}.$$

For e^{tQ} where $Q = \begin{bmatrix} \alpha t & \beta t \\ -\beta t & \alpha t \end{bmatrix} = \underbrace{\begin{bmatrix} \alpha t & 0 \\ 0 & \alpha t \end{bmatrix} + \begin{bmatrix} 0 & \beta t \\ -\beta t & 0 \end{bmatrix}}_{\text{commute}}$, we have:

$$\begin{aligned} e^{tQ} &= \exp\left(\begin{bmatrix} \alpha t & 0 \\ 0 & \alpha t \end{bmatrix} + \begin{bmatrix} 0 & \beta t \\ -\beta t & 0 \end{bmatrix}\right) \\ &= \exp\begin{bmatrix} \alpha t & 0 \\ 0 & \alpha t \end{bmatrix} \cdot \exp\begin{bmatrix} 0 & \beta t \\ -\beta t & 0 \end{bmatrix} \\ &= \exp\begin{bmatrix} \alpha t & 0 \\ 0 & \alpha t \end{bmatrix} \cdot \begin{bmatrix} \cos\beta t & \sin\beta t \\ -\sin\beta t & \cos\beta t \end{bmatrix} \end{aligned}$$

Note that $\begin{bmatrix} \cos\beta t & \sin\beta t \\ -\sin\beta t & \cos\beta t \end{bmatrix}$ is a rotation matrix. If $\mathbf{x}$ is unit, then $\begin{bmatrix} \cos\beta t & \sin\beta t \\ -\sin\beta t & \cos\beta t \end{bmatrix}\mathbf{x}$ is also unit. Therefore, $\left\|\begin{bmatrix} \cos\beta t & \sin\beta t \\ -\sin\beta t & \cos\beta t \end{bmatrix}\right\| = 1$ which implies:

$$\begin{aligned} \left\|e^{tQ}\right\| &\leq \left\|\exp\begin{bmatrix} \alpha t & 0 \\ 0 & \alpha t \end{bmatrix}\right\| \left\|\begin{bmatrix} \cos\beta t & \sin\beta t \\ -\sin\beta t & \cos\beta t \end{bmatrix}\right\| \\ &\leq \underbrace{e^{\alpha t}}_{\text{from previous part}} \cdot 1 = e^{\alpha t}. \end{aligned}$$

For e^{tJ} where $J = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}$, we have

$$e^{tJ} = \begin{bmatrix} e^{\lambda t} & te^{\lambda t} \\ 0 & e^{\lambda t} \end{bmatrix} = e^{\lambda t}\begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}$$

Consider $\mathbf{x} = (x_1, x_2)$ with $|\mathbf{x}| = 1$, then

$$\begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 + tx_2 \\ x_2 \end{bmatrix}$$

$$\begin{aligned} \left| \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right|^2 &= (x_1 + tx_2)^2 + x_2^2 \\ &= x_1^2 + 2tx_1x_2 + t^2x_2^2 + x_2^2 \\ &= 1 + 2tx_1x_2 + t^2x_2^2 \\ &\le 1 + t(x_1^2 + x_2^2) + t^2(x_1^2 + x_2^2) \qquad \text{(since } 2x_1x_2 \le x_1^2 + x_2^2\text{)} \\ &= 1 + t + t^2 \le 1 + 2t + t^2 = (1+t)^2. \end{aligned}$$

Therefore, $\left\| \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix} \right\| \le (1+t)$, and so $\left\| e^{tJ} \right\| \le (1+t)e^{\lambda t}$. □

Exercise 3.3. Suppose A is a $d \times d$ real matrix with distinct real eigenvalues $\lambda_1, \ldots, \lambda_d$. Show that $\left\| e^{tA} \right\| \le e^{\max\{\lambda_1, \ldots, \lambda_d\}t}$ for any $t > 0$.

Exercise 3.4. Suppose A is a $d \times d$ real matrix whose real eigenvalues are all negative, and whose complex eigenvalues all have negative real parts. Prove that there exists $\delta > 0$ such that

$$\left\| e^{tA} \right\| \le e^{-\delta t}$$

for any $t > 0$.

Let A be a 2×2 real matrix and consider the planar linear system $\mathbf{x}' = A\mathbf{x}$. We are going to argue that the stability of the origin is determined by the real parts of the eigenvalues of A according to the following table:

Eigenvalues of A		Signs	The origin is:
distinct real	(λ_1, λ_2)	$(-, -)$	asymptotically stable
		$(-, 0)$ or $(0, -)$	stable
		$(+, *)$ or $(*, +)$	unstable
complex	$\alpha \pm \beta i$	$\alpha < 0$	asymptotically stable
		$\alpha = 0$	stable
		$\alpha > 0$	unstable
repeated real	λ	$-$	asymptotically stable
		0 and $A \ne 0$	unstable
		0 and $A = 0$	stable
		$+$	unstable

Table 1. Stability of planar linear systems ($*$ can be any real number)

We split the proof into two cases. One part assumes A has distinct real eigenvalues or complex eigenvalues, another assumes A has repeated real eigenvalues. The latter case is a bit more subtle.

Case 1: A has distinct real or complex eigenvalues

Depending on whether A has real or complex eigenvalues, there exists a matrix K and an invertible matrix P such that $A = PKP^{-1}$, where K is either a diagonal matrix or a complex canonical form with the same eigenvalues as A. The flow of the system is given by $\Phi_t(\mathbf{x}_0) = e^{tA}\mathbf{x}_0 = e^{tPKP^{-1}}\mathbf{x}_0 = Pe^{tK}P^{-1}\mathbf{x}_0$.

When the real parts of the eigenvalues of K are all non-positive, from the estimate proved in Lemma 3.3, we have $\left\|e^{tK}\right\| \leq e^{-\mu t}$ for some $-\mu \leq 0$ (where $-\mu$ is either $\max\{\lambda_1, \lambda_2\}$ or α depending on whether eigenvalues of A, and hence of K, are real or complex). Therefore:

$$\begin{aligned}|\Phi_t(\mathbf{x}_0)| &= \left|Pe^{tK}P^{-1}\mathbf{x}_0\right| \\ &\leq \|P\|\|e^{tK}\|\|P^{-1}\|\,|\mathbf{x}_0| \\ &\leq \|P\|\|P^{-1}\|\,|\mathbf{x}_0|\,e^{-\mu t}.\end{aligned}$$

To show that the origin is stable, we first note that $e^{-\mu t} \leq 1$ for any $t \in [0, \infty)$, and so

$$|\Phi_t(\mathbf{x}_0)| \leq \|P\|\|P^{-1}\|\,|\mathbf{x}_0|\,.$$

Given any $\varepsilon > 0$, one can find a $\delta < \frac{\varepsilon}{\|P\|\|P^{-1}\|}$ such that whenever $\mathbf{x}_0 \in B_\delta(\mathbf{0})$, we have:

$$|\Phi_t(\mathbf{x}_0)| \leq \|P\|\|P^{-1}\|\,|\mathbf{x}_0| \underbrace{\leq}_{\mathbf{x}_0 \in B_\delta(\mathbf{0})} \|P\|\|P^{-1}\|\delta \underbrace{<}_{\text{by choice of } \delta} \varepsilon.$$

In other words, $\Phi_t(\mathbf{x}_0) \in B_\varepsilon(\mathbf{0})$ for any $t \in [0, \infty)$. By the definition of stability, the origin is stable.

When the real parts of the eigenvalues of A (and hence of K) are all negative, then we further have

$$|\Phi_t(\mathbf{x}_0)| \leq \|P\|\|P^{-1}\|\|\mathbf{x}_0|e^{-\mu t} \to 0$$

as $t \to \infty$. In other words, $\Phi_t(\mathbf{x}_0) \to \mathbf{0}$ as $t \in \infty$. Therefore, the origin is asymptotically stable.

To show that the origin is unstable if one of the eigenvalues of A has positive real part, it is easier to use some particular solutions instead of matrix exponentials. From Theorems 1.9 and 1.16, by letting $c_2 = 0$, either one of the following is a solution to the system (depending on whether A has distinct real or complex eigenvalues):

$$\begin{aligned}\mathbf{x}(t) &= c_1 e^{\lambda_1 t}\mathbf{v}_1 && \text{(real)} \\ \mathbf{x}(t) &= c_1 e^{\alpha t}\begin{bmatrix}\cos\beta t \\ -\sin\beta t\end{bmatrix} && \text{(complex)}\end{aligned}$$

where $c_1 \in \mathbb{R}$, and for the real case, $\lambda_1 > 0$ is an eigenvalue of A, $\mathbf{v}_1$ is a **unit** eigenvector of A; and for the complex case, $\alpha > 0$ is the real part of the eigenvalue and $\beta \in \mathbb{R}$ is the imaginary part. To show instability from the definition, we fix $\varepsilon = 1$ and take an arbitrary small $\delta > 0$, and we need to find an initial condition $\mathbf{x}_0 \in B_\delta(\mathbf{0})$ such that $\Phi_t(\mathbf{x}_0)$ will eventually leave $B_\varepsilon(\mathbf{0})$. We argue only the real case since the complex case is similar. Recall that $\mathbf{v}_1$ is unit, given any $\delta > 0$, we consider the solution $\mathbf{x}(t) = \frac{\delta}{2}e^{\lambda_1 t}\mathbf{v}_1$. Then $|\mathbf{x}(0)| = \frac{\delta}{2} < \delta$ and so $\mathbf{x}(0) \in B_\delta(\mathbf{0})$. However, when $t > \frac{1}{\lambda_1}\log\frac{2}{\delta}$,

$$|\mathbf{x}(t)| > \frac{\delta}{2}e^{\lambda_1 \cdot \frac{1}{\lambda_1}\log\frac{2}{\delta}}\,|\mathbf{v}_1| = 1 = \varepsilon$$

and so $\mathbf{x}(t) \notin B_\varepsilon(\mathbf{0})$. By the definition, $\mathbf{0}$ is unstable.

Case 2: A has a repeated real eigenvalue

This case is more subtle than the distinct real or complex cases because the estimate $\left\|e^{tJ}\right\| \leq (1+t)e^{\lambda t}$, where J is a Jordan canonical form, is not as sharp as those for the diagonal or complex canonical form. Denote λ the only eigenvalue of A.

When $\lambda < 0$, then $A = PKP^{-1}$ where K is either $D := \begin{bmatrix}\lambda & 0 \\ 0 & \lambda\end{bmatrix}$ or $J := \begin{bmatrix}\lambda & 1 \\ 0 & \lambda\end{bmatrix}$, P is some invertible matrix. The former case was settled in Case 1. Assuming we have $K = J$, then by Lemma 3.3, we have:

$$\left|e^{tA}\mathbf{x}_0\right| \leq \|P\|\|P^{-1}\|\,|\mathbf{x}_0| \cdot (1+t)e^{\lambda t}.$$

The subtlety here is that the upper bound for $(1+t)e^{\lambda t}$ is not as trivial as in Case 1. Let $f(t) = (1+t)e^{\lambda t}$. From elementary calculus, one can show f is increasing on $(-\infty, \frac{\lambda+1}{-\lambda}]$ and is decreasing on $[\frac{\lambda+1}{-\lambda}, \infty)$. Therefore, $f(t)$ has an absolute maximum and so it is bounded from above by a constant C. Thus, we have:

$$|\Phi_t(\mathbf{x}_0)| \leq C\|P\|\|P^{-1}\| \, |\mathbf{x}_0|$$

for any $t \in [0, \infty)$. Given any $\varepsilon > 0$, choose $\delta < \frac{\varepsilon}{C\|P\|\|P^{-1}\|}$, then whenever $\mathbf{x}_0 \in B_\delta(\mathbf{0})$, we have:

$$|\Phi_t(\mathbf{x}_0)| \leq C\|P\|\|P^{-1}\| \cdot \frac{\varepsilon}{C\|P\|\|P^{-1}\|} = \varepsilon.$$

Hence $\Phi_t(\mathbf{x}_0) \in B_\varepsilon(\mathbf{0})$ for any $t \in [0, \infty)$. Also, by l'Hopital's rule:

$$\lim_{t\to+\infty} (1+t)e^{\lambda t} = \lim_{t\to+\infty} \frac{1+t}{e^{-\lambda t}} = \lim_{t\to+\infty} \frac{1}{-\lambda e^{-\lambda t}} = 0.$$

Therefore, $\Phi_t(\mathbf{x}_0) \to \mathbf{0}$ as $t \to +\infty$, and hence $\mathbf{0}$ is asymptotically stable.

If $\lambda = 0$, there are two possibilities. If $A = 0$ then all solutions to the system are stationary, which is clearly stable (but not asymptotically stable). For the other case, the phase portrait is a family of parallel lines as shown in Figure 1.14. The general solution is of the form:

$$\mathbf{x}(t) = P\left(c_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} t \\ 1 \end{bmatrix}\right), \quad c_1, c_2 \in \mathbb{R},$$

where P is some invertible matrix. In particular, $\mathbf{x}(t) = c_2 P \begin{bmatrix} t \\ 1 \end{bmatrix}$ is a solution to the system with initial condition $\mathbf{x}(0) = c_2 P \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ for any $c_2 \in \mathbb{R}$. To show the origin is unstable, we again fix $\varepsilon = 1$, and consider an arbitrarily small $\delta > 0$, choose $|c_2| < \frac{\delta}{\|P\|}$, then $|\mathbf{x}(0)| \leq |c_2| \, \|P\| < \delta$ and so the initial condition $\mathbf{x}(0) \in B_\delta(\mathbf{0})$. However, as $t \to +\infty$, $\mathbf{x}(t) = c_2 P \begin{bmatrix} t \\ 1 \end{bmatrix}$ is unbounded and must leave the ball $B_\varepsilon(\mathbf{0})$ after some finite time. Therefore, the origin is unstable.

If $\lambda > 0$, then the fact that the origin is unstable can be proved similarly as in the distinct real case. Let $\mathbf{v}$ be a unit eigenvector of A (with eigenvalue λ), then

$$\mathbf{x}(t) = ce^{\lambda t}\mathbf{v}$$

is a solution to the system for any $c \in \mathbb{R}$. The rest of the argument is exactly the same as in the distinct real case.

Exercise 3.5. Suppose A is a $d \times d$ matrix with distinct real eigenvalues $\lambda_1, \ldots, \lambda_d$. Consider the system $\mathbf{x}' = A\mathbf{x}$. Show that:

- If λ_i's are all non-positive, then the origin is stable.
- If λ_i's are all (strictly) negative, then the origin is asymptotically stable.
- If at least one of the λ_i's is positive, then the origin is unstable.

Exercise 3.6. Let φ_t be the flow of a nonlinear autonomous system on $\mathbb{R}^d$ with an equilibrium point $\mathbf{x}^*$. Suppose there exists a $\mathbf{x}_0 \in \mathbb{R}^d$ such that $\varphi_t(\mathbf{x}_0) \to \mathbf{x}^*$ as $t \to -\infty$ and $\varphi_t(\mathbf{x}_0)$ is unbounded as $t \to +\infty$. Show that $\mathbf{x}^*$ is unstable.

3.2. Linearization

In the previous section, we have completely understood the stability of the origin for planar linear systems. The results can be easily extended to the higher dimensional cases where the matrix A has distinct real or complex eigenvalues, although the case for repeated real or complex eigenvalues in higher dimensions will involve some sophisticated linear algebra (Jordan decomposition).

One important way to study the stability of a nonlinear system is by comparing them with an approximated linear system. As an illustration, let's consider the following system

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

Clearly, $(0,0)$ is an equilibrium point. Near $(0,0)$, the quadratic term y^2 is considerably smaller than the linear terms x and $-y$. Therefore, it is expected that the phase portrait will resemble the linear system

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

where the quadratic term y^2 is dropped. Sketches of their phase portraits reveal that they do look similar around $(0,0)$ and both are of saddle types, but they deviate more and more away from $(0,0)$. See Figures 3.3 and 3.4.

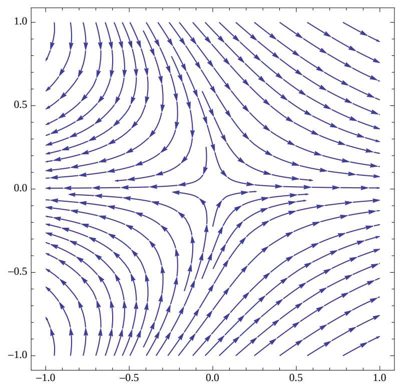


Figure 3.3. The phase portrait of the system: $x' = x + y^2, \quad y' = -y$.

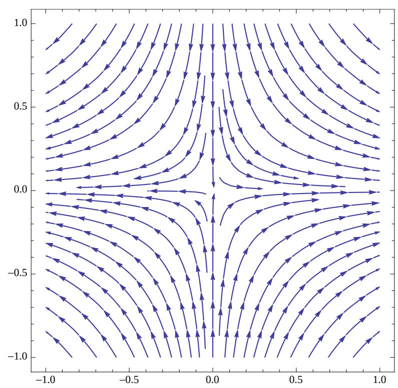


Figure 3.4. The phase portrait of the system: $x' = x, \quad y' = -y$.

Let's look at another example:

$$x' = -x + y + y^2$$
$$y' = -2y + x^3$$

Similarly, $(0,0)$ is an equilibrium point near which the nonlinear terms y^2 and x^3 are very small compared to the linear terms. Therefore, we expect its phase portrait near $(0,0)$ should look similar to the linear system

$$x' = -x + y$$
$$y' = -2y$$

Figures 3.5 and 3.6 illustrate that this prediction is indeed true.

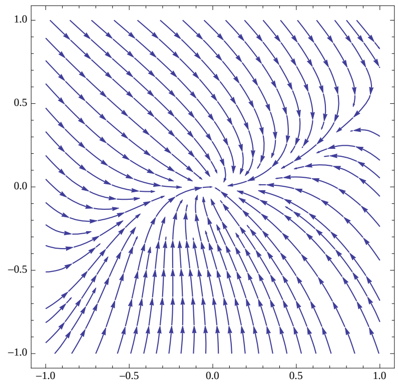


Figure 3.5. The phase portrait of the system: $x' = -x + y + y^2, \quad y' = -2y + x^3$.

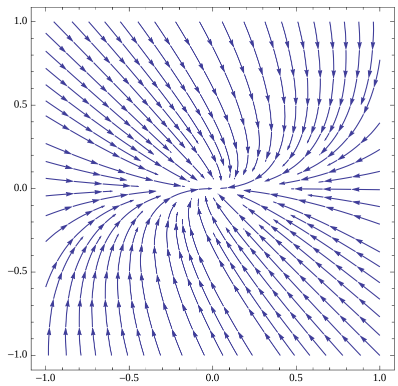


Figure 3.6. The phase portrait of the system: $x' = -x + y, \quad y' = -2y$.

However, there are non-examples where dropping nonlinear terms will change the phase portrait substantially. Let's consider the system

$$x' = x^2$$
$$y' = -y$$

and after dropping the nonlinear term x^2, we get the linear system:

$$x' = 0$$
$$y' = -y$$

You may see Figures 3.7 and 3.8 that their phase portraits look substantially different even around $(0, 0)$. In particular, the origin in the nonlinear system is unstable whereas in the linear system is stable.

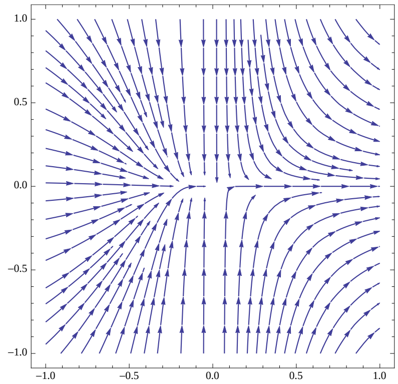


Figure 3.7. The phase portrait of the system: $x' = x^2, \quad y' = -y$.

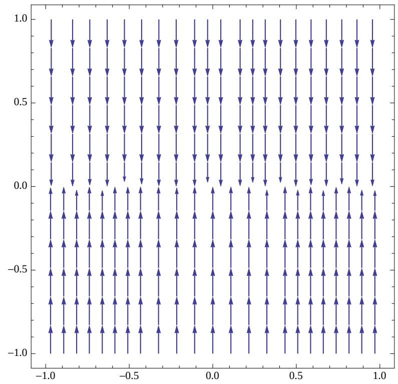


Figure 3.8. The phase portrait of the system: $x' = 0, \quad y' = -y$.

It turns out that the reason this is a "bad" example is because of the eigenvalues of the linear system, which are 0 and -1. In the two "good" examples we have previously seen, the eigenvalues of the linear systems all have non-zero real parts. We will call the origin in the "bad" example as a *non-hyperbolic* equilibrium whereas in the "good" examples as a *hyperbolic* equilibrium. In the next few sections, we will see why this hyperbolicity issue matters.

3.2.0.1. Differentiability. The process of dropping higher order terms is called **linearization**. We will define linearization in a more rigorous way soon so that we can deal with more complicated systems such as $x' = \sin x + \tan y, \quad y' = \tan x + \sin y$ where the meaning of "dropping nonlinear terms" is not as obvious as in previous examples.

In order to make sense of linearization formally, we need to introduce the concept of differentiability in a rigorous way. In single-variable calculus, a function $f(x)$ is differentiable at $x = x_0$ if the limit:

$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

exists, and if this limit exists, it is called the *derivative* of f at x_0 and is denoted by $f'(x_0)$.

Before we introduce the concept of differentiability for multivariable functions, we first reformulate the differentiability of single-variable functions in an alternative, but equivalent, way. Suppose $f'(x_0)$ exists, then:

$$\begin{aligned}
\lim_{x\to x_0} \frac{f(x)-f(x_0)}{x-x_0} &= f'(x_0) \\
\lim_{x\to x_0} \frac{f(x)-f(x_0)}{x-x_0} - f'(x_0) &= 0 \\
\lim_{x\to x_0} \frac{f(x)-f(x_0)-f'(x_0)\cdot(x-x_0)}{x-x_0} &= 0.
\end{aligned}$$

If one defines $L(x) = f(x_0) + f'(x_0)\cdot(x-x_0)$, which is the tangent line to the graph of $f(x)$ at x_0, then:

$$\lim_{x\to x_0} \frac{f(x)-L(x)}{x-x_0} = 0.$$

In other words, $f(x) - L(x)$ goes to 0 *faster* than $x - x_0$ does, and so the tangent line at x_0 is an 'effective' approximation of $f(x)$ provided that x is near x_0.

Conversely, if $f(x)$ can be 'effectively' approximated near x_0 by a straight-line $L(x) = f(x_0) + m(x-x_0)$ through the point $(x_0,\ f(x_0))$, then one can show that $f'(x_0)$ exists and the slope m of the line must be $f'(x_0)$. By 'effective' we mean:

$$\begin{aligned}
\lim_{x\to x_0} \frac{f(x)-L(x)}{x-x_0} &= 0 \\
\text{and so: } \lim_{x\to x_0} \frac{f(x)-f(x_0)-m(x-x_0)}{x-x_0} &= 0 \\
\lim_{x\to x_0} \frac{f(x)-f(x_0)}{x-x_0} - m &= 0.
\end{aligned}$$

Therefore, the limit $\lim_{x\to x_0} \dfrac{f(x)-f(x_0)}{x-x_0}$ exists and is equal to m.

From the above discussion, we see that one can restate the definition of differentiability of single-variable functions as:

$f(x)$ is differentiable at x_0 if and only if there exists a straight-line $L(x) = f(x_0) + m(x-x_0)$ through $(x_0,\ f(x_0))$ such that:

$$\lim_{x\to x_0} \frac{f(x)-L(x)}{x-x_0} = 0.$$

This equivalent form of differentiability can be easily generalized to higher dimensions. As we will mostly work with two dimensional systems for the rest of the course, we will only state the definition of two-variable functions.

Definition 3.4 (Differentiability). A two-variable function $f(x,y)$ is a differentiable at (x_0,y_0) if and only if there exists a function of the form $L(x,y) = f(x_0,y_0) + a(x-x_0) + b(y-y_0)$ such that:

$$\lim_{(x,y)\to(x_0,y_0)} \frac{|f(x,y)-L(x,y)|}{|(x,y)-(x_0,y_0)|} = 0.$$

□

If such an $L(x,y)$ exists, it is necessary that $a = \frac{\partial f}{\partial x}(x_0,y_0)$ and $b = \frac{\partial f}{\partial y}(x_0,y_0)$. To show the former, we fix $y = y_0$ and consider:

$$\lim_{x\to x_0} \frac{|f(x,y_0) - L(x,y_0)|}{|(x,y_0)-(x_0,y_0)|} = 0$$
$$\lim_{x\to x_0} \frac{|f(x,y_0) - f(x_0,y_0) - a(x-x_0)|}{|x-x_0|} = 0$$
$$\lim_{x\to x_0} \left|\frac{f(x,y_0) - f(x_0,y_0)}{x-x_0} - a\right| = 0$$

By the definition of partial derivatives:

$$\frac{\partial f}{\partial x}(x_0,y_0) := \lim_{x\to x_0} \frac{f(x,y_0) - f(x_0,y_0)}{x-x_0}$$

it exists and is equal to a. Similarly, one can also show $b = \frac{\partial f}{\partial y}(x_0,y_0)$.

If such an L exists, we must have

$$L(x,y) = f(x_0,y_0) + \frac{\partial f}{\partial x}(x_0,y_0)\cdot(x-x_0) + \frac{\partial f}{\partial y}(x_0,y_0)\cdot(y-y_0).$$

From your multivariable calculus class, the graph of L is the tangent plane to the function $f(x,y)$ at the point $(x,y) = (x_0,y_0)$. Therefore, a two-variable function $f(x,y)$ is said to be differentiable at (x_0,y_0) if the tangent plane at (x_0,y_0) 'effectively' approximates the function around (x_0,y_0), in a sense that the gap between the graph and the tangent plane is going to zero faster than $\sqrt{(x-x_0)^2+(y-y_0)^2}$ does.

Although it may be complicated to check differentiability for multivariable functions since it involves evaluation of a multivariable limit, fortunately one can show all C^1 functions are differentiable (but not vice versa):

Theorem 3.5 (C^1 implies Differentiability). *Let Ω be an open domain in $\mathbb{R}^2$. If a function $f : \Omega \to \mathbb{R}^m$ is C^1 on Ω, then it is differentiable at every point on Ω (or one can simply say differentiable on Ω).*

Proof. Take an arbitrary point $(x_0,y_0) \in \Omega$. We consider:

$$\begin{aligned} f(x,y) - f(x_0,y_0) &= f(x,y) - f(x_0,y) + f(x_0,y) - f(x_0,y_0) \\ &= \frac{\partial f}{\partial x}(\xi,y)\cdot(x-x_0) + \frac{\partial f}{\partial y}(x_0,\zeta)\cdot(y-y_0) \quad \text{(mean-value theorem)} \end{aligned}$$

Here ξ is some number between x and x_0, and ζ is some number between y and y_0.

Define

$$L(x,y) := f(x_0,y_0) + \frac{\partial f}{\partial x}(x_0,y_0)\cdot(x-x_0) + \frac{\partial f}{\partial y}(x_0,y_0)\cdot(y-y_0),$$

then

$$\begin{aligned} &\frac{|f(x,y) - L(x,y)|}{|(x,y)-(x_0,y_0)|} \\ &= \frac{\left|\left(\frac{\partial f}{\partial x}(\xi,y) - \frac{\partial f}{\partial x}(x_0,y_0)\right)\cdot(x-x_0) + \left(\frac{\partial f}{\partial y}(x_0,\zeta) - \frac{\partial f}{\partial y}(x_0,y_0)\right)\cdot(y-y_0)\right|}{|(x,y)-(x_0,y_0)|} \\ &\le \frac{\left|\frac{\partial f}{\partial x}(\xi,y) - \frac{\partial f}{\partial x}(x_0,y_0)\right||x-x_0| + \left|\frac{\partial f}{\partial y}(x_0,\zeta) - \frac{\partial f}{\partial y}(x_0,y_0)\right||y-y_0|}{\sqrt{(x-x_0)^2+(y-y_0)^2}} \\ &\le \left|\frac{\partial f}{\partial x}(\xi,y) - \frac{\partial f}{\partial x}(x_0,y_0)\right| + \left|\frac{\partial f}{\partial y}(x_0,\zeta) - \frac{\partial f}{\partial y}(x_0,y_0)\right|. \end{aligned}$$

The last inequality follows from the fact that $|x - x_0| \leq \sqrt{(x-x_0)^2 + (y-y_0)^2}$ and $|y - y_0| \leq \sqrt{(x-x_0)^2 + (y-y_0)^2}$.

As $x \to x_0$, the quantity ξ which is between x and x_0 also approaches x_0. Similarly, $\zeta \to y_0$ as $y \to y_0$. By continuity of the partials $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$, we have $\frac{\partial f}{\partial x}(\xi, y) \to \frac{\partial f}{\partial x}(x_0, y_0)$ and $\frac{\partial f}{\partial y}(x_0, \zeta) \to \frac{\partial f}{\partial y}(x_0, y_0)$ as $(x, y) \to (x_0, y_0)$. By squeezing principle applied to the above inequality, we proved:

$$\lim_{(x,y)\to(x_0,y_0)} \frac{|f(x,y) - L(x,y)|}{|(x,y) - (x_0,y_0)|} = 0.$$

Therefore, f is differentiable at (x_0, y_0). Since (x_0, y_0) is arbitrarily chosen, f is differentiable on Ω.

□

Remark 3.6. C^1 implies differentiability, but not all differentiable functions are C^1. Readers may consult the classic book *Calculus on Manifolds* by Spivak for counter-examples and for more thorough discussions about differentiability. □

Remark 3.7. C^1 functions are commonly called *continuously differentiable* functions. □

Next we express the definition differentiability using little-o notations.

Definition 3.8 (Big-O and little-o)**.** Given two real-valued multivariable functions g, h defined around a point $\mathbf{x}_0$, we say:

- $g = O(h)$ if there exist constants $C, \varepsilon > 0$ such that $|g(\mathbf{x})| \leq C|h(\mathbf{x})|$ for any $x \in B_\varepsilon(\mathbf{x}_0)$.
- $g = o(h)$ if $\frac{g(\mathbf{x})}{h(\mathbf{x})} \to 0$ as $\mathbf{x} \to \mathbf{x}_0$.

Alternatively, $g = O(h)$ and $g = o(h)$ can be denoted by $g \in O(h)$ and $g \in o(h)$ respectively. □

When we say $k = f + o(h)$, we mean $k = f + g$ for some $g \in o(h)$. Similar for $k = f + O(h)$.

Using the little-o notation, one can then restate the differentiability definition in the following equivalent way:

A real-valued multivariable function $f(x, y)$ is differentiable at (x_0, y_0) if and only if both $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ exist at (x_0, y_0) and:

$$\begin{aligned} &f(x,y) \\ &= \underbrace{f(x_0,y_0) + f_x(x_0,y_0)\cdot(x-x_0) + f_y(x_0,y_0)\cdot(y-y_0)}_{L(x,y)} + o\left(\sqrt{(x-x_0)^2+(y-y_0)^2}\right). \end{aligned}$$

3.2.0.2. Linearized systems. A vector field $\mathbf{F} : \mathbb{R}^2 \to \mathbb{R}^2$ is said to be differentiable at (x_0, y_0) if each component of $\mathbf{F}$ is differentiable at (x_0, y_0). Denote the components of $\mathbf{F}$ by: $\mathbf{F}(x, y) = \begin{bmatrix} u(x,y) \\ v(x,y) \end{bmatrix}$. Suppose $\mathbf{F}(x, y)$ is differentiable at (x_0, y_0). For simplicity, we denote $\mathbf{x} = (x, y)$ and $\mathbf{x}_0 = (x_0, y_0)$ then:

$$\begin{aligned} u(\mathbf{x}) &= u(\mathbf{x}_0) + u_x(\mathbf{x}_0)\cdot(x-x_0) + u_y(\mathbf{x}_0)\cdot(y-y_0) + o\left(|\mathbf{x}-\mathbf{x}_0|\right) \\ v(\mathbf{x}) &= v(\mathbf{x}_0) + v_x(\mathbf{x}_0)\cdot(x-x_0) + v_y(\mathbf{x}_0)\cdot(y-y_0) + o\left(|\mathbf{x}-\mathbf{x}_0|\right) \end{aligned}$$

Rewrite them in a vector form:

$$\mathbf{F}(\mathbf{x}) = \mathbf{F}(\mathbf{x}_0) + \begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix}_{\mathbf{x}=\mathbf{x}_0} \cdot (\mathbf{x} - \mathbf{x}_0) + \begin{bmatrix} o\left(|\mathbf{x}-\mathbf{x}_0|\right) \\ o\left(|\mathbf{x}-\mathbf{x}_0|\right) \end{bmatrix}.$$

When $\mathbf{x}$ is very close to $\mathbf{x}_0$, the little-o terms can be neglected and the vector field $\mathbf{F}$ is approximately equal to:

$$\mathbf{F}(\mathbf{x}) \approx \mathbf{F}(\mathbf{x}_0) + \begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix}_{\mathbf{x}=\mathbf{x}_0} \cdot (\mathbf{x} - \mathbf{x}_0).$$

If we let $\mathbf{x}_0 = \mathbf{x}^*$ which is an equilibrium point of the system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$, then near the equilibrium point $\mathbf{x}^*$, the nonlinear system can be approximated by the linear system:

$$\underbrace{(\mathbf{x} - \mathbf{x}^*)'}_{=\mathbf{x}'} = \underbrace{\begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix}_{\mathbf{x}=\mathbf{x}^*}}_{\text{a constant matrix}} \cdot (\mathbf{x} - \mathbf{x}^*).$$

In this connection, we define:

Definition 3.9 (Jacobian Matrix). Let $\mathbf{F} : \mathbb{R}^2 \to \mathbb{R}^2$ be a differentiable vector field whose components are given by $\mathbf{F}(x, y) = \begin{bmatrix} u(x,y) \\ v(x,y) \end{bmatrix}$. The Jacobian matrix of $\mathbf{F}$ at a point $\mathbf{x}_0$ is defined as:

$$D\mathbf{F}_{\mathbf{x}_0} := \begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix}_{\mathbf{x}=\mathbf{x}_0}.$$

□

Definition 3.10 (Linearization). Given an autonomous system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ in $\mathbb{R}^d$ with $\mathbf{x}^*$ as one of the equilibrium points, its **linearization** at $\mathbf{x}^*$, or its **linearized system** at $\mathbf{x}^*$, is the system:

$$\mathbb{X}' = (D\mathbf{F}_{\mathbf{x}^*})\mathbb{X}$$

where $\mathbb{X} := \mathbf{x} - \mathbf{x}^*$, and $D\mathbf{F}_{\mathbf{x}^*}$ is the Jacobian matrix at $\mathbf{x}^*$. □

The linearized system at an equilibrium point $\mathbf{x}^*$ of a nonlinear system reveals the local behaviors of the phase portrait near $\mathbf{x}^*$. The linearized system is much easier to study since its stability is determined by the eigenvalues of the matrix $D\mathbf{F}_{\mathbf{x}^*}$. In the next few sections, we will prove, with rigorous proofs, that a nonlinear system does resemble its linearization near the equilibrium points in a number of ways including stability and the phase portrait type. However, this resemblance is subject to one crucial condition, namely the equilibrium point $\mathbf{x}^*$ has to be *hyperbolic*:

Definition 3.11 (Hyperbolic Equilibrium Point). An equilibrium point $\mathbf{x}^*$ of a system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ is said to be **hyperbolic** if all eigenvalues of the Jacobian matrix $D\mathbf{F}_{\mathbf{x}^*}$ at $\mathbf{x}^*$ have non-zero real parts. □

Example 3.3. The system

$$\begin{aligned} x' &= y \\ y' &= \cos x \end{aligned}$$

has infinitely many equilibrium points, namely $(k\pi + \frac{\pi}{2}, 0)$ for any integer k.

The Jacobian matrix of the vector field $\mathbf{F}(x, y) = \begin{bmatrix} y \\ \cos x \end{bmatrix}$ at an arbitrary point (x, y) is given by:

$$D\mathbf{F}_{(x,y)} = \begin{bmatrix} 0 & 1 \\ -\sin x & 0 \end{bmatrix},$$

and so at the equilibrium points $(k\pi + \frac{\pi}{2}, 0)$ it is given by:

$$D\mathbf{F}_{(k\pi+\frac{\pi}{2},0)} = \begin{bmatrix} 0 & 1 \\ (-1)^{k-1} & 0 \end{bmatrix}.$$

When k is odd, the eigenvalues of $D\mathbf{F}_{(k\pi+\frac{\pi}{2},0)}$ are -1 and 1. Therefore, $(k\pi + \frac{\pi}{2}, 0)$ are hyperbolic equilibrium points when k is odd.

However, when k is even, the eigenvalues of $D\mathbf{F}_{(k\pi+\frac{\pi}{2},0)}$ are $\pm i$, whose real parts are both 0. Therefore, $(k\pi + \frac{\pi}{2}, 0)$ are non-hyperbolic equilibrium points when k is even. □

Exercise 3.7. For each of the following systems, find all equilibrium point(s) and determine whether they are hyperbolic or not.

(1) $x' = \sin x, \quad y' = \cos y$

(2) $x' = x + y^2, \quad y' = 2y$

(3) $x' = \log(1 + y^2), \quad y' = e^x - 1$

Exercise 3.8. Consider the system

$$\begin{aligned} x' &= x^2 + y \\ y' &= x - y + a \end{aligned}$$

where a is a parameter.

(1) Find all equilibrium point(s). Express your answers in terms of a.

(2) Determine whether each equilibrium point is hyperbolic.

3.3. Poincaré-Lyapunov's Theorem

A linear system $\mathbf{x}' = A\mathbf{x}$ is asymptotically stable if and only if all eigenvalues of A have strictly negative real parts. In this section, we will show that a nonlinear system has an asymptotically stable equilibrium if its linearized system at this point is asymptotically stable (or equivalently, the eigenvalues of the Jacobian matrix all have negative real parts). On the other hand, if all eigenvalues of the Jacobian matrix have (strictly) positive real parts, one can also show that the nonlinear system is unstable. This result is known as the Poincaré-Lyapunov's Theorem.

As such, linearization is a very effective method to test whether a nonlinear system is stable or not. However, this test has a limitation, as it fails when any one of the eigenvalues of the Jacobian matrix has zero real part.

Theorem 3.12 (Poincaré-Lyapunov's Theorem: stable case)**.** *Let $\mathbf{F} : \mathbb{R}^2 \to \mathbb{R}^2$ be a C^1 vector field. Consider the autonomous system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$. Suppose $\mathbf{x}^*$ is an equilibrium point of the system and that all eigenvalues of $D\mathbf{F}_{\mathbf{x}^*}$ have (strictly) negative real parts, then the equilibrium point $\mathbf{x}^*$ of the nonlinear system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ is asymptotically stable.*

To prepare for the proof of the theorem, we first perform a linear change of variables such that the matrix of the linearized system becomes one of the canonical forms. We will show that the stability of the transformed system is equivalent to the stability of the original system. As such, it suffices to consider the case when the Jacobian matrix is in a canonical form.

Given a differentiable vector field $\mathbf{F} : \mathbb{R}^2 \to \mathbb{R}^2$ with an equilibrium point $\mathbf{x}^*$, then the differentiability condition asserts that $\mathbf{F}(\mathbf{x}) = (D\mathbf{F}_{\mathbf{x}^*})(\mathbf{x} - \mathbf{x}^*) + o(|\mathbf{x} - \mathbf{x}^*|)$ where $o(|\mathbf{x} - \mathbf{x}^*|)$ represents a vector field whose components are all in $o(|\mathbf{x} - \mathbf{x}^*|)$.

From Chapter 1, the 2×2 matrix $D\mathbf{F}_{\mathbf{x}^*}$ admits a canonical decomposition:

$$D\mathbf{F}_{\mathbf{x}^*} = PKP^{-1}$$

where P is invertible and K is a one of the following canonical forms:

$$\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}, \quad \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}, \quad \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}$$

Therefore, the system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ can be written as:

$$\begin{aligned} \mathbf{x}' &= PKP^{-1}(\mathbf{x} - \mathbf{x}^*) + o(|\mathbf{x} - \mathbf{x}^*|) \\ P^{-1}\mathbf{x}' &= KP^{-1}(\mathbf{x} - \mathbf{x}^*) + P^{-1} \cdot o(|\mathbf{x} - \mathbf{x}^*|) \\ \left(P^{-1}(\mathbf{x} - \mathbf{x}^*)\right)' &= K\left(P^{-1}(\mathbf{x} - \mathbf{x}^*)\right) + P^{-1} \cdot o(|\mathbf{x} - \mathbf{x}^*|) \\ \mathbf{y}' &= K\mathbf{y} + P^{-1} \cdot o(|\mathbf{x} - \mathbf{x}^*|) && (\text{Let } \mathbf{y} := P^{-1}(\mathbf{x} - \mathbf{x}^*)) \end{aligned}$$

As $\mathbf{x} \to \mathbf{x}^*$, $\mathbf{y} \to \mathbf{0}$, and it can be verified that:

$$\begin{aligned} P^{-1} \cdot o(|\mathbf{x} - \mathbf{x}^*|) &= P^{-1} \cdot o(|P\mathbf{y}|) \\ &= P^{-1} \cdot o(\|P\| \, |\mathbf{y}|) && (\text{since } |P\mathbf{y}| \leq \|P\| \, |\mathbf{y}|) \\ &= o(|\mathbf{y}|) && (\text{see exercise below}) \end{aligned}$$

Exercise 3.9. Show, from the definition, that $P^{-1} \cdot o(|\mathbf{x} - \mathbf{x}^*|) = o(|\mathbf{y}|)$.

Therefore, the nonlinear system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ is equivalent to $\mathbf{y}' = K\mathbf{y} + o(|\mathbf{y}|)$ via the change of variable $\mathbf{y} = P^{-1}(\mathbf{x} - \mathbf{x}^*)$. Needless to say, the system $\mathbf{y}' = K\mathbf{y} + o(|\mathbf{y}|)$ is easier to work with since the matrix K is a canonical form. The following lemma shows that the stability of $\mathbf{x}^*$ of the $\mathbf{x}$-system is equivalent to the stability of $\mathbf{0}$ of the $\mathbf{y}$-system.

Therefore, it suffices to study systems whose linearization is in one of the canonical forms. The transformation $\mathbf{y} = P^{-1}(\mathbf{x} - \mathbf{x}^*)$ consists of a translation followed by a linear map represented by P^{-1}. Clearly, translation preserves stability and asymptotic stability (Exercise 3.10). We next show that an invertible linear map transforms an asymptotically stable $\mathbf{0}$ to an asymptotically stable $\mathbf{0}$ in the new system as well.

Lemma 3.13. *Let $\mathbf{0}$ be an equilibrium point of a system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ where $\mathbf{F}$ is differentiable. Suppose P is an invertible matrix, then via the change of variables $\mathbf{y} = P^{-1}\mathbf{x}$, the equilibrium point $\mathbf{0}$ is asymptotically stable in the $\mathbf{x}$-system if and only if $\mathbf{0}$ is asymptotically stable in the $\mathbf{y}$-system $\mathbf{y}' = P^{-1}\mathbf{F}(P\mathbf{y})$.*

Proof. First suppose $\mathbf{0}$ is asymptotically stable in the $\mathbf{x}$-system. We want to show $\mathbf{0}$ is asymptotically stable in the $\mathbf{y}$-system. Given any $\varepsilon > 0$, then $P\left(B_\varepsilon^{\mathbf{y}}(\mathbf{0})\right)$ is an open ellipse in the $\mathbf{x}$-plane containing $\mathbf{0}$. Here the superscript $\mathbf{y}$ of $B_\varepsilon^{\mathbf{y}}(\mathbf{0})$ means the ball is in the $\mathbf{y}$-plane. Then, there exists another $\varepsilon' > 0$ such that the ball $B_{\varepsilon'}^{\mathbf{x}}(\mathbf{0}) \subset P\left(B_\varepsilon^{\mathbf{y}}(\mathbf{0})\right)$. By the asymptotic stability of $\mathbf{0}$ in the $\mathbf{x}$-system, there exists $\delta' > 0$ such that whenever $\mathbf{x}_0 \in B_{\delta'}^{\mathbf{x}}(\mathbf{0})$, we have $\varphi_t^{\mathbf{x}}(\mathbf{x}_0) \in B_{\varepsilon'}^{\mathbf{x}}(\mathbf{0})$ for any $t \geq 0$, and as $t \to +\infty$, we have $\varphi_t^{\mathbf{x}}(\mathbf{x}_0) \to \mathbf{0}$. Here $\varphi_t^{\mathbf{x}}$ denotes the flow of the $\mathbf{x}$-system. Transform the ball $B_{\delta'}^{\mathbf{x}}(\mathbf{0})$ to the $\mathbf{y}$-system: one get an open ellipse $P^{-1}\left(B_{\delta'}^{\mathbf{x}}(\mathbf{0})\right)$ in the $\mathbf{y}$-plane containing $\mathbf{0}$. Then, there exists $\delta > 0$ such that $B_\delta^{\mathbf{y}}(\mathbf{0}) \subset P^{-1}\left(B_{\delta'}^{\mathbf{x}}(\mathbf{0})\right)$. Whenever, $\mathbf{y}_0 \in B_\delta^{\mathbf{y}}(\mathbf{0})$, we have $\mathbf{y}_0 \in P^{-1}\left(B_{\delta'}^{\mathbf{x}}(\mathbf{0})\right)$ and so $P\mathbf{y}_0 \in B_{\delta'}^{\mathbf{x}}(\mathbf{0})$. Therefore, by the choice of δ' we have $\varphi_t^{\mathbf{x}}(P\mathbf{y}_0) \in B_{\varepsilon'}^{\mathbf{x}}(\mathbf{0})$ for all $t \geq 0$, and as $t \to +\infty$, we have $\varphi_t^{\mathbf{x}}(P\mathbf{y}_0) \to \mathbf{0}$. Since $P\mathbf{y}_0$ in the $\mathbf{x}$-plane corresponds to $\mathbf{y}_0$ in the $\mathbf{y}$-plane, $\varphi_t^{\mathbf{x}}(P\mathbf{y}_0)$ in the $\mathbf{x}$-plane corresponds to $\varphi_t^{\mathbf{y}}(\mathbf{y}_0)$ in the $\mathbf{y}$-plane. In other words, $\varphi_t^{\mathbf{y}}(\mathbf{y}_0) = P^{-1}\left(\varphi_t^{\mathbf{x}}(P\mathbf{y}_0)\right)$ for all $t \geq 0$. Since $\varphi_t^{\mathbf{x}}(P\mathbf{y}_0) \in B_{\varepsilon'}^{\mathbf{x}}(\mathbf{0}) \subset P\left(B_\varepsilon^{\mathbf{y}}(\mathbf{0})\right)$, we have $\varphi_t^{\mathbf{y}}(\mathbf{y}_0) = P^{-1}\left(\varphi_t^{\mathbf{x}}(P\mathbf{y}_0)\right) \in B_\varepsilon^{\mathbf{y}}(\mathbf{0})$. It completes the part that $\mathbf{0}$ is stable in the $\mathbf{y}$-system. To show it is asymptotically stable, we let $t \to +\infty$, then by $\varphi_t^{\mathbf{x}}(P\mathbf{y}_0) \to \mathbf{0}$, we have $\varphi_t^{\mathbf{y}}(\mathbf{y}_0) = P^{-1}\left(\varphi_t^{\mathbf{x}}(P\mathbf{y}_0)\right) \to \mathbf{0}$.

The proof of the converse is almost the same by considering $\mathbf{x} = P\mathbf{y}$. □

Exercise 3.10. Let $\mathbf{x}^*$ be an equilibrium point of a system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$, and consider the translation $\mathbb{X} := \mathbf{x} - \mathbf{x}^*$. Show that $\mathbf{x}^*$ is asymptotically stable in the $\mathbf{x}$-system if and only if $\mathbf{0}$ is asymptotically stable in the $\mathbb{X}$-system $\mathbb{X}' = \mathbf{F}(\mathbb{X} + \mathbf{x}^*)$.

Combining Lemma 3.13 and Exercise 3.10, we have proved:

Corollary 3.14. *Let $\mathbf{x}^*$ be an equilibrium point of a system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$, and consider the affine linear transformation $\mathbf{y} = P^{-1}(\mathbf{x} - \mathbf{x}^*)$. Then, $\mathbf{x}^*$ is asymptotically stable in the $\mathbf{x}$-system if and only if $\mathbf{0}$ is asymptotically stable in the $\mathbf{y}$-system $\mathbf{y}' = P^{-1}\mathbf{F}(\mathbf{x}^* + P\mathbf{y})$.*

Now we get back to the discussion of linearizations. Given a differentiable vector $\mathbf{F} : \mathbb{R}^2 \to \mathbb{R}^2$ with equilibrium point $\mathbf{x}^*$. The system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ has the following asymptotic form:

$$\mathbf{x}' = (D\mathbf{F}_{\mathbf{x}^*})(\mathbf{x} - \mathbf{x}^*) + o(|\mathbf{x} - \mathbf{x}^*|).$$

The 2×2 Jacobian matrix $D\mathbf{F}_{\mathbf{x}^*}$ admits a canonical decomposition $D\mathbf{F}_{\mathbf{x}^*} = PKP^{-1}$ where P is invertible and K is one of the canonical forms, then under the affine linear transformation $\mathbf{y} = P^{-1}(\mathbf{x} - \mathbf{x}^*)$, the system is transformed into the form:

$$\mathbf{y}' = K\mathbf{y} + o(|\mathbf{y}|)$$

as per the discussion preceding Lemma 3.13.

Since $\mathbf{x}^*$ is asymptotically stable in the $\mathbf{x}$-system if and only if $\mathbf{0}$ is asymptotically stable in the $\mathbf{y}$-system, from now on we can focus on the $\mathbf{y}$-system (whose linearization at $\mathbf{0}$ is given by K).

Proof of Theorem 3.12. Suppose the Jacobian matrix of $\mathbf{F}$ at $\mathbf{x}^*$ has a canonical form decomposition:

$$D\mathbf{F}_{\mathbf{x}^*} = PKP^{-1}$$

where P is invertible and K is one of the canonical forms. The matrix K has the same eigenvalue as $D\mathbf{F}_{\mathbf{x}^*}$ and is one of the following forms according to whether the eigenvalues of $D\mathbf{F}_{\mathbf{x}^*}$ is distinct real, complex or repeated:

$$\begin{bmatrix} -\lambda_1 & 0 \\ 0 & -\lambda_2 \end{bmatrix}, \quad \begin{bmatrix} -\alpha & \beta \\ -\beta & -\alpha \end{bmatrix}, \quad \begin{bmatrix} -\lambda & 1 \\ 0 & -\lambda \end{bmatrix}.$$

By our assumption in the theorem, λ_1, λ_2, α and λ are all positive real numbers. We divide the proof into three cases, each of which correspond to K being one of the above canonical forms.

Recall that to show $\mathbf{x}^*$ is asymptotically stable, it suffices to show $\mathbf{0}$ is asymptotically stable in the system of the form:

$$\mathbf{y}' = K\mathbf{y} + o(|\mathbf{y}|).$$

Case 1: $K = \begin{bmatrix} -\lambda_1 & 0 \\ 0 & -\lambda_2 \end{bmatrix}$

We assume without loss of generality that $-\lambda_1 < -\lambda_2 < 0$. We first show $\mathbf{0}$ is stable in the $\mathbf{y}$-system. We first show that Consider:

$$\begin{aligned}
\frac{d}{dt}|\mathbf{y}|^2 &= 2\mathbf{y}\cdot\mathbf{y}' \\
&= 2\mathbf{y}\cdot(K\mathbf{y} + o(|\mathbf{y}|)) \\
&= -2(\lambda_1 y_1^2 + \lambda_2 y_2^2) + 2\mathbf{y}\cdot o(|\mathbf{y}|) \\
&\le -2\lambda_2(y_1^2 + y_2^2) + o(|\mathbf{y}|^2) \\
&= -2\lambda_2|\mathbf{y}|^2 + o(|\mathbf{y}|^2) \\
&= (-2\lambda_2 + o(1))|\mathbf{y}|^2.
\end{aligned}$$

Hence, by the definition of little-oh, there exists $\delta > 0$, such that whenever $|\mathbf{y}| < \delta$, we have

$$-2\lambda_2 + o(1) < -2\lambda_2 + \lambda_2 = -\lambda_2 < 0.$$

Therefore, when $\mathbf{y}_0 \in B_\delta(\mathbf{0})$, we have $\frac{d}{dt}|\varphi_t(\mathbf{y}_0)|^2 \le -\lambda_2|\varphi_t(\mathbf{y}_0)|^2 < 0$ and so $|\varphi_t(\mathbf{y}_0)|$ is strictly decreasing for $t \ge 0$. In other words, $\varphi_t(\mathbf{y}_0) \in B_\delta(\mathbf{0})$ for all $t \ge 0$ and $B_\delta(\mathbf{0})$ is a 'trapping set' of the system in a sense that solution curves that start inside the ball will be trapped inside the ball forever. It follows immediately from the definition that $\mathbf{0}$ is stable.

The fact that $\mathbf{0}$ is asymptotically stable follows from:

$$\begin{aligned}
\frac{d}{dt}|\varphi_t(\mathbf{y}_0)|^2 &\le -\lambda_2|\varphi_t(\mathbf{y}_0)|^2 \\
\frac{d}{dt}\left(e^{\lambda_2 t}|\varphi_t(\mathbf{y}_0)|^2\right) &= e^{\lambda_2 t}\frac{d}{dt}|\varphi_t(\mathbf{y}_0)|^2 + \lambda_2 e^{\lambda_2 t}|\varphi_t(\mathbf{y}_0)|^2 \\
&= e^{\lambda_2 t}\left(\frac{d}{dt}|\varphi_t(\mathbf{y}_0)|^2 + \lambda_2|\varphi_t(\mathbf{y}_0)|^2\right) \\
&\le e^{\lambda_2 t}\cdot 0 = 0 \\
e^{\lambda_2 t}|\varphi_t(\mathbf{y}_0)|^2 &\le e^{\lambda_2\cdot 0}|\varphi_0(\mathbf{y}_0)|^2 = |\mathbf{y}_0|^2
\end{aligned}$$

for any $t \ge 0$. Therefore,

$$|\varphi_t(\mathbf{y}_0)| \le |\mathbf{y}_0|\, e^{-\frac{\lambda_2}{2}t} \to 0 \quad \text{as } t \to +\infty.$$

Case 2: $K = \begin{bmatrix} -\alpha & \beta \\ -\beta & -\alpha \end{bmatrix}$

Similar to the previous case, we show that $\mathbf{0}$ in the $\mathbf{y}$-system:

$$\mathbf{y}' = K\mathbf{y} + o\left(|\mathbf{y}|\right)$$

is asymptotically stable.

$$\begin{aligned} \frac{d}{dt}\left|\mathbf{y}\right|^2 &= 2\mathbf{y}\cdot\mathbf{y}' = 2\mathbf{y}\cdot\left(K\mathbf{y} + o\left(|\mathbf{y}|\right)\right) \\ &= -2\alpha\left(y_1^2 + y_2^2\right) + 2\mathbf{y}\cdot o\left(|\mathbf{y}|\right) \\ &\leq -2\alpha\left|\mathbf{y}\right|^2 + o(|\mathbf{y}|^2) = \left(-2\alpha + o(1)\right)\left|\mathbf{y}\right|^2. \end{aligned}$$

Using this inequality, one can proceed as in Case 1 (with α in place of λ_2) to show $\mathbf{0}$ is asymptotically stable.

Case 3: $K = \begin{bmatrix} -\lambda & 1 \\ 0 & -\lambda \end{bmatrix}$

In this case, we make a further change of variables. Let:

$$\mathbf{z} := \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{\lambda} \end{bmatrix}\mathbf{y}.$$

Denote the components of $\mathbf{z}$ by $\mathbf{z} := \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$. Then, it can be directly verified that:

$$\begin{aligned} \frac{d}{dt}\left|\mathbf{z}\right|^2 &= \frac{d}{dt}\left(y_1^2 + \lambda y_2^2\right) \\ &= -2\lambda\left(y_1^2 + \lambda y_2^2\right) + o(|\mathbf{y}|^2) \\ &= -2\lambda\left|\mathbf{z}\right|^2 + o(|\mathbf{z}|^2). \end{aligned}$$

Again, one apply a similar argument as in Cases 1 and 2 to show $\mathbf{0}$ is asymptotically stable in the $\mathbf{z}$-system. Finally, Lemma 3.13 shows $\mathbf{0}$ is also asymptotically stable in the $\mathbf{y}$-system.

□

Let $\mathbf{F} : \mathbb{R}^2 \to \mathbb{R}^2$ be a C^1 vector field. Consider the two systems:

$$\mathbf{x}'(t) = \mathbf{F}(\mathbf{x}(t)), \quad \mathbf{y}'(t) = -\mathbf{F}(\mathbf{y}(t)).$$

Graphically, the vector fields of the $\mathbf{x}$- and $\mathbf{y}$-systems are in the opposite direction, and so the flow of the $\mathbf{y}$-system should be the backward flow of the $\mathbf{x}$-system. We will use this observation to establish the linearization test when the Jacobian matrix has positive real parts. Precisely, given that $\mathbf{x}^*$ is an equilibrium point of the $\mathbf{x}$-system and that all eigenvalues of $D\mathbf{F}_{\mathbf{x}^*}$ have positive real parts. Then, all eigenvalues of $D(-\mathbf{F})_{\mathbf{x}^*}$ have negative real parts, and by Theorem 3.12, the $\mathbf{y}$-system is asymptotically stable around the equilibrium point $\mathbf{x}^*$. Solution curves near $\mathbf{x}^*$ in the $\mathbf{y}$-system are tending *towards* $\mathbf{x}^*$, and so those in the $\mathbf{x}$-system are tending *away from* $\mathbf{x}^*$. This establish instability of $\mathbf{x}^*$ in the $\mathbf{x}$-system. Let's state and prove this result rigorously:

Theorem 3.15 (Poincaré-Lyapunov's Theorem: unstable case). *Let $\mathbf{F} : \mathbb{R}^2 \to \mathbb{R}^2$ be a C^1 vector field. Consider the autonomous system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$. Suppose $\mathbf{x}^*$ is an equilibrium point of the system and that all eigenvalues of $D\mathbf{F}_{\mathbf{x}^*}$ have (strictly) positive real parts, then the equilibrium point $\mathbf{x}^*$ of the nonlinear system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ is unstable.*

Proof. Let φ_t be the flow of the system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$, and ψ_t be the flow of the system $\mathbf{y}'(t) = -\mathbf{F}(\mathbf{y}(t))$, i.e. the backward flow of φ_t. Then, for any $t \in \mathbb{R}$ and $\mathbf{x}_0 \in \mathbb{R}^2$ such that $\varphi_t(\mathbf{x}_0)$ is defined, we have $\psi_{-t}(\mathbf{x}_0) = \varphi_t(\mathbf{x}_0)$.

The linearization of the $\mathbf{y}$-system at $\mathbf{x}^*$ is given by the matrix $-D\mathbf{F}_{\mathbf{x}^*}$, whose eigenvalues have negative real parts by the given condition. Therefore, by Theorem 3.12, the equilibrium point $\mathbf{x}^*$ is asymptotically stable in the $\mathbf{y}$-system. In particular, one can pick a $\mathbf{x}_0 \in \mathbb{R}^2$, sufficiently close to $\mathbf{x}^*$ (but $\mathbf{x}_0 \neq \mathbf{x}^*$), such that $|\psi_t(\mathbf{x}_0) - \mathbf{x}^*| \to 0$ as $t \to \infty$.

Next we verify that $\mathbf{x}^*$ is unstable in the $\mathbf{x}$-system from the definition. Fix an $\varepsilon > 0$ such that $\mathbf{x}_0 \notin B_\varepsilon(\mathbf{x}^*)$. For any $\delta > 0$, since $\psi_t(\mathbf{x}_0) \to \mathbf{x}^*$ as $t \to +\infty$, there exists a sufficiently large $T > 0$ such that $\psi_T(\mathbf{x}_0) \in B_\delta(\mathbf{x}^*)$. Then, by flowing along the $\mathbf{x}$-system from $\psi_T(\mathbf{x}_0)$ for T unit time forward, one should expect to get back to $\mathbf{x}_0$ which is outside the ε-ball $B_\varepsilon(\mathbf{x}^*)$. Precisely, we have:

$$\varphi_T(\psi_T(\mathbf{x}_0)) = \psi_{-T}(\psi_T(\mathbf{x}_0)) = \psi_{-T+T}(\mathbf{x}_0) = \psi_0(\mathbf{x}_0) = \mathbf{x}_0.$$

To summarize, there exists an $\varepsilon > 0$ such that for any $\delta > 0$, one can find a point $\mathbf{z}_0 := \psi_T(\mathbf{x}_0) \in B_\delta(\mathbf{x}^*)$ such that $\varphi_T(\mathbf{z}_0) = \mathbf{x}_0 \notin B_\varepsilon(\mathbf{x}^*)$, which is exactly the definition of instability of $\mathbf{x}^*$ for the $\mathbf{x}$-system.

□

To sum up, given a planar nonlinear system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ with a hyperbolic equilibrium point $\mathbf{x}^*$, one can determine the stability of $\mathbf{x}^*$ by finding eigenvalues of $D\mathbf{F}_{\mathbf{x}^*}$. If they are both negative, then $\mathbf{x}^*$ is asymptotically stable. If they are both positive, then it is unstable. It is natural to ask whether the phase portrait will resemble a saddle near $\mathbf{x}^*$ if one of the eigenvalues is positive and another is negative, as one saw in the linear case. The answer is affirmative, but with a much more delicate proof than Theorems 3.12 and 3.15. It is a consequence of the Stable Curve Theorem which will be discussed in the next section.

3.4. Stable Curve Theorem

In this section, we will settle the last (and the most difficult) case of hyperbolic equilibrium point, namely the case when the Jacobian matrix has a mix of eigenvalue signs. We will prove that the phase portrait resembles a saddle and hence is unstable. It is a consequence of the following celebrated theorem:

Theorem 3.16 (Stable Curve Theorem)**.** *Let $\mathbf{F} : \mathbb{R}^2 \to \mathbb{R}^2$ be a C^1 vector field. Consider the nonlinear planar system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ with φ_t as its flow. Suppose $\mathbf{x}^*$ is an equilibrium point and the Jacobian matrix $D\mathbf{F}_{\mathbf{x}^*}$ has a positive eigenvalue λ and a negative eigenvalue $-\mu$. Then:*

- *there exists a point $\mathbf{x}_s$ near $\mathbf{x}^*$ such that $\varphi_t(\mathbf{x}_s) \to \mathbf{x}^*$ as $t \to +\infty$; and*
- *there exists a point $\mathbf{x}_u$ near $\mathbf{x}^*$ such that $\varphi_t(\mathbf{x}_u) \to \mathbf{x}^*$ as $t \to -\infty$.*

Corollary 3.17. *Under the same assumption as in Theorem 3.16, the equilibrium point $\mathbf{x}^*$ is unstable.*

Exercise 3.11. Prove Corollary 3.17 using the Stable Curve Theorem.

Proof of Theorem 3.16. We first outline the proof, and then give the delicate detail. Similar to the proof of Theorem 3.12, one can transform the system, via a translation and a linear map, the system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ becomes[2].

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} -\mu & 0 \\ 0 & \lambda \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + o\left(\sqrt{x^2 + y^2}\right),$$

or equivalently,

$$\begin{aligned} x' &= -\mu x + h_1(x, y) \\ y' &= \lambda y + h_2(x, y). \end{aligned}$$

where $h_1,\ h_2 \in o\left(\sqrt{x^2 + y^2}\right)$ as $(x, y) \to (0, 0)$. Therefore, we can assume without loss of generality that $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ is in this form.

It can be verified that this system is equivalent to the integral system:

$$\begin{aligned} x(t) &= e^{-\mu t}\left(x(0) + \int_0^t e^{\mu s} h_1(x(s), y(s))\,ds\right) \\ y(t) &= e^{\lambda t}\left(y(0) + \int_0^t e^{-\lambda s} h_2(x(s), y(s))\,ds\right). \end{aligned}$$

In order to find a suitable initial condition $(x(0), y(0))$ to give a stable solution such that $x(t) \to 0$ and $y(t) \to 0$ as $t \to \infty$, we need to pick $y(0)$ very judiciously. It is because $y(t) = e^{\lambda t}\left(y(0) + \int_0^t e^{-\lambda s} h_2(x(s), y(s))ds\right)$ and $e^{\lambda t} \to \infty$ as $t \to \infty$. In hopes of having $\lim_{t\to\infty} y(t) = 0$, we require:

$$\lim_{t\to\infty}\left(y(0) + \int_0^t e^{-\lambda s} h_2(x(s), y(s))\,ds\right) = 0$$

to compensate the growth of the $e^{\lambda t}$ term. In other words, we require:

$$y(0) := -\int_0^\infty e^{-\lambda s} h_2(x(s), y(s))\,ds.$$

[2] Here we denote (x, y) as the components of $\mathbf{x}$ instead of (x_1, x_2) since we will deal with sequences $x_1(t),\ x_2(t),\ x_3(t),\ \cdots$ in the proof and we will use subscripts for the sequence index.

With this choice of $y(0)$, the integral system can be rewritten as

$$x(t) = e^{-\mu t}\left(x(0) + \int_0^t e^{\mu s} h_1(x(s), y(s))\,ds\right)$$
$$y(t) = -e^{\lambda t}\int_t^\infty e^{-\lambda s} h_2(x(s), y(s))\,ds.$$

To prove the stable part of the theorem, we will show the above system has a solution provided that the initial-value $x(0)$ is sufficiently small, and that this solution $(x(t), y(t)) \to (0,0)$ as $t \to \infty$.

Similar to the existence of ODEs, we define the following coupled iteration sequence:

$$x_0(t) \equiv 0$$
$$y_0(t) \equiv 0$$
$$x_n(t) = e^{-\mu t}\left(x(0) + \int_0^t e^{\mu s} h_1(x_{n-1}(s), y_{n-1}(s))\,ds\right)$$
$$y_n(t) = -e^{\lambda t}\int_t^\infty e^{-\lambda s} h_2(x_{n-1}(s), y_{n-1}(s))\,ds.$$

The goal is now to prove both $x_n(t)$ and $y_n(t)$ converges uniformly on $t \in [0,\infty)$ as $n \to \infty$, then the limit functions $x_\infty(t)$ and $y_\infty(t)$ will solve the integral system, and therefore the equivalent differential system has a solution $(x_\infty(t), y_\infty(t))$. An additional argument will show $x_\infty(t) \to 0$ and $y_\infty(t) \to 0$ as $t \to \infty$, and therefore the solution $(x_\infty(t), y_\infty(t))$ traces out a stable curve near $(0,0)$.

The analysis part for showing $x_n(t)$ and $y_n(t)$ converges uniformly on $t \in [0,\infty)$ goes as follows: we will find $\varepsilon > 0$ small enough such that whenever the initial-value $|x(0)| < \varepsilon/4$, we have both

$$|x_k(t) - x_{k-1}(t)| \le \frac{1}{2^{k-1}} e^{-\mu t/2}|x(0)|$$
$$|y_k(t) - y_{k-1}(t)| \le \frac{1}{2^{k-1}} e^{-\mu t/2}|x(0)|$$

for any integer $k > 0$. Then it will follow from Weierstrass' M-test that both $\sum_{k=1}^\infty (x_k(t) - x_{k-1}(t))$ and $\sum_{k=1}^\infty (y_k(t) - y_{k-1}(t))$ converges uniformly on $t \in [0,\infty)$. By the telescoping method (similar to the existence theorem of ODEs), it implies $x_n(t)$ and $y_n(t)$ converge uniformly on $t \in [0,\infty)$ as $n \to \infty$. The proof consists of four steps:

Step 1: Restrict the domain to a small ball $\sqrt{x^2+y^2} < \varepsilon$ for some $\varepsilon > 0$, such that the Lipschitz constants of $h_1(x,y)$ and $h_2(x,y)$ are not so large:

First we claim that

$$\frac{\partial h_1}{\partial x}, \quad \frac{\partial h_1}{\partial y}, \quad \frac{\partial h_2}{\partial x}, \quad \frac{\partial h_2}{\partial y} \to 0$$

as $(x,y) \to (0,0)$. We give the proof for $\frac{\partial h_1}{\partial x}$ only since the other three can be proved in exactly the same way. As the vector field is C^1, the component $f_1(x,y) := -\mu x + h_1(x,y)$ has continuous first partial derivatives. Therefore, we have.

$$\frac{\partial h_1}{\partial x}(x,y) \to \frac{\partial h_1}{\partial x}(0,0)$$

as $(x,y) \to (0,0)$. It suffices to show $\frac{\partial h_1}{\partial x}(0,0) = 0$, which is equivalent to $\frac{\partial f_1}{\partial x}(0,0) = -\mu$. To verify this:

$$\begin{aligned} \frac{\partial f_1}{\partial x}(0,0) &= \lim_{s\to 0} \frac{f_1(s,0) - f_1(0,0)}{s} \\ &= \lim_{s\to 0} \frac{-\mu s + h_1(s,0)}{s} \\ &= -\mu + \lim_{s\to 0} \frac{h_1(s,0)}{s} \\ &= -\mu. \end{aligned}$$

The last step uses the fact that $h_1(x,y) = o\left(\sqrt{x^2+y^2}\right)$ and so $h_1(s,0) = o(|s|)$.

Therefore, we have $\frac{\partial h_1}{\partial x} \to 0$ as $(x,y) \to (0,0)$, and similarly, $\frac{\partial h_1}{\partial y}$, $\frac{\partial h_2}{\partial x}$, $\frac{\partial h_2}{\partial y} \to 0$ as $(x,y) \to (0,0)$.

As a result, there exists $\varepsilon > 0$ such that whenever $(x,y) \in B_\varepsilon(\mathbf{0})$, we have[3]

$$\left|\frac{\partial h_1}{\partial x}(x,y)\right|, \left|\frac{\partial h_1}{\partial y}(x,y)\right|, \left|\frac{\partial h_2}{\partial x}(x,y)\right|, \left|\frac{\partial h_2}{\partial y}(x,y)\right| < \frac{\mu}{4\sqrt{2}}.$$

By the mean value theorem, for any (x,y) and $(\overline{x},\overline{y})$ in the ball $B_\varepsilon(\mathbf{0})$, we have (for $i = 1,2$):

$$|h_i(x,y) - h_i(\overline{x},\overline{y})| \le \frac{\mu}{4\sqrt{2}}\sqrt{(x-\overline{x})^2 + (y-\overline{y})^2}.$$

Step 2: Next we prove the "core" part by induction. Pick $|x(0)| < \varepsilon/4$, we claim:

$$\begin{aligned} |x_k(t) - x_{k-1}(t)| &\le \frac{1}{2^{k-1}} e^{-\mu t/2}|x(0)| \\ |y_k(t) - y_{k-1}(t)| &\le \frac{1}{2^{k-1}} e^{-\mu t/2}|x(0)| \end{aligned}$$

for any $k > 0$.

For $k = 1$, note that $x_0(t) = 0$ and $y_0(t) = 0$, we have:

$$\begin{aligned} |x_1(t) - x_0(t)| &= \left| e^{-\mu t} x(0) + e^{-\mu t} \int_0^t e^{\mu s} h_1(x_0(s), y_0(s)) | ds \right| \\ &\le e^{-\mu t}|x(0)| + e^{-\mu t} \int_0^t e^{\mu s} |h_1(0,0)| ds \\ &= e^{-\mu t}|x(0)| + 0 \\ &\le e^{-\mu t/2}|x(0)| = \frac{1}{2^{1-1}} e^{-\mu t/2}|x(0)| \\ |y_1(t) - y_0(t)| &= \left| e^{\lambda t} \int_t^\infty e^{-\lambda t} h_2(x_0(s), y_0(s)) \right| \\ &= e^{\lambda t} \left| \int_t^\infty e^{-\lambda t} h_2(0,0) ds \right| = 0 \end{aligned}$$

Therefore, the claim is true when $k = 1$.

[3] It may not be clear at this stage why we want the Lipschitz constants to be less than $\frac{\mu}{4\sqrt{2}}$, but you will see why later. As in many other analysis proofs, things have to be understood backward.

Now assume the claim is true when $k = 1, 2, \cdots, j$. One consequence is that

$$\begin{aligned}
|x_j(t)| &= |x_j(t) - x_0(t)| \\
&= \left|\sum_{k=1}^{j}(x_k(t) - x_{k-1}(t))\right| && \text{(telescoping)} \\
&\le \sum_{k=1}^{j} |x_k(t) - x_{k-1}(t)| && \text{(triangle inequality)} \\
&\le \sum_{k=1}^{j} \frac{1}{2^{k-1}} e^{-\mu t/2}|x(0)| && \text{(induction assumption)} \\
&\le e^{-\mu t/2}|x(0)| \sum_{k=1}^{\infty} \frac{1}{2^{k-1}} \\
&\le 2e^{-\mu t/2}|x(0)| && \text{(sum of the geometric series)} \\
&\le 2|x(0)| < \varepsilon/2.
\end{aligned}$$

Similarly, we have $|y_j(t)| < \varepsilon/2$. Therefore, the point $(x_j(t), y_j(t))$ lies in the ball $B_\varepsilon(\mathbf{0})$ so that we can apply the Lipschitz continuity of h_1 and h_2 later in the proof.

Now, consider the case $k = j + 1$:

$$\begin{aligned}
|x_{j+1}(t) - x_j(t)| &= \left|e^{-\mu t} \int_0^t e^{\mu s}(h_1(x_j(s), y_j(s)) - h_1(x_{j-1}(s), y_{j-1}(s)))ds\right| \\
&\le e^{-\mu t} \int_0^t e^{\mu s}|h_1(x_j(s), y_j(s)) - h_1(x_{j-1}(s), y_{j-1}(s))|ds \\
&\le e^{-\mu t} \int_0^t e^{\mu s} \frac{\mu}{4\sqrt{2}} \sqrt{|x_j(s) - x_{j-1}(s)|^2 + |y_j(s) - y_{j-1}(s)|^2} ds,
\end{aligned}$$

where the last step follows from the Lipschitz continuity of h_1 and that $(x_j(s), y_j(s))$ and $(x_{j-1}(s), y_{j-1}(s))$ are in the ball $B_\varepsilon(\mathbf{0})$.

By the induction assumption, $|x_j(s) - x_{j-1}(s)| < \frac{1}{2^{j-1}} e^{-\mu s/2}|x(0)|$ and $|y_j(s) - y_{j-1}(s)| < \frac{1}{2^{j-1}} e^{-\mu s/2}|x(0)|$, and so:

$$\begin{aligned}
|x_{j+1}(t) - x_j(t)| &\le e^{-\mu t} \int_0^t \frac{\mu e^{\mu s}}{4\sqrt{2}} \cdot \frac{\sqrt{2}}{2^{j-1}} e^{-\mu s/2}|x(0)|ds \\
&\le e^{-\mu t} \cdot \frac{\mu|x(0)|}{4 \cdot 2^{j-1}} \int_0^t e^{\mu s/2} ds \\
&= e^{-\mu t} \cdot \frac{\mu|x(0)|}{4 \cdot 2^{j-1}} \cdot \frac{2}{\mu}(e^{\mu t/2} - 1) \\
&< e^{-\mu t} \cdot \frac{\mu|x(0)|}{4 \cdot 2^{j-1}} \cdot \frac{2}{\mu} e^{\mu t/2} = \frac{1}{2^j} e^{-\mu t/2}|x(0)|
\end{aligned}$$

Similarly,

$$\begin{aligned}
|y_{j+1}(t) - y_j(t)| &= \left| e^{\lambda t} \int_t^\infty e^{-\lambda s}(h_2(x_j(s), y_j(s)) - h_2(x_{j-1}(s), y_{j-1}(s)))ds \right| \\
&\le e^{\lambda t} \int_t^\infty e^{-\lambda s} |h_2(x_j(s), y_j(s)) - h_2(x_{j-1}(s), y_{j-1}(s))| ds \\
&\le e^{\lambda t} \int_t^\infty e^{-\lambda s} \cdot \frac{\mu}{4\sqrt{2}} \sqrt{|x_j(s) - x_{j-1}(s)|^2 + |y_j(s) - y_{j-1}(s)|^2} ds \\
&\le e^{\lambda t} \int_t^\infty e^{-\lambda s} \cdot \frac{\mu}{4\sqrt{2}} \cdot \frac{\sqrt{2}}{2^{j-1}} e^{-\mu s/2} |x(0)| ds \\
&= e^{\lambda t} \cdot \frac{\mu |x(0)|}{4 \cdot 2^{j-1}} \int_t^\infty e^{-(\lambda + \frac{\mu}{2})s} ds \\
&= e^{\lambda t} \cdot \frac{\mu |x(0)|}{4 \cdot 2^{j-1}} \frac{1}{\lambda + \frac{\mu}{2}} (e^{-(\lambda + \frac{\mu}{2})t} - 1) \\
&< e^{\lambda t} \cdot \frac{\mu |x(0)|}{4 \cdot 2^{j-1}} \frac{1}{\lambda + \frac{\mu}{2}} e^{-(\lambda + \frac{\mu}{2})t} = \frac{\mu}{\lambda + \frac{\mu}{2}} \cdot \frac{1}{4 \cdot 2^{j-1}} \cdot |x(0)| e^{-\mu t/2} \\
&\le \frac{\mu}{\mu/2} \cdot \frac{1}{4 \cdot 2^{j-1}} \cdot |x(0)| e^{-\mu t/2} = \frac{1}{2^j} e^{-\mu t/2} |x(0)|.
\end{aligned}$$

Hence the claim is true for $k = j+1$. By induction, the claim holds for any integer $k > 0$.

Step 3: Show uniform convergence and complete the proof (of the stable part).

From Step 2 we have

$$|x_k(t) - x_{k-1}(t)| \le \frac{e^{-\mu t/2}}{2^{k-1}} |x(0)| \le \frac{|x(0)|}{2^{k-1}}$$

for any $k > 0$ and $t \in [0, \infty)$. The geometric series test implies $\sum_{k=1}^\infty \frac{|x(0)|}{2^{k-1}}$ converges. Therefore, the Weierstrass's M-test shows the series $\sum_{k=1}^\infty (x_k(t) - x_{k-1}(t))$ converges uniformly on $t \in [0, \infty)$. Since

$$x_n(t) = \sum_{k=1}^n (x_k(t) - x_{k-1}(t)) \quad \text{(recall that } x_0(t) = 0\text{)},$$

the sequence $x_n(t)$ converges uniformly on $t \in [0, \infty)$ to some function $x_\infty(t)$ as $n \to \infty$. Exactly the same argument proves $y_n(t) \to y_\infty(t)$ uniformly on $t \in [0, \infty)$ as $n \to \infty$.

Given that we have uniform convergence, one can switch the integral sign and the limit. Let $n \to \infty$ on both sides of:

$$\begin{aligned}
x_n(t) &= e^{-\mu t} \left(x(0) + \int_0^t e^{\mu s} h_1(x_{n-1}(s), y_{n-1}(s)) ds \right) \\
y_n(t) &= -e^{\lambda t} \int_t^\infty e^{-\lambda s} h_2(x_{n-1}(s), y_{n-1}(s)) ds,
\end{aligned}$$

one can get

$$\begin{aligned}
x_\infty(t) &= e^{-\mu t} \left(x(0) + \int_0^t e^{\mu s} h_1(x_\infty(s), y_\infty(s)) ds \right) \\
y_\infty(t) &= -e^{\lambda t} \int_t^\infty e^{-\lambda s} h_2(x_\infty(s), y_\infty(s)) ds.
\end{aligned}$$

Therefore $(x_\infty(t), y_\infty(t))$ solves the integral system, and hence the differential system.

The last thing to check is that this particular solution is indeed a stable solution. In fact, it follows from the inequalities:

$$\begin{aligned}|x_k(t) - x_{k-1}(t)| &\le \frac{1}{2^{k-1}} e^{-\mu t/2} |x(0)| \\ |y_k(t) - y_{k-1}(t)| &\le \frac{1}{2^{k-1}} e^{-\mu t/2} |x(0)|.\end{aligned}$$

Again use the fact that

$$x_n(t) = \sum_{k=1}^{n} (x_k(t) - x_{k-1}(t)),$$

we get

$$\begin{aligned}|x_n(t)| &\le \sum_{k=1}^{n} |x_k(t) - x_{k-1}(t)| \\ &\le \sum_{k=1}^{\infty} |x_k(t) - x_{k-1}(t)| \\ &\le \sum_{k=1}^{\infty} \frac{1}{2^{k-1}} e^{-\mu t/2} |x(0)| = 2e^{-\mu t/2} |x(0)|.\end{aligned}$$

Let $n \to \infty$, we have $|x_\infty(t)| \le 2e^{-\mu t/2}|x(0)|$ for any $t \in [0, \infty)$. Clearly it implies $x_\infty(t) \to 0$ as $t \to \infty$. Similarly, one can show $|y_n(t)| \le 2e^{-\mu t/2}|x(0)|$ and the same result holds for $y_\infty(t)$.

Now that $(x_\infty(t), y_\infty(t)) \to (0, 0)$ as $t \to \infty$, so they form a stable solution. Take $\mathbf{x}_s = (x_\infty(0), y_\infty(0))$, then $\varphi_t(\mathbf{x}_s) = (x_\infty(t), y_\infty(t)) \to (0, 0)$ as $t \to \infty$.

Step 4: Finally, mimic the above with $-t$ in place of t to prove the existence of an unstable curve.

□

3.4.0.1. Summary. To summarize, given a nonlinear system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ with an equilibrium point $\mathbf{x}^*$, if all eigenvalues of the Jacobian matrix $D\mathbf{F}_{\mathbf{x}^*}$ have non-zero real parts, then the point $\mathbf{x}^*$ is said to be hyperbolic, and the stability of the point $\mathbf{x}^*$ is the same as the stability of **0** of the linearized system $\mathbb{X}' = (D\mathbf{F}_{\mathbf{x}^*})\mathbb{X}$. However, this is not necessarily true if one of the eigenvalues of $D\mathbf{F}_{\mathbf{x}^*}$ has zero real parts, i.e. non-hyperbolic.

3.5. Lyapunov Functions

When an equilibrium point is not hyperbolic, the linearization method fails to conclude whether the point is stable or not. In this section, we introduce the Lyapunov's Second Method which can be used to determine the stability of a non-hyperbolic equilibrium point.

Definition 3.18 (Lyapunov Functions)**.** Let $\mathbf{F} : \mathbb{R}^d \to \mathbb{R}^d$ be a C^1 vector field, and the system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ has an equilibrium point $\mathbf{x}^*$. Let $L : U \to \mathbb{R}$ be a C^1 function defined on an open set U containing $\mathbf{x}^*$. Suppose all of the following holds:

(1) $L(\mathbf{x}^*) = 0$; and
(2) $L(\mathbf{x}) > 0$ for any $\mathbf{x} \in U \backslash \{\mathbf{x}^*\}$; and
(3) For any solution curve $\mathbf{x}(t)$ in U, we have

$$\frac{d}{dt} L(\mathbf{x}(t)) \leq 0 \quad \text{for any } t \geq 0 \text{ such that } \mathbf{x}(t) \neq \mathbf{x}^*,$$

then L is called a *Lyapunov function* for $\mathbf{x}^*$.

If the function L further satisfies:

3'. For any solution curve $\mathbf{x}(t)$ in U, we have

$$\frac{d}{dt} L(\mathbf{x}(t)) < 0 \quad \text{for any } t \geq 0 \text{ such that } \mathbf{x}(t) \neq \mathbf{x}^*,$$

then L is called a *strict Lyapunov function* for $\mathbf{x}^*$. □

The following theorem showcases the significance of Lyapunov functions:

Theorem 3.19 (Lyapunov's Second Method)**.** *Let* $\mathbf{F} : \mathbb{R}^d \to \mathbb{R}^d$ *be a* C^1 *vector field, and the system* $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ *has an equilibrium point* $\mathbf{x}^*$*. Then,*

- *If there exists a Lyapunov function* L *for* $\mathbf{x}^*$*, then* $\mathbf{x}^*$ *is stable.*
- *If there exists a strict Lyapunov function* L *for* $\mathbf{x}^*$*, then* $\mathbf{x}^*$ *is asymptotically stable.*

Before we learn the proof of Theorem 3.19, let's first look at a couple of examples of Lyapunov Functions and how Theorem 3.19 can be used to show stability of an equilibrium point.

Example 3.4. Consider the ODE system:

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

Clearly, $(0,0)$ is an equilibrium point. The linearization at $\mathbf{0}$ is represented by the matrix

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

which has eigenvalues 0 and -2. Therefore, $\mathbf{0}$ is **not** hyperbolic and so the linearization method discussed in the previous section fails to conclude the stability of $\mathbf{0}$.

Let $L(x, y) = x^2 + y^2$. We are about to show that it is a strict Lyapunov function for $\mathbf{0}$ and so Theorem 3.19 concludes $\mathbf{0}$ is asymptotically stable:

Clearly conditions (1) and (2) hold since $L(0,0) = 0$ and $L(x,y) > 0$ when $(x,y) \neq (0,0)$. To check condition (3'), we consider:

$$\begin{aligned}
\frac{d}{dt}L(x,\,y) &= \frac{d}{dt}(x^2+y^2) \\
&= 2x \cdot x' + 2y \cdot y' \\
&= 2x \cdot (-x+y+xy) + 2y \cdot (x-y-x^2-y^3) \\
&= -2x^2 + 4xy - 2y^2 - 2y^4 \\
&= -2(x-y)^2 - 2y^4 && \text{(factorization)} \\
&< 0 && \text{(when } (x,y) \neq (0,0)\text{)}
\end{aligned}$$

Therefore, L is a strictly Lyapunov function and $\mathbf{0}$ is asymptotically stable. □

Example 3.5. Consider the system:

$$\begin{aligned}
x' &= y \\
y' &= -kx - y^3(1+x^2)
\end{aligned}$$

where $k > 0$ is a constant. We are going to show that

$$L(x,y) := kx^2 + y^2$$

is a Lyapunov function for the equilibrium point $\mathbf{0}$.

Clearly, conditions (1) and (2) are satisfied. To show it satisfies condition (3), we consider:

$$\begin{aligned}
\frac{d}{dt}L(x,y) &= \frac{d}{dt}(kx^2+y^2) \\
&= 2kx \cdot x' + 2y \cdot y' \\
&= 2kxy + 2y(-kx - y^3(1+x^2)) \\
&= -y^4(1+x^2) \leq 0.
\end{aligned}$$

Therefore, $L(x,y)$ is a Lyapunov function and so $(0,0)$ is stable.

Note that $L(x,y)$ is not a strict Lyapunov function since $\frac{dL}{dt} = -y^4(1+x^2)$ can be zero even when $(x,y) \neq (0,0)$. In fact, $\frac{dL}{dt} = 0$ whenever $y = 0$ and x is any real number. □

Next we give the proof of the Lyapunov's Second Method. The key idea of the stable part of the method is that the level region of the form $\{\mathbf{x} \in \mathbb{R}^d : L(\mathbf{x}) \leq \alpha\}$ will "trap" the trajectory inside the region, i.e. if the trajectory starts there initially, it will remain so along the flow. Figure 3.9 shows the trajectories of the system in Example 3.4 on the level set diagram of the Lyapunov function $L(x,y) = x^2 + y^2$.

Proof of Theorem 3.19. First suppose there exists a Lyapunov's function $L : U \to \mathbb{R}$ defined on an open set U containing $\mathbf{x}^*$. Given any $\varepsilon > 0$, by the continuity of L and the fact that the boundary of the ball $B_\varepsilon(\mathbf{x}^*)$, denoted by $\partial B_\varepsilon(\mathbf{x}^*)$, is closed and bounded, the extreme-value theorem shows L attains a minimum α on the boundary set $\partial B_\varepsilon(\mathbf{x}^*)$. By Property 2 of Lyapunov's functions, we must have $\alpha > 0$. Then, the region $\mathcal{O} = \{\mathbf{x} \in \mathbb{R}^d : L(\mathbf{x}) < \alpha\}$ is in the interior of the ball $B_\varepsilon(\mathbf{x}^*)$, and by Property 1 of Lyapunov's function and the fact that $\alpha > 0$, we know $\mathbf{x}^* \in \mathcal{O}$. Given any $\mathbf{x}_0 \in \mathcal{O}$, Property 3 of Lyapunov's function asserts that $L(\varphi_t(\mathbf{x}_0))$ must be decreasing as t increases, thus $\varphi_t(\mathbf{x}_0)$ must stay in the region $\mathcal{O}$. This shows $\mathbf{x}^*$ is stable.

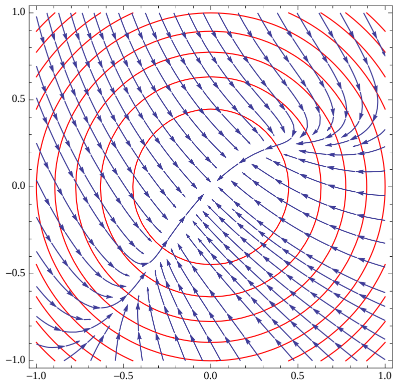


Figure 3.9. The phase portrait of the system: $x' = -x + y + xy, \quad y' = x - y - x^2 - y^3$ on the level sets of $L(x, y) = x^2 + y^2$. The solution curves travel in a direction which decreases the value of L. Therefore, each ball bounded by a circle contour is a trapping region of the trajectories.

Now assume further that $L : U \to \mathbb{R}$ is a strict Lyapunov's function. We need to show $\mathbf{x}^*$ is asymptotically stable, i.e. given any $\mathbf{x}_0$ sufficiently close to $\mathbf{x}^*$, we have $\varphi_t(\mathbf{x}_0) \to \mathbf{x}^*$ as $t \to \infty$.

In order to prove

$$\lim_{t\to\infty} \varphi_t(\mathbf{x}_0) = \mathbf{x}^*,$$

one common approach is to show that if t_n is a sequence of times such that $t_n \to \infty$ as $n \to \infty$ and that $\varphi_{t_n}(\mathbf{x}_0)$ converges to a limit $\mathbf{z}$, then the limit $\mathbf{z}$ must be $\mathbf{x}_0$. To prove this, we take an arbitrary $s > 0$, and we will show $L(\varphi_s(\mathbf{z})) = L(\mathbf{z})$ for any $s > 0$. Then, it will imply

$$\frac{d}{ds} L(\varphi_s(\mathbf{z})) = 0$$

for any $s > 0$. By Property 3' of strict Lyapunov's function, the only chance it can happen is that $\mathbf{z} = \mathbf{x}^*$.

Here we give the proof of $L(\varphi_s(\mathbf{z})) = L(\mathbf{z})$: given any $s > 0$, consider a new sequence of times $\{s + t_n\}_{n=1}^{\infty}$. Pick a subsequence $\{t_{k_n}\}_{n=1}^{\infty}$ of $\{t_n\}_{n=1}^{\infty}$ such that $s + t_n \le t_{k_n}$ for each n, then we have $\varphi_{t_{k_n}}(\mathbf{x}_0) \to \mathbf{z}$ as $n \to \infty$. By the continuity of L and φ_s, we have:

$$\begin{aligned}
L(\mathbf{z}) &= \lim_{n\to\infty} L(\varphi_{t_{k_n}}(\mathbf{x}_0)) && \text{(continuity of } L\text{)}\\
&\le \lim_{n\to\infty} L(\varphi_{s+t_n}(\mathbf{x}_0)) && \text{(since } s + t_n \le t_{k_n}\text{)}\\
&= \lim_{n\to\infty} L(\varphi_s(\varphi_{t_n}(\mathbf{x}_0))) && \text{(since } \varphi_s \circ \varphi_{t_n} = \varphi_{s+t_n}\text{)}\\
&= L(\varphi_s(\mathbf{z})) && \text{(by continuity of } L \text{ and } \varphi_s\text{)}\\
&\le L(\mathbf{z}) && \text{(by Property 3')}
\end{aligned}$$

Overall, we proved $L(\varphi_s(\mathbf{z})) = L(\mathbf{z})$ for any $s > 0$. By the discussion above, we proved $\mathbf{z} = \mathbf{x}^*$. This shows $\mathbf{x}^*$ is asymptotically stable.

□

The key idea for establishing asymptotic stability in the above proof is to show $L(\varphi_s(\mathbf{z})) = L(\mathbf{z})$ for any $s > 0$ where $\mathbf{z}$ is the limit of $\varphi_{t_n}(\mathbf{x}_0)$. Property 3' of strict Lyapunov functions forces $\mathbf{z}$ must be the equilibrium point $\mathbf{x}^*$. In fact, in order to show $\mathbf{z} = \mathbf{x}^*$, L does not have to be a *strict* Lyapunov function, but simply a Lyapunov function, provided that the equilibrium solution $\mathbf{x}(t) \equiv \mathbf{x}^*$ is the *only* solution such that $\frac{d}{dt} L(\mathbf{x}(t)) = 0$.

Theorem 3.20 (LaSalle's Principle)**.** *Let* $\mathbf{F} : \mathbb{R}^d \to \mathbb{R}^d$ *be a* C^1 *vector field, and the system* $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ *has an equilibrium point* $\mathbf{x}^*$*. If there exists a Lyapunov function* L *for* $\mathbf{x}^*$ *such that the only possible solution* $\mathbf{x}(t)$ *for* $\frac{d}{dt}L(\mathbf{x}(t)) = 0$ *for any* $t \in [0, \infty)$ *is the equilibrium solution* $\mathbf{x}^*$*, then* $\mathbf{x}^*$ *is asymptotically stable.*

Proof. The proof is almost identical to the proof of Theorem 3.19. After establishing $L(\varphi_s(\mathbf{z})) = L(\mathbf{z})$ for any $s > 0$ (and hence $\frac{d}{ds}L(\varphi_s(\mathbf{z})) = 0$), the hypothesis of this theorem implies $\varphi_s(\mathbf{z})$ must be the equilibrium solution $\varphi_s(\mathbf{z}) = \mathbf{x}^*$ for any $s > 0$.

□

Example 3.6 (Revisited)**.** Consider the system

$$
\begin{aligned}
x' &= y \\
y' &= -kx - y^3(1 + x^2)
\end{aligned}
$$

where $k > 0$. This system was considered in Example 3.5 and we showed $L = kx^2 + y^2$ is a Lyapunov function for $(0, 0)$:

$$\frac{d}{dt}L(x, y) = -y^4(1 + x^2).$$

L is merely a Lyapunov function and so Theorem 3.19 can only concludes that $(0, 0)$ is stable. However, one can use the LaSalle's Principle to show in fact $(0, 0)$ is asymptotically stable:

Whenever $\frac{d}{dt}L(x(t), y(t)) = 0$ for any $t \geq 0$, we have $-y(t)^4(1 + x(t)^2) = 0$ for any t. Therefore, $y(t) \equiv 0$, and by substituting it to the system, we get:

$$
\begin{aligned}
x' &= 0 \\
0 &= -kx
\end{aligned}
$$

which implies $x(t) \equiv 0$ as well. Therefore, $\mathbf{x}(t) \equiv \mathbf{0}$ is the only possible solution for $L(\mathbf{x}(t)) \equiv$ constant. By LaSalle's Principle, $\mathbf{0}$ is asymptotically stable. □

3.5.0.1. Tips of finding a Lyapunov function. Finding a Lyapunov function may sometimes be a challenging task that may require trial-and-error. The following example demonstrates some tips of finding a suitable Lyapunov function.

Example 3.7. Consider the system:

$$
\begin{aligned}
x' &= -x^3 + y \\
y' &= -2(x^3 + y^3)
\end{aligned}
$$

Clearly $(0, 0)$ is an equilibrium point. We want to determine whether it is stable. We guess a Lyapunov function is of the form $L(x, y) = Ax^m + By^n$. In order for L to satisfy Properties 1 and 2, we need:

$$
\begin{aligned}
&A,\ B > 0, \text{ and} \\
&m,\ n \text{ are even.}
\end{aligned}
$$

We compute:

$$
\begin{aligned}
\frac{dL}{dt} &= mAx^{m-1} \cdot x' + nBy^{n-1} \cdot y' \\
&= mAx^{m-1}(-x^3 + y) - 2nBy^{n-1}(x^3 + y^3) \\
&= -mAx^{m+2} + mAx^{m-1}y - 2nBx^3y^{n-1} - 2nBy^{n+2}
\end{aligned}
$$

As far as A, $B > 0$, m and n are even, the terms $-mAx^{m+2}$ and $-2nBy^{n+2}$ are negative as desired. However, the terms $mAx^{m-1}y$ and $-2nBx^3y^{n-1}$ are neither positive or negative definite, so we want to choose some suitable A, B, m and n so that they cancel out each other. Namely, we require:

$$\begin{aligned} m - 1 &= 3 \\ 1 &= n - 1 \\ mA &= 2nB \end{aligned}$$

Clearly, $m = 4$, $n = 2$, $A = B = 1$ is a solution. With this choice, $L(x, y) = x^4 + y^2$ is a strict Lyapunov function since

$$\frac{dL}{dt} = -4x^6 - 4y^4 < 0 \text{ whenever } (x, y) \neq (0, 0).$$

Therefore, $(0, 0)$ is asymptotically stable. □

Exercise 3.12. For each of the following systems, show that $(0, 0)$ is asymptotically stable. Note that some of them can be handled by linearization.

(1) $x' = -4x - y, \quad y' = -2x - 5y - 2y\sin x$
(2) $x' = x + y + x^2y, \quad y' = -x + y\cos x$
(3) $x' = xy - 2x^2y^3 - x^3, \quad y' = -y - \frac{1}{2}x^2 + x^3y^2$
(4) $x' = -2(x^3 + y^3), \quad y' = x - 2y^3$

Theorem 3.19 asserts that if there exists a strict Lyapunov function for an equilibrium point $\mathbf{x}^*$, then $\mathbf{x}^*$ is asymptotically stable. It is natural to answer whether the converse is true, i.e. if $\mathbf{x}^*$ is asymptotically stable, does it always exist a strict Lyapunov function? The answer is yes, which was proved by Massera in 1950s. The proof is beyond the scope of this lecture note, but it is much easier to prove a *partial* converse by imposing a condition on the flow φ_t. See the following exercise:

Exercise 3.13. Consider a system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ with flow denoted by φ_t. Suppose $\mathbf{0}$ is an asymptotically stable equilibrium point and assume φ_t satisfies:

$$|\varphi_t(\mathbf{x})| \leq f(t)\,|\mathbf{x}| \quad \text{for all } t \geq 0$$

where $f(t) : [0, \infty) \to \mathbb{R}$ is a continuous function satisfying:

$$\int_0^\infty |f(t)|^2 \, dt < \infty.$$

For example, $f(t) = e^{-t}$ satisfies this condition.

Define

$$L(\mathbf{x}) := \int_0^\infty |\varphi_s(\mathbf{x})|^2 \, ds.$$

(1) Check that L is well-defined, i.e. the integral defining L is finite.
(2) Show that for $t > 0$ and $\mathbf{x} \in \mathbb{R}^d$, we have:

$$L(\varphi_t(\mathbf{x})) = \int_t^\infty |\varphi_s(\mathbf{x})|^2 \, ds.$$

(3) Show that

$$\frac{d}{dt} L(\varphi_t(\mathbf{x})) = -\,|\varphi_t(\mathbf{x})|^2$$

and that L is a strict Lyapunov function for $\mathbf{0}$.

3.6. Gradient and Hamiltonian Systems

Using Lyapunov's Second Method (Theorem 3.19) to determine the stability of an equilibrium point is effective in a sense that one does not need to know the solution to the ODE system. However, it may sometimes be challenging to come up with a suitable Lyapunov function, and sometimes it is a matter of experience and trial-and-error. However, there are two types of ODE systems, *gradient* and *Hamiltonian* systems, which always come along with possible *candidates* of Lyapunov functions.

3.6.1. Gradient Systems.

Let's begin with gradient systems:

Definition 3.21 (Gradient System). A *gradient* system is an ODE system of the form:

$$\mathbf{x}' = -\nabla f(\mathbf{x})$$

where $f : \mathbb{R}^d \to \mathbb{R}$ is a C^2-function. The function f is often called a *potential function* of the system. □

Remark 3.22. Recall that ∇f is defined as the following vector field:

$$\nabla f(\mathbf{x}) = \left(\frac{\partial f}{\partial x_1}(\mathbf{x}), \ldots, \frac{\partial f}{\partial x_d}(\mathbf{x})\right).$$

□

Remark 3.23. A system of the form $\mathbf{x}' = \nabla f(\mathbf{x})$ is also a gradient system since one rewrite it as $\mathbf{x}' = -\nabla(-f(\mathbf{x}))$. However, it is a convention to take the form $\mathbf{x}' = -\nabla f(\mathbf{x})$ since then the potential function f, as we will see, will be decreasing along trajectories. □

The equilibrium points of a gradient system $\mathbf{x}' = -\nabla f(\mathbf{x})$ are the critical points of the function f. The significance of gradient systems in terms of stability is that the potential function is a possible *candidate* of a Lyapunov function, in a sense that Property 3 of Lyapunov functions always hold, although Properties 1 and 2 may not.

Lemma 3.24. *The potential function $f(\mathbf{x})$ of the gradient system $\mathbf{x}' = -\nabla f(\mathbf{x})$ satisfies:*

$$\frac{d}{dt} f(\mathbf{x}(t)) = -\left|\nabla f(\mathbf{x}(t))\right|^2 \leq 0$$

for any solution curve $\mathbf{x}(t)$. Therefore, f must satisfy Property 3 of Lyapunov functions.

Proof. As always, we denote the components of $\mathbf{x}$ by $(x_1, \ldots, x_d)$, then f is a function of $x_1, \ldots, x_d$ and each of x_i's is a function of t when traveling along any trajectory. By chain rule, we have:

$$\begin{aligned}
\frac{d}{dt} f(\mathbf{x}(t)) &= \sum_{i=1}^{d} \frac{\partial f}{\partial x_i} \cdot \frac{dx_i}{dt} \\
&= \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_d}\right) \cdot (x_1', \ldots, x_d') \\
&= \nabla f(\mathbf{x}) \cdot \mathbf{x}' \\
&= \nabla f(\mathbf{x}) \cdot (-\nabla f(\mathbf{x})) \\
&= -\left|\nabla f(\mathbf{x})\right|^2.
\end{aligned}$$

□

The following result tells us when the potential function f is truly a Lyapunov function:

Theorem 3.25. *Consider the gradient system $\mathbf{x}' = -\nabla f(\mathbf{x})$ with an equilibrium point $\mathbf{x}^*$, i.e. a critical point of the function f. Suppose $\mathbf{x}^*$ is an isolated local minimum point of f, then $L(\mathbf{x}) := f(\mathbf{x}) - f(\mathbf{x}^*)$ is a strict Lyapunov function for $\mathbf{x}^*$ and hence $\mathbf{x}^*$ is asymptotically stable.*

Proof. Since $\mathbf{x}^*$ is an isolated local minimum point of f, there exists a small ball $B_\varepsilon(\mathbf{x}^*)$ such that:

$$f(\mathbf{x}) > f(\mathbf{x}^*) \quad \text{for any } \mathbf{x} \in B_\varepsilon(\mathbf{x}^*) \text{ and } \mathbf{x} \neq \mathbf{x}^*.$$

Then, $L(\mathbf{x}^*) = 0$ and $L(\mathbf{x}) = f(\mathbf{x}) - f(\mathbf{x}^*) > 0$ for any $\mathbf{x} \in B_\varepsilon(\mathbf{x}^*)$ and $\mathbf{x} \neq \mathbf{x}^*$. Furthermore, by Lemma 3.24, we have:

$$\frac{d}{dt}L(\mathbf{x}(t)) = \frac{d}{dt}f(\mathbf{x}(t)) = -\left|\nabla f(\mathbf{x}(t))\right|^2.$$

Since $\mathbf{x}^*$ is the only critical point of f on the ball $B_\varepsilon(\mathbf{x}^*)$, $\nabla f(\mathbf{x}) \neq \mathbf{0}$ unless $\mathbf{x} = \mathbf{x}^*$. It shows

$$\frac{d}{dt}L(\mathbf{x}(t)) < 0 \text{ for any } \mathbf{x} \in B_\varepsilon(\mathbf{x}^*) \text{ and } \mathbf{x} \neq \mathbf{x}^*.$$

Therefore, $L : B_\varepsilon(\mathbf{x}^*) \to \mathbb{R}$ is a strict Lyapunov function for $\mathbf{x}^*$.

□

Example 3.8. Consider the system

$$\begin{aligned} x' &= -2x(x-1)(2x-1) \\ y' &= -2y \end{aligned}$$

It is a gradient system as one can verify that the potential function can be taken to be $f(x,y) = x^2(x-1)^2 + y^2$. There are three equilibrium points, namely $(0,0)$, $(\frac{1}{2},0)$ and $(1,0)$.

Both $(0,0)$ and $(1,0)$ are isolated local minimum points of f, since $f(0,0) = 0$ and $f(x,y) > 0$ for any $(x,y) \in B_{1/2}(0,0)$, and likewise $f(1,0) = 0$ and $f(x,y) > 0$ for any $(x,y) \in B_{1/2}(1,0)$. Therefore, $f(x,y)$ is a strict Lyapunov function for both $(0,0)$ and $(1,0)$, and hence $(0,0)$ and $(1,0)$ are both asymptotically stable.

However, $(\frac{1}{2},0)$ is not a local minimum point of f as $f(x,0) = x^2(x-1)^2$ has a maximum point at $x = \frac{1}{2}$ and $f(\frac{1}{2},y) = \frac{1}{16} + y^2$ has a minimum point at $y = 0$. Therefore, $f(x,y) - f(\frac{1}{2},0)$ is not a Lyapunov function for $(\frac{1}{2},0)$. In fact, by the linearization at $(\frac{1}{2},0)$ is given by the matrix:

$$\begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix}$$

whose eigenvalues have mixed signs. By Theorem 3.16 (Stable Curve Theorem), the phase portrait near $(\frac{1}{2},0)$ resembles a saddle and hence is unstable. □

3.6.1.1. Criterion for gradient systems. Consider the following system:

$$\begin{aligned} x' &= -y + 10x^5 \\ y' &= x + 3y^6 \end{aligned}$$

One can show it is *not* a gradient system by attempting to solve the equations:

$$-y + 10x^5 = -\frac{\partial f}{\partial x}$$
$$x + 3y^6 = -\frac{\partial f}{\partial y}$$

The first equation implies $f(x, y) = -xy + \frac{5}{3}x^6 + g(y)$ where $g(y)$ is an arbitrary function of y. Differentiating f with respect to y, we get:

$$\frac{\partial f}{\partial y} = -x + g'(y).$$

However, comparing with the second equation above, we get:

$$x + 3y^6 = -x + g'(y) \;\Rightarrow\; 2x = g'(y) - 3y^6.$$

The LHS is a function of x while the RHS is a function of y. It is impossible. Therefore, there is no f to make the ODE system a gradient system.

In some systems such as:

$$x' = x^2 e^{x^2+y^2}$$
$$y' = y^2 e^{x^2+y^2},$$

it may not be feasible to solve $x^2e^{x^2+y^2} = -\frac{\partial f}{\partial x}$ and $y^2e^{x^2+y^2} = -\frac{\partial f}{\partial y}$ for f since it involves integrating $x^2e^{x^2}$ by x. Prior experience of single variable calculus should have told you that it is not possible. Fortunately, one can still show whether it is a gradient system or not using the following elegant test:

Proposition 3.26. *Given a C^1 vector field $\mathbf{F}$ defined everywhere on $\mathbb{R}^d$, there exists a potential function $f : \mathbb{R}^d \to \mathbb{R}$ such that $\mathbf{F} = -\nabla f$ if and only if the Jacobian matrix $D\mathbf{F}_\mathbf{x}$ is symmetric for any $\mathbf{x} \in \mathbb{R}^d$.*

Proof. ($\Rightarrow$) Suppose $\mathbf{F} = -\nabla f$, then $F_i = -\frac{\partial f}{\partial x_i}$ for each $i = 1, \ldots, d$. The (i, j)-th component of the Jacobian matrix $D\mathbf{F}$ is:

$$\frac{\partial F_i}{\partial x_j} = -\frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right) = -\frac{\partial^2 f}{\partial x_j \partial x_i}.$$

From multivariable calculus, we know the second derivatives commute, i.e. $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$, so $\frac{\partial F_j}{\partial x_i} = \frac{\partial F_i}{\partial x_j}$ and the (i, j)-th and (j, i)-th components of $D\mathbf{F}$ are equal. Therefore, $D\mathbf{F}$ is symmetric.

($\Leftarrow$) Given $D\mathbf{F}_\mathbf{x}$ is symmetric for any $\mathbf{x} \in \mathbb{R}^d$, we need to find the potential function f such that $\mathbf{F} = -\nabla f$. We define f as the following line integral:

$$f(\mathbf{x}) := -\int_L \mathbf{F} \cdot d\mathbf{r}$$

where L is the straight line segment from $\mathbf{0}$ to $\mathbf{x}$, i.e. L is parametrized by $\mathbf{r}(t) = t\mathbf{x}$, $0 \le t \le 1$. We will show that f indeed satisfies $\mathbf{F} = -\nabla f$.

Precisely, f is given by:

$$f(x_1, \ldots, x_d) = -\sum_{i=1}^{d} \int_0^1 x_i F_i(tx_1, \ldots, tx_d)\,dt.$$

We need to show $F_j = -\frac{\partial f}{\partial x_j}$ for any $j = 1, \ldots, j$, so we consider:

$$\begin{aligned}
-\frac{\partial f}{\partial x_j} &= \frac{\partial}{\partial x_j} \sum_{i=1}^{d} \int_0^1 x_i F_i(tx_1, \ldots, tx_d)dt \\
&= \sum_{i=1}^{d} \int_0^1 \frac{\partial}{\partial x_j} \left(x_i F_i(tx_1, \ldots, tx_d) \right) dt \\
&= \sum_{i=1}^{d} \int_0^1 \delta_{ij} F_i(tx_1, \ldots, tx_d)dt \\
&\quad + \sum_{i=1}^{d} \int_0^1 x_i \frac{\partial}{\partial x_j} F_i(tx_1, \ldots, tx_d)dt && \text{(product rule)} \\
&= \int_0^1 F_j(tx_1, \ldots, tx_d)dt + \sum_{i=1}^{d} \int_0^1 x_i t (\partial_j F_i)(tx_1, \ldots, tx_d)dt && \text{(chain rule)}
\end{aligned}$$

For simplicity we denote $\frac{\partial F_i}{\partial x_j}$ as $\partial_j F_i$. Since $D\mathbf{F}$ is symmetric, its (i, j)-th and (j, i)-th components are equal, i.e. $\partial_j F_i = \partial_i F_j$. Therefore, the latter term of the above computation becomes:

$$\begin{aligned}
&\sum_{i=1}^{d} \int_0^1 x_i t (\partial_j F_i)(tx_1, \ldots, tx_d)dt \\
&= \sum_{i=1}^{d} \int_0^1 x_i t (\partial_i F_j)(tx_1, \ldots, tx_d)dt \\
&= \int_0^1 t \frac{\partial}{\partial t} F_j(tx_1, \ldots, tx_d)dt && \text{(chain rule, reversed)} \\
&= \left[tF_j(tx_1, \ldots, tx_d) \right]_{t=0}^{t=1} - \int_0^1 F_j(tx_1, \ldots, tx_d)dt && \text{(integration by parts)} \\
&= F_j(x_1, \ldots, x_d) - \int_0^1 F_j(tx_1, \ldots, tx_d)dt.
\end{aligned}$$

Therefore, we have shown $-\frac{\partial f}{\partial x_j}(x_1, \ldots, x_d) = F_j(x_1, \ldots, x_d)$ for any $(x_1, \ldots, x_d) \in \mathbb{R}^d$. In other words, $\mathbf{F}(\mathbf{x}) = -\nabla f(\mathbf{x})$ for any $\mathbf{x} \in \mathbb{R}^d$.

□

Let's try to apply Proposition 3.26 to the vector field:

$$\mathbf{F}(x, y) = \begin{bmatrix} x^2 e^{x^2+y^2} \\ y^2 e^{x^2+y^2} \end{bmatrix}.$$

By direct computation, the Jacobian matrix of $\mathbf{F}$ is:

$$D\mathbf{F}(x, y) = \begin{bmatrix} * & 2x^2 y e^{x^2+y^2} \\ 2xy^2 e^{x^2+y^2} & * \end{bmatrix}$$

where we omit the diagonal entries marked with $*$'s – since they do not affect whether $D\mathbf{F}$ is symmetric or not.

Clearly, $D\mathbf{F}$ is not a symmetric matrix, so Proposition 3.26 asserts that $\mathbf{F}$ cannot be written as $-\nabla f$ for some potential function f. In other words, the system

$$\begin{aligned} x' &= x^2 e^{x^2+y^2} \\ y' &= y^2 e^{x^2+y^2} \end{aligned}$$

is not a gradient system. Therefore, one should not waste time to find the potential function f by integration, since it is not possible!

Exercise 3.14. Determine whether each of the following systems is a gradient system. If so, find a potential function f and determine whether it is a strict Lyapunov function for $(0, 0)$.

(1) $x' = x + 2y, \quad y' = -y$

(2) $x' = x^2 - 2xy, \quad y' = y^2 - x^2$

(3) $x' = -\sin^2 x \sin y, \quad y' = -2 \sin x \cos x \cos y$

3.6.2. Hamiltonian Systems.

Hamiltonian systems, to be defined below, often arise in classical mechanics. While a gradient system is associated with a potential function (or potential energy) f, a Hamiltonian system is often associated with the total energy H, called the Hamiltonian function of the system. To ease our discussion, we will only deal with two dimensional Hamiltonian systems in this note.

Definition 3.27 (Hamiltonian System)**.** A two dimensional Hamiltonian system is of the form:

$$\begin{aligned} p' &= -\frac{\partial H}{\partial q} \\ q' &= \frac{\partial H}{\partial p} \end{aligned}$$

where $H(p, q) : \mathbb{R}^2 \to \mathbb{R}$ is a C^2 function, called a Hamiltonian function.

Remark 3.28. Here we use (p, q) to denote the coordinates of $\mathbb{R}^2$ instead of the usual (x, y) because of the physics origin of Hamiltonian systems. In classical mechanics, p represents *momentum* and q represents *position*. □

The significance of a Hamiltonian function H in terms of stability is that H is a possible *candidate* of a (non-strict) Lyapunov function.

Lemma 3.29. *The Hamiltonian function $H(p, q)$ of the system:*

$$\begin{aligned} p' &= -\frac{\partial H}{\partial q} \\ q' &= \frac{\partial H}{\partial p} \end{aligned}$$

satisfies:

$$\frac{d}{dt} H(\mathbf{x}(t)) = 0$$

for any solution curves $\mathbf{x}(t)$. In other words, the value of H is constant along any trajectory in the phase portrait.

Proof. Let $\mathbf{x}(t) = (p(t), q(t))$. By chain rule,

$$\begin{aligned}\frac{d}{dt}H(p(t), q(t)) &= \frac{\partial H}{\partial p} \cdot \frac{dp}{dt} + \frac{\partial H}{\partial q} \cdot \frac{dq}{dt} \\ &= \frac{\partial H}{\partial p} \cdot \left(-\frac{\partial H}{\partial q}\right) + \frac{\partial H}{\partial q} \cdot \frac{\partial H}{\partial p} \\ &= 0.\end{aligned}$$

□

Therefore, H always satisfies Property 3 of Lyapunov functions. In order for H to be a Lyapunov function, we need the following conditions stated in the following theorem:

Theorem 3.30. *Consider the Hamiltonian system*

$$\begin{aligned}p' &= -\frac{\partial H}{\partial q} \\ q' &= \frac{\partial H}{\partial p}\end{aligned}$$

Suppose $\mathbf{x}^*$ is an equilibrium point of the system and is an isolated local minimum point of H, then $L(\mathbf{x}) := H(\mathbf{x}) - H(\mathbf{x}^*)$ is a Lyapunov function for $\mathbf{x}^*$ and hence $\mathbf{x}^*$ is stable.

Proof. The proof is almost identical to the gradient system case (Theorem 3.25) and so it is omitted here.

□

Example 3.9. The following is a classic example (undamped harmonic oscillator) of a Hamiltonian system:

$$\begin{aligned}p' &= -kq \\ q' &= \frac{p}{m}\end{aligned}$$

where $m, k > 0$ are constants. It can be easily verified that $H(p, q) = \frac{1}{2m}p^2 + \frac{k}{2}q^2$ is a Hamiltonian function (which represents the total energy of the system). Clearly, H is a Lyapunov function for $(0, 0)$ and so the origin is a stable equilibrium. □

Example 3.10. Consider the system:

$$\begin{aligned}p' &= q - q^2 \\ q' &= p\end{aligned}$$

By setting $q - q^2 = -\frac{\partial H}{\partial q}$ and $p = \frac{\partial H}{\partial p}$, one can easily find that $H(p, q) = \frac{p^2}{2} + \frac{q^3}{3} - \frac{q^2}{2}$ is a Hamiltonian function for the system.

The system has two equilibrium points: $(0, 0)$ and $(0, 1)$. The linearization at $(0, 0)$ is given by the matrix:

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

whose eigenvalues are 1 and -1. Therefore, the Stable Curve Theorem asserts that $(0, 0)$ is unstable.

However, the linearization method fails to conclude the stability of $(0,1)$ at which the linearized system is given by the matrix:

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

that has eigenvalues $0 \pm i$. Therefore, $(0,1)$ is not a hyperbolic equilibrium point. Fortunately, the system is a Hamiltonian system and so $L(p,q) := H(p,q) - H(0,1)$ is a possible candidate for a Lyapunov function.

Let's verify that $L(p,q) = \frac{p^2}{2} + \frac{q^3}{3} - \frac{q^2}{2} + \frac{1}{6}$ is indeed a Lyapunov function for $(0,0)$. Clearly, $\frac{p^2}{2} \geq 0$ and equality holds if and only if $p = 0$. We let $f(q) = \frac{q^3}{3} - \frac{q^2}{2} + \frac{1}{6}$, then one can verify that $f(1) = 0$, $f'(1) = 0$ and $f''(1) = 2 > 0$. Therefore, $q = 1$ is a local minimum point for $f(q)$. In other words, one can find a small $\varepsilon > 0$ such that $f(q) > f(0) = 0$ for $q \in (1-\varepsilon, 1+\varepsilon)$ and $q \neq 0$.

Combining with the p-term, we have $L(p,q) \geq 0$ when (p,q) is close to $(0,1)$, and equality holds when $(p,q) = (0,1)$. Therefore L is a Lyapunov function for $(0,1)$ and hence $(0,1)$ is stable. □

3.6.2.1. Criterion for two-dimensional Hamiltonian systems. Similar to gradient systems, there is a convenient way to determine whether a given two dimensional system is a Hamiltonian system.

Proposition 3.31. *Given a C^1 vector field $\mathbf{F}$ defined everywhere on $\mathbb{R}^2$, the system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ is a Hamiltonian system if and only if the trace of the Jacobian matrix $D\mathbf{F}_{\mathbf{x}}$ is zero for any $\mathbf{x} \in \mathbb{R}^2$.*

We sketch the proof in the exercise below and let readers complete the detail:

Exercise 3.15. Consider a C^1 vector field $\mathbf{F}(p,q) = \begin{bmatrix} F_1(p,q) \\ F_2(p,q) \end{bmatrix}$ defined everywhere on $\mathbb{R}^2$, and the system:

$$\begin{aligned} p' &= F_1(p,q) \\ q' &= F_2(p,q) \end{aligned}$$

Complete the proof of Proposition 3.31 based on the following outline:

(1) Prove the $(\Rightarrow)$-part, i.e. assume there exists a Hamiltonian function $H(p,q)$ such that $F_1 = -\frac{\partial H}{\partial q}$ and $F_2 = \frac{\partial H}{\partial p}$, show that the trace of $D\mathbf{F}$ is zero.

(2) For the $(\Leftarrow)$-part, we assume the trace of $D\mathbf{F}$ is zero, i.e. $\frac{\partial F_1}{\partial p} + \frac{\partial F_2}{\partial q} = 0$. Define H as the following line integral:

$$H(p,q) := \int_L -F_1 dq + F_2 dp$$

where L is the straight-line segment from $(0,0)$ to (p,q), parametrized by $\mathbf{r}(t) = (tp, tq)$, $0 \leq t \leq 1$. Show that H is explicitly given by:

$$H(p,q) = \int_0^1 \left(-F_1(tp,tq)q + F_2(tp,tq)p\right) dt.$$

(3) Prove that $\frac{\partial H}{\partial p}(p,q) = F_2(p,q)$ and $-\frac{\partial H}{\partial q}(p,q) = F_1(p,q)$ for any $(p,q) \in \mathbb{R}^2$, hence completing the proof of the $(\Leftarrow)$-part. You may need to use the chain rule twice. Look at the proof of Proposition 3.26 as a reference.

Exercise 3.16. For each of the following systems, determine whether it is a Hamiltonian system or not. If so, find the Hamiltonian function and determine whether it is a Lyapunov function for $(0,0)$.

(1) $p' = p + 2q, \quad q' = -q$

(2) $p' = p^2 - 2pq, \quad q' = q^2 - 2pq$

(3) $p' = -\sin^2 p \sin q, \quad q' = -2\sin p \cos p \cos q$

Chapter 4

Periodicity

A solution $\mathbf{x}(t)$ to an ODE system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ is said to be **periodic** if the trajectory goes back to the original position after some finite time. While it is easy to determine periodic solutions in *some* systems, it is in general difficult to find an explicit expression of periodic solutions in many systems.

In this chapter, we introduce an important theorem (the Poincaré-Bendixson's Theorem) which can be used to show the *existence* of periodic solutions. The Poincaré-Bendixson's Theorem is significant in both pure and applied mathematics. It is a consequence of the Jordan Curve Theorem, a celebrated result in topology. Practically, it not only tells us whether a periodic solution exists, but also suggests where it is located. Furthermore, the Poincaré-Bendixson's Theorem is significant in chaotic theory in a sense that it shows there is no chaotic behavior on the plane. It prompted mathematicians to consider non-planar or higher dimensional systems when studying chaotic behavior.

4.1. Periodic Solutions

Definition 4.1 (Periodic Solutions)**.** Given an autonomous system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$, a solution $\mathbf{x}(t)$ of the system is said to be a **periodic solution**, or a **closed orbit**, if there exists a time $T > 0$ such that $\mathbf{x}(t + T) = \mathbf{x}(t)$ for any $t \in \mathbb{R}$.

Remark 4.2. If one denotes φ_t the flow of the system, then in terms of flow notations, a periodic solution is a trajectory $\varphi_t(\mathbf{x}_0)$ satisfying $\varphi_{t+T}(\mathbf{x}_0) = \varphi_t(\mathbf{x}_0)$ for all $t \in \mathbb{R}$. □

Remark 4.3. Assume the vector field $\mathbf{F}$ is C^1 so that the uniqueness theorem holds for the system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ and hence we have $\varphi_t \circ \varphi_s = \varphi_{t+s}$ for any legitimate t and s. Then, in order for $\varphi_t(\mathbf{x}_0)$ to be periodic, it suffices to have a time $T > 0$ such that $\varphi_T(\mathbf{x}_0) = \mathbf{x}_0$, since it automatically implies $\varphi_{t+T}(\mathbf{x}_0) = \varphi_t(\mathbf{x}_0)$ for any $t \in \mathbb{R}$. □

Remark 4.4. An equilibrium solution is *a fortiori* periodic since T can be taken to be any positive number. A non-equilibrium periodic solution can be called a **non-trivial periodic solution**. □

As mentioned in the introduction of this chapter, it is in general difficult to solve for periodic solutions explicitly. However, there are some exceptions. The planar linear systems that give a center phase portrait are clear examples. Another is the following example due to Hopf:

Example 4.1 (Hopf)**.** Consider the system:

$$x' = x - y - x(x^2 + y^2)$$
$$y' = x + y - y(x^2 + y^2)$$

While it seems difficult to solve the system using Cartesian coordinates, it is much nicer if one converts it into polar coordinates. Under the transformation rule $r^2 = x^2 + y^2$ and $\tan\theta = \frac{y}{x}$, we leave it as an exercise for readers to verify that the above system can be rewritten as:

$$r' = r(1 - r^2)$$
$$\theta' = 1$$

Therefore, if the initial data $\mathbf{x}_0$ is on the unit circle $r = 1$, then it will stay on it for all time. In polar coordinates, this solution can be explicitly written as $r(t) = 1$ and $\theta(t) = t + \theta_0$ where θ_0 is the initial angle from the positive x-axis. Convert this solution back to Cartesian coordinates, it is written as: $\mathbf{x}(t) = (\cos(t + \theta_0), \sin(t + \theta_0))$. Clearly, $T = 2\pi$ is the period of the solution, i.e. $\mathbf{x}(t + 2\pi) = \mathbf{x}(t)$ for any $t \in \mathbb{R}$.

In general, if the initial data has polar coordinates (r_0, θ_0), then the solution to the system is given by

$$r(t) = \frac{e^t}{\sqrt{\left(\frac{1}{r_0^2} - 1\right) + e^{2t}}}, \quad \theta(t) = t + \theta_0.$$

Therefore, the trajectories off the unit circle are never periodic since $r(t)$ is either strictly decreasing (when $r_0 > 1$) or strictly increasing (when $0 < r_0 < 1$). In either case, the radius $r(t) \to 1$ as $t \to +\infty$. Therefore, these trajectories are approaching to the unit circle. See Figure 4.1 for the phase portrait. □

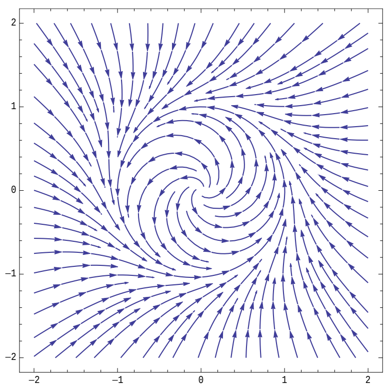


Figure 4.1. The phase portrait of the system in Example 4.1: $r' = r(1 - r^2), \quad \theta' = 1$

Unlike the Hopf's system (Example 4.1), there are many systems whose solution are difficult to solve, let alone finding periodic solutions. Many of these systems are nonetheless of important scientific significance. The following is such an example. Consider the system:

$$x' = -x + ay + x^2 y$$
$$y' = b - ay - x^2 y$$

where $a,\ b > 0$ are two parameters.

This system governs the *glycolysis* inside a human body. Here x is the concentration of ADP (adenosine diphosphate) and y is the concentration of F6P (fructose 6-phosphate). The rate of change of each of the two chemicals are governed by the above ODE system. For instance, one can see that the increase of y will lead to slower growth rate of y and higher growth rate of x. In order to maintain a sustainable metabolism, it will be ideal if the concentrations of these two chemicals exhibit a periodic pattern. Mathematically speaking, one hopes that the solution curves to the system are periodic, or at least "asymptotically periodic" just like the trajectories near the unit circle in the Hopf's system. The phase portrait of the glycolysis system (with $a = \frac{1}{10}$ and $b = \frac{1}{2}$) is shown in Figure 4.2. The phase portrait suggests that there should be a periodic solution. However, due to some unavoidable numerical errors of the plotting software, the periodic solution cannot be clearly shown in the diagram. We will show that such a periodic solution indeed *exists* using the Poincaré-Bendixson's Theorem in the next section.

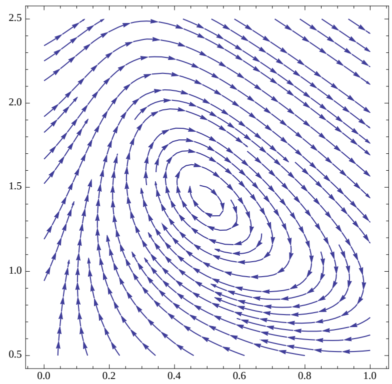


Figure 4.2. The phase portrait of the system: $\frac{dx}{dt} = -x + \frac{1}{10}y + x^2 y$ $\frac{dy}{dt} = \frac{1}{2} - \frac{1}{10}y - x^2 y$.

Before we get to the above-mentioned Poincaré-Bendixson's Theorem, let's look at the flip side of the coin that some systems never have (non-trivial) periodic solutions.

Gradient systems are such examples:

Proposition 4.5. *The only periodic solutions for any gradient system* $\mathbf{x}' = -\nabla f$ *are the equilibrium solutions.*

Proof. Suppose $\mathbf{x}(t)$ is a periodic solution of the system $\mathbf{x}' = -\nabla f$ with period $T > 0$. Then,

$$\begin{aligned}
f(\mathbf{x}(T)) - f(\mathbf{x}(0)) &= \int_0^T \frac{d}{dt} f(\mathbf{x}(t))dt && \text{(Fundamental Theorem of Calculus)} \\
&= \int_0^T \sum_{i=1}^d \frac{\partial f}{\partial x_i} \frac{dx_i}{dt} dt && \text{(chain rule)} \\
&= \int_0^T \sum_{i=1}^d \frac{\partial f}{\partial x_i} \left(-\frac{\partial f}{\partial x_i} \right) dt && \text{(since } \mathbf{x}' = -\nabla f) \\
&= -\int_0^T \sum_{i=1}^d \left(\frac{\partial f}{\partial x_i} \right)^2 dt \\
&= -\int_0^T |\nabla f|^2 \, dt.
\end{aligned}$$

Since we assume $\mathbf{x}(t)$ is periodic with period T, we have $\mathbf{x}(T) = \mathbf{x}(0)$, and so we have

$$0 = \int_0^T |\nabla f|^2 \, dt$$

which implies $\nabla f(\mathbf{x}(t)) = \mathbf{0}$ for any $t \in [0, T]$. In other words, $\mathbf{x}(t)$ is an equilibrium solution of the system.

□

Another result to rule out the periodic solutions of some systems is the following:

Theorem 4.6 (Bendixson-Dulac's Theorem). *Let $\mathbf{F} : \mathbb{R}^2 \to \mathbb{R}^2$ be a C^1 vector field on $\mathbb{R}^2$. Denote the components of $\mathbf{F}$ by:*

$$\mathbf{F}(x, y) = \begin{bmatrix} F_1(x, y) \\ F_2(x, y) \end{bmatrix}.$$

If there exists a C^1 scalar function $h(x, y) : \Omega \to \mathbb{R}$ defined on a simply-connected region $\Omega \subset \mathbb{R}^2$, such that

$$\frac{\partial(hF_1)}{\partial x} + \frac{\partial(hF_2)}{\partial y}$$

is positive in Ω, then the system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ does not have any non-trivial periodic solution inside Ω.

Proof. Suppose $\mathbf{x}(t)$ is a periodic solution with period $T > 0$. We will use the Green's Theorem to derive a contradiction.

The solution curve $\mathbf{x}(t)$, $0 \leq t \leq T$ forms a closed loop. Denote this closed curve by C and the region enclosed by R, then the Green's Theorem shows:

$$\int_C -hF_2 dx + hF_1 dy = \iint_R \frac{\partial(hF_1)}{\partial x} + \frac{\partial(hF_2)}{\partial y} dA$$

which is positive by the hypothesis of the theorem.

However, under the system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$, we have $dx = x'dt = hF_1 dt$ and $dy = y'dt = hF_2 dt$. Therefore,

$$\int_C -hF_2 dx + hF_1 dy = \int_C (-h^2 F_1 F_2 + h^2 F_1 F_2) dt = 0.$$

Clearly, it is a contradiction.

□

Example 4.2. The following system is a *modified* predator-prey model:

$$\begin{aligned} x' &= x(1 - ax - by) \\ y' &= y(1 + cx - dy) \end{aligned}$$

where a, b, c, $d > 0$ are parameters, and x and y represent the population of two species. Think about which species is the predator and which is the prey! Using the Bendixson-Dulac's Theorem, one can show this system does not have non-trivial periodic solutions:

Let $\Omega = \{(x, y) \in \mathbb{R}^2 : x > 0 \text{ and } y > 0\}$ which is a simply-connected region. Define $h(x, y) = -\frac{1}{xy}$, then

$$\frac{\partial(h \cdot x(1 - ax - by))}{\partial x} + \frac{\partial(h \cdot y(1 + cx - dy))}{\partial y} = \frac{a}{y} + \frac{d}{x} > 0$$

for any $(x, y) \in \Omega$. □

Remark 4.7. There is no general tip on the choice of a suitable h. The use of Bendixson-Dulac's Theorem requires some trial-and-errors and experience. □

Exercise 4.1. For each of the following systems, determine whether it has a non-trivial periodic solution. If so, find a periodic solution explicitly. If not, prove that it does not have any periodic solution. (Hint: for second order systems, rewrite them into first-order systems first)

(1) $x'' = -\omega_1^2 x$ where ω_1 is a constant.

(2) $x'' = -\omega_1^2 x, \quad y'' = -\omega_2^2 y$ where ω_i's are constants such that $\frac{\omega_1}{\omega_2}$ is rational.

(3) $x'' = -\omega_1^2 x, \quad y'' = -\omega_2^2 y$ where ω_i's are constants such that $\frac{\omega_1}{\omega_2}$ is irrational.

(4) $x' = x^2 - 4x^3 y, \quad y' = y^2 - x^4$

(5) $x' = \sin y, \quad y' = \cos x + y$

(6) $x' = \frac{1}{1+x^2+y^2}, \quad y' = \frac{y}{1+x^2+y^2}$

Exercise 4.2. Let $\mathbf{F} : \mathbb{R}^d \to \mathbb{R}^d$ be a C^1 vector field. Suppose $f : \mathbb{R}^d \to \mathbb{R}$ is a C^1 scalar function such that for any solution curve $\mathbf{x}(t)$ of the system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$, we have

$$\frac{d}{dt} f(\mathbf{x}(t)) \leq 0 \quad \text{for any } t \in \mathbb{R}.$$

Prove that any periodic solution (if there is any) must lie on a level set of f.

4.2. Poincaré-Bendixson's Theorem: applications

While the periodic solution in the Hopf's system (Example 4.1) is easy to find once the system is converted into polar coordinates, it is usually extremely difficult to do so in most systems. Even for scientific relevant systems such as the glycolysis model (shown in Figure 4.2 of Chapter 1) where periodic solutions are of practical importance, the closed orbit can be roughly seen in the phase portrait but it cannot be well illustrated there. It is due to numerical errors which unavoidably appear in the graph plotting process by the software.

In this section, we will use an important result in qualitative theory of ODEs, the Poincaré-Bendixson's Theorem, to prove that a periodic solution really exists in some planar ODE systems including the glycolysis model. While the theorem cannot tell what is the explicit expression of the periodic solution, it gives us an idea of where the closed orbit is located in the phase portrait. Theoretically speaking, the proof of the theorem by itself is a beautiful one too.

The statement and the proof of the Poincaré-Bendixson's Theorem involve some topological concepts such as openness and closedness. It is recommended for readers to review these concepts discussed in Chapter 2 before going ahead.

Theorem 4.8 (Poincaré-Bendixson's Theorem)**.** *Let $\mathbf{F} : \mathbb{R}^2 \to \mathbb{R}^2$ be a C^1 vector field in $\mathbb{R}^2$ and consider the system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$. Suppose K is a set in $\mathbb{R}^2$ such that:*

(1) K is closed and bounded;

(2) the system has no equilibrium point in K; and

(3) K contains a forward trajectory of the system, i.e. there exists $\mathbf{x}_0 \in K$ such that $\varphi_t(\mathbf{x}_0) \in K$ for any $t \geq 0$. Here φ_t denotes the flow of the system.

Then, the system has a non-trivial closed orbit in K.

Yes! The theorem seems to good to be true. In order to guarantee a periodic solution, one simply needs to exhibit a forward trajectory which is trapped inside K. This forward trajectory by itself needs not be periodic, but the theorem shows that if such a trajectory exists, then it will warrant a closed orbit for the system provided that K fulfills the assumption of the theorem!

We will give the proof of the Poincaré-Bendixson's Theorem in the next subsection. Meanwhile let's go through some examples to illustrate the use of the theorem.

One typical technique for applying the Poincaré-Bendixson's Theorem is to construct a **trapping region** in the phase portrait, so that trajectories starting from any point in the region will stay there for any positive time.

Example 4.3. The Hopf's system in Example 4.1 has an explicit periodic solution, namely the unit circle. Let's pretend we don't know this and try to use the Poincaré-Bendixson's Theorem to prove that a periodic solution exists!

Let K be the following closed and bounded subset of $\mathbb{R}^2$:

$$K = \left\{ \mathbf{x} \in \mathbb{R}^2 : \frac{1}{2} \leq |\mathbf{x}| \leq 2 \right\}$$

which is an annular region with outer radius 2 and inner radius $\frac{1}{2}$. The boundary of K consists of a circle of radius $\frac{1}{2}$ and a circle of radius 2, both centered at the origin. See Figure 4.3.

Under the Hopf's system, $r' = r(1-r^2)$. Therefore, on the boundary circle $\{r = 2\}$, we have $r' = r(1-r^2) = -6 < 0$, and hence trajectories hitting $\{r = 2\}$ will decrease it's distance from the origin as t increases. On the other hand, on another boundary circle $\{r = \frac{1}{2}\}$, we have $r' = r(1-r^2) = \frac{3}{8} > 0$, and hence trajectories hitting $\{r = \frac{1}{2}\}$ will increase r as t increases. These show any trajectories in the annular region K will stay in K for any future time. Also, as the system is C^1, by Theorem 2.31 there will not be any finite-time singularity, so the trajectories trapped inside must be defined for all time $t \in [0, \infty)$.

Now that K is closed and bounded. The only equilibrium point of the system, the origin, is not in K. From the above discussion, K contains many forward trajectories (in particular it contains at least one). All these fulfill the conditions of the Poincaré-Bendixson's Theorem, so the system has a non-trivial closed orbit in K. Of course, the closed orbit as we figured out before is the unit circle. $\square$

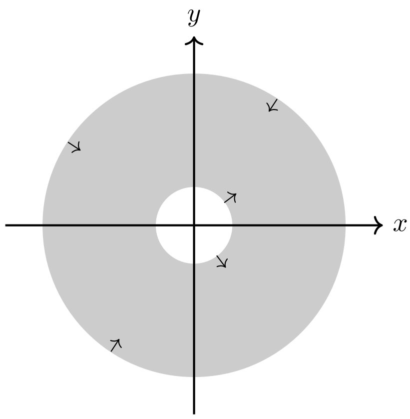


Figure 4.3. The trapping region K, shaded in gray, of the Hopf's system in Example 4.3 with sample of vectors on the boundaries.

Example 4.4. Consider the system:

$$\begin{aligned} x' &= y \\ y' &= -x + y(1 - x^2 - 1.01y^2) \end{aligned}$$

Unlike the Hopf's system, this system is not easy to be solved explicitly. However, its phase portrait looks like a slightly distorted Hopf's system portrait, and it seems the same annular region $K = \{\frac{1}{2} \le r \le 2\}$ we considered before should also be a trapping region for this system. We will verify that it is indeed the case:

Under this system, one can easily verify that:

$$\frac{d}{dt}(x^2 + y^2) = y^2(1 - x^2 - 1.01y^2).$$

On the boundary circle $\{r = 2\}$, we have

$$\frac{d}{dt}(x^2 + y^2) \le y^2(1 - x^2 - y^2) = y^2(1 - (x^2 + y^2)) = y^2(1 - 2^2) \le 0.$$

Note that we used the fact that $-1.01y^2 \le -y^2$.

On another boundary circle $\{r = \frac{1}{2}\}$, we have

$$\frac{d}{dt}(x^2 + y^2) \ge y^2(1 - 1.01x^2 - 1.01y^2) = y^2(1 - 1.01 \times (0.5)^2) \ge 0.$$

Therefore, K is a trapping region and it is closed and bounded. An easy calculation shows the origin is the only equilibrium point for the system, and it is not inside K. The Poincaré-Bendixson's Theorem shows there is a closed orbit in K. See Figure 4.4 for the diagram showing the phase portrait inside the trapping region K. □

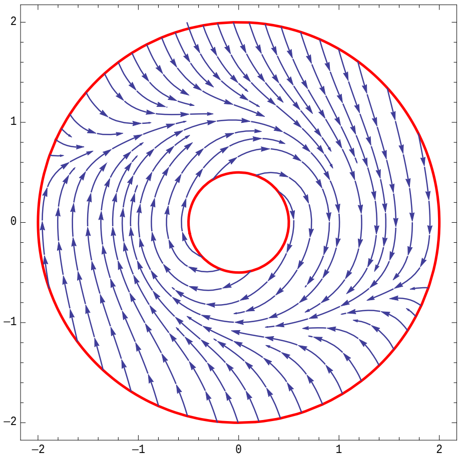


Figure 4.4. The solution curves inside the trapping region K in Example 4.4.

Example 4.5. Here we give an example where the trapping region is not an annular region. It is the glycolysis model we mentioned at the beginning of this chapter:

$$x' = -x + ay + x^2 y$$
$$y' = b - ay - x^2 y$$

where $a,\ b > 0$ are two parameters.

The phase portrait with a specific choice of (a, b) was shown in Figure 4.2. We will show the quadrilateral region K with vertex $(0, 0)$, $(b + \frac{b}{a}, 0)$, $(b, \frac{b}{a})$ and $(0, \frac{b}{a})$ is a trapping region for the system. See Figure 4.5 for the sketch of the region.

To show it is a trapping region, we need to show the vector field $\mathbf{F}(x, y) = \begin{bmatrix} -x + ay + x^2 y \\ b - ay - x^2 y \end{bmatrix}$ is pointing into the region near the boundary, or equivalently, $\mathbf{F} \cdot \mathbf{n} > 0$ where $\mathbf{n}$ is an inward normal vector of the boundary.

There are four boundary components, three of which are either horizontal or vertical. Let's verify two of them and the other two are left as an exercise.

On the boundary segment joining $(0, 0)$ and $(b + \frac{b}{a}, 0)$, we have $y = 0$ (when x is varying) and the inward normal vector $\mathbf{n}$ is $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$, and we have:

$$\mathbf{F}(x, 0) \cdot \mathbf{n} = (-x, b) \cdot (0, 1) = b > 0.$$

Hence $\mathbf{F}$ is pointing inward on this boundary component.

The boundary component joining $(b + \frac{b}{a}, 0)$ and $(b, \frac{b}{a})$ can be expressed as $y = -x + b + \frac{b}{a}$, with $x \in [b, b + \frac{b}{a}]$, and the inward normal vector $\mathbf{n}$ is $(-1, -1)$. Therefore,

$$\mathbf{F}(x, y) \cdot \mathbf{n} = -(-x + ay + x^2 y) - (b - ay - x^2 y) = x - b \geq 0$$

since $x \geq b$.

After verifying the other two boundary components, we can conclude K is a trapping region. K is closed and bounded. Unfortunately, there is an equilibrium point $(b, \frac{b}{a+b^2})$ which is inside K! One cannot apply the Poincaré-Bendixson's Theorem with this K directly. However, it is still possible to show existence of periodic solution if $(b, \frac{b}{a+b^2})$ can be shown to be unstable, since then one can drill a small open ball $B_\varepsilon((b, \frac{b}{a+b^2}))$ inside K and $K \backslash B_\varepsilon((b, \frac{b}{a+b^2}))$ is a closed and bounded trapping region for the system not containing any equilibrium point. The Poincaré-Bendixson's Theorem hence shows there is a periodic solution inside the region $K \backslash B_\varepsilon((b, \frac{b}{a+b^2}))$.

Unfortunately, the equilibrium is unstable only for some pairs of (a, b), for instance $a = \frac{1}{10}$ and $b = \frac{1}{2}$ (as you should verify as an exercise). There are many other good pairs, but it is quite tedious (yet possible) to figure out all possible pairs for which the equilibrium point is unstable. See Figure 4.6 for the phase portrait inside the trapping region for this pair of (a, b). $\square$

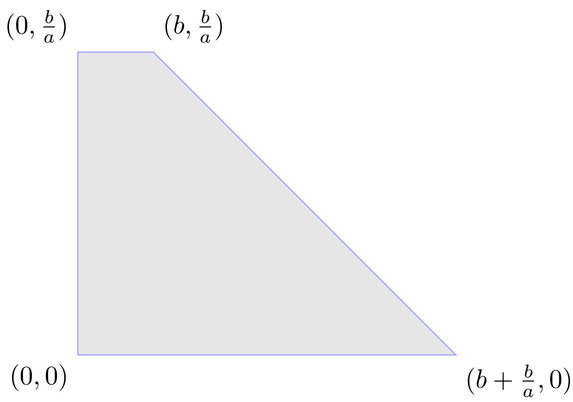


Figure 4.5. The trapping region in Example 4.5.

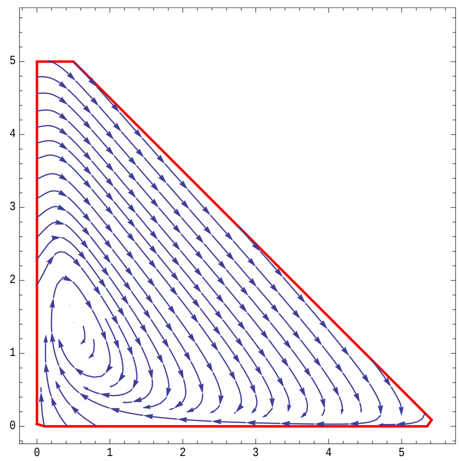


Figure 4.6. The phase portrait inside the trapping region in Example 4.5 when $(a, b) = (\frac{1}{10}, \frac{1}{2})$.

Exercise 4.3. Complete the verification that K defined in Example 4.5 is a trapping region for the system for any pair of $a, b > 0$. Find another pair of (a, b), other than $(\frac{1}{10}, \frac{1}{2})$, such that K satisfies all conditions of the Poincaré-Bendixson's Theorem.

Exercise 4.4. Show that the system has a (non-trivial) periodic solution:

$$x' = -x(x^2 + y^2 - 3x - 1) + y$$
$$y' = -y(x^2 + y^2 - 3x - 1) - x$$

[Hint: first convert the system into polar coordinates, then find a suitable pair of r_1 and r_2 such that $K = \{r_1 \leq |\mathbf{x}| \leq r_2\}$ is a trapping region for the system.]

Exercise 4.5. Show that if a (two-dimensional) Hamiltonian system with a Hamiltonian function H has the property that the set $K = \{(p, q) : a \leq H(p, q) \leq b\}$ is non-empty, closed and bounded for some a and b, then K contains an equilibrium or a closed orbit (or both).

4.3. Poincaré-Bendixson's Theorem: the proof

This section is devoted to the proof of the Poincaré-Bendixson's Theorem. The key idea follows closely from Chapter 10 in Hirsch-Smale-Devaney's book but here we will minimize the technicality in the exposition while keeping the proof rigorous.

4.3.1. Limit Sets.

An important concept in the proof of the Poincaré-Bendixson's Theorem is the $\alpha-$ and $\omega-$ limit sets to be defined below.

Let φ_t be the flow of the Hopf's system discussed in Example 4.1. Consider the trajectory $\varphi_t(\mathbf{x}_0)$ for some point $\mathbf{x}_0 \in \mathbb{R}^2$ with polar coordinates $(r_0, 0)$. As we computed before, the trajectory is given in polar coordinates by:

$$r(t) = \frac{e^t}{\sqrt{\left(\frac{1}{r_0^2} - 1\right) + e^{2t}}}, \quad \theta(t) = t,$$

or equivalently in (x, y)-coordinates:

$$\varphi_t(\mathbf{x}_0) = (x(t), y(t)) = \frac{e^t}{\sqrt{\left(\frac{1}{r_0^2} - 1\right) + e^{2t}}}(\cos t, \sin t).$$

Although the scaling factor $\frac{e^t}{\sqrt{\left(\frac{1}{r_0^2} - 1\right) + e^{2t}}}$ approaches to 1 as $t \to +\infty$, the limit $\lim_{t\to+\infty} \varphi_t(\mathbf{x}_0)$ does not exist because the trigonometric functions $\cos t$ and $\sin t$ are oscillating between -1 and 1 rather than converging to specific numbers.

However, if one substitute t by a suitable time sequence $\{t_n\}_{n=1}^{\infty}$ which approaches to $+\infty$ as $n \to \infty$, then one can possibly talk about convergence of $\varphi_{t_n}(\mathbf{x}_0)$ as $n \to \infty$. For example, if we let $t_n = 2\pi n$, then

$$\begin{aligned}\varphi_{t_n}(\mathbf{x}_0) &= \frac{e^{2\pi n}}{\sqrt{\left(\frac{1}{r_0^2} - 1\right) + e^{4\pi n}}}(\cos(2\pi n), \sin(2\pi n)) \\ &= \frac{e^{2\pi n}}{\sqrt{\left(\frac{1}{r_0^2} - 1\right) + e^{4\pi n}}}(1, 0).\end{aligned}$$

Letting $n \to \infty$ gives $\varphi_{t_n}(\mathbf{x}_0) \to (1, 0)$ as $n \to \infty$.

That says, although we do not have convergence for $\varphi_t(\mathbf{x}_0)$ when t is regarded as a *continuous* parameter, we can still talk about a *discrete* notion of convergence by substituting t by a suitable sequence t_n. The cost is that now the "limit" may not be unique. For instance, if one choose $t_n = 2\pi n + \theta_0$ where θ_0 is any fixed angle, then one should verify that $\varphi_{t_n}(\mathbf{x}_0) \to (\cos\theta_0, \sin\theta_0)$ as $n \to +\infty$, which is another point on the unit circle.

Under this generalized notion of limits, we no longer say $\varphi_t(\mathbf{x}_0)$ converges to a particular *point*, but rather say $\varphi_t(\mathbf{x}_0)$ approaches to a *set*. This motivates the following definition:

Definition 4.9 (Limit Points and Limit Sets). Let φ_t be the flow of an ODE system on $\mathbb{R}^d$. A point $\mathbf{y} \in \mathbb{R}^d$ is called an ω-**limit point** of $\mathbf{x}_0$ if there exists a time sequence $t_n \to +\infty$ as $n \to \infty$ such that $\varphi_{t_n}(\mathbf{x}_0) \to \mathbf{y}$ as $n \to \infty$. The ω-**limit set of** $\mathbf{x}_0$, denoted by $\omega(\mathbf{x}_0)$, is the set of all possible ω-limit points of $\mathbf{x}_0$. Precisely,

$$\omega(\mathbf{x}_0) = \{\mathbf{y} \in \mathbb{R}^d : \exists t_n \to +\infty \text{ as } n \to \infty \text{ such that } \varphi_{t_n}(\mathbf{x}_0) \to \mathbf{y} \text{ as } n \to \infty\}.$$

A point $\mathbf{z} \in \mathbb{R}^d$ is called an α-**limit point** of $\mathbf{x}_0$ if there exists a time sequence $t_n \to -\infty$ as $n \to \infty$ such that $\varphi_{t_n}(\mathbf{x}_0) \to \mathbf{y}$ as $n \to \infty$. The α-**limit set of** $\mathbf{x}_0$, denoted by $\alpha(\mathbf{x}_0)$, is the set of all possible α-limit points of $\mathbf{x}_0$. Precisely,

$$\alpha(\mathbf{x}_0) = \{\mathbf{z} \in \mathbb{R}^d : \exists t_n \to -\infty \text{ as } n \to \infty \text{ such that } \varphi_{t_n}(\mathbf{x}_0) \to \mathbf{z} \text{ as } n \to \infty\}.$$

□

Remark 4.10. The letters α and ω are chosen because they are the first and the last Greek alphabet respectively. □

Example 4.6. Recall that the flow of the Hopf's system is given by

$$\varphi_t(\mathbf{x}_0) = (x(t), y(t)) = \frac{e^t}{\sqrt{\left(\frac{1}{r_0^2} - 1\right) + e^{2t}}}(\cos t, \sin t),$$

where $\mathbf{x}_0 = (r_0, 0)$ in (x, y)-coordinates and $r_0 > 0$.

As discussed before, there exists a sequence of times $t_n = 2\pi n + \theta_0 \to +\infty$ as $n \to \infty$ such that $\varphi_{t_n}(\mathbf{x}_0) \to (\cos\theta_0, \sin\theta_0)$. One can pick θ_0 to be any angle , so any point on the unit circle is an ω-limit point of $\mathbf{x}_0$. Conversely, any ω-limit point of $\mathbf{x}_0$ must be on the unit circle since $|\varphi_{t_n}(\mathbf{x}_0)| \to 1$ as $n \to \infty$ for any sequence $t_n \to +\infty$. Therefore, the ω-limit set of $\mathbf{x}_0$ is the unit circle. Symbolically, we denote it by:

$$\omega(\mathbf{x}_0) = \{\mathbf{y} \in \mathbb{R}^2 : |\mathbf{y}| = 1\}.$$

The α-limit points of $\mathbf{x}_0$ is bit more subtle than their ω-counterparts. If $\mathbf{x}_0 = (r_0, 0)$ is chosen such that $0 < r_0 < 1$, then $\frac{1}{r_0^2} - 1$ is positive and so $\sqrt{\left(\frac{1}{r_0^2} - 1\right) + e^{2t}}$ is defined for all time t. Therefore it makes sense to talk about $\varphi_{t_n}(\mathbf{x}_0)$ for sequences $t_n \to -\infty$. One can verify that in this case $\varphi_{t_n}(\mathbf{x}_0) \to (0, 0)$ for any sequence $t_n \to -\infty$, and so $\mathbf{0}$ is the only α-limit point of $\mathbf{x}_0 = (r_0, 0)$. Symbolically, it is defined by:

$$\alpha(\mathbf{x}_0) = \{\mathbf{0}\} \quad \text{when } 0 < r_0 < 1.$$

However, $\mathbf{x}_0 = (r_0, 0)$ with $r_0 > 1$. The square-root $\sqrt{\left(\frac{1}{r_0^2} - 1\right) + e^{2t}}$ is undefined when t is sufficiently negative. It is therefore forbidden to substitute t by a sequence t_n that goes to $-\infty$. Therefore, there is no α-limit point for this $\mathbf{x}_0$, and symbolically we say:

$$\alpha(\mathbf{x}_0) = \emptyset \quad \text{when } r_0 > 1.$$

□

Exercise 4.6. If $\mathbf{x}_0$ lies on the unit circle, then what are the limit sets $\omega(\mathbf{x}_0)$ and $\alpha(\mathbf{x}_0)$ under the Hopf's system? How about $\omega(\mathbf{0})$ and $\alpha(\mathbf{0})$?

Exercise 4.7. Show that if $\varphi_t(\mathbf{x}_0)$ is a periodic solution, then both $\omega(\mathbf{x}_0)$ and $\alpha(\mathbf{x}_0)$ are equal to the trajectory $\{\varphi_t(\mathbf{x}_0)\}_{t\in(-\infty,\infty)}$ itself.

If $\mathbf{x}^*$ is an equilibrium point, then what are $\omega(\mathbf{x}^*)$ and $\alpha(\mathbf{x}^*)$?

We will mostly deal with ω-limits in the rest of the chapter. While it is possible to determine the limit sets for the Hopf's system where the flow map can be explicitly stated, it is in general difficult to determine limit sets for most nonlinear systems. In the rest of the chapter, we will deal with limit sets in a qualitative way rather than finding them explicitly.

The following lemma presents some important facts about ω-limit sets. They will be used often when establishing the Poincaré-Bendixson's Theorem.

Lemma 4.11. *Let φ_t be the flow of a C^1-system on $\mathbb{R}^d$. Suppose $\mathbf{y} \in \omega(\mathbf{x})$ for some $\mathbf{x} \in \mathbb{R}^d$, then we have:*

(1) $\varphi_s(\mathbf{y}) \in \omega(\mathbf{x})$ for any s (as long as $\varphi_s(\mathbf{y})$ exists).

(2) If $\mathbf{z} = \varphi_s(\mathbf{y})$ for some fixed s, i.e. $\mathbf{z}$ is on the trajectory through $\mathbf{y}$, then $\mathbf{z} \in \omega(\mathbf{x})$.

(3) If $\mathbf{w} \in \omega(\mathbf{y})$, i.e. $\mathbf{w}$ is an ω-limit point of $\mathbf{y}$, then we also have $\mathbf{w} \in \omega(\mathbf{x})$. [In other words, $\mathbf{w} \in \omega(\mathbf{y})$ and $\mathbf{y} \in \omega(\mathbf{x})$ imply $\mathbf{w} \in \omega(\mathbf{x})$.]

Proof. Given that $\mathbf{y} \in \omega(\mathbf{x})$, there exists a sequence of times $t_n \to +\infty$ such that $\lim_{n\to\infty} \varphi_{t_n}(\mathbf{x}) = \mathbf{y}$.

Parts 1 and 2 are easy consequences of the continuity of φ_s proved in Proposition 2.51 in Chapter 2:

To prove (1), we consider $\varphi_{s+t_n}(\mathbf{y}) = \varphi_s(\varphi_{t_n}(\mathbf{y}))$. Since φ_s is continuous, we have:

$$\lim_{n\to\infty} \varphi_{s+t_n}(\mathbf{x}) = \lim_{n\to\infty} \varphi_s(\varphi_{t_n}(\mathbf{x})) = \varphi_s\left(\lim_{n\to\infty} \varphi_{t_n}(\mathbf{x})\right) = \varphi_s(\mathbf{y}).$$

Therefore, $\varphi_s(\mathbf{y}) \in \omega(\mathbf{x})$ and the associated time sequence is $s + t_n$.

For (2), we consider:

$$\begin{aligned} z &= \varphi_s(\mathbf{y}) && \text{(given)} \\ &= \varphi_s\left(\lim_{n\to\infty} \varphi_{t_n}(\mathbf{x})\right) && \text{(given)} \\ &= \lim_{n\to\infty} \varphi_s(\varphi_{t_n}(\mathbf{x})) && (\varphi_s \text{ is continuous}) \\ &= \lim_{n\to\infty} \varphi_{t_n+s}(\mathbf{x}). \end{aligned}$$

Therefore, z is an ω-limit point of $\mathbf{x}$ since there exists a time sequence $t_n + s \to +\infty$ such that $\varphi_{t_n+s}(\mathbf{x}) \to \mathbf{z}$. It completes the proof of (2).

For (3), since $\mathbf{w} \in \omega(\mathbf{y})$ there exists a sequence of times $s_k \to \infty$ such that $\varphi_{s_k}(\mathbf{y}) \to \mathbf{w}$ as $k \to \infty$. Given that $\lim_{n\to\infty} \varphi_{t_n}(\mathbf{x}) = \mathbf{y}$ and by the continuity of φ_{s_k}, we have $\lim_{n\to\infty} \varphi_{s_k+t_n}(\mathbf{x}) = \varphi_{s_k}(\mathbf{y})$ for each fixed k. We pick a subsequence t_{n_k} of t_n such that for each k, we have

$$\left|\varphi_{s_k+t_{n_k}}(\mathbf{x}) - \varphi_{s_k}(\mathbf{y})\right| < \frac{1}{k}.$$

Consider the sequence of times $s_k + t_{n_k}$, we then have:

$$\begin{aligned}\left|\varphi_{s_k+t_{n_k}}(\mathbf{x}) - \mathbf{w}\right| &= \left|\varphi_{s_k+t_{n_k}}(\mathbf{x}) - \varphi_{s_k}(\mathbf{y}) + \varphi_{s_k}(\mathbf{y}) - \mathbf{w}\right| \\ &\leq \left|\varphi_{s_k+t_{n_k}}(\mathbf{x}) - \varphi_{s_k}(\mathbf{y})\right| + \left|\varphi_{s_k}(\mathbf{y}) - \mathbf{w}\right| \\ \left|\varphi_{s_k+t_{n_k}}(\mathbf{x}) - \mathbf{w}\right| &\leq \underbrace{\frac{1}{k}}_{\to 0} + \underbrace{\left|\varphi_{s_k}(\mathbf{y}) - \mathbf{w}\right|}_{\to 0}.\end{aligned}$$

As $k \to +\infty$, we have $\left|\varphi_{s_k+t_{n_k}}(\mathbf{x}) - \mathbf{w}\right| \to 0$, or in other words $\varphi_{s_k+t_{n_k}}(\mathbf{x}) \to \mathbf{w}$. Therefore, $\mathbf{w} \in \omega(\mathbf{x})$ and its associated time sequence is $s_k + t_{n_k}$.

□

4.3.1.1. Closedness and boundedness of the trapping region. Recall there are two conditions for the trapping region K in the statement of the Poincaré-Bendixson's Theorem, namely K has to be closed and bounded. These two conditions have important implications in terms of limit sets.

Suppose $\varphi_t(\mathbf{x}_0)$ is a forward trajectory contained in K entirely. If K were not bounded, then $\varphi_t(\mathbf{x}_0)$ may diverge to infinity as $t \to +\infty$ then there is no ω-limit point to talk about. The boundedness of K guarantees there is at least one ω-limit point of $\mathbf{x}_0$. In fact, it is a consequence of the following famous theorem in analysis:

Theorem 4.12 (Bolzano-Weierstrass's Theorem). *If S is a bounded infinite set in $\mathbb{R}^d$, then there exists a sequence $\mathbf{s}_n \in S$ such that $\mathbf{s}_n$ converges to a limit $\mathbf{s}_0$ in $\mathbb{R}^d$ as $n \to \infty$.*

We omit the proof here. A standard proof can be found in any basic analysis textbook, and is normally taught in the first course of analysis.

Now applying the Bolzano-Weierstrass's Theorem to our scenario. The forward trajectory $\varphi_t(\mathbf{x}_0)$ is an infinite set (unless $\mathbf{x}_0$ is an equilibrium point, but then $\omega(\mathbf{x}_0)$ is $\{\mathbf{x}_0\}$ itself). If it is completely inside a bounded set K, then the trajectory is a bounded infinite set so the theorem implies there exists a sequence $\varphi_{t_n}(\mathbf{x}_0)$ that converges to a limit $\mathbf{y}$ in $\mathbb{R}^d$. Consequently, $\omega(\mathbf{x})$ contains at least one point $\mathbf{y}$.

The boundedness of K guarantees the forward trajectory has at least one ω-limit point. However, boundedness *alone* cannot guarantee the limit point must be in K. That's why we need to combine closedness with boundedness. The following is a "common-sense" fact in analysis and point-set topology:

Proposition 4.13. *Let K be a closed set in $\mathbb{R}^d$. If $\mathbf{x}_n$ is a sequence in K and that $\mathbf{x}_n \to \mathbf{y}$ as $n \to \infty$, then the limit $\mathbf{y}$ must be in K.*

Proof. We prove by contradiction. Suppose $\mathbf{y}$ is not in K, then $\mathbf{y} \in \mathbb{R}^d \backslash K$. Since K is closed, the complement $\mathbb{R}^d \backslash K$ is open. By the definition of openness, there exists a ball $B_\varepsilon(\mathbf{y})$ that is contained inside $\mathbb{R}^d \backslash K$.

The sequence $\mathbf{x}_n \to \mathbf{y}$ as $n \to \infty$, so $\mathbf{x}_n$ will eventually enter the ball $B_\varepsilon(\mathbf{y})$ for sufficiently large n. However, it is not possible since all $\mathbf{x}_n$'s are in K but the ball $B_\varepsilon(\mathbf{y})$ is disjoint from K.

Therefore, we must have $\mathbf{y} \in K$.

□

Combining closedness and boundedness of K, the forward trajectory $\varphi_t(\mathbf{x}_0)$ trapped inside K must have at least one ω-limit point, and all ω-limit points must be in K. This is significant since then our proof "game" will be confined in the trapping region K.

Remark 4.14. In **finite-dimensional** spaces such as $\mathbb{R}^d$, a closed and bounded set can be called a *compact* set. □

4.3.2. Local Sections and Flow Boxes. The Poincaré-Bendixson's Theorem requires the trapping region K has no equilibrium point for the system. We will explore why this is needed in this subsection.

From now on, we will restrict the discussion to planar system only. Let $\mathbf{x}_0$ be a non-equilibrium point of a C^1-system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$. The vector $\mathbf{F}(\mathbf{x}_0)$ at $\mathbf{x}_0$ is non-zero, and so there is a straight line, denoted by $l(\mathbf{x}_0)$, passing through $\mathbf{x}_0$ and is perpendicular to $\mathbf{F}(\mathbf{x}_0)$. Pick a point $\mathbf{x}$ on this line $l(\mathbf{x}_0)$, then one can tell whether $\mathbf{F}(\mathbf{x})$ is pointing at the same side of the line as $\mathbf{F}(\mathbf{x}_0)$ by considering the dot product $\mathbf{F}(\mathbf{x}) \cdot \mathbf{F}(\mathbf{x}_0)$. If the dot product is positive, then $\mathbf{F}(\mathbf{x})$ points at the same side of the line as $\mathbf{F}(\mathbf{x}_0)$.

Since $\mathbf{F}(\mathbf{x}_0) \cdot \mathbf{F}(\mathbf{x}_0) = |\mathbf{F}(\mathbf{x}_0)|^2 > 0$, by continuity of the vector field, the dot product $\mathbf{F}(\mathbf{x}) \cdot \mathbf{F}(\mathbf{x}_0)$ must be positive as far as $\mathbf{x}$ is sufficiently close to $\mathbf{x}_0$. Consequently, one can find a line segment $\mathcal{S}_{\mathbf{x}_0}$ of $l(\mathbf{x}_0)$ such that at every point $\mathbf{x}$ on this line segment $\mathcal{S}_{\mathbf{x}_0}$, the vector field $\mathbf{F}(\mathbf{x})$ is pointing at the same side of $l(\mathbf{x}_0)$ as $\mathbf{F}(\mathbf{x}_0)$. This line segment is called:

Definition 4.15 (Local Sections). Let $\mathbf{x}_0$ be a non-equilibrium point of a C^1-system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$. A **local sections** $\mathcal{S}_{\mathbf{x}_0}$ is a line segment passing through $\mathbf{x}_0$ and perpendicular to $\mathbf{F}(\mathbf{x}_0)$ such that $\mathbf{F}(\mathbf{x}) \cdot \mathbf{F}(\mathbf{x}_0) > 0$ for any $\mathbf{x} \in \mathcal{S}_{\mathbf{x}_0}$. □

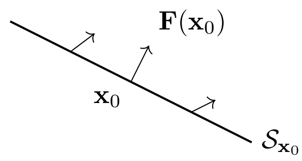


Figure 4.7. A local section $\mathcal{S}_{\mathbf{x}_0}$ based at $\mathbf{x}_0$.

Given a local section $\mathcal{S}_{\mathbf{x}_0}$, one can construct a **flow box** at $\mathbf{x}_0$ to be described below. As the vector field $\mathbf{F}(\mathbf{x})$ is pointing at the same side of the local section $\mathcal{S}_{\mathbf{x}_0}$ when $\mathbf{x}$ is sufficiently close to $\mathbf{x}_0$. One can expect there is a neighborhood $\mathcal{V}$ of $\mathbf{x}_0$ so that the trajectories inside $\mathcal{V}$ are *flowing in approximately parallel directions* as shown in Figure 4.8. A flow box has two edges: the in-edge and the out-edge. A flow box is characterized by the following properties:

(1) Any trajectory must enter the flow box $\mathcal{V}$ through its in-edge.
(2) After a trajectory enters the flow box $\mathcal{V}$, it must intersect the local section $\mathcal{S}_{\mathbf{x}_0}$ exactly once before leaving $\mathcal{V}$.
(3) Any trajectory must leave the flow box $\mathcal{V}$ through its out-edge.

The formal construction of flow boxes is bit technical so we omit it here. Readers may consult Section 10.2 of Hirsch-Smale-Devaney's book for both the formal definition and the existence proof of flow boxes using the Implicit Function Theorem. In order to understand the key idea of the Poincaré-Bendixson's Theorem, it is more important to keep in mind the *geometric intuition* of flow boxes, rather than knowing the formal definition or why flow boxes must exist.

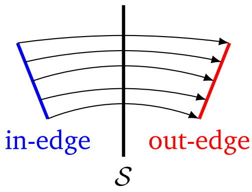


Figure 4.8. An example of a flow box and a local section.

4.3.3. Jordan Curve Theorem and Consequences. Limit sets, local sections and flow boxes are three key ingredients in the proof of the Poincaré-Bendixson's Theorem. In this subsection, we will first state (but not prove) a celebrated result in topology, the Jordan Curve Theorem. It will lead to two important consequences about limit sets and local sections.

The statement of the Jordan Curve Theorem, stated below, sounds quite trivial and you may wonder why such an obvious statement can be qualified as a theorem. Nonetheless, the proof requires an advanced concept called Homology which is usually taught in graduate level Algebraic Topology course.

Theorem 4.16 (Jordan Curve Theorem)**.** *Any continuous simple closed curve C in the plane $\mathbb{R}^2$ divides the plane into two disjoint components, i.e. there exist two disjoint connected open sets U and V such that $\mathbb{R}^2 \backslash C = U \cup V$. Moreover, one of the U and V is bounded and the other one must be unbounded.*

The first consequence of the Jordan Curve Theorem is about monotonicity:

Lemma 4.17. *Let φ_t be the flow of a C^1 planar system, and let $\mathcal{S}$ be any local section. Consider a trajectory $\varphi_t(\mathbf{y})$ from a point $\mathbf{y}$ in $\mathbb{R}^2$. If $t_1 < t_2 < t_3$ are times at which the trajectory $\varphi_t(\mathbf{y})$ intersects $\mathcal{S}$, then the intersection points $\varphi_{t_1}(\mathbf{y})$, $\varphi_{t_2}(\mathbf{y})$ and $\varphi_{t_3}(\mathbf{y})$ must be in monotonic order on the local section $\mathcal{S}$ (see Figures 4.9 and 4.10 for an example and a non-example of monotonically ordered points).*

Proof. First construct a continuous simple closed curve C by gluing the part of the trajectory $\varphi_t(\mathbf{y})$ for $t \in [t_1, t_2]$ and the line segment joining $\varphi_{t_1}(\mathbf{y})$ and $\varphi_{t_2}(\mathbf{y})$. The Jordan Curve Theorem asserts that C divides the plane $\mathbb{R}^2$ into two disjoint open sets U and V. Assume without loss of generality that the trajectory $\varphi_t(\mathbf{y})$ enters the region U shortly after t_2. Suppose at a later time t_3, the trajectory intersects $\mathcal{S}$ in the middle of $\varphi_{t_1}(\mathbf{y})$ and $\varphi_{t_2}(\mathbf{y})$ (let's call this $1-3-2$ configuration), then $\mathcal{S}$ being a local section implies the trajectory must come from another region V shortly before t_3 (see Figure 4.11). However, it is impossible since U and V are disjoint. It rules out the $1-3-2$ arrangement on the local section $\mathcal{S}$. Similarly, one can also rule out the $3-1-2$ configuration by the same argument. Therefore, the only possibility is $1-2-3$, which is exactly what we need to show.

□

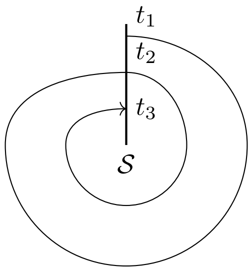


Figure 4.9. $\varphi_{t_i}(\mathbf{x}_0)$'s are monotonically ordered on $\mathcal{S}$

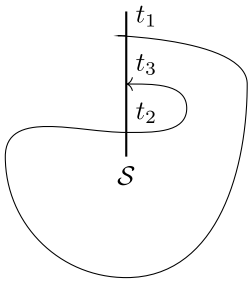


Figure 4.10. $\varphi_{t_i}(\mathbf{x}_0)$'s are **not** monotonically ordered on $\mathcal{S}$

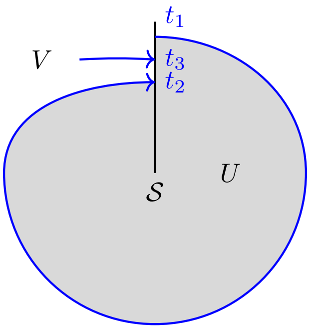


Figure 4.11. The trajectory in blue is a hypothetical trajectory that gives a $1-3-2$ configuration. This configuration is ruled out by the Jordan Curve Theorem.

If we further assume that the point $\mathbf{y}$ of Lemma 4.17 is an ω-limit point of another point $\mathbf{x}$, then we have a stronger result:

Lemma 4.18. *Let φ_t be the flow of a C^1 planar system, and let $\mathcal{S}$ be any local section. If a trajectory $\varphi_t(\mathbf{y})$ starts from a point $\mathbf{y} \in \omega(\mathbf{x})$ for some $\mathbf{x} \in \mathbb{R}^2$, then the trajectory $\varphi_t(\mathbf{y})$ intersects $\mathcal{S}$ at at most one point.*

Remark 4.19. The trajectory can intersect $\mathcal{S}$ for infinitely many *times*, but the lemma shows the intersection *point* must be the same every time. □

Proof. We prove by contradiction. Suppose $\varphi_t(\mathbf{y})$ intersects $\mathcal{S}$ at two different points $\mathbf{z}$ and $\mathbf{w}$. One can then find two disjoint flow boxes, $\mathcal{V}$ based at $\mathbf{z}$ and another $\mathcal{W}$ based at $\mathbf{w}$.

As $\mathbf{z}$ and $\mathbf{w}$ are on the trajectory from $\mathbf{y}$, they are both ω-limit points of $\mathbf{x}$ by Lemma 4.11. As a result, there exist sequences $t_n \to +\infty$ and $s_n \to +\infty$ such that $\varphi_{t_n}(\mathbf{x}) \to \mathbf{z}$ and $\varphi_{s_n}(\mathbf{x}) \to \mathbf{w}$ as $n \to \infty$.

There must be infinitely many $\varphi_{t_n}(\mathbf{x})$'s in $\mathcal{V}$ since $\varphi_{t_n}(\mathbf{x})$ converges to $\mathbf{z}$ which is inside $\mathcal{V}$. Therefore, by the property of a flow box, the trajectory $\varphi_t(\mathbf{x})$ must enter the flow box for infinitely many times and intersect the $\mathcal{V}$-portion of the local section $\mathcal{S}$ for infinitely many times. See Figure 4.12.

Similarly, the trajectory $\varphi_t(\mathbf{x})$ must intersect the $\mathcal{W}$-portion of the local section $\mathcal{S}$ for infinitely many times. However, Lemma 4.17 shows $\varphi_t(\mathbf{x})$ must intersect the local section $\mathcal{S}$ in monotonic order. It is impossible to have this trajectory intersecting the $\mathcal{V}$- and $\mathcal{W}$-portions of the local section $\mathcal{S}$ both for infinitely many times and overall in a monotonic order. It leads to a contradiction. Therefore, $\varphi_t(\mathbf{y})$ cannot intersect $\mathcal{S}$ at two different points, and hence it can only intersect $\mathcal{S}$ at at most one point.

□

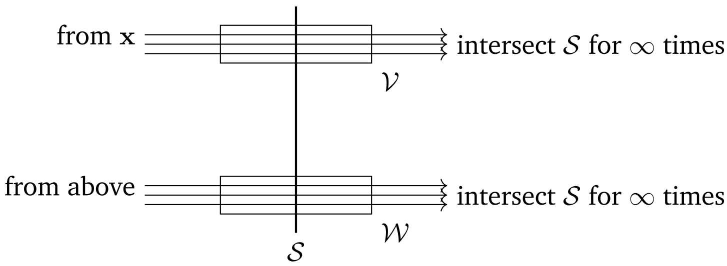


Figure 4.12. The trajectory from $\mathbf{x}$ cannot intersect $\mathcal{S}$ first in $\mathcal{V}$ and then $\mathcal{W}$ both for infinitely many times in a monotonic manner. This leads to a contradiction. Note that the trajectory from $\mathbf{y}$ is not shown in the figure since it is not relevant.

4.3.4. Completion of the Proof: A Tale of Three Points. Finally, with Lemma 4.18, we are ready to give the proof of the Poincaré-Bendixson's Theorem.

Proof of Theorem 4.8. Recall that the set-up is that there is a closed and bounded set K in $\mathbb{R}^2$ that contains a forward trajectory $\varphi_t(\mathbf{x})$ for $t \in [0, \infty)$. By the closedness and boundedness of K, the limit set $\omega(\mathbf{x})$ is non-empty (by Bolzano-Weierstrass) and is contained inside K (by closedness). The absence of equilibrium point in K and that $\omega(\mathbf{x}) \subset K$ guarantee every point on $\omega(\mathbf{x})$ has a local section and a flow box around the point.

Now let $\mathbf{y}$ be any point in $\omega(\mathbf{x})$. **The key idea of proving the theorem is to show that $\varphi_t(\mathbf{y})$ is a periodic solution**. In order to prove this, consider a point $\mathbf{z} \in \omega(\mathbf{y})$. Since $\mathbf{y} \in \omega(\mathbf{x})$, Lemma 4.11 shows $\varphi_t(\mathbf{y}) \in \omega(\mathbf{x}) \subset K$ for any $t \geq 0$. Consequently, the closedness of K implies $\omega(\mathbf{y}) \subset K$, and so $\mathbf{z} \in K$.

Now that $\mathbf{z} \in K$, it is not an equilibrium point and so there exists a local section $\mathcal{S}$ and a flow box $\mathcal{V}$ based at $\mathbf{z}$. Now the proof is completed by applying Lemma 4.18: since $\mathbf{y}$ is an ω-limit point of $\mathbf{x}$, the lemma shows that the trajectory $\varphi_t(\mathbf{y})$ intersects the local section $\mathcal{S}$ at at most one point. Since $\mathbf{z}$ is an ω-limit point of $\mathbf{y}$, it implies $\varphi_t(\mathbf{y})$ must enter the flow box $\mathcal{V}$ for at infinitely many times, and intersect $\mathcal{S}$ for infinitely many times but every time the intersection point must be the same. Therefore, one can pick two different times s and t, where $s < t$, such that $\varphi_t(\mathbf{y}) = \varphi_s(\mathbf{y})$, which implies $\varphi_{t-s}(\mathbf{y}) = \mathbf{y}$. In other words, the trajectory $\varphi_t(\mathbf{y})$ is periodic with a period $t - s > 0$. It completes the proof.

□

We would like to end this section by stating (but not proving) two extended forms of the Poincaré-Bendixson's Theorem. In the version we stated in Theorem 4.8, if K is a closed and bounded set in $\mathbb{R}^2$, contains no equilibrium point of the system and contains a forward trajectory $\varphi_t(\mathbf{x})$, then the system has a periodic solution. From the proof, we have seen that the periodic solution is given by $\varphi_t(\mathbf{y})$ where $\mathbf{y} \in \omega(\mathbf{x})$. By Lemma 4.11, $\mathbf{y} \in \omega(\mathbf{x})$ implies $\varphi_t(\mathbf{y}) \in \omega(\mathbf{x})$ for any t, and so the periodic trajectory $\varphi_t(\mathbf{y})$ must be a subset of $\omega(\mathbf{x})$.

In fact, one can show that the periodic trajectory $\varphi_t(\mathbf{y})$ is all of $\omega(\mathbf{x})$:

Theorem 4.20 (Poincaré-Bendixson's Theorem: extension 1)**.** *Assume K and the system satisfy all conditions of Theorem 4.8, and that the forward trajectory is given by $\varphi_t(\mathbf{x}) \in K$, then the system has a non-trivial closed orbit in K and the image of the closed orbit is equal to $\omega(\mathbf{x})$.*

Proof. Readers may consult Section 10.5 in Hirsch-Smale-Devaney's book, or Chapter 7 of *Ordinary Differential Equations: Qualitative Theory* by Barreria and Valls.

□

Another extension of the Poincaré-Bendixson's Theorem concerns about the possibility of having equilibrium points inside the trapping region K:

Theorem 4.21 (Poincaré-Bendixson's Theorem: extension 2)**.** *Assume K is a closed and bounded set in $\mathbb{R}^2$ and it contains a forward trajectory $\varphi_t(\mathbf{x})$ of a planar system, then the limit set $\omega(\mathbf{x})$ must be one of the following:*

(1) a non-trivial closed orbit; or

(2) an equilibrium point; or

(3) a connected set of finitely many equilibrium points together with non-periodic trajectories connecting them.

Proof. Readers may consult Teschl's book for a proof of this extension.

□

4.3.5. Limit Cycles and the Hilbert's 16th Problem. We would like to end this course by keeping readers informed of a long-standing unsolved problem. In order to state the problem, we need to introduce:

Definition 4.22 (Limit Cycles)**.** A limit cycle γ is a non-trivial periodic solution to an ODE system and it is a limit set of a point not lying on γ, i.e. $\gamma = \omega(\mathbf{x})$ or $\gamma = \alpha(\mathbf{x})$ for some $\mathbf{x} \in \mathbb{R}^2 \backslash \gamma$. □

The unit circle of the Hopf's system is a limit cycle because it is the ω-limit set of the trajectories spiraling towards it. However, the closed orbits of a center phase portrait for linear system are not limit cycles because they are not limit sets for nearby trajectories but rather they are limit sets of itself.

The Poincaré-Bendixson's Theorem (extension 1) asserts that for a planar system if a non-periodic forward trajectory from $\mathbf{x}$ is trapped inside a closed and bounded set K which contains no equilibrium, then the trajectory limits to a periodic solution. The image of the closed orbit is equal to $\omega(\mathbf{x})$. This closed orbit is a limit cycle. Therefore, the Poincaré-Bendixson's Theorem gives a lower bound (i.e. at least one) of the number of limit cycles in this scenario.

However, it is an extremely difficult problem to determine an upper bound on the number of limit cycles for planar systems. The problem was proposed by David Hilbert in 1900, among a list of 23 problems. The one about limit cycles, listed as the 16th problem, is stated as follows:

Problem 4.23 (Hilbert's 16th Problem). Let $P(x, y)$ and $Q(x, y)$ be two real polynomials of degree n. Consider the planar system

$$x' = P(x, y)$$
$$y' = Q(x, y).$$

Find an upper bound (in terms of n) of the number of limit cycles to this system. □

As of today (January 10, 2020), the problem is unsolved for all $n > 1$. The best result obtained so far is by Yulii Ilyashenko and Jean Écalle in 1991/92 who proved that these systems have only finitely many limit cycles.

– The End of the Course –

Appendix

A.1. Uniform Convergence

This appendix reviews some basic knowledge about uniform convergence.

A.1.0.1. Pointwise convergence. The Picard's iteration sequence $\mathbf{x}_n(t)$ is a sequence of *functions*, not a sequence of *numbers*. There are two 'variables' involves: one is an integer n and another is the time t. Pointwise convergence, or pointwise limit, means we are taking $n \to \infty$ while regarding t as a constant.

Definition A.1 (Pointwise Convergence). Given a sequence of functions $\{\mathbf{x}_n(t)\}_{n=1}^{\infty}$ where t is defined on an interval I, we say $\mathbf{x}_n(t)$ converges *pointwise* to a function $\mathbf{x}_\infty(t)$ on I if for each fixed $t \in I$, the sequence $\{\mathbf{x}_n(t)\}$ regarding t constant converges to $\mathbf{x}_\infty(t)$ as $n \to \infty$, or equivalently, for each $t \in I$

$$|\mathbf{x}_n(t) - \mathbf{x}_\infty(t)| \to 0 \quad \text{as } n \to \infty.$$

Example A.1. The sequence of functions $x_n(t) = (1 - t^2)^n$ converges pointwise on $[0, 1]$ to the function:

$$x_\infty(t) = \begin{cases} 0 & \text{if } 0 < t \leq 1 \\ 1 & \text{if } t = 0 \end{cases}$$

Therefore, it is possible for a sequence of continuous functions converges pointwise to a discontinuous function! □

Example A.2. Denote a sequence of functions as follows:

$$x_n(t) = \begin{cases} n^2 t & \text{if } 0 \leq t \leq 1/n \\ -n^2 t + 2n & \text{if } 1/n \leq t < 2/n \\ 0 & \text{if } 2/n \leq t \leq 1 \end{cases}$$

The graph of the function can be found in Figure A.1. By the definition, $x_n(0) = 0$ for any n, therefore $x_n(0) \to 0$ as $n \to \infty$. For any $t \in (0, 1]$, one can always find a large enough N such that $2/N \leq t \leq 1$, and therefore for any $n \geq N$, we have $x_n(t) = 0$ and so $x_n(t) \to 0$ as $n \to \infty$. Therefore, $x_n(t) \to 0$ for $t \in [0, 1]$ as $n \to \infty$.

From the graph of the function $x_n(t)$, one can easily see that:

$$\int_0^1 x_n(t)dt = 1 \quad \text{for any } n \geq 2$$

but

$$\int_0^1 \lim_{n\to\infty} x_n(t)dt = \int_0^1 0dt = 0.$$

Therefore, this example shows it is possible that:

$$\lim_{n\to\infty} \int_0^1 x_n(t)dt = 1 \neq 0 = \int_0^1 \lim_{n\to\infty} x_n(t)dt$$

and so the limit and integral signs cannot be always switched! The step we 'cheated' in page 41 needs to be justified! □

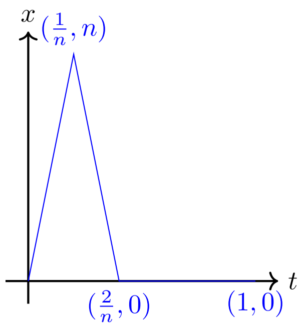


Figure A.1. The graph of the function $x_n(t)$ defined in Example A.2.

A.1.0.2. Uniform convergence: definition. We ask: Is there any condition for the sequence of a function which allows us to switch the limit and integral signs?

The answer is positive: we require a *stronger* type of convergence called *uniform convergence*. To start with, we define a norm of a function.

Definition A.2 (L^∞-Norm of a Function). Given a bounded function $\mathbf{x}(t) : I \to \mathbb{R}^d$ defined on an interval I, the L^∞*-norm* of the function over the interval I is defined as:

$$\|\mathbf{x}\|_\infty := \sup\{|\mathbf{x}(t)| : t \in [a, b]\}.$$

In other words, $\|\mathbf{x}\|_\infty$ measures the largest magnitude of $\mathbf{x}(t)$ among all t in the given interval $[a, b]$.

Remark A.3. Note that the L^∞-norm $\|\cdot\|_\infty$ depends on the time interval I. Larger time interval may give a larger L^∞-norm. If the time interval I plays an essential role in some of our arguments, you should indicate on which time interval the L^∞-norm is defined, by either declaring the interval in the text, or write $\|\cdot\|_{\infty, I}$. □

Remark A.4. While the magnitude $|\mathbf{x}(t)|$ depends on t, the L^∞-norm $\|\mathbf{x}\|_\infty$ doesn't. Therefore, it is recommended *not* to write the t-variable for the function $\mathbf{x}$ inside the L^∞-norm, i.e write $\|\mathbf{x}\|_\infty$ but not $\|\mathbf{x}(t)\|_\infty$. □

Remark A.5. The name L^∞-norm refers to the fact that it is the limit of the L_p-norm as $p \to \infty$. The L_p-norm of a continuous function $\mathbf{x}(t)$ on the time interval $[a, b]$ is defined as:

$$\|\mathbf{x}\|_p := \left(\int_a^b |\mathbf{x}(t)|^p dt\right)^{1/p}.$$

It can be shown by some analysis argument (omitted here) that $\lim_{p\to\infty} \|\mathbf{x}\|_p = \|\mathbf{x}\|_\infty$ □

Example A.3. The L^∞-norm over the interval $[0, 2\pi]$ of the functions $f(t) = \sin t$ and $g(t) = \sin t + \cos t$ are:

$$\|f\|_\infty = 1, \quad \|g\|_\infty = \sqrt{2}.$$

□

Example A.4. The following sequence of functions was investigated before:

$$x_n(t) = \begin{cases} n^2 t & \text{if } 0 \le t \le 1/n \\ -n^2 t + 2n & \text{if } 1/n \le t < 2/n \\ 0 & \text{if } 2/n \le t \le 1 \end{cases}$$

From the graph of $x_n(t)$, one can see easily that $\|x_n\|_\infty = n$ for any $n \ge 2$. □

We are now ready to state the definition of uniform convergence:

Definition A.6 (Uniform Convergence)**.** We say a sequence of functions $\mathbf{x}_n(t) : I \to \mathbb{R}^d$ converges *uniformly* on I to a function $\mathbf{y}(t) : I \to \mathbb{R}^d$ if:

$$\|\mathbf{x}_n - \mathbf{y}\|_{\infty, I} \to 0 \quad \text{as } n \to \infty.$$

Remark A.7. Just like Lipschitz continuity, the notion of uniform convergence always get tied with the time interval because the L^∞-norm depends on the time interval chosen. It is possible for a sequence converges uniformly to a limit on one interval but not on a larger one. Therefore, it is crucial to indicate the interval whenever we talk about uniform convergence: always say $\mathbf{x}_n(t)$ converges uniformly on I to a function, rather than just saying $\mathbf{x}_n(t)$ converges uniformly to that function. □

Remark A.8. Uniform convergence implies pointwise convergence. Precisely, if $\mathbf{x}_n$ converges uniformly on $[a, b]$ to a function $\mathbf{y} : I \to \mathbb{R}^d$, then for any fixed $t \in I$, the sequence $\mathbf{x}_n(t)$ converges to $\mathbf{y}(t)$ as $n \to \infty$. This can be argued easily by squeezing principle: for each fixed $t \in I$, we have:

$$|\mathbf{x}_n(t) - \mathbf{y}(t)| \le \sup\{|\mathbf{x}_n(s) - \mathbf{y}(s)| : s \in I\} =: \|\mathbf{x}_n - \mathbf{y}\|_\infty.$$

If $\mathbf{x}_n$ converges uniformly on $[a, b]$ to $\mathbf{y}$, then the right-hand side tends to 0 as $n \to \infty$ by the definition of uniform convergence. The squeezing principle shows the left-hand side also tends to 0.

Therefore, we say uniform convergence is *stronger* than pointwise convergence. □

Example A.5. The sequence $f_n(t) := \sin \frac{t}{n}$ converges uniformly on $[0, 2\pi]$ to 0 as $n \to \infty$. It can be argued as follows: since uniform convergence implies pointwise convergence to the same limit function, the only candidate uniform convergence limit for $f_n(t)$ must be the pointwise limit 0. To show it is indeed the case, we consider:

$$\begin{aligned} |f_n(t) - 0| &= \left|\sin \frac{t}{n} - \sin 0\right| \\ &= |\cos \xi| \left|\frac{t}{n} - 0\right| && \text{(mean-value theorem)} \\ &\le \left|\frac{t}{n}\right| \le \frac{2\pi}{n}. && \text{(since } t \in [0, 2\pi]) \end{aligned}$$

The ξ above is some number in $[0, \frac{t}{n}]$.

Therefore, $|f_n(t) - 0| \leq \frac{2\pi}{n}$ for any $t \in [0, 2\pi]$, and so

$$\|f_n - 0\|_\infty := \sup\{|f_n(t) - 0| : t \in [0, 2\pi]\} \leq \frac{2\pi}{n} \to 0 \quad \text{as } n \to \infty.$$

Therefore, by squeezing principle $\|f_n - 0\|_\infty \to 0$ as $n \to \infty$. By the definition of uniform convergence, f_n converges uniformly on $[0, 2\pi]$ to 0 as $n \to \infty$. □

Example A.6. However, the sequence of 'triangle' functions $x_n(t)$ defined in Example A.2 does not converges uniformly on $[0, 1]$. Since uniform convergence implies pointwise convergence, the only candidate for the uniform convergence limit must equal to the pointwise convergence limit, which is 0 for this example.

However, $\|x_n - 0\|_\infty = n$ which does not converge to 0 as $n \to \infty$. Therefore, x_n does not converge uniformly on $[0, 1]$ to any limit function. □

A.1.0.3. Weierstrass's M-test. Given an infinite series of functions $\sum_{n=1}^{\infty} \mathbf{a}_n(t)$, defined on a time interval I, we say it converges pointwise on I if for every fixed $t \in I$, we have

$$\sum_{n=1}^{N} \mathbf{a}_n(t) \to \sum_{n=1}^{\infty} \mathbf{a}_n(t) \quad \text{as } N \to \infty,$$

regarding the series as a sequence indexed by N. Likewise, we say $\sum_{n=1}^{\infty} \mathbf{a}_n$ converges uniformly on I if

$$\left\| \sum_{n=1}^{N} \mathbf{a}_n - \sum_{n=1}^{\infty} \mathbf{a}_n \right\|_{\infty, I} \to 0 \quad \text{as } N \to \infty.$$

There is a test, called the Weierstrass's M-test, one can use to prove a series converges uniformly on a given interval I. This test is particularly useful when dealing uniform convergence of a series since it can bypass the calculation of the N-th partial sums $\sum_{n=1}^{N} \mathbf{a}_n(t)$ and their the L^∞-norms.

Theorem A.9 (Weierstrass's M-test)**.** *Let $I = [a, b]$ be a time interval. Let $\sum_{n=1}^{\infty} \mathbf{a}_n(t)$ be an infinite series of functions on I. Suppose:*

(1) for each n, there exists a real number M_n such that $\|\mathbf{a}_n\|_{\infty, I} \leq M_n$; and

(2) the infinite series of real numbers $\sum_{n=1}^{\infty} M_n$ converges to a finite number,

then $\sum_{n=1}^{\infty} \mathbf{a}_n$ converges uniformly on I.

Remark A.10. We call it M-test because it is a convention to use M_n to denote the upper bound for $\|\mathbf{a}_n\|_\infty$. □

Proof of Theorem A.9. Since $\sum_{n=1}^{\infty} M_n$ converges and $\|\mathbf{a}_n\|_\infty \leq M_n$, by comparison test, we know $\sum_{n=1}^{\infty} \|\mathbf{a}_n\|_\infty$ converges too. For each $t \in I$, the magnitude $|\mathbf{a}_n(t)| \leq \|\mathbf{a}_n\|_\infty$ and so by comparison test again, we know $\sum_{n=1}^{\infty} |\mathbf{a}_n(t)|$ converges. The absolute convergence test on $\mathbb{R}^d$ shows $\sum_{n=1}^{\infty} \mathbf{a}_n(t)$ converges for each $t \in I$. This shows pointwise convergence.

To prove uniform convergence, one considers the partial sums:

$$\left\| \sum_{n=1}^{N} \mathbf{a}_n - \sum_{n=1}^{\infty} \mathbf{a}_n \right\|_\infty = \left\| \sum_{n=N+1}^{\infty} \mathbf{a}_n \right\|_\infty \leq \sum_{n=N+1}^{\infty} \|\mathbf{a}_n\|_\infty \leq \sum_{n=N+1}^{\infty} M_n.$$

Since $\sum_{n=1}^{\infty} M_n$ converges, we have:

$$\sum_{n=N+1}^{\infty} M_n = \sum_{n=1}^{\infty} M_n - \sum_{n=1}^{N} M_n \to 0 \quad \text{as } N \to \infty.$$

By squeezing principle, we proved:

$$\left\| \sum_{n=1}^{N} \mathbf{a}_n - \sum_{n=1}^{\infty} \mathbf{a}_n \right\|_{\infty} \to 0 \quad \text{as } N \to \infty$$

which means, by definition, the infinite series $\sum_{n=1}^{\infty} \mathbf{a}_n$ converges uniformly on I. □

Example A.7. To apply the Weierstrass's M-test, one should:

(i) bound each $\mathbf{a}_n(t)$ by a real number M_n, which depends only on n, but not on t. Then,

$$|\mathbf{a}_n(t)| \leq M_n \text{ for each } t \in I \;\Rightarrow \|\mathbf{a}_n\|_{\infty} := \sup\{|\mathbf{a}_n(t)| : t \in I\} \leq M_n$$

which shows these M_n's satisfy the first condition of the test; then

(ii) show that the infinite series of numbers $\sum_{n=1}^{\infty} M_n$ converges using some of the series tests you learned in calculus course.

Here are some examples of using the Weierstrass's M-test:

(1) The series of functions $\sum_{n=1}^{\infty} \frac{\sin nt}{n^2}$ converges uniformly on $[0, 2\pi]$: since for each n, we have

$$\left| \frac{\sin nt}{n^2} \right| \leq \frac{1}{n^2} =: M_n \quad \text{for any } t \in [0, 2\pi],$$

and $\sum_{n=1}^{\infty} M_n = \sum_{n=1}^{\infty} \frac{1}{n^2}$ converges by p-test (where $p = 2$).

(2) The series of functions $\sum_{n=1}^{\infty} \frac{t^n}{n!}$ converges uniformly on $[-T, T]$ where $T > 0$ is a fixed real number: since for each n, we have

$$\left| \frac{t^n}{n!} \right| \leq \frac{T^n}{n!} =: M_n,$$

and $\sum_{n=1}^{\infty} M_n$ converges by ratio test:

$$\lim_{n\to\infty} \frac{T^{n+1}/(n+1)!}{T^n/n!} = \lim_{n\to\infty} \frac{T}{n} = 0 < 1.$$

Note that although this series converges uniformly on every closed and bounded interval $[-T, T]$ where T can be as large as we want, this series does *not* converge uniformly on $\mathbb{R}$.

(3) The series of vector-valued functions $\displaystyle\sum_{n=1}^{\infty} \begin{bmatrix} (\sin nt)/n^2 \\ t^n/n! \end{bmatrix}$ converges uniformly on $[-1, 1]$: since for each n, we have

$$\left| \begin{bmatrix} (\sin nt)/n^2 \\ t^n/n! \end{bmatrix} \right| = \left(\left| \frac{\sin nt}{n^2} \right|^2 + \left| \frac{t^n}{n!} \right|^2 \right)^{1/2} \leq \left(\frac{1}{n^4} + \frac{1}{(n!)^2} \right)^{1/2} =: M_n$$

for any $t \in [-1, 1]$. For large n, the term $\frac{1}{n^4}$ dominates over $\frac{1}{(n!)^2}$ and so $M_n \simeq \sqrt{\frac{1}{n^4}}$ as $n \to \infty$. Since $\sum_{n=1}^{\infty} \frac{1}{n^2}$ converges by p-test, $\sum_{n=1}^{\infty} M_n$ converges as well. One may give a more rigorous argument by limit comparison test: show that $\lim_{n\to\infty} \frac{M_n}{1/n^2}$ converges to a non-zero finite number, so that $\sum_{n=1}^{\infty} M_n$ and $\sum_{n=1}^{\infty} \frac{1}{n^2}$ must either both converge or both diverge.

□

Exercise A.1. Show that each of the following series of functions converges uniformly on the interval indicated:

(a) $\sum_{n=1}^{\infty} \frac{\cos nt}{n!}$ on $t \in [-T, T]$ where $T > 0$ is any fixed real number.

(b) $\sum_{n=1}^{\infty} \frac{t^n}{2^n}$ on $t \in [-1, 1]$.

(c) $\sum_{n=1}^{\infty} \frac{t^n}{n}$ on $t \in [0, \frac{1}{2}]$.

(d) $\displaystyle\sum_{n=1}^{\infty} \begin{bmatrix} \cos(nt)/(n!) \\ t^n/n \end{bmatrix}$ on $t \in [0, \frac{1}{2}]$.

A.1.0.4. Consequences of uniform convergence. We list some important consequences of having uniform convergence for a sequence or series of functions.

Theorem A.11. *Let I be a time interval. Suppose $\mathbf{x}_n : I \to \mathbb{R}^d$ is a sequence of integrable functions defined on I such that $\mathbf{x}_n$ converges uniformly on I to a limit function $\mathbf{y} : I \to \mathbb{R}^d$, then:*

(1) given any ***finite numbers*** *α and β such that $[\alpha, \beta] \subset I$, we have*

$$\lim_{n\to\infty} \int_{\alpha}^{\beta} \mathbf{x}_n(t)dt = \int_{\alpha}^{\beta} \lim_{n\to\infty} \mathbf{x}_n(t)dt = \int_{\alpha}^{\beta} \mathbf{y}(t)dt,$$

(2) if $\mathbf{x}_n$'s are all continuous on I, then the limit function $\mathbf{y}$ is also continuous on I.

Proof. By the definition of uniform convergence, we have $\|\mathbf{x}_n - \mathbf{y}\|_{\infty} \to 0$ as $n \to \infty$. To prove the first part, we consider:

$$\begin{aligned} \left| \int_{\alpha}^{\beta} \mathbf{x}_n(t)dt - \int_{\alpha}^{\beta} \mathbf{y}(t)dt \right| &= \left| \int_{\alpha}^{\beta} (\mathbf{x}_n(t) - \mathbf{y}(t))\, dt \right| \\ &\leq \int_{\alpha}^{\beta} |\mathbf{x}_n(t) - \mathbf{y}(t)|\, dt \\ &\leq \int_{\alpha}^{\beta} \|\mathbf{x}_n - \mathbf{y}\|_{\infty}\, dt \\ &= \|\mathbf{x}_n - \mathbf{y}\|_{\infty} \cdot (\beta - \alpha). \end{aligned}$$

In the last equality, we have used the fact the L^{∞}-norm does not depend on t. Since $\|\mathbf{x}_n - \mathbf{y}\|_{\infty} \to 0$ as $n \to \infty$, by squeezing principle, we have:

$$\lim_{n\to\infty} \int_{\alpha}^{\beta} \mathbf{x}_n(t)\, dt = \int_{\alpha}^{\beta} \mathbf{y}(t)\, dt$$

as desired.

To prove that $\mathbf{y}$ is continuous on I, take any arbitrary $t_0 \in I$ and consider $\mathbf{y}(t) - \mathbf{y}(t_0)$. For any n, we have:

$$\begin{aligned} |\mathbf{y}(t) - \mathbf{y}(t_0)| &= |\mathbf{y}(t) - \mathbf{x}_n(t) + \mathbf{x}_n(t) - \mathbf{x}_n(t_0) + \mathbf{x}_n(t_0) - \mathbf{y}(t_0)| \\ &\leq |\mathbf{y}(t) - \mathbf{x}_n(t)| + |\mathbf{x}_n(t) - \mathbf{x}_n(t_0)| + |\mathbf{x}_n(t_0) - \mathbf{y}(t_0)| \\ &\leq \|\mathbf{y} - \mathbf{x}_n\|_{\infty} + |\mathbf{x}_n(t) - \mathbf{x}_n(t_0)| + \|\mathbf{x}_n - \mathbf{y}\|_{\infty}. \end{aligned}$$

Given any $\varepsilon > 0$, since $\|\mathbf{x}_n - \mathbf{y}\|_\infty \to 0$ as $n \to \infty$, there is a large $N > 0$ such that $\|\mathbf{x}_N - \mathbf{y}\|_\infty < \varepsilon/3$. Since $\mathbf{x}_N$ is continuous on I, there exists $\delta > 0$ such that whenever $|t - t_0| < \delta$ we have $|\mathbf{x}_N(t) - \mathbf{x}_N(t_0)| < \varepsilon/3$, and therefore:

$$|\mathbf{y}(t) - \mathbf{y}(t_0)| \leq \|\mathbf{y} - \mathbf{x}_n\|_\infty + |\mathbf{x}_N(t) - \mathbf{x}_N(t_0)| + \|\mathbf{x}_n - \mathbf{y}\|_\infty \leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.$$

It shows $\mathbf{y}$ is continuous at t_0. Since t_0 is arbitrary on I, we conclude $\mathbf{y}$ is continuous on I. □

The two examples discussed in page 143 does not converge uniformly on their indicated interval, since one does not have a continuous limit function, another violates the interchange of the limit and integral signs.

Remark A.12. Note that (1) in Theorem A.11 holds only when $[\alpha, \beta]$ is a bounded interval. To deal with the swapping of limit and integral signs for imporper integrals, other tools such as Lebesgue's Dominated Convergence Theorem is needed.

For differentiations, note that uniform convergence of differentiable functions $\mathbf{x}_n(t)$ does NOT guarantee that we have $\lim_{n\to\infty} \dfrac{d}{dt}\mathbf{x}_n(t) = \dfrac{d}{dt} \lim_{n\to\infty} \mathbf{x}_n(t)$. The conditions required are listed in the theorem below:

Theorem A.13. *Let I be a time interval. Suppose $\mathbf{x}_n(t) : I \to \mathbb{R}^d$ is a sequence of differentiable functions such that:*

(1) $\mathbf{x}_n'$ converges uniformly on I to some function $\mathbf{y} : I \to \mathbb{R}^d$, and

(2) $\mathbf{x}(t)$ converges pointwise on I to some function $\mathbf{x}_\infty : I \to \mathbb{R}^d$,

then we have $\frac{d}{dt}\mathbf{x}_\infty(t) = \mathbf{y}(t)$ on I. In other words, we have

$$\frac{d}{dt} \lim_{n\to\infty} \mathbf{x}_n(t) = \lim_{n\to\infty} \mathbf{x}_n'(t).$$

One good example to demonstrate the use of the theorem is the justification of $\frac{d}{dt}e^{tA} = Ae^{tA}$ in Section 1.3.