Perturbation Theory

(Lecture Notes)

Hildeberto Jardón Kojakhmetov

Contents

Preface

These lecture notes are intended for the 2022 course Perturbation Theory at the University of Groningen. You will notice that, although the course is called Perturbation Theory, we cover also a few perturbation methods. The purpose of that is to provide an overview of what one would usually refer to "a perturbation analysis" in science. My goal with the course, and with the lecture notes, is that you will have a thorough introduction to some classical topics of perturbation methods and of perturbation theory. Hopefully, that will provide you with enough basis to explore further related topics.

These lecture notes are mostly based on the references in the bibliography, but more so on [**13**] and [**5**]. Having said that, and although I have tried to make these notes as pedagogical as possible, *this is not a textbook*. Thus, I definitely encourage you to consult the references as well.

Please communicate any comments, suggestions, corrections, etc. to h.jardon.kojakhmetov@rug.nl. I want to thank all those interested in the course, and especially those who have pointed-out several typos and mistakes.

CHAPTER I

Introduction and Motivation

In this course we are interested in understanding, for example, problems defined by the following differential equation

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = f(\boldsymbol{x}, \varepsilon), \tag{1}$$

where $\boldsymbol{x} \in \mathbb{R}^n$ denotes the state of a system, ε is a small parameter[1], and $f : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ is a sufficiently smooth vector field. The main idea of perturbation theory can be described as follows:

Assume that the dynamics of the unperturbed system

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = f(\boldsymbol{x}, 0), \tag{2}$$

are well-understood. What can we say about the behavior of (1) for ε sufficiently small, as $t \to \infty$?

Here, by "well-understood" we mean that, for example, the solutions of (2) are explicitly known, or that one has sufficiently good knowledge about the behavior of its solutions.

To start fixing ideas, let us see our first example.

EXAMPLE I.1. Consider the ODE

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \underbrace{-ax + \varepsilon h(x)}_{f(x,\varepsilon)}, \tag{3}$$

with $x \in \mathbb{R}$, and $a > 0$. If we set $\varepsilon = 0$, then we obtain the unperturbed equation

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -ax, \tag{4}$$

which has the analytic solution

$$x(t) = x(t_0)\exp(-a(t - t_0)). \tag{5}$$

This solution has an exponentially decaying behavior as shown in figure 1.

[1] That is $0 < |\varepsilon| \ll 1$

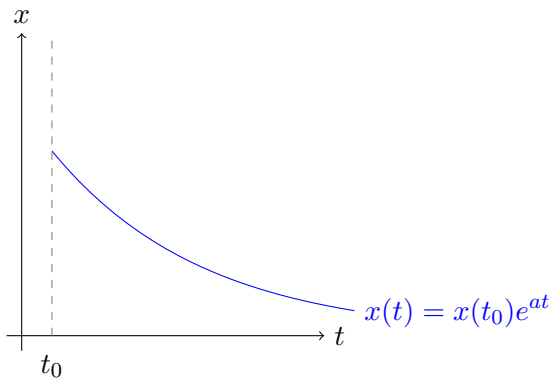


FIGURE 1. Solution of the unperturbed system (4).

We now would like to know, which functions h lead to a qualitatively similar behavior as in Figure 1 for ε sufficiently small.

In this course, among other topics, we will learn about some of the methods that can be used to answer the aforementioned question. For now, let us "naively" assume that we can write the solution of the perturbed system (3) as:

$$x_\varepsilon(t) = x_0(t) + \varepsilon x_1(t) + \varepsilon^2 x_2(t) + \cdots, \tag{6}$$

where $x_0(t)$ is given by (5), and the functions $x_i(t)$, for $i > 0$, are to be found. Regarding the initial condition, let us for simplicity assume that $x_\varepsilon(t_0) = x_0(t_0) =: x_0^*$ and $x_i(t_0) =: x_i^* = 0$ for all $i > 0$. Substituting (6) into (3) we get:

$$\frac{\mathrm{d}x_\varepsilon}{\mathrm{d}t} = -ax_\varepsilon + \varepsilon h(x_\varepsilon).$$

Before going further with the substitution, it is worth noting that we can expand $h(x_\varepsilon)$ for x_ε near x_0 as[a]:

$$\begin{aligned} h(x_\varepsilon) &\approx h(x_0) + h'(x_0)(x_\varepsilon - x_0) + h''(x_0)(x_\varepsilon - x_0)^2 + \cdots \\ &= h(x_0) + h'(x_0)(\varepsilon x_1 + \varepsilon^2 x_2 + \cdots) + \varepsilon^2 h''(x_0)(\varepsilon x_1 + \varepsilon^2 x_2 + \cdots)^2 + \cdots \\ &= h(x_0) + \varepsilon h'(x_0)x_1 + \varepsilon^2 h(x_0)x_2 + O(\varepsilon^3), \end{aligned}$$

where $\mathcal{O}(\varepsilon^3)$ denotes terms that are multiplied by ε^k with $k \geq 3$ (we will see the formal definition of the big-O symbol later). Thus, continuing with the substitution in (3):

$$\begin{aligned} \frac{\mathrm{d}x_\varepsilon}{\mathrm{d}t} &= -ax_\varepsilon + \varepsilon h(x_\varepsilon) \\ \frac{\mathrm{d}(x_0 + \varepsilon x_1 + \varepsilon^2 x_2 + \cdots)}{\mathrm{d}t} &= -a(x_0 + \varepsilon x_1 + \varepsilon^2 x_2 + \cdots) + \varepsilon h(x_0) + \varepsilon^2 h'(x_0)x_1 + O(\varepsilon^3). \end{aligned}$$

Since x_0 is a solution of the unperturbed system, that is $\dfrac{\mathrm{d}x_0}{\mathrm{d}t} = -ax_0$, we can further reduce, and identify terms that multiply the same power of ε as:

$$\varepsilon\frac{\mathrm{d}x_1}{\mathrm{d}t} + \varepsilon^2\frac{\mathrm{d}x_2}{\mathrm{d}t} + O\left(\varepsilon^3\right) = \varepsilon\big(-ax_1 + h(x_0)\big) + \varepsilon^2\big(-ax_2 + h'(x_0)x_1\big) + O\left(\varepsilon^3\right).$$

That is, for each power of ε we have the differential equations:

$$\begin{aligned}\frac{\mathrm{d}x_0}{\mathrm{d}t} &= -ax_0\\ \frac{\mathrm{d}x_1}{\mathrm{d}t} &= -ax_1 + h(x_0)\\ \frac{\mathrm{d}x_2}{\mathrm{d}t} &= -ax_2 + h'(x_0)x_1\\ &\vdots\end{aligned}$$

If we refer to the above equations as the 0th-degree equations, 1st-degree equation, and so forth, we notice that the i-th equation is a scalar non-autonomous ODE, where the time-dependent terms depend only on the solutions of kth-degree equations with $k < i$. Thus, if one can solve each of these equations, then one obtains the perturbed problem's solution (6).

To conclude this example, let us mention that a central problem when dealing with the approximations as in this example is to know whether the solution written as in (6) converges to the true analytic solution. Another important observation is that a solution as proposed in (6) is not always valid, and the choice of "anzats" usually depends on the problem at hand. The method we have employed in this example is called **Method of Series Expansion**, and is one of the most widely used methods.

[a] let us omit the argument t

EXERCISE I.1. *Consider the perturbed problem of Example I.1 with $h(x) = x^2$. The corresponding solution can be found analytically, find it. Next, assume a series solution of the form $x_\varepsilon = x_0 + \varepsilon x_1$, that is up to first degree in ε, and compute x_1. Using a computer, plot and compare the analytic solution with x_0 (the solution of the unperturbed system) and the solution $x_\varepsilon = x_0 + \varepsilon x_1$ choosing different values of $|\varepsilon| \ll 1$.*

EXAMPLE I.2. One important question regarding perturbations is whether, for example, an equilibrium persists under sufficiently small perturbations. So, let us consider the system

$$\frac{\mathrm{d}x}{\mathrm{d}t} = f(x, \varepsilon),$$

where $x \in \mathbb{R}^n$, ε is a small parameter, and f is $\mathcal{C}^r$-smooth, $r \geq 1$. Assume that $x^* \in \mathbb{R}^n$ is an equilibrium point of the unperturbed problem $\dfrac{\mathrm{d}x}{\mathrm{d}t} = f(x, 0)$, that is $f(x^*, 0) = 0$. It follows from the *implicit function theorem* that if $\dfrac{\partial f}{\partial x}(x^*, 0)$ is non-singular (has full rank), then there exists a unique local function $x_\varepsilon = x(\varepsilon)$ such that $f(x_\varepsilon, \varepsilon) = 0$. In this case we say that *the equilibrium point persists.*

EXERCISE I.2. *For the following scalar systems, locate the equilibria of the perturbed and unperturbed problems. What is the difference between them?*

- $\dfrac{\mathrm{d}x}{\mathrm{d}t} = x + \varepsilon x^2$

- $\dfrac{\mathrm{d}x}{\mathrm{d}t} = x^2 + \varepsilon x$

EXAMPLE I.3 (A projectile problem). Consider the differential equation

$$\frac{\mathrm{d}^2 y}{\mathrm{d}t^2} = -\frac{1}{(1+\varepsilon y)^2}, \qquad \text{for } t > 0. \tag{7}$$

This equation represents, after some re-scaling (see [**13**, section 1.1]) the height of an object projected radially upward the surface of the Earth. In particular, the small parameter $\varepsilon > 0$ provides a relative measurement of the height of the projectile compared with the radius of the Earth. Let us then assume that $y(0) = 0$ (the projectile is initially located at the surface of the Earth), and $\dfrac{\mathrm{d}y}{\mathrm{d}t}(0) = 1$ (some initial velocity).

Let us start with the unperturbed problem $\varepsilon = 0$. In such a case (7) reduces to

$$\frac{\mathrm{d}^2 y}{\mathrm{d}t^2} = -1,$$

which is essentially telling us that the projectile is subject to a constant downward force. The particular solution with the initial conditions $y_0(0) = 0$ and $\dfrac{\mathrm{d}y_0}{\mathrm{d}t}(0) = 1$ is:

$$y_0(t) = \frac{t}{2}(2 - t).$$

Recall that with the subscript 0 we denote the solution of the unperturbed problem.

Next, since $\varepsilon > 0$ let us suppose that we can write the solution of the perturbed problem (7) as

$$y_\varepsilon(t) = y_0(t) + \varepsilon y_1(t) + \varepsilon^2 y_2(t) + \cdots . \tag{8}$$

At this moment, let us look at the initial conditions and recall that:

$$y_\varepsilon(0) = y_0(0) + \varepsilon y_1(0) + \cdots = 0$$

and

$$\frac{\mathrm{d}y_\varepsilon}{\mathrm{d}t}(0) = \frac{\mathrm{d}y_0}{\mathrm{d}t}(0) + \varepsilon \frac{\mathrm{d}y_1}{\mathrm{d}t}(0) = 1.$$

Thus, we shall assume that $y_i(0) = \dfrac{\mathrm{d}y_i}{\mathrm{d}t}(0) = 0$ for all $i > 0$.

Substituting (8) in (7) we get[a]:

$$\frac{\mathrm{d}^2(y_0(t) + \varepsilon y_1(t) + \varepsilon^2 y_2(t) + \cdots)}{\mathrm{d}t^2} = -\frac{1}{(1 + \varepsilon(y_0(t) + \varepsilon y_1(t) + \varepsilon^2 y_2(t) + \cdots))^2}$$

$$\frac{\mathrm{d}^2 y_0(t)}{\mathrm{d}t^2} + \varepsilon \frac{\mathrm{d}^2 y_1(t)}{\mathrm{d}t^2} + \varepsilon^2 \frac{\mathrm{d}y_2(t)}{\mathrm{d}t} + \cdots \approx -1 + 2\varepsilon(y_0(t) + \varepsilon y_1(t) + \cdots) - 3\varepsilon^2 (y_0(t) + \varepsilon y_1(t) + \cdots)^2.$$

So, matching terms with the same degree of ε we get the equations:

$$\begin{aligned}
\frac{\mathrm{d}^2 y_0(t)}{\mathrm{d}t^2} &= -1, & y_0(0) = 0,\ \frac{\mathrm{d}y_0}{\mathrm{d}t}(0) = 1 \\
\frac{\mathrm{d}^2 y_1(t)}{\mathrm{d}t^2} &= 2y_0(t), & y_1(0) = \frac{\mathrm{d}y_1}{\mathrm{d}t}(0) = 0 \\
\frac{\mathrm{d}^2 y_2(t)}{\mathrm{d}t^2} &= 2y_1(t) - 3y_0(t), & y_2(0) = \frac{\mathrm{d}y_2}{\mathrm{d}t}(0) = 0 \\
&\vdots
\end{aligned},$$

where the first equation has already been solved above. The equation for $y_1(t)$ then reads as:

$$\frac{\mathrm{d}^2 y_1(t)}{\mathrm{d}t^2} = t(2-t),$$

which has the particular solution

$$y_1(t) = \frac{t^3}{12}(4-t).$$

Consequently, the equation for $y_2(t)$ reads as:

$$\frac{\mathrm{d}^2 y_2(t)}{\mathrm{d}t^2} = \frac{t^3}{4}(4-t) - \frac{3t}{2}(2-t),$$

which has the particular solution

$$y_2(t) = -\frac{t^3}{360}(180 - 45t - 12t^2 + 2t^3).$$

Further solutions $y_i(t)$, for $i > 2$, can be computed in a similar way.
With the previous computations we have that the solution of the perturbed problem (7) is given by:

$$y_\varepsilon(t) = \frac{t}{2}(2-t) + \frac{\varepsilon t^3}{12}(4-t) - \frac{\varepsilon^2 t^3}{360}(180 - 45t - 12t^2 + 2t^3) + \cdots \tag{9}$$

Notice that, at least from the few solutions we have computed, the solutions y_i seem to have the trend $y_i \to 0$ as $i \to \infty$. This hints to the possibility of the series (9) to be convergent. See in Figure a comparison of the solutions in this example.

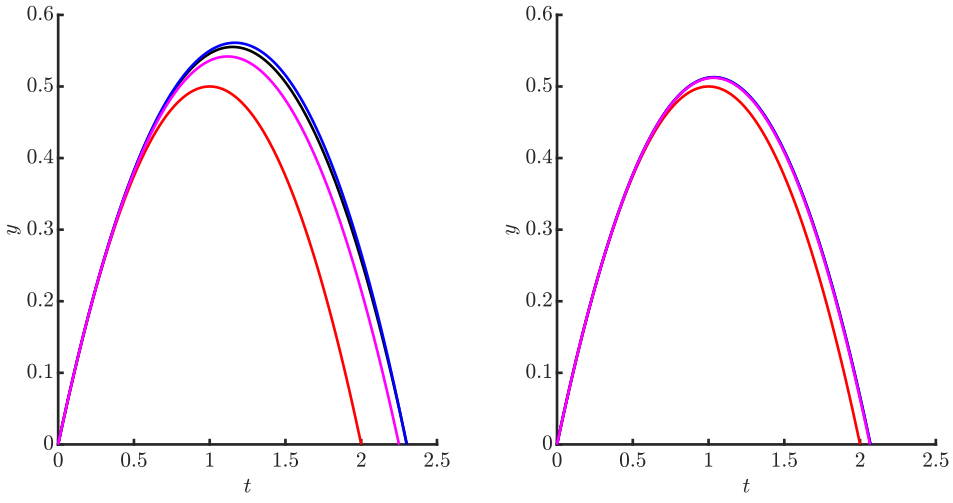


FIGURE 2. Comparison of the approximations in this example. The black, red, blue, and magenta curves correspond to the numerical, y_0, $y_0 + \varepsilon y_1$, and $y_0 + \varepsilon y_1 + \varepsilon^2 y_2$ solutions. On the left we used $\varepsilon = 0.2$ while in the right $\varepsilon = 0.05$.

[a] See also exercise I.3

EXERCISE I.3. *In the example I.3 we used the approximation:*

$$\frac{1}{(1+x)^2} \approx 1 - 2x + 3x^2 + \cdots,$$

for x small. Show that the full expansion is, in fact,

$$\frac{1}{(1+x)^2} = 1 + \sum_{k=1}^{\infty} (-1)^k (k+1) x^k.$$

Is the series convergent for $|x| \ll 1$? Answer: yes, but only for $|x| < 1$.

Hint: *Since $|x| \ll 1$ one can propose the ansatz $\frac{1}{(1+x)^2} \approx 1 + a_1 x + a_2 x^2 + \cdots$ (notice that this series is simply the Taylor series for x near 0, and thus the a_i's are the Taylor coefficients). Compute a few of the coefficients a_i. Once you observe a pattern, try to find a closed formula for each of the a_i coefficients. To check for convergence, you may want to use on of the many convergence tests, such as the "Ratio test".*

REMARK I.1. This exercise shows that series expansions can also be used in other contexts beyond ODEs.

EXAMPLE I.4 (Forced oscillator). Consider a harmonic oscillator subject to a small periodic force, that is

$$\frac{\mathrm{d}^2\theta}{\mathrm{d}t^2} + \omega_0^2 \theta = \varepsilon \sin(\omega t), \tag{10}$$

where $\theta \in [0, 2\pi)$ is the angle of the oscillator, ω_0 its natural frequency, ω is the frequency of the force, and $|\varepsilon| \ll 1$.

The solution of the homogeneous part

$$\frac{\mathrm{d}^2\theta}{\mathrm{d}t^2} + \omega_0^2 \theta = 0$$

is

$$\theta_h(t) = a \cos(\omega_0 t) + b \sin(\omega_0 t),$$

where a and b are constants that depend on the initial conditions. To find a particular solution for the inhomogeneous part, let us propose the particular solution

$$\theta_p(t) = c \sin(\omega t), \tag{11}$$

where c is a constant to be found. Substituting (11) into (10) we find that

$$c = \frac{\varepsilon}{\omega_0^2 - \omega^2},$$

and therefore, the general solution of (10) is given by

$$\theta(t) = a\cos(\omega_0 t) + b\sin(\omega_0 t) + \frac{\varepsilon}{\omega_0^2 - \omega^2}\sin(\omega t), \tag{12}$$

with $a = \theta(0)$ and $b = \dfrac{1}{\omega_0}\left(\dfrac{\mathrm{d}\theta}{\mathrm{d}t}(0) - \dfrac{\varepsilon\omega}{\omega_0^2 - \omega^2}\right)$.
Notice that the solution (12) is defined only whenever $\omega \neq \omega_0$. However, the limit of $\theta(t)$ as $\omega \to \omega_0$ is well-defined. Indeed:

$$\begin{aligned}
\lim_{\omega\to\omega_0}\theta(t) &= \lim_{\omega\to\omega_0}\left[a\cos(\omega_0 t) + b\sin(\omega_0 t) + \frac{\varepsilon}{\omega_0^2-\omega^2}\sin(\omega t)\right] \\
&= \lim_{\omega\to\omega_0}\left[\theta(0)\cos(\omega_0 t) + \frac{1}{\omega_0}\left(\frac{\mathrm{d}\theta}{\mathrm{d}t}(0) - \frac{\varepsilon\omega}{\omega_0^2-\omega^2}\right)\sin(\omega_0 t) + \frac{\varepsilon}{\omega_0^2-\omega^2}\sin(\omega t)\right] \\
&= \theta(0)\cos(\omega_0 t) + \frac{1}{\omega_0}\frac{\mathrm{d}\theta}{\mathrm{d}t}(0)\sin(\omega_0 t) + \varepsilon\lim_{\omega\to\omega_0}\left[\frac{\sin(\omega t)}{\omega_0^2-\omega^2} - \frac{\omega\sin(\omega_0 t)}{\omega_0(\omega_0^2-\omega^2)}\right] \\
&= \theta(0)\cos(\omega_0 t) + \frac{1}{\omega_0}\frac{\mathrm{d}\theta}{\mathrm{d}t}(0)\sin(\omega_0 t) + \varepsilon\lim_{\omega\to\omega_0}\left[\frac{\omega_0\sin(\omega t) - \omega\sin(\omega_0 t)}{\omega_0(\omega_0^2-\omega^2)}\right] \\
&= \theta(0)\cos(\omega_0 t) + \frac{1}{\omega_0}\frac{\mathrm{d}\theta}{\mathrm{d}t}(0)\sin(\omega_0 t) + \varepsilon\lim_{\omega\to\omega_0}\left[\frac{\omega_0 t\cos(\omega t) - \sin(\omega_0 t)}{-2\omega_0\omega}\right] \\
&= \theta(0)\cos(\omega_0 t) + \frac{1}{\omega_0}\frac{\mathrm{d}\theta}{\mathrm{d}t}(0)\sin(\omega_0 t) - \varepsilon\frac{t\cos(\omega_0 t)}{2\omega_0} + \varepsilon\frac{\sin(\omega_0 t)}{2\omega_0^2}.
\end{aligned} \tag{13}$$

Notice that the solution for $\omega \neq \omega_0$, given by (12), is bounded. Moreover, if $\dfrac{\omega}{\omega_0} \in \mathbb{Q}$, then the solution $\theta(t)$ is periodic, and quasiperiodic otherwise. In contrast, the solution for $\omega = \omega_0$ grows proportionally to εt. So, if we let $\delta = \omega - \omega_0$ be another parameter, we notice that: a) the solution for $\varepsilon = 0$ and for $\varepsilon \neq 0$ but small can be largely different. The same is true for the case $\delta = 0$ and δ small. See figure 3.

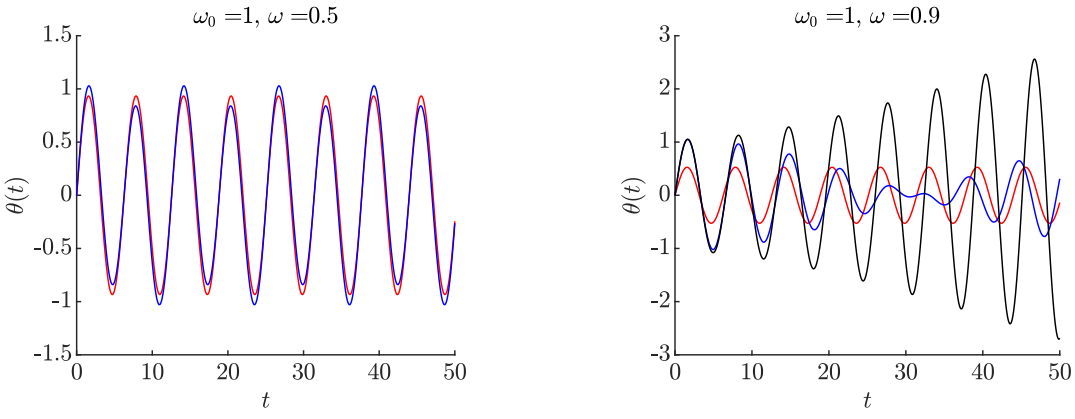


FIGURE 3. Comparison of the analytic solution (12) and its approximation under several assumptions. In both graphs $\varepsilon = 0.1$, the solution (12) is shown in blue, and the solution for $\varepsilon = 0$ in red. On the right we also plot the limit solution (13). Notice that its amplitude grows linearly, this effect is called *resonance*.

EXERCISE I.4. *Implement, on a computer program, the distinct solutions of the forced oscillator of Example I.4 and test several setups for ε and δ comparing your observations.*

EXAMPLE I.5 (Example I.4 continued - action angle variables). We continue with Example I.4, but we now consider an arbitrary oscillator (and not necessarily a harmonic oscillator). For convenience, let us recall that the equation we study is

$$\frac{\mathrm{d}^2\theta}{\mathrm{d}t^2} = f(\theta) + \varepsilon \sin(\omega t). \tag{14}$$

We shall assume that, for $\varepsilon = 0$, the origin $\theta = 0$ is a stable equilibrium point.

LEMMA I.1. *Consider (14) with $\varepsilon = 0$. If $f(0) = 0$ and $\frac{\mathrm{d}f}{\mathrm{d}\theta}(0) < 0$, then $\theta = 0$ is a stable equilibrium point.*

PROOF. When $\varepsilon = 0$, (14) simply reads as

$$\frac{\mathrm{d}^2\theta}{\mathrm{d}t^2} = f(\theta),$$

which can be re-written as a system (with $(x_1, x_2) = (\theta, \frac{\mathrm{d}\theta}{\mathrm{d}t})$)

$$\begin{aligned} \frac{\mathrm{d}x_1}{\mathrm{d}t} &= x_2 \\ \frac{\mathrm{d}x_2}{\mathrm{d}t} &= f(x_1). \end{aligned} \tag{15}$$

The equilibrium for (15) is, indeed, given by $x_2 = 0$ and $f(x_1) = f(\theta) = 0$. Linearization of (15) at the origin gives

$$\begin{bmatrix} \frac{\mathrm{d}x_1}{\mathrm{d}t} \\ \frac{\mathrm{d}x_2}{\mathrm{d}t} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \frac{\mathrm{d}f}{\mathrm{d}x_1}(0) & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Notice that if $\frac{\mathrm{d}f}{\mathrm{d}x_1}(0) > 0$, then the origin is a saddle, while when $\frac{\mathrm{d}f}{\mathrm{d}x_1}(0) < 0$, the origin is a center. In the latter case, the origin is indeed stable (but, of course, not asymptotically stable). □

The most important observation in this example is that (14) has a constant of motion for $\varepsilon = 0$.

LEMMA I.2. *Consider (14) with $\varepsilon = 0$ and let $V(\theta)$ be a (potential) function such that $\frac{\mathrm{d}V(\theta)}{\mathrm{d}\theta} = -f(\theta)$. Then the function $H = \frac{1}{2}\left(\frac{\mathrm{d}\theta}{\mathrm{d}t}\right)^2 + V(\theta)$ is constant along solutions of (14) with $\varepsilon = 0$.*

PROOF. Indeed

$$\begin{aligned}\frac{\mathrm{d}H}{\mathrm{d}t} &= \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{2}\left(\frac{\mathrm{d}\theta}{\mathrm{d}t}\right)^2 + V(\theta)\right)\\ &= \frac{\mathrm{d}\theta}{\mathrm{d}t}\frac{\mathrm{d}^2\theta}{\mathrm{d}t^2} + \frac{\mathrm{d}V}{\mathrm{d}\theta}\frac{\mathrm{d}\theta}{\mathrm{d}t}\\ &= \frac{\mathrm{d}\theta}{\mathrm{d}t}\left(\frac{\mathrm{d}^2\theta}{\mathrm{d}t^2} + \frac{\mathrm{d}V}{\mathrm{d}\theta}\right)\\ &= \frac{\mathrm{d}\theta}{\mathrm{d}t}\left(\frac{\mathrm{d}^2\theta}{\mathrm{d}t^2} - f(\theta)\right)\\ &= 0.\end{aligned}$$

□

Notice that from the assumption $\frac{\mathrm{d}f}{\mathrm{d}\theta}(0) < 0$ and $\frac{\mathrm{d}V(\theta)}{\mathrm{d}\theta} = -f(\theta)$, we can say that $V(\theta) \sim \frac{1}{2}\omega_0^2\theta^2$ (where ω_0 and the $\frac{1}{2}$ factor are chosen just for convenience). The previous observation, together with the fact that $H = H(\theta, \frac{\mathrm{d}\theta}{\mathrm{d}t}) = \frac{1}{2}\left(\frac{\mathrm{d}\theta}{\mathrm{d}t}\right)^2 + V(\theta)$ is constant along the solutions of the unperturbed problem, imply that the orbits in the phase-space $\left(\theta, \frac{\mathrm{d}\theta}{\mathrm{d}t}\right)$ are closed near the origin $(\theta, \frac{\mathrm{d}\theta}{\mathrm{d}t}) = (0, 0)$.

Since (at least locally) the orbits are closed, and the function H is constant along such orbits, we can imagine that we can parameterize the solutions, not only by their $\left(\theta, \frac{\mathrm{d}\theta}{\mathrm{d}t}\right)$-parametrization, but also by (H, ψ), where $\psi \in [0, 2\pi)$. In that way, for each constant value of H, $\frac{\mathrm{d}H}{\mathrm{d}t} = 0$ and the corresponding closed orbit is parametrized by the angle ψ. Moreover, one can even find a parametrization where the angular velocity along each closed curve is constant. In summary, one can find a parametrization $\left(\theta, \frac{\mathrm{d}\theta}{\mathrm{d}t}\right) \mapsto (I, \phi)$, where (14) (for $\varepsilon = 0$) is given by

$$\begin{aligned}\frac{\mathrm{d}I}{\mathrm{d}t} &= 0\\ \frac{\mathrm{d}\phi}{\mathrm{d}t} &= g(I),\end{aligned} \tag{16}$$

where the function g is specified once $f(\theta)$ is fixed. The coordinates (I, ϕ) are called *action-angle* coordinates[a]. In these coordinates $I = 0$ correspond to $\theta = \frac{\mathrm{d}\theta}{\mathrm{d}t} = 0$ and $g(0) = -\omega_0$. Notice indeed that the orbits of (16) are circles, and for I near 0, the angular speed along each circle is close to ω_0, as it occurs for the harmonic oscillator.

Since the action-angle coordinates remind us of polar coordinates, it will be convenient to work with the complex variable $z = I\exp(\imath\phi)$. From (16), and using the chain rule, we find that z satisfies the differential equation

$$\frac{\mathrm{d}z}{\mathrm{d}t} = \imath z g(|z|),$$

where one should recall that, for $\varepsilon = 0$, $|z|$ is constant along solutions near the origin. Therefore, the corresponding solution is

$$z(t) = z(0)\exp(\imath g(|z|)t). \tag{17}$$

In this way, we can re-write (14) as

$$\frac{\mathrm{d}z}{\mathrm{d}t} = \imath z g(|z|) + \varepsilon \sin(wt) F(z, \bar{z}), \tag{18}$$

where the function $F(z, \bar{z})$ is introduced to account for the fact that we have not specified the nonlinear function $f(\theta)$. Naturally, a particular choice of $f(\theta)$ leads to a particular function $F(z, \bar{z})$.

A particularly convenient way to qualitatively understand the dynamics of (18) is to use a Poincaré map. Since the right-hand side of (18) is $\dfrac{2\pi}{\omega}$ periodic, it suffices to look at the iteration of the Poincaré map $P_\varepsilon : z(0) \mapsto z\left(\dfrac{2\pi}{\omega}\right)$. If $\varepsilon = 0$, we know from (17) that

$$P_0(z) = z\exp\left(\frac{2\pi\imath g(|z|)}{\omega}\right).$$

In particular, we notice that

$$\begin{aligned} P_0(0) &= 0 \\ \frac{\mathrm{d}P_0}{\mathrm{d}z}(0) &= \exp\left(\frac{2\pi\imath g(|0|)}{\omega}\right) = \exp\left(\frac{-2\pi\imath\omega_0}{\omega}\right). \end{aligned}$$

This means, from the implicit function theorem, that for $\varepsilon \neq 0$ sufficiently small, and provided that $\dfrac{\omega_0}{\omega} \notin \mathbb{Z}$, the Poincaré map P_ε has a fixed point near 0. Let $z^* = z^*(\varepsilon)$ denote such a fixed point and let $Z = z - z^*$. It follows from our previous arguments that, for $\varepsilon \neq 0$ small and $\dfrac{\omega_0}{\omega} \notin \mathbb{Z}$, the Poincaré map is of the form

$$P_\varepsilon(Z) = Z\exp\left(-\frac{2\pi\imath\omega_0}{\omega}\right) + R(Z, \bar{Z}, \varepsilon),$$

where the function R represents higher order terms. From the form of $P_\varepsilon(Z)$ we conclude that near the origin, and ε small, the Poincaré map P_ε looks like a translation along circles. See some examples in figure 4.

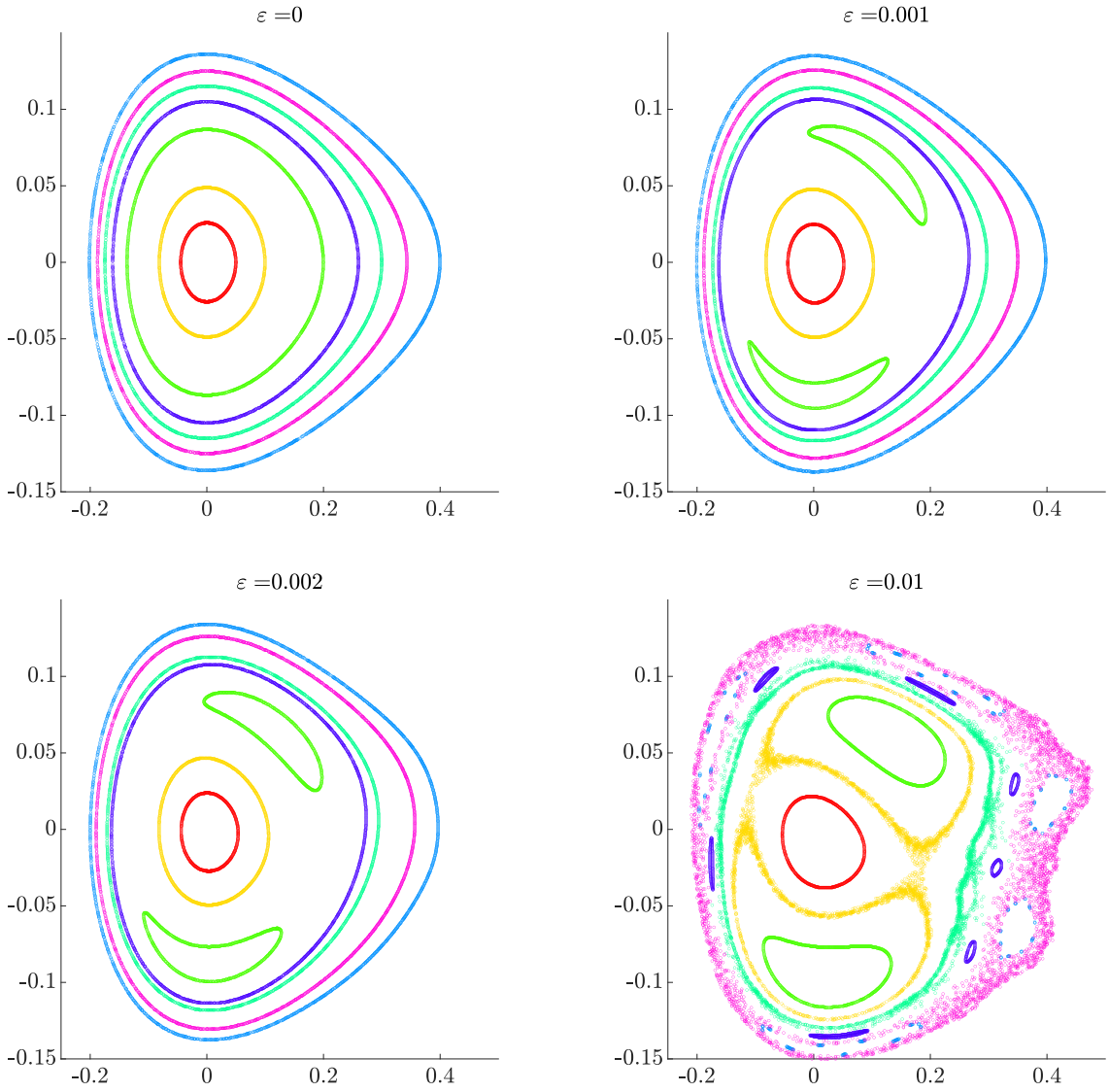


FIGURE 4. A few plots of the Poincaré section of (18) for $f(\theta) = -\omega_0^2\theta + \theta^2 - \theta^3$, $\omega_0 = 0.55$, $\omega = 1$. Each colored orbit corresponds to a single initial condition, and we show about 1000 iterations of the Poincaré map after some time has passed allowing each trajectory to approach an attractor. Notice that for $\varepsilon = 0$ all orbits shown are periodic. When $\varepsilon > 0$, some of such orbits persist, while others get destroyed. Indeed one can observe that for $\varepsilon = 0.01$ a few islands appear surrounded by what appear to be chaotic clouds.

[a] A more formal and general description of action-angle coordinates can be found in the course "Hamiltonian mechanics".

EXERCISE I.5. *Implement on a computer program an algorithm to visualize the Poincaré map of the previous example. Test with different choices of ω_0, ω, and ε. In particular, compare the time series of orbits corresponding to single close rings, pockets, and clouds.*

I.1. Further exercises for this chapter

(1) Let $f(x)$ and $g(x)$ be smooth functions. Let $h(x) = \dfrac{f(x)}{1+\varepsilon g(x)}$. Write an approximation of the form $h(x) \approx f(x) + \varepsilon h_1(x) + \varepsilon^2 h_2(x) + \cdots$ and explicitly obtain h_1 and h_2.

(2) Using Taylor's formula find the first three terms of the approximation of $f(\varepsilon) = \sin(\exp(\varepsilon))$ for ε near 0. Graph $f(\varepsilon)$ and its approximation, and compare them.

(3) Using Taylor's formula find the first three terms of the approximation of $f(\varepsilon) = \dfrac{\exp(\varepsilon)}{1-\varepsilon}$ for ε near 0. Graph $f(\varepsilon)$ and its approximation, and compare them.

(4) Using Taylor's formula approximate the integral $\displaystyle\int \exp(x^k)\mathrm{d}x$, for any $k \in \mathbb{N}$, $k > 1$.

(5) Consider the scalar equation $\dfrac{\mathrm{d}x}{\mathrm{d}t} = -x + \varepsilon \sin(t)$ with initial condition $x(0) = 1$. Propose an approximation of the form $x_\varepsilon(t) = x_0(t) + \varepsilon x_1(t)$ where $x_0(t)$ is the solution of the unperturbed problem $\dfrac{\mathrm{d}x}{\mathrm{d}t} = -x$. Find the term $x_1(t)$.

CHAPTER II

Some basic notions and definitions

II.1. Order symbols

In this section we define the Landau symbols. These will help us formalize and compare the behavior of functions of the perturbation parameter ε as $\varepsilon \to 0$.

Definition II.1 (Big-Oh). Let $f(\varepsilon)$ and $g(\varepsilon)$ be (real valued) functions and $0 \leq \varepsilon_0 \ll 1$ be a small constant. We say that "f is big Oh of g as $\varepsilon \to \varepsilon_0$", and write $f = \mathcal{O}(g)$ as $\varepsilon \to \varepsilon_0$, if there are positive constants k and ε_1 such that

$$|f(\varepsilon)| \leq k|g(\varepsilon)|, \qquad \text{for } \varepsilon_0 < \varepsilon < \varepsilon_1$$

The previous definition means that if $f(\varepsilon) = \mathcal{O}(g(\varepsilon))$, then the absolute value of $f(\varepsilon)$ is bounded, up to a constant, by the absolute value of $g(\varepsilon)$. Equivalently we can say that if

$$\lim_{\varepsilon \to \varepsilon_0} \frac{|f(\varepsilon)|}{|g(\varepsilon)|} < \infty,$$

then $f = \mathcal{O}(g)$ as $\varepsilon \to \varepsilon_0$.

Definition II.2 (Small-Oh). Let $f(\varepsilon)$ and $g(\varepsilon)$ be (real valued) functions and $0 \leq \varepsilon_0 \ll 1$ be a small constant. We say that "f is small Oh of g as $\varepsilon \to \varepsilon_0$", and write $f = o(g)$ as $\varepsilon \to \varepsilon_0$, if for every constant $\delta > 0$, there is a constant ε_2 such that

$$|f(\varepsilon)| \leq \delta|g(\varepsilon)|, \qquad \text{for } \varepsilon_0 < \varepsilon < \varepsilon_2$$

The previous definition means that if $f(\varepsilon) = o(g(\varepsilon))$, then $g(\varepsilon)$ dominates $f(\varepsilon)$ as $\varepsilon \to \varepsilon_0$. Equivalently we can say that if

$$\lim_{\varepsilon \to \varepsilon_0} \frac{|f(\varepsilon)|}{|g(\varepsilon)|} = 0,$$

then $f = o(g)$ as $\varepsilon \to \varepsilon_0$.

Example II.1. All the orders here are as $\varepsilon \to 0$.

(1) Consider $f(\varepsilon) = \varepsilon^2$ and $g_1(\varepsilon) = \varepsilon$, $g_2(\varepsilon) = -a\varepsilon^2 + \varepsilon^3$, $a > 0$. Then

$$\lim_{\varepsilon \to 0} \frac{f}{g_1} = \lim_{\varepsilon \to 0} \varepsilon = 0 \qquad \Longrightarrow \qquad f = o(g)$$

$$\lim_{\varepsilon \to 0} \frac{f}{g_2} = \lim_{\varepsilon \to 0} \frac{\varepsilon^2}{|-a\varepsilon^2 + \varepsilon^3|} = \frac{1}{a} \qquad \Longrightarrow \qquad f = \mathcal{O}(g)$$

(2) Consider $f(\varepsilon) = \varepsilon \sin(1 + \varepsilon^{-1})$ and $g(\varepsilon) = \varepsilon$. In this case we cannot use the limit criterion. But it is clear that $\dfrac{|f|}{|g|} \leq 1$ for $0 < \varepsilon$. Hence $f = \mathcal{O}(g)$.

(3) Consider $f(\varepsilon) = \sin(\varepsilon)$ and $g(\varepsilon) = \varepsilon$. Then

$$\lim_{\varepsilon\to 0}\frac{f}{g} = \lim_{\varepsilon\to 0}\frac{\displaystyle\sum_{k=0}^{\infty}(-1)^k\frac{\varepsilon^{2k+1}}{(2k+1)!}}{\varepsilon} = 1 \qquad \Longrightarrow\ f = \mathcal{O}(g).$$

EXERCISE II.1. *Let* $f = \exp\left(-\frac{1}{\varepsilon}\right)$. *Prove that for all* $k \in \mathbb{N}$, $f = o(\varepsilon^k)$. *In this case we say that* f *is "transcendentally small with respect to powers of* ε*".*

EXERCISE II.2. *Prove the following statements :*

(1) $f = \mathcal{O}(1)$ *as* $\varepsilon \to \varepsilon_0 \iff f$ *is bounded as* $\varepsilon \to \varepsilon_0$.

(2) $f = o(1)$ *as* $\varepsilon \to \varepsilon_0 \iff f \to 0$ *as* $\varepsilon \to \varepsilon_0$.

(3) $f = o(g)$ *as* $\varepsilon \to \varepsilon_0 \implies f = \mathcal{O}(g)$ *as* $\varepsilon \to \varepsilon_0$. *Is the converse true?*

(4) $f = \mathcal{O}(\varepsilon^\alpha)$ *as* $\varepsilon \to 0 \implies f = o(\varepsilon^\beta)$ *as* $\varepsilon \to 0$ *for any* $\beta < \alpha$.

(5) $o(\mathcal{O}(h)) = \mathcal{O}(o(h)) = o(h)$ *as* $\varepsilon \to 0$. *(Here* $h = h(\varepsilon)$*)*

EXERCISE II.3. *Let* $f_1 = \mathcal{O}(g_1)$ *and* $f_2 = \mathcal{O}(g_2)$ *as* $\varepsilon \to 0$. *Show that*

(1) $f_1 + f_2 = \mathcal{O}(\max\{g_1, g_2\})$

(2) $f_1 f_2 = \mathcal{O}(g_1 g_2)$

II.2. Asymptotic approximations

In this section we formalize what we mean by an asymptotic approximation.

DEFINITION II.3. Given $f(\varepsilon)$ and $g(\varepsilon)$, we say that $g(\varepsilon)$ is an *asymptotic approximation* of $f(\varepsilon)$ as $\varepsilon \to \varepsilon_0$ whenever $f = g + o(g)$ as $\varepsilon \to \varepsilon_0$ and we write $f \sim g$ as $\varepsilon \to \varepsilon_0$.

The above definition means that g is an asymptotic approximation of f if the error $f - g$ is of higher order than the approximation itself. Let us now see a few examples.

EXAMPLE II.2. Let $f(\varepsilon) = \sin(\varepsilon)$ and $\varepsilon = 0$. We know that $f(\varepsilon) = \varepsilon - \frac{1}{6}\varepsilon^3 + \mathcal{O}(\varepsilon^5)$. Then:

(1) $f(\varepsilon) - \varepsilon = -\frac{1}{6}\varepsilon^3 + \mathcal{O}(\varepsilon^5) = o(\varepsilon) \implies f \sim \varepsilon$

(2) $f(\varepsilon) - \left(\varepsilon - \frac{1}{6}\varepsilon^3\right) = \frac{1}{120}\varepsilon^5 + \mathcal{O}(\varepsilon^7) = o\left(\varepsilon - \frac{1}{6}\varepsilon^3\right) \implies f \sim \varepsilon - \frac{1}{6}\varepsilon^3$

EXERCISE II.4. *Related to the previous example where* $f(\varepsilon) = \sin(\varepsilon)$. *Is it true that* $f \sim \varepsilon + 2\varepsilon^2$*?* *Answer: yes.*

EXAMPLE II.3. Let $f(x,\varepsilon) = x + \exp\left(-\frac{x}{\varepsilon}\right)$ with $x \in (0,1)$ and small ε. First, we wonder if $f \sim x$ as $\varepsilon \to 0$. Indeed

$$f - x = \exp\left(-\frac{x}{\varepsilon}\right), \text{ and } \lim_{\varepsilon\to 0} \frac{\exp\left(-\frac{x}{\varepsilon}\right)}{x} = 0,$$

for any fixed value of $x \in (0,1)$. So indeed $f \sim x$. Notice, however, that $f(0) = 1$, which means that the approximation is worse the closer x is to 0, no matter the value of ε. In other words, if we want the approximation to be "good" the closer x is to 0, the smaller ε must be, see figure 1.

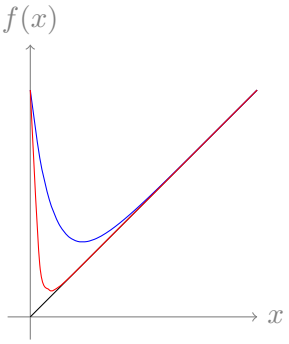


FIGURE 1. Comparison between the approximation $f(x) \sim x$ (in black) and $f(x)$ for $\varepsilon = 0.1$ (blue) and $\varepsilon = 0.025$ (red).

EXERCISE II.5. *Consider the function* $f(x) = \sin(\pi x) + \varepsilon^3$ *for* $0 \le x \le \frac{1}{2}$. *Is it true that* $f \sim \sin(\pi x)$ *as* $\varepsilon \to 0$*? Justify your answer sufficiently.*

II.3. Asymptotic expansions

In the previous section we have noted a few facts: a) asymptotic approximations are not unique, and b) we do not get that much information about the accuracy of the approximation. In this section we address precisely such issues.

DEFINITION II.4.

(1) The functions $\phi_1(\varepsilon)$, $\phi_2(\varepsilon)$, ... form *an asymptotic sequence*[1] as $\varepsilon \to \varepsilon_0$ if and only if $\phi_{j+1} = o(\phi_j)$ as $\varepsilon \to \varepsilon_0$ for all $j = 1, 2, \ldots$.

(2) If $\phi_1(\varepsilon)$, $\phi_2(\varepsilon)$, ... is an asymptotic sequence, then $f(\varepsilon)$ has *an asymptotic expansion* up to order ϕ_n (w.r.t. the given sequence) if and only if

$$f = \sum_{j=1}^{n} a_j \phi_j + o(\phi_n),$$

as $\varepsilon \to \varepsilon_0$ and where the coefficients a_j are independent of ε.

REMARK II.1.

- The functions forming the asymptotic sequence in the previous definition are usually called *gauge functions, basis functions, or scale*, depending on the framework.

[1] We may also say that the functions are well-ordered

- It is also common to write an asymptotic expansion as
$$f = \sum_{j=1}^{n} a_j \phi_j + \mathcal{O}(\phi_{n+1}),$$
to highlight the concept "the next term in the expansion is smaller than the previous one".
- If f has an asymptotic expansion for all $n \in \mathbb{N}$ one usually writes
$$f \sim \sum_{j=1}^{\infty} a_j \phi_j$$

EXAMPLE II.4. Examples of gauge functions are:

(1) $\phi_1 = \varepsilon^{r_1}$, $\phi_2 = \varepsilon^{r_2}$, $\phi_3 = \varepsilon^{r_3}$, ..., with $r_1 < r_2 < r_3 < \cdots$.
(2) $\phi_1 = 1$, $\phi_2 = \exp\left(-\frac{1}{\varepsilon}\right)$, $\phi_3 = \exp\left(-\frac{2}{\varepsilon}\right)$, ...
(3) $\phi_k = \varepsilon |\ln \varepsilon|^{-k}$, $k \in \mathbb{N}$

EXERCISE II.6. *Verify that the functions given in the previous example indeed form an asymptotic sequence. You may need to re-arrange some.*

Now a question that arises is how to find an asymptotic expansion for a given function $f(\varepsilon)$? A fairly common first approach is to use Taylor's expansion.

EXAMPLE II.5.

(1) Let $f(\varepsilon) = \exp(\varepsilon)$. Then, the Taylor expansion for ε near 0 is
$$\exp(\varepsilon) = 1 + \varepsilon + \frac{\varepsilon^2}{2} + \cdots.$$
Therefore $\exp(\varepsilon) \sim 1 + \varepsilon + \mathcal{O}(\varepsilon^2)$.

(2) Let $f(\varepsilon) = \dfrac{\cos \varepsilon}{\varepsilon}$. For this example we cannot immediately use Taylor's expansion because the function $f(\varepsilon)$ is not defined at $\varepsilon = 0$. However, what we can do is expand simply the regular term $\cos \varepsilon$ to obtain:
$$f(\varepsilon) \sim \frac{1}{\varepsilon}\left(1 - \frac{\varepsilon^2}{2} + \cdots\right).$$
Indeed the expansion is not defined at $\varepsilon = 0$ either. As an exercise, you can corroborate the validity of this expansion numerically.

(3) Let $f(\varepsilon) = \dfrac{\sqrt{1+\varepsilon}}{\sin(\sqrt{\varepsilon})}$. Then we note that, Taylor expanding each term:
$$\sqrt{1+\varepsilon} \sim 1 + \frac{\varepsilon}{2} + \cdots$$
$$\sin(\sqrt{\varepsilon}) \sim \varepsilon^{1/2} - \frac{\varepsilon^{3/2}}{6} + \cdots.$$

Then, we can write:

$$f \sim \frac{1+\frac{\varepsilon}{2}+\cdots}{\varepsilon^{1/2}-\frac{\varepsilon^{3/2}}{6}+\ldots} = \frac{1}{\varepsilon^{1/2}}\frac{1+\frac{\varepsilon}{2}+\cdots}{1-\frac{\varepsilon}{6}+\cdots} \sim \frac{1}{\varepsilon^{1/2}}\left(1+\frac{\varepsilon}{2}+\cdots\right)\left(1+\frac{\varepsilon}{6}+\cdots\right)$$
$$\sim \frac{1}{\varepsilon^{1/2}}\left(1+\frac{2\varepsilon}{3}+\mathcal{O}(\varepsilon^2)\right).$$

In the previous examples, the gauge functions appeared naturally from the method used, Taylor's series. However, one may attempt to use a specified asymptotic sequence for the expansions. The overall procedure goes as follows: suppose that the gauge functions $\phi_1, \phi_2, \ldots$ are given, and that an expansion of the form $f \sim a_1\phi_1 + a_2\phi_2$ is sought. Since, by assumption, $f = a_1\phi_1 + o(\phi_1)$, and assuming that we can divide by ϕ_1, we have that $a_1 = \lim_{\varepsilon\to\varepsilon_0} \frac{f}{\phi_1}$. This procedure can be repeated again at each scale obtaining:

$$\begin{aligned} a_1 &= \lim_{\varepsilon\to\varepsilon_0} \frac{f}{\phi_1} \\ a_2 &= \lim_{\varepsilon\to\varepsilon_0} \frac{f-a_1\phi_1}{\phi_2} \\ a_3 &= \lim_{\varepsilon\to\varepsilon_0} \frac{f-a_1\phi_1-a_2\phi_2}{\phi_3} \\ &\vdots \end{aligned}$$

Naturally here we are assuming that the gauge functions are nonzero at $\varepsilon = \varepsilon_0$ and that the limits exist.

EXAMPLE II.6. Consider $f(\varepsilon) = \frac{1}{1+\varepsilon} + \exp\left(-\frac{1}{\varepsilon}\right)$. Let $\phi_k = \varepsilon^{k-1}$, $k = 1, 2, \ldots$. From the above formulas we have:

$$\begin{aligned} a_1 &= \lim_{\varepsilon\to 0} \frac{f}{1} = 1 \\ a_2 &= \lim_{\varepsilon\to 0} \frac{f-1}{\varepsilon} = \lim_{\varepsilon\to 0}\left(\frac{-1}{1+\varepsilon} + \frac{\exp\left(-\frac{1}{\varepsilon}\right)}{\varepsilon}\right) = -1 \\ &\vdots \end{aligned}$$

Thus, we can conclude that $f \sim 1 - \varepsilon + O(\varepsilon^2)$. Notice that the exponential term has no contribution in the expansion!

EXERCISE II.7. *Give a few examples of different functions with the same asymptotic expansion.*

We finish this section with a brief digression on a couple of operations that will frequently appear during the course, differentiation and integration.

Given a function $f(x,\varepsilon) \sim a_1(x)\phi_1(\varepsilon) + a_2(x)\phi_2(\varepsilon) + \cdots$ as $\varepsilon \to 0$, we shall generally *assume* that there exist functions $b_1(x)$, $b_2(x)$, ... such that

$$\frac{\mathrm{d}}{\mathrm{d}x} f(x,\varepsilon) \sim b_1(x)\phi_1(\varepsilon) + b_2(x)\phi_2(\varepsilon) + \cdots$$

as $\varepsilon \to 0$, where $b_k = \dfrac{\mathrm{d}a_k}{\mathrm{d}x}$.

EXERCISE II.8. *Notice that the above assumption is not necessarily true and may depend on the choice of gauge functions. Take for example* $f = \exp\left(-\frac{x}{\varepsilon}\right)\sin\left(\frac{x}{\varepsilon}\right)$. *Compute an expansion in terms of powers of* ε. *Next compute the derivative of* f *and check whether* $\frac{\mathrm{d}f}{\mathrm{d}x}$ *has an expansion in terms of powers of* ε.

On the other hand, regarding integration, given that $f(x,\varepsilon) \sim a_1(x)\phi_1(\varepsilon) + a_2(x)\phi_2(\varepsilon) + \cdots$ as $\varepsilon \to \varepsilon_0$, and assuming that the functions $a_k(x)$ are integrable in the interval $x \in [a,b]$ for all k, then it holds that

$$\int_a^b f(x,\varepsilon)\mathrm{d}x \sim \left(\int_a^b a_1(x)\mathrm{d}x\right)\phi_1(\varepsilon) + \left(\int_a^b a_2(x)\mathrm{d}x\right)\phi_2(\varepsilon) + \cdots,$$

as $\varepsilon \to \varepsilon_0$.

EXAMPLE II.7.

(1) Let $f(\varepsilon) = \displaystyle\int_0^1 \exp(\varepsilon x^2)\mathrm{d}x$. Since $\exp(\varepsilon x^2) \sim 1 + \varepsilon x^2 + \cdots$ for $x \in [0,1]$, then we have that $f(\varepsilon) \sim \displaystyle\int_0^1 (1 + \varepsilon x^2 + \cdots)\mathrm{d}x = 1 + \frac{\varepsilon}{3} + \cdots$.

(2) Let $f(\varepsilon) = \displaystyle\int_0^1 \frac{\mathrm{d}x}{\varepsilon^2 + x^2}$. Notice that the integrand $\dfrac{1}{\varepsilon^2+x^2} \sim \dfrac{1}{x^2} - \dfrac{\varepsilon^2}{x^4} + \cdots$. The coefficients $a_k(x)$ are not integrable in the interval $x \in [0,1]$ (there is a singularity at $x = 0$). Of course in this case that is not really an issue because we can simply compute the integral, which results in $f(\varepsilon) = \dfrac{1}{\varepsilon}\arctan\left(\dfrac{1}{\varepsilon}\right)$. As an exercise, expand $f(\varepsilon) = \dfrac{1}{\varepsilon}\arctan\left(\dfrac{1}{\varepsilon}\right)$.

(3) Consider the function $f(\varepsilon) = \displaystyle\int_0^{\pi/3} \frac{\mathrm{d}x}{\varepsilon^2 + \sin x}$. The integrand has the expansion (using Taylor for $\varepsilon \sim 0$):

$$\frac{1}{\varepsilon^2 + \sin x} \sim \frac{1}{\sin x}\left(1 - \frac{\varepsilon^2}{\sin x} + \cdots\right)$$

As before, the coefficients are not integrable in an interval containing 0, and now it is considerably more difficult to explicitly integrate the function. Notice that for the gauge functions to be well-ordered (to form an asymptotic sequence) we require $\dfrac{\varepsilon^2}{\sin x} \ll 1$, which implies $\varepsilon^2 \ll x$ for x near $x = 0$ (the singularity). Thus, let $\delta \in (0,1)$ be such that $\varepsilon^2 \ll \delta \ll 1$ and let us split the integral as

$$f(\varepsilon) = \int_0^{\delta} \frac{\mathrm{d}x}{\varepsilon^2 + \sin x} + \int_{\delta}^{\pi/3} \frac{\mathrm{d}x}{\varepsilon^2 + \sin x}.$$

The second integral is now well-defined:

$$\begin{aligned}\int_\delta^{\pi/3} \frac{\mathrm{d}x}{\varepsilon^2+\sin x} &\sim \int_\delta^{\pi/3} \frac{1}{\sin x}\left(1-\frac{\varepsilon^2}{\sin x}+\cdots\right)\mathrm{d}x \\ &= \ln\left(\frac{1}{\sqrt{3}}\right) - \underbrace{\ln\left(\tan\left(\frac{\delta}{2}\right)\right)}_{\sim \ln\delta - \ln 2 + \frac{\delta^2}{12}+\cdots} + \varepsilon^2\left(\frac{1}{\sqrt{3}} - \underbrace{\cot(\delta)}_{\sim\frac{1}{\delta}+\cdots}\right)+\cdots \\ &\sim \ln\left(\frac{2}{\sqrt{3}}\right)+\frac{\varepsilon^2}{\sqrt{3}}-\ln\delta-\frac{\varepsilon^2}{\delta}-\frac{\delta^2}{12}+\cdots\end{aligned}$$

For the first integral, let y be defined by $x=\varepsilon^2 y$. Then

$$\begin{aligned}\int_0^{\delta} \frac{\mathrm{d}x}{\varepsilon^2+\sin x} &= \varepsilon^2\int_0^{\delta/\varepsilon^2}\frac{\mathrm{d}y}{\varepsilon^2+\sin(\varepsilon^2 y)} \\ &\sim \varepsilon^2\int_0^{\delta/\varepsilon^2}\frac{\mathrm{d}y}{\varepsilon^2+\varepsilon^2 y-\frac{\varepsilon^6 y^3}{3!}+\cdots} \\ &= \varepsilon^2\int_0^{\delta/\varepsilon^2}\frac{\mathrm{d}y}{\varepsilon^2(1+y-\frac{\varepsilon^4 y^3}{3!}+\cdots)} \\ &\sim \int_0^{\delta/\varepsilon^2}\left(\frac{1}{1+y}+\frac{\varepsilon^4 y^3}{3!(1+y)^2}+\cdots\right)\mathrm{d}y \qquad \leftarrow \frac{1}{1+y-ay^3}=\sum_{n=0}^{\infty}\frac{a^n\left(\frac{y^3}{1+y}\right)^n}{1+y} \text{ for } |a|\ll 1 \\ &= \ln\left(1+\frac{\delta}{\varepsilon^2}\right)+\frac{\varepsilon^4}{6}\left(\frac{1}{2}\frac{\delta^2}{\varepsilon^4}-2\frac{\delta}{\varepsilon^2}+3\ln\left(1+\frac{\delta}{\varepsilon^2}\right)+\frac{\varepsilon^2}{\delta+\varepsilon^2}-1\right)+\cdots \\ &\sim \ln\left(\frac{\delta}{\varepsilon^2}\right)+\frac{\varepsilon^2}{\delta}+\cdots+\frac{\delta^2}{12}+\cdots \\ &= \ln(\delta)-2\ln(\varepsilon)+\frac{\varepsilon^2}{\delta}+\frac{\delta^2}{12}+\cdots.\end{aligned}$$

Adding the two integrals we then get

$$f(\varepsilon)\sim \ln\left(\frac{2}{\sqrt{3}}\right)-2\ln(\varepsilon)+\mathcal{O}(\varepsilon^2).$$

II.4. Further exercises for this chapter

(1) Find the values of α (if any) such that for the following functions it holds that $f = \mathcal{O}(\varepsilon^\alpha)$ as $\varepsilon \to 0$. Repeat for $f = o(\varepsilon^\alpha)$ as $\varepsilon \to 0$.
   (a) $f = (1+\varepsilon^2)^{1/2}$
   (b) $f = \varepsilon \sin(\varepsilon)$
   (c) $f = \varepsilon \ln(\varepsilon)$
   (d) $f = \sqrt{x+\varepsilon}$ with $x \in [0,1]$

(2) Suppose $f = \mathcal{O}(g)$. Is it true that $\exp(f) = \mathcal{O}(\exp(g))$?

(3) Suppose that $f(\varepsilon) = o(g(\varepsilon))$ for small ε and where f and g are continuous. Which of the following is true?
   (a) $\displaystyle\int_0^\varepsilon f \mathrm{d}\varepsilon = o\left(\int_0^\varepsilon g \mathrm{d}\varepsilon\right)$
   (b) $\displaystyle\int_0^\varepsilon f \mathrm{d}\varepsilon = o\left(\int_0^\varepsilon |g| \mathrm{d}\varepsilon\right)$

(4) Assume that $f \sim \displaystyle\sum_{k=1}^{\infty} a_k \varepsilon^{\alpha_k}$. Find (at least up to $k=2$) the appropriate powers α_k's with $\alpha_k < \alpha_{k+1}$ and nonzero constants a_k for the following functions:
   (a) $f = \dfrac{1}{1-\exp(\varepsilon)}$
   (b) $f = 1 + \varepsilon - 2\ln(1+\varepsilon) - \dfrac{1}{1+\varepsilon}$
   (c) $f = \displaystyle\int_0^\varepsilon \sin(x + \varepsilon x^2) \mathrm{d}x$

(5) Find the first two terms in the expansion of the function $f(\varepsilon) = \displaystyle\int_0^{\pi/4} \frac{\mathrm{d}x}{\varepsilon^2 + \sin^2 x}$.

(6) Find the first three terms of the asymptotic approximation of the Stieltjes function $S(\varepsilon) = \displaystyle\int_0^\infty \frac{\exp(-t)}{1+\varepsilon t} \mathrm{d}t$.

Hints: First, find the expansion of the integrand for small ε. From here conclude that it is necessary to let $t \ll \dfrac{1}{\varepsilon}$. Next, split the integral similar to the example in this section. Now you can argue that the second integral is bounded, while for the first integral you can use the expansion you just did to approximate it. You should arrive to $S \sim 1 - \varepsilon + 2\varepsilon^2 + \cdots$ (what is the fourth term of the expansion?)

(7) Let $f(\varepsilon)$ and $g(\varepsilon)$ be positive and assume that $f \sim g$ as $\varepsilon \to 0$. Show that $f^\alpha \sim g^\alpha$ for $\alpha > 0$. Is it true that $\exp(f) \sim \exp(g)$? Why?

(8) All functions in this exercise are continuous and nonzero in a full neighborhood of $\varepsilon = 0$.
   (a) Show that if $f \sim g$ as $\varepsilon \to 0$ then $g \sim f$ as $\varepsilon \to 0$.
   (b) If $f \sim g$ and $h \sim k$, both as $\varepsilon \to 0$, is it true that $f + h \sim g + k$ as $\varepsilon \to 0$?

(9) Find a two term approximation for the roots of $x^2 + x - \varepsilon = 0$.

CHAPTER III

Perturbation Methods

III.1. Matched Asymptotic Expansions

In this section we shall learn about a widely used perturbation method called *Matched asymptotic expansions*. The best way to get to know the method is via an example:

EXAMPLE III.1. Let us consider the second order ODE

$$\varepsilon \frac{\mathrm{d}^2 y}{\mathrm{d}t^2} + 2\frac{\mathrm{d}y}{\mathrm{d}t} + 2y = 0, \tag{19}$$

with initial and boundary conditions $y(0) = 0$ and $y(1) = 1$, and $0 < \varepsilon \ll 1$. The most important observation at this moment is that when we take the limit $\varepsilon \to 0$, then the equation is not anymore of second order, but of first. This carries several problems: one of them being the fact that the initial and the boundary conditions cannot be satisfied. Indeed, if we set $\varepsilon = 0$:

$$2\frac{\mathrm{d}y}{\mathrm{d}t} + 2y = 0, \tag{20}$$

which has solution $y(t) = y(0)\exp(-t)$. If $y(0) = 0$, then $y(t) = 0$ for all $t > 0$.
Problems of this sort are usually called "singular perturbation problem". However, these terms are not unified, and we will deal with a particular class of singular perturbation problems in chapter VII.
To find approximations of the solution of the ODE, we will proceed in several steps. These steps are applied analogously to a wide variety of problems.

Step 1. Outer solution: To begin, as we have already seen before, we will assume that the solution can be expanded in terms of ε, that is:

$$y_\varepsilon(t) = y_0(t) + \varepsilon y_1(t) + \varepsilon^2 y_2(t) + \cdots . \tag{21}$$

Substituting (21) in (19) we get:

$$\underbrace{\varepsilon\left(\frac{\mathrm{d}^2 y_0}{\mathrm{d}t^2} + \varepsilon\frac{\mathrm{d}^2 y_1}{\mathrm{d}t^2} + \cdots\right)}_{①} + \underbrace{2\left(\frac{\mathrm{d}y_0}{\mathrm{d}t} + \varepsilon\frac{\mathrm{d}y_1}{\mathrm{d}t} + \cdots\right)}_{②} + \underbrace{(y_0 + \varepsilon y_1 + \cdots)}_{③} = 0. \tag{22}$$

Notice that ② and ③ have leading order terms of order $\mathcal{O}(1)$ (those that do not multiply ε). In this case we say that ② and ③ are *balanced*. The terms of order $\mathcal{O}(1)$ give (20). The corresponding general solution is

$$y_0(t) = c\exp(-t), \tag{23}$$

where c is an arbitrary constant. As noted above, with this equation we can at most satisfy one of the given conditions, either $y(0) = 0$ or $y(1) = 1$.

The intuition now is that, probably (but we will see that this is indeed true), the solution y_0 is valid for a large interval in $t \in [0,1]$, but that either near $t=0$, or near $t=1$, the solution must be approximated in some other way. Let us assume, that we are missing the approximation near $t=0$. In the next step we will find a better approximation of a thin layer near $t=0$. To distinguish the solutions on different regions of t we will name them. We shall call (21) "the outer solution" (and (23) the first term of the outer solution).

Step 2. Boundary layer: Based on the previous assumption, that there is a *boundary layer* at $t=0$ where we want to obtain a more appropriate approximation, we introduce the so-called "boundary-layer coordinate"

$$\tilde{t} = \frac{t}{\varepsilon^\alpha},$$

where $\alpha > 0$. Let $Y(\tilde{t})$ denote the solution under the new time-coordinate. Using the boundary-layer coordinate, (19) transforms to

$$\varepsilon^{1-2\alpha}\frac{\mathrm{d}^2 Y}{\mathrm{d}\tilde{t}} + 2\varepsilon^{-\alpha}\frac{\mathrm{d}Y}{\mathrm{d}\tilde{t}} + 2Y = 0. \tag{24}$$

If we want to now satisfy the initial condition $y(0)=0$, then we let $Y(0)=0$. Next we assume that the solution of (24) has an expansion of the form

$$Y_\varepsilon(\tilde{t}) = \sum_{k=0}^{\infty} \varepsilon^{\beta_k} Y_k(\tilde{t}), \tag{25}$$

with $0 = \beta_0 < \beta_1 < \beta_2 < \cdots$. Substituting (25) in (24) we get:

$$\underbrace{\varepsilon^{1-2\alpha}\left(\frac{\mathrm{d}^2 Y_0}{\mathrm{d}\tilde{t}^2} + \cdots\right)}_{①} + \underbrace{2\varepsilon^{-\alpha}\left(\frac{\mathrm{d}Y_0}{\mathrm{d}\tilde{t}} + \cdots\right)}_{②} + \underbrace{2(Y_0 + \cdots)}_{③} = 0. \tag{26}$$

Our job now is to find a balancing relation between the terms in (26). Notice, first of all, that the balance between ② and ③ has already been considered in the outer solution, so we must not consider it again. We are left with the following two options:

(1) If ① would be balanced with ③, making ② the higher order term, then we would need to impose $1-2\alpha = 0$, implying $\alpha = \frac{1}{2}$. This would make ① and ③ of order $\mathcal{O}(1)$ and ② of order $\mathcal{O}(\varepsilon^{-1/2})$. But notice that this is a contradiction because $\mathcal{O}(1)$ is in fact a higher order than $\mathcal{O}(\varepsilon^{-1/2})$ as $\varepsilon \to 0$[a]. We conclude that this case is not possible.

(2) If ① would be balanced with ②, making ③ the higher order term, then we would need to impose $1-2\alpha = -\alpha$, implying that $\alpha = 1$. In this case ① and ② are of order $\mathcal{O}(\varepsilon^{-1})$ and ③ of order $\mathcal{O}(1)$. Thus, indeed ③ is of higher order as $\varepsilon \to 0$.

Setting $\alpha = 1$, the leading order terms (those with order $\mathcal{O}(\varepsilon^{-1})$) of (26) correspond to:

$$\frac{\mathrm{d}^2 Y_0}{\mathrm{d}\tilde{t}^2} + 2\frac{\mathrm{d}Y_0}{\mathrm{d}\tilde{t}} = 0. \tag{27}$$

The general solution of (27) is

$$Y_0(\tilde{t}) = A(1 - \exp(-2\tilde{t})).$$

We call Y_0 the (first or leading term of the) inner solution.

Let us now return to the initial and boundary conditions. Recall that we are assuming that the outer solution (23) is valid away from $t = 0$ and thus we impose it to satisfy $y(1) = 1$. Then the particular solution is $y_0(t) = \exp(1 - t)$. On the other hand, we assume that the inner solution satisfies $Y_0(0) = 0$. However, notice that this condition is automatically satisfied. This, and the fact that the boundary-layer solution contains a term of the outer solution will be used in the matching process of the next step. See also figure 1.

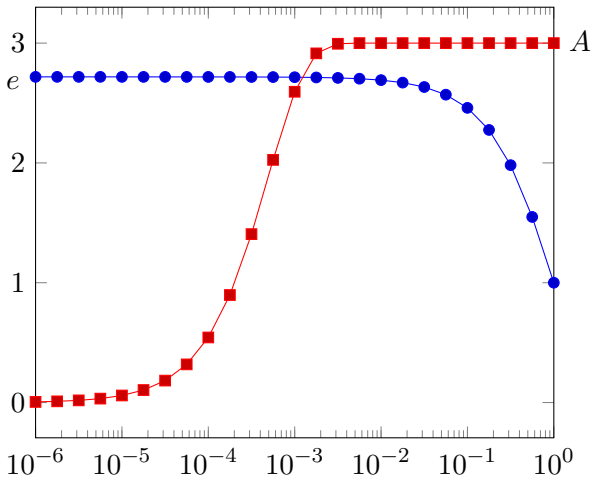


FIGURE 1. Sketch of the outer solution y_0 (blue) and of the boundary-layer or inner solution Y_0 (red). Notice that if one would set $A = e$, then the two solutions would "overlap" over some interval.

Step 3. Matching: Recall that, in principle, both expansions y_0 and Y_0 are approximations of the same function. Thus, it is reasonable to expect that in the transition region (away from $t = 0$ and of $t = 1$) the two approximations coincide. This is achieved by asking that the value of Y_0 as $\tilde{t} \to \infty$ (as it leaves the boundary layer) approaches the value of y_0 as $t \to 0$ (as it approaches the boundary layer). See Figure 2. Formally this is:

$$\lim_{\tilde{t}\to\infty} Y_0 = \lim_{t\to 0} y_0,$$

which immediately implies that $A = e$.

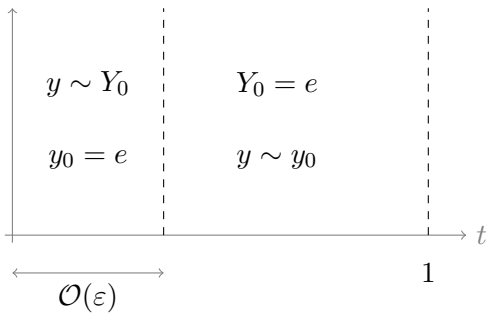


FIGURE 2. Sketch of the inner ($\mathcal{O}(\varepsilon)$) and outer regions. Within each of such regions the solution y has different approximations.

Step 4. Composite Expansion: The final step is to bring the two obtained approximations together, recall that neither (y_0 and Y_0) can be used over the whole interval $t \in [0, 1]$. Since via the matching step we know that both approximation are constant (and equal to e) away from their intervals of applicability, we can create a uniform solution by adding the two approximations and subtracting what is common between them, namely:

$$\begin{aligned} y(t) &\sim y_0(t) + Y_0(\tilde{t}) - e \\ &= \underbrace{\exp(1-t)}_{y_0} + \underbrace{\exp(1) - \exp(1 - \frac{2t}{\varepsilon})}_{Y_0(t/\varepsilon)} - \exp(1) \\ &= \exp(1-t) - \exp\left(1 - \frac{2t}{\varepsilon}\right). \end{aligned} \tag{28}$$

See Figure 3 for a comparison between the analytic solution

$$y(t) = c_1 \exp\left(-\frac{(\sqrt{1-2\varepsilon}+1)t}{\varepsilon}\right) + c_2 \exp\left(\frac{(\sqrt{1-2\varepsilon}-1)t}{\varepsilon}\right)$$

with $c_1 = \dfrac{1}{\exp\left(-\frac{(\sqrt{1-2\varepsilon}+1)}{\varepsilon}\right) - \exp\left(\frac{(\sqrt{1-2\varepsilon}-1)}{\varepsilon}\right)}$ and $c_2 = -c_1$, and the approximation (28).

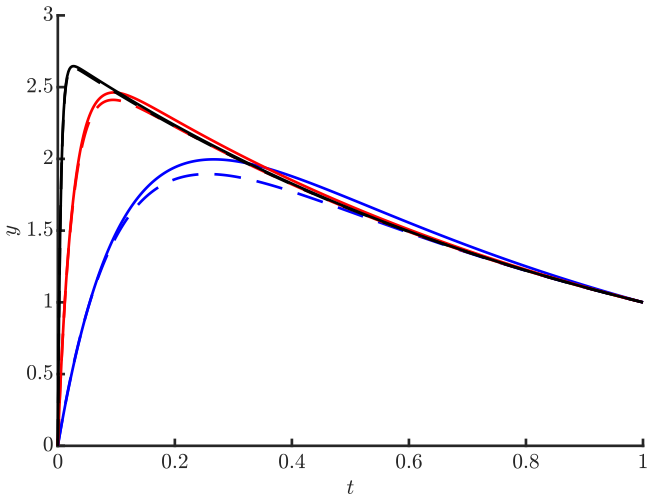


FIGURE 3. Comparison between the analytic solution (in solid) and the approximation (28) (in dashed) for $\varepsilon = 0.2,\ 0.05,\ 0.01$ in blue, red and black respectively.

[a]Recall that intuitively "a is of higher order than b" means that $a \ll b$.

REMARK III.1. The previous example contains, in some sense, all steps needed in the methodology known as "Matched Asymptotic Expantions". It is important to remark, however, that complications may appear in different situations. A common one is when the limits used in the matching process do not exist. Through more examples we will see some of such complications. Check further examples in [**13**].

EXAMPLE III.2 (Example III.1 continued (computation of the second term)). Usually, we will be satisfied by computing the first terms of the expansion of the solution for a perturbation problem. However, computing the second term may be insightful, as it can give us an idea of the error of the first-term approximation. In this example, we compute the second expansion term for the system of Example III.1.

The $\mathcal{O}(\varepsilon)$ terms from (22) correspond to

$$2\frac{\mathrm{d}y_1}{\mathrm{d}t} + 2y_1 = -\frac{\mathrm{d}^2 y_0}{\mathrm{d}t^2}, \qquad y_1(1) = 0.$$

The corresponding (particular) solution is:

$$y_1 = \frac{1}{2}(1-t)\exp(1-t).$$

Similarly, from (26), the $\mathcal{O}(1)$ terms obtained by setting $\beta_1 = 1$ lead to:

$$\frac{\mathrm{d}^2 Y_1}{\mathrm{d}\tilde{t}^2} + 2\frac{\mathrm{d}Y_1}{\mathrm{d}\tilde{t}} = -2Y_0, \qquad Y_1(0) = 0.$$

The corresponding (general) solution is:

$$Y_1 = B(1-\exp(-2\tilde{t})) - \tilde{t}\exp(1)(1+\exp(-2\tilde{t})),$$

where B is an arbitrary constant and will be used when matching.

For the matching, let us introduce an "intermediate variable" $t_\eta = \dfrac{t}{\varepsilon^\beta}$ where $0 < \beta < 1$, and we assume that $\eta = \eta(\varepsilon)$ is the interval where the inner and outer solutions coincide. We choose such a value for β because we want that the intermediate variable t_η lies between the outer scale $\mathcal{O}(1)$ and the inner scale $\mathcal{O}(\varepsilon)$.

Notice that, by definition, we have $t = \varepsilon^\beta t_\eta$ and $\tilde{t} = \varepsilon^{\beta-1} t_\eta$. Next we have:

$$\begin{aligned}
y_{\text{outer}} &\sim y_0 + \varepsilon y_1 + \cdots \\
&= \exp(1-t) + \frac{\varepsilon}{2}(1-t)\exp(1-t) + \cdots \\
&= \exp(1-\varepsilon^\beta t_\eta) + \frac{\varepsilon}{2}(1-\varepsilon^\beta t_\eta)\exp(1-\varepsilon^\beta t_\eta) + \cdots \\
&\sim \exp(1)\left(1 - \varepsilon^\beta t_\eta + \frac{\varepsilon^{2\beta} t_\eta^2}{2} + \cdots\right) + \frac{\varepsilon}{2}(1-\varepsilon^\beta t_\eta)\exp(1)\left(1 - \varepsilon^\beta t_\eta + \frac{\varepsilon^{2\beta} t_\eta^2}{2} + \cdots\right) \\
&= \exp(1)\left(1 - \varepsilon^\beta t_\eta + \frac{1}{2}\varepsilon + \cdots\right).
\end{aligned}$$

Through similar computations we obtain:

$$\begin{aligned} y_{\text{inner}} &\sim Y_0 + \varepsilon Y_1 + \cdots \\ &\sim \exp(1)(1 - \varepsilon^\beta t_\eta) + B\varepsilon + \cdots . \end{aligned}$$

Matching is achieved by setting $B = \frac{1}{2}\exp(1)$.

Since the above computations are independent of the precise value of β, let us choose $\beta = \frac{1}{2}$ for the composite expansion, which therefore reads as:

$$\begin{aligned} y &\sim y_0 + \varepsilon y_1 + Y_0 + \varepsilon Y_1 - \left(\exp(1)\left(1 - \sqrt{\varepsilon}t_\eta + \frac{1}{2}\varepsilon\right)\right) + \cdots \\ &= \underbrace{\exp(1-t) + \frac{\varepsilon}{2}(1-t)\exp(1-t)}_{y_0+\varepsilon y_1} + \\ &\quad \underbrace{\exp(1)\left(1 - \exp\left(-\frac{2t}{\varepsilon}\right) + \frac{\varepsilon}{2}\left(1 - \exp\left(\frac{-2t}{\varepsilon}\right)\right)\right) - t\left(1 + \exp\left(\frac{-2t}{\varepsilon}\right)\right)}_{Y_0+\varepsilon Y_1} - \\ &\quad \underbrace{\left(\exp(1)\left(1 - \sqrt{\varepsilon}t_\eta + \frac{1}{2}\varepsilon\right)\right)}_{\text{common}} + \cdots \\ &\sim \exp(1-t) - (1+t)\exp\left(-\frac{2t}{\varepsilon}\right) + \frac{\varepsilon}{2}\left((1-t)\exp(1-t) - \exp\left(1 - \frac{2t}{\varepsilon}\right)\right). \end{aligned}$$

We can now notice that the difference between the 1-term and the 2-term expansions are of order $\mathcal{O}(\varepsilon)$, as $\varepsilon \to 0$, in the interval $t \in [0, 1]$.

EXERCISE III.1. *In the previous example(s), we chose the boundary layer at $t = 0$. Repeat the computations by choosing the boundary layer at $t = 1$ and compare the results. Pay particular attention to the matching step, what do you notice?*

Hint: use a boundary coordinate of the form $\hat{t} = \dfrac{t-1}{\varepsilon^\alpha}$.

REMARK III.2. As already mentioned, in this section we have presented already the essence of the "Matched Asymptotic Expansions" method. As it is evident from the examples already treated, the method may present several complications depending on the particular problem under study. In the following sections we present a few examples of slightly more complicated scenarios. See [13] for even more examples.

III.1.1. Boundary layers. In this section we treat examples where more than one boundary layer must be accounted for.

EXAMPLE III.3. Let us consider

$$\varepsilon^2 \frac{\mathrm{d}^2 y}{\mathrm{d}t^2} + \varepsilon t \frac{\mathrm{d}y}{\mathrm{d}t} - y = -\exp(t), \qquad 0 < t < 1, \tag{29}$$

with boundary conditions $y(0) = 2$, $y(1) = 1$.

Step 1. Outer expansion: As before, we let the outer expansion be given as:

$$y_{\text{outer}} \sim y_0 + \cdots .$$

Thus, the first term of the outer expansion is simply

$$y_0 = \exp(t). \tag{30}$$

We now notice that y_0 cannot satisfy either of the boundary conditions. This is an indication that there are boundary layers *at each end of the interval* $t \in [0,1]$.

Step 2. Boundary Layers and Balancing:

a) Let us begin with the boundary layer at $t = 0$ by introducing a boundary layer coordinate $\tilde{t} = \dfrac{t}{\varepsilon^\alpha}$. As before, we denote by $Y = Y(\tilde{t})$ the re-scaled solution. Thus (29) now reads as:

$$\underbrace{\varepsilon^{2-2\alpha}\frac{\mathrm{d}^2 Y}{\mathrm{d}\tilde{t}^2}}_{①} + \underbrace{\varepsilon\tilde{t}\frac{\mathrm{d}Y}{\mathrm{d}\tilde{t}}}_{②} - \underbrace{Y}_{③} = \underbrace{-\exp(\varepsilon^\alpha \tilde{t})}_{④}$$

$$\sim -\left(1 + \varepsilon^\alpha \tilde{t} + \frac{\varepsilon^{2\alpha}\tilde{t}^2}{2} + \cdots\right).$$

We notice that, since $\alpha > 0$, the convenient balancing is between the terms ①, ③, and ④ by setting $\alpha = 1$ (a smaller value of α would lead to a balance already accounted in the outer expansion, while a larger value would not lead to any balancing). Thus, by setting $\alpha = 1$ and to leading terms considering that $Y \sim Y_0 + \cdots$, we have:

$$\frac{\mathrm{d}^2 Y_0}{\mathrm{d}\tilde{t}^2} - Y_0 = -1, \qquad Y_0(0) = 2. \tag{31}$$

The corresponding general solution is

$$Y_0(\tilde{t}) = 1 + A\exp(-\tilde{t}) + (1 - A)\exp(\tilde{t}),$$

where A is an arbitrary constant to be used for matching. Due to the definition $\tilde{t} = \dfrac{t}{\varepsilon}$ and because the boundary layer is located at $t = 0$, the matching condition is $\lim\limits_{\tilde{t}\to\infty} Y_0(\tilde{t}) = \lim\limits_{t\to 0} y_0(t)$, which leads to $A = 1$, and so

$$Y_0(\tilde{t}) = 1 + \exp(-\tilde{t}).$$

b) Next, let us look at the boundary layer at $t = 1$. For this let us introduce the boundary-layer coordinate:

$$\hat{t} = \frac{t - 1}{\varepsilon^\beta}.$$

Accordingly, we denote by $\hat{Y} = \hat{Y}(\hat{t})$ the solution in this layer. Notice that now $\hat{t} \in (-\infty, 0]$. Re-scaling accordingly in (29) leads to:

$$\varepsilon^{2-2\beta}\frac{\mathrm{d}^2\hat{Y}}{\mathrm{d}\hat{t}^2} + (1 + \varepsilon^\beta\hat{t})\varepsilon^{1-\beta}\frac{\mathrm{d}\hat{Y}}{\mathrm{d}\hat{t}} - \hat{Y} = \underbrace{-\exp(1 + \varepsilon^\beta\hat{t})}_{\sim -\exp(1)(1+\varepsilon^\beta\hat{t}+\cdots)}.$$

By choosing $\beta = 1$ we can, in fact, balance all terms obtaining, up to leading order terms:

$$\frac{\mathrm{d}^2\hat{Y}_0^2}{\mathrm{d}\hat{t}^2} + \frac{\mathrm{d}\hat{Y}_0}{\mathrm{d}\hat{t}} - \hat{Y}_0 = -\exp(1), \qquad \hat{Y}_0(0) = 1. \tag{32}$$

The corresponding general solution is

$$\hat{Y}_0(\hat{t}) = \exp(1) + B\exp(\lambda_+\hat{t}) + (1 - \exp(1) - B)\exp(\lambda_-\hat{t}),$$

where B is an arbitrary constant to be used for matching and $\lambda_\pm = \dfrac{-1 \pm \sqrt{5}}{2}$. The matching requirement now is analogous to the previous boundary layer. Namely, we require that the value of $\hat{Y}_0(\hat{t})$ as it leaves the boundary layer is equal to $y_0(t)$ as it approaches the boundary layer. In other words we now look for the constant B that allows the following equality to hold:

$$\lim_{\hat{t}\to-\infty} \hat{Y}_0(\hat{t}) = \lim_{t\to 1} y_0(t).$$

The previous condition implies that $B = 1 - \exp(1)$. Hence:

$$\hat{Y}_0(\hat{t}) = \exp(1) + (1 - \exp(1))\exp(\lambda_+\hat{t}).$$

Step 3. Composite expansion: We are now ready to derive the composite expansion. Analogous to previous cases, the composite expansion is obtained as:

$$y \sim y_{\text{outer}} + y_{\text{inner}} - \text{“common terms”}.$$

In more detail we now have, up to 1-st order terms:

$$\begin{aligned} y &\sim y_0(t) + Y_0(t/\varepsilon) + \hat{Y}_0\left(\frac{t-1}{\varepsilon}\right) - y_0(0) - y_0(1) \\ &\sim \exp(t) + \exp\left(-\frac{t}{\varepsilon}\right) + (1 - \exp(1))\exp\left(\frac{\lambda_+(t-1)}{\varepsilon}\right) + \cdots . \end{aligned}$$

Remark III.3. Notice in the previous example that both inner equations ((31) and (32)) have at least one common term with the outer equation (30). This is always evidence that the matching may work.

Exercise III.2. *Consider the differential equation*

$$\varepsilon\frac{\mathrm{d}^2y}{\mathrm{d}t^2} + \varepsilon\frac{\mathrm{d}y}{\mathrm{d}t} - \exp(y) = -2 - t, \qquad 0 < t < 1,$$

and boundary conditions $y(0) = 0$ *and* $y(1) = 1$.

(1) Show that the outer solution is given by $y_{outer} \sim \ln(t+2)$. *Is any boundary condition satisfied by the outer solution? From here argue that there are two boundary layers, one at* $t = 0$ *and another at* $t = 1$.

(2) For the boundary layer at $t = 0$ *propose the boundary-coordinate as* $\tilde{t} = \dfrac{t}{\sqrt{\varepsilon}}$. *Justify this choice and find the first term of the inner expansion* $Y_0(\tilde{t})$. *Hint: the solution* Y_0 *is given implicitly.*

(3) *For the boundary layer at $t = 1$ propose a boundary-layer coordinate $\hat{t} = \dfrac{t-1}{\sqrt{\varepsilon}}$. Justify this choice and find the first term of the inner expansion $\hat{Y}_0(\hat{t})$. Hint: the solution $\hat{Y}_0$ is given implicitly.*

(4) *The composite expansion is now given by: $y \sim y_0(t) + Y_0(\tilde{t}) + \hat{Y}_0\left(\hat{t}\right) - y_0(0) - y_0(1)$. Show that this leads to:*

$$y \sim \ln\left(\frac{t+2}{6}\right) + Y_0(\tilde{t}) + \hat{Y}_0(\hat{t})$$

EXAMPLE III.4. Matched asymptotic expansions is not limited to boundary value problems. Let us now consider (an auto-catalytic system):

$$\begin{aligned} \varepsilon \frac{\mathrm{d}u}{\mathrm{d}t} &= \exp(-t) - uv^2 - u \\ \frac{\mathrm{d}v}{\mathrm{d}t} &= uv^2 + u - v, \end{aligned} \tag{33}$$

with initial conditions $u(0) = v(0) = 1$ and $0 < \varepsilon \ll 1$. We assume that, due to the exponential term, there is a boundary layer at $t = 0$. This means that we shall find an inner solution that is valid near $t = 0$ and an outer solution valid away from $t = 0$.

For the outer solution let us assume that $u_{\text{outer}} \sim u_0 + \varepsilon u_1 + \cdots$ and $v_{\text{outer}} \sim v_0 + \varepsilon v_1 + \cdots$. Substituting these expansions in (33), and considering only the $\mathcal{O}(1)$ terms we have:

$$\begin{aligned} 0 &= \exp(-t) - u_0 v_0^2 - u_0 \\ \frac{\mathrm{d}v_0}{\mathrm{d}t} &= u_0 v_0^2 + u_0 - v_0. \end{aligned}$$

The solution of this first approximation system is

$$\begin{aligned} v_0(t) &= (t + a)\exp(-t) \\ u_0(t) &= \frac{\exp(-t)}{v_0^2 + 1}, \end{aligned}$$

where a is an arbitrary constant. Notice indeed that the initial condition cannot be satisfied by the outer solution.

For the layer near $t = 0$, let us consider the re-scaling variable $\tau = \dfrac{t}{\varepsilon}$. As usual we denote by $U = U(\tau) \sim U_0(\tau) + \cdots$ and $V = V(\tau) \sim V_0(\tau) + \cdots$ the inner solutions. Substituting the inner solutions in (33) leads to (up to terms of order $\mathcal{O}(1)$):

$$\begin{aligned} \frac{\mathrm{d}U_0}{\mathrm{d}\tau} &= 1 - U_0 V_0^2 - U_0 \\ \frac{\mathrm{d}V_0}{\mathrm{d}\tau} &= 0, \end{aligned}$$

with $U_0(0) = V_0(0) = 1$. The corresponding particular solution is

$$\begin{aligned} U_0(\tau) &= \frac{1}{2}(1 + \exp(-2\tau)) \\ V_0(\tau) &= 1. \end{aligned}$$

To match the solution we notice that the condition to be satisfied is $\lim_{\tau\to\infty} U_0(\tau) = \lim_{t\to 0} u_0(t)$ and $\lim_{\tau\to\infty} V_0(\tau) = \lim_{t\to 0} v_0(t)$ leading to $a = 1$. Finally, the composite expansion reads as:

$$u \sim \frac{\exp(-t)}{(t+1)^2\exp(-2t)+1} + \frac{\exp\left(\frac{-2t}{\varepsilon}\right)}{2}$$
$$v \sim (t+1)\exp(-t).$$

EXERCISE III.3. *Consider the differential equation*

$$\varepsilon^3\frac{\mathrm{d}^2y}{\mathrm{d}t^2} + t^3\frac{\mathrm{d}y}{\mathrm{d}t} - \varepsilon y = t^3, \qquad 0 < t < 1,$$

with boundary conditions $y(0) = 1$ *and* $y(1) = 3$.

(1) Simulate the differential equation using a computer. Does the simulation hint at the boundary layers?

Hint: Try with $\varepsilon = 0.001$ *and look near* $t = 0$.

(2) Conclude from the previous simulation that there are two boundary layers near $t = 0$.

(3) Show that the outer expansion is $y_{\text{outer}} \sim t + 2$ *and plot it on top of the simulation of item 1.*

(4) Consider now a time re-scaling of the form $\tilde{t} = \dfrac{t}{\varepsilon^\alpha}$. *Denoting by* $Y = Y(\tilde{t})$ *the re-scaled solution, obtain the re-scaled differential equation. Argue that, besides the balancing leading to the outer solution, there are two distinguished limits: one for* $\alpha = 1$ *and the other for* $\alpha = \dfrac{1}{2}$. *What portions of the solution do each of these limits correspond to?*

(5) Compute the first-term expansions for each layer ($\alpha = 1$ *and* $\alpha = \dfrac{1}{2}$*). Then, show that the composite expansion is of the form:*

$$y \sim \begin{cases} t + \exp\left(-\dfrac{t}{\varepsilon}\right) + 2\exp\left(-\dfrac{\varepsilon}{2t^2}\right), & 0 < t \le 1, \\ 1, & t = 0. \end{cases}$$

III.1.2. Interior layers. In the previous sections we have seen examples where the boundary layers occur at the end-points of the the interval of interest. In this section we give an example of a problem with an *interior layer*.

EXAMPLE III.5. Let us consider

$$\varepsilon\frac{\mathrm{d}^2y}{\mathrm{d}t^2} - y\frac{\mathrm{d}y}{\mathrm{d}t} + y = 0, \qquad 0 < t < 1, \tag{34}$$

with boundary conditions $y(0) = 1$ and $y(1) = -1$. The outer expansion is proposed as usual, namely: $y_{\text{outer}} \sim y_0 + \cdots$. Substituting in (34) and accounting only for the first term of the expansion we get:

$$y_0\frac{\mathrm{d}y_0}{\mathrm{d}t} - y_0 = 0. \tag{35}$$

The ODE (35) has solutions $y_0 = 0$ or $y_0 = t + a$, where a is an arbitrary constant. This implies that, depending on the inner solution we will have to be careful with the choice of the outer solution to match to.

Now, *how do we know there "must be" an interior layer?*. To realize this, let us rewrite (34) as

$$\varepsilon \frac{\mathrm{d}^2 y}{\mathrm{d}t^2} = y \frac{\mathrm{d}y}{\mathrm{d}t} - y = y\left(\frac{\mathrm{d}y}{\mathrm{d}t} - 1\right), \tag{36}$$

and let us consider first the outer solution $y \sim y_0 = 0$. If we were to use such an outer solution, then due to the boundary condition, we would need boundary layers at $t = 0$ and $t = 1$. In the boundary near $t = 0$ we would require $\dfrac{\mathrm{d}y}{\mathrm{d}t} < 0$, $y > 0$ and $\dfrac{\mathrm{d}^2 y}{\mathrm{d}t^2} > 0$, however no solution of (36) can satisfy that. A similar situation occurs if one would hope for a boundary layer at $t = 1$, see a schematic representation of this argument in figure 4 (it is left as an exercise to arrive to the same conclusion if one would take the first term of the outer expansion $y_0 = t + a$).

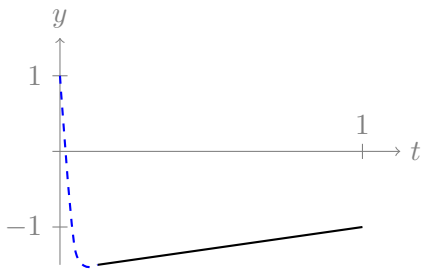


FIGURE 4. Hypothetical situation if one would look for a boundary layer at $t = 0$. This situation is not compatible with the original equation. In particular, it is not possible to satisfy the behavior on the boundary layer near $t = 0$ (blue dashed curve).

From the previous arguments, now we ask ourselves if an interior layer would be compatible with the ODE. For this, we take the outer solution $y_0 = t + a$ (the solution $y_0 = 0$ is of no help here). The equation $y_0 = t + a$ is a straight line with positive slope. Let $t_0 \in (0, 1)$ be some interior layer. To the left of t_0 (that is for $0 \leq t < t_0$) we let

$$y_0 = t + 1$$

noticing that it satisfies the initial condition. To the right of t_0 (that is for $t_0 < t \leq 1$) we take the outer solution

$$y_0 = -2 + t$$

noticing that it satisfies the boundary condition. Within the layer region, and to the left of t_0, we would have $\dfrac{\mathrm{d}y}{\mathrm{d}t} < 0$, $y > 0$ and $\dfrac{\mathrm{d}^2 y}{\mathrm{d}t^2} < 0$, which is compatible with (36), see a sketch in figure 5. A similar argument applies on the right of t_0. What is important to notice is that, before engaging into blindly computing boundary layers, it is worth stopping for a moment and qualitatively analyzing where it is most likely to find the inner layers.

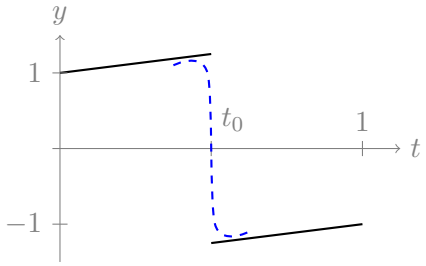


FIGURE 5. Sketch of the interior layer at $t = t_0$.

From the previous arguments, now we propose the interior-layer coordinate

$$\tilde{t} = \frac{t - t_0}{\varepsilon^\alpha}.$$

As already suggested above, we will now consider two outer regions, one for $0 \leq t < t_0$ and another for $t_0 < t \leq 1$. As usual, let $Y = Y(\tilde{t})$ denote the solution in the re-scaled coordinate. Then (36) now reads as:

$$\varepsilon^{1-2\alpha}\frac{\mathrm{d}^2 Y}{\mathrm{d}\tilde{t}^2} = \varepsilon^{-\alpha} Y \frac{\mathrm{d}Y}{\mathrm{d}\tilde{t}} - Y.$$

Choosing $\alpha = 1$ and letting $Y \sim Y_0 + \cdots$, we get the balanced equation

$$\frac{\mathrm{d}Y_0^2}{\mathrm{d}\tilde{t}^2} = Y_0 \frac{\mathrm{d}Y_0}{\mathrm{d}\tilde{t}}. \tag{37}$$

Integrating once the previous equation one gets

$$\frac{\mathrm{d}Y_0}{\mathrm{d}\tilde{t}} = \frac{1}{2}Y_0^2 + A.$$

Depending on the sign of A one can get different solutions as follows:

$$Y_0 = \begin{cases} B\dfrac{1 - D\exp(B\tilde{t})}{1 + D\exp(B\tilde{t})}, & A < 0 \\ \dfrac{2}{C - \tilde{t}}, & A = 0 \\ B\tan\left(C - B\dfrac{\tilde{t}}{2}\right), & A > 0, \end{cases}$$

where B, C, and D are further arbitrary constants.

Since within the interior layer we would like the derivative of y (or of Y) to be negative, it seems reasonable to choose the solution to (37) as

$$Y_0(\tilde{t}) = B\frac{1 - D\exp(B\tilde{t})}{1 + D\exp(B\tilde{t})},$$

with B and D different from zero. Without loss of generality we may take $B > 0$.

Next, the matching conditions we want to satisfy are:

$$\lim_{\tilde{t}\to-\infty} Y(\tilde{t}) = \lim_{t\to t_0^-} y_0(t)$$
$$\lim_{\tilde{t}\to\infty} Y(\tilde{t}) = \lim_{t\to t_0^+} y_0(t),$$

which leads to the system

$$B = t_0 + 1$$
$$B = 2 - t_0,$$

with solution $B = \dfrac{3}{2}$ and $t_0 = \dfrac{1}{2}$.

The previous matching process has allowed us to fix the constant B, however the constant D is still arbitrary. We can now resolve this issue by imposing that $y_0(t_0) = 0$ or equivalently that $Y_0(0) = 0$ (refer back to the sketch...). From such a condition we get that $D = 1$, thus

$$Y_0(\tilde{t}) = \frac{3}{2}\frac{1 - \exp\left(\frac{3}{2}\tilde{t}\right)}{1 + \exp\left(\frac{3}{2}\tilde{t}\right)}.$$

It only rests to compute the composite expansion. For this example, it is very helpful that the outer expansion on both sides of t_0 is given by the same expression. Thus we have:

$$y \sim y_0(t) + Y_0\left(\frac{t-1/2}{\varepsilon}\right) - \underbrace{\text{“common terms”}}_{y_0(1/2)} + \cdots$$

$$\sim (t+a) + \frac{3}{2}\frac{1-\exp\left(\frac{3}{2}\left(\frac{t-1/2}{\varepsilon}\right)\right)}{1+\exp\left(\frac{3}{2}\left(\frac{t-1/2}{\varepsilon}\right)\right)} - \left(\frac{1}{2}+a\right) + \cdots$$

$$\sim t - \frac{1}{2} + \frac{3}{2}\frac{1-\exp\left(\frac{3}{2}\left(\frac{t-1/2}{\varepsilon}\right)\right)}{1+\exp\left(\frac{3}{2}\left(\frac{t-1/2}{\varepsilon}\right)\right)} + \cdots.$$

III.1.3. Corner layers. In the examples we have seen before, the layers are “roughly speaking” given by intervals where the solution to the ODE rapidly changes. However, it is possible that what changes rapidly is the derivative instead. In these cases, as we will see below, the solution seems to have a corner.

Example III.6. Let us consider the ODE

$$\varepsilon\frac{\mathrm{d}^2y}{\mathrm{d}t^2} + \left(t-\frac{1}{2}\right)\left(t+\frac{1}{2}\right)\frac{\mathrm{d}y}{\mathrm{d}t} - \left(t+\frac{1}{2}\right)y = 0, \qquad 0<t<1, \tag{38}$$

with boundary conditions $y(0)=2$, $y(1)=3$. Notice, indeed, that the coefficient of $\dfrac{\mathrm{d}y}{\mathrm{d}t}$ is zero (and changes sign) at $t=\dfrac{1}{2}$. For this reason we can already anticipate that the layer is located at $t=\dfrac{1}{2}$.

As usual, we consider that the outer expansion is of the form $y_{\text{outer}} \sim y_0 + \cdots$. To leading terms we have

$$\left(t-\frac{1}{2}\right)\left(t+\frac{1}{2}\right)\frac{\mathrm{d}y_0}{\mathrm{d}t} - \left(t+\frac{1}{2}\right)y_0 = 0$$

$$\left(t-\frac{1}{2}\right)\frac{\mathrm{d}y_0}{\mathrm{d}t} - y_0 = 0,$$

and therefore

$$y_0 = A\left(t-\frac{1}{2}\right),$$

with A an arbitrary constant. Again, it is not possible to satisfy the two boundary conditions with only one integration constant.

To locate and analyze the solution in the layer, let us introduce the variable $\tilde{t} = \dfrac{t-t_0}{\varepsilon^\alpha}$. Notice that with the outer solution we can write a solution satisfying one of the boundary conditions as:

$$y_{\text{outer}} \sim \begin{cases} -4\left(t-\dfrac{1}{2}\right), & 0 \le t < t_0 \\ 6\left(t-\dfrac{1}{2}\right), & t_0 \le t \le 1 \end{cases} \tag{39}$$

To locate the value of t_0 we simply notice that if $t_0 = \frac{1}{2}$ then (39) is continuous (although not differentiable). For other choice of t_0 (39) would be discontinuous (now it should be clear why this case is called "corner layer").

To find the inner expansion we proceed as usual. Let $Y = Y(\tilde{t})$ denote the re-scaled solution, and then (38) transforms to

$$\varepsilon^{1-2\alpha}\frac{\mathrm{d}^2 Y}{\mathrm{d}\tilde{t}^2} + (\varepsilon^\alpha \tilde{t})(\varepsilon^\alpha \tilde{t} + 1)\varepsilon^{-\alpha}\frac{\mathrm{d}Y}{\mathrm{d}t} - (\varepsilon^\alpha \tilde{t} + 1)Y = 0$$

$$\varepsilon^{1-2\alpha}\frac{\mathrm{d}^2 Y}{\mathrm{d}\tilde{t}^2} + (\tilde{t} + \varepsilon^\alpha \tilde{t}^2)\frac{\mathrm{d}Y}{\mathrm{d}t} - (\varepsilon^\alpha \tilde{t} + 1)Y = 0.$$

Now we can notice that the appropriate balancing is achieved by letting $\alpha = \frac{1}{2}$. Next, contrary to the previous examples, let us assume that $y_{\text{inner}} = Y \sim \varepsilon^{\gamma_0} Y_0 + \cdots$. The constant γ_0 will become useful later. To leading terms we thus get

$$\frac{\mathrm{d}^2 Y_0}{\mathrm{d}t^2} + \tilde{t}\frac{\mathrm{d}Y}{\mathrm{d}\tilde{t}} - Y_0 = 0,$$

which has solution:

$$Y_0(\tilde{t}) = B\tilde{t} + C\left(\exp\left(-\frac{\tilde{t}^{\,2}}{2}\right) + \tilde{t}\int_0^{\tilde{t}} \exp\left(-\frac{s^2}{2}\right)\mathrm{d}s\right).$$

Notice that for $|\tilde{t}|$ large:

$$Y_0 \sim \begin{cases} B\tilde{t} + C\sqrt{\frac{\pi}{2}}\tilde{t}, & \tilde{t} > 0 \\ B\tilde{t} - C\sqrt{\frac{\pi}{2}}\tilde{t}, & \tilde{t} < 0 \end{cases}$$

and therefore

$$y_{\text{inner}} \sim \begin{cases} \varepsilon^{\gamma_0}\tilde{t}\left(B + C\sqrt{\frac{\pi}{2}}\right), & \tilde{t} > 0 \\ \varepsilon^{\gamma_0}\tilde{t}\left(B - C\sqrt{\frac{\pi}{2}}\right), & \tilde{t} < 0. \end{cases}$$

Notice, however, that y_{inner} is unbounded as $|\tilde{t}| \to \infty$. So, to match the outer and the inner expansions, let us introduce yet another layer variable $\hat{t} = \frac{t - \frac{1}{2}}{\varepsilon^\kappa}$. With such re-scaling we get:

$$y_{\text{outer}} \sim \begin{cases} -4\varepsilon^\kappa \hat{t}, & \hat{t} < 0, \\ 6\varepsilon^\kappa \hat{t}, & \hat{t} > 0 \end{cases}$$

and

$$y_{\text{inner}} \sim \begin{cases} \varepsilon^{\gamma_0+\kappa-1/2}\hat{t}\left(B - C\sqrt{\frac{\pi}{2}}\right), & \hat{t} < 0 \\ \varepsilon^{\gamma_0+\kappa-1/2}\hat{t}\left(B + C\sqrt{\frac{\pi}{2}}\right), & \hat{t} > 0. \end{cases}$$

As anticipated above, we can now match the expansions by letting $\gamma_0 = \frac{1}{2}$, obtaining the matching conditions

$$B - C\sqrt{\frac{\pi}{2}} = -4$$
$$B + C\sqrt{\frac{\pi}{2}} = 6,$$

leading to $B = 1$ and $C = 5\sqrt{\frac{2}{\pi}}$. Therefore, the inner expansion is of the form

$$y_{\text{inner}} \sim \varepsilon^{1/2}\left(\tilde{t} + 5\sqrt{\frac{\pi}{2}}\left(\exp\left(-\frac{\tilde{t}^{\,2}}{2}\right) + \tilde{t}\int_0^{\tilde{t}} \exp\left(-\frac{s^2}{2}\right) \mathrm{d}s\right)\right).$$

III.2. Other methods

There exist a further variety of asymptotic methods depending on the problem at hand. Several of them, however, work under a similar philosophy as in the previous sections. In this section we briefly present a few other methods using an example. Further details and examples can be found in [**13**].

III.2.1. The WKB method.

The name WKB stems from the last names Wentzels, Kramers, and Brillouin, who were among the many scientist that popularized the method. This method is applicable whenever, for some reason, one can predict that the behavior of the solution is exponential.

Example III.7. Let us consider the equation

$$\varepsilon^2 \frac{\mathrm{d}^2 y}{\mathrm{d}t^2} - q(t)y = 0, \tag{40}$$

where q is some smooth function. To motivate the ansatz we will make later, let us first consider the case where $q(t) = q$ is a constant. In that case the general solution is

$$y(t) = c_1 \exp\left(\frac{\sqrt{q}}{\varepsilon}t\right) + c_2 \exp\left(-\frac{\sqrt{q}}{\varepsilon}t\right).$$

The main assumption of the WKB method is that the exponential behavior of the solution when q is constant, can be extended to approximate the solution of (40). In other words, when using the WKB method we assume that the solution of (40) is of the form

$$y \sim \exp\left(\frac{h(t)}{\varepsilon^\alpha}\right)(y_0(t) + \varepsilon^\alpha y_1(t) + \cdots). \tag{41}$$

It follows from (41) that

$$y' \sim \exp\left(\frac{h(t)}{\varepsilon^\alpha}\right)\left(\frac{1}{\varepsilon^\alpha}\frac{\mathrm{d}h}{\mathrm{d}t}(y_0 + \varepsilon y_1 + \cdots) + \frac{\mathrm{d}y_0}{\mathrm{d}t} + \varepsilon\frac{\mathrm{d}y_1}{\mathrm{d}t} + \cdots\right)$$

and

$$\frac{\mathrm{d}^2y}{\mathrm{d}t^2} \sim \exp\left(\frac{h(t)}{\varepsilon^\alpha}\right)\left[\frac{1}{\varepsilon^{2\alpha}}\left(\frac{\mathrm{d}h}{\mathrm{d}t}\right)^2(y_0+\varepsilon y_1+\cdots)+\frac{2}{\varepsilon^\alpha}\frac{\mathrm{d}h}{\mathrm{d}t}\left(\frac{\mathrm{d}y_0}{\mathrm{d}t}+\varepsilon\frac{\mathrm{d}y_1}{\mathrm{d}t}+\cdots\right)+\right.$$
$$\left.\frac{1}{\varepsilon^\alpha}\frac{\mathrm{d}^2h}{\mathrm{d}t^2}(y_0+\varepsilon y_1+\cdots)+\frac{\mathrm{d}^2y_0}{\mathrm{d}t^2}+\varepsilon\frac{\mathrm{d}^2y_1}{\mathrm{d}t^2}+\cdots\right]$$

Substituting these expansions in (40) we get

$$\varepsilon^2\left(\frac{1}{\varepsilon^{2\alpha}}\left(\frac{\mathrm{d}h}{\mathrm{d}t}\right)^2Y+\frac{2}{\varepsilon^\alpha}\frac{\mathrm{d}h}{\mathrm{d}t}\frac{\mathrm{d}Y}{\mathrm{d}t}+\frac{1}{\varepsilon^\alpha}\frac{\mathrm{d}^2h}{\mathrm{d}t^2}Y+\frac{\mathrm{d}^2Y}{\mathrm{d}t^2}\right)-q(t)Y=0, \tag{42}$$

where $Y=y_0+\varepsilon y_1+\cdots$.

Notice that the exponential term has canceled out. This occurred because the equation we are considering is linear. By further inspection we find that to balance the ε terms an appropriate choice is $\alpha=1$.

For the terms of order $\mathcal{O}(1)$ we have:

$$\left(\frac{\mathrm{d}h}{\mathrm{d}t}\right)^2-q(t)=0, \tag{43}$$

which has as solution

$$h(t)=\pm\int_0^t\sqrt{q(s)}\mathrm{d}s.$$

Notice indeed that if q is constant we obtain the same argument in the exponential as in our initial analysis.

For the terms of order $\mathcal{O}(\varepsilon)$ we have

$$\left(y_1\left(\frac{\mathrm{d}h}{\mathrm{d}t}\right)^2+2\frac{\mathrm{d}y_0}{\mathrm{d}t}\frac{\mathrm{d}h}{\mathrm{d}t}+y_0\frac{\mathrm{d}^2h}{\mathrm{d}t^2}\right)-q(t)y_1=0$$
$$y_1\underbrace{\left(\left(\frac{\mathrm{d}h}{\mathrm{d}t}\right)^2-q(t)\right)}_{=0}+2\frac{\mathrm{d}y_0}{\mathrm{d}t}\frac{\mathrm{d}h}{\mathrm{d}t}+y_0\frac{\mathrm{d}^2h}{\mathrm{d}t^2}=0$$
$$2\frac{\mathrm{d}y_0}{\mathrm{d}t}\frac{\mathrm{d}h}{\mathrm{d}t}+y_0\frac{\mathrm{d}^2h}{\mathrm{d}t^2}=0.$$

Let $u=y_0^2\dfrac{\mathrm{d}h}{\mathrm{d}t}$. It follows that $\dfrac{\mathrm{d}u}{\mathrm{d}t}=y_0\left(2\dfrac{\mathrm{d}y_0}{\mathrm{d}t}\dfrac{\mathrm{d}h}{\mathrm{d}t}+y_0\dfrac{\mathrm{d}^2h}{\mathrm{d}t^2}\right)=0$. Therefore, y_0 is given by

$$y_0=c\left(\frac{\mathrm{d}h}{\mathrm{d}t}\right)^{-1/2}$$
$$=cq(t)^{-1/4},$$

where we have used the equation for the $\mathcal{O}(1)$ terms and c is some arbitrary constant. With this, we have that up to first order terms, the solution of (40) can be approximated by

$$y\sim q(t)^{-1/4}\left(c_1\exp\left(\frac{1}{\varepsilon}\int_0^t\sqrt{q(s)}\mathrm{d}s\right)+c_2\exp\left(-\frac{1}{\varepsilon}\int_0^t\sqrt{q(s)}\mathrm{d}s\right)\right),$$

where the (possibly complex) constants c_1 and c_2 help us match the boundary conditions. It is evident now that for the above solution to be valid, one would require $q(t) \neq 0$. Points where $q(t) = 0$ are called *turning points*.

Let us now consider a particular expression for the function $q(t)$. Let $q(t) = -\exp(2t)$, and suppose that the boundary conditions are $y(0) = a$ and $y(1) = b$. Then (by re-labeling the constants $(-1)^{-1/4}c_i \mapsto c_i$, $i = 1, 2$) we have:

$$y \sim \exp\left(-\frac{t}{2}\right)\left(c_1 \exp\left(\imath\frac{\exp(t)}{\varepsilon}\right) + c_2 \exp\left(-\imath\frac{\exp(t)}{\varepsilon}\right)\right),$$

which can be rewritten as

$$y \sim \exp\left(-\frac{t}{2}\right)\left(\bar{c}_1 \cos\left(\frac{\exp(t)}{\varepsilon}\right) + \bar{c}_2 \sin\left(\frac{\exp(t)}{\varepsilon}\right)\right),$$

where $\bar{c}_1$ and $\bar{c}_2$ are new (complex) constants. We can compare this expansion with the analytic solution, namely

$$y(t) = AJ_0\left(\frac{\exp(t)}{\varepsilon}\right) + BY_0\left(\frac{\exp(t)}{\varepsilon}\right),$$

where A and B are constants defined by the boundary conditions and J_0 and Y_0 are Bessel functions. See figure 6 for a comparison.

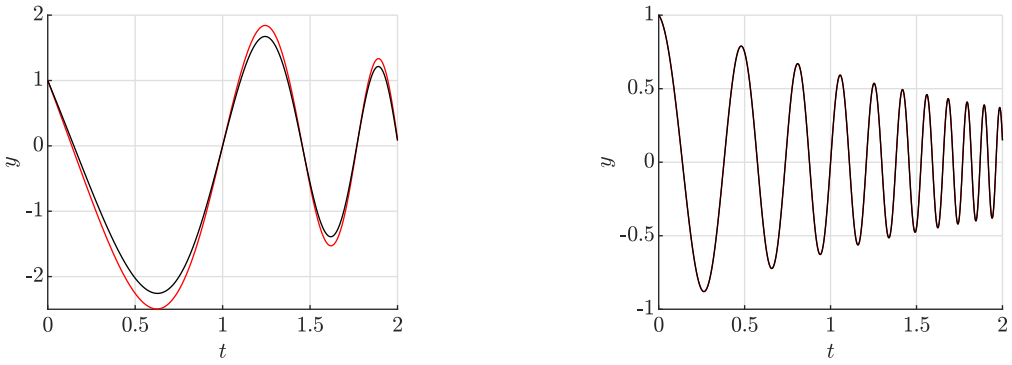


FIGURE 6. Comparison between the approximate solution (red) and the exact solution (black) for $\varepsilon = 0.5$ on the left and $\varepsilon = 0.1$ on the right. The boundary conditions are $y(0) = 1$ and $y(1) = 0$.

We now proceed with finding the error of the approximation y_0. This is done by computing the second term in the expansion y_1 from the terms of order $\mathcal{O}(\varepsilon^2)$ in (42). Accordingly we have:

$$\frac{\mathrm{d}^2 y_0}{\mathrm{d}t^2} + \frac{\mathrm{d}^2 h}{\mathrm{d}t^2}y_1 + 2\frac{\mathrm{d}h}{\mathrm{d}t}\frac{\mathrm{d}y_1}{\mathrm{d}t} + \left(\frac{\mathrm{d}h}{\mathrm{d}t}\right)^2 y_2 = q(t)y_2.$$

Using (43) we further have

$$\frac{\mathrm{d}^2 y_0}{\mathrm{d}t^2} + \frac{\mathrm{d}^2 h}{\mathrm{d}t^2}y_1 + 2\frac{\mathrm{d}h}{\mathrm{d}t}\frac{\mathrm{d}y_1}{\mathrm{d}t} = 0. \tag{44}$$

Using the identities: $y_0 = q^{-1/4}$, $\dfrac{\mathrm{d}h}{\mathrm{d}t} = \sqrt{q}$ and assuming that $y_1(t) = y_0(t)w(t)$, we have:

$$\frac{\mathrm{d}^2 y_0}{\mathrm{d}t^2} + \frac{\mathrm{d}^2 h}{\mathrm{d}t^2} y_1 + 2\frac{\mathrm{d}h}{\mathrm{d}t}\frac{\mathrm{d}y_1}{\mathrm{d}t} = 0$$
$$\frac{5}{16}\left(\frac{\mathrm{d}q}{\mathrm{d}t}\right)^2 - \frac{1}{4}q\frac{\mathrm{d}^2 q}{\mathrm{d}t^2} + 2q^{5/2}\frac{\mathrm{d}w}{\mathrm{d}t} = 0.$$

Therefore (44) holds with solution $y_1 = y_0 w$ if w satisfies

$$\frac{\mathrm{d}w}{\mathrm{d}t} = \frac{1}{8}\frac{1}{q^{3/2}}\frac{\mathrm{d}^2 q}{\mathrm{d}t^2} - \frac{5}{32}\frac{1}{q^{5/2}}\left(\frac{\mathrm{d}q}{\mathrm{d}t}\right)^2. \tag{45}$$

Notice that $\dfrac{1}{8}\dfrac{\mathrm{d}}{\mathrm{d}t}\left(\dfrac{1}{q^{3/2}}\dfrac{\mathrm{d}q}{\mathrm{d}t}\right) = \dfrac{1}{8}\dfrac{1}{q^{3/2}}\dfrac{\mathrm{d}^2 q}{\mathrm{d}t^2} - \dfrac{6}{32}\dfrac{1}{q^{5/2}}\left(\dfrac{\mathrm{d}q}{\mathrm{d}t}\right)^2$. Therefore, (45) can be rewritten as:

$$\frac{\mathrm{d}w}{\mathrm{d}t} = \frac{1}{8}\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{q^{3/2}}\frac{\mathrm{d}q}{\mathrm{d}t}\right) + \frac{1}{32}\frac{1}{q^{5/2}}\left(\frac{\mathrm{d}q}{\mathrm{d}t}\right)^2,$$

which leads to

$$w(t) = D + \frac{1}{8}\frac{1}{q^{3/2}}\frac{\mathrm{d}q}{\mathrm{d}t} + \frac{1}{32}\int_0^t \frac{1}{q(s)^{5/2}}\left(\frac{\mathrm{d}q(s)}{\mathrm{d}s}\right)^2 \mathrm{d}s,$$

where D is an arbitrary constant (a similar expression is found if one would consider $\dfrac{\mathrm{d}h}{\mathrm{d}t} = -\sqrt{q}$). Since the expansion, up to the second term reads as $y \sim y_0 + \varepsilon y_1 = y_0 + \varepsilon y_0 w = y_0(1+\varepsilon w)$, we have that a good approximation is guaranteed if $|\varepsilon w| \ll 1$, that is, if

$$\varepsilon\left(|D| + \frac{1}{8}\left|\frac{1}{q^{3/2}}\frac{\mathrm{d}q}{\mathrm{d}t}\right| + \frac{1}{32}\int_0^t \left|\frac{1}{q(s)^{5/2}}\left(\frac{\mathrm{d}q(s)}{\mathrm{d}s}\right)^2\right| \mathrm{d}s\right) \ll 1$$

holds (for the functions $|\cdot|$ means the ∞-norm). In particular, we notice that such a bound can be achieved if $q(t) \neq 0$ in the interval of interest.

We end this section by summarizing that the WKB method is useful whenever we assume that the behavior of the solutions is exponential. This is the usual case for linear systems, thus it is frequently applied in such cases. Much of the difficulties, of course, arise depending on the function $q(t)$. In particular, if q takes a zero value somewhere in the interval of interest, the above methodology may fail. See more details in [**13**]. A few nonlinear problems may be treated with the WKB method as well.

III.2.2. Poincaré-Lindstedt (or multiple time scales) method.

We recall that the method of matched asymptotic expansions consists of constructing different solutions for appropriate intervals of the independent variable t. In contrast, the Poincaré-Lindstedt method assumes a (single) generalized solution that is itself transformed, by introducing new coordinates, to adapt to each of the layers.

Example III.8. Let us consider the ODE

$$\frac{\mathrm{d}^2 y}{\mathrm{d}t^2} + \varepsilon\frac{\mathrm{d}y}{\mathrm{d}t} + y = 0, \qquad t \geq 0, \tag{46}$$

with initial conditions $y(0) = 0$ and $\dfrac{\mathrm{d}y}{\mathrm{d}t}(0) = 1$.
Since the parameter ε does not multiply the highers derivative, it is safe to assume that the solution admits a regular power series expansion of the form

$$y_\varepsilon(t) \sim y_0(t) + \varepsilon y_1(t) + \cdots . \tag{47}$$

Substituting (47) in (46) we get

$$\frac{\mathrm{d}^2 y_0}{\mathrm{d}t^2} + y_0 + \varepsilon\left(\frac{\mathrm{d}^2 y_1}{\mathrm{d}t^2} + \frac{\mathrm{d}y_0}{\mathrm{d}t} + y_1\right) + \mathcal{O}(\varepsilon^2) = 0.$$

Thus, at order $\mathcal{O}(1)$ we have

$$\frac{\mathrm{d}^2 y_0}{\mathrm{d}t^2} + y_0 = 0, \qquad y_0(0) = 0,\ \frac{\mathrm{d}y_0}{\mathrm{d}t}(0) = 1,$$

which has (particular) solution

$$y_0(t) = \sin(t).$$

Similarly, at order $\mathcal{O}(\varepsilon)$ we have

$$\frac{\mathrm{d}^2 y_1}{\mathrm{d}t^2} + y_1 = -\cos(t), \qquad y_1(0) = 0,\ \frac{\mathrm{d}y_1}{\mathrm{d}t}(0) = 0,$$

which has (particular) solution

$$y_1(t) = -\frac{1}{2} t \sin(t),$$

and therefore:

$$y_\varepsilon \sim \sin(t) - \frac{\varepsilon}{2} t \sin(t). \tag{48}$$

Notice that the expansion we just computed grows linearly with t. In fact, if we compare y_ε with the analytic (or even numeric) solution of (46), as in Figure 7, we see that our expansion is less accurate once $\varepsilon t \approx 1$.

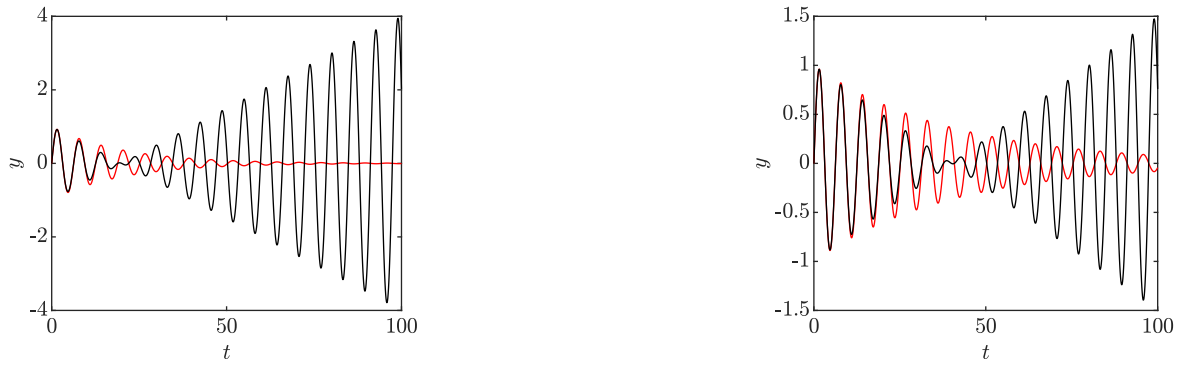


FIGURE 7. Comparison between the analytic solution of (46) (in red) and the approximation y_ε (in black) for $\varepsilon = 0.1$ (left) and $\varepsilon = 0.05$ (right). Notice that the approximation seems to be reasonable for $t < \dfrac{1}{\varepsilon}$.

In a qualitative sense, the discrepancy of our approximation stems from the second term in (48). If such an expansion is to be well-ordered, then we require $\varepsilon t \ll 1$.

To overcome the failed approximation for large t, consider new time parameters

$$\tau_1 = t$$
$$\tau_2 = \varepsilon^\alpha t.$$

> The time τ_2 is usually referred to as "slow time scale".

The important assumption here is that the two variables τ_1 and τ_2 are independent. This means that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} &\mapsto \frac{\mathrm{d}\tau_1}{\mathrm{d}t}\frac{\partial}{\partial\tau_1} + \frac{\mathrm{d}\tau_2}{\mathrm{d}t}\frac{\partial}{\partial\tau_2} \\ &= \frac{\partial}{\partial\tau_1} + \varepsilon^\alpha\frac{\partial}{\partial\tau_2} \\ \frac{\mathrm{d}^2}{\mathrm{d}t^2} &\mapsto \frac{\partial^2}{\partial\tau_1^2} + 2\varepsilon^\alpha\frac{\partial^2}{\partial\tau_1\partial\tau_2} + \varepsilon^{2\alpha}\frac{\partial^2}{\partial\tau_2^2}.\end{aligned}$$

In this way (46) transforms to

$$\left(\frac{\partial^2}{\partial\tau_1^2} + 2\varepsilon^\alpha\frac{\partial^2}{\partial\tau_1\partial\tau_2} + \varepsilon^{2\alpha}\frac{\partial^2}{\partial\tau_2^2}\right)y + \varepsilon\left(\frac{\partial}{\partial\tau_1} + \varepsilon^\alpha\frac{\partial}{\partial\tau_2}\right)y + y = 0, \tag{49}$$

where now $y = y(\tau_1, \tau_2)$, and therefore, the initial conditions now read as $y(0,0) = 0$ and $\left(\frac{\partial}{\partial\tau_1} + \varepsilon^\alpha\frac{\partial}{\partial\tau_2}\right)y(0,0) = 1$.

At this moment one should not despair. Although it may look like we have over complicated ourselves (after all we now have a PDE), the advantages will become evident shortly.

We now proceed with a similar methodology as usual. Since now y is a function of two variables, we assume an expansion of the form

$$y = y_\varepsilon(\tau_1, \tau_2) \sim y_0(\tau_1, \tau_2) + \varepsilon y_1(\tau_1, \tau_2) + \cdots. \tag{50}$$

Substituting (50) into (49) we get:

$$\begin{aligned}&\left(\frac{\partial^2}{\partial\tau_1^2} + 2\varepsilon^\alpha\frac{\partial^2}{\partial\tau_1\partial\tau_2} + \varepsilon^{2\alpha}\frac{\partial^2}{\partial\tau_2^2}\right)(y_0(\tau_1,\tau_2) + \varepsilon y_1(\tau_1,\tau_2) + \cdots) + \\ &\quad\varepsilon\left(\frac{\partial}{\partial\tau_1} + \varepsilon^\alpha\frac{\partial}{\partial\tau_2}\right)(y_0(\tau_1,\tau_2) + \varepsilon y_1(\tau_1,\tau_2) + \cdots) + (y_0(\tau_1,\tau_2) + \varepsilon y_1(\tau_1,\tau_2) + \cdots) = 0,\end{aligned} \tag{51}$$

From the terms of order $\mathcal{O}(1)$ we get

$$\left(\frac{\partial^2}{\partial\tau_1^2} + 1\right)y_0 = 0,$$

with initial conditions $y_0(0,0) = 0$, $\frac{\partial y_0}{\partial\tau_1}(0,0) = 1$. The general solution is given by

$$y_0 = c_1(\tau_2)\sin(\tau_1) + c_2(\tau_2)\cos(\tau_1), \tag{52}$$

where, to satisfy the initial conditions, we have $c_1(0) = 1$, $c_2(0) = 0$.

At this moment, recall that the term $\mathcal{O}(\varepsilon t)$ in (48) comes from the equation $\frac{\mathrm{d}^2y_0}{\mathrm{d}t^2} + y_1 + \frac{\mathrm{d}y_0}{\mathrm{d}t} = 0$. We want to avoid such a (secular) term. Inspecting (51) we notice that at order $\mathcal{O}(\varepsilon)$ a similar

expression appears. However, if we would choose $\alpha = 1$, then a new term, namely $2\varepsilon \dfrac{\partial^2}{\partial\tau_1\partial\tau_2} y_0$ also appears, and it will help us resolve our problem. Thus, the appropriate balance is achieved by letting $\alpha = 1$ leading to (at $\mathcal{O}(\varepsilon)$):

$$\frac{\partial^2}{\partial\tau_1^2} y_1 + 2\frac{\partial^2}{\partial\tau_1\partial\tau_2} y_0 + \frac{\partial}{\partial\tau_1} y_0 + y_1 = 0$$
$$\left(\frac{\partial^2}{\partial\tau_1^2} + 1\right) y_1 = -2\frac{\partial^2}{\partial\tau_1\partial\tau_2} y_0 - \frac{\partial}{\partial\tau_1} y_0.$$

From (52) we further have

$$\left(\frac{\partial^2}{\partial\tau_1^2} + 1\right) y_1 = -\left(2\frac{\mathrm{d}c_1}{\mathrm{d}\tau_2} + c_1\right)\cos(\tau_1) + \left(2\frac{\mathrm{d}c_2}{\mathrm{d}\tau_2} + c_2\right)\sin(\tau_1). \tag{53}$$

The general solution of (53) is

$$y_1(\tau_1,\tau_2) = -\frac{1}{2}\left(2\frac{\mathrm{d}c_1}{\mathrm{d}\tau_2} + c_1\right)\tau_1\sin(\tau_1) - \frac{1}{2}\left(2\frac{\mathrm{d}c_2}{\mathrm{d}\tau_2} + c_2\right)\tau_1\cos(\tau_1) + c_3(\tau_2)\sin(\tau_1) + c_4(\tau_2)\cos(\tau_1).$$

We observe that the secular terms ($\tau\sin(\cdot)$ and $\tau\cos(\cdot)$) still appear. However, we can eliminate them by appropriately choosing the functions c_1 and c_2. Indeed we want to solve:

$$2\frac{\mathrm{d}c_1}{\mathrm{d}\tau_2} + c_1 = 0$$
$$2\frac{\mathrm{d}c_2}{\mathrm{d}\tau_2} + c_2 = 0,$$

which have general solutions

$$c_1(\tau_2) = A_1\exp\left(-\frac{1}{2}\tau_2\right)$$
$$c_2(\tau_2) = A_2\exp\left(-\frac{1}{2}\tau_2\right).$$

Recalling that the initial conditions impose $c_1(0) = 1$ and $c_2(0) = 0$ we finally get

$$c_1(\tau_2) = \exp\left(-\frac{1}{2}\tau_2\right)$$
$$c_2(\tau_2) = 0.$$

Therefore, since $\tau_2 = \varepsilon t$, we have obtained the leading term expansion

$$y_\varepsilon \sim \exp\left(-\frac{\varepsilon t}{2}\right)\sin(t) + \cdots. \tag{54}$$

See in Figure 8 plots for the error between the analytic solution and (54).

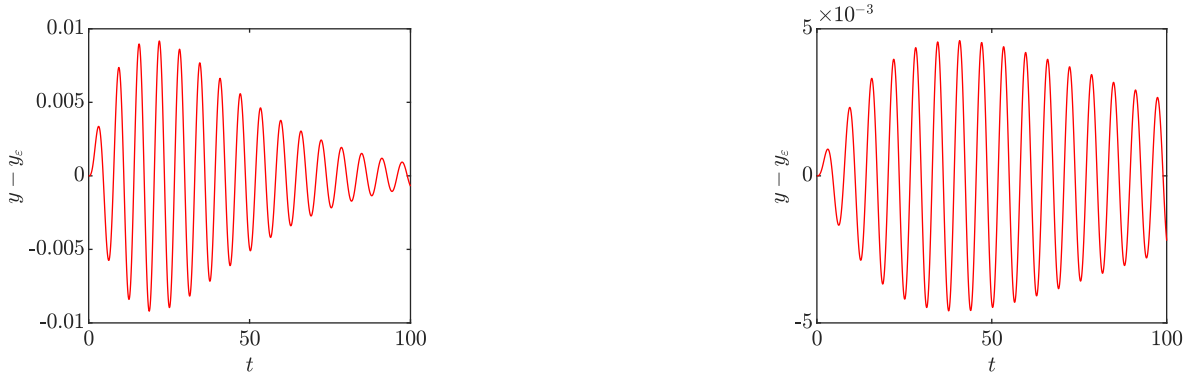


FIGURE 8. Difference between the analytic solution of (46) and the approximation y_ε (54) for $\varepsilon = 0.1$ (left) and $\varepsilon = 0.05$ (right). Notice that indeed the approximation error is already quite small for sufficiently large time.

EXERCISE III.4. *Find that the two term expansion for the previous example is*

$$y_\varepsilon \sim \exp\left(-\frac{\varepsilon t}{2}\right)\left(\sin(t) - \frac{1}{2}\varepsilon^2 t\cos(t)\right).$$

Plot the error between this expansion and the analytic solution.

Hints: what is required is to find y_1.

The analytic solution is $y(t) = \dfrac{1}{\sqrt{1-\frac{\varepsilon^2}{4}}}\exp\left(-\dfrac{\varepsilon t}{2}\right)\sin\left(t\sqrt{1-\dfrac{\varepsilon^2}{4}}\right)$

In this example we have seen the basic way the method of multiple time scales works. Notice that the method helped us to obtain a single solution valid on a sufficiently large interval (in contrast to matched expansion where we need to compute a solution on different layers). Moreover we were able to resolve the issue induced by the appearance of secular terms. There are several generalization of the method, see [**13**]. For example, two time scales may not be enough, and one may be required to define $\tau_1 = t$, $\tau_2 = \varepsilon t$, $\tau_3 = \varepsilon^2 t$ and so on. Furthermore, the new time variables may have a more complicated dependence on ε, for example $\tau_1 = (1+\sum_{k=1}^{\infty}\omega_k\varepsilon^k)t$, where the coefficients ω_k are found during the solving process (this is known as Lindstedt's method). Finally, we mention that the correct scaling is not always evident, and it may be necessary to start with something like $\tau_1 = \varepsilon^\alpha t$, $\tau_2 = \varepsilon^\beta t$ with $\alpha < \beta$

EXERCISE III.5. *Solve the same problem as in the previous example by assuming* $\tau_1 = (1+\omega_1\varepsilon)t$. *Compare your result with that of the example.*

III.3. Further exercises for this chapter

(1) For the following second order systems, find a composite expansion of the solution and sketch it. To verify, it is recommended to compare the approximation with a numerical solution of the differential equation.

(a) $\varepsilon\dfrac{\mathrm{d}^2y}{\mathrm{d}t^2}+\varepsilon(t+1)\dfrac{\mathrm{d}y}{\mathrm{d}t}-y=t-1$, for $t\in[0,1]$ and $y(0)=0$, $y(1)=-1$.

(b) $\varepsilon\dfrac{\mathrm{d}^2y}{\mathrm{d}t^2}-exp(t)y=f(t)$, for $t\in[0,1]$ and $y(0)=1$, $y(1)=-1$.

(c) $\varepsilon\dfrac{\mathrm{d}^2y}{\mathrm{d}t^2}-y\left(\dfrac{\mathrm{d}y}{\mathrm{d}t}+1\right)=0$, for $t\in[0,1]$ and $y(0)=3$, $y(1)=3$.

(2) The Michaelis-Menten model for enzyme catalyzed reactions is given by

$$\begin{aligned}\frac{\mathrm{d}S}{\mathrm{d}t}&=-S+(\mu+S)P\\ \varepsilon\frac{\mathrm{d}P}{\mathrm{d}t}&=S-(\kappa+S)P,\end{aligned}$$

where $S(0)=1$ and $P(0)=0$. The variables $S(t)$ and $P(t)$ are the concentrations of the substrate and of the product, respectively, of the catalyzed reaction, and μ, κ are positive constants with $\mu<\kappa$.

(a) Find the first term of the expansion in the outer layer
(b) Find the first term of the expansion in the initial layer
(c) Find a composite expansion.

(3) The Poisson-Nerst-Planck model for the flow of ions through a membrane is given by

$$\begin{aligned}\frac{\mathrm{d}p}{\mathrm{d}x}+p\frac{\mathrm{d}\phi}{\mathrm{d}x}&=-\alpha\\ \frac{\mathrm{d}n}{\mathrm{d}x}-n\frac{\mathrm{d}\phi}{\mathrm{d}x}&=-\beta\\ \varepsilon^2\frac{\mathrm{d}^2\phi}{\mathrm{d}x^2}&=-p+n,\end{aligned}$$

where $x\in[0,1]$, the variables p and n describe the concentration of ions with valency 1 and -1 respectively, and ϕ is the potential. Let the boundary conditions be $\phi(0)=1$, $\phi(1)=0$, $p(0)=4$, and $n(0)=1$. Suppose that $\alpha\neq\beta$ are positive constants satisfying

$$\kappa=\frac{\alpha+\beta}{\sqrt{p(0)n(0)}}<1.$$

(a) Assume that there is a boundary layer at $x=0$. Derive the outer and boundary-layer approximations. Is it possible to match the approximations? Provide enough reasoning.
(b) From the previous item argue that there should be another boundary layer at $x=1$. Derive the corresponding approximation and a composite solution. Show that the approximate values of the concentrations at $x=1$ are

$$\begin{aligned}p(1)&\sim p(0)\exp(\phi(0))(1-\kappa)^{2\beta/(\alpha+\beta)}\\ n(1)&\sim n(0)\exp(\phi(0))(1-\kappa)^{2\alpha/(\alpha+\beta)}.\end{aligned}$$

(4) Find the first term expansion of the solution for the following problems. Whenever possible, it is recommended to compare the approximation with a numerical solution. Note that for the

nonlinear problems the solutions may either be defined implicitly, or the transition layer may contain an undetermined constant. Both are correct and should not be worked-out further.

(a) $\varepsilon\dfrac{\mathrm{d}^2y}{\mathrm{d}t^2} = -\left(t^2 - \dfrac{1}{4}\right)\dfrac{\mathrm{d}y}{\mathrm{d}t}$ for $t \in [0,1]$ with $y(0) = 1$ and $y(1) = -1$.

(b) $\varepsilon\dfrac{\mathrm{d}^2y}{\mathrm{d}t^2} = y\dfrac{\mathrm{d}y}{\mathrm{d}t} - y^3$ for $t \in [0,1]$ with $y(0) = 3/5$ and $y(1) = -2/3$.

(c) $\varepsilon\dfrac{\mathrm{d}^2y}{\mathrm{d}t^2} + y(1+y^2)\dfrac{\mathrm{d}y}{\mathrm{d}t} - \dfrac{1}{2}y = 0$ for $t \in [0,1]$ with $y(0) = -1$ and $y(1) = 1$.

(5) Consider the problem

$$\varepsilon\frac{\mathrm{d}^2y}{\mathrm{d}t^2} = y\frac{\mathrm{d}y}{\mathrm{d}t}$$

for $t \in [0,1]$ with $y(0) = a$ and $y(1) = -a$ and $a > 0$.

(a) Argue sufficiently to justify that $y(1/2) = 0$.

(b) Find a composite expansion of the solution.

(c) Show that the exact solution is of the form

$$y(t) = A\frac{1 - B\exp\left(\frac{Ax}{\varepsilon}\right)}{1 + B\exp\left(\frac{Ax}{\varepsilon}\right)},$$

where, for $\varepsilon > 0$ small, $A = A(\varepsilon) \sim a\left(1 + 2\exp\left(\dfrac{-a}{2\varepsilon}\right)\right)$ and $B = B(\varepsilon) \sim \exp\left(\dfrac{-a}{2\varepsilon}\right)$.

(6) Consider the problem

$$\varepsilon\frac{\mathrm{d}^2y}{\mathrm{d}t^2} = -f(t)\frac{\mathrm{d}y}{\mathrm{d}t},$$

for $t \in [0,1]$ with $y(0) = a$ and $y(1) = -b$ and $a, b > 0$. Moreover $f(t)$ is smooth with $\dfrac{\mathrm{d}f}{\mathrm{d}t}(t) > 0$ and $f(t_0) = 0$ for some $t_0 \in (0,1)$.

(a) Explain why there is at least one point in the interval $0 < t < 1$ for which $y(t) = 0$.

(b) Find the exact solution of the problem and obtain the equation that must be satisfied to determine the point t for which $y(t) = 0$. Is the solution to such an equation unique?

(c) Find a two term expansion for the solution of the equation $y(t) = 0$ of the previous item. Note that the second term is defined implicitly.

(7) Consider the problem

$$\varepsilon\frac{\mathrm{d}^2y}{\mathrm{d}t^2} = t^2\left(1 - \left(\frac{\mathrm{d}y}{\mathrm{d}t}\right)^2\right),$$

for $t \in [0,1]$ with $y(0) = 1$ and $y(1) = 1$.

(a) Assume that there is a corner-layer solution. Argue why such an assumption is reasonable. How many outer-solutions are there? (there is more than one possible one, but it is possible to rule out one)

(b) Find the corner-layer solution and construct a composite expansion.

(8) Use the WKB method to find an approximate solution of the following problems:
(a) $\varepsilon\dfrac{\mathrm{d}^2y}{\mathrm{d}t^2}+2\dfrac{\mathrm{d}y}{\mathrm{d}t}+2y=0$ for $t\in[0,1]$ with $y(0)=0$ and $y(1)=1$. Compare your solution with the expansion obtained using Matched asymptotic expansions (first example of section III.1).
(b) $\varepsilon\dfrac{\mathrm{d}^2y}{\mathrm{d}t^2}+\left(t-\dfrac{1}{2}\right)\dfrac{\mathrm{d}y}{\mathrm{d}t}+y=0$ for $t\in[0,1]$ with $y(0)=2$ and $y(1)=3$.

(9) Consider the problem

$$\frac{\mathrm{d}^2y}{\mathrm{d}t^2}+k^2(\varepsilon t)y=0,$$

for $t\geq 0$ and with $y(0)=a$ and $\dfrac{\mathrm{d}y}{\mathrm{d}t}(0)=b$ and where $k(\cdot)$ is a smooth function. Make the change of coordinates $\tau=\varepsilon t$ and then use the WKB method to construct a first-term approximation of the solution.

(10) A common approach to use the WKB method (when it is not clear that the solution has an exponential behavior) is to perform the change of variables $y(t)=\exp(w(t))$. Consider the following problem

$$\varepsilon^2\frac{\mathrm{d}^2y}{\mathrm{d}t^2}-q(t,\varepsilon)y=0.$$

(a) What equation must be satisfied by $w(t)$?
(b) Suppose that $q(t,\varepsilon)\sim q_0(t)+\varepsilon q_1(t)$, where $q_0\neq 0$. Propose that $w\sim\varepsilon^{-\alpha}(w_0(t)+\varepsilon^\beta w_1(t)+\cdots)$ and find the first two terms in the expansion of w. Afterwards, find the resulting expansion for y.
(c) Suppose that $q(t,\varepsilon)\sim\varepsilon q_0(t)+\varepsilon^2 q_1(t)$, where $q_0\neq 0$. Find the first two terms in the expansion of w, and then determine the resulting expansion for y

(11) For the next two problems, find the first-term expansion of the solutions that is valid for large t. By this we mean that a time-scaling should be introduced to remove the first secular term appearing in a regular expansion.
(a) $\dfrac{\mathrm{d}^2y}{\mathrm{d}t^2}+\varepsilon\left(\dfrac{\mathrm{d}y}{\mathrm{d}t}\right)^3y=0$ for $t\in[0,1]$ with $y(0)=0$ and $\dfrac{\mathrm{d}y}{\mathrm{d}t}(0)=1$.
(b) $\varepsilon\dfrac{\mathrm{d}^2y}{\mathrm{d}t^2}+\dfrac{\mathrm{d}y}{\mathrm{d}t}+y=0$ for $t\in[0,1]$ with $y(0)=0$ and $\dfrac{\mathrm{d}y}{\mathrm{d}t}(0)=1$.

(12) The equation for an oscillator can be written as

$$\frac{\mathrm{d}^2y}{\mathrm{d}t^2}+\frac{\mathrm{d}V(y)}{\mathrm{d}y}=0,$$

for $t>0$ and where $V(y)$ is a potential function. Consider initial conditions $y(0)=\varepsilon$ and $\dfrac{\mathrm{d}y}{\mathrm{d}t}(0)=0$. Find the first-term expansion of the solution that is valid for large t for the following potential functions:

(a) $V(y)=-\cos(y)$ (the classical pendulum).

(b) $V(y) = (1 - \exp(-\alpha y))^2$ with $\alpha > 0$ (the Morse oscillator).

(c) $V(y) = \exp(y) - y$ (the Toda oscillator).

(d) $V(y) = (1 + y)^{-12} - (1 + y)^{-6}$ (the Lennard-Jones oscillator).

CHAPTER IV

Normal Forms

In this chapter we study an important aspect of perturbation theory. Namely, we now study what is known as *normal forms*. Briefly speaking, a normal form (up to some notion of equivalence) is the "simplest" representative of a certain class of problems, in our case, ODEs. A normal form is obtained after successive changes of coordinates. As we will see, in some sense we attempt to simplify, as much as possible, the Taylor expansion of a vector field near, say an equilibrium point.

Before going into further details, let us see a first useful normal form.

THEOREM IV.1 (Flow-box theorem). *Consider a smooth nonlinear system given by* $\dfrac{\mathrm{d}x}{\mathrm{d}t} = f(x)$, $x \in \mathbb{R}^n$, *and assume that* $f(x^*) \neq 0$. *Then, there exists a neighborhood* U *of* x^* *with local coordinates* $y = (y_1, \dots, y_n)$ *such that in this coordinates the original system has the form*

$$\begin{aligned} \frac{\mathrm{d}y_1}{\mathrm{d}t} &= 1 \\ \frac{\mathrm{d}y_2}{\mathrm{d}t} &= 0 \\ &\vdots \\ \frac{\mathrm{d}y_n}{\mathrm{d}t} &= 0. \end{aligned}$$

PROOF. A proof can be found in [22], Chapter 5, Theorem 7. □

The previous theorem can also be regarded as a "straightening" or "rectification" of the vector field, see [2]. For the rest of this chapter, we are mostly interested in normal forms near equilibrium points, and especially of singularities. Let us first describe the reason.

Assume we deal with a certain vector field

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= Ax + f(x, y) \\ \frac{\mathrm{d}y}{\mathrm{d}t} &= By + g(x, y), \end{aligned} \tag{55}$$

where $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, the matrices A and B are of appropriate dimensions, and the functions f and g vanish together with their derivatives at the origin. Moreover, let us assume that A has eigenvalues with zero real parts while B is hyperbolic. In fact, without loss of generality, let us simply assume that B has eigenvalues with strictly negative part (thus the center manifold is attracting). Since the center manifold W^c is tangent to the center eigenspace, i.e. to $\{y = 0\}$, we can express W^c as a graph

$$W^c = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m \,|\, y = h(x)\},$$

with $h(0) = 0$ and $\dfrac{\partial h}{\partial x}(0) = 0$. Because the center manifold is invariant, the (local) flow restricted to W^c is therefore governed by

$$\frac{\mathrm{d}x}{\mathrm{d}t} = Ax + f(x, h(x)). \tag{56}$$

Knowledge about the behavior of (56) together with the fact that W^c is attracting, gives us sufficiently good information of about the dynamics of (55). Thus, indeed it suffices to restrict our attention to systems of the form (56).

Next, given (56) we proceed to obtain a simpler representative by performing "near identity" transformations $\tilde{x} = x + \psi(x)$ with the goal of eliminating as much as possible, the nonlinear terms appearing in $f(x, h(x))$.

Although our main interest is on bifurcations, let us first exemplify the normal form procedure by answering the question: when can a nonlinear system be reduced to its linear part?

Let us start with a system of the form

$$\frac{\mathrm{d}x}{\mathrm{d}t} = Ax + \cdots, \tag{57}$$

where A has distinct eigenvalues.

DEFINITION IV.1. The spectrum of A, $\operatorname{spec} A = \{\lambda_1, \ldots, \lambda_n\}$, is said to be resonant if

$$\lambda_k = < m, \lambda >, \tag{58}$$

where $m = (m_1, \ldots, m_n) \in \mathbb{N}^n$, $m_i \geq 0$ and $\sum_{i=1}^{n} m_i \geq 2$, holds for some $\lambda_k \in \operatorname{spec}(A)$ and $\lambda = (\lambda_1, \ldots, \lambda_n)$. The number $|m| = \sum_{i=1}^{n} m_i$ is called "the order of the resonance".

EXERCISE IV.1. *The Hopf bifurcation is characterized by a pair of conjugated eigenvalues $\pm\imath\lambda$ at the bifurcation point. Is this pair resonant?*

We now have a fundamental result due to Poincaré.

THEOREM IV.2. *If the eigenvalues of the matrix A are nonresonant, then the nonlinear system (57) can be reduced to the linear equation $\frac{\mathrm{d}y}{\mathrm{d}t} = Ay$ by a near identity (formal[1]) change of coordinates $x = y + \mathcal{O}(||y||^2)$.*

The proof of this theorem can be consulted in [4]. Instead of repeating here the proof, let us provide some intuition as of how the proof, and the normal form computation works.

Let us consider the linear system

$$\frac{\mathrm{d}y}{\mathrm{d}t} = Ay,$$

[1]Given as power series.

and consider a near identity transformation $x = y + h(y)$, where $h(y) = \sum_{k\geq 2} h_k(y)$ with $h_k(y)$ a vector with entries being a homogeneous polynomial of degree k. Under such a transformation we obtain:

$$\begin{aligned}\frac{\mathrm{d}x}{\mathrm{d}t} &= \frac{\mathrm{d}y}{\mathrm{d}t} + \frac{\partial h}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t}\\ &= \left(I + \frac{\partial h}{\partial y}\right) A(x-h)\\ &= \left(I + \frac{\partial h}{\partial x} + \cdots\right) A(x-h)\\ &= Ax + \left[\frac{\partial h}{\partial x}Ax - Ah\right] + \cdots\\ &= Ax + \sum_{k\geq 2}\left[\frac{\partial h_k}{\partial x}Ax - Ah_k\right] + \cdots .\end{aligned}$$

Notice that for each k the term $[Ax, h_k(x)] = \dfrac{\partial h_k}{\partial x}Ax - Ah_k$ is of degree k. The important observation is that if we can match $[Ax, h_k(x)]$ to the k-th higher order terms of $\dot{x} = Ax + f(x)$, then there exists a transformation that eliminates such a term (just go backwards in the above line of thought). It turns out that *if A is non-resonant* then the so called *homological equation*

$$[Ax, h] = f,$$

where $[\cdot,\cdot]$ denotes the Lie bracket, has a solution for any vector field f of degree at least 2.

In the case of resonance (58), the homogical equation cannot be solved for those monomials x^m (using multi-index notation). Such monomials are called *resonant*.

THEOREM IV.3 (Poincaré-Dulac). *If the eigenvalues of the matrix A are resonant, then the non-linear system (57) can be reduced to the linear equation $\dfrac{\mathrm{d}y}{\mathrm{d}t} = Ay + g(y)$ by a near identity (formal) change of coordinates $x = y + \mathcal{O}(||y||^2)$, where all the monomials of $g(y)$ are resonant. More precisely, for every non resonance $\lambda_k =< m, \lambda >$, the nonlinear monomial remaining in the normal form is $x^m \hat{e}_k$.*

For the proof see [4].

EXAMPLE IV.1. Consider the system

$$\begin{aligned}\frac{\mathrm{d}x_1}{\mathrm{d}t} &= \lambda x_1 + \cdots\\ \frac{\mathrm{d}x_2}{\mathrm{d}t} &= k\lambda x_2 + \cdots ,\end{aligned} \tag{59}$$

for some $k \in \mathbb{N}$ and $k > 1$. In this case the resonant monomial is x_1^k and therefore we can reduce (59) to

$$\begin{aligned}\frac{\mathrm{d}x_1}{\mathrm{d}t} &= \lambda x_1\\ \frac{\mathrm{d}x_2}{\mathrm{d}t} &= k\lambda x_2 + x_1^k.\end{aligned}$$

EXERCISE IV.2. *Perform the computations to verify this example.*

EXERCISE IV.3. *Is a matrix with at least one eigenvalue with zero real part resonant?*

EXERCISE IV.4. *Obtain the normal form, modulo formal change of coordinates, of the system*

$$\frac{\mathrm{d}x_1}{\mathrm{d}t} = x_1 + \cdots$$
$$\frac{\mathrm{d}x_2}{\mathrm{d}t} = -x_2 + \cdots .$$

In the case one is interested in a linearization about a hyperbolic equilibrium point, there are stronger results, of which we mention the following two.

THEOREM IV.4 (Sternberg). *Let X and Y be $\mathcal{C}^\infty$ vector fields on $\mathbb{R}^n$ with 0 as a hyperbolic equilibrium point. Suppose that there exists a formal transformation taking the Taylor series of X at 0 to that of Y. Then there exists a $\mathcal{C}^\infty$-diffeomorphism transforming X to Y.*

The above theorem is telling us that "two vector fields that are formally equivalent about a hyperbolic equilibrium point are smoothly equivalent". Furthermore, suppose X is a vector field whose linear part at 0 has no resonances. Then, one can conclude that X is smoothly linearizable. The following theorem tells us that the same conclusion is possible under resonance conditions, as long as the transformation is only $\mathcal{C}^0$.

THEOREM IV.5 (Hartman-Grobman). *Let X be a $\mathcal{C}^\infty$ vector field with 0 being a hyperbolic equilibrium point. Then, X is (near 0) topologically equivalent to it linear part.*

So far we have performed what is known as "formal normal form". That is we have performed a formal change of coordinates by which we mean that the transformation is given as a power series. Moreover, we have not addressed the question of whether such a transformation is convergent. The question whether the normal form is analytic of $\mathcal{C}^\infty$ is more subtle, and shall not be discussed in this course.

Let us return to the nonlinear system

$$\frac{\mathrm{d}x}{\mathrm{d}t} = f(x). \tag{60}$$

Here, parameters may be included trivially. Let $A = Df(0)x$ denote the linear part of (60). Let $\mathcal{H}_k$ denote the space of homogeneous polynomial vector fields of degree k. We define *the adjoint map induced by* A as $\mathrm{ad}_A : \mathcal{H}_k \to \mathcal{H}_k$ given by

$$\mathrm{ad}_A(h) = [Ax, h].$$

Now we state the main result regarding normal forms.

THEOREM IV.6. *Consider (60) with $f(0) = 0$ and $Df(0) = A$ such that f is of class $\mathcal{C}^r$. Choose a complement $\mathcal{G}_k$ for $\mathrm{ad}_A(\mathcal{H}_k)$ in $\mathcal{H}_k$ so that $\mathcal{H}_k = \mathrm{ad}_A(\mathcal{H}_k) + \mathcal{G}_k$. Then, there is a formal near identity change of coordinates that transforms (60) into $\dfrac{\mathrm{d}y}{\mathrm{d}t} = Ay + \displaystyle\sum_{i=1}^{r} g_i(y) + R_r$, where $g_i \in \mathcal{G}_i$ and $R_r = c(|y|^r)$.*

For a proof you can see [**10**].

EXAMPLE IV.2 (The focus singularity). For the normal form of the focus singularity we have a linearization given by the matrix

$$A = \begin{bmatrix} 0 & -\omega_0 \\ \omega_0 & 0 \end{bmatrix}.$$

where the eigenvalues are $\lambda_{1,2} = \pm\omega_0 \imath$, $\omega_0 \neq 0$.
We will see in the next chapter that we can look at the complexified version, namely

$$\frac{\mathrm{d}z}{\mathrm{d}t} = \imath\omega_0 z + \cdots$$
$$\frac{\mathrm{d}\bar{z}}{\mathrm{d}t} = -\imath\omega_0 \bar{z} + \cdots$$

It suffices to look at the equation for z, since the equation for $\bar{z}$ is obtained by conjugation.
In this case we have a resonance condition of the form $\lambda_1 + \lambda_2 = 0$, or equivalently $(m_1 - m_2) = 1$ with $m_1 + m_2 \geq 2$. This means that the resonant monomials are of the form $z^{m_1}\bar{z}^{m_1-1} = |z|^{2(m_1-1)}z$, and so the corresponding normal form reads as

$$\frac{\mathrm{d}z}{\mathrm{d}t} = \imath\omega_0 z + c|z|^2 z + \cdots, \tag{61}$$

where c is some constant (possibly complex).
Although the above computation was simple and straightforward, let us exemplify the normal form procedure if one would not go to complex coordinates. For clarity, we are now considering the system

$$\frac{\mathrm{d}x_1}{\mathrm{d}t} = -x_2 + f_1(x_1, x_2)$$
$$\frac{\mathrm{d}x_2}{\mathrm{d}t} = x_1 + f_2(x_1, x_2),$$

where for simplicity we set $\omega_0 = 1$.
Let us first look at the action of ad_A on $\mathcal{H}_2$. The space $\mathcal{H}_2$ is generated by the basis

$$\left\{ x_1^2\frac{\partial}{\partial x_1}, x_1x_2\frac{\partial}{\partial x_1}, x_2^2\frac{\partial}{\partial x_1}, x_1^2\frac{\partial}{\partial x_2}, x_1x_2\frac{\partial}{\partial x_2}, x_2^2\frac{\partial}{\partial x_2} \right\} \tag{62}$$

Therefore, it is enough to look at the action of ad_A into any of such elements. Let h_i be any of such base monomials. Then

$$\begin{aligned} \mathrm{ad}_A(h_i) &= [Ax, h_i] \\ &= \begin{bmatrix} \frac{\partial h_i^1}{\partial x_1} & \frac{\partial h_i^1}{\partial x_2} \\ \frac{\partial h_i^2}{\partial x_1} & \frac{\partial h_i^2}{\partial x_2} \end{bmatrix} \begin{bmatrix} -x_2 \\ x_1 \end{bmatrix} - \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} h_i^1 \\ h_i^2 \end{bmatrix} \end{aligned}$$

and, for example, $\mathrm{ad}_A\left(x_1^2\frac{\partial}{\partial x_1}\right) = \left(-2x_1x_2\frac{\partial}{\partial x_1}, -x_1^2\frac{\partial}{\partial x_2}\right)$. Since $\mathcal{H}_k$ is a linear space, and $\mathrm{ad}_A : \mathcal{H}_k \to \mathcal{H}_k$, we can represent the action of ad_A on $\mathcal{H}_k$ by a matrix. For the case of $\mathcal{H}_2$, the

corresponding 6×6 matrix representing the action of ad_A on $\mathcal{H}_2$ is

$$\begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 0 \\ -2 & 0 & 2 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & -2 & 0 & 2 \\ 0 & 0 & -1 & 0 & -1 & 0 \end{pmatrix}.$$

In this matrix the i-th column corresponds to the coefficients of the image of ad_A for the basis ordered as in (62). For example, $x_1^2 \frac{\partial}{\partial x_1} \sim (1,0,0,0,0,0)^\top$, and so $\mathrm{ad}_A(x_1^2) \sim (0,-2,0,-1,0,0)^\top \sim -2x_1x_2\frac{\partial}{\partial x_1} - x_1^2\frac{\partial}{\partial x_2}$. Similarly, $x_1x_2\frac{\partial}{\partial x_2} \sim (0,0,0,0,1,0)^\top$ and therefore $\mathrm{ad}_A\left(x_1x_2\frac{\partial}{\partial x_2}\right) = x_1x_2\frac{\partial}{\partial x_1} + (x_1^2 - x_2^2)\frac{\partial}{\partial x_2}$.

The above representation matrix is invertible. This implies (or rather verifies) that the homological equation can be solved for any quadratic vector field. Thus, every quadratic term of (60) can be removed.

Next, we look at the action of ad_A on $\mathcal{H}_3$, which has basis

$$\left\{ x_1^3\frac{\partial}{\partial x_1}, x_1^2x_2\frac{\partial}{\partial x_1}, x_1x_2^2\frac{\partial}{\partial x_1}, x_2^3\frac{\partial}{\partial x_1}, x_1^3\frac{\partial}{\partial x_2}, x_1^2x_2\frac{\partial}{\partial x_2}, x_1x_2^2\frac{\partial}{\partial x_2}, x_2^3\frac{\partial}{\partial x_2} \right\}.$$

That is, $\mathcal{H}_3$ has dimension 8, and the matrix representing the action of ad_A on $\mathcal{H}_3$ reads as

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ -3 & 0 & 2 & 0 & 0 & 1 & 0 & 0 \\ 0 & -2 & 0 & 3 & 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & -3 & 0 & 2 & 0 \\ 0 & 0 & -1 & 0 & 0 & -2 & 0 & 3 \\ 0 & 0 & 0 & -1 & 0 & 0 & -1 & 0 \end{pmatrix}$$

One can check that the kernel of the above matrix is spanned by

$$\left\{ \begin{pmatrix} -1 \\ 0 \\ -1 \\ 0 \\ 0 \\ -1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \\ -1 \\ 0 \\ -1 \\ 0 \end{pmatrix} \right\}.$$

This means that a complement of $\mathrm{ad}_A(\mathcal{H}_3)$ can be chosen to be given by the vector fields

$$-x_1(x_1^2 + x_2^2)\frac{\partial}{\partial x_1} - x_2(x_1^2 + x_2^2)\frac{\partial}{\partial x_2},$$

and

$$x_2(x_1^2+x_2^2)\frac{\partial}{\partial x_1}-x_1(x_1^2+x_2^2)\frac{\partial}{\partial x_2}.$$

Therefore, applying the normal form theorem IV.6 we conclude that there exists a coordinate transformation which transform asystem of the form

$$\begin{aligned}\frac{\mathrm{d}x_1}{\mathrm{d}t}&=-x_2+\cdots\\ \frac{\mathrm{d}x_2}{\mathrm{d}t}&=x_1+\cdots\end{aligned}$$

into the system

$$\begin{aligned}\frac{\mathrm{d}u}{\mathrm{d}t}&=-v+(-au+bv)(u^2+v^2)\\ \frac{\mathrm{d}v}{\mathrm{d}t}&=u-(av+bu)(u^2+v^2),\end{aligned}\tag{63}$$

for some constants $a\in\mathbb{R}$ and $b\in\mathbb{R}$.

EXERCISE IV.5. *Corroborate that the normal forms (61) and (63) are indeed equivalent.*

REMARK IV.1. We emphasize that the choice of the complementary space of $\mathrm{ad}_A(\mathcal{H}_k)$ is far from unique. Some choices may be more convenient than others. The choice we did for the above example is rather convenient when we want to rewrite the system in polar coordinates.

We finish this chapter by addressing the question of how to obtain normal forms of a bifurcation. In the case we are interested in the family

$$\frac{\mathrm{d}x}{\mathrm{d}t}=F(x,\mu),$$

such that $F(x,0)=f(x)$, we can simply consider the extended system

$$\begin{aligned}\frac{\mathrm{d}x}{\mathrm{d}t}&=F(x,\mu)\\ \frac{\mathrm{d}\mu}{\mathrm{d}t}&=0.\end{aligned}$$

One can proceed as above with the normal form calculations by taking now a coordinate change $x=y+H(x,\mu)$ where $H(x,\mu)=(h(x,\mu),\mu)$, which evidently leaves the equation $\frac{\mathrm{d}\mu}{\mathrm{d}t}=0$ invariant. In practice, what changes now is that we assume that the coefficients of the power series of h depend on μ. Families of vector fields as above are relevant when studying bifurcations.

EXAMPLE IV.3 (Saddle-node bifurcation). Consider the family of scalar systems

$$\frac{\mathrm{d}x}{\mathrm{d}t}=F(x,\mu),\qquad x\in\mathbb{R},\ \mu\in\mathbb{R},$$

where $F(0,0)=0$, $\frac{\partial F}{\partial x}(0,0)=0$ and $\frac{\partial F}{\partial \mu}\neq 0$.
We consider the extended system

$$\begin{aligned}\frac{\mathrm{d}x}{\mathrm{d}t}&=F(x,\mu)\\ \frac{\mathrm{d}\mu}{\mathrm{d}t}&=0.\end{aligned}$$

The linearization at the origin is given by the matrix

$$A = \begin{bmatrix} 0 & a \\ 0 & 0 \end{bmatrix},$$

where $a = \dfrac{\partial F}{\partial \mu}(0,0) \neq 0$.

We now consider the action of ad_A on $\tilde{\mathcal{H}}_2 = \mathrm{span}\left\{x^2\dfrac{\partial}{\partial x}, \mu x\dfrac{\partial}{\partial x}, \mu^2\dfrac{\partial}{\partial x}, 0, 0, 0\right\}$, because we are assuming that the change of coordinates leaves $\dfrac{\mathrm{d}\mu}{\mathrm{d}t}$ invariant. Noticing that

$$\mathrm{ad}_A\left(x^2\frac{\partial}{\partial x}\right) = 2a\mu x\frac{\partial}{\partial x}$$
$$\mathrm{ad}_A\left(\mu x\frac{\partial}{\partial x}\right) = a\mu^2\frac{\partial}{\partial x}$$
$$\mathrm{ad}_A\left(\mu^2\frac{\partial}{\partial x}\right) = 0,$$

we find that a complement of ad_A on $\tilde{\mathcal{H}}_2$ is given by $x^2\dfrac{\partial}{\partial x}$. Thus, using the normal form theorem IV.6 we conclude that we can transform the initial system into the normal form

$$\frac{\mathrm{d}x}{\mathrm{d}t} = a\mu - x^2 + \cdots,$$

where the choice of sign in front of x^2 is just for convenience. This equation is know as "the normal form of the saddle-node bifurcation". To provide further details, let $a = 1$. In this case the leading part of the normal form reads as

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \mu - x^2. \tag{64}$$

We notice therefore that if $\mu > 0$, then there are two hyperbolic equilibria $x^* = \pm\sqrt{\mu}$. For $\mu = 0$ such equilibria collide, while for $\mu < 0$ there are no equilibria anymore. The previous behavior, depending on the parameter μ, can be captured in a bifurcation diagram as shown in Figure .

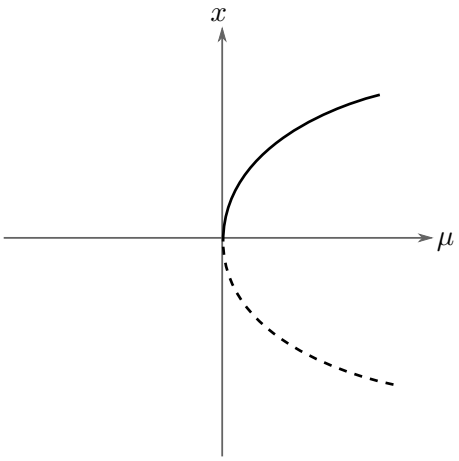


FIGURE 1. Bifurcation diagram of the saddle-node bifurcation (64). The parabola corresponds to the equilibria given by $x^2 = \mu$. For $\mu > 0$ the equilibrium point $x^* = \sqrt{\mu}$ is attracting while the equilibrium point $x^* = -\sqrt{\mu}$ is repelling (solid and dashed curves respectively).

IV.1. Further exercises for this chapter

(1) Compute the next non-zero term in the normal form of the saddle-node bifurcation. Does including such a term in the normal form change the local behavior (sufficiently near the origin) of the system?

(2) Consider the family of scalar systems

$$\frac{\mathrm{d}x}{\mathrm{d}t} = F(x,\mu), \qquad x \in \mathbb{R},\ \mu \in \mathbb{R},$$

where $F(0,0) = 0$, $\dfrac{\partial F}{\partial x}(0,0) = 0$, $\dfrac{\partial F}{\partial \mu} = 0$ and $\dfrac{\partial^2 F}{\partial \mu^2} \neq 0$. Obtain a normal form up to quadratic terms. What type of bifurcation is this? Draw the corresponding bifurcation diagram and the phase portrait. What happens if the "non-degeneracy condition" $\dfrac{\partial^2 F}{\partial \mu^2} \neq 0$ is replaced for $\dfrac{\partial^2 F}{\partial \mu \partial x} \neq 0$?

(3) Consider the family of scalar systems

$$\frac{\mathrm{d}x}{\mathrm{d}t} = F(x,\mu), \qquad x \in \mathbb{R},\ \mu \in \mathbb{R},$$

where $F(-x,\mu) = -F(x,\mu)$, $F(0,0) = 0$, $\dfrac{\partial F}{\partial x}(0,0) = 0$, $\dfrac{\partial F}{\partial \mu} = 0$ and $\dfrac{\partial^2 F}{\partial \mu \partial x} \neq 0$. Notice that now the condition $\dfrac{\partial^2 F}{\partial x^2}(0,0) = 0$ *must* be imposed, and thus one may assume $\dfrac{\partial^3 F}{\partial x^3}(0,0) \neq 0$. Explain the reason.

Obtain a normal form up to cubic terms. What type of bifurcation is this? Draw the corresponding bifurcation diagram and the phase portrait.

(4) Consider the family of planar systems

$$\frac{\mathrm{d}x}{\mathrm{d}t} = F(x,\mu), \qquad x \in \mathbb{R}^2,\ \mu \in \mathbb{R},$$

such that the system at $\mu = 0$ is given by (63). Obtain a normal form (up to degree 3). What type of bifurcation is this? Draw the corresponding bifurcation diagram and the phase portraits for topologically different members of the family.

Hint: notice that $\dfrac{\partial F}{\partial x}(0) \neq 0 \in \mathbb{R}^2$.

CHAPTER V

Singularities and Bifurcations for planar systems

In this chapter we restrict ourselves to autonomous planar systems

$$\frac{\mathrm{d}x}{\mathrm{d}t} = f(x), \tag{65}$$

where $x \in \mathbb{R}^2$ and f is a sufficiently smooth (at least twice differentiable) vector field. Some singularities and bifurcations for scalar systems are discussed in the next chapter.

Remark V.1. For basic background and terminology on dynamical systems and ODEs, refer to Appendices A.2 and A.3.

Recall that x^* is an equilibrium of (65) if $f(x^*) = 0$.

Definition V.1. We say that an equilibrium point x^* of (65) is:

- *Hyperbolic* if the eigenvalues of $D_x f(x^*)$ have nonzero real parts.
- *Non-hyperbolic* if at least one of the eigenvalues of $D_x f(x^*)$ have zero real parts.
- *Elliptic* if the eigenvalues of of $D_x f(x^*)$ are purely imaginary with nonzero imaginary part.
- *Nilpotent* if both eigenvalues of $D_x f(x^*)$ are exactly zero.

The classification of equilibrium points is important because it provides qualitative information of the asymptotic behavior of the orbits of (65) near the equilibrium point. This is better understood by linearizing (65) near x^*. Indeed, using Taylor's formula, it follows that:

$$\frac{\mathrm{d}y}{\mathrm{d}t} = D_x f(x^*)y + \mathcal{O}(||y||^2), \tag{66}$$

where $y = x - x^*$ and $A := D_x f(x^*)$ is known as the *Jacobian*. The solution of the linear part of (66) is given by

$$y(t) = \exp(A(t - t_0))y(t_0). \tag{67}$$

Now, from (67) it is evident that knowledge of the eigenvalues of the Jacobian matrix suffices to know the solution $y(t)$ which is a valid approximation of the solution of (65) close enough to the equilibrium x^*. For example:

Proposition V.1. *Consider (65), let x^* be an isolated equilibrium point, and denote by $\lambda_{1,2}$ the eigenvalues of $D_x f(x^*)$. Assume that $\Re(\lambda_1) < 0$ and $\Re(\lambda_2) < 0$. Then, there exists a neighborhood U of x^* such that the ω-limit set of any point $x \in U$ is x^*, that is $\omega(x) = x^*$ $\forall x \in U$.*

Proof. Let us first consider that the eigenvalues are real (negative) and $\lambda_1 \neq \lambda_2$. Then, because the Jacobian can be diagonalized, the linearized system reads as

$$\frac{\mathrm{d}y_1}{\mathrm{d}t} = \lambda_1 y_1 + h_1(y)$$
$$\frac{\mathrm{d}y_2}{\mathrm{d}t} = \lambda_2 y_2 + h_2(y),$$

where $|h_i(y)| \leq M||y||^2$, $i = 1, 2$, for sufficiently small $||y||$ and $M > 0$. Let us define the Lyapunov function

$$V(t) = \frac{1}{2}\left(y_1(t)^2 + y_2(t)^2\right).$$

Notice that $V = 0$ if and only if $y = 0$ and is positive otherwise. Next, differentiating with respect to time:

$$\begin{aligned}
\frac{\mathrm{d}V(t)}{\mathrm{d}t} &= y_1\frac{\mathrm{d}y_1}{\mathrm{d}t} + y_2\frac{\mathrm{d}y_2}{\mathrm{d}t} \\
&= y_1(\lambda_1 y_1 + h_1(y)) + y_2(\lambda_2 y_2 + h_2(y)) \\
&= \lambda_1 y_1^2 + \lambda_2 y_2^2 + y_1 h_1(y) + y_2 h_2(y) \\
&\leq \lambda_1 y_1^2 + \lambda_2 y_2^2 + (y_1 + y_2)M(y_1^2 + y_2^2) \\
&= 2(\lambda_1 + \lambda_2 + M(y_1 + y_2))V(t) \\
&\leq 2(\lambda_1 + \lambda_2 + M(|y_1| + |y_2|))V(t).
\end{aligned}$$

Thus, there is a sufficiently small constant $r > 0$ such that the term $\lambda_1 + \lambda_2 + M(|y_1| + |y_2|)$ is strictly negative for all $0 \leq ||y|| \leq r$. This means that the disc $B_r = \left\{y \in \mathbb{R}^2 : ||y|| \leq r\right\}$ is positively (or forward) invariant. Thus $\frac{\mathrm{d}V(t)}{\mathrm{d}t}|_{B_r} = 0$ if and only if $y = 0$. Therefore, every trajectory with initial condition within B_r converges towards the origin $y = 0$. □

Exercise V.1.

- *Complete the above proof by relaxing the condition of real eigenvalues.*
- *Show that if $\Re(\lambda_1) > 0$ and $\Re(\lambda_2) > 0$, then there exists a neighborhood U of x^* such that $\alpha(x) = x^*$ $\forall x \in U$.*

For the saddle case, we have the following important theorem generalizing the situation in linear case (recall theorem A.7).

Theorem V.1 (Stable Manifold Theorem). *Consider (65), let x^* be an isolated equilibrium point, and denote by $\lambda_{1,2}$ the eigenvalues of $D_x f(x^*)$. Assume that the eigenvalues are real with $\lambda_1 < 0 < \lambda_2$. Then*

(1) there exists a curve $W^s(x^*)$ tangent at x^* to E_1 (the eigenspace of λ_1), such that $\omega(x) = x^*$ for all $x \in W^s(x^*)$. We call $W^s(x^*)$ the stable manifold *of x^*.*

(2) there exists a curve $W^u(x^*)$ tangent at x^* to E_2 (the eigenspace of λ_2), such that $\alpha(x) = x^*$ for all $x \in W^u(x^*)$. We call $W^u(x^*)$ the unstable manifold *of x^*.*

To summarize this section, we emphasize that there are three qualitatively different hyperbolic equilibria: *sinks* that correspond to points attracting every nearby solution, *sources* that correspond to points repelling every nearby solution, and *saddles* that attract and repel solutions. See figure 1.

Exercise V.2. *Prove that a hyperbolic equilibrium point remains hyperbolic under sufficiently small smooth perturbations.*

Hint: consider a system given by $\frac{\mathrm{d}x}{\mathrm{d}t} = Dx + \varepsilon h(x)$, where D is a diagonal matrix with nonzero diagonal elements and h is smooth and satisfies $h(0) = 0$. Assume that the eigenvalues of the perturbed problem depend smoothly on ε for ε sufficiently small (which is true indeed [20]).

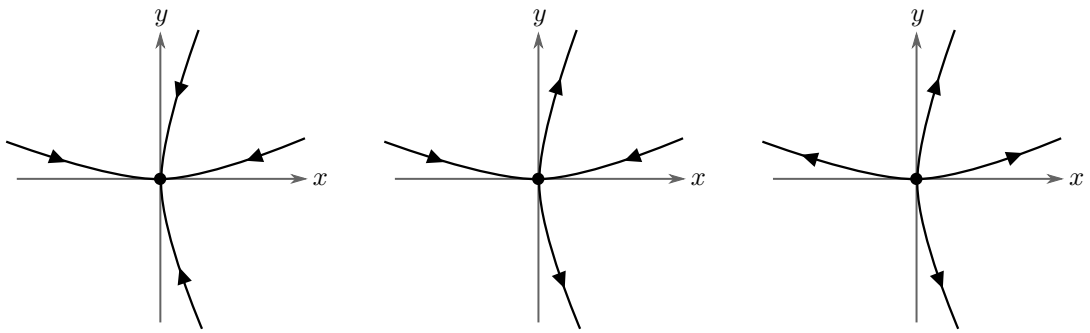


FIGURE 1. Sketches of the 3 different types of equilibria. Left: a sink, for which the stable manifold is the whole plane and the unstable manifold is empty. Middle: a saddle, for which the stable and unstable manifolds are 1-dimensional each. Right: a source, for which the stable manifold is empty, and the unstable manifold is the whole plane.

Note: this is not the most general case (which is more cumbersome to prove), but it does give a general idea of what one means by "persistence" under small perturbations.

We now turn our attention to periodic orbits.

DEFINITION V.2. A *periodic orbit* is an orbit that forms a closed curve, denoted by Γ, in a region $D \subset \mathbb{R}^2$ of the phase-space. Alternatively, if $x_0 \in D$ is not an equilibrium point, and there is a $T > 0$ such that $\Phi_T(x_0) = x_0$, then the orbit through x_0 is a periodic orbit with period T. We call a T such that $\Phi_t(x) \neq x_0$ for all $0 < t < T$ (and $\Phi_T(x) = x_0$) the *least or minimal period*.

A trajectory contained in Γ shall be denoted by $\gamma = \gamma(t)$. To analyze the dynamics near a periodic orbit, let us perform the change of coordinates $y = x - \gamma$. Then (65) is rewritten as:

$$\frac{\mathrm{d}y}{\mathrm{d}t} = f(\gamma + y) - f(\gamma). \tag{68}$$

The linearization of (68) at $y \sim 0$ is given by

$$\frac{\mathrm{d}y}{\mathrm{d}t} = A(t)y, \tag{69}$$

where $A(t) = \dfrac{\partial f}{\partial x}(\gamma(t))$. Notice that $A(t)$ is periodic with minimal period T, that is $A(t+T) = A(t)$. Solutions of (69) can be written as $y(t) = M(t)y(0)$, where $M(t)$ is a fundamental matrix[1]. It follows from Floquet's theory[2] that any such $M(t)$ can be written as

$$M(t) = P(t)\exp(Kt),$$

where $P(t+T) = P(t)$, with $P(0) = I$ (the identity matrix), and K is a constant matrix. This implies that the asymptotic behavior of the solutions of (69) depends only on the eigenvalues of the constant matrix K, which are called *the characteristic exponents of* Γ. The matrix $M(T) = \exp(KT)$ is called the *monodromy matrix*, while its eigenvalues are called *characteristic multipliers*.

[1]Recall that a fundamental matrix of an ODE is formed by putting on the columns linearly independent solutions of the ODE.
[2]We will not discuss this theory during the course, but you are referred to [11].

PROPOSITION V.2. *The characteristic exponents of* Γ *are given by* 0 *and*

$$\frac{1}{T}\int_0^T \left(\frac{\partial f_1}{\partial x_1}+\frac{\partial f_2}{\partial x_2}\right)\gamma(t)\mathrm{d}t.$$

EXERCISE V.3. *Prove the previous proposition.*
Hints:

(1) Show that $\dot\gamma$ *satisfies* (69)*, that is, that it satisfies the equation* $\dfrac{\mathrm{d}\dot\gamma}{\mathrm{d}t}=A(t)\dot\gamma$.

(2) From the previous step, show that $\dot\gamma(0)=\exp(KT)\dot\gamma(0)$*. Notice that this means that* $\exp(KT)$ *has an eigenvalue equal to* 1*. Thus, show that the previous observation implies that* K *has an eigenvalue equal to* 0.

(3) Next, it is known that any fundamental matrix $M(t)$ *satisfies the equation*

$$\frac{\mathrm{d}}{\mathrm{d}t}\det M(t)=\mathrm{Tr}(A(t))\det M(t). \tag{70}$$

Do not prove (70)*, but use it (together with the fact that* $P(0)=I$*) to show that*

$$\det\exp(KT)=\det M(T)=\exp\left(\int_0^T \mathrm{Tr}(A(t))\mathrm{d}t\right).$$

(4) Finally, argue from: a) the determinant is the product of the eigenvalues, b) by taking the logarithm of $\det(\exp(KT))$*, c) by using the identity* $\log(\det M)=\mathrm{Tr}(\log M)$ *for any matrix* M*, and d) the definition of* $A(t)$*, that the result follows.*

REMARK V.2. In a sufficiently small neighborhood of Γ the non-zero characteristic exponent of Γ gives information on its stability: if it is positive/negative then Γ is locally repelling/attracting. In fact, we say that a periodic orbit Γ is hyperbolic if the characteristic exponent $\dfrac{1}{T}\displaystyle\int_0^T \left(\frac{\partial f_1}{x_1}+\frac{\partial f_2}{x_2}\right)\gamma(t)\mathrm{d}t$ is nonzero.

We will now present one of the fundamental results for planar vector fields. First we need another definition.

DEFINITION V.3. Let x_1^* and x_2^* be two saddles. Assume that there is a point $x_0\in W^u(x_1^*)\cap W^s(x_2^*)$ (that is the unstable manifold of x_1^* and the stable manifold of x_2^* intersect). Thus, the orbit through x_0, γ_{x_0} is contained in $W^u(x_1^*)\cap W^s(x_2^*)$. This implies that $\alpha(x)=x_1^*$ and $\omega(x)=x_2^*$ for any $x\in\gamma_{x_0}$. Such a connection is called heteroclinic if $x_1^*\neq x_2^*$ and homoclinic if $x_1^*=x_2^*$. When no distinction is necessary, we refer to both of them as "saddle connection". See figure 2.

We are ready now to provide a full classification of limit sets for vector fields in the plane:

THEOREM V.2 (Poincaré-Bendixson). *Let* D *be a compact positively invariant region containing a finite number of equilibrium points. For any* $x\in D$*, the* ω*-limit set* $\omega(x)$ *is one of the following:*

(1) an equilibrium point, or
(2) a periodic orbit, or
(3) a set consisting of a finite number of equilibrium points $x_1^*,\dots,x_k^*$ *and orbits* Γ_k *such that* $\alpha(\Gamma_k)=x_i^*$ *and* $\omega(\Gamma_k)=x_j$.

PROOF. A proof can be found in Chapter 16 of [7]. □

A very useful property related to limit sets is that the phase space may be arranged by regions sharing the same limit sets.

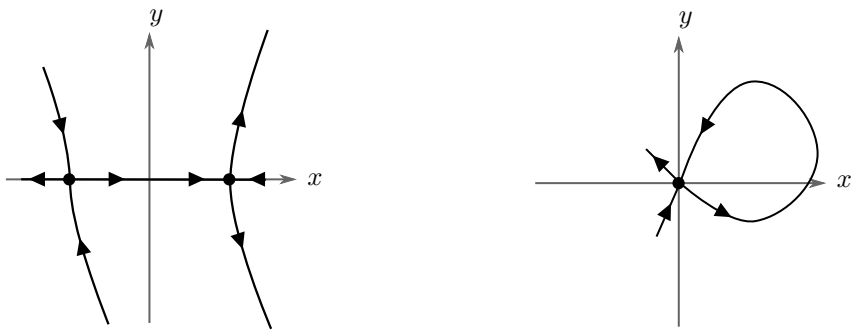


FIGURE 2. Sketch of the two different types of saddle connections. On the left a heteroclinic connection and on the right an homoclinic loop.

DEFINITION V.4. Let $\bar{\omega} = \omega(x)$ be the ω-limit set of a point $x \in D$. The *basin of attraction* of $\bar{\omega}$ is the set

$$\mathcal{A}(\bar{\omega}) = \{y \in D \,:\, \omega(y) = \bar{\omega}\}\,.$$

EXERCISE V.4. *Related to the previous definition, prove that if D is a positively invariant set containing $\bar{\omega}$ then the sets $\mathcal{A}(\bar{\omega}) \cap D$, $D \backslash \mathcal{A}(\bar{\omega})$ and $\partial\mathcal{A}(\bar{\omega}) \cap D$ are positively invariant. Make a sketch of each of such sets.*

At this moment, it is worth recalling that in this course we are interested on perturbations of a "well-known" unperturbed problem. Sometimes such perturbations change drastically the qualitative behavior of the problem, some other times, the perturbed and unperturbed problems are qualitatively the same. Thus, we now discuss an important class of vector fields.

DEFINITION V.5. Let $D \subset \mathbb{R}^2$ be a compact region. Denote by $\mathcal{X}^k(D)$ the set of all C^k vector fields on D that point inwards on the boundary of D (thus D is positively invariant under the flow generated by any $X \in \mathcal{X}^k(D)$).

- For a vector field $X = X_1(x_1, x_2)\dfrac{\partial}{\partial x_1} + X_2(x_1, x_2)\dfrac{\partial}{\partial x_2}$ the norm
$$||X||_{\mathcal{C}^k} = \sup_{x\in D} \max_{j=1,2} \max_{\substack{0\le p_1+p_2\le k \\ p_1,p_2\ge 0}} \left| \frac{\partial^{p_1+p_2} X_j}{\partial^{p_1} x_1 \partial^{p_2} x_2} \right|$$
is called *the $\mathcal{C}^k$-norm*. The resulting topology on $\mathcal{C}^k(D)$ is called *the $\mathcal{C}^k$-topology*.
- Let X and Y be two vector fields on $\mathcal{X}^k(D)$. We say that Y is a *$\mathcal{C}^k$-small perturbation* of X if $||Y - X||_{\mathcal{C}^k}$ is small for certain k.

The above definition allows us to formalize what one usually means by "a small perturbation". However classification under $\mathcal{C}^0$-smallness can be too coarse, while classification under $\mathcal{C}^1$-smallness can be too fine.

EXAMPLE V.1. Consider the scalar functions $f(x) = -x$ and $g(x) = -x + \varepsilon\sqrt{|x|}$, with $D = [-1, 1]$. Then

$$||g - f||_{\mathcal{C}^0} = \sup_{x\in D} \varepsilon\sqrt{|x|} = \varepsilon.$$

However, even if the vector fields generated by f and g are $\mathcal{C}^0$-close, they are not qualitatively similar: it suffices to notice that $\dot{x} = f(x)$ has one equilibrium point, while $\dot{x} = g(x)$ has two, for any $\varepsilon > 0$.

As we can see from the above digression, we are generally interested in comparing the qualitative behavior of two systems.

DEFINITION V.6. Two vector fields $X \in \mathcal{X}^k(D)$ and $Y \in \mathcal{X}^k(D)$ are said to be *topologically equivalent* (or $\mathcal{C}^0$-equivalent), if there is a homeomorphism[3] $h : D \to D$, that takes orbits of X onto orbits of Y by preserving the direction (but not necessarily the parametrization) of time. Equivalently, if we denote by ϕ_t and ψ_t the flows of X and Y respectively, the vector fields X and Y are topologically equivalent if there is a homeomorphism h and a monotonously increasing bijection $\tau : \mathbb{R} \to \mathbb{R}$ such that

$$\psi_{\tau(t)}(x) = h \circ \phi_t \circ h^{-1}(x),$$

for all $x \in D$ and $t \geq 0$.

REMARK V.3. One can also define $\mathcal{C}^k$-equivalence by requiring the map h in the previous definition to be a $\mathcal{C}^k$ diffeomorphism.

EXAMPLE V.2. Consider the one-dimensional systems $\dfrac{\mathrm{d}x}{\mathrm{d}t} = ax$ and $\dfrac{\mathrm{d}y}{\mathrm{d}t} = by$ with $0 > a > b$. The corresponding solutions are $x(t) = \exp(at)x(0)$ and $y(t) = \exp(bt)y(0)$. Consider the homeomorphism

$$h(x) = \begin{cases} x^{b/a}, & x > 0 \\ 0, & x = 0, \\ -|x|^{b/a}, & x < 0. \end{cases}$$

The map h serves as a homeomorphism making the two vector fields topologically equivalent.

EXERCISE V.5. *Consider the one-dimensional systems* $\dfrac{\mathrm{d}x}{\mathrm{d}t} = ax$ *and* $\dfrac{\mathrm{d}y}{\mathrm{d}t} = by$ *with* $a \neq b$ *nonzero constants. Is there any choice of* a, b *such that the two vector fields are* $\mathcal{C}^1$*-equivalent?*
Answer: no.

EXERCISE V.6. *Consider the linear system* $\dot{x} = \begin{bmatrix} 1 & 0 \\ 0 & 1+\varepsilon \end{bmatrix} x$*. Are the perturbed and unperturbed systems topologically equivalent? Are the perturbed and unperturbed systems differentiably (*$\mathcal{C}^k$ *with* $k \geq 1$*) equivalent?*
Answer: yes, no.

EXERCISE V.7. *Show that, up to topological equivalence, hyperbolic equilibrium points in the plane are classified in three categories: sinks, sources, and saddles. Hint: assume that the eigenvalues are real and simple, and show equivalence only of the linear parts.*

[3]We recall that a homeomorphism is a map that is continuous, bijective, and with continuous inverse

EXERCISE V.8. *Show that, up to differentiable equivalence, hyperbolic equilibrium points do not have a finite classification. Hint: it suffices to show this in dimension* 1*. Show in that case that differentiable equivalence requires that the eigenvalues of the linear part coincide.*

From the previous examples and exercises we now see the utility of the following definition.

DEFINITION V.7 (Structural stability). A vector filed $X \in \mathcal{C}^k(D)$, with $k \geq 1$ is said to be *structurally stable* if there exists an $\varepsilon > 0$ such that every other vector field $Y \in \mathcal{C}^k(D)$ with $||Y - X||_{\mathcal{C}^k} < \varepsilon$ is topologically equivalent to X.

Notice that the concept of structural stability balances out $\mathcal{C}^1$-small perturbations with $\mathcal{C}^0$-equivalence. This will lead to important results.

THEOREM V.3 (Peixoto-Andronov-Pontryagin theorem). *A vector field $X \in \mathcal{X}^1(D)$ is structurally stable if and only if*

(1) X has finitely many equilibrium points, all being hyperbolic,
(2) X has finitely many periodic orbits, all being hyperbolic,
(3) X has no homoclinic nor heteroclinic connections.

Furthermore, the set of structurally stable vector fields is dense[4] *in $\mathcal{X}^1(D)$.*

REMARK V.4. We do not discuss a proof of the previous theorem, but mention a few important comments:

(1) Notice that the conditions to determine that a vector field is structurally stable are rather simple. Yet, they are not always easy to check, especially the third condition.
(2) The fact that structurally stable vector fields are dense is extremely important. It implies that for any structurally unstable vector field, one can find a sufficiently small perturbation that makes it structurally stable. It also implies that "typical" vector fields have only hyperbolic equilibria and/or periodic orbits without saddle connections.

V.1. Singularities of codimension 1

In this section we turn our attention to *structurally unstable* vector fields. There are two main reasons for this: structurally unstable vector fields represent boundaries between different classes of structurally stable vector fields. Thus, structurally unstable vector fields may give us information about transitions between different classes of equivalent vector fields. Another important reason is that although typical vector fields may no be structurally unstable, it is possible (as we shall see) that one-parameter families of vector fields (that is curves in $\mathcal{X}^k(D)$) do contain structurally unstable vector fields.

REMARK V.5. In these notes, we restrict ourselves to *local* bifurcations of *equilibria* of planar systems. Namely, we study the saddle-node bifurcation, and the Hopf bifurcation. A more extensive discussion on bifurcation theory can be found in, for example, [**10**].

For convenience, let us denote by $\mathcal{S}_0$ the set of structurally unstable vector fields in $\mathcal{X}^k(D)$.

DEFINITION V.8. A subset $\mathcal{S} \subset \mathcal{X}^k(D)$ is said to be a $\mathcal{C}^r$ submanifold of codimension 1 if there exists an open set $\mathcal{U} \subset \mathcal{X}^k(D)$ and a $\mathcal{C}^r$-function $H : \mathcal{U} \to \mathbb{R}$ such that $DH(X) \neq 0$ in $\mathcal{U}$ and $\mathcal{S} = \{f \in \mathcal{U} : H(X) = 0\}$.

[4] A subset A of a topological space V is said to be dense in X if every point of X either belongs to A or else is arbitrarily "close" to a member of A.

The above definition can be applied to characterize the set of structurally unstable vector fields. For example $H : \mathcal{X}^k(D) \to \mathbb{R}$ could be given by an eigenvalue of a linearization matrix. It turns-out that $\mathcal{S}_0$ is not a submanifold, but as we will see below, a subset of $\mathcal{S}_0$ is indeed a submanifold.

DEFINITION V.9. A structurally unstable vector field $X \in \mathcal{S}_0$ is called *singular of codimension* 1 if there exists a neighborhood U_0 of X in $\mathcal{S}_0$ such that every other vector field $Y \in U_0$ is topologically equivalent to X. The set of singular vector fields of codimension 1 shall be denoted by $\mathcal{S}_1$. Elements of $\mathcal{S}_1$ are called singularities (of codimension 1).

It turns out that $\mathcal{S}_1$ is a submanifold of codimension 1 of $\mathcal{X}^k(D)$. This indeed implies that for $X \in \mathcal{S}_1$, $\mathcal{S}_1$ divides a small neighborhood of X into two regions of different equivalence classes of vector fields. As a consequence, if $X \in \mathcal{S}^1$, then there is a family of vector fields $X_\lambda \in \mathcal{X}^k(D)$ (depending smoothly on λ) such that for all $Y \in \mathcal{X}^K(D)$ with $||Y - X||_{\mathcal{C}^k}$ small enough, Y is topologically equivalent to X_λ for some $\lambda \in \mathbb{R}$.

EXAMPLE V.3. Consider the subspace of linear vector fields in the plane

$$L(D) = \left\{ X \in \mathcal{X}^k(D) \,|\, X = ax\frac{\partial}{\partial x} + by\frac{\partial}{\partial y} \right\}.$$

In this case, the set of structurally unstable vector fields $\mathcal{S}_0$ is given by the subclass of vector fields that have at least one zero eigenvalue, i.e., such that $ab = 0$. However, the the set of singular vector fields of codimension 1, $\mathcal{S}_1$, is given by the vector fields with strictly one zero eigenvalue (and not two). Let $X_0 \in \mathcal{S}_1$, for example $X_0 = 0\frac{\partial}{\partial x} + b\frac{\partial}{\partial y}$, with $b > 0$. Let X_λ denote 1-parameter family $X_\lambda = \lambda\frac{\partial}{\partial x} + b\frac{\partial}{\partial y}$. Of course $X_0 = X_\lambda|_{\lambda=0}$. Any vector field in the neighborhood of X_0 is topologically equivalent to some element of the family X_λ. See a sketch in

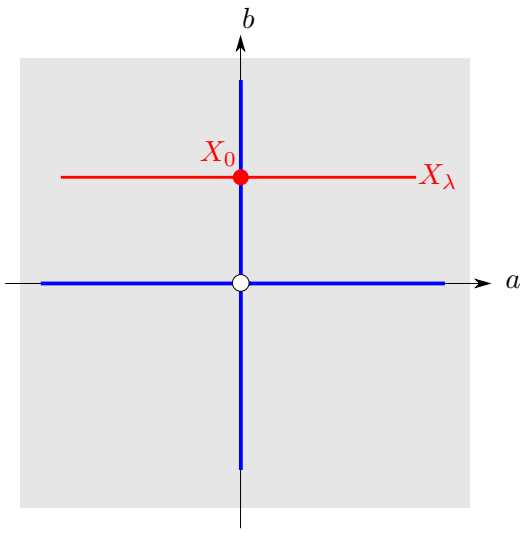


FIGURE 3. Schematic representation of L and the different subspaces. The space L can be divided by vector fields such that $ab = 0$, the blue lines. Each of the blue lines is a codimension 1 subset of L, i.e. $\mathcal{S}_1$. Notice that these lines divide L into four regions of topologically equivalent vector fields. However, vector fields in different quadrants are not topologically equivalent. The origin is the vector field such that $a = b = 0$ and therefore $\mathcal{S}_0$ is the union the different $\mathcal{S}_1$'s and the origin. For any $X_0 \in \mathcal{S}_1$ there is a one parameter family X_λ such that any vector field near X_0 is topologically equivalent to some member of X_λ. The previous fails if X_0 is the zero vector field.

DEFINITION V.10. Let $X \in \mathcal{S}_1$. An unfolding of X is a family of vector fields $X_\lambda \in \mathcal{X}^k(D)$, $\lambda \in \mathbb{R}^p$, depending smoothly on λ and such that $X = X_0$.

PROPOSITION V.3. *Let $X \in \mathcal{S}_1$. There exists a 1-parameter unfolding X_λ such that for every $\mathcal{C}^k$-close vector field Y to X, Y is topologically equivalent to some element of the unfolding X_λ.*

PROOF. Let $Z \in \mathcal{X}^k(D)$ be such that $DH(X)Z > 0$. Such a vector field always exists because $DH(X) \neq 0$. Let ε be sufficiently small and define the subsets

$$\begin{aligned}
\mathcal{W}_0 &= \{Y \in \mathcal{S}_1 \,:\, ||Y - X||_{\mathcal{C}^k} < \varepsilon\} \\
\mathcal{W}_+ &= \{Y_\lambda = Y_0 + \lambda Z \,:\, Y_0 \in \mathcal{W}_0,\, 0 < \lambda < \varepsilon\} \\
\mathcal{W}_- &= \{Y_\lambda = Y_0 + \lambda Z \,:\, Y_0 \in \mathcal{W}_0,\, -\varepsilon < \lambda < 0\}\,.
\end{aligned}$$

It follows from definition that $H(Y_0) = 0$ for all $Y_0 \in \mathcal{W}_0$. Furthermore, for sufficiently small ε we have that

$$\frac{\mathrm{d}}{\mathrm{d}\lambda} H(Y_\lambda) = DH(Y\lambda)Z > 0.$$

Therefore, by continuity, one concludes that $H > 0$ in $\mathcal{W}_+$ and $H < 0$ in $\mathcal{W}_-$. Hence, all $Y \in \mathcal{W}_+$ are structurally stable, and since $\mathcal{W}_+$ is open, they are topologically. Analogously, one can argue that all $Y \in \mathcal{W}_-$ are topologically equivalent. We thus conclude that any family Y_λ is an unfolding of X because it contains elements in $\mathcal{W}_0$, $\mathcal{W}_+$ and $\mathcal{W}_-$. □

REMARK V.6. Unfoldings are very useful in many contexts. For dynamical systems, they are useful to describe the behavior of a vector field near a nonhyperbolic equilibrium point, and their perturbations. Unfoldings play a major role in Singularity Theory, so your are encouraged to take that course.

REMARK V.7. In the rest of this section we detail the saddle-node and the Hopf bifurcations, both of codimension 1. There are not all of them, but we discuss them due to their importance. See the further comments at the end of this chapter.

V.1.1. Saddle-node bifurcation of equilibria. The first singularity of codimension 1 that we study is the one related to a simple 0 eigenvalue. Notice that a matrix

$$\begin{bmatrix} 0 & * \\ 0 & a \end{bmatrix}$$

with $a \neq 0$ can be diagonalized to

$$\begin{bmatrix} 0 & 0 \\ 0 & a \end{bmatrix}.$$

Thus, in appropriate coordinates, a vector field with a simple zero eigenvalue at the origin can be written as:

$$\begin{aligned} \frac{\mathrm{d}x_1}{\mathrm{d}t} &= f_1(x_1, x_2) \\ \frac{\mathrm{d}x_2}{\mathrm{d}t} &= ax_2 + f_2(x_1, x_2), \end{aligned} \tag{71}$$

where the functions f_i are sufficiently smooth, satisfying $f_1(0) = f_2(0) = 0$ and $\dfrac{\partial f_i}{\partial x_j}(0) = 0$ for $i, j = 1, 2$.

The following result generalizes the stable manifold theorem, see also theorem A.8.

THEOREM V.4. *Regarding (71):*

- *There exists an invariant curve W^c, called "a center manifold", that is tangent to the x_1-axis (the center eigenspace) at the origin.*
- *If $a > 0$, then there is a unique unstable manifold W^u tangent to the x_2-axis at the origin.*
- *If $a < 0$, then there is a unique stable manifold W^u tangent to the x_2-axis at the origin.*

Although, in general, center manifolds are not unique, any choice of center manifold can locally be described as a graph of a function, i.e. by the equation $x_2 = h(x_1)$. Since the center manifold is a solution of (71) we have that h satisfies:

$$\begin{aligned} \frac{\mathrm{d}x_2}{\mathrm{d}t} &= \frac{\partial h}{\partial x_1}\frac{\mathrm{d}x_1}{\mathrm{d}t} \\ ax_2 + f_2(x_1, x_2) &= \frac{\partial h}{\partial x_1} f_1(x_1, x_2) \\ ah + f_2(x_1, h(x_1)) &= \frac{\partial h}{\partial x_1} f_1(x_1, h(x_1)). \end{aligned}$$

Since we do not posses any further information on f_1, f_2, it suffices for now to know that $h = \mathcal{O}(x_1^2)$ as $x_1 \to 0$ since the center manifold is tangent to the x_1-axis. What is rather important though is to know if the center manifold attracts or repels nearby orbits. For this purpose the $y = x_2 - h(x_1)$ describe the distance of an arbitrary solution to the center manifold. It follows that y satisfies:

$$\begin{aligned} \frac{\mathrm{d}y}{\mathrm{d}t} &= \frac{\mathrm{d}x_2}{\mathrm{d}t} - \frac{\partial h}{\partial x_1}\frac{\mathrm{d}x_1}{\mathrm{d}t} \\ &= ax_2 + f_2(x_1, x_2) - \frac{\partial h}{\partial x_1} f_1(x_1, x_2) \\ &= a(y + h) + f_2(x_1, y + h) - \frac{\partial h}{\partial x_1} f_1(x_1, y + h) \\ &\sim ay + ah + f_2(x_1, h) - \frac{\partial h}{\partial x_1} f_1(x_1, h) + \mathcal{O}(x_1)y \\ &= (a + \mathcal{O}(x_1))y, \end{aligned}$$

for $x_1 \sim 0$. It follows that the center manifold is attracting or repelling if $a < 0$ or $a > 0$ respectively. Moreover, to have a qualitative understanding of the dynamics near the origin, it suffices to know the dynamics on the center manifold (because the other direction is hyperbolic and we know that depending on the sign of a the center manifold is attracting or repelling). So, the dynamics restricted to the center manifold are given by

$$\dot{x}_1 = f_1(x_1, h(x_1)) = c_1 x_1^2 + \mathcal{O}(x_1^3), \tag{72}$$

where $c = \frac{1}{2}\frac{\partial^2 f_1}{\partial x_1^2}(0)$ (recall that $f_1(x_1, x_2) = \mathcal{O}(||x||^2)$ and $h = \mathcal{O}(x_1^2)$). It follows then that the flow on the center manifold is as shown in figure 4.

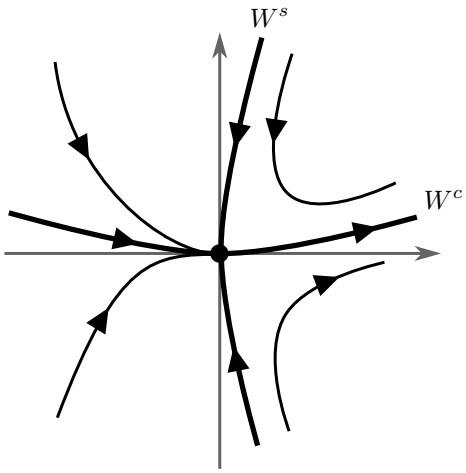


FIGURE 4. Local phase portrait at a saddle-node singularity. In this case the flow on the center manifold approaches the equilibrium point form one side, and goes away from from the equilibrium on the other side.

DEFINITION V.11. Consider the planar system

$$\frac{\mathrm{d}x}{\mathrm{d}t} = f(x),$$

with $x \in \mathbb{R}^2$. Assume that x^* is a nonhyperbolic equilibrium point and that the linearization of f at x^* has eigenvalues 0 and $a \neq 0$, and that (72) satisfies $c \neq 0$. Then x^* is an *elementary saddle-node*.

The saddle-node bifurcation receives several other names, among which we mention: fold bifurcation, tangent bifurcation, limit point, and turning point.

Next, we would like to understand what are the effects of small perturbations to (71). For this, let us consider a vector field $G_\lambda(x_1, x_2) = G(x_1, x_2, \lambda) = \begin{bmatrix} g_1(x_1, x_2, \lambda) \\ g_2(x_1, x_2, \lambda) \end{bmatrix}$, such that

$$\begin{aligned} g_1(x_1, x_2, 0) &= f_1(x_1, x_2) \\ g_2(x_1, x_2, 0) &= ax_2 + f_2(x_1, x_2). \end{aligned}$$

In this situation it is convenient to look at the "extended" system

$$\begin{aligned} \frac{\mathrm{d}x_1}{\mathrm{d}t} &= g_1(x_1, x_2, \lambda) \\ \frac{\mathrm{d}x_2}{\mathrm{d}t} &= g_2(x_1, x_2, \lambda) \\ \frac{\mathrm{d}\lambda}{\mathrm{d}t} &= 0, \end{aligned} \tag{73}$$

where we consider λ as a "dummy" or trivial variable. Of course (73) coincides with (71) when $\lambda = 0$. The origin $(x_1, x_2, \lambda) = (0, 0, 0)$ is nonhyperbolic, and the linearization is given by the matrix

$$A = \begin{bmatrix} 0 & 0 & * \\ 0 & a & * \\ 0 & 0 & 0 \end{bmatrix},$$

and thus the eigenvalues are $\{0, a, 0\}$. The center manifold theorem tells us that there is a 2-dimensional invariant (center) manifold, which is attracting if $a < 0$ and repelling if $a > 0$. Locally, such a manifold can be described by the equation $x_2 = h(x_1, \lambda)$. The dynamics restricted to the center manifold are given by

$$\frac{\mathrm{d}x_1}{\mathrm{d}t} = g_1(x_1, h(x_1, \lambda), \lambda) = G(x_1, \lambda),$$

where we know, from the case $\lambda = 0$ studied previously, that G satisfies:

$$G(0,0) = 0, \qquad \frac{\partial G}{\partial x_1}(0,0) = 0, \qquad \frac{\partial^2 G}{\partial x_1^2}(0,0) = 2c,$$

and we recall that $c \neq 0$. The above properties tell us that the graph of $G(x_1, 0)$ is locally a parabola tangent to the x_1-axis at the origin. Therefore, the graph of $G(x, \lambda) = G(x, 0) + \mathcal{O}(\lambda)$ is still locally a parabola, but can have zero, one, or two intersection points with the x_1-axis for λ small.

Indeed, let $L(x_1, \lambda) = \dfrac{\partial G}{\partial x_1}(x_1, x_2)$. Since $L(0,0) = 0$ and $\dfrac{\partial L}{\partial x_1} = 2c \neq 0$, it follows from the implicit function theorem that $L(x_1, \lambda) = 0$ has locally a unique solution given by the graph of a function $x_1 = \sigma(\lambda)$ with $\sigma(0) = 0$, i.e. $L(\sigma(\lambda), \lambda) = 0$ for sufficiently small λ. Consider next a function $K(y, \lambda) = G(y + \sigma(\lambda), \lambda)$. For y small, the function K is a small shift of the function G. Moreover, for y close to 0 we can use Taylor's formula to write:

$$\begin{aligned} K(y, \lambda) &= G(\sigma(\lambda), \lambda) + \frac{\partial G}{\partial y}(\sigma(\lambda), \lambda) y + \frac{1}{2}\frac{\partial^2 G}{\partial y^2}(\sigma(\lambda), \lambda) y^2 + \mathcal{O}(y^3, \lambda) \\ &= G(\sigma(\lambda), \lambda) + L(\sigma(\lambda), \lambda) y + c y^2 + \mathcal{O}(y^2, \lambda) + \mathcal{O}(y^3, \lambda) \\ &= G(\sigma(\lambda), \lambda) + y^2 (c + R_1(y, \lambda)) \\ \frac{\partial K}{\partial y}(y, \lambda) &= y(2c + R_2(y, \lambda)), \end{aligned}$$

where the functions R_1 and R_2 vanish at the origin. Let $H(\lambda) = \dfrac{G(\sigma(\lambda), \lambda)}{c} = \dfrac{K(0, \lambda)}{c}$. We then have that $H(0) = 0$ and if (y, λ) are sufficiently small so that $|R_1(y, \lambda)| < \dfrac{c}{2}$, the following relation holds:

$$H(\lambda) + \frac{y^2}{2} \leq \frac{K(y, \lambda)}{c} \leq H(\lambda) + \frac{3y^2}{2}.$$

This immediately implies that if $H(\lambda)$ is positive, negative, or zero, the function K has zero, one, or two intersection points with the x_1-axis for λ sufficiently small.

Example V.4. We provide three examples:

(1) Let $G(x_1, \lambda) = \lambda + x_1^2$. This is the case, for example, for the system

$$\begin{aligned} \frac{\mathrm{d}x_1}{\mathrm{d}t} &= \lambda + x_1^2 + \cdots \\ \frac{\mathrm{d}x_2}{\mathrm{d}t} &= -x_2 + \cdots \end{aligned} \tag{74}$$

In this case, for (x_1, x_2) near the origin and $\lambda \sim 0$, $G = 0$ has two roots for $\lambda < 0$ and no roots for $\lambda > 0$. Consequently, (74) has two equilibria $(x_1^*, x_2^*) = (\pm\sqrt{-\lambda}, 0)$ for $\lambda < 0$ and no equilibria for $\lambda > 0$.

This is the generic case, and is known as the *saddle-node bifurcation*. See a sketch of the corresponding dynamics in figure 5.

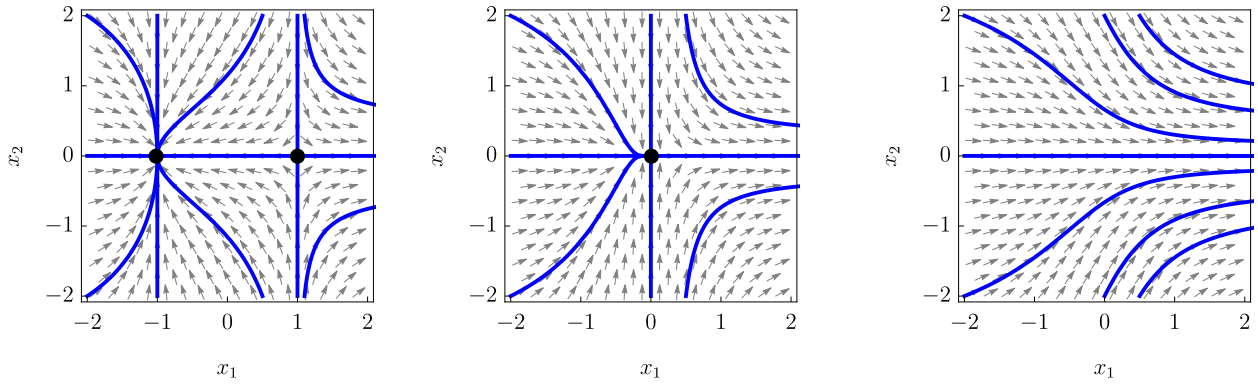


FIGURE 5. Phase portraits of (74) for $\lambda < 0$, $\lambda = 0$, and $\lambda > 0$ from left to right. These sketches represent the unfolding of a saddle-node singularity, where for $\lambda < 0$ (left) there is a saddle and a node, for $\lambda = 0$ (middle) the two singularities collide, and for $\lambda > 0$ (right) the equilibria have disappeared.

(2) Let $G(x_1, \lambda) = -\lambda^2 + x_1^2$. This is the case, for example, for the system

$$\frac{\mathrm{d}x_1}{\mathrm{d}t} = -\lambda^2 + x_1^2 + \cdots$$
$$\frac{\mathrm{d}x_2}{\mathrm{d}t} = -x_2 + \cdots$$

In this case, there are always two equilibria $(x_1^*, x_2^*) = (\pm\lambda, 0)$ for all $\lambda \neq 0$. Notice though that the stability of the equilibria change depending on the sign of λ. This can be seen from the Jacobian evaluated at the equilibrium points

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

This bifurcation is known as *the transcritical bifurcation.*

EXERCISE V.9. *Sketch the phase portraits for $\lambda < 0$, $\lambda = 0$, and $\lambda > 0$ for this example.*

(3) Let $G(x_1, \lambda) = -\lambda x_1 + x_1^2$. This is the case, for example, for the system

$$\frac{\mathrm{d}x_1}{\mathrm{d}t} = -\lambda x_1 + x_1^2 + \cdots$$
$$\frac{\mathrm{d}x_2}{\mathrm{d}t} = -x_2 + \cdots$$

EXERCISE V.10. *Show that this is also a transcritical bifurcation by drawing the corresponding phase portrait and comparing with the previous item.*

The last two examples are not generic (they are tangent to $\mathcal{S}_1$. This means that small $\mathcal{C}^1$ perturbations break the bifurcation generating, usually, a pair of saddle-node bifurcations).

EXERCISE V.11. *Consider a perturbation of the transcritical model*

$$\begin{aligned}\frac{\mathrm{d}x_1}{\mathrm{d}t} &= -\lambda^2 + x_1^2 + \lambda\\ \frac{\mathrm{d}x_2}{\mathrm{d}t} &= -x_2,\end{aligned}$$

and show qualitatively that there are, indeed, a couple of saddle-node bifurcations.

We conclude this section by mentioning that, through our previous analysis, one concludes that the singularity (71) admits the local unfolding

$$\begin{aligned}\frac{\mathrm{d}x_1}{\mathrm{d}t} &= \lambda + x_1^2\\ \frac{\mathrm{d}x_2}{\mathrm{d}t} &= -x_2.\end{aligned} \tag{75}$$

Although we do not provide a formal proof of our claim, it suffices to have the idea that every other unfolding of (71) is topologically equivalent to (75). Precise definitions of unfoldings and related topics are the subject of the course Singularity Theory.

V.1.2. Hopf bifurcation. In the previous section we considered a planar system for which the linearization has a simple zero eigenvalue at the nonhyperbolic equilibrium point. Now we consider the case where the nonhyperbolic equilibrium point is elliptic, i.e., it has a pair of purely imaginary eigenvalues with strictly nonzero imaginary part. In appropriate coordinates we can thus write:

$$\begin{aligned}\frac{\mathrm{d}x_1}{\mathrm{d}t} &= -ax_2 + f_1(x_1, x_2)\\ \frac{\mathrm{d}x_2}{\mathrm{d}t} &= ax_1 + f_2(x_1, x_2),\end{aligned}$$

where, as before, f_1 and f_2, as well as their derivatives, vanish at the origin.

To proceed with the analysis, let us introduce complex coordinates $z = x_1 + \imath x_2$. Thus:

$$\begin{aligned}\frac{\mathrm{d}z}{\mathrm{d}t} &= \frac{\mathrm{d}x_1}{\mathrm{d}t} + \imath\frac{\mathrm{d}x_2}{\mathrm{d}t}\\ &= -ax_2 + f_1(x_1, x_2) + \imath a x_1 + \imath f_2(x_1, x_2)\\ &= \imath a z + F(z, \bar{z}).\end{aligned}$$

Next, we will attempt to simplify the function F via the normal form procedure. For this, it is worth noting that, due to the properties of f_1 and of f_2, such functions have the local expansion

$$\begin{aligned}f_1(x_1, x_2) &\sim \sum_{2\le u+v\le 3} \alpha_{uv} x_1^u x_2^v + \mathcal{O}(||x||^4)\\ f_2(x_1, x_2) &\sim \sum_{2\le u+v\le 3} \beta_{uv} x_1^u x_2^v + \mathcal{O}(||x||^4),\end{aligned}$$

where α_{uv} and β_{uv} are some real coefficients. This implies that we can analogously write:

$$F(z, \bar{z}) \sim \sum_{2\le u+v\le 3} F_{uv} z^u \bar{z}^v + \mathcal{O}(|z|^4),$$

for some complex coefficients F_{uv}. For shortness of notation let $F_k = \sum_{u+v=k} F_{uv} z^u \bar{z}^v$.

Our goal now is to eliminate some of the higher order terms that appear in F through some convenient changes of coordinates. Let $h_2 = h_2(z, \bar{z})$ be a homogeneous polynomial of degree 2, and define the variable $w = z + h_2$. It follows that:

$$\begin{aligned}
\frac{\mathrm{d}w}{\mathrm{d}t} &= \frac{\mathrm{d}z}{\mathrm{d}t} + \frac{\partial h_2}{\partial z}\frac{\mathrm{d}z}{\mathrm{d}t} + \frac{\partial h_2}{\partial \bar{z}}\frac{\mathrm{d}\bar{z}}{\mathrm{d}t} \\
&= \imath a z + F(z,\bar{z})\frac{\partial h_2}{\partial z}(\imath a z + F(z,\bar{z})) + \frac{\partial h_2}{\partial \bar{z}}(-\imath a \bar{z} + \bar{F}(z,\bar{z})) \\
&= \imath a z + \imath a\left(z\frac{\partial h_2}{\partial z} - \bar{z}\frac{\partial h_2}{\partial \bar{z}}\right) + F(z,\bar{z}) + \frac{\partial h_2}{\partial z}F(z,\bar{z}) + \frac{\partial h_2}{\partial \bar{z}}\bar{F}(z,\bar{z}) \\
&= \imath a z + \imath a\left(z\frac{\partial h_2}{\partial z} - \bar{z}\frac{\partial h_2}{\partial \bar{z}}\right) + F_2(z,\bar{z}) + \underbrace{\sum_{k\geq 3} F_k(z,\bar{z}) + \frac{\partial h_2}{\partial z}F(z,\bar{z}) + \frac{\partial h_2}{\partial \bar{z}}\bar{F}(z,\bar{z})}_{\text{degree} \geq 3}.
\end{aligned}$$

Suppose that h_2 satisfies:

$$\imath a\left(h_2 - z\frac{\partial h_2}{\partial z} + \bar{z}\frac{\partial h_2}{\partial \bar{z}}\right) = F_2(z,\bar{z}).$$

Then, by substituting h_2, we could eliminate the quadratic terms in F_2, and only be left with the leading part and higher order terms of degree at least 3. In other words, we would further simplify to

$$\frac{\mathrm{d}w}{\mathrm{d}t} = \imath a w + \underbrace{\sum_{k\geq 3} F_k(z,\bar{z}) + \frac{\partial h_2}{\partial z}F(z,\bar{z}) + \frac{\partial h_2}{\partial \bar{z}}\bar{F}(z,\bar{z})}_{\text{degree} \geq 3}.$$

The above procedure can, in fact, be done iteratively, as we now describe. Let

$$h = h(z,\bar{z}) = \sum_{2\leq u+v} h_{uv} z^u \bar{z}^v = \sum_{2\leq k} h_k,$$

where $k = u + v$. It follows that

$$\begin{aligned}
z\frac{\partial h}{\partial z} &= \sum_{2\leq u+v} u h_{uv} z^u \bar{z}^v \\
\bar{z}\frac{\partial h}{\partial \bar{z}} &= \sum_{2\leq u+v} v h_{uv} z^u \bar{z}^v.
\end{aligned}$$

Let $w = z + h$ (such a transformation is called "near identity transformation"), then

$$\begin{aligned}
\frac{\mathrm{d}w}{\mathrm{d}t} &= \frac{\mathrm{d}z}{\mathrm{d}t} + \frac{\partial h}{\partial z}\frac{\mathrm{d}z}{\mathrm{d}t} + \frac{\partial h}{\partial \bar{z}}\frac{\mathrm{d}\bar{z}}{\mathrm{d}t} \\
&= \imath a z + \sum_{2\leq u+v} F_{uv} z^u \bar{z}^v + \imath a\left(z\frac{\partial h}{\partial z} - \bar{z}\frac{\partial h}{\partial \bar{z}}\right) + \frac{\partial h}{\partial z}F(z,\bar{z}) + \frac{\partial h}{\partial \bar{z}}\bar{F}(z,\bar{z}) \\
&= \imath a w + \sum_{2\leq u+v} F_{uv} z^u \bar{z}^v + \imath a\left(-h + z\frac{\partial h}{\partial z} - \bar{z}\frac{\partial h}{\partial \bar{z}}\right) + \frac{\partial h}{\partial z}F(z,\bar{z}) + \frac{\partial h}{\partial \bar{z}}\bar{F}(z,\bar{z}) \\
&= \imath a w + \sum_{2\leq u+v} F_{uv} z^u \bar{z}^v + \imath a\left(\sum_{2\leq u+v}(-1+u-v)h_{uv} z^u \bar{z}^v\right) + \frac{\partial h}{\partial z}F(z,\bar{z}) + \frac{\partial h}{\partial \bar{z}}\bar{F}(z,\bar{z}).
\end{aligned} \tag{76}$$

Here we notice that, for each pair (u, v), we can eliminate (by an appropriate choice of h_{uv}) the monomials $z^u \bar{z}^v$ as long as the nonresonance condition

$$v \neq u - 1,$$

is satisfied. At degree 2, there are no resonances, thus we can completely eliminate the terms for which $u+v=2$. Indeed, the quadratic terms can be eliminated by choosing

$$h_{20} = -F_{20}, \qquad h_{11} = F_{11}, \qquad h_{02} = \frac{1}{3}F_{02}.$$

At degree three the pair $(2,1)$ is resonant. This means that the monomial $z^2\bar{z}$ cannot be eliminated, at least with the coordinate transformation we have proposed. A similar argument follows at every degree.

Notice $z^2\bar{z} = |z|^2 z$. Thus (76) can be reduced to:

$$\frac{\mathrm{d}w}{\mathrm{d}t} = \imath a w + b_{21}|w|^2 w + \mathcal{O}(|w|^4), \tag{77}$$

where the coefficient b_{21} depends on the coefficients F_{uv}.

Let us now rewrite (77) in polar coordinates by employing $w = r\exp(\imath\theta)$. Thus:

$$\begin{aligned} \frac{\mathrm{d}w}{\mathrm{d}t} &= \frac{\mathrm{d}r}{\mathrm{d}t}\exp(\imath\theta) + \imath r\frac{\mathrm{d}\theta}{\mathrm{d}t}\exp(\imath\theta) \\ \imath a r\exp(\imath\theta) + b_{21}r^3\exp(\imath\theta) + \mathcal{O}(r^4) &= \frac{\mathrm{d}r}{\mathrm{d}t}\exp(\imath\theta) + \imath r\frac{\mathrm{d}\theta}{\mathrm{d}t}\exp(\imath\theta), \end{aligned}$$

which leads to

$$\begin{aligned} \frac{\mathrm{d}r}{\mathrm{d}t} &= \Re(b_{21})r^3 + \mathcal{O}(r^4), \\ \frac{\mathrm{d}\theta}{\mathrm{d}t} &= a + \Im(b_{21})r^2 + \mathcal{O}(r^3). \end{aligned} \tag{78}$$

Similar to what we did for the saddle-node singularity, we have the following definition:

Definition V.12. Consider the planar system

$$\frac{\mathrm{d}x}{\mathrm{d}t} = f(x),$$

and assume that x^* is a nonhyperbolic equilibrium point and that the linearization of f at x^* has a pair of purely imaginary eigenvalues $\pm\imath a$ with $a \neq 0$. If the normal form (78) satisfies $\Re(b_{21}) \neq 0$ then x^* is called an *elementary focus*.

Analogous to the saddle-node, we now consider perturbations of the elementary focus. We are thus interested on a family of vector fields, depending on one parameter λ, such that for $\lambda = 0$ we have an elementary focus. It then makes sense to consider that the linear part of the family has eigenvalues $\alpha(\lambda) \pm \imath\beta(\lambda)$ such that $\alpha(0) = 0$ and $\beta(0) = a \neq 0$. A representative of a family with such properties is:

$$\begin{aligned} \frac{\mathrm{d}x_1}{\mathrm{d}t} &= \alpha(\lambda)x_1 - \beta(\lambda)x_2 + g_1(x_1, x_2, \lambda) \\ \frac{\mathrm{d}x_2}{\mathrm{d}t} &= \beta(\lambda)x_1 + \alpha(\lambda)x_2 + g_2(x_1, x_2, \lambda). \end{aligned}$$

The normal form

$$\begin{aligned} \frac{\mathrm{d}r}{\mathrm{d}t} &= \alpha(\lambda)r + \Re(b_{21}(\lambda))r^3 + \cdots \\ \frac{\mathrm{d}\theta}{\mathrm{d}t} &= \beta(\lambda) + \Im(b_{21}(\lambda))r^2 + \cdots . \end{aligned}$$

can be obtained in an analogous way as above. This unfolding, provided with the non-degeneracy condition $\Re(b_{21}(\lambda)) \neq 0$, is known as the *Hopf bifurcation*.

EXAMPLE V.5. Consider

$$\begin{aligned} \frac{\mathrm{d}r}{\mathrm{d}t} &= \lambda r \pm r^3 \\ \frac{\mathrm{d}\theta}{\mathrm{d}t} &= a, \end{aligned} \tag{79}$$

which are truncated forms of the supercritical $(-)$ and of the subcritical $(+)$ Hopf bifurcation.

For the supercritical case, the equation $(\lambda r - r^3) = 0$ has solutions $r = 0$ and $r = \pm\sqrt{\lambda}$. Therefore, for $\lambda < 0$ the trajectories spiral toward $r = 0$, while for $\lambda > 0$ the trajectories spiral towards a limit cycle of radius $\sqrt{\lambda}$ (notice that $r^* = \sqrt{\lambda}$ is attracting for $\lambda > 0$). See a sketch in figure 6.

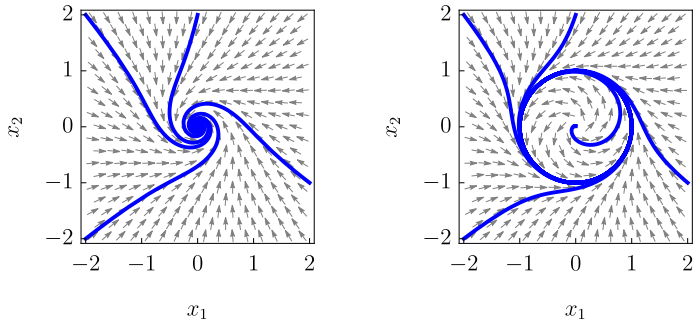


FIGURE 6. Phase portrait, in cartesian coordinates, for the supercritical Hopf bifurcation for $\lambda < 0$ on the left and $\lambda > 0$ on the right. For this sketch we have chosen $a > 0$, and changing its sign changes the direction of the rotations as can be seen from (79).

For the subcritical case, the equation $(\lambda r + r^3) = 0$ has solutions $r = 0$ and $r = \pm\sqrt{-\lambda}$. Therefore, for $\lambda > 0$ the trajectories spiral away from $r = 0$, while for $\lambda < 0$ there is an unstable limit cycle of radius $\sqrt{\lambda}$ (notice that $r^* = \sqrt{-\lambda}$ is repelling for $\lambda < 0$). See a sketch in figure 7.

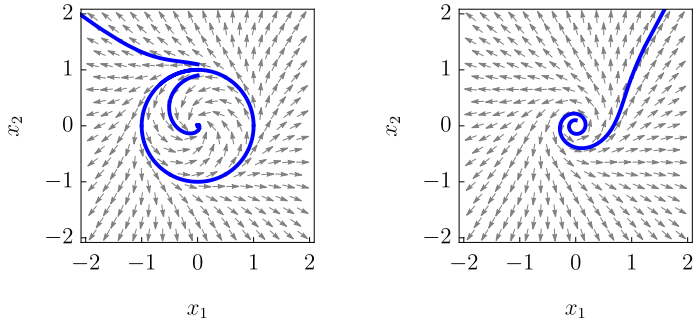


FIGURE 7. Phase portrait, in cartesian coordinates, for the subcritical Hopf bifurcation for $\lambda < 0$ on the left and $\lambda > 0$ on the right. For this sketch we have chosen $a > 0$, and changing its sign changes the direction of the rotations as can be seen from (79).

V.2. Some extra comments

The following comments are provided (without proof) to put the contents of the previous section in a larger context.

- If in a saddle-node bifurcation, the unstable manifold of the saddle makes a loop and connects with the stable manifold of the node, the as the equilibria collide, a periodic orbit arises. This is a result due to Andronov, Leontovich and Shilnikov, and is depicted in figure 8.

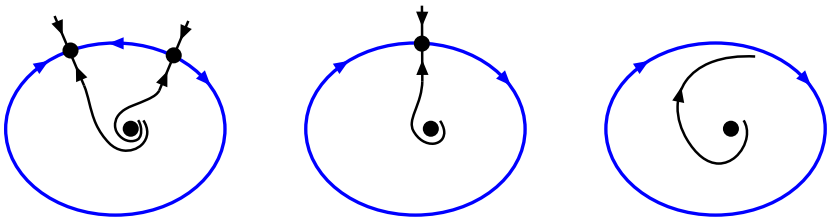


FIGURE 8. A saddle node bifurcation when the center manifold (middle) forms a loop.

- Just as equilibria can present saddle-node bifurcations, periodic orbits can also undergo such a bifurcation. Essentially, two hyperbolic periodic orbits, one of saddle type and one of node type, collide and disappear as the bifurcation ensues. Equivalently, a periodic orbit γ can be seen as a fixed point x^* of a Poincar'e map Π with section transverse to the periodic orbit. If the eigenvalue of $D\Pi(x^*)$ is exactly 1, then the periodic orbit γ (equivalently the fixed point x^*) is non-hyperbolic. Small perturbations of the latter situation create either two fixed points (one saddle and one node) or a regular flow. Generic unfoldings of the saddle-node singularity for periodic orbits are sketched in figure 9.

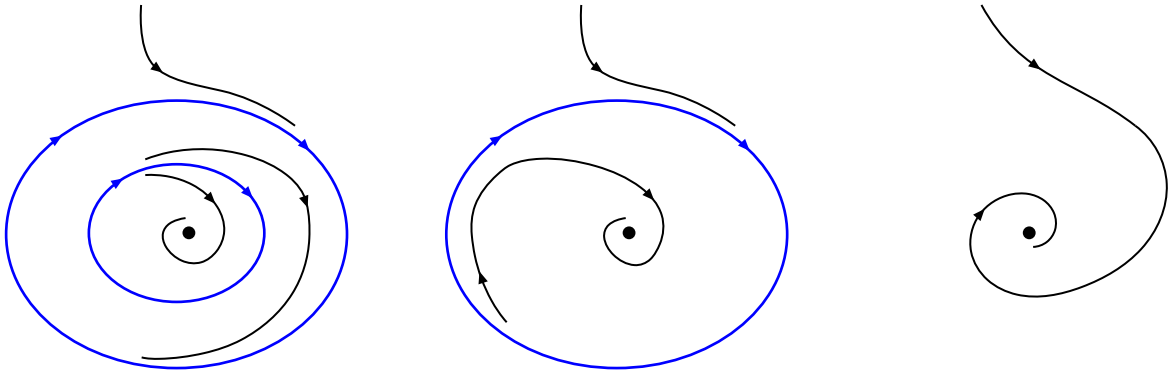


FIGURE 9. Sketch of a saddle-node bifurcation of periodic orbits.

- One should emphasize that the saddle-node bifurcation is robust because of the presence of a node. The situation is completely different when two saddles interact. Regarding saddle connections, one can consider what happens when a heteroclinic connection or a homoclinic loop is perturbed. Sketches of such situations are depicted in figure 10. In this context, [**14**] contains a recent account of important results.

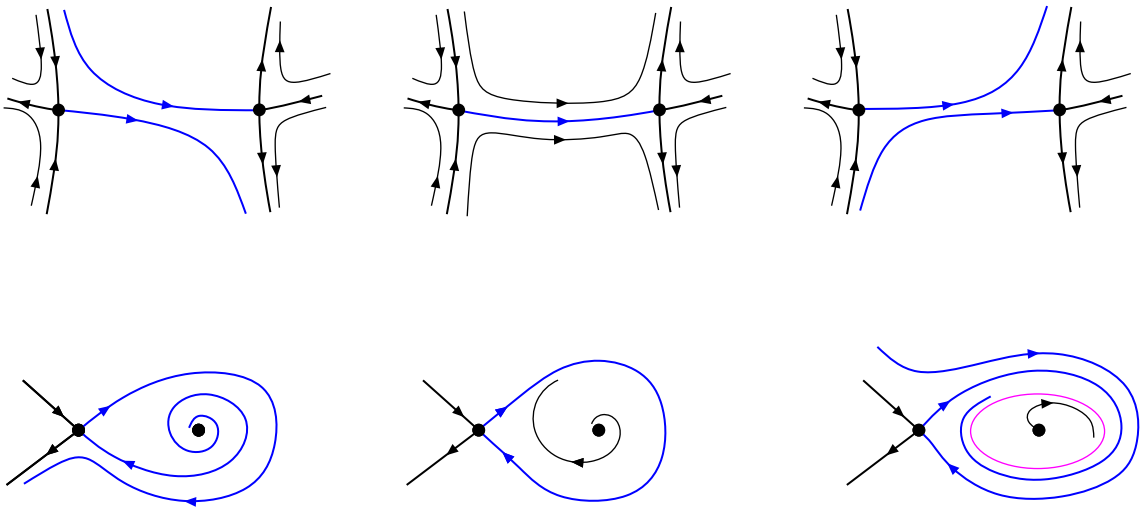


FIGURE 10. Above: perturbation of a heteroclinic connection. Below: perturbation of a homoclinic connection.

- The classification of singularities of higher codimension (generic for higher dimensional families) or in higher dimensions (greater than two) is considerably more complicated and even incomplete. For example in the plane, codimension 2 singularities include the pitchfork and the Bogdanov-Takens singularities. In higher dimensions, say 3, the situation can be much more subtle. To start, a result analogous to the Poincaré-Bendixson theorem does not exist. More importantly, however, is that structurally stable vector fields are not "typical" in higher dimensions, which indicates that vector fields in higher dimensions can be very complicated. Here, by typical, we mean that structurally stable vector fields are not dense in the space of vector fields (in higher dimensions).

V.3. Further exercises for this chapter

(1) For the following scalar equations, find all the equilibrium points and sketch the phase-portrait by determining the local stability of the equilibrium points. (Note: it may not be possible to explicitly find the equilibrium points, in that case, make a sketch of their relative position, and describe the qualitative behavior in a neighborhood of the equilibrium point(s)).

(a) $\dfrac{\mathrm{d}x}{\mathrm{d}t} = 3x^2 - 9$

(b) $\dfrac{\mathrm{d}x}{\mathrm{d}t} = 1 + \dfrac{1}{2}\cos x$

(c) $\dfrac{\mathrm{d}x}{\mathrm{d}t} = \exp(-x)\sin x$

(d) $\dfrac{\mathrm{d}x}{\mathrm{d}t} = \exp(x) - \cos x$

(2) For the following scalar equations, sketch the qualitative behavior of the vector field as the parameter λ is varied. Find the value $\lambda = \lambda^*$ at which a bifurcation occurs, which type of bifurcation is it?

(a) $\dfrac{\mathrm{d}x}{\mathrm{d}t} = 1 + \lambda x + x^2$

(b) $\dfrac{\mathrm{d}x}{\mathrm{d}t} = \lambda - \cosh x$

(c) $\dfrac{\mathrm{d}x}{\mathrm{d}t} = \lambda^2 + x^2$

(d) $\dfrac{\mathrm{d}x}{\mathrm{d}t} = x(\lambda - \exp(x))$

(e) $\dfrac{\mathrm{d}x}{\mathrm{d}t} = \lambda x + 4x^3$

(3) (A preamble to normal forms) Consider the scalar equation

$$\frac{\mathrm{d}x}{\mathrm{d}t} = ax - x^2 + bx^3 + \mathcal{O}(x^4),$$

where a and b are arbitrary constants with $a \neq 0$. Our goal is to find a near identity transformation $y = x + h(x)$ that eliminates the cubic term. Let $y = x + cx^3 + \mathcal{O}(x^4)$, where c is a constant. Write the system in the new coordinate y and choose c so that in the new equation, the cubic term disappears. Does this procedure work for even higher order terms? that is, to eliminate monomials of the form x^k with $k > 3$?

Hint: the transformation $y = x + cx^3 + \mathcal{O}(x^4)$ can be inverted as $x = y + dy^3 + \mathcal{O}(y^4)$, where d is some constant. What is the value of d?

(4) Although we have mainly seen vector fields on $\mathbb{R}$ and $\mathbb{R}^2$, this exercise explores vector fields on the circle. In these exercises $\theta \in \mathbb{S}^1$, that is, you may assume that $\theta \in [0, 2\pi)$. and identify 0 with 2π. Find equilibrium points and sketch the corresponding phase portrait (on the unit circle).

(a) $\dfrac{\mathrm{d}\theta}{\mathrm{d}t} = 1 + 2\cos\theta$

(b) $\dfrac{\mathrm{d}\theta}{\mathrm{d}t} = \sin(k\theta)$, where $k \in \mathbb{N}$.

(5) Consider the linear planar system

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \underbrace{\begin{bmatrix} a & b \\ c & d \end{bmatrix}}_{A} x.$$

What are the conditions for the entries of the matrix A to guarantee that there is a unique hyperbolic equilibrium point at the origin? Classify the hyperbolic equilibrium point according to values of the elements of A.

(6) Consider the linear system

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \begin{bmatrix} \lambda & a \\ 0 & \lambda \end{bmatrix},$$

where $\lambda \neq 0$ and $a \neq 0$. Assume $\lambda < 0$. What is the corresponding stable eigenspace? Sketch the corresponding phase-portrait.

(7) Which of the following systems is structurally stable? Why?

(a) $\begin{cases} \dot{x} &= x - 2y \\ \dot{y} &= -3x + 6y \end{cases}$

(b) $\begin{cases} \dot{x} &= 3x + y \\ \dot{y} &= -x \end{cases}$

(c) $\begin{cases} \dot{x} &= x + 2y \\ \dot{y} &= -x - y \end{cases}$

(8) Prove that the systems

$$\frac{\mathrm{d}x_1}{\mathrm{d}t} = -x_1$$
$$\frac{\mathrm{d}x_2}{\mathrm{d}t} = -x_2$$

and

$$\frac{\mathrm{d}x_1}{\mathrm{d}t} = -x_1$$
$$\frac{\mathrm{d}x_2}{\mathrm{d}t} = -2x_2$$

are topologically equivalent. Are they smoothly equivalent?

(9) (Gradient systems can't have periodic orbits) Consider a differential equation in the plane given by

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\nabla V(x),$$

where $V : \mathbb{R}^2 \to \mathbb{R}$ is a smooth function. These type of systems are known as *gradient systems*. Show that such gradient systems cannot present periodic orbits.

Hint: proceed by contradiction. Assume that the system has a T-periodic orbit γ and consider the variation of V in one period. That is evaluate $\displaystyle\int_0^T \frac{\mathrm{d}V}{\mathrm{d}t}\mathrm{d}t$. Such variation should be zero along γ, is that possible? Did any of the arguments depend on the dimension of x?

(10) Consider the planar systems

$$\frac{\mathrm{d}x_1}{\mathrm{d}t} = \lambda + x_1^2$$
$$\frac{\mathrm{d}x_2}{\mathrm{d}t} = -x_2$$

and

$$\frac{\mathrm{d}y_1}{\mathrm{d}t} = \sigma_1 + \sigma_2 y_1 + y_1^2$$
$$\frac{\mathrm{d}y_2}{\mathrm{d}t} = -y_2.$$

The first one is the truncated form of a saddle-node bifurcation. For the second one, is there a transformation $y_1 \mapsto y_1 + h(y_1, \sigma_2)$ that eliminates the term $\sigma_2 y_1$? If the answer is affirmative, can you say something about the qualitative behavior of the second system given that you already know the behavior of the first one?

(11) Show that the system

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \lambda x - y + xy^2$$
$$\frac{\mathrm{d}y}{\mathrm{d}t} = x + \lambda y + y^3,$$

undergoes a Hopf bifurcation at the origin as λ varies. What is the type of the bifurcation (sub / super critical)?

(12) Consider the system

$$\frac{\mathrm{d}x_1}{\mathrm{d}t} = -x_2 - x_1x_2 + 2x_2^2$$
$$\frac{\mathrm{d}x_2}{\mathrm{d}t} = x_1 - x_1^2x_2.$$

Similar to what we did in the analysis of the Hopf bifurcation, introduce a complex variable $z = x_1 + \imath x_2$ and obtain the corresponding differential equation. Obtain a truncated normal form up to the first non-zero coefficient. For such a truncated form, determine the stability of the origin.

CHAPTER VI

Regular Perturbation Theory

In this chapter we shall consider perturbation problems of the form:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = f(x) + \varepsilon g(x,t,\varepsilon), \tag{80}$$

where $x \in \mathbb{R}^n$, f and g are of class $\mathcal{C}^r$ with $r \geq 1$, ε is a small parameter and g is bounded and T-periodic in t.

The main question we are interested in addressing is: if x^* is a stable equilibrium point of the unperturbed problem, does the perturbed problem have a stable T-periodic solution in a neighborhood of x^*? Let us first formalize what we mean by a stable periodic orbit.

Definition VI.1. A periodic solution $\gamma(t)$ of (80) is *stable* if for every $\delta > 0$ there exists a $\mu > 0$ such that if $||x(0) - \gamma(0)|| < \mu$, then $||x(t) - \gamma(t)|| < \delta$ for all $t \geq 0$.

Proposition VI.1. *Let x^* be a hyperbolic equilibrium point of the unperturbed problem, that is (80) with $\varepsilon = 0$. Then, the perturbed problem (80) has a stable periodic solution for $\varepsilon > 0$ sufficiently small if and only if x^* is stable.*

Proof. Without loss of generality we assume that $x^* = 0 \in \mathbb{R}^n$. Let $A = D_x f(0)$, which by hypothesis is hyperbolic. Denote by $\phi_{t,\varepsilon}$ the flow induced by the perturbed problem.

Notice that the map $\phi_{T,\varepsilon}$ can be regarded as a Poincaré map on the extended space $\{(x,t) \in \mathbb{R}^n \times \mathbb{R}\}$ with Poincaré section $\{t = 0\}$. Since the perturbed problem is T-periodic, we can in fact identify the planes $\{t = 0\}$ and $\{t = T\}$.

Since $\phi_{t,0} : x \mapsto \exp(At)x$, we notice that $\dfrac{\partial \phi_{t,0}}{\partial x}(0) = \exp(At)$. Let us define the function $\Phi = \Phi(x,\varepsilon)$ by $\Phi = \phi_{T,\varepsilon} - x$. Thus, the function Φ satisfies

$$\Phi(0,0) = 0$$

and

$$\frac{\partial \Phi}{\partial x}(0,0) = \exp(AT) - I,$$

where I is the identity matrix in $\mathbb{R}^n$. Notice now that the existence of a periodic solution for ε sufficiently small is given by the solution of $\Phi(x,\varepsilon) = 0$. Next, since A is hyperbolic, all the eigenvalues of $\exp(AT)$ have modulus different from 1, which in turn means that $\dfrac{\partial \Phi}{\partial x}(0,0)$ has no zero eigenvalues. The implicit function theorem tells us that there is a unique solution of $\Phi = 0$ for $\varepsilon > 0$ sufficiently small. In other words there is a function $\hat{x} = \hat{x}(\varepsilon)$ such that $\phi_{T,\varepsilon}(\hat{x}) = \hat{x}$, with $\hat{x}(0) = 0$, for $\varepsilon > 0$ sufficiently small. This means that the Poincaré map $\phi_{T,\varepsilon}$ has a fixed point $\hat{x}$ for $\varepsilon > 0$ sufficiently small. Such a fixed point corresponds precisely with the periodic orbit λ of the perturbed problem.

Next, regarding stability, we recall that the map $\phi_{T,\varepsilon}$ is a contraction if and only if $\dfrac{\partial \phi_{T,\varepsilon}}{\partial x}(\hat{x})$ has all its eigenvalues within the unit circle. Since the eigenvalues of 1-parameter families of matrices depend continuously on the parameter, it follows that if $\varepsilon > 0$ is sufficiently small and $\dfrac{\partial \phi_{T,0}}{\partial x}(0)$ has

all its eigenvalues within the unit circle, then $\frac{\partial \phi_{T,\varepsilon}}{\partial x}(\hat{x})$ has all its eigenvalues within the unit circle. Thus, it suffices to look at the eigenvalues of $\frac{\partial \phi_{T,0}}{\partial x}(0) = \exp(AT)$. Such eigenvalues are within the unite circle if and only if A has all its eigenvalues with negative real part.

The above arguments immediately imply that, within a small neighborhood of γ it holds that $||x(t) - \gamma(t)|| < c||x(0) - \gamma(0)||$ for some constant $c > 0$ and all $t = kT$, $k \in \mathbb{N}$, if and only if A has all its eigenvalues with negative real part. It now rests to show that $||x(t) - \gamma(t)||$ is small for all $t \in (0, T)$ (and thus for all $t > 0$). Let, for simplicity, $y(t) = x(t) - \gamma(t)$. Then $\frac{\mathrm{d}y}{\mathrm{d}t} = f(y + \gamma) - f(\gamma) + \varepsilon G(t)$, where $G(t)$ is some bounded function. It follows that:

$$\begin{aligned} y(t) &= y(0) + \int_0^t [f(y(s) + \gamma(s)) - f(\gamma(s))]\,\mathrm{d}s + \varepsilon \int_0^t G(s)\mathrm{d}s \\ ||y(t)|| &\leq ||y(0)|| + \int_0^t ||f(y(s) + \gamma(s)) - f(\gamma(s))||\mathrm{d}s + \varepsilon \int_0^t ||G(s)||\mathrm{d}s \\ &\leq ||y(0)|| + \int_0^t L||y(s)||\mathrm{d}s + \varepsilon M t \\ &\leq ||y(0)||\exp(Lt) + \frac{\varepsilon M}{L}(\exp(Lt) - 1) \leq ||y(0)||\exp(LT) + \frac{\varepsilon M}{L}(\exp(LT) - 1), \end{aligned}$$

where L is a local Lipschitz constant, for the last line we have used Gronwall's inequality. □

Although the above description is of high importance, we are mostly interested in perturbations of systems that have a nonhyperbolic equilibrium point. Due to their importance, we will study such problems in the context of Hamiltonian Systems.

VI.1. Basics of Hamiltonian systems

In this section we briefly present some basic terminology regarding Hamiltonian systems. Most of the results contained here are proven in the course Hamiltonian Mechanics. The perturbation results we shall cover are the basic ones, and much development exists in the field. A good starting point is [1].

Let $H : \mathbb{R}^{2n} \times \mathbb{R} \to \mathbb{R}$ be a sufficiently smooth function ($\mathbb{R}^{2n}$ can be replaced by an open subset of $\mathbb{R}^{2n}$ or a $2n$-dimensional smooth manifold). We call n the number of degrees of freedom, and H the Hamiltonian (or Hamiltonian function). Local coordinates on (a subset of) $\mathbb{R}^n$ are denoted by $(q, p) = (q_1, \ldots, q_n, p_1, \ldots, p_n)$. The (canonical) *equations of motion* associated with the Hamiltonian H are defined by

$$\begin{aligned} \frac{\mathrm{d}q_i}{\mathrm{d}t} &= \frac{\partial H}{\partial p_i} \\ \frac{\mathrm{d}p_i}{\mathrm{d}t} &= -\frac{\partial H}{\partial q_i}. \end{aligned} \tag{81}$$

Example VI.1. One of the simplest examples of Hamiltonian systems are given by the Hamiltonian

$$H = \frac{1}{2}||p||^2 + V(q),$$

with corresponding equation

$$\frac{\mathrm{d}q_i}{\mathrm{d}t} = p_i$$
$$\frac{\mathrm{d}p_i}{\mathrm{d}t} = -\frac{\partial V(q)}{\partial q_i},$$

which may remind you of (some) mechanical systems and thus H is the "total energy" of the system (with $(p_i = m_i v_i)$).

Notice that (81) can be written as

$$\begin{bmatrix} \frac{\mathrm{d}q_i}{\mathrm{d}t} \\ \frac{\mathrm{d}p_i}{\mathrm{d}t} \end{bmatrix} = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix} \nabla H,$$

where I is the n-dimensional identity matrix. Such a skew-symmetric characteristic of the equations of motion is called "symplectic structure" and leads to some important properties of Hamiltonian systems.

PROPOSITION VI.2. *The following are satisfied by a Hamiltonian system (81).*

(1) H is a constant of motion.

(2) The flow of the Hamiltonian equations of motion preserves the volume form

$$\mathrm{d}q\mathrm{d}p = \mathrm{d}q_1 \cdots \mathrm{d}q_n \mathrm{d}p_1 \cdots \mathrm{d}p_n.$$

(3) A curve $\gamma : t \to \mathbb{R}^{2n}$ given by $\gamma(t) = \{(q(t), p(t)) : t_0 \le t \le t_1\}$ is a solution of (81) if and only if the integral

$$\int_\gamma (p \cdot \mathrm{d}q - H)\mathrm{d}t = \int_{t_0}^{t_1} \left(p(t) \cdot \frac{\mathrm{d}q(t)}{\mathrm{d}t} - H(q(t), p(t)) \right) \mathrm{d}t$$

is stationary with respect to variations of γ with fixed end points. (This is called Hamilton's invariance principle)

EXERCISE VI.1. *Prove the previous proposition. Hints:*

(1) Compute $\frac{\mathrm{d}H}{\mathrm{d}t}(q, p)$, and show that it is equal to zero.

(2) Let $M(0)$ be a compact subset in $\mathbb{R}^{2n}$ and denote by $M(t)$ the image of $M(0)$ under the flow ϕ_t of the Hamiltonian system. Consider the extended system

$$\frac{\mathrm{d}q}{\mathrm{d}t} = \frac{\partial H}{\partial p}$$
$$\frac{\mathrm{d}p}{\mathrm{d}t} = -\frac{\partial H}{\partial q}$$
$$\frac{\mathrm{d}t}{\mathrm{d}t} = 1.$$

Denote the extended vector field by $f = f(q, p, t)$, and define the cylinder

$$\mathcal{C} = \left\{ (M(s), s) \in \mathbb{R}^{2n} \times \mathbb{R} \, : \, 0 \le s \le t \right\}.$$

Compute

$$\int_{M(t)} \mathrm{d}q\,\mathrm{d}p - \int_{M(0)} \mathrm{d}q\,\mathrm{d}p = \int_{\mathcal{C}} (\nabla f)\mathrm{d}q\,\mathrm{d}p\,\mathrm{d}t,$$

and show that it is equal to zero. You may want to use Gauss' divergence theorem.

(3) *Let* $\delta(t) = \left\{(\xi(t),\eta(t)) \in \mathbb{R}^{2n} : t_0 \le t \le t_1\right\}$ *be a curve such that* $\delta(t_0) = \delta(t_1) = 0$. *Using the notation* $S(\gamma) = \int_\gamma (p\cdot \mathrm{d}q - H)\mathrm{d}t$ *we have*

$$S(\gamma+\varepsilon\delta) = \int_{t_0}^{t_1} \left(p\frac{\mathrm{d}q}{\mathrm{d}t} + \varepsilon p\frac{\mathrm{d}\xi}{\mathrm{d}t} + \varepsilon\eta\frac{\mathrm{d}q}{\mathrm{d}t} + \varepsilon^2\eta\frac{\mathrm{d}\xi}{\mathrm{d}t} - H(q+\varepsilon\xi, p+\varepsilon\eta)\right)\mathrm{d}t.$$

Compute $\dfrac{\mathrm{d}}{\mathrm{d}\varepsilon} S(\gamma+\varepsilon)|_{\varepsilon=0}$ *and show that it vanishes if and only if* $\gamma(t) = (q(t),p(t))$ *satisfies the Hamiltonian equations of motion. (You need to integrate* $\int p\dfrac{\mathrm{d}\xi}{\mathrm{d}t}\mathrm{d}t$ *by parts and recall that* $\xi = 0$ *at the end points).*

The following definition plays a central role in the study of Hamiltonian systems.

DEFINITION VI.2. Let $f = f(q,p)$ and $g = g(q,p)$ be two differantiable functions. The Poisson bracket of f and g is defined as

$$\{f,g\} = \sum_{i=1}^{n} \frac{\partial f}{\partial q_i}\frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i}\frac{\partial q}{\partial q_i}.$$

The Poisson bracket satisfies the following properties:

(1) it is anti-symmetric, i.e., $\{g,f\} = -\{f,g\}$,
(2) it is bi-linear, i.e., $\{\alpha f + \beta g, h\} = \alpha\{f,h\} + \beta\{g,h\}$ and $\{f, \alpha g + \beta h\} = \alpha\{f,g\} + \beta\{f,h\}$ for any constants $\alpha, \beta \in \mathbb{R}$,
(3) satisfies the Jacobi identity, i.e., $\{\{f,g\},h\} + \{\{h,f\},g\} + \{\{g,h\},f\} = 0$.
(4) satisfies the Leibnitz rule, i.e., $\{fg,h\} = g\{f,h\} + f\{g,h\}$.

Notice that, in particular,

$$\frac{\mathrm{d}f}{\mathrm{d}t} = \sum_{i=1}^{n} \frac{\partial f}{\partial q_i}\frac{\mathrm{d}q_i}{\mathrm{d}t} + \frac{\partial f}{\partial p_i}\frac{\mathrm{d}p_i}{\mathrm{d}t} = \{f,H\}.$$

Thus, it is evident that $\dfrac{\mathrm{d}H}{\mathrm{d}t} = \{H,H\} = 0$.

DEFINITION VI.3. A function $F = F(p,q)$ is *a constant of motion* (or a first integral) if and only if $\{F,H\} = 0$.

Therefore, H is always a constant of motion. It is not difficult to check that if F and G are two constants of motion, then $\{F,G\}$ is also a constant of motion.

From the above arguments, it follows that the solutions of (81) belong to the invariant sets H =constant. Depending on the dimension of the problem, the motion on such invariant sets may be very complicated. However, as a general rule, the more constants of motion one knows, the easier the analysis may become because one is able to find smaller invariant manifolds where the trajectories evolve.

As we have seen already, changes of coordinates usually allow us to simplify the analysis of a problem. In the context of Hamiltonian systems, one would like to use changes of coordinates that

preserve the symplectic structure of the equations of motion. Transformations that preserve such a symplectic structure are called *canonical transformations*.

PROPOSITION VI.3. *Given a canonical transformation $(q,p) \mapsto (Q,P)$, the new equations of motion are*

$$\begin{aligned} \frac{\mathrm{d}Q_i}{\mathrm{d}t} &= \frac{\partial K}{\partial P_i} \\ \frac{\mathrm{d}P_i}{\mathrm{d}t} &= -\frac{\partial K}{\partial Q_i}, \end{aligned}$$

where $K = K(Q,P)$ is the (new) Hamiltonian expressed in the new variables (Q,P).

EXAMPLE VI.2.

- Let $Q = q + a$ and $P = p + b$, where a, b are constant vectors. It is straightforward to check that this is a canonical transformation.
- Let $Q = q + f(p)$ and $P = p$. Then, $H(q,p) = H(Q - f(p), p) = H(Q - f(P), P) = K(Q,P)$, and using the chain rule we have $\dfrac{\partial H}{\partial q} = \dfrac{\partial K}{\partial Q}$ and $\dfrac{\partial H}{\partial p} = \dfrac{\partial K}{\partial Q}\dfrac{\partial f}{\partial p} + \dfrac{\partial K}{\partial P}$. Since $\dfrac{\mathrm{d}Q}{\mathrm{d}t} = \dfrac{\partial H}{\partial p} - \dfrac{\partial f}{\partial p}\dfrac{\partial H}{\partial q}$ and $\dfrac{\mathrm{d}P}{\mathrm{d}t} = -\dfrac{\partial H}{\partial q}$ we indeed have that in the new coordinates
$$\begin{bmatrix} \dfrac{\mathrm{d}Q_i}{\mathrm{d}t} \\ \dfrac{\mathrm{d}P_i}{\mathrm{d}t} \end{bmatrix} = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix} \nabla K.$$
- Let $\begin{bmatrix} Q \\ P \end{bmatrix} = A \begin{bmatrix} q \\ p \end{bmatrix}$, where $AJA^\top = J$ with $J = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}$. For simplicity let $X = (Q,P)$ and $x = (q,p)$. Then $\dfrac{\mathrm{d}X}{\mathrm{d}t} = A\dfrac{\mathrm{d}x}{\mathrm{d}t} = AJ\nabla H(x) = AJA^\top \nabla H(X) = J\nabla H(x)$.

  A matrix A satisfying $AJA^\top = J$. This example shows that any (linear) symplectic transformation is a canonical transformation.

A useful approach to obtain canonical transformations is via *generating functions*. A generating function is a smooth function $G = G(q,Q)$ (there are other types of generating functions, which shall not be discussed here) satisfying the non-degeneracy condition $\det \dfrac{\partial^2 G}{\partial q \partial Q} \neq 0$. We define the new coordinate P implicitly by the relations

$$\begin{aligned} p(q,Q) &= \frac{\partial G}{\partial q}(q,Q) \\ P(q,Q) &= -\frac{\partial G}{\partial Q}(q,Q). \end{aligned} \tag{82}$$

Due to the non-degeneracy condition, the equation $p = \dfrac{\partial G}{\partial q}(q,Q)$ can be "inverted" to obtain $Q = Q(q,p)$, that is, the new variable Q is a function of the "old" variables (q,p)[1]. Therefore, (82) indeed defines (implicitly) a transformation $(q,p) \mapsto (Q,P)$.

[1]Recall the inverse function theorem: Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be $\mathcal{C}^1$ on some open set around a point x^*, and suppose that $\det D_x f(x^*) \neq 0$. Then, there is an open set U containing x^* and an open set V containing $f(x^*)$ such that $f : U \to V$ has a continuous inverse $f^{-1} : V \to U$, which is differentiable for all $y \in V$.

EXAMPLE VI.3. Suppose $G = qQ$. Thus $p = Q$ and $P = -q$. Thus this generating function defines the canonical transformation $(q, p) \mapsto (p, -q)$.

The usefulness of canonical transformations is better witnessed in a general example.

EXAMPLE VI.4. Consider a 1-DOF Hamiltonian system, i.e. $(q, p) \in \mathbb{R}^2$ and $H : \mathbb{R}^2 \to \mathbb{R}$. As mentioned before, since H is a constant of motion, the solutions of the equations of motion evolve in level sets given by $H =$constant. In this case, the level sets are 1-dimensional. Thus, in fact, the orbits themselves are given by the level sets of H. For the analysis, it would be greatly convenient if one could find new coordinates where one of them is constant along each invariant level set. Such coordinates are called *action-angle variables*. We thus proceed to construct a canonical transformation that allows us to obtain action-angle variables.
Let us assume that each level curve $\gamma_h = \{(q, p) \in \mathbb{R}^2 : H(q, p) = h\}$ is bounded. We define *the action* of the level curve as

$$I(h) = \frac{1}{2\pi} \int_{\gamma_h} p \, \mathrm{d}q.$$

Further assuming that $\dfrac{\mathrm{d}I}{\mathrm{d}h} \neq 0$, we can write the inverse of the action, i.e. $h = h(I)$. Suppose that $p(h, q)$ is a solution of $H(q, p) = h$. Then, we can define the generating function

$$G(I, q) = \int_{q_0}^{q} p(h(I), \tilde{q}) \mathrm{d}\tilde{q}.$$

Thus, we define the (canonical) transformation $(q, p) \mapsto (I, \phi)$ implicitly by

$$p(I, q) = \frac{\partial G}{\partial q}$$
$$\phi(I, q) = \frac{\partial G}{\partial I}.$$

To see that ϕ is indeed an angle, we can consider its variation during one whole period, i.e.:

$$\Delta\phi = \int_{\gamma_h} \frac{\partial \phi}{\partial q} \mathrm{d}q = \int_{\gamma_h} \frac{\partial}{\partial q} \left(\frac{\partial G}{\partial I} \right) \mathrm{d}q = \frac{\partial}{\partial I} \int_{\gamma_h} p \mathrm{d}q = \frac{\partial}{\partial I} (2\pi I) = 2\pi.$$

That is $\phi \to 2\pi$ as $q \to q_0$ along an orbit. The Hamiltonian in the action-angle variables takes now the form

$$K(\phi, I) = h(I),$$

and the equations of motion are

$$\frac{\mathrm{d}\phi}{\mathrm{d}t} = \frac{\partial K}{\partial I} = \frac{h(I)}{I}$$
$$\frac{\mathrm{d}I}{\mathrm{d}t} = -\frac{\partial K}{\partial \phi} = 0.$$

To fix ideas, consider a harmonic oscillator with $H(q, p) = \dfrac{1}{2}(p^2 + q^2)$. Each level curve $\{H = h\}$ is a circle satisfying $q^2 + p^2 = 2h$, thus on each level curve $p = \pm\sqrt{2h - q^2}$.
The action variable thus reads as

$$I = \frac{1}{2\pi} \int_{\gamma} p \, \mathrm{d}q = h,$$

where used the simple fact that $\int_\gamma p\,\mathrm{d}q$ is the area of a circle of radius $\sqrt{2h}$. We see, therefore, that the action variable is constant (in this case equal to h).
As for the angle variable, we can take $\phi = \theta$ so that the variables (I, ϕ) are defined by

$$q(I,\theta) = \sqrt{2I}\sin\theta$$
$$p(I,\theta) = \sqrt{2I}\cos\theta.$$

Indeed, just by definition one has:

$$G(I,q) = \int_{q_0}^{q} \sqrt{2h - q^2}\mathrm{d}\tilde{q} = 2I\int_0^\theta \cos^2\tilde{\theta}\mathrm{d}\tilde{\theta} = I(\theta - \sin\theta\cos\theta).$$

Therefore

$$\phi = \frac{\partial G}{\partial I} = \theta + \sin\theta\cos\theta + I(1+\cos^2\theta - \sin^2\theta)\frac{\partial\theta}{\partial I} = \theta + \sin\theta\cos\theta + 2I\cos^2\theta\left(-\frac{1}{2I}\frac{\sin\theta}{\cos\theta}\right) = \theta$$

The Hamiltonian in the action-angle variables is therefore $K = K(\phi, I) = I$, and the equations of motion now read as

$$\frac{\mathrm{d}\phi}{\mathrm{d}t} = 1$$
$$\frac{\mathrm{d}I}{\mathrm{d}t} = 0.$$

In the previous example we took advantage of the fact that the Hamiltonian is *always* a conserved quantity. In higher dimensions, it is not always true that there are more conserved quantities.

Definition VI.4. A Hamiltonian system with n degrees of freedom is *integrable*, if it has n conserved quantities $H = J_1, J_2, \ldots, J_n$ such that the following are satisfied:

- $\{J_i, J_j\} = 0$ for all (i, j),
- let $h = (h_1, \ldots, h_n)$ and define $M_h = \big\{(q,p) \in \mathbb{R}^{2n} : J_i(q,p) = h_i,\ i = 1, \ldots, n\big\}$; the gradients of J_i are linearly independent for every point in M_h.

Functions satisfying the above definition are said to be in *involution*.

Theorem VI.1 (Liouville-Arnold). *If a Hamiltonian system is integrable such that M_h is compact and connected, then M_h is diffeomorphic to an n-dimensional torus $\mathbb{T}^n = \mathbb{S}^1 \times \cdots \times \mathbb{S}^1$. Moreover, there exist action-angle coordinates $(\phi_1, \ldots, \phi_n, I_1, \ldots, I_n) \in \mathbb{T}^n \times \mathbb{R}^n$ and frequencies $\omega_1, \ldots, \omega_n$ such that the equations of motion are given by*

$$\frac{\mathrm{d}\phi_i}{\mathrm{d}t} = \omega_i$$
$$\frac{\mathrm{d}I_i}{\mathrm{d}t} = 0,$$

and thus the corresponding solutions are simply given by

$$\phi_i(t) = \omega_i t + \phi_i(0)$$
$$I_i(t) = I_i(0).$$

EXAMPLE VI.5. Let $n = 1$ and assume that $H(q,p) = H(q+1,p)$. As a (not so particular) example we can take

$$H = \frac{1}{2}p^2 + V(q),$$

where $V(q+1) = V(q)$. For any level set $\{H = h\}$, we have $p = \pm\sqrt{2(h - V(q))} \neq 0$. The corresponding phase-space is the cylinder $C = \mathbb{T} \times \mathbb{R}$, and so each trajectory wraps around the cylinder. For any point $(q,p) \in C$, the action variable is defined as the integral

$$I = \int_0^1 p \, \mathrm{d}q,$$

where the integral is taken along the level curve passing through (q,p). The angle variable θ is defined by

$$\theta = \frac{1}{T}\int_0^q \frac{\mathrm{d}q}{p} = \frac{\int_0^x p \, \mathrm{d}q}{\int_0^1 p \, \mathrm{d}q}.$$

Since the motion of integrable systems is "relatively simple", we now study perturbations of integrable systems.

Let us consider a Hamiltonian of the form

$$H(\phi, I, \varepsilon) = H_0(I) + \varepsilon H_1(I, \phi, \varepsilon),$$

where the system defined by H_0 is integrable. The associated equations of motion are

$$\begin{aligned} \frac{\mathrm{d}\phi_i}{\mathrm{d}t} &= \frac{\partial H_0}{\partial I_i} + \varepsilon \frac{\partial H_1}{\partial I_i} \\ \frac{\mathrm{d}I_i}{\mathrm{d}t} &= -\varepsilon \frac{\partial H_1}{\partial \phi_i}. \end{aligned} \tag{83}$$

Notice that (83) is of the form (81) but time-independent. There are some particular cases where we can indeed reformulate (83) as a periodic perturbation. We describe a few examples of such cases:

Case 1: Assume that $\dfrac{\partial H_0}{\partial I_1} \neq 0$ for all $I \in \mathbb{R}^n$ and that $\dfrac{\partial H_1}{\partial I_1}$ is bounded. Then, for ε sufficiently small, ϕ_1 is monotonous in t. Thus, one can "replace" t by ϕ_1 obtaining then

$$\frac{\mathrm{d}\phi_j}{\mathrm{d}t} = \frac{\frac{\partial H_0}{\partial I_j} + \varepsilon \frac{\partial H_1}{\partial I_j}}{\frac{\partial H_0}{\partial I_1} + \varepsilon \frac{\partial H_1}{\partial I_1}}, \qquad j = 2, \ldots, n$$

$$\frac{\mathrm{d}I_i}{\mathrm{d}t} = -\varepsilon \frac{\frac{\partial H_1}{\partial \phi_i}}{\frac{\partial H_0}{\partial I_1} + \varepsilon \frac{\partial H_1}{\partial I_1}}, \qquad i = 1, \ldots, n,$$

which is a system of $2n-1$ equations. Notice that this is a time-varying system, which indeed is periodic in ϕ_1 because ϕ_1 is an angle.

Case 2: Assume now that $\gamma(t) = (q(t), p(t))$ is a T-periodic solution of a (not necessarily integrable) Hamiltonian system. Let $y(t) = x(t) - \gamma(t)$, where $x(t) = (q(t), p(t))$. Using Taylor expansion

we know that y satisfies

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \underbrace{\begin{pmatrix} \dfrac{\partial^2 H}{\partial q \partial p} & \dfrac{\partial^2 H}{\partial p^2} \\ -\dfrac{\partial^2 H}{\partial q^2} & -\dfrac{\partial^2 H}{\partial p \partial q} \end{pmatrix}\Bigg|_{\gamma}}_{A(t)} y + g(t,y),$$

where $g \in \mathcal{O}(||y||^2)$ for $||y||$ small. Let $U(t)$ be the solution of $\dfrac{\mathrm{d}U}{\mathrm{d}t} = A(t)U$ with $U(0)$ the $2n \times 2n$ identity matrix. Since $A(t)$ is T-periodic, there is a T-periodic matrix $P(t)$ and a constant matrix B such that[2] $U(t) = P(t)\exp(tB)$. Therefore, P satisfies the differential equation

$$\frac{\mathrm{d}P}{\mathrm{d}t} = A(t)P - PB.$$

Let $z(t)$ be defined by the change of coordinates $y = P(t)z$. Then

$$\frac{\mathrm{d}z}{\mathrm{d}t} = Bz - P^{-1}g(t, Pz).$$

Finally, let $z = \varepsilon w$ with $\varepsilon > 0$ sufficiently small. Notice that for ε small, the variable w is a "zoom-in" into $y \sim 0$, or equivalently $x \sim \gamma$. The corresponding differential equation in w now reads as

$$\frac{\mathrm{d}w}{\mathrm{d}t} = Bw + \varepsilon G(t, w, \varepsilon), \tag{84}$$

where $G = \dfrac{1}{\varepsilon^2} P^{-1} g(t, \varepsilon P w)$ with $G = \mathcal{O}(||w^2||)$. Indeed:

$$||G|| \leq \frac{1}{\varepsilon^2}||P^{-1}|| \cdot \underbrace{||g(t, \varepsilon P w)||}_{=\mathcal{O}(||(\varepsilon P w)^2||)} \leq \frac{1}{\varepsilon^2}||P^{-1}|| \cdot (\varepsilon^2 ||P||^2 ||w||^2) \leq K||w||^2,$$

for some positive constant K (the last inequality holds because P is periodic). System (84) is now of the form (80).

Example VI.6. Consider a system of the form

$$\frac{\mathrm{d}x}{\mathrm{d}t} = Ax + \varepsilon g(t),$$

$g(t)$ is T-periodic and $A = [a_i]$ is diagonal. Then, each component solution is given by

$$x_i(t) = \exp(a_i t)x_i(0) + \varepsilon \int_0^t \exp(a_i(t-\tau))g_i(\tau)\mathrm{d}\tau.$$

Since each g_i is periodic, we may assume that the g_i's are given in Fourier series as

$$g_i(s) = \sum_{k \in \mathbb{Z}} c_{ik} \exp\left(\frac{2\pi}{T} \imath k s\right).$$

Notice therefore that to compute the solution $x(t)$ we need to solve integrals of the form

$$\int_0^t \exp\left(a_i(t-s)\right) \exp\left(\frac{2\pi}{T}\imath k s\right) \mathrm{d}s = \exp(a_i t) \int_0^t \exp\left(\left(\frac{2\pi}{T}\imath k - a_i\right) s\right) \mathrm{d}s.$$

[2] This is Floquet's theorem.

Depending on the eigenvalues a_i we have that the integral has solution

$$\begin{cases} \dfrac{\exp\left(\frac{2\pi}{T}\imath k t\right) - \exp(a_i t)}{\frac{2\pi}{T}\imath k - a_i}, & \text{if } a_i \neq \dfrac{2\pi}{T}\imath k \\ t\exp(a_i t), & \text{if } a_i = \dfrac{2\pi}{T}\imath k \end{cases}.$$

So:

- If $\Re(a_i) \neq 0$ then the solution $x_i(t)$ grows or decreases exponentially (as for the unperturbed case).
- If $\Re(a_i) = 0$ then the solution present resonance: if a_i is a multiple of $\frac{2\pi}{T}\imath$, then the solutions flow linearly with t. Otherwise, the solution remains bounded, but the amplitude may be large depending on the denominator $\frac{2\pi}{T}\imath k - a_i$. A similar situation occurs with the secular terms in Example III.8

In the next section we are going to study a perturbation method that allows us to provide estimates of the perturbed solution on systems where the higher order terms are periodic.

VI.2. Averaging method

The averaging method is applicable to systems of the form

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \varepsilon g(x, t, \varepsilon), \tag{85}$$

where g is T-periodic in T, $\mathcal{C}^r$-smooth and bounded. We will later see how to associate such an equation to a perturbation problem of the form (80).

Definition VI.5. Given (85), the associated *autonomous averaged system* is defined as

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \varepsilon \underbrace{\frac{1}{T}\int_0^T g(y, t, 0)\mathrm{d}t}_{=:\bar{g}(y)}. \tag{86}$$

In this context we have:

Theorem VI.2. *Consider (85) and its associated average (86). There exists a $\mathcal{C}^r$ change of coordinates $x = y + \varepsilon w(y, t, \varepsilon)$, where w is T-periodic in t, transforming (85) into*

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \varepsilon\bar{g}(y) + \varepsilon^2 g_1(y, t, \varepsilon),$$

where g_1 is T-periodic in t. Moreover:

(1) Let $x(t)$ and $y(t)$ be solutions of (85) and of (86) with initial conditions $x_0 = x(0)$ and $y_0 = y(0)$ respectively. If $||x_0 - y_0|| = \mathcal{O}(\varepsilon)$, then $||x(t) - y(t)|| = \mathcal{O}(\varepsilon)$ on a time scale $t \sim \frac{1}{\varepsilon}$.

(2) If y^* is a hyperbolic equilibrium point of (86), then, for ε sufficiently small, the perturbed system (85) has a unique hyperbolic orbit γ_ε (with the same stability properties as y^*) in a small neighborhood of y^*.

Proof. See [10]. □

Let us now see some examples of the application of the averaging theorem.

EXAMPLE VI.7. Consider the scalar system $\dfrac{\mathrm{d}x}{\mathrm{d}t} = \varepsilon x \sin^2 t$. We then have

$$\begin{aligned}\bar{g}(y) &= \frac{1}{\pi}\int_0^{\pi} y \sin^2 t \mathrm{d}t \\ &= \frac{y}{2}.\end{aligned}$$

In this case we can even compute the change of coordinates:

$$\begin{aligned} x &= y + \varepsilon w \\ \frac{\mathrm{d}x}{\mathrm{d}t} &= \frac{\mathrm{d}y}{\mathrm{d}t} + \varepsilon \frac{\mathrm{d}w}{\mathrm{d}t} \\ \varepsilon x \sin^2 t &= \frac{\varepsilon y}{2} + \varepsilon \frac{\mathrm{d}w}{\mathrm{d}t} \\ \frac{\mathrm{d}w}{\mathrm{d}t} &= -\frac{y}{2} + y \sin^2 t + \mathcal{O}(\varepsilon) \\ w &= -\frac{y}{4}\sin(2t) + \mathcal{O}(\varepsilon).\end{aligned}$$

Therefore, under the coordinate transformation $x = y - \varepsilon \dfrac{y}{4}\sin(2t)$ we obtain:

$$\begin{aligned}\frac{\mathrm{d}x}{\mathrm{d}t} &= \left(1 - \frac{\varepsilon}{4}\sin(2t)\right)\frac{\mathrm{d}y}{\mathrm{d}t} - \varepsilon\frac{y}{2}\cos(2t) \\ \frac{\mathrm{d}y}{\mathrm{d}t} &= \frac{4}{4 - \varepsilon\sin(2t)}\left(\varepsilon(y+\varepsilon w)\sin^2 t + \varepsilon\frac{y}{2}\cos(2t)\right) \\ &= \frac{4}{4 - \varepsilon\sin(2t)}\left(\varepsilon\left(y - \varepsilon\frac{y}{4}\sin(2t) + \mathcal{O}(\varepsilon^2)\right)\left(\frac{1}{2} - \frac{1}{2}\cos(2t)\right) + \varepsilon\frac{y}{2}\cos(2t)\right) \\ &= \frac{4}{4 - \varepsilon\sin(2t)}\left(\varepsilon\frac{y}{2} + \varepsilon^2\frac{y}{8}\left(\cos(2t) - 1\right)\sin(2t) + \mathcal{O}(\varepsilon^3)\right) \\ &= \left(1 + \varepsilon\frac{\sin(2t)}{4} + \mathcal{O}(\varepsilon^2)\right)\left(\varepsilon\frac{y}{2} + \varepsilon^2\frac{y}{8}\left(\cos(2t) - 1\right)\sin(2t) + \mathcal{O}(\varepsilon^3)\right) \\ &= \varepsilon\frac{y}{2} + \varepsilon^2\frac{y\sin(4t)}{16} + \mathcal{O}(\varepsilon^3).\end{aligned}$$

See a comparison between the solution of the original equation and the averaged system.

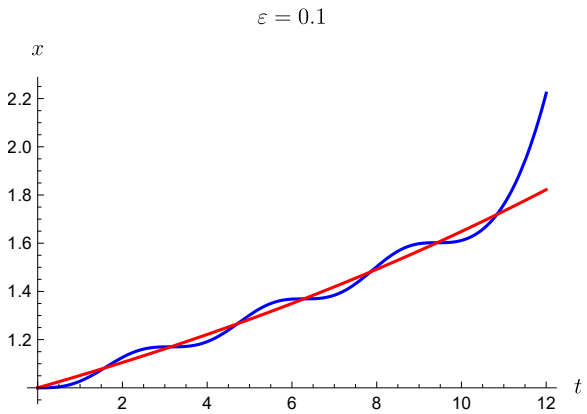


FIGURE 1. Comparison between the original equation and its average. A solution with $x(0) = 1$ of the original equation is plotted in blue, while for the average equation with the same initial condition, the solution is plotted in red. Notice that up to $t \sim 10 = \frac{1}{\varepsilon}$, both solutions are close to each other.

In the context of Hamiltonian systems, we are interested *almost integrable* systems, meaning that one looks at Hamiltonians of the form

$$H(I, \phi) = H_0(I) + \varepsilon h_1(\phi, I),$$

where $(I, \phi) \in \mathbb{R}^n \times \mathbb{T}^n$. The corresponding equations of motion are

$$\begin{aligned} \frac{\mathrm{d}\phi_i}{\mathrm{d}t} &= \frac{\partial H_0}{\partial I_i} + \varepsilon \frac{\partial H_1}{\partial I_i} \\ \frac{\mathrm{d}I_i}{\mathrm{d}t} &= -\varepsilon \frac{\partial H_1}{\partial \phi_i}. \end{aligned} \tag{87}$$

The motion prescribed by (87) is described by fast rotations (ϕ_i) with slow drifts (I_i) along the cylinder.

EXAMPLE VI.8. Let $H_0(I)$ denote the Hamiltonian of an integrable system with two degrees of freedom (depending only on the action). This kind of Hamiltonians appear, for instance, in some problems in celestial mechanics. Consider a small perturbation of the form

$$H(\phi_1, \phi_2, I_1, I_2, \varepsilon) = H_0(I_1) + \varepsilon H_1(\phi_1, \phi_2, I_1, I_2, \varepsilon). \tag{88}$$

For $\varepsilon = 0$, the equations of motion are simply

$$\begin{aligned} \frac{\mathrm{d}\phi_1}{\mathrm{d}t} &= \frac{\partial H_1}{\partial I_1} \\ \frac{\mathrm{d}\phi_2}{\mathrm{d}t} &= 0 \\ \frac{\mathrm{d}I_1}{\mathrm{d}t} &= 0 \\ \frac{\mathrm{d}I_2}{\mathrm{d}t} &= 0. \end{aligned}$$

That is, ϕ_1 rotates with constant speed $\omega(I_1) = \dfrac{\partial H_1}{\partial I_1}$ and all other variables are fixed. Assuming that $\omega(I_1) \neq 0$, and for $\varepsilon > 0$ sufficiently small, we can write

$$\begin{aligned} \frac{\mathrm{d}\phi_2}{\mathrm{d}\phi_1} &= \varepsilon \frac{\frac{\partial H_1}{\partial I_2}}{\omega(I_1) + \varepsilon \frac{\partial H_1}{\partial I_1}} \\ \frac{\mathrm{d}I_1}{\mathrm{d}\phi_1} &= -\varepsilon \frac{\frac{\partial H_1}{\partial \phi_1}}{\omega(I_1) + \varepsilon \frac{\partial H_1}{\partial I_1}} \\ \frac{\mathrm{d}I_2}{\mathrm{d}\phi_1} &= -\varepsilon \frac{\frac{\partial H_1}{\partial \phi_2}}{\omega(I_1) + \varepsilon \frac{\partial H_1}{\partial I_1}}. \end{aligned}$$

From (85) we have that the corresponding averaged system reads as:

$$
\begin{aligned}
\frac{\mathrm{d}\phi_2}{\mathrm{d}\phi_1} &= \varepsilon \frac{1}{2\pi\omega(I_1)} \int_0^{2\pi} \frac{\partial H_1}{\partial I_2}(\phi_1, \phi_2, I_1, I_2, 0)\mathrm{d}\phi_1 \\
\frac{\mathrm{d}I_1}{\mathrm{d}\phi_1} &= -\varepsilon \frac{1}{2\pi\omega(I_1)} \int_0^{2\pi} \frac{\partial H_1}{\partial \phi_1}(\phi_1, \phi_2, I_1, I_2, 0)\mathrm{d}\phi_1 = 0 \\
\frac{\mathrm{d}I_2}{\mathrm{d}\phi_1} &= -\varepsilon \frac{1}{2\pi\omega(I_1)} \int_0^{2\pi} \frac{\partial H_1}{\partial \phi_2}(\phi_1, \phi_2, I_1, I_2, 0)\mathrm{d}\phi_1
\end{aligned} \tag{89}
$$

We can take another, more intuitive approach. Let us consider the *averaged Hamiltonian*

$$
\bar{H}(\phi_2, I_1, I_2, \varepsilon) = \frac{1}{2\pi} \int_0^{2\pi} H(\phi_1, \phi_2, I_1, I_2, \varepsilon)\mathrm{d}\phi_1 = H_0(I_1) + \varepsilon \bar{H}_1(\phi_2, I_1, I_2, \varepsilon).
$$

The important observation here is that the averaged Hamiltonian does not depend on the ϕ_1 (the averaged fast variable). Notice that the corresponding equations of motion are

$$
\begin{aligned}
\frac{\mathrm{d}\phi_1}{\mathrm{d}t} &= \omega(I_1) + \varepsilon \frac{\partial \bar{H}_1}{\partial I_1} \\
\frac{\mathrm{d}\phi_2}{\mathrm{d}t} &= \varepsilon \frac{\partial \bar{H}_1}{\partial I_2} \\
\frac{\mathrm{d}I_1}{\mathrm{d}t} &= 0 \\
\frac{\mathrm{d}I_2}{\mathrm{d}t} &= -\varepsilon \frac{\partial \bar{H}_1}{\partial \phi_2}.
\end{aligned} \tag{90}
$$

Notice indeed in (89) that $\dfrac{\mathrm{d}I_1}{\mathrm{d}t} = 0$ because the right-hand side is the integral of a 2π-periodic function over the whole period.

Up to first order in ε the systems (89) and (90) are equivalent. That is "for Hamiltonian systems (88), the averaging method, with respect to the fast variable ϕ_1 can be applied directly to the Hamiltonian". This procedure leads to the constant of motion $\bar{H}$ and I_1. The previous, is true for time of order $\mathcal{O}(1/\varepsilon)$ from the averaging theorem. Notice also that via applying averaging, we have reduced the dimension of the problem. The original was 4-dimensional (2-DOF), but the averaged (90) has only 1-DOF (ϕ_2, I_2) because $\bar{H}$ is independent of ϕ_1 and I_1 is a parameter. The quantities $\bar{H}$ and I_1 are called *adiabatic invariants*. In general an adiabatic invariant is a quantity that changes sufficiently little on sufficiently large time intervals. A generalization of this "averaging in frequency" for systems of higher dimensions can be found in [**3**].

EXAMPLE VI.9 (Weakly nonlinear forced oscillations). Consider the second order equation

$$
\frac{\mathrm{d}^2\theta}{\mathrm{d}t^2} + \omega_0^2\theta = \varepsilon f\left(\theta, \frac{\mathrm{d}\theta}{\mathrm{d}t}, t\right), \tag{91}
$$

where $\theta \in \mathbb{S}^1$ and f is T-periodic in t. We let ω denote the frequency of the driving force f. For $\varepsilon = 0$, the system corresponds to an integrable Hamiltonian system with one degree of freedom θ. Let $x = \left(\theta, \dfrac{\mathrm{d}\theta}{\mathrm{d}t}\right)$. The solutions of the unperturbed problem are given by

$$
x(t) = \exp(B(\omega_0)t)x_0(0),
$$

where

$$\exp(B(\omega_0)t) = \begin{bmatrix} \cos(\omega_0 t) & -\sin(\omega_0 t) \\ -\omega_0 \sin(\omega_0 t) & -\omega_0 \cos(\omega_0 t) \end{bmatrix}.$$

Consider the change of coordinates $x = \exp(B(\omega/k)t)u$, where $u = (u_1, u_2)^\top$.
This transformation is called *the van der Pol transformation* and allows us to put the original system into one that is suitable for averaging. Indeed, in the new coordinates (91) reads as

$$\begin{aligned} \frac{\mathrm{d}u_1}{\mathrm{d}t} &= -\frac{k}{\omega}\left[\frac{\omega^2 - k^2\omega_0^2}{k^2}x + \varepsilon f\left(\theta, \frac{\mathrm{d}\theta}{\mathrm{d}t}, t, \varepsilon\right)\right]\sin\left(\frac{\omega}{k}t\right) \\ \frac{\mathrm{d}u_2}{\mathrm{d}t} &= -\frac{k}{\omega}\left[\frac{\omega^2 - k^2\omega_0^2}{k^2}x + \varepsilon f\left(\theta, \frac{\mathrm{d}\theta}{\mathrm{d}t}, t, \varepsilon\right)\right]\cos\left(\frac{\omega}{k}t\right), \end{aligned} \tag{92}$$

where for convenience we did not substitute x and $\dfrac{\mathrm{d}x}{\mathrm{d}t}$, but they can be written in terms of u via the above transformation. Suppose that $\omega^2 - k^2\omega_0^2 = \mathcal{O}(\varepsilon)$. Then (92) is indeed in the form required to perform averaging. The transformed system (92) allows us to study, via averaging for example, problems near resonances, that is when the driving frequency ω is close to the natural frequency ω_0.
To gain more insight, let us choose $f = \gamma\cos(\omega t) - \delta\dfrac{\mathrm{d}\theta}{\mathrm{d}t} - \alpha\theta^3$ corresponding to the *Duffing oscillator*. Let $\omega_0^2 - \omega^2 = \varepsilon\Omega$. Then, via the van der Pol transformation (setting $k = 1$) we get

$$\begin{aligned} \frac{\mathrm{d}u_1}{\mathrm{d}t} &= \frac{\varepsilon}{\omega}\left[\Omega(u_1C - u_2S) - \omega\delta(u_1S + u_2C) + \alpha(u_1C - u_2S)^3 - \gamma C\right]S \\ \frac{\mathrm{d}u_2}{\mathrm{d}t} &= \frac{\varepsilon}{\omega}\left[\Omega(u_1C - u_2S) - \omega\delta(u_1S + u_2C) + \alpha(u_1C - u_2S)^3 - \gamma C\right]C, \end{aligned} \tag{93}$$

where for brevity we use $C = \cos(\omega t)$ and $S = \sin(\omega t)$.
Averaging (93) over one period $T = \dfrac{2\pi}{\omega}$ we get

$$\begin{aligned} \frac{\mathrm{d}u_1}{\mathrm{d}t} &= \frac{\varepsilon}{2\omega}\left(-\omega\delta u_1 - \Omega u_2 - \frac{3}{4}\alpha(u_1^2 + u_2^2)u_2\right), \\ \frac{\mathrm{d}u_2}{\mathrm{d}t} &= \frac{\varepsilon}{2\omega}\left(-\omega\delta u_2 + \Omega u_1 + \frac{3}{4}\alpha(u_1^2 + u_2^2)u_1 - \gamma\right), \end{aligned}$$

or in polar coordinates (r, ϕ):

$$\begin{aligned} \frac{\mathrm{d}r}{\mathrm{d}t} &= \frac{\varepsilon}{2\omega}(-\omega\delta r - \gamma\sin\phi), \\ r\frac{\mathrm{d}\phi}{\mathrm{d}t} &= \frac{\varepsilon}{2\omega}(\Omega r + \frac{3}{4}\alpha r^3 - \gamma\cos\phi). \end{aligned} \tag{94}$$

If we fix the parameters (α, δ, γ) we can obtain (numerically) the equilibrium points of the averaged system as shown in figure 2.

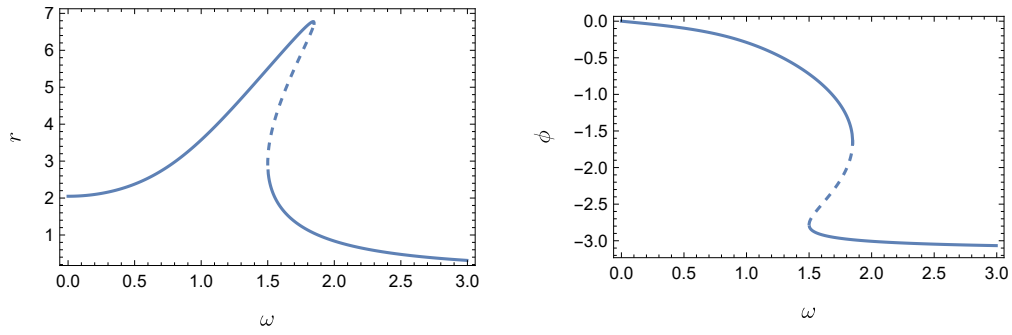


FIGURE 2. Numerically computed equilibrium points for the averaged system (94). The solid lines correspond to stable equilibria, while the dashed line corresponds to saddles. We have used the parameters: $\omega_0 = 1$, $\varepsilon\alpha = 0.07$, $\varepsilon\gamma = 2.5$, $\varepsilon\delta = .2$.

According to the averaging theorem, we would expect that solutions corresponding to fixed points of the averaged system are translated into stable periodic orbits. Compare with the phase-portraits in figure 3.

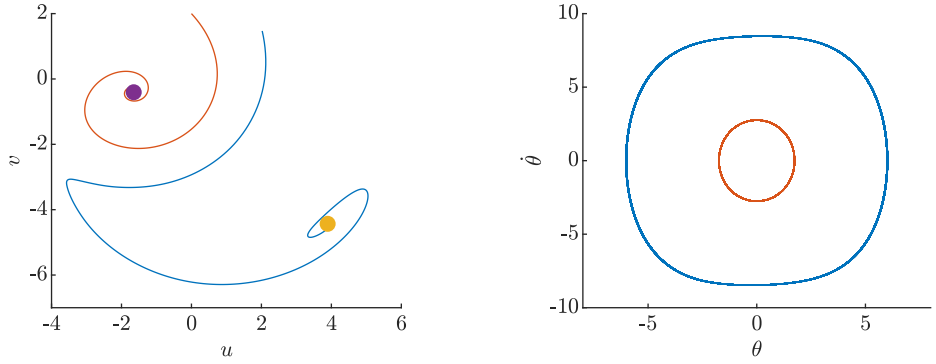


FIGURE 3. Left: phase-portrait of the averaged system for $\omega = 1.6$. In this region the averaged system 3 hyperbolic equilibria, two of which are stable. Right: the corresponding simulation for the Duffing oscillator. The shown periodic orbits are locally stable as predicted by the averaging theorem.

EXAMPLE VI.10. Consider the van der Pol equation

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = -x + \varepsilon(1 - x^2)\frac{\mathrm{d}x}{\mathrm{d}t}.$$

This model describes the motion of a harmonic oscillator with small nonlinear friction. The unperturbed equation is clearly $\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = -x$. Each orbit of this system is a circle given by $\left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)^2 + x^2 = h$, which are concentric circles of radius $\sqrt{h}$. Thus, we define I as

$$I = \frac{1}{2\pi}(\pi h) = \frac{h}{2} = \frac{\left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)^2 + x^2}{2}.$$

On the other hand, ϕ is defined by $\phi = \arg\left(x + \imath\dfrac{\mathrm{d}x}{\mathrm{d}t}\right)$, keeping in mind that we also can define equivalently the relations $x = \sqrt{2I}\cos\phi$, $\dfrac{\mathrm{d}x}{\mathrm{d}t} = \sqrt{2I}\sin\phi$. In this way the equations of motion in action angle variables read as

$$\frac{\mathrm{d}I}{\mathrm{d}t} = 2\varepsilon I(1 - I\cos^2\phi)\sin^2\phi$$
$$\frac{\mathrm{d}\phi}{\mathrm{d}t} = -1 + \varepsilon(1 - I\cos^2\phi)\sin\phi\cos\phi.$$

Averaging the I-equation we get

$$\frac{\mathrm{d}J}{\mathrm{d}t} = \frac{J}{2}\varepsilon(2 - J).$$

This equation has two equilibria: $J = 0$ and $J = 2$, which are hyperbolic. The equilibrium $J = 0$ is repelling, while the equilibrium $J = 2$ is attracting. We therefore conclude that for ε small, the original system has a stable limit cycle close to the circle $\left(\dfrac{\mathrm{d}x}{\mathrm{d}t}\right)^2 + x^2 = 4$, see Figure 4.

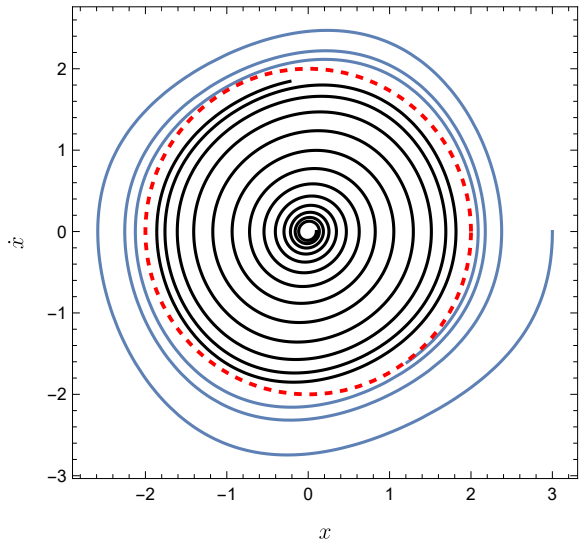


FIGURE 4. Phase portrait of the van der Pol equation for $\varepsilon = 0.1$. Notice that trajectories converge to a limit cycle that is close to the circle $\left(\dfrac{\mathrm{d}x}{\mathrm{d}t}\right)^2 + x^2 = 4$.

VI.2.1. Averaging and local bifurcations. In this section we briefly mention an important result relating the averaging theorem with bifurcations.

THEOREM VI.3. *Consider a μ-parameter family of equations*

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \varepsilon g(x, t, \varepsilon, \mu), \qquad \mu \in \mathbb{R}, \tag{95}$$

and its associated average

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \varepsilon \bar{g}(y, \mu). \tag{96}$$

If at $\mu = \mu_0$ (96) undergoes a saddle or a Hopf bifurcation, then for $\mu \sim \mu_0$ and ε sufficiently small, the Poincaré map of (95) undergoes a saddle or a Hopf bifurcation.

PROOF. See Theorem 4.3.1 in [10]. □

EXAMPLE VI.11. For the Duffing oscillator in the previous example, one may consider ω as the bifurcation parameter. We notice that for the averaged system, the equilibria undergo a saddle-node bifurcation as ω is varied (Figure 2) and so do the periodic orbits for ε small.

VI.2.2. Lie-Deprit series. We now explore another method applicable to Hamiltonian systems. In this method, we start with a analytic Hamiltonian

$$H(q,p,\varepsilon)=\sum_{k\geq 0}\frac{\varepsilon^k}{k!}H_k(p,q). \tag{97}$$

The overall idea is to carry out a near identity transformation $(q,p)\mapsto(q,p)+W(q,p,\varepsilon)$, such that up to some order the transformed Hamiltonian is integrable.

The function W is also assumed to be analytic and given by

$$W(q,p,\varepsilon)=\sum_{k\geq 0}\frac{\varepsilon^k}{k!}W_k(p,q),$$

where the W_k's are yet to be determined. The idea to obtain such function W is for it to satisfy the system (observe that it has a Hamiltonian structure)

$$\begin{aligned}\frac{\mathrm{d}q}{\mathrm{d}\varepsilon}&=\frac{\partial W}{\partial p}\\ \frac{\mathrm{d}p}{\mathrm{d}\varepsilon}&=-\frac{\partial W}{\partial q},\end{aligned} \tag{98}$$

with $q(0)=Q$ and $p(0)=P$. Of course, for $\varepsilon=0$ we simply have the identity transformation. Assuming that we know the W_n's the problem now is to express H in terms of (Q,P). We now describe how this is done.

Consider for a moment a Hamiltonian $H=H(q,p)$ and $f=f(q,p)$ as smooth function. We define *the Lie derivative* generated by H as the map

$$\mathcal{L}_H : f\mapsto\{f,H\}.$$

We further make use of the notation $\mathcal{L}_H^k(f)=\mathcal{L}_H(\mathcal{L}_H^{k-1}(f))$, where $\mathcal{L}_H^0(f)=f$.

Notice that for an analytic function $f(p,q)$ we have

$$\frac{\mathrm{d}f}{\mathrm{d}t}(q,p)=\left(\sum_{i=1^n}\frac{\partial f}{\partial q_i}\frac{\mathrm{d}q_i}{\mathrm{d}t}\right)(q,p)=\frac{\partial f}{\partial p_i}\frac{\mathrm{d}p_i}{\mathrm{d}t}(q,p)=\{f,H\}(q,p)=\mathcal{L}_H(f)(q,p).$$

In fact, $\dfrac{\mathrm{d}^k f}{\mathrm{d}t^k}=\mathcal{L}_H^k(f)$. Thus, Taylor's formula can be rewritten as

$$f(q,p)=\sum_{k\geq 0}\frac{t^k}{k!}\mathcal{L}_H^k(f)(q(0),p(0))=\exp\left(t\mathcal{L}_H\right)(f)(q(0),p(0)),$$

for t small. If f is analytic, then the above series converges for t sufficiently small.

If f explicitly depends on time, that is $f=f(q,p,t)$, then

$$\frac{\mathrm{d}f}{\mathrm{d}t}(q,p,t)=\{f,H\}+\frac{\partial f}{\partial t}.$$

For brevity, let us use the notation $\Delta_H(f) = \{f, H\} + \dfrac{\partial f}{\partial t}$, and similar to $\mathcal{L}_H^k$, let $\Delta_H^k(f) = \Delta_H(\Delta_H^{k-1}(f))$. Thus

$$f(q,p,t) = \frac{t^k}{k!}\Delta_H^k(f)(q(0),p(0)) = \exp(t\Delta_H)(f)(q(0),p(0)).$$

With the notation introduced so far we then have that the solutions of (98) can be give as

$$q(Q,P,\varepsilon) = \exp(\varepsilon\Delta_W)(Q)$$
$$p(Q,P,\varepsilon) = \exp(\varepsilon\Delta_W)(P).$$

The advantage of the previous notation is that the inverse transformation is simply

$$Q(q,p,\varepsilon) = \exp(-\varepsilon\Delta_W)(q)$$
$$P(q,p,\varepsilon) = \exp(-\varepsilon\Delta_W)(p).$$

PROPOSITION VI.4. *Consider an analytic function*

$$f(q,p,\varepsilon) = \sum_{k\geq 0}\frac{\varepsilon^k}{k!}f_k(q,p).$$

Let (Q,P) be new variables defined by (98) such that

$$W(q,p,\varepsilon) = \sum_{k\geq 0}\frac{\varepsilon^k}{k!}W_k(q,p).$$

Then, if $F = F(Q,P,\varepsilon)$ denotes the transformation of f into the new coordinates, that is $F(Q,P,\varepsilon) = f(q(Q,P),p(Q,P),\varepsilon)$, then

$$F = \sum_{k\geq 0}\frac{\varepsilon^k}{k!}F_k(Q,P),$$

where to determine the F_k's one proceeds as follows: let $f_k^0(Q,P) = f_k(Q,P)$, and define functions $f_k^j(Q,P)$ recursively as

$$f_k^j(Q,P) = f_{k+1}^{j-1} + \sum_{m=0}^{k}\binom{k}{m}\left\{f_{k-m}^{j-1}, W_m\right\}(Q,P).$$

Then $F_k(Q,P) = f_0^k(Q,P)$.

For Hamiltonian systems with H given as in (97) we can follow the diagram
to obtain the transformed Hamiltonian $K(Q,P,\varepsilon)$. For example

$$\begin{aligned} K_0(Q,P) &= H_0(Q,P)\\ K_1(Q,P) &= H_1(Q,P) + \{H_0, W_0\}(Q,P)\\ K_2(Q,P) &= H_2(Q,P) + 2\{H_1, W_0\}(Q,P) + \{H_0, W_1\}(Q,P) + \{\{H_0, W_0\}, W_0\}(Q,P) \end{aligned} \tag{99}$$

where, in practice, we choose W_1 such that K_1 is as simple as possible. One proceeds similarly for the higher order terms. Notice that this algorithm is suitable for symbolic computations.

EXAMPLE VI.12. We consider again a Hamiltonian of the form

$$H(\phi_1,\phi_2,I_1,I_2,\varepsilon) = H_0(I_1) + \sum_{k\geq 1}\frac{\varepsilon^k}{k!}H_k(\phi_1,\phi_2,I_1,I_2),$$

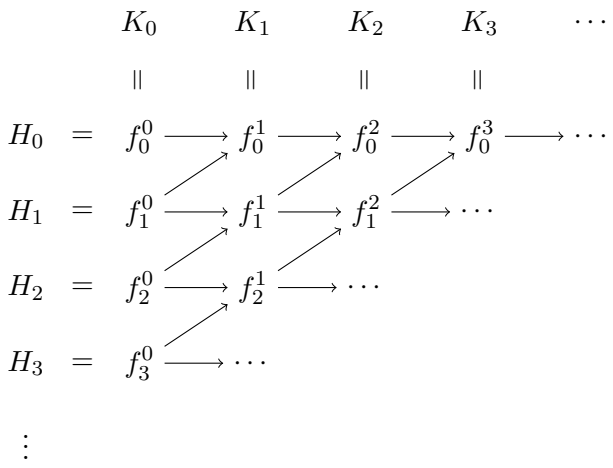


FIGURE 5. A diagram to compute the higher order terms of the transformed Hamiltonian $K = K_0 + \varepsilon K_1 + \cdots$.

where $\omega_0(I) = \dfrac{\partial H_0}{\partial I_1} \neq 0$. Our goal is to determine W such that in the transformed Hamiltonian $K(\psi, J)$ the dependence on ϕ_1 is eliminated. We use (J, ψ) as the new action-angle variables. From (99) we see that since H_0 does not depend on ϕ_1 neither does K_0. For K_1 we have

$$K_1(\psi, J) = H_1(\psi, J) - \frac{\partial H_0}{\partial J_1}\frac{\partial W_0}{\partial \psi_1}(\psi, J).$$

Ideally we would like K_1 to be independent of ψ_1. We can achieve this, for example, if W_0 is periodic in ψ_1. We can do that if we define

$$W_0(\psi_1, \psi_2, J_1, J_2) = \frac{1}{\omega_0(J_1)} \int_0^{\psi_1} \big(H_1(\theta_1, \psi_2, J_1, J_2) - \bar{H}_1(\psi_2, J_1, J_2)\big)\, \mathrm{d}\theta_1,$$

where $\bar{H}_1$ denotes the average of H_1 over ψ_1. Under such a choice of W_0 we get

$$K_1(\psi_1, \psi_2, J_1, J_2) = \bar{H}_1(\psi_2, J_1, J_2).$$

So, up to the first order, the Lie-Deprit method coincides with the averaging method. However, if one would continue with the computations, we can use similar arguments as before to find that

$$K(\psi_1, \psi_2, J_1, J_2) = K_0(J_1) + \sum_{k=1}^{n} \varepsilon^k K_k(\psi_2, J_1, J_2) + \varepsilon^{n+1} R(\psi_1, \psi_2, J_1, J_2).$$

The corresponding equations of motion have, in particular, the term $\dfrac{\mathrm{d}J_1}{\mathrm{d}t} = \mathcal{O}(\varepsilon^{n+1})$. This shows that J_1 is an adiabatic invariant for time of order $\mathcal{O}(1/\varepsilon^{n+1})$.

EXAMPLE VI.13. Consider the 1-DOF system defined by

$$H = \frac{1}{2}p^2 + \frac{1}{2}\omega_0^2 q^2 + \frac{\varepsilon}{4}\omega_0^2 q^4 + \frac{\varepsilon^2}{8} a \omega_0^3 q^6,$$

which models a weakly forced an-harmonic pendulum. The unperturbed Hamiltonian $H_0 = \dfrac{1}{2}p^2 + \dfrac{1}{2}\omega_0^2 q^2$ corresponds to a simple oscillator. We can define action angle variables (I, ϕ)

through the relations

$$q = \sqrt{\frac{2I}{\omega_0}}\cos\phi$$
$$p = -\sqrt{2\omega_0 I}\sin\phi.$$

In these new coordinates the Hamiltonian reads as

$$H(\phi, I) = \omega_0 I + \varepsilon \underbrace{I^2\cos^4\phi}_{H_1} + \varepsilon^2 \underbrace{aI^3\cos^6\phi}_{H_2}.$$

We now attempt to obtain a simplified Hamiltonian via the Lie-Deprit method. We already know that $K_0 = H_0 = H_0(I)$. For the first perturbation term we have

$$\begin{aligned} K_1 &= H_1 + \{H_0, W_0\} \\ &= H_1 - \frac{\partial H_0}{\partial J}\frac{\partial W_0}{\partial \psi} \\ &= I^2\cos^4\phi - \omega_0\frac{\partial W_0}{\partial \psi} \end{aligned}$$

Naturally, one could think that the best is to choose $K_1 = 0$. The problem with such a choice is that W_0 would have secular terms, thus being unbounded as $t \to \infty$. Instead, let $K_1 = \bar{H}_1 = \frac{1}{\pi}\int_0^\pi H_1 \mathrm{d}\phi = \frac{3}{8}I^2$. In this way we have

$$W_0 = \frac{1}{\omega_0}\int_0^\psi \left(I^2\cos^4\phi - \frac{3}{8}I^2\right)\mathrm{d}\phi = \frac{I^2}{32\omega_0}(8\sin(2\psi) + \sin(4\psi)).$$

With such a choice of W_0 we now have that the new Hamiltonian reads as

$$K = K_0 + \varepsilon K_1 + \mathcal{O}(\varepsilon^2) = \omega_0 J + \frac{3}{8}\varepsilon J^2 + \mathcal{O}(\varepsilon^2).$$

Thus, for the truncated Hamiltonian $K = \omega_0 J + \frac{3}{8}\varepsilon J^2$ the equations of motion are

$$\frac{\mathrm{d}J}{\mathrm{d}t} = 0$$
$$\frac{\mathrm{d}\psi}{\mathrm{d}t} = \omega_0 + \frac{3}{4}\varepsilon J.$$

The corresponding solutions are simply

$$J(t) = J(0)$$
$$\psi(t) = \psi(0) + \left(\omega_0 + \frac{3}{4}\varepsilon J(0)\right)t.$$

Let us further compute K_2. Recall that $K_2 = H_2 + 2\{H_1, W_0\} + \{H_0, W_1\} + \{\{H_0, W_0\}, W_0\}$, therefore, after some computations:

$$K_2 = \frac{J^3(3(5a\omega_0 - 11)\cos(2\psi) + 6(a\omega_0 - 1)\cos(4\psi) + (a\omega + 1)\cos(6\psi) + 10a\omega_0 - 17)}{32\omega_0} - \omega_0\frac{\partial W_1}{\partial \psi}.$$

Now we can simply choose

$$K_2 = \frac{(10a\omega_0 - 17)J^3}{32\omega_0}.$$

Indeed in that case one can check that W_1 is bounded. Therefore, up to quadratic terms we have the Hamiltonian

$$K = \omega_0 J + \frac{3}{8}\varepsilon J^2 + \varepsilon^2 \frac{(10a\omega_0 - 17)J^3}{32\omega_0} + \cdots,$$

with corresponding evolution given by

$$J(t) = J(0)$$
$$\psi(t) = \psi(0) + \left(\omega_0 + \frac{3}{4}\varepsilon J + \varepsilon^2 \frac{3(10a\omega_0 - 17)J^2}{32\omega_0}\right) t$$

VI.3. KAM Theory

In this section we describe some basic results related to the asymptotic behavior of small perturbations of integrable systems. We shall restrict ourselves to 2-DOF.

Thus, let us consider $H = H(q_1, q_2, p_1, p_2)$, $(q, p) \in \mathbb{R}^2 \times \mathbb{R}^2$. We assume that the Hamiltonian system admits a periodic solution $\gamma(t)$ and our goal is to determine its stability. Recall that $H = h$ is a conserved quantity. This means that the flow of the Hamiltonian system "lives" in an invariant 3-dimensional manifold.

Without loss of generality, let us assume that q_1 is an angle variable. We now know that this can be achieved by a canonical transformation. Moreover, we assume that[3], locally in a neighborhood of γ,

$$\frac{\mathrm{d}q_1}{\mathrm{d}t} = \frac{\partial H}{\partial p_1} > 0.$$

By the inverse function theorem, we know that H can be inverted with respect to p_1 so that

$$p_1 = P(H, q_1, q_2, p_2).$$

Consequently, we notice now that the invariant manifolds where the trajectories evolve can be parametrized by only (q_1, p_1, p_2). Let us now take a Poincaré section at some value $q_1 = q_1^*$. Since $\frac{\mathrm{d}q_1}{\mathrm{d}t} > 0$, the flow is transverse to any section $q_1 =$constant. Thus, without loss of generality, let the section be at $\Sigma = \{q_1 = 0 = 0 \mod 2\pi\}$ and consider the Poincaré map

$$\Pi : (q_2, p_2, q_1 = 0) \mapsto (q_2, p_2, q_1 = 2\pi).$$

This map is, effectively, 2-dimensional and depends on the constant $H = h$. What we are doing is eliminating the time dependence because we can use q_1 to measure (a rescaled) time. Indeed, noticing that $H(q_1, q_2, P, p_2) = h$ implies $\frac{\partial H}{\partial q_2} = \frac{\partial H}{\partial P}\frac{\partial P}{\partial q_2}$ and $\frac{\partial H}{\partial p_2} = \frac{\partial H}{\partial P}\frac{\partial P}{\partial p_2}$ we can now write the equations of motion as

$$\frac{\mathrm{d}q_2}{\mathrm{d}q_1} = \frac{\frac{\partial H}{\partial p_2}}{\frac{\partial H}{\partial p_1}} = -\frac{\partial P}{\partial p_2}$$
$$\frac{\mathrm{d}p_2}{\mathrm{d}q_1} = -\frac{\frac{\partial H}{\partial q_2}}{\frac{\partial H}{\partial p_1}} = \frac{\partial P}{\partial q_2}.$$

This h-family of Hamiltonian systems is called *the reduced Hamiltonian system*. For this reduced system P plays the role of a time-dependent (in reality q_1-periodic) Hamiltonian. We thus know that the Poincaré map Π is area-preserving. In practical terms, this means that $|\det(J\Pi)| = 1$. Let

[3] The case $\frac{\partial H}{\partial p_1} < 0$ is completely analogous.

$x^* = (q_2^*, p_2^*)$ denote $x^* = \gamma \cap \Sigma$. Since the map is 2-dimensional, we have the following possibilities for the fixed point:

Hyperbolic : in this case $\dfrac{\partial \Pi}{\partial x}(x^*)$ has real eigenvalues $\lambda_1 \neq \pm 1$ and λ_2 so that $\lambda_1 \lambda_2 = \pm 1$. Thus the fixed point is hyperbolic.

Elliptic: in this case $\dfrac{\partial \Pi}{\partial x}(x^*)$ has complex eigenvalues $\exp(\pm 2\pi\imath\theta) \neq \pm 1$. The fixed point is thus called elliptic. This is the case we are going to consider further.

Parabolic:: in this case $\dfrac{\partial \Pi}{\partial x}(x^*)$ has eigenvalues ± 1. This situation usually arises in bifurcations.

EXAMPLE VI.14. Consider the linear 2-DOF Hamiltonian system defined by

$$H(q_1, p_1, q_2, p_2) = \frac{p_1^2 + \omega_1^2 q_1^2}{2} + \frac{p_2^2 + \omega_2^2 q_2^2}{2}.$$

This is the Hamiltonian of an uncoupled system

$$\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} + \omega_1^2 x = 0$$
$$\frac{\mathrm{d}^2 y}{\mathrm{d}t^2} + \omega_2^2 y = 0,$$

with nonzero natural frequencies ω_1 and ω_2.
For $h > 0$, each level set $\{H = h\}$ is bounded, let us use action angle coordinates (I, θ) leading to the relation

$$q_2 = \sqrt{\frac{2I}{\omega_2}} \sin\theta$$
$$p_2 = \sqrt{2\omega_2 I} \cos\theta.$$

These coordinates are chosen so that $\dfrac{\mathrm{d}I}{\mathrm{d}t} = 0$ and $\dfrac{\mathrm{d}\theta}{\mathrm{d}t} = \omega_2$. In these new coordinates the Hamiltonian reads as

$$H(q_1, p_1, I, \theta) = \frac{p_1^2 + \omega_1^2 q_1^2}{2} + \omega_2 I = h.$$

Since $\omega_2 \neq 0$ it follows that

$$I = \frac{1}{\omega_2}\left(h - \frac{p_1^2 + \omega_1^2 q_1^2}{2}\right).$$

From here we find that the reduced Hamiltonian system reads as

$$\frac{\mathrm{d}q_1}{\mathrm{d}\theta} = \frac{p_1}{\omega_2}$$
$$\frac{\mathrm{d}p_1}{\mathrm{d}\theta} = -\frac{\omega_1^2}{\omega_2} q_1, \tag{100}$$

which is well-defined whenever $p_1^2 + \omega_1^2 q_1^2 < 2h$. Since the equations do not depend explicitly in θ, we can compute the Poincaré map easily. The solutions of (100), for a section at $\theta = 0$, are

given by

$$q_1(\theta) = q_1(0)\cos\left(\frac{\omega_1}{\omega_2}\theta\right) + \frac{\omega_2}{\omega_1}p_1(0)\sin\left(\frac{\omega_1}{\omega_2}\theta\right)$$
$$p_1(\theta) = -\frac{\omega_1}{\omega_2}q_1(0)\sin\left(\frac{\omega_1}{\omega_2}\theta\right) + p_1(0)\cos\left(\frac{\omega_1}{\omega_2}\theta\right)$$

Then, the Poincaré map reads as

$$\Pi(q_1,p_1) = \begin{bmatrix} \cos\left(\frac{2\pi\omega_1}{\omega_2}\right) & \frac{\omega_2}{\omega_1}\sin\left(\frac{2\pi\omega_1}{\omega_2}\right) \\ -\frac{\omega_1}{\omega_2}\sin\left(\frac{2\pi\omega_1}{\omega_2}\right) & \cos\left(\frac{2\pi\omega_1}{\omega_2}\right) \end{bmatrix} \begin{bmatrix} q_1 \\ p_1 \end{bmatrix}.$$

The map Π has a unique fixed point $x^* = (q_1^*, p_1^*) = (0,0)$ of elliptic type. This means that x^* is surrounded by invariant curves, which are intersections of the 2-torus with Σ. The Poincaré map is periodic if $\frac{\omega_1}{\omega_2} \in \mathbb{Q}$, and is dense (fills the invariant circle densely) if $\frac{\omega_1}{\omega_2} \in \mathbb{R}\backslash\mathbb{Q}$. This means that the full system has invariant 2-tori, with periodic orbits in the fist case, and dense orbits in the second.

As we have seen in the previous example, the linearization of Π at the fixed point x^* is a rotation of angle $2\pi\theta$ (up to a rescaling for θ). In complex coordinates the Poincaré map reads as

$$z \mapsto \hat{z} = \exp(2\pi\imath\theta)z + g(z,\bar{z}), \tag{101}$$

where g is of the form

$$g(z,\bar{z}) = \sum_{u+v=2}^{r} g_{uv} z^u \bar{z}^v + \mathcal{O}(|z|^{r+1}),$$

assuming that H is of class $\mathcal{C}^r$.

If the term g was zero, the map $z \mapsto \exp(2\pi\imath\theta)z$ would be a rotation, implying that the periodic orbit γ is stable. It is therefore natural to check whether we can eliminate the term g by a change of coordinates. Let us then consider a near identity transformation

$$z = \zeta + \alpha\zeta^j\bar{\zeta}^k, \qquad 2 \le j+k \le r.$$

Notice that $g(z,\bar{z}) = g(\zeta,\bar{\zeta}) + \mathcal{O}(|\zeta|^{j+k+1})$. Therefore, the map (101) becomes:

$$\hat{z} = \hat{\zeta} + \alpha\hat{\zeta}^j\bar{\hat{\zeta}}^k = \exp(2\pi\imath\theta)\zeta + \alpha\exp(2\pi\imath\theta)\zeta^j\bar{\zeta}^k + g(\zeta,\bar{\zeta}) + \mathcal{O}(|\zeta|^{j+k+1}).$$

In particular, the previous equation tells us that $\hat{\zeta} = \exp(2\pi\imath\theta)\zeta + \mathcal{O}(|\zeta|^2)$. It follows that $\bar{\hat{z}} = \exp(-2\pi\imath\theta)\bar{\zeta}$ and therefore

$$\hat{\zeta} = \exp(2\pi\imath\theta)\zeta + \alpha\left(\exp(2\pi\imath\theta) - \exp(2\pi\imath(j-k)\theta)\right)\zeta^j\bar{\zeta}^k + \sum_{u+v=2}^{r} g_{uv}\zeta^u\bar{\zeta}^v + \mathcal{O}(|\zeta|^{j+k+1}).$$

Thus, we see that we can eliminate the monomial $g_{jk}\zeta^j\bar{\zeta}^k$ if we choose

$$\alpha = \frac{g_{jk}}{\exp(2\pi\imath(j-k)\theta) - \exp(2\pi\imath\theta)} = \frac{g_{jk}\exp(-2\pi\imath\theta)}{\exp(2\pi\imath(j-k-1)\theta) - 1}.$$

Naturally, such a transformation would introduce new terms, but those are of higher order. Therefore, one can proceed at each degree (first eliminating terms of degree two, then those of degree three,

etc.) as long as

$$\exp(2\pi\imath(j-k-1)\theta) \neq 1.$$

This condition reminds us of those resonances in the chapter of normal forms. Indeed, the terms $g_{jk}z^j\bar{z}^k$ such that $\exp(2\pi\imath(j-k-1)\theta) = 1$ are called resonant, and cannot be eliminated via the proposed change of coordinates. We now distinguish two types of resonances:

(1) The terms $g_{(k+1)k}z^{k+1}\bar{z}^k = g_{(k+1)k}|z|^{2k}z$.
(2) If $\theta = \dfrac{p}{q} \in \mathbb{Q}$, then the terms for which $j-k-1 = nq$, $n \in \mathbb{N}$, are also resonant. That is, the terms $g_{(nq+k+1)k}|z|^{2k}z^{nq+1}$ cannot be eliminated via the normal form procedure.

For the case where θ is irrational, the Poincaré map (101) reads as

$$\hat{\zeta} = \exp(2\pi\imath\theta)\zeta + \sum_{j=1}^{m} c_j|\zeta|^{2j}\zeta + \mathcal{O}(|\zeta|^{r+1}),$$

where $m = \left[\dfrac{r}{2} - 1\right]$. Such a map is known as the *Birkhoff normal form*. Furthermore, an elliptic point is called *non-degenerate* if the first coefficient of the corresponding Birkhoff normal form, c_1, is nonzero.

Let us take another look at the term α. Notice that since the denominator is of the form $\exp(2\pi\imath n\theta) - 1$, for some integer n, such a term may become arbitrarily small even in the irrational case (this is because for irrational θ all numbers $\exp(2\pi\imath n\theta)$ are dense in the unit circle). This problem is referred to as *small denominators*.

In the following, let us assume that the Hamiltonian is at least 4-times differentiable, and that $\exp(2\pi\imath n\theta) \neq 1$ for $n = 1, 2, 3, 4$. In this case the Birkhoff normal form reads as

$$\hat{\zeta} = \exp(2\pi\imath\theta)\zeta + c_1|\zeta|^2\zeta + \mathcal{O}(|\zeta|^5),$$

Let action-angle coordinates (I, ϕ) be defined $\zeta = I\exp(\imath\phi)$. Then

$$\begin{aligned}\hat{\zeta} = \hat{I}\exp(\imath\hat{\phi}) &= \exp(2\pi\imath\theta)I\exp(\imath\phi) + c_1I^3\exp(\imath\phi) + \cdots \\ &= \exp(2\pi\imath\theta)\exp(\imath\phi)I(1 + c_1I^2\exp(-2\pi\imath\theta)) + \cdots .\end{aligned}$$

It follows that

$$|\hat{\zeta}|^2 = \hat{I}^2 = I^2 + 2I^4\Re(c_1\exp(-2\pi\imath\theta)) + \cdots . \tag{102}$$

Using such an expression for $\hat{I}$ we can also compute

$$\exp(\imath\hat{\phi}) = \exp(2\pi\imath\theta + \imath\phi)(1 + \imath I^2\Im(c_1\exp(-2\pi\imath\theta))) + \cdots . \tag{103}$$

From (102) and (103) we obtain[4]:

$$\begin{aligned}\hat{\phi} &= \phi + 2\pi\theta + \Im(c_1\exp(-2\pi\imath\theta))I^2 + \cdots , \\ \hat{I} &= I + \Re(c_1\exp(-2\pi\imath\theta))I^3 + \cdots .\end{aligned} \tag{104}$$

Since the Poincaré map is area-preserving, it holds that $\Re(c_1\exp(-2\pi\imath\theta)) = 0$. Thus we further simplify (104) to

$$\begin{aligned}\hat{\phi} &= \phi + \Omega(I) + f(\phi, I) \\ \hat{I} &= I + g(\phi, I),\end{aligned} \tag{105}$$

[4]Expanding the square root of (102) in I and taking logarithm in (103)

where[5] $\Omega(I) = 2\pi\theta + \dfrac{c_1 \exp(-2\pi\imath\theta)}{\imath} I^2$, and f and g are higher-order terms. Naturally, if the system would have higher differentiability, the higher-order terms would be smaller, and Ω would include some extra terms from the normal form.

We now return to the issue of small denominators. Notice that if one is to expect that the successive changes of coordinates converge, one would require that the small denominators are bounded away from 0. In other words, one would expect that θ is "badly approximated" by rationals. These numbers are called *Diophantine*.

DEFINITION VI.6 (Diophantine numbers). A number $\omega \in \mathbb{R}\backslash\mathbb{Q}$ is called *Diophantine* of type (C, τ) for some real numbers $C > 0$, $\tau \geq 1$ (possibly depending on ω) if the inequality

$$\left|\omega - \frac{m}{n}\right| \geq \frac{C}{|n|^{1+\tau}},$$

holds for all relatively prime numbers[6] (n, m), $n \neq 0$.

A particular class of Diophantine numbers are the *algebraic numbers*.

DEFINITION VI.7 (Algebraic number). An irrational number ω is called *algebraic* of order $n \geq 2$ if there exists and n-degree polynomial with integer coefficients $P(z) = a_n z^n + \cdots + a_1 z + a_0$, with $a_n \neq 0$ such that $P(\omega) = 0$.

For example, $\omega = \sqrt{2}$ is an algebraic number, it is a solution of $P(z) = z^2 - 2$. Liouville proved that algebraic numbers are indeed Diophantine. On the other hand, there are Diophantine numbers that are not algebraic. But in fact, it even holds that most real numbers are Diophantine [4].

LEMMA VI.1 ([4]). *Let $\tau \geq 1$. For almost every real ω there exists a $C = C(\omega, \tau)$ such that*

$$\left|\omega - \frac{m}{n}\right| \geq \frac{C}{|n|^{1+\tau}}.$$

Given a Diophantine number ω of type (C, τ) we can compute a lower bound for the small denominator. First, notice that for $n \neq 0$

$$|\exp(2\pi\imath n\omega) - 1|^2 = 2 - 2\cos(2\pi n\omega) = 4\sin^2(\pi n\omega),$$

and let m be the integer closest to $n\omega$. Therefore, $\sin(\pi n\omega) = \sin(\pi n\omega - \pi m)$, and noticing that $|\sin\theta| \geq \dfrac{2}{\pi}|\theta|$ for $|\theta| \leq \dfrac{\pi}{2}$ we have:

$$|\exp(2\pi\imath n\omega) - 1| = 2|\sin(\pi n\omega - \pi m)| \geq 4|n\omega - m| \geq 4\frac{C}{|n|^\tau}.$$

Coming back to (105) we notice that if f and g are zero, namely for the system

$$\begin{aligned} \hat{\phi} &= \phi + \Omega(I) \\ \hat{I} &= I, \end{aligned} \tag{106}$$

then the dynamics consist of simple rotations because I is constant, and ϕ increases with an amount depending on I. If (106) was structurally stable, then we would be able to find a coordinate change $(\phi, I) \mapsto (\psi, J)$ such that the dynamics of (105) reduce to simple rotations. This, in general, is not possible. However, the following theorem due to Moser tells us that for certain initial conditions, such a reduction is possible.

[5] This expression for Ω follows from $\Re(c_1 \exp(-2\pi\imath\theta)) = 0$.

[6] Relatively prime numbers are integers that do not have common factors other than 1.

THEOREM VI.4 (Moser [25, 12]). *Assume that the map (106) is $\mathcal{C}^r$-smooth, $r \geq 4$, in a strip $a \leq I \leq b$. Assume that $\Omega(I)$ satisfies (the so-called twist condition)*

$$\frac{\mathrm{d}\Omega}{\mathrm{d}I} \geq W > 0,$$

for $a \leq I \leq b$. Then, for every $\delta > 0$ there exists an $\varepsilon > 0$ (depending on δ and r) such that if $\omega \in [\Omega(a) + C, \Omega(b) - C]$ is Diophantine of type (C, τ) for some $\tau \in \left(1, \frac{r-1}{2}\right)$, $C > 0$, and $||f||_{\mathcal{C}^r} + ||g||_{\mathcal{C}^r} < \varepsilon W C^2$, then the perturbed map (105) admits an invariant curve of the form

$$\phi = \psi + u(\psi, \omega)$$
$$I = \Omega^{-1}(2\pi\omega) + v(\psi, \omega),$$

with $0 \leq \psi \leq 2\pi$ and where u, v are 2π-periodic in ψ and differentiable satisfying

$$||u||_{\mathcal{C}^1} + ||v||_{\mathcal{C}^1} < \delta.$$

The dynamics in such an invariant curve is given by the (circle) map

$$\psi \mapsto \hat{\psi} = \psi + 2\pi\omega.$$

REMARK VI.1. In the above theorem, the norm is defined as

$$||f||_{\mathcal{C}^r} = \sup_{\substack{\phi \\ a \leq I \leq b}} \max_{i+j \leq r} \left| \frac{\partial^{i+j} f}{\partial \phi^i \partial I^j} \right|.$$

An important consequence of Moser's theorem is the following.

COROLLARY VI.1. *Let x^* be an elliptic fixed point of a $\mathcal{C}^4$ area-preserving map Π in the plane. Assume that $\dfrac{\partial \Pi}{\partial x}(x^*)$ has eigenvalues $\lambda = \exp(2\pi\imath\theta)$ and $\bar{\lambda}$ such that $|\lambda| = 1$ and $\lambda^q \neq 1$ for $q = 1, 2, 3, 4$. Assume further that the coefficient c_1 of the corresponding Birkhoff normal form is not zero. Then x^* is stable.*

PROOF. Before starting it is worth noticing that the non-resonant condition is "only" excluding eigenvalues $\lambda = \pm 1$, $\lambda = \pm\imath$ and $\lambda = \exp\left(\pm\frac{2\pi}{3}\imath\right)$.

From the analysis we have performed above, we know that the dynamics of the map Π near x^* are given by

$$\hat{\phi} = \phi + \Omega(I) + f(\phi, I)$$
$$\hat{I} = I + g(\phi, I).$$

Consider a small strip $\varepsilon \leq I \leq 2\varepsilon$[7]. Let J be defined by $I = \varepsilon J$. Then

$$\hat{\phi} = \phi + \Omega(\varepsilon J) + \tilde{f}(\phi, J)$$
$$\hat{J} = J + \tilde{g}(\phi, J),$$

where $\tilde{f}(\phi, J) = f(\phi, \varepsilon I)$ and $\tilde{g}(\phi, J) = \dfrac{1}{\varepsilon} g(\phi, \varepsilon I)$. For this map the domain is $1 \leq J \leq 2$. Furthermore,

$$\left| \frac{\mathrm{d}}{\mathrm{d}J} \Omega(\varepsilon J) \right| = 2\varepsilon^2 |c_1| J \geq 2\varepsilon^3 |c_1|.$$

[7]Notice that in this strip f and g are indeed $\mathcal{C}^4$ even though at $I = 0$ f may only be $\mathcal{C}^3$ due to the singularity of the polar change of coordinates at the origin

Since $c_1 \neq 0$, then Moser's theorem ensures the existence of invariant curves in the strip, implying the stability of x^*.

□

Let us see now a specific example.

EXAMPLE VI.15. Consider the standard map given by

$$\begin{aligned} \hat{\phi} &= \phi + I + \varepsilon \sin\phi \\ \hat{I} &= I + \varepsilon \sin\phi, \end{aligned} \tag{107}$$

with $\phi \in [0, 2\pi)$.

This discrete map arises as the Poincaré map of the "kicked rotator" [21].

Notice that the phase-space is, topologically, a cylinder. The Jacobian of the map is given by

$$J = \begin{bmatrix} 1 + \varepsilon\cos\phi & 1 \\ \varepsilon\cos\phi & 1 \end{bmatrix},$$

which has determinant 1 and hence the map is area-preserving. On the other hand, the eigenvalues of J are given by $\lambda_\pm = 1 \pm \sqrt{\varepsilon\cos\phi} + \mathcal{O}(\varepsilon)$. Thus, for ε sufficiently small the eigenvalues satisfy the non-resonant conditions of Corollary VI.1.
Moreover, we readily see that the twist condition $\dfrac{\mathrm{d}\Omega}{\mathrm{d}I} > 0$ is satisfied (In this case $\Omega(I) = \dfrac{I}{2\pi}$).
For $\varepsilon = 0$, I is constant and ϕ increases each iteration by I. Therefore, we distinguish two cases:

(1) if $\dfrac{I}{2\pi} = \dfrac{m}{n} \in \mathbb{Q}$, then an orbit starting at a point (ϕ, I) is periodic with period n,

(2) if $\dfrac{I}{2\pi}$ is irrational, then an orbit starting at a point (ϕ, I) is not periodic and fills the curve $I =$constant densely. These orbits are called quasi-periodic.

For $\varepsilon > 0$ sufficiently small, Moser's Theorem gives us sufficient conditions for the existence of an invariant curve given by

$$\begin{aligned} \phi &= \psi + u(\psi, \omega), \\ I &= 2\pi\omega + v(\psi, \omega), \end{aligned} \tag{108}$$

for every Diophantine number ω of type (C, τ) provided that $\varepsilon \leq \varepsilon_0 C^2$ (here ε_0 is replacing ε in the theorem).

In some sense, Moser's theorem is telling us that for ε sufficiently small, most of the invariant curves for $\varepsilon = 0$ survive.

In figure 6 we show some simulations of the standard map (107) for different values of ε where it is evident that many such invariant curves exist for ε small. The invariant curves (108) receive the name of *rotational invariant curves*.

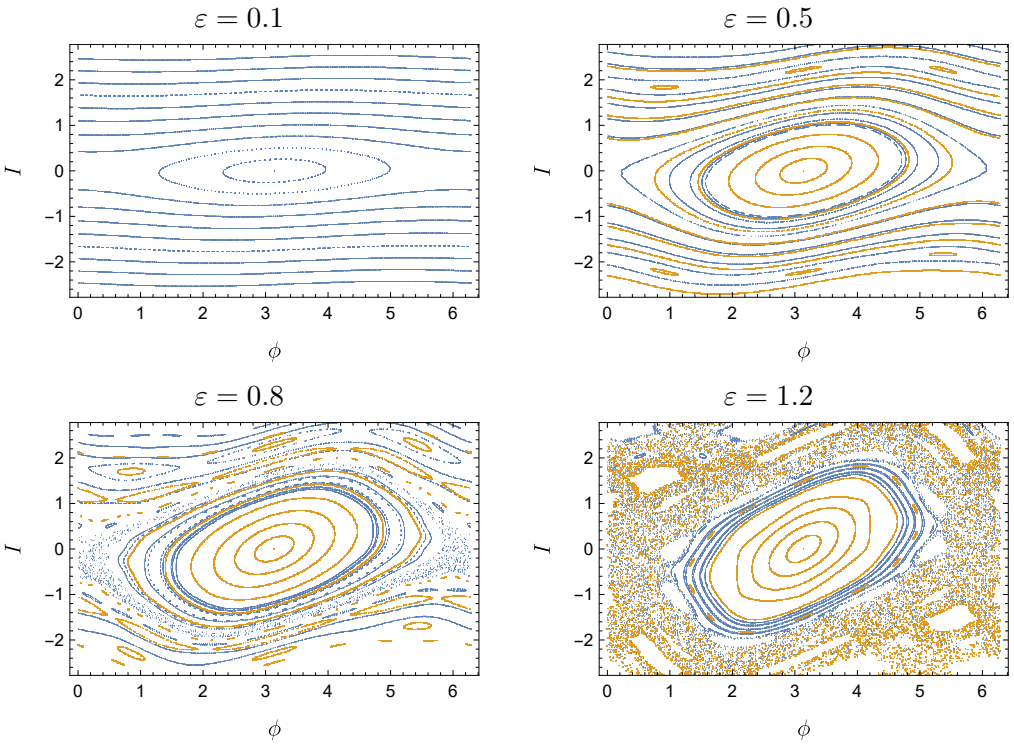


FIGURE 6. Simulations of (107) for different values of ε. Notice that near the fixed point, there exist invariant curves predicted by Moser's theorem which correspond to those curves "that traverse horizontally the phase-space" The closed orbits are due to resonances, while the "dust" indicates chaos.

From Moser's theorem it follows that the dynamics on each rotational curve is given by $\psi \mapsto \psi + 2\pi\omega$. Therefore, the k-th iterate for an initial condition (ϕ_0, I_0) reads as

$$\phi_k = \psi_0 + 2\pi k\omega + u(\psi_0 + 2\pi k\omega, \omega)$$
$$I_k = 2\pi\omega + v(\psi_0 + 2\pi k\omega, \omega).$$

To each orbit $\gamma_n = \{(\phi_n, I_n) = \Pi^n(\phi_0, I_0)\}\,|_{n\in\mathbb{Z}}$ we associate the rotation number

$$\bar{\Omega} = \frac{1}{2\pi}\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}(\phi_{k+1} - \phi_k).$$

For the standard map (107), the rotation number reads as

$$\bar{\Omega} = \frac{1}{2\pi}\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}(I_k + \varepsilon\sin\phi_k).$$

Notice that for $\varepsilon = 0$, indeed $\bar{\Omega} = \dfrac{I_0}{2\pi}$.
We finish this example by addressing the following issue: what happens to the periodic orbits that we know exist for $\varepsilon = 0$? The following theorem gives us the answer:

THEOREM VI.5 (Poincaré-Birkhoff). *Let Π be an area-preserving twist map admitting two rotational invariant curves γ_1 and γ_2, with respective rotation numbers ω_1 and ω_2. For every*

rational number $\frac{m}{n} \in (\omega_1, \omega_2)$*, there exist at least two periodic orbits with rotation number* $\frac{m}{n}$*, and contained in the domain between* γ_1 *and* γ_2*.*

Roughly speaking, periodic orbits are seen as "the centers of rings" (elliptic fixed points) which are locally stable, and the intersection of separatrices of "ring regions" (hyperbolic fixed points).

An in-depth treatment of standard maps, in much greater generality, can be found in [**21**]. To test the standard map you can program it yourselves or visit https://ibiblio.org/e-notes/Chaos/stdmap.htm.

EXERCISE VI.2. *What is the consequence of Moser's theorem regarding Example VI.14? Corroborate your arguments with appropriate simulations.*

Let us now turn our attention to another type of KAM-result. In particular, we now consider Hamiltonians of the form

$$H(I, \phi) = H_0(I) + \varepsilon H_1(I, \phi, \varepsilon), \tag{109}$$

with n-degrees of freedom. Before going into the main result, let us see how the problem of small denominators, and the related resonances, appear.

Let us suppose that we want to eliminate the angle dependence in H_1. We saw in the section of Lie-Deprit series that the first terms after the proposed change of coordinates is given by

$$K_1(J, \psi) = H_1(J, \psi) + \{H_0, W_0\}(J, \psi) = H_1(J, \psi) - \sum_{j=1}^{n} \underbrace{\frac{\partial H_0}{\partial I_j}(J)}_{\Omega_j(J)} \frac{\partial W_0}{\partial \psi_j}(J, \psi).$$

Let us write H_1 and W_0 in Fourier series, i.e.,

$$H_1(J, \psi) = \sum_{k \in \mathbb{Z}^n}^{n} H_{1k}(J) \exp(\imath k \psi)$$

$$W_0(J, \psi) = \sum_{k \in \mathbb{Z}^n}^{n} W_{0k}(J) \exp(\imath k \psi).$$

Then, K_1 further reads as

$$K_1(J, \psi) = \sum_{k \in \mathbb{Z}^n}^{n} H_{1k}(J) \exp(\imath k \psi) - \sum_{j=1}^{n} \Omega_j(J) \sum_{k \in \mathbb{Z}^n}^{n} \imath k_j W_{0k} \exp(\imath k \psi).$$

This means that, to achieve our goal, we want to solve an equation of the form

$$H_{1k} \exp(\imath k \psi) = \sum_{j=1}^{n} \Omega_j(J) \imath k_j W_{0k}(J) \exp(\imath k \psi),$$

for each $k \in \mathbb{Z}^n$. Thus, we can make the choice

$$W_{0k} = \frac{H_{1k(J)}}{\imath k \Omega(J)},$$

where $k\Omega(J) = \sum_{j=1}^{n} k_j \Omega_j(J)$. If such a term is equal to zero for some k, then we say that it is resonant, and as usual, cannot be eliminated. However, just as we have seen before, the denominator $k\Omega(J)$ may become arbitrarily small, unless, again, we impose a Diophantine condition on $\Omega_j(J)$. This problem is addressed by the following theorem.

THEOREM VI.6 (Arnol'd). *Assume that the Hamiltonian (109) is analytic and satifies*

$$\left|\det \frac{\partial^2 H_0}{\partial I^2}(I)\right| \geq W > 0,$$

in a neighborhood of the torus $I = I_0$. Let $\omega = \Omega(I_0) \in \mathbb{R}^n$ satisfy the Diophantine condition

$$|\omega k| \geq \frac{C}{|k|^\tau},$$

for all $k \in \mathbb{Z}^n \backslash 0^n$. Then, if ε is sufficiently small, the Hamiltonian system admits a quasiperiodic solution with frequency ω. This solution lies on an analytic torus filled by the solution. The distance from this torus to the unperturbed torus $I = I_0$ goes to zero as $\varepsilon \to 0$.

The previous KAM-theorem is telling us that for a nearly-integrable Hamiltonian system, the invariant tori of the unperturbed system corresponding of Diophantine frequencies, persist under sufficiently small perturbations.

EXERCISE VI.3. *Consider Hamiltonian $H(I, \phi) = H_0(I) + \varepsilon H_1(I, \phi)$. The purpose is to find a near identity canonical transformation[a] $\Phi : (I, \phi) \to (\tilde{I}, \tilde{\phi})$ such that in the new coordinates the Hamiltonian reads as $\tilde{H} = \tilde{H}_0(\tilde{I}) + \varepsilon^2 \tilde{H}_1(\tilde{I}, \tilde{\phi})$.*

[a] Its Jacobian is a symplectic matrix.

EXAMPLE VI.16. This example is taken from [9].
Let us consider the planar system

$$x_{n+1} = a x_n f(x_n, y_n)$$
$$y_{n+1} = c x_n (1 - f(x_n, y_n)).$$

This equation, called May's model, is used to model host-parasite dynamics, and thus x_n represents the host density and y_n the parasite density at generation n. The function f represents the fraction of hosts x_n not parasitized (and accordingly $1 - f$ represents the fraction of hosts parasitized). For the model, the parameter a represents the net rate at which the number of hosts increase in the absence of parasites, while the parameter c represents the average number of adult female parasites emerging from each host parasitized. To be more specific, for this example let us take the function

$$f = \left(1 + b\frac{y_n}{k}\right)^{-k}.$$

For this function, b models the "area reached by each parasite" and k is a parameter accounting for parasite clumping. The presence of these many parameters complicates the analysis. Thus,

let us simplify the model by assuming $b = k = 1$, and $c = a$. Thus the corresponding model is

$$\begin{aligned} x_{n+1} &= \frac{ax_n}{1+y_n} \\ y_{n+1} &= \frac{ax_n y_n}{1+y_n}. \end{aligned} \tag{110}$$

Notice that $y_{n+1} = ax_n - x_{n+1}$. Therefore, we can eliminate y_n and obtain an equation purely in x_n, namely:

$$x_{n+1} = \frac{ax_n}{1 + ax_{n-1} - x_n}.$$

The first relevant result to obtain is that the model, under appropriate coordinates, is area-preserving.

LEMMA VI.2. *The model* (110) *is area-preserving under logarithmic coordinates.*

PROOF. The Jacobian associated to (110) is

$$J(x,y) = \begin{bmatrix} \dfrac{a}{1+y} & -\dfrac{ax}{(1+y)^2} \\ \dfrac{ay}{1+y} & \dfrac{ax}{(1+y)^2} \end{bmatrix}.$$

Clearly $\det J(x,y) = \dfrac{a^2 x}{(1+y^2)}$, and therefore $\det J(0,0) = 0$. So, although J is not invertible at the origin, it certainly is everywhere else.
Let us define new coordinates $u = \ln x$ and $v = \ln y$. In these coordinates (110) transforms to

$$\begin{pmatrix} u \\ v \end{pmatrix} \mapsto \begin{pmatrix} \ln a + u - \ln(1+\exp(v)) \\ \ln a + u + v - \ln(1+\exp(v)) \end{pmatrix}, \tag{111}$$

and the corresponding Jacobian is

$$J(u,v) = \begin{bmatrix} 1 & -\dfrac{\exp(v)}{1+\exp(v)} \\ 1 & 1 - \dfrac{\exp(v)}{1+\exp(v)} \end{bmatrix}.$$

We can now readily see that the Jacobian in logarithmic coordinates satisfies $\det J(u,v) = 1$. □

Next, let us look at the fixed points. It is readily seen that (111) has the fixed point $(u^*, v^*) = (0, \ln(a-1))$. For this fixed point to be defined we let $a > 1$. Notice that this fixed point corresponds to the fixed point $(x^*, y^*) = (1, a-1)$ in the original coordinates. The original system (110) has another fixed point at the origin, but it corresponds to a fixed point at infinity for the logarithmic system, so we won't look at it.

EXERCISE VI.4. *Show, however, that the fixed point* $(x,y) = (0,0)$ *is a saddle provided that* $a > 0$.

Evaluating $J(u^*, v^*)$ we get

$$J(u^*, v^*) = \begin{bmatrix} 1 & -\dfrac{a-1}{a} \\ 1 & \dfrac{1}{a} \end{bmatrix}.$$

Consequently, we know that (u^*, v^*) is an elliptic point with eigenvalues

$$\lambda_{1,2} = \exp(\pm \imath\theta), \qquad \cos\theta = \frac{a+1}{2a}.$$

From the beginning of the proof of Corollary VI.1, we know that the eigenvalues $\lambda_{1,2}$ are not roots of unity for $q = 1, 2, 3, 4$. To apply the corollary, it only rests to find the first coefficient of the Birkhoff normal form. Notice that this coefficient will depend on the coefficient a. Thus, it is not surprising to expect that c_1 (in the Birkhoff normal form) is, generically, nonzero. The actual computation of c_1 us quite tedious, but can be obtained from the formulas in [28]. In Figure 7 we see a few orbits for $a = 1.7$.

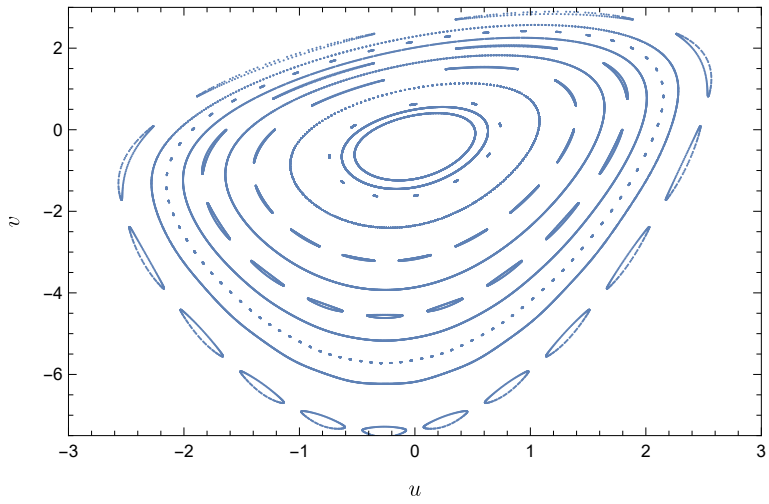


FIGURE 7. A simulation of (111). Notice the rings (quasiperiodic orbits) near the fixed point located at the origin. These are the one predicted by Moser's theorem. Notice as well the periodic orbit of period 10 after the first two rings.

VI.4. Further exercises for this chapter

(1) Consider an holomorphic map $F : \mathbb{C} \to \mathbb{C}$ of the form $F(z) = \lambda z + f(z)$ with $f(0) = f'(0) = 0$ and assume that f is given by a series $f(z) = \sum_{j\geq 2} f_j z^j$. Consider a near identity transformation $\phi(z) = z + \sum_{j\geq 2} \phi_j z^j$. Show that, if the transformation ϕ where to linearize F, the problem of small divisors appear.

(2) Study the equation $\dfrac{\mathrm{d}x}{\mathrm{d}t} = -\varepsilon x \cos t$ via the method of averaging. Compare the averaged and exact solutions.

(3) Consider the system

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \varepsilon(a + b\cos\phi)$$
$$\frac{\mathrm{d}\phi}{\mathrm{d}t} = \omega.$$

Find the averaged system, and compare numerically the solutions of the original system and the averaged one.

(4) Study the nonlinear systems

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \varepsilon(x - x^2)\sin^2 t$$

and

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \varepsilon\left(x\sin^2 t - \frac{x^2}{2}\right)$$

with the averaging method. Compare the respective solutions (averaged and full).

(5) Using the averaging method, study the "original" van der Pol equation

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + \frac{\varepsilon}{\omega}(x^2 - 1)\frac{\mathrm{d}x}{\mathrm{d}t} + x = \varepsilon\gamma\cos(\omega t),$$

with $1 - \omega^2 = \mathcal{O}(\varepsilon)$.

(6) Using the method of Lie-Deprit series analyze the parametrically driven an-harmonic pendulum

$$H = \frac{1}{2}p^2 + \frac{1}{2}\omega_0^2(1 + \alpha\sin(\omega t))q^2 + \frac{\varepsilon}{4}\omega^2 q^4.$$

In this model both ε and α are small. However, assume that $\alpha \ll \varepsilon$. This assumption allows you to relate the leading order terms of the Hamiltonian with that of example VI.13. In other words, the higher order term now is $\frac{1}{2}\alpha\omega_0^2\sin(\omega t)q^2$.

Hint: when computing K_2 you will find that if you choose $K_2 = 0$, then the problem of small denominators will appear when computing W_2. You can keep this choice for $\omega \leq 2\omega_0$, but for $\omega \sim \omega_0$ the choice of K_2 will depend on the resonant term $\sin(\omega t - 2\omega_0)$.

(7) Consider a 2-DOF Hamiltonian $H(I_1, I_2, \phi_1, \phi_2) = H_0(I_1, I_2) + \varepsilon H_1(I_1, I_2, \phi_1, \phi_2)$. Let $H_0 = a_1 I_1 + a_2 I_2$ with $a = (a_1, a_2) \in \mathbb{R}^2$ and $\phi = (\phi_1, \phi_2) \in \mathbb{T}^2$.

- Describe the dynamics of the unperturbed system.
- Consider a resonant perturbation $H_1 = -\sin(2\pi k \cdot \phi)$ where $k \in \mathbb{Z}\setminus\{0\}$ is chosen so that $k \cdot a = 0$. Find the trajectory passing through $(\psi_1, \psi_2, I_1, I_2) = (0, 0, I_1, I_2)$ and explain why the invariant tori of the unperturbed system do not persist for any arbitrarily small ε.
- What happens if you impose Diophantine conditions for the perturbation?

CHAPTER VII

(Geometric) Singular Perturbation Theory

In this chapter we consider the so-called *slow-fast* systems. These are "singularly perturbed" ODEs of the form

$$\begin{aligned} \varepsilon \frac{\mathrm{d}x}{\mathrm{d}t} &= f(x, y, \varepsilon) \\ \frac{\mathrm{d}y}{\mathrm{d}t} &= g(x, y, \varepsilon), \end{aligned} \tag{112}$$

where $x \in \mathbb{R}^{n_f}$, $y \in \mathbb{R}^{n_s}$, $0 < \varepsilon \ll 1$, and f and g are assumed to be sufficiently smooth.

> Although the distinction between regular and singular perturbations is rather subtle and sometimes imprecise, here the term singular refers to the fact that in the limit $\varepsilon = 0$, the system (112) is not an ODE anymore.

Re-scaling time by $t = \varepsilon\tau$ we obtain the equivalent system

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}\tau} &= f(x, y, \varepsilon) \\ \frac{\mathrm{d}y}{\mathrm{d}\tau} &= \varepsilon g(x, y, \varepsilon). \end{aligned} \tag{113}$$

For $\varepsilon > 0$, the only difference between (112) and (113) is their time parametrization, that is their orbits are the same, hence their equivalence. The time parameter t is usually called the "slow time" and therefore (112) is called the slow equation. Similarly τ is the "fast time" and (113) the fast equation.

The overall idea of Geometric Singular Perturbation Theory (GSPT) is to study the limit of (112) and (113), and from there provide qualitative (or even quantitative) description for the perturbed problem. The limit, as $\varepsilon \to 0$ of (112) is

$$\begin{aligned} 0 &= f(x, y, 0) \\ \frac{\mathrm{d}y}{\mathrm{d}t} &= g(x, y, 0), \end{aligned}$$

which is not an ODE anymore, but a constrained differential equation (CDE) [**26**], also called differential-algebraic equations (DAE). We notice that for a CDE, the solutions of $\dfrac{\mathrm{d}y}{\mathrm{d}t} = g(x, y, 0)$ are required to satisfy the algebraic constraint $f = 0$. Generically, that is near points where $\mathrm{D}_x f|_{\{f=0\}}$ is full rank, the set

$$\mathcal{C}_0 = \{(x, y) \in \mathbb{R}^{n_f} \times \mathbb{R}^{n_s} \mid f(x, y, 0) = 0\}$$

is an n_s-dimensional manifold. This manifold is called *the critical manifold*[1]. For points $p \in \mathcal{C}_0$ for which $\mathrm{D}_x f(p)$ is full rank, the implicit function theorem tells us that $\mathcal{C}_0$ is locally given as the graph

[1]The name critical manifold is misleading because, as we will see, the set $\mathcal{C}_0$ is not always a manifold. The name remains like that due to historical reasons. If one would like to be more formal, one should simply call $\mathcal{C}_0$ *the critical set*.

of a smooth function $x = h(y)$. In this case, the dynamics on the critical manifold are given by

$$\frac{\mathrm{d}y}{\mathrm{d}t} = g(h(y), y, 0) = G(y), \tag{114}$$

which is called the "reduced slow equation". The flow of (114), that is the flow on the critical manifold, is called *the slow flow*. In principle, the reduced slow equation is easier to solve than the original (112). Thus, a central question we want to answer is: how are the solutions of the reduced slow equation (114) related to those of the slow-fast system (112)?

On the other hand, the limit of (113) as $\varepsilon \to 0$ reads as

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}\tau} &= f(x, y, 0) \\ \frac{\mathrm{d}y}{\mathrm{d}\tau} &= 0, \end{aligned} \tag{115}$$

which is called *the layer equation*. We notice that the layer equation is effectively an ODE $\frac{\mathrm{d}x}{\mathrm{d}\tau} = f(x, y, 0)$ with slow variable y having the role of a parameter. Notice that for the layer equation, the critical manifold corresponds to equilibria. Moreover, regular points of $\mathcal{C}_0$ correspond to hyperbolic equilibria, while singular points correspond to non-hyperbolic equilibria. This distinction will play a fundamental role in the theory.

Regular perturbation theory tells us that orbits of (113) remain close to those of (115) for time τ of order $\mathcal{O}(1)$, that is for time t of order $\mathcal{O}(\varepsilon)$. The goal of this chapter is to provide some perturbation results that describe the dynamics of a slow-fast system for time t of order $\mathcal{O}(1)$.

VII.1. Fenichel's Theory

In this section we focus on the case where the unperturbed problem indeed gives a good enough approximation of the dynamics of the perturbed problem for large time. To fix ideas, let us see first an example.

EXAMPLE VII.1. Consider the planar slow-fast system

$$\begin{aligned} \varepsilon \frac{\mathrm{d}x}{\mathrm{d}t} &= y^2 - x \\ \frac{\mathrm{d}y}{\mathrm{d}t} &= -y, \end{aligned} \tag{116}$$

or in its fast-time parametrization

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}\tau} &= y^2 - x \\ \frac{\mathrm{d}y}{\mathrm{d}\tau} &= -\varepsilon y. \end{aligned}$$

The corresponding critical manifold is

$$\mathcal{C}_0 = \left\{ (x, y) \in \mathbb{R}^2 \,|\, y^2 - x = 0 \right\}.$$

We notice that the critical manifold is everywhere regular because $\frac{\partial}{\partial x}(y^2 - x) = -1$ for all $p \in \mathcal{C}_0$. In this case, the reduced slow equation is simply

$$\frac{\mathrm{d}y}{\mathrm{d}t} = -y,$$

and therefore the slow flow is

$$y(t) = y(0)\exp(-t).$$

On the other hand, the layer equation reads as

$$\frac{\mathrm{d}x}{\mathrm{d}\tau} = y^2 - x$$
$$\frac{\mathrm{d}y}{\mathrm{d}\tau} = 0.$$

As mentioned before, every point of the critical manifold is an equilibrium point of the layer equation. Moreover, since $\dfrac{\partial}{\partial x}(y^2 - x) = -1$, each equilibrium point is stable. The limit behavior, for $\varepsilon = 0$, is sketched in figure 1

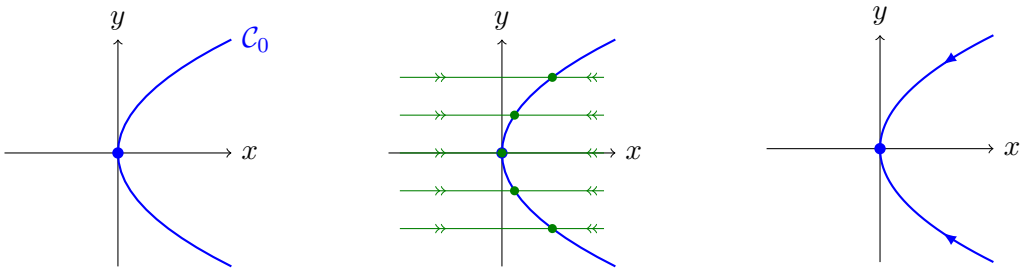


FIGURE 1. Critical manifold, layer dynamics and slow flow for (116).

Let us now compare the solutions of the unperturbed and perturbed problems. The layer equation has the solution

$$\hat{x}(\tau) = y(0)^2 + (x(0) - y(0)^2)\exp(-\tau),$$

where we introduce the $\hat{x}$ notation only to make the distinction between the solutions clearer. On the other hand, the fast equation has solution

$$x(\tau) = \frac{y(0)^2\exp(-2\varepsilon\tau)}{1-2\varepsilon} + \left(x(0) - \frac{y(0)^2}{1-2\varepsilon}\right)\exp(-\tau).$$

If we write such solutions in the slow time scale we get

$$\begin{aligned}\hat{x}(t) &= y(0)^2 + (x(0) - y(0)^2)\exp(-t/\varepsilon)\\ x(t) &= \frac{y(0)^2\exp(-2t)}{1-2\varepsilon} + \left(x(0) - \frac{y(0)^2}{1-2\varepsilon}\right)\exp(-t/\varepsilon)\\ &= \frac{y^2(t)}{1-2\varepsilon} + \left(x(0) - \frac{y(0)^2}{1-2\varepsilon}\right)\exp(-t/\varepsilon).\end{aligned}$$

For $t = \mathcal{O}(\varepsilon)$ we see that $|\hat{x} - x| = \mathcal{O}(\varepsilon)$, as predicted by regular perturbation theory. For larger time, the term $\exp(-t/\varepsilon)$ is exponentially small. For example, for time $t = k\varepsilon|\log\varepsilon|$, we have that $\exp(-t/\varepsilon) = \varepsilon^k$ and for larger times, such exponential goes to zero faster than any power of ε. Therefore, for large times $x(t) \sim y^2(t) + \mathcal{O}(\varepsilon)$, meaning that the flow is $\mathcal{O}(\varepsilon)$ close to the critical manifold. See a numerical simulation in figure 2.

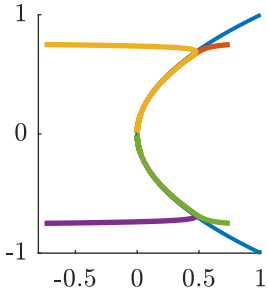


FIGURE 2. Simulation of (116) for $\varepsilon = 0.01$

Fenichel's theorem give conditions on general systems so that the behavior described in the previous example holds. Before we state the theorem, we need the concept of normal hyperbolicity.

DEFINITION VII.1. A point $p \in \mathcal{C}_0$ is called hyperbolic, if the $n_f \times n_f$ matrix $\mathrm{D}_x f(x,y,0)|_{(p)}$ has all its eigenvalues with nonzero real part. The critical manifold $\mathcal{C}_0$ is called normally hyperbolic if all points $p \in \mathcal{C}_0$ are hyperbolic. A normally hyperbolic critical manifold is called attracting / repelling / or of saddle type, if the eigenvalues of $\mathrm{D}_x f(x,y,0)|_{(p)}$ have real part negative / positive / or both.

Points $p \in \mathcal{C}_0$ that are not hyperbolic are called non-hyperbolic.

THEOREM VII.1 (Fechichel). *Suppose that $\mathcal{S}_0$ is a* compact *normally hyperbolic subset of the critical manifold $\mathcal{C}_0$. Then for $\varepsilon > 0$ sufficiently small the foll0wing hold:*

- *There exists a locally invariant manifold $\mathcal{S}_\varepsilon$ diffeomorphic to $\mathcal{S}_0$.*
- *$\mathcal{S}_\varepsilon$ lies within distance (Hausdorff) $\mathcal{O}(\varepsilon)$ for $\mathcal{S}_0$.*
- *The flow on $\mathcal{S}_\varepsilon$ converges to the slow flow (on $\mathcal{S}_0$) as $\varepsilon \to 0$.*
- *$\mathcal{S}_\varepsilon$ is normally hyperbolic and has the same stability properties as $\mathcal{S}_0$.*

REMARK VII.1.

- The proof of Fenichel's theorem follows a series of strong results and is contained in [8]. Notice that this article is in fact called "Geometric Singular Perturbation Theory". Nowadays, what we call Geometric Singular Perturbation Theory includes many more results, since one can also deal with non-hyperbolic critical manifolds, see [19].
- Usually, the manifold $\mathcal{S}_\varepsilon$ is not unique. However, away from the boundaries of $\mathcal{S}_0$, all the manifolds satisfying Fenichel's theorem lie within distance $\mathcal{O}(\exp(-l/\varepsilon))$ for some $k > 0$. Any choice of such manifolds is called *the slow manifold*.
- A similar version of Fenichel's theorem is Tikhonov's theorem [27], which deals with the case of attracting critical manifolds.

EXAMPLE VII.2. In the previous example, notice that if one choose an initial condition $x(0) = \dfrac{y(0)^2}{1-2\varepsilon}$, then the dynamics evolve only on the slow time-scale and $x(t) = \dfrac{y(t)^2}{1-2\varepsilon}$ would be the solution. Hence the slow manifold is given by

$$\mathcal{S}_\varepsilon = \left\{ (x,y) \in \mathbb{R}^2 \,|\, x = \frac{y^2}{1-2\varepsilon} \right\}.$$

We now derive a general formula for the slow flow.

PROPOSITION VII.1. *Let $p \in \mathcal{C}_0$ be such that $\mathrm{D}_x f(x,y,0)|_{(p)}$ is full rank*[2]*. Then, there exists a neighborhood $V \subset \mathcal{C}_0$ of p such that the slow subsystem for $\varepsilon = 0$ on V is given by*

$$\begin{aligned}\frac{\mathrm{d}x}{\mathrm{d}t} &= -(\mathrm{D}_x f(q,0))^{-1}(\mathrm{D}_y f(q,0))g(q,0)\\ \frac{\mathrm{d}y}{\mathrm{d}t} &= g(q,0),\end{aligned}$$

for all $q \in V$.

PROOF. Since $\det \mathrm{D}_x f(x,y,0)|_{(p)} \neq 0$, we can find a neighborhood of p, call it $V \subset \mathcal{C}_0$, such that $\mathrm{D}_x f(x,y,0)|_{(q)}$ is invertible for all $q \in V$. Implicit differentiation of $f(x,y,0)$ with respect to t yields:

$$(\mathrm{D}_x f)\frac{\mathrm{d}x}{\mathrm{d}t} + (\mathrm{D}_y f)\frac{\mathrm{d}y}{\mathrm{d}t} = (\mathrm{D}_x f)\frac{\mathrm{d}x}{\mathrm{d}t} + (\mathrm{D}_y f)g(x,y,0) = 0,$$

from which the result follows. □

REMARK VII.2. One should note that, in practice, computing the slow flow can be very difficult as analytically solving the equation $f(x,y,0) = 0$ can be highly non-trivial.

EXAMPLE VII.3 (van der Pol). Consider the unforced van der Pol oscillator

$$\begin{aligned}\varepsilon\frac{\mathrm{d}x}{\mathrm{d}t} &= y - \frac{x^3}{3} + x\\ \frac{\mathrm{d}y}{\mathrm{d}t} &= -x.\end{aligned}$$

The critical manifold is

$$\mathcal{C}_0 = \left\{(x,y) \in \mathbb{R}^2 \,|\, y = \frac{x^3}{3} - x\right\}.$$

Notice that $\mathrm{D}_x f(x,y,0) = -x^2 + 1$, meaning that $(x,y) = \left(\pm 1, \mp\frac{2}{3}\right)$ are non-hyperbolic points. It then follows that the critical manifold has three components:

$$\begin{aligned}\mathcal{C}_0^{a,-} &= \{(x,y) \in \mathcal{C}_0 \,|\, x < -1\},\\ \mathcal{C}_0^{r} &= \{(x,y) \in \mathcal{C}_0 \,|\, -1 < x < 1\},\\ \mathcal{C}_0^{a,+} &= \{(x,y) \in \mathcal{C}_0 \,|\, x > 1\}.\end{aligned}$$

All of the above branches are normally hyperbolic, and the superscript a stands for "attracting" and r for "repelling".
Away from the non-hyperbolic points, the slow flow is given by

$$\begin{aligned}\frac{\mathrm{d}x}{\mathrm{d}t} &= \frac{x}{1-x^2}\\ \frac{\mathrm{d}y}{\mathrm{d}t} &= -x.\end{aligned}$$

Notice that the first equation suffices, as is purely given with respect to x (one could of course parametrize the slow flow in terms of y, just for this example parametrizing it in terms of x is easier). We readily see that the slow flow is not well-defined at the non-hyperbolic points.

[2]Notice that this is weaker than hyperbolicity. The matrix $\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}$ is full rank but has eigenvalues $\pm\imath$.

Qualitative pictures of the slow flow (in x and in y) are simple to obtain. From those we can sketch the flow in $\mathcal{C}_0$. All these are shown in figure 3.

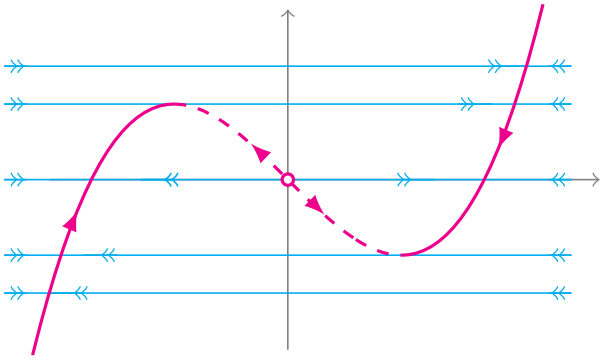


FIGURE 3. Singular limit for the unforced van der Pol equation.

A drawback of the slow-subsystem as given in Proposition VII.1 is that the equation

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -(\mathrm{D}_x f(q,0))^{-1}(\mathrm{D}_y f(q,0))g(q,0)$$

does not define a vector field on $\mathbb{R}^{n_f}$ whenever $\mathrm{D}_x f(q,0)$ loses rank (as in the Example). This can be remedied by defining the *desingularized* vector field

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\det(\mathrm{D}_x f(q,0))(\mathrm{D}_x f(q,0))^{-1}(\mathrm{D}_y f(q,0))g(q,0). \tag{117}$$

Notice that now we have a well-defined vector field. However, to relate the flow of (117) with the slow flow, one needs to pay particular attention to the sign of $\det(\mathrm{D}_x f(q,0))$.

VII.2. Singularities of the critical manifold

In the previous section we have seen that the relationship between the unperturbed and perturbed solutions near normally hyperbolic points of the critical manifold is relatively easy. In this section we briefly explore what happens in a neighborhood of a class of non-hyperbolic points. In particular, we will only focus on *fold points*, which are one of the simplest kinds of non-hyperbolic points in slow-fast systems. Moreover, from now on we restrict ourselves to planar slow-fas systems (except for a few examples). A thorough account of this theory can be found in [**19**].

The analysis near non-hyperbolic points can be seen as the problem of "dynamic bifurcations". Consider again the van der Pol equation, but now in its fast formulation:

$$\begin{aligned}\frac{\mathrm{d}x}{\mathrm{d}\tau} &= y - \frac{x^3}{3} + x\\ \frac{\mathrm{d}y}{\mathrm{d}\tau} &= -\varepsilon x.\end{aligned}$$

We can see y as a "slowly varying parameter". Notice that the scalar equation

$$\frac{\mathrm{d}x}{\mathrm{d}\tau} = y - \frac{x^3}{3} + x$$

undergoes a saddle node bifurcation for the parameter $y = \pm\frac{2}{3}$. Since this bifurcation is also know as fold bifurcation, the points $(x,y) = \left(\pm 1, \mp\frac{2}{3}\right)$ are called *fold points*.

We recall that $p \in \mathcal{C}_0$ is called singular if $\mathrm{D}_x f(x,y,0)|_{(p)}$ is not full rank.

EXAMPLE VII.4. The simplest example of a singular point presents in a planar slow-fast system given by

$$\frac{\mathrm{d}x}{\mathrm{d}\tau} = \underbrace{y - x^2}_{=f(x,y)}$$
$$\frac{\mathrm{d}y}{\mathrm{d}\tau} = \varepsilon g(x,y,\varepsilon),$$

where $(x,y) \in \mathbb{R} \times \mathbb{R}$. The critical manifold is the parabola

$$\mathcal{C}_0 = \left\{(x,y) \in \mathbb{R}^2 \,|\, y = x^2\right\}.$$

Notice that at the origin $\mathrm{D}_x f(0) = -2x|_{x=0} = 0$, hence a non-hyperbolic point. The example equation has the non-degeneracy condition $\dfrac{\partial^2 f}{\partial x^2}|_0 \neq 0$. See a sketch in figure 4

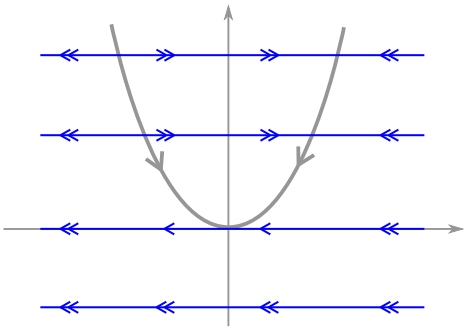


FIGURE 4. Schematic of a fold singularity. In this case we simply put $y' < 0$ close to the singularity.

Notice that the fast equation corresponds to the normal form of the saddle-node bifurcation with y playing the role of a parameter. The type of points (see a formal definition below) satisfying $f(p^*,0) = 0$, $\dfrac{\partial f(x,y,0)}{\partial x}(p) = 0$, and $\dfrac{\partial^2 f(x,y,0)}{\partial x^2}(p) \neq 0$ are called *fold points*.

EXERCISE VII.1. *Corroborate that the two non-hyperbolic points of the van der Pol equation are indeed fold points.*

For general slow-fast systems, fold points are defined as follows.

DEFINITION VII.2 (Fold point). Let $p \in \mathcal{C}_0$. The point p is called a *fold point*, if

$$\mathrm{D}_x f(x,y,0)|_p$$

is of rank $n_f - 1$. A fold point is called *non-degenerate* if for left and right eigenvectors u, v, of $\mathrm{D}_x f(x,y,0)$ one has

$$u \cdot (\mathrm{D}_x f(x,y,0)|_p) \neq 0$$

or

$$(\mathrm{D}_x f(x,y,0)|_p) \cdot v \neq 0.$$

As mentioned above, the importance of fold points is that they are (one of the) simplest singularities of critical manifolds. Therefore, when studying non-linear slow-fast systems, it will not be surprising

if we encounter fold points. Therefore, in the following section we describe a very powerful geometric technique that allow us to study the dynamics of slow-fast systems near such type of points (and many other non-hyperbolic points). Fold points (in particular regarding the van der Pol equation) have been studied using asymptotic methods as well, see [**24, 23**].

VII.3. The blow-up method

In this section we describe a technique that nowadays form a fundamental part of GSPT. For pedagogical purposes, we shall describe it only in the context of planar systems and use it to study a slow-fast system near a generic fold point. For further details and applications see [**19, 16**].

VII.3.1. Blow-up for a single time scale planar system. In this section we introduce the blow-up method in its classical context, that is, to desingularize a nilpotent equilibrium point[3] of a planar vector field. For a detailed exposition see [**19**, Chapter 7] and references therein. Here we shall only treat an example to highlight the main idea of the method. Later, we will see how this transformation also fits into the study of slow-fast systems.

Let us consider the planar ordinary differential equation (ODE)

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= y \\ \frac{\mathrm{d}y}{\mathrm{d}t} &= x^3 + xy. \end{aligned} \tag{118}$$

We note that the origin $(x, y) = (0, 0)$ is a unique equilibrium point and that the linearization of (118) at the origin is given by the matrix

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

Thus, the origin is a non-hyperbolic equilibrium point and, moreover, is nilpotent. Our goal is to qualitatively describe the orbits of (118) in a small neighborhood of the origin. However, not only the linearization offers no useful information, but center manifold reduction is not suitable since in this case the center manifold corresponds to the whole phase-space. So, what we are going to use is a suitable change of coordinates, known as *blow-up*, which will induce a new system with only hyperbolic equilibrium points, and therefore can be analyzed by dynamical systems tools.

Let us consider a *weighted polar change of coordinates*

$$\phi : \mathbb{S}^1 \times I \to \mathbb{R}^2, \qquad \phi(\theta, r) = (r\cos\theta, r^2 \sin\theta), \tag{119}$$

where $I \subseteq \mathbb{R}$ is an interval containing the origin and $\theta \in [0, 2\pi]$. At the end of this section we clarify the reason to choose a weighted polar change of coordinates, for now let us proceed with the example.

The change of coordinates defined by $(x, y) = (r\cos\theta, r^2\sin\theta)$ defines a new ODE, namely

$$\begin{aligned} \dot\theta &= \frac{r\left(1 + \sin\theta - 4\sin^2\theta - \sin^3\theta + \sin^4\theta\right)}{\sin^2\theta + 1} \\ \dot r &= \frac{r^2}{\sin^2\theta + 1}\cos\theta\sin\theta\left(\sin\theta - \sin^2\theta + 2\right). \end{aligned} \tag{120}$$

Note that the change of coordinates defined by ϕ maps the circle $\mathbb{S}^1 \times \{0\}$ to the origin in the plane[4]. Moreover, since ϕ is a diffeomorphism for $\{r > 0\}$, orbits of (118) in a small neighborhood

[3] We recall that an equilibrium point of a vector field is called nilpotent if the linearization of the vector field at such a point is given by a matrix with only zero eigenvalues.

[4] Equivalently ϕ^{-1} maps the origin in the plane to the circle $\mathbb{S}^1 \times \{0\}$.

of the origin correspond to orbits of (120) in a small neighborhood of $\mathbb{S}^1 \times \{0\}$. Note however that (120) vanishes along $\mathbb{S}^1 \times \{0\}$. To overcome this we can divide the right-hand side of (120) by r. This operation does not change the qualitative properties of the orbits in the region $\mathbb{S}^1 \times \{r > 0\}$. Thus, it shall suffice to study *the desingularized system*

$$\begin{aligned} \dot{\theta} &= \frac{1}{\sin^2\theta + 1}\left(1 + \sin\theta - 4\sin^2\theta - \sin^3\theta + \sin^4\theta\right) \\ \dot{r} &= \frac{r}{\sin^2\theta + 1}\cos\theta\sin\theta\left(\sin\theta - \sin^2\theta + 2\right), \end{aligned} \tag{121}$$

which does not vanish any more along $\mathbb{S}^1 \times \{0\}$. The most important fact is that *orbits of* (121) *near* $\mathbb{S}^1 \times \{0\}$ *correspond to orbits of* (118) *near the origin.*

It is now straightforward to show that (121) has four *hyperbolic* saddle equilibrium points, namely $p_1 = (-\arcsin(\sqrt{2} - 1), 0)$, $p_2 = (\arcsin(\sqrt{5}/2 - 1/2), 0)$, $p_3 = (\pi - \arcsin(\sqrt{5}/2 - 1/2), 0)$ and $p_4 = (\pi + \arcsin(\sqrt{2} - 1), 0)$. Since the aforementioned equilibrium points are hyperbolic it follows from linear analysis that the phase portrait of (121) in a small neighborhood of $\mathbb{S}^1 \times \{0\}$ is as show in Figure 5.

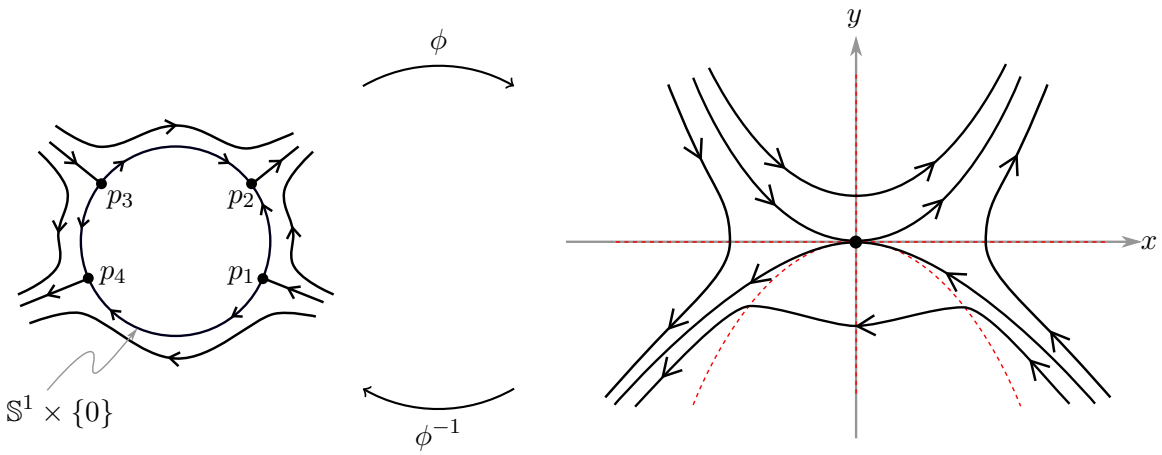


FIGURE 5. Blow-up analysis of (118). On the left we show the phase-portrait of (121) in a small neighborhood of $\mathbb{S}^1 \times \{0\}$, where four hyperbolic saddle points are found. On the right we show the corresponding orbits of (118), where from a qualitative perspective, the circle $\mathbb{S}^1 \times \{0\}$ "blows-down" to the origin and all other orbits of (118) are equivalent to orbits of (121). To provide more detail on the flow of (118) away from the origin we have made use of the corresponding nullclines, shown as dashed-red curves.

We finish this section with some important remarks:

- The procedure we exemplified above is known as *the blow-up method*. In some sense, the transformation ϕ^{-1} "blows the origin up to a circle". The advantage of blowing up is that one obtains a new system which is simpler to analyze. We recall that, in the above example, (118) has a nilpotent equilibrium point at the origin while (121) has four hyperbolic equilibrium points along $\mathbb{S}^1 \times \{0\}$, which are simpler to study with standard techniques of dynamical systems. Once the blown-up system is understood we then "blow-down" the phase-portrait of (121) resulting in a qualitative description of the original system (118).
- In the example presented above we have used a weighted version of a polar change of coordinates. Usually one then refers to the transformation as a *quasi-homogeneous* blow-up to emphasize that the weights in the transformation are distinct from 1. The advantage of using a quasi-homogeneous

blow-up instead of a homogeneous one is that we can desingularize the origin in just one step. The reader can check that if one uses $(x, y) = (r\cos\theta, r\sin\theta)$ instead of (119), the blown-up system then has a pair of nilpotent singularities located at $(\theta, r) = (0, 0)$ and $(\theta, r) = (\pi, 0)$. In turn, the blow-up method can be applied once more to such pair of points, see [**19**, Chapter 7].

VII.3.2. The blow-up method for slow-fast systems. In the previous section we sketched the idea of the blow-up method to desingularize a nilpotent singularity of planar vector fields. In this section we describe the blow-up method as is nowadays commonly used for the analysis of slow-fast systems with non-hyperbolic singularities.

Let us first rewrite the ε-family of vector fields (113) on $\mathbb{R}^{m+n}$ as a single vector field on $\mathbb{R}^{m+n+1}$ of the form

$$\begin{aligned} x' &= f(x, y, \varepsilon) \\ y' &= \varepsilon g(x, y, \varepsilon) \\ \varepsilon' &= 0. \end{aligned} \tag{122}$$

Furthermore, let us assume that the origin $(x, y, \varepsilon) = (0, 0, 0)$ is an equilibrium point and that $\mathrm{D}_x f(0, 0, 0)$ has all its eigenvalues equal to zero. This means that the origin is a nilpotent singularity of (122) and, as such, the blow-up method can be adapted to desingularize the origin of (122).

Remark VII.3. It is worth noting that nilpotent singularities are a subset of non-hyperbolic singularities. Thus, not all non-hyperbolic singularities of slow-fast systems may be studied with the blow-up method. In particular, in all slow-fast systems with one-dimensional fast direction ($x \in \mathbb{R}$), a non-hyperbolic singularity is nilpotent. In other cases where the singularity is non-hyperbolic but not nilpotent, a preliminary transformation may bring a slow-fast system into a suitable form to be analyzed via the blow-up method.

Although there are several (equivalent) versions and improvements of the blow-up method, we restrict to the quasihomogeneous case as it is more commonly used nowadays. For further information see [**19**, Chapter 7] and references therein.

Let $X : \mathbb{R}^{m+n+1} \to \mathbb{R}^{m+n+1}$ be the vector field, which in coordinates is defined by (122), and let $\mathbb{S}^N$ denote the N-th dimensional sphere[5]. Next, we can formally define the blow-up transformation most commonly used in slow-fast systems:

Definition VII.3 (Quasihomogeneous blow-up). Consider a vector field $X : \mathbb{R}^{m+n+1} \to \mathbb{R}^{m+n+1}$ defined by (122) and assume that $X(0) = 0$. Let $\alpha = (\alpha_1, \dots, \alpha_m) \in \mathbb{N}_0^m$, $\beta = (\beta_1, \dots, \beta_m) \in \mathbb{N}_0^n$ and $\gamma \in \mathbb{N}_0$. Let the generalized polar transformation $\phi : \mathbb{S}^{m+n} \times I \to \mathbb{R}^{m+n+1}$ be defined by

$$\phi(\bar{x}, \bar{y}, \bar{\varepsilon}, r) = (r^\alpha \bar{x}, r^\beta \bar{y}, r^\gamma \bar{\varepsilon}) = (x, y, \varepsilon), \tag{123}$$

where $(\bar{x}, \bar{y}, \bar{\varepsilon}) = (\bar{x}_1, \dots, \bar{x}_m, \bar{y}_1, \dots, \bar{y}_n, \bar{\varepsilon}) \in \mathbb{S}^{m+n}$, $r \in I$, and $I \subseteq \mathbb{R}$ is an interval containing the origin. Here we use the multi-index notation $r^\alpha \bar{x} = (r^{\alpha_1}\bar{x}_1, \dots, r^{\alpha_m}\bar{x}_m)$, and similarly for $r^\beta \bar{y}$. The *quasihomogeneous* blow-up of the vector field X, denoted as $\bar{X}$, is defined by

$$\bar{X} = \mathrm{D}\phi^{-1}|_{(\bar{x}, \bar{y}, \bar{\varepsilon}, r)} \circ X \circ \phi(\bar{x}, \bar{y}, \bar{\varepsilon}, r). \tag{124}$$

We note that ϕ maps the sphere $\mathcal{B}_0 := \mathbb{S}^{m+n} \times \{0\}$ to the origin in $\mathbb{R}^{m+n+1}$, while ϕ^{-1} maps $0 \in \mathbb{R}^{m+n+1}$ to $\mathcal{B}_0$. Hence, the operation ϕ^{-1} is called (quasihomogeneous) blow-up while ϕ is called (quasihomogeneous) blow-down. The word quasihomogeneous reflects the fact that the exponents

[5]To use spheres, cylinders, or related spaces as blown-up spaces is often very convenient, yet not necessary.

appearing in (123) are not necessarily the same. We omit the term "quasihomogeneous" when all exponents (α, β, γ) are equal to 1.

It follows from (124) that $\bar{X}$ and X are conjugate for $r > 0$, meaning that there exists a one-to-one mapping between trajectories of X and trajectories of $\bar{X}$ outside $\mathcal{B}_0$. Moreover, it can be shown that $\bar{X}$ is well defined at $r = 0$ [**19**]. Due to the presence of non-hyperbolic singularities, and depending on the choice of the exponents, it is usually the case that the system denoted by $\bar{X}$ vanishes on $\mathcal{B}_0$. In fact, let $j_\ell(X)$ denote the ℓ-jet of X at the origin. If $j_\ell(X) = 0$ for $\ell = 0, 1, \dots, k$ and $j_{k+1}(X) \neq 0$, then we define the *desingularized vector field* $\tilde{X} = \frac{1}{r^k}\bar{X}$. Now $\tilde{X}$ does not vanish at $\mathcal{B}_0$. Since $\bar{X}$ and $\tilde{X}$ are smoothly equivalent for $r > 0$, all the information obtained from $\tilde{X}$ is equivalent to that of $\bar{X}$ outside $\mathcal{B}_0$. However, since $\tilde{X}$ does not vanish any more along $\{r = 0\}$, we may try to infer the dynamics of $\tilde{X}$ for $r > 0$ small from the restriction $\tilde{X}|_{\{r=0\}}$. This greatly simplifies the analysis, since usually we find that $\tilde{X}$ has semi-hyperbolic singularities, hyperbolic singularities, or no singularities at all. Finally, due to the equivalences between $X, \bar{X}$, and $\tilde{X}$, we conclude that the flow of $\tilde{X}$ for $r > 0$ sufficiently small provides a complete description of the flow of X for $\varepsilon > 0$ sufficiently small.

When we study high dimensional problems, say for $m + n > 2$, working with polar coordinates can become cumbersome. Then, we rather work in charts that cover the blow-up space. In each of the charts we can define local coordinates and a corresponding *local* vector field. In practice, what we do to define local coordinates in a chart is to fix one of the blow-up coordinates to ± 1. This approach is called "directional blow-up". For example, to perform a blow-up in the $\bar{\varepsilon}$-direction we would define new coordinates according to $\phi : \mathbb{R}^{n+m+1} \to \mathbb{R}^{n+m+1}$ given by

$$\phi(\bar{x}, \bar{y}, \bar{\varepsilon}, r) = (r^\alpha \bar{x}, r^\beta \bar{y}, r^\gamma),$$

that is by fixing $\bar{\varepsilon} = 1$. Similarly, we can define blow-ups in any of the other directions.

REMARK VII.4. The chart $K := \{\bar{\varepsilon} = 1\}$ is the most important one and it is called *the rescaling chart*, *the family chart* or *the central chart*. The rest of the charts are often referred to as *phase-directional charts*.

Directional blow-ups induce local vector fields on each of the (Euclidean) charts. Once the analysis of the relevant local vector fields is performed, one can overlap suitable regions of the charts and match the flow on such charts via the so-called *matching maps* (or *transition maps*) to describe the dynamics all around $\mathbb{S}^{m+n} \times I$. In particular, this process allows us to track invariant objects, principally center manifolds [**6**], across the blow-up space. A schematic representation of the blow-up map is provided in Figure 6.

VII.3.3. Blow-up analysis of a folded singularity. In this section we consider the planar slow-fast system

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}\tau} &= -y + x^2 + \mathcal{O}(\varepsilon, xy, y^2, x^2) \\ \frac{\mathrm{d}y}{\mathrm{d}\tau} &= \varepsilon(-1 + \mathcal{O}(x, y, \varepsilon)). \end{aligned} \tag{125}$$

Notice that the critical manifold is, locally, a parabola

$$\mathcal{C}_0 = \left\{(x, y) \in \mathbb{R}^2 \,:\, y = x^2\right\}.$$

Our objective is to provide a sketch of the proof of the following theorem.

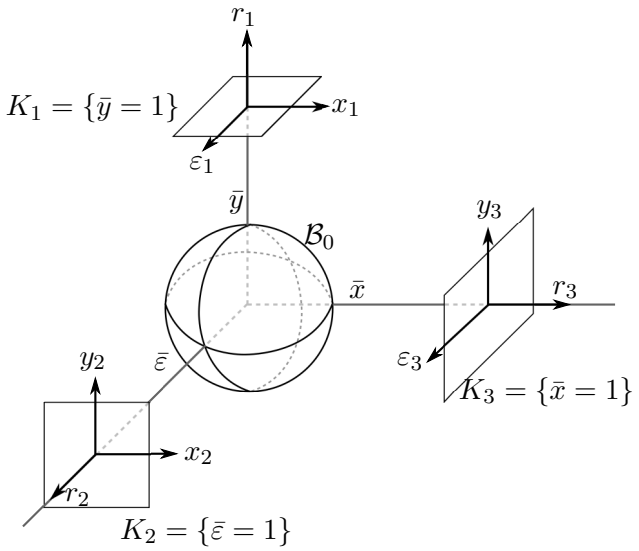


FIGURE 6. Sketch of the blown-up space and of some of the directional charts. In practice, via the blow-up method, we study local vector fields defined in the charts, and then "glue" trajectories and other invariant objects together to describe the dynamics in a small neighborhood of $\mathcal{B}_0$, which in turn provides the dynamics of a slow-fast system around the origin for $\varepsilon > 0$ sufficiently small.

THEOREM VII.2 ([18]). *Consider* (125) *and define the sections*

$$\Sigma^{\text{en}} = \left\{(x, y) \in \mathbb{R}^2 \,|\, x \in I,\, y = \rho^2\right\},$$
$$\Sigma^{\text{ex}} = \left\{(x, y) \in \mathbb{R}^2 \,|\, x = \rho,\, y \in \mathbb{R}\right\},$$

where $\rho > 0$ is small and I is a small suitable interval so that Σ^{en} intersects transversally the attracting part of the critical manifold. Let $\Pi : \Sigma^{\text{en}} \to \Sigma^{\text{ex}}$ be the transition map for the flow of (125)*. Then, there exists $\varepsilon_0 > 0$ such that the following assertions hold for $\varepsilon \in (0, \varepsilon_0]$:*

(F1) The manifold $\mathcal{S}^{\text{a}}_\varepsilon$ passes through Σ^{ex} at a point $(\rho, h(\varepsilon))$, where $h(\varepsilon) \in \mathcal{O}(\varepsilon^{2/3})$.
(F2) The transition Π is a contraction with contraction rate $\mathcal{O}(\exp(-C/\varepsilon))$, where $C > 0$.

A description of the dynamics near a generic fold point can be seen in figure 7.

For the blow-up analysis, we define the blow-up as

$$x = \bar{r}\bar{x},\ y = \bar{r}^2\bar{y},\ \varepsilon = \bar{r}^3\bar{\varepsilon}. \tag{126}$$

According to the entry and exit sections defined above, we define the charts $K_1 = \{\bar{y} = 1\}$, $K_2 = \{\bar{\varepsilon} = 1\}$, and $K_3 = \{\bar{x} = 1\}$.

The strategy to prove the theorem is as follows: first we consider the dynamics in the entry chart K_1. In K_1 one studies the dynamics along $\mathcal{S}^{\text{a}}_0$ approaching the fold. Next we study the dynamics of the chart K_2, where we "zoom-in" into the fold. Finally we consider the dynamics in the chart K_3 corresponding to the dynamics leaving a small neighborhood of the fold along the fast fibers.

VII.3.4. Analysis in the entry-chart K_1. According to the blow-up (126), the local coordinates in this chart $(r_1, x_1, \varepsilon_1)$ are defined as

$$x = r_1 x_1,\ y = r_1^2,\ \varepsilon = r_1^3 \varepsilon_1.$$

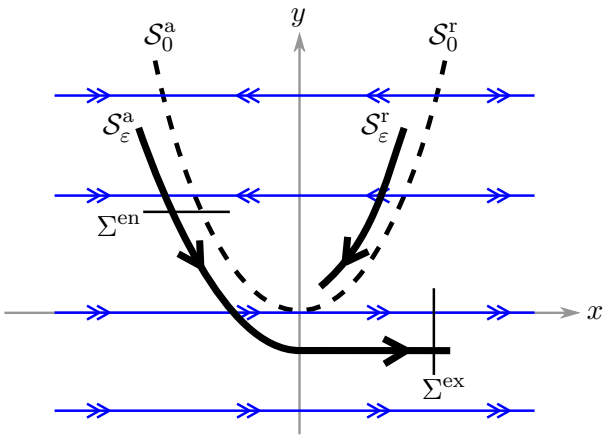


FIGURE 7. Schematic of a slow-fast system near a generic fold point. Up to leading order terms, the critical manifold $\mathcal{C}_0$, shown in dashed, is given by $\mathcal{C}_0 = \{(x,y) \in \mathbb{R}^2 \,|\, y = x^2\}$. The (blue) lines with double arrows depict the dynamics of the layer equation. Thus, $\mathcal{C}_0$ has an attracting ($\mathcal{S}_0^{\mathrm{a}}$) and a repelling branch ($\mathcal{S}_0^{\mathrm{r}}$). Away from the fold point, Fenichel's theorem shows that, for $\varepsilon > 0$ sufficiently small, $\mathcal{S}_0^{\mathrm{a}}$ and $\mathcal{S}_0^{\mathrm{r}}$ are smoothly perturbed to invariant manifolds (in this case trajectories) $\mathcal{S}_\varepsilon^{\mathrm{a}}$ and $\mathcal{S}_\varepsilon^{\mathrm{r}}$ respectively. The analysis (via the blow-up method) shows that $\mathcal{S}_\varepsilon^{\mathrm{a}}$ can be extended beyond the fold point as depicted in the figure. In particular, one can show that the distance between the x-axis and the intersection $\mathcal{S}_\varepsilon^{\mathrm{a}} \cap \Sigma^{\mathrm{ex}}$ is of order $\mathcal{O}(\varepsilon^{2/3})$. This is an example of a delayed loss of stability, or delayed bifurcation.

This change of coordinates induces the (desingularized) local vector field

$$\begin{aligned} x_1' &= -1 + x_1^2 + \frac{1}{2}\varepsilon_1 x_1 + \mathcal{O}(r_1) \\ r_1' &= \frac{1}{2} r_1 \varepsilon_1 (-1 + \mathcal{O}(r_1)) \\ \varepsilon_1' &= \frac{3}{2}\varepsilon_1^2 (1 + \mathcal{O}(r_1)). \end{aligned} \tag{127}$$

We notice that the sets $\{r_1 = 0\}$, $\{\varepsilon_1 = 0\}$ and their intersection $\{r_1 = \varepsilon_1 = 0\}$ are invariant. In the set $\ell_1 = \{r_1 = \varepsilon_1 = 0\}$ the dynamics are given by

$$x_1' = -1 + x_1^2.$$

Therefore, there are two equilibrium points $p_1^{\pm} = (\pm 1, 0, 0)$. Restricted to ℓ_1, both points are hyperbolic, and p_1^- is attracting while p_1^+ repelling.

The dynamics on the invariant plane $\{\varepsilon = 0\}$ read as

$$\begin{aligned} x_1' &= -1 + x_1^2 + \mathcal{O}(r_1) \\ r_1' &= 0. \end{aligned}$$

In this case, the lines $M_1^{\pm} = \{(x_1, r_1, \varepsilon_1) = (\pm 1, r_1, 0)\}$ are sets of normally hyperbolic equilibria. with M_1^- being attracting and M_1^+ repelling.

The dynamics on the invariant plane $\{r_1 = 0\}$ read as

$$\begin{aligned} x_1' &= -1 + x_1^2 + \frac{1}{2}\varepsilon_1 x_1 \\ r_1' &= \frac{3}{2}\varepsilon_1^2. \end{aligned}$$

For this system, the equilibria are $p_1^{\pm}$, but they are semi-hyperbolic. It is not difficult to check that each equilibrium point possesses a center manifold $N_1^{\pm}$ tangent to the eigenvector $(-1, 0, 4)$. In the case of N_1^-, the flow along N_1^- is directed away from p_1^-. Moreover N_1^- is unique (this fact will be important later). Similarly, the flow along N_1^+ is directed away from p_1^+ but it is not unique.

The 1-dimensional manifolds $N_1^{\pm}$ extend to 2-dimensional center manifolds, denoted by $M_1^{\pm}$ respectively, sufficiently close to ℓ_1, with the same stability properties and flow induced by $N_1^{\pm}$, as sketched in figure 8.

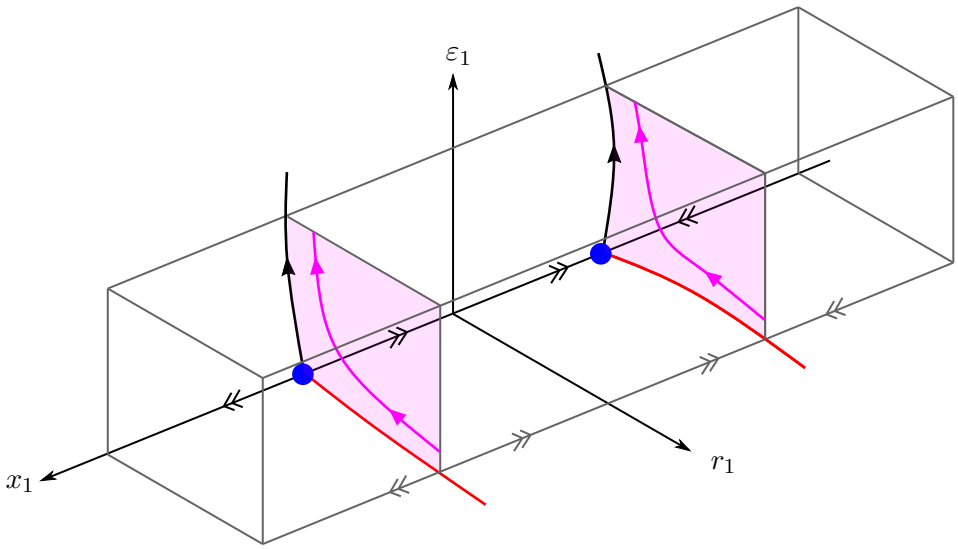


FIGURE 8. Sketch of the flow in chart K_1.

Let us define the sections

$$\Sigma_1^{\text{en}} = \{(x_1, r_1, \varepsilon_1) \in D_1 \,:\, r_1 = \rho\}$$
$$\Sigma_1^{\text{ex}} = \{(x_1, r_1, \varepsilon_1) \in D_1 \,:\, \varepsilon_1 = \delta\},$$

where D_1 is a small $3D$-rectangle

$$D_1 = \left\{(x_1, r_1, \varepsilon_1) \in \mathbb{R}^3 \,:\, x_1 \in \mathbb{R},\, 0 \le r_1 \le \rho,\, 0 \le \varepsilon_1 \le \delta\right\}.$$

We denote by $\Pi_1 : \Sigma_1^{\text{en}} \to \Sigma_1^{\text{ex}}$ the map induced by the flow of (127). For a sufficiently small region D_1, such a map is well defined. Moreover, by direct integration of the ε_1-equation in (127), the transition time from a point $p = (x_1, \rho, \varepsilon_1) \in \Sigma_1^{\text{en}}$ to the point $\Pi_1(p) \in \Sigma_1^{\text{ex}}$ is given by

$$T_1 = \frac{2}{3}\left(\frac{1}{\varepsilon_1} - \frac{1}{\delta}\right)(1 + \mathcal{O}(\rho)).$$

VII.3.5. Analysis in the central-chart K_2. The local coordinates in this chart are given by

$$x = r_2 x + 2,\; y = r_2^2 y_2,\; \varepsilon = r_2^3,$$

which lead to the (desingularized) local vector field

$$\begin{aligned} x_2' &= x_2^2 - y_2 + \mathcal{O}(r_2) \\ y_2' &= -1 + \mathcal{O}(r_2) \\ r_2' &= 0. \end{aligned} \tag{128}$$

For $r_2 = 0$, one obtains the Riccati equation

$$\begin{aligned} x_2' &= x_2^2 - y_2 \\ y_2' &= -1. \end{aligned} \tag{129}$$

The solutions of a Riccati equation can be expressed in terms of Bessel and/or Airy functions. What is most important for our purposes is the following:

PROPOSITION VII.2. *The Riccati equation (129) has the following properties:*

(1) Every orbit has a horizontal asymptote $y = y_r$, where y_r depends on the orbit, such that $x \to \infty$ as $y \to y_r$ from above.

(2) There exists a unique orbit γ_2 that can be parametrized as $(x_2, s(x_2))$, $x \in \mathbb{R}$. The orbit γ_2 is asymptotic to the left branch of the parabola $\{y_2 = x_2^2\}$ as $x_2 \to -\infty$ and has horizontal asymptote $y = -\Omega_0 < 0$ as $x_2 \to \infty$.

(3) The function $s(x)$ has the asymptotic expansion

$$s(x_2) = x_2^2 + \frac{1}{2x_2} + \mathcal{O}(x_2^{-4}),$$

as $x_2 \to -\infty$, and

$$s(x_2) = -\Omega_0 + \frac{1}{x_2} + \mathcal{O}(x_2^{-3}),$$

as $x_2 \to \infty$

(4) The constant Ω_0 is the smallest positive zero of

$$J_{-1/3}\left(\frac{2}{3}z^{3/2}\right) + J_{1/3}\left(\frac{2}{3}z^{3/2}\right),$$

where $J_\bullet$ are Bessel functions of the first kind.

(5) All orbits to the right of γ_2 are backward asymptotic to the left branch of the parabola $\{y_2 = x_2^2\}$ as $x_2 \to -\infty$, while all orbits of to the left of γ_2 have a horizontal asymptote $y_2 > y_r$ as $x_2 \to -\infty$.

The statements of the previous proposition are sketched in figure 9.

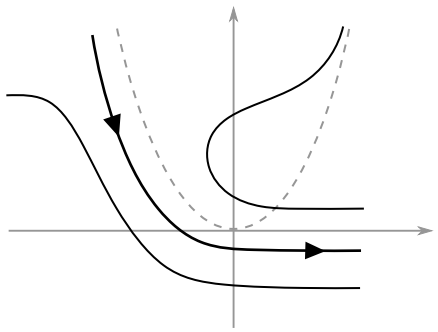


FIGURE 9. Flow in chart K_2.

Let us now define the sections

$$\begin{aligned} \Sigma_2^{\text{en}} &= \left\{(x_2, y_2, r_2) \in \mathbb{R}^3 \,:\, y_2 = \delta^{-2/3}\right\} \\ \Sigma_2^{\text{en}} &= \left\{(x_2, y_2, r_2) \in \mathbb{R}^3 \,:\, x_2 = \delta^{-1/3}\right\}. \end{aligned}$$

The reason of such a choice will be evident later when we glue together the dynamics on each of the charts.

Let $\Pi_2 : \Sigma_2^{\text{en}} \to \Sigma_2^{\text{ex}}$ be the transition map induced by the flow of (128), and let $q = \Sigma^{\text{en}} \cap \gamma_2$. Form the properties of the Riccati equation it follows that

$$\Pi_2(q) = (\delta^{-1/3}, -\Omega_0 + \delta^{1/3} + \mathcal{O}(\delta), 0).$$

Moreover, using regular perturbation arguments one can show that a small neighborhood of q is mapped (diffeomorphically) to a small neighborhood of $\Pi_2(q)$.

VII.3.6. Analysis in the exit-chart K_3. The local coordinates in this chart are given by:

$$x = r_3,\ y = r_3^2 y_3,\ \varepsilon = r_3^3 \varepsilon_3.$$

The corresponding desingularized vector fields then reads as

$$\begin{aligned} r_3' &= r_3 F(r_3, y_3, \varepsilon_3) \\ y_3' &= \varepsilon_3(-1 + \mathcal{O}(r_3)) - 2y_3 F(r_3, y_3, \varepsilon_3) \\ \varepsilon_3' &= -3\varepsilon_3 F(r_3, y_3, \varepsilon_3), \end{aligned} \tag{130}$$

where $F(r_3, y_3, \varepsilon_3) = 1 - y_3 + \mathcal{O}(r_3)$.

It is straightforward to check that the origin is a hyperbolic equilibrium point with eigenvalues $(\lambda_1, \lambda_2, \lambda_3) = (1, -2, -3)$. Notice that there is a resonance given by $\lambda_2 = \lambda_1 + \lambda_3$, and recall that this is a difficulty for the linearization. We will deal with this issue shortly.

As is previous charts, let us define the sections

$$\begin{aligned} \Sigma_3^{\text{en}} &= \{(r_3, y_3, \varepsilon_3) \in \mathbb{R}_{\geq 0} \times \mathbb{R} \times \mathbb{R}_{\geq 0} \,:\, r_3 \in [0, \rho],\ y_3 \in [-\beta_3, \beta_3],\ \varepsilon = \delta\} \\ \Sigma_3^{\text{ex}} &= \{(r_3, y_3, \varepsilon_3) \in \mathbb{R}_{\geq 0} \times \mathbb{R} \times \mathbb{R}_{\geq 0} \,:\, r_3 = \rho,\ y_3 \in [-\beta_3, \beta_3],\ \varepsilon \in [0, \delta]\}, \end{aligned}$$

where, for now, it may look arbitrary the repeated use of the parameters, but this will make sense in the next section. Let $\Pi_3 : \Sigma^{\text{en}} \to \Sigma^{\text{ex}}$ denote the map induces by the flow of (130). Our goal is to obtain a formula for Π_3 accounting for the resonance. For this, let us first divide (130) by F (which close to the origin is simply a smooth equivalence), obtaining:

$$\begin{aligned} r_3' &= r_3 \\ y_3' &= -2y_3 - \frac{\varepsilon_3}{1 - y_3} + r_3 \varepsilon_3 G(r_3, y_3, \varepsilon_3) \\ \varepsilon_3' &= -3\varepsilon_3, \end{aligned} \tag{131}$$

where G_3 is a sufficiently smooth function. We notice that the first resonant monomial is $r_3\varepsilon_3$ (because of $\lambda_2 = (1, 0, 1) \cdot (\lambda_1, \lambda_2, \lambda_3)$). Thus, we know that there is a near identity transformation $y_3 = h(r_3, \tilde{y}_3, \varepsilon_3) = \tilde{y}_3 + \mathcal{O}(r_3 y_3 \varepsilon_3)$ transforming (131) to

$$\begin{aligned} r_3' &= r_3 \\ \tilde{y}_3' &= -2\tilde{y}_3 - \varepsilon_3 + \mathcal{O}(r_3\varepsilon_3) \\ \varepsilon_3' &= -3\varepsilon_3. \end{aligned}$$

Now, we notice that $r_3\varepsilon_3 = r_3(0)\varepsilon_3(0)\exp(-2t)$, which coincides with the linear coefficient of $\tilde{y}_3$. Thus, integrating the $\tilde{y}_3$ we have

$$\tilde{y}_3 = (\tilde{y}_3(0) - \varepsilon_3(0))\exp(-2t) + \mathcal{O}(t_3 r_3(0)\varepsilon_3(0)\exp(-2t)),$$

where by t_3 we indicate the time in this chart. The transition time for the map Π_3 can be computed from $r_3' = r_3$, giving:

$$T = \ln\left(\frac{\rho}{r_i}\right),$$

where r_i denotes the coordinate of r_3 at Σ^{en}. Recalling that $\varepsilon_3(0) = \delta$ we have

$$\tilde{y}_3(T) = (\tilde{y}_3(0) - \delta)\left(\frac{r_i}{\rho}\right)^2 + \mathcal{O}\left(\frac{r_i^3}{\rho^2}\ln\left(\frac{\rho}{r_i}\right)\right).$$

We notice that in the previous equation we do not write the term δ inside the big-Oh because it is a fixed constant. Also, notice that $\mathcal{O}(r_3^3\ln(r_3^{-1})) = \mathcal{O}(r_3^3\ln(r_3))$. If we denote by $\tilde{h}$ the inverse of the function h, we then have that the transition map Π_3 reads as

$$\Pi_3(r_3, y_3, \delta) = \begin{pmatrix} \rho \\ \left(\tilde{h}(r_3, y_3, \delta) - \delta\right)\left(\frac{r_3}{\rho}\right)^2 + \mathcal{O}(r_3^3\ln r_3) \\ \delta\left(\frac{r_3}{\rho}\right)^3 \end{pmatrix}.$$

VII.3.7. Gluing the local results. We are now in position to prove Theorem VII.2. We do this using changes of coordinates between the charts given by maps $\kappa_{ij} : K_i \to K_j$ given by:

$$\begin{aligned}
&\kappa_{12}: && x_2 = x_1\varepsilon_1^{-1/3}, && y_2 = \varepsilon_1^{-2/3}, && r_2 = r_1\varepsilon_1^{1/3}, && \varepsilon_1 > 0,\\
&\kappa_{21}: && x_1 = x_2 y_2^{-1/2}, && r_1 = r_2 y_2^{1/2}, && \varepsilon_1 = y_2^{-3/2}, && y_2 > 0,\\
&\kappa_{23}: && r_3 = r_2 x_2, && y_3 = y_2 x_2^{-1}, && \varepsilon_3 = x_2^{-3}, && x_2 > 0,\\
&\kappa_{32}: && x_2 = \varepsilon^{-1/3}, && y_2 = y_3\varepsilon_2^{-2/3}, && r_2 = r_2\varepsilon_3^{1/3}, && \varepsilon_3 > 0.
\end{aligned} \tag{132}$$

We now have the following lemma.

Lemma VII.1.

(1) The unique branch N_1^- is chart K_1 (in $r_1 = 0$ and $\varepsilon > 0$) is equal to $\gamma_1 = \kappa_{21}(\gamma_2)$ (whenever κ_{21} is defined).
(2) The orbit $\gamma_3 = \kappa_{23}(\gamma_2)$ lies in the plane $\{r_3 = 0\}$, converges to the origin in K_3 as $\varepsilon_3 \to 0$, and is tangent at the origin (in K_3) to the y_3-axis.

Proof.

(1) First, we notice since we are considering the left branch of γ_2, we assume $x_2 \to -\infty$. It follows form (132) that

$$\gamma_1 = \left\{(x_1, r_1 = 0, \varepsilon_1) \ : \ x_1 = x_2\left(x_2^2 + \frac{1}{2x_2} + \mathcal{O}(x_2^{-4})\right)^{-1/2}, \varepsilon_1 = \left(x_2^2 + \frac{1}{2x_2} + \mathcal{O}(x_2^{-4})\right)^{-3/2}\right\}.$$

Next, we expand the above expression in terms of x_2 as $x_2 \to -\infty$ obtaining

$$\gamma_1 = \left\{(x_1, 0, \varepsilon_1) \ : \ x_1 = -1 + \frac{1}{4x_2^3} + \mathcal{O}(x_2^{-4}), \varepsilon_1 = -\frac{1}{x_2^3} + \mathcal{O}(x_2^{-4})\right\},$$

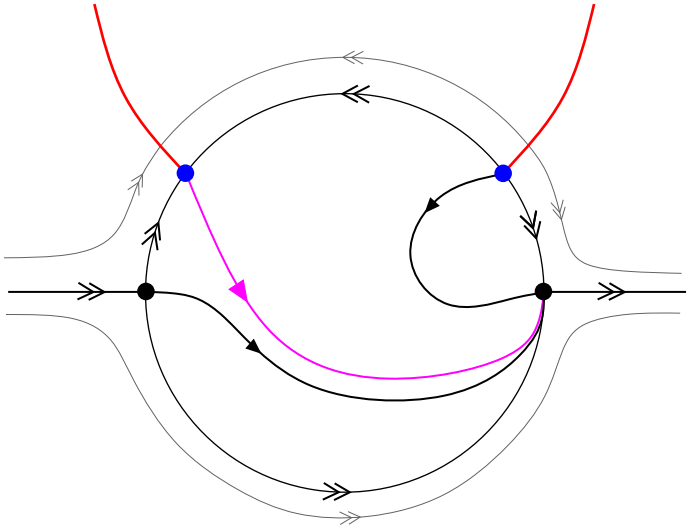


FIGURE 10. Blown-up flow (inside the circle). Outside the circle we show the (blown-up) layer dynamics for $\bar{\varepsilon} = 0$.

from where it follows that $\gamma_1 \to p_1^-$ as $\varepsilon_1 \to 0^+$ (or $x_2 \to -\infty$) and that γ_1 is tangent to $(1, 0, -4)$. This fact, together with the uniqueness of N_1^- shows the statement.

(2) In a completely analogous way as in the previous item we obtain that

$$\gamma_3 = \left\{(r_3 = 0, y_3, \varepsilon_3) \ : \ y_3 = -\Omega_0 \varepsilon_3^{2/3} + \varepsilon_3 + \mathcal{O}(\varepsilon_3^{4/3})\right\},$$

which shows the statement.

□

The previous lemma is schematized in figure 10

Next, let us define the map $\Pi : \sigma_1^{\text{en}} \to \Sigma_3^{\text{ex}}$ by

$$\Pi = \Pi_3 \circ \kappa_{23} \circ \Pi_2 \circ \kappa 12 \circ \Pi_1.$$

At this moment it is convenient to notice that Σ_1^{en} is the blow-up of Σ^{en} in chart K_1, $\Sigma_2^{\text{en}} = \kappa_{12}(\Sigma_1^{\text{ex}})$, $\Sigma_3^{\text{en}} = \kappa_{23}(\Sigma_2^{\text{ex}})$, and Σ_2^{ex} is the blow-up of Σ^{ex} in chart K_3.

From the analysis we did in the chart K_1, it follows that $\Pi_1(D_1 \cap M_1^-) \subset \Sigma_1^{\text{ex}}$ is a smooth curve transverse to $\{r_1 = 0\}$. Thus, $\kappa_{12}\left(\Pi_1(D_1 \cap M_1^-)\right) \subset \Sigma_2^{\text{en}}$ is a smooth curve transverse to $\{r_2 = 0\}$. From the analysis in K_2 we then know that $\Pi_2\left(\kappa_{12}\left(\Pi_1(D_1 \cap M_1^-)\right)\right)$ has the form

$$\left\{x_2 = \delta^{-1/3},\, y_2 = h_2^{\text{ex}}(r_2),\, r_2\right\}$$

with $r_2 \in [0, \rho\delta^{1/3}]$, and h_2^{ex} smooth. Such a curve, under the transformation κ_{23} reads as

$$\{r_3, h_3^{\text{en}}(r_3), \delta\}$$

with $(0, h_3^{\text{en}}(0), \delta) = \kappa_{23}(\gamma_2 \cap \Sigma_2^{\text{ex}})$. Then, the analysis performed in chart K_3 shows that $\Pi(D_1 \cap M_1^-)$ has the form

$$\{r_3 = \rho,\, y_3 = h_3^{\text{ex}}(\varepsilon_3), \varepsilon_3\}$$

where $h_3^{\text{ex}} = \mathcal{O}(\varepsilon^{2/3})$. This proves the first statement of theorem VII.2. The second statement (which we do not detail) follows from the local stability of the equilibria in charts K_1 and K_3 and the fact that all other transformations are diffeomorphisms.

EXERCISE VII.2. *Show that* $h_3^{out} = -\Omega_0 \varepsilon_3^{2/3} + \mathcal{O}(\varepsilon_3 \ln \varepsilon_3)$.

The proof is completed by blowing-down.

APPENDIX A

Background

A.1. Taylor series

Let a function $f : \mathbb{R}^n \to \mathbb{R}$ be k-times differentiable, $k \geq 1$. By $\mathrm{d}f(\boldsymbol{a}) : \mathbb{R}^n \to \mathbb{R}$ we denote the differential of f at a point $\boldsymbol{a} \in \mathbb{R}^n$, and it is given by

$$\mathrm{d}f(\boldsymbol{a})(\boldsymbol{v}) = \frac{\partial f}{\partial x_1}(\boldsymbol{a})v_1 + \cdots + \frac{\partial f}{\partial x_n}(\boldsymbol{a})v_n,$$

where $\boldsymbol{v} = (v_1, \ldots, v_n)$.

Let $\alpha \in \mathbb{N}^n$ and $\boldsymbol{x} \in \mathbb{R}^n$ and consider the multi-index notation

$$|\alpha| = \sum_{i=1}^{n} \alpha_i,$$

$$\alpha! = \prod_{i=1}^{n} \alpha_i!,$$

$$\boldsymbol{x}^\alpha = \prod_{i=1}^{n} x_i^{\alpha_i}.$$

We define the notation

$$\mathrm{D}^\alpha f = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}},$$

for $|\alpha| \leq k$. Then the multivariable version of Taylor's theorem lets us right

$$f(\boldsymbol{x}) = \sum_{|\alpha| \leq k} \frac{\mathrm{D}^\alpha f(\boldsymbol{a})}{\alpha!}(\boldsymbol{x} - \boldsymbol{a})^\alpha + \sum_{|\alpha| = k+1} h_\alpha(\boldsymbol{x})(\boldsymbol{x} - \boldsymbol{a})^\alpha,$$

where $\lim_{\boldsymbol{x} \to \boldsymbol{a}} h_\alpha(\boldsymbol{x}) = 0$. Of course, for $x \in \mathbb{R}$ the above formula is simply

$$f(x) = f(a) + \frac{\partial f}{\partial x}(a)(x - a) + \frac{\partial^2 f}{\partial x^2}(a)\frac{x - a}{2} + \cdots,$$

and for $\boldsymbol{x} \in \mathbb{R}^2$ it reads as

$$f(\boldsymbol{x}) = f(\boldsymbol{a}) + \frac{\partial f}{\partial x_1}(\boldsymbol{a})v_1 + \frac{\partial f}{\partial x_2}(\boldsymbol{a})v_2 + \frac{\partial^2 f}{\partial x_1^2}(\boldsymbol{a})\frac{v_1^2}{2!} + + \frac{\partial^2 f}{\partial x_1 \partial x_2}(\boldsymbol{a})v_1 v_2 + + \frac{\partial^2 f}{\partial x_2^2}(\boldsymbol{a})\frac{v_2^2}{2!} + \cdots,$$

where $\boldsymbol{v} = \boldsymbol{x} - \boldsymbol{a}$.

A.2. Dynamical Systems

In this section we recall some basic concepts of dynamical systems.

DEFINITION A.1 (Dynamical System). A dynamical system is a triplet (M, T, Φ), where M is a set (usually called the phase-space), $T \subseteq \mathbb{R}$ is the time set, and $\Phi : T \times M \to M$ is the evolution operator and satisfies:

$$\Phi(0, x) = x, \qquad \forall x \in M,$$

and

$$\Phi(s, \Phi(t, x)) = \Phi(s + t, x), \qquad \forall x \in M,\ s, t \in T.$$

For $T = \mathbb{Z}$, the dynamical system is called discrete, and for $T = \mathbb{R}$ it is called continuous.

EXAMPLE A.1 (Examples of discrete and continuous dynamical systems).

- Map iterations give rise to discrete dynamical systems. For example, consider the map

$$\phi(x) = \begin{cases} 2x, & \text{if} \quad 0 \leq x < \dfrac{1}{2} \\ 2x - 1, & \text{if} \quad \dfrac{1}{2} \leq x < 1, \end{cases}$$

  with $x \in [0, 1)$. If we take $T = \mathbb{Z}_+$ and $M = [0, 1)$ then the evolution operator $\Phi : \mathbb{Z}_+ \times M \to M$ can be defined by $\Phi : (t, x) \mapsto \phi^t(x) := \underbrace{\phi \circ \phi \circ \cdots \circ \phi}_{t\text{-times}}(x)$ providing a dynamical system. You may recall from your dynamical systems course that $\phi(x)$ is called *the doubling map*, and it has several interesting properties. You can simulate such a dynamical system and see what happens if the initial condition $\Phi(0, x)$ is rational or irrational.
- Ordinary differential equations, for example, give rise to continuous time dynamical systems, see section A.3. In the case of the simple pendulum (Example A.5) we have the ODE

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} x_2 \\ -\dfrac{g}{l} \sin x_1 + f(t) \end{bmatrix},$$

  where $x \in \mathbb{S}^1$ and $x_2 \in \mathbb{R}$. Therefore, the phase-space is $M = \mathbb{S}^1 \times \mathbb{R}$ and $T = \mathbb{R}$. On the other hand, the evolution operator is given by $\Phi : (t, (x_1(0), x_2(0))) \mapsto (x_1(t), x_2(t))$, where $(x_1(t), x_2(t))$ corresponds to the particular solution of the ODE with initial conditions $(x_1(0), x_2(0))$.

REMARK A.1. A usual assumption, that we shall adopt without further recalling, is that $\Phi_t(x) = \Phi(t, x)$ is, at the very least, a continuous function for every $t \in T$.

We now have a series of important terminology:

DEFINITION A.2. Given a dynamical system (M, T, Φ):

- For fixed $x \in M$, the function $\Phi_x(t) = \Phi(x, t)$, with $t \in T$, is called *the flow through* x and its graph is called *trajectory through* x.
- For fixed $x \in M$, the set $\gamma_x = \{\Phi(t, x) \in M \,|\, t \in T\}$ is called, the orbit through x.
- A subset $S \subseteq M$ is called Φ-invariant if for all $x \in S$ and $t \in T$, $\Phi(t, x) \in S$. When the context is clear, we simply say that a set is invariant.

- A point $x \in M$ such that $\gamma_x = x$ is called *a fixed point*.
- A point $x \in M$ is called *periodic* if there is a $\tilde{t} > 0$, $\tilde{t} \in T$ such that $\Phi(x, \tilde{t}) = x$. Any such $\tilde{t}$ is called *a period*, but $\tilde{t}$ is called *minimal period* if $\Phi(t, x) \neq x$ for all $t \in (0, \tilde{t})$. The corresponding trajectory and orbit (through x) are also called periodic.
- Given $x \in M$, we call:

$$\alpha(x) = \left\{ m \in M \,|\, \exists \{t_n\}_{n\in\mathbb{N}} \text{ with } \lim_{n\to\infty} t_n = -\infty, \text{ and } \lim_{n\to\infty} \Phi(t_n, x) = m \right\}$$
$$\omega(x) = \left\{ m \in M \,|\, \exists \{t_n\}_{n\in\mathbb{N}} \text{ with } \lim_{n\to\infty} t_n = \infty, \text{ and } \lim_{n\to\infty} \Phi(t_n, x) = m \right\},$$

  the α-limit set of x and the ω-limit set of x respectively.

  Notice that these sets are invariant. If x is a fixed point, then $\alpha(x) = \omega(x) = x$. Let γ be a periodic orbit, then $\alpha(x) = \omega(x) = \gamma$ for every $x \in \gamma$.

We now turn our attention to the important concept of stability.

Definition A.3. Let x be a fixed point of a dynamical system (M, T, Φ), with M being a normed space[1] and let $|\cdot|$ denote the corresponding norm.

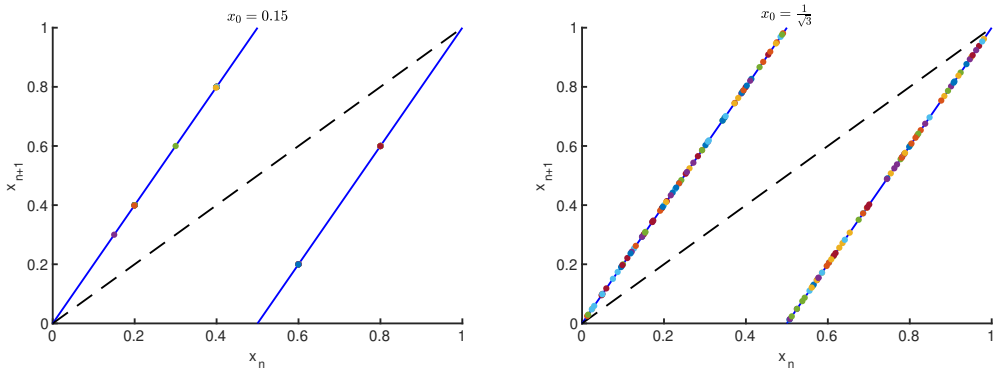
- The point x is called *Lyapunov stable* if given $\varepsilon > 0$, there exists a $\delta(\varepsilon) > 0$ such that for every $y \in M$ with $|y - x| < \delta(\varepsilon)$, it holds that $|\Phi(t, y) - \Phi(t, x)| < \varepsilon$ for all $t > 0$, $t \in T$. If the previous inequality does not hold, then the point x is called *unstable*.
- The point x is called *asymptotically stable* if x is Lyapunov stable, and moreover, for all $y \in M$ with $|y - x| < \delta(\varepsilon)$, $\lim_{t\to\infty} \Phi_t(y) = x$.

Example A.2.

- Consider the doubling map introduced above, namely

$$\phi(x) = \begin{cases} 2x, & \text{if } \quad 0 \leq x < \dfrac{1}{2} \\ 2x - 1, & \text{if } \quad \dfrac{1}{2} \leq x < 1, \end{cases}$$

  with $x \in [0, 1)$, and the dynamical system defined by iterations of ϕ. Such a dynamical system has the unique fixed point $x = 0$, which is unstable. Every rational point $x \in (0, 1)$, that is $x = \dfrac{p}{q}$, where p and q are positive integers with $p < q$, is a periodic point. The latter means that for every such point there is a periodic orbit passing through it. On the other hand, every orbit passing through an irrational point $x \in (0, 1)$ fills the entire interval as $t \to \infty$, and we say that the orbit is dense.



[1] otherwise one can adapt the definitions by using "neighborhoods"

FIGURE 1. Examples of the doubling map. On the left, convergence towards a periodic orbit $\{0.6, 0.2, 0.4, 0.8\}$, and on the right a dense orbit where we show a trajectory for $n = 150$ iterations.

- Consider the (unforced) pendulum

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} x_2 \\ -\frac{g}{l} \sin x_1 \end{bmatrix},$$

where $x \in \mathbb{S}^1$ and $x_2 \in \mathbb{R}$. This systems has fixed points: $(x_1, x_2) = (0, 0)$ and $(x_1, x_2) = (\pi, 0)$. The fixed point $(0, 0)$ is Lyapunov stable, while the fixed point $(\pi, 0)$ is unstable.

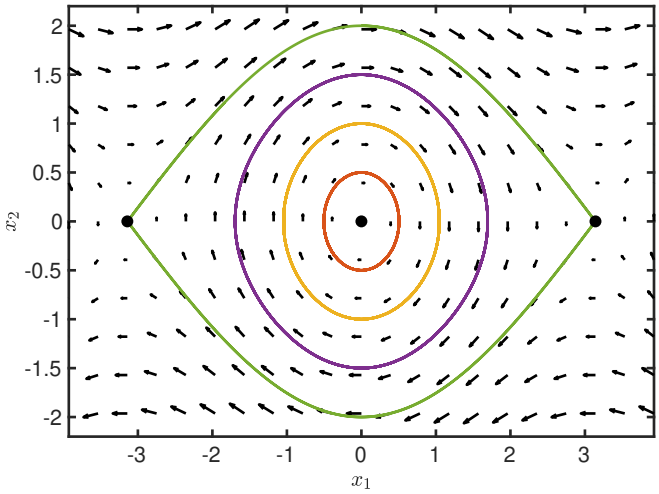


FIGURE 2. Phase portrait of an unforced pendulum.

A.3. Differential Equations

DEFINITION A.4 (Differential Equation (DE)). A *differential equation* is an equation involving one or several independent variables and the derivatives of one or several functions with respect to those variables. An **Ordinary Differential Equation (ODE)** is a differential equation depending on **one** independent variable. A differential equation involving more than one independent variable is called **Partial Differential Equation (PDE)**.

EXAMPLE A.3. Let $a \neq 0$ be some scalar constant. A scalar ordinary differential equation (ODE) is, for example, the following:

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = ax(t), \tag{133}$$

where $t \in \mathbb{R}$ is the independent variable, in this case *time*, and $x(t)$ is the *dependent variable* (because it depends on t). You may recall that the solution of (133) is

$$x(t) = x(0)e^{at},$$

where $x(0)$ is the *initial condition*, that is, the value of $x(t)$ at $t = 0$. Depending on the value of the constant a, the solution either approaches 0 or diverges, as shown in Figure 3.

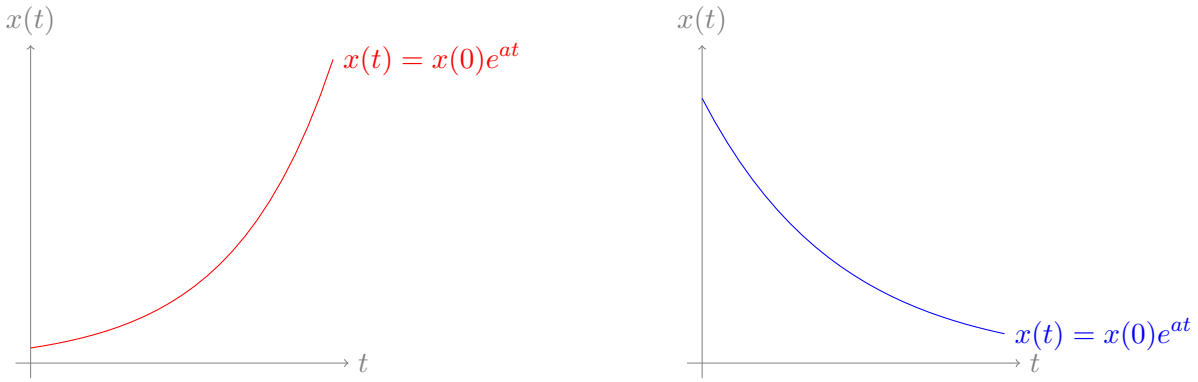


FIGURE 3. Plots of $x(t) = x(0)e^{at}$ for $a > 0$ on the left and $a < 0$ on the right.

The above plots are classical examples of exponential growth / decay.

REMARK A.2. In the example above, the notation $\frac{\mathrm{d}x(t)}{\mathrm{d}t}$ means "the derivative of the function $x(t)$ with respect to t". Such a derivative is a function. Thus, another common notation is $\frac{\mathrm{d}x}{\mathrm{d}t}(t)$, or even $\frac{\mathrm{d}x(t)}{\mathrm{d}t}(t)$. However, most of the times when there is no room for confusion, it is also enough to write $\frac{\mathrm{d}x}{\mathrm{d}t}$.

EXAMPLE A.4. Let $u = u(x,t)$ be a scalar function that depends on two scalar independent variables x, and t. The *one dimensional wave equation*

$$\frac{\partial^2 u}{\partial t^2}(x,t) = c^2 \frac{\partial^2 u}{\partial x^2}(x,t)$$

is an example of a partial differential equation (PDE). One can check that a solution of the wave equation is given by:

$$u(x,t) = f(x-ct) + g(x+ct), \tag{134}$$

where the functions f and g are at least twice differentiable and are the so-called *left* and *right* travelling waves, respectively. Indeed, to verify that (134) is, as claimed, the general solution it suffices to let $z_1 = x - ct$, $z_2 = x + ct$ and compute:

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial^2 f}{\partial z_1^2} + \frac{\partial^2 g}{\partial z_2^2} \qquad \frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 f}{\partial z_1^2} + c^2 \frac{\partial^2 g}{\partial z_2^2}$$

REMARK A.3. Notice the use of ∂ in the previous example. This is done to clearly distinguish between a 'total' derivative and a 'partial' derivative.

Along these notes, we will use the customary notation $\dot{x}(t) = \dfrac{\mathrm{d}x}{\mathrm{d}t}(t)$. Moreover, as long as there is no room for confusion, we shall omit the argument and thus write $\dot{x} = \dfrac{\mathrm{d}x}{\mathrm{d}t}$. In this way, we also use the notation: $\ddot{x} = \dfrac{\mathrm{d}^2 x}{\mathrm{d}t^2}$; and when more derivatives are involved: $x^{(n)} = \dfrac{\mathrm{d}^n x}{\mathrm{d}t^n}$ for some positive $n > 2$.

If $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{R}^m$ is a vector, then $\dot{\boldsymbol{x}} = (\dot{x}_1, \ldots, \dot{x}_n)$, and similarly for the higher derivatives.

DEFINITION A.5 (Order). The highest derivative appearing in a differential equation is called the *order of the differential equation*.

According to the previous definition, the ODE of example A.3 is of first order, while the PDE of example A.4 is of second order.

In general, an n-th order ODE is a function of the form:

$$F\left(t, x, \dot{x}, \ddot{x}, \ldots, x^{(n)}\right) = 0. \tag{135}$$

We say that an ODE is given in *explicit form* if (135) can be rewritten as

$$x^{(n)} = f\left(t, x, \dot{x}, \ddot{x}, \ldots, x^{(n-1)}\right), \tag{136}$$

otherwise we say it is given *implicitly* (or in implicit form). An explicit ODE can always be transformed into a system of first order differential equations. Indeed, consider (136) and define $u = (u_1, u_2, \ldots, u_n)$ by:

$$u_1 = x,\ u_2 = \dot{x}, \ldots,\ u_n = x^{(n-1)}.$$

Thus, we can write the n-dimensional first order system:

$$\dot{u} = \begin{bmatrix} \dot{u}_1 \\ \dot{u}_2 \\ \vdots \\ \dot{u}_n \end{bmatrix} = \begin{bmatrix} u_2 \\ u_3 \\ \vdots \\ f(t, u_1, u_2, \ldots, u_n) \end{bmatrix}.$$

EXAMPLE A.5. Consider the equation of the forced simple pendulum: $\ddot{\theta}+\frac{g}{l}\sin\theta = f(t)$. Letting $x = (x_1, x_2)^\top = (\theta, \dot{\theta})^\top$ we can rewrite the second order ODE as the 2-dimensional system of ODEs:

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\end{bmatrix} = \begin{bmatrix}x_2\\ -\frac{g}{l}\sin x_1 + f(t)\end{bmatrix}.$$

In general **a system of ordinary differential equations** is given by a set of equations of the form:

$$\begin{aligned}
F_1(t, x_1, \dot{x}_1, \dots, x_1^{(n)}, x_2, \dot{x}_2, \dots, x_2^{(n)}, \dots, x_m, \dot{x}_m, \dots, x_m^{(n)}) &= 0\\
F_2(t, x_1, \dot{x}_1, \dots, x_1^{(n)}, x_2, \dot{x}_2, \dots, x_2^{(n)}, \dots, x_m, \dot{x}_m, \dots, x_m^{(n)}) &= 0\\
&\vdots\\
F_k(t, x_1, \dot{x}_1, \dots, x_1^{(n)}, x_2, \dot{x}_2, \dots, x_2^{(n)}, \dots, x_m, \dot{x}_m, \dots, x_m^{(n)}) &= 0.
\end{aligned}$$

However, it is very common in models of natural phenomena that one is interested in an explicit first-order system of the form

$$\begin{aligned}
\dot{x}_1 &= f_1(t, x_1, \dots, x_n)\\
&\vdots\\
\dot{x}_n &= f_n(t, x_1, \dots, x_n),
\end{aligned}$$

or in compact form

$$\dot{\boldsymbol{x}} = f(t, \boldsymbol{x}), \tag{137}$$

where $\boldsymbol{x} = (x_1, \dots, x_n)^\top$, $f : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ is given component-wise by $f = (f_1, \dots, f_n)$, and each $f_i = f_i(t, \boldsymbol{x})$ is, say, differentiable in all its arguments (but see Theorem A.2 below). Of course, some more general expressions exist, but as we will see through these notes, most of the models we shall consider are either of the form (136) or (137).

DEFINITION A.6 (Solution). Consider an ODE (135) or (136). An n-times differentiable function on an open interval $x : I \to \mathbb{R}$, $I \subset \mathbb{R}$, is a solution of the ODE (135) or (136) if

$$F(t, x(t), \dot{x}(t), \dots, x^{(n)}(t)) = 0,$$

or

$$x^{(n)}(t) = f(t, x(t), \dot{x}(t), \dots, x^{(n-1)}(t)), \qquad t \in I,$$

respectively.

REMARK A.4. An intuitive way to think of a solution of an ODE is as a function that "satisfies" the ODE. That is, a function that after substitution into the ODE (either (135) or (136)) makes the equation hold.

EXAMPLE A.6 (A simple oscillator). Consider the harmonic oscillator $\ddot{\theta} = -\theta$, or equivalently by defining $(x_1, x_2) = (\theta, \dot{\theta})$:

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= -x_1. \end{aligned} \tag{138}$$

A solution is given by $\theta(t) = a\sin t + b\cos t$, for some arbitrary constants a, b. Equivalently, for (138), a solution is $\{x_1(t) = a\sin t + b\cos t,\ x_2(t) = a\cos t - b\sin t\}$. Indeed it is straightforward to verify that any of such solutions satisfy the corresponding differential equation.

Let us provide some further observations regarding solutions:

- Let $\boldsymbol{x}(t) = (x_1(t), \ldots, x_n(t))^\top$ denote a solution of an n-dimensional first-order system of ODEs $\dot{\boldsymbol{x}} = f(t, \boldsymbol{x})$, $\boldsymbol{x} \in \mathbb{R}^n$. The *initial condition* is the value of the solution at some initial time, usually denoted by t_0. That is, the initial condition corresponds to $\boldsymbol{x}(t_0) = \boldsymbol{x}_0$ for some $\boldsymbol{x}_0 \in \mathbb{R}^n$. A solution of an ODE satisfying $\boldsymbol{x}(t_0) = \boldsymbol{x}_0$ is called *particular solution*.
- By the equivalence between n-th order ODEs and n-dimensional first-order systems, we see that an initial condition for an n-th order ODE is given by specifying the initial values of all the $(n-1)$ derivatives of the dependent variable.
- A solution of an n-th order ODE is called *general*, if it contains n arbitrary parameters. For example, the solution of Example A.6 has the pair (a, b), making it a general solution.
- A solution of an ODE is called *complete* if all particular solutions can be obtained from the general solution, by taking appropriate values of the parameters.
- A particular solution that is not obtained from a (parameter dependent) general solution is called *singular*[2].
- For a time independent ODE $\dot{\boldsymbol{x}} = f(\boldsymbol{x})$, $\boldsymbol{x} \in \mathbb{R}^n$, a point $\boldsymbol{x}^*$ such that $f(\boldsymbol{x}^*) = 0$ is called *an equilibrium point*. Equilibrium points are solutions of and ODE.

EXAMPLE A.7 (Example A.6 continued). The ODE $\ddot{\theta}(t) = \theta(t)$ is of second order and the solution $\theta(t) = a\sin t + b\cos t$ has two arbitrary parameters a and b. Thus, the given solution is a general solution. Let $t_0 = 0$. Thus, the initial conditions are given by $\theta(0) = b$ and $\dot{\theta}(0) = a$. Notice that, for (138), the aforementioned initial conditions correspond also to $x_1(0) = b$ and $x_2(0) = a$ as expected. Thus one could also write the general solution as

$$\theta(t) = \dot{\theta}(0)\sin t + \theta(0)\cos t.$$

A particular solution is obtained by choosing values of $\theta(0)$ and of $\dot{\theta}(0)$.

EXAMPLE A.8 (The logistic model). The logistic model is broadly used to model population dynamics. Such a model is given by

$$\frac{\mathrm{d}P}{\mathrm{d}t} = kP\left(1 - \frac{P}{M}\right),$$

where $P = P(t)$ denotes the population size at time t, M is the so-called "carrying capacity" accounting for the maximum population that can be sustained, and k is a reproduction rate.

[2]A word of caution: as we will see in these notes, terms like 'singular', 'regular', 'singularity', among others mean different things in different contexts. One must always be sure of the precise working terminology when necessary.

The general solution is:

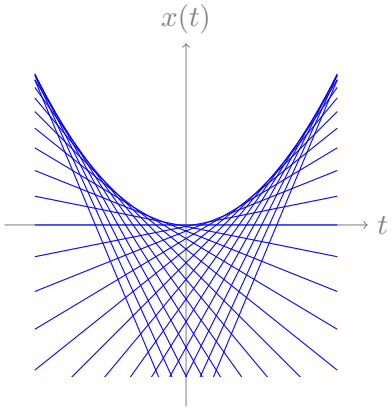
$$P(t) = \frac{M}{1 + Ae^{-kt}}, \qquad A = \frac{M - P(0)}{P(0)}.$$

However this general solution is not complete, since the solution $P(t) = 0$, which also satisfies the given ODE, cannot be obtained by any choice of the initial condition $P(0) > 0$ (notice that A is only defined for $P(0) > 0$). Of course, in this case, we can simply rewrite the solution as

$$P(t) = \frac{P(0)M}{P(0) + (A - P(0))e^{-kt}},$$

which is complete. Not all solutions become complete by simply rewriting it, as we see in the next example.

EXAMPLE A.9. Consider the equation $\left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)^2 - 4t\frac{\mathrm{d}x}{\mathrm{d}t} + 4x = 0$. The general solution, *but not complete*, is $x(t) = 2ct - c^2$, for some arbitrary constant $c \in \mathbb{R}$. Notice that such solutions are straight lines. Indeed, the solutions for different choices of c look like in the following picture:



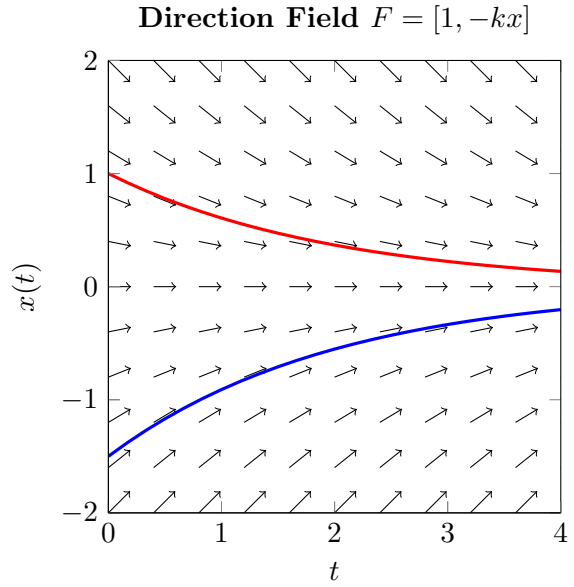
However, the function $x(t) = t^2$, corresponding to the envelope of the straight lines is also a valid singular solution.

Notice from the solution $x(t) = 2ct - c^2$, or also from the picture, that there are two solutions (straight-lines) passing through each point (t_0, x_0) with $x_0 < t_0^2$. On the other hand, there are no solutions for initial (t_0, x_0) such that $x_0 > t_0^2$ (above the parabola $x = t^2$).

Most of the times, it is not possible to solve a differential equation, that is to find an analytical solution. In some cases, the ***direction field*** can give us good enough graphical information to understand the overall behavior of the solutions.

To exemplify how to obtain a direction field, let us consider a differential equation $x'(t) = f(t, x)$ with $(t, x) \in U \subset \mathbb{R}$. We build a gird of point in U and at each point $(t, x) \in U$ we draw a little line segment with base at (t, x) and slope $f(t, x)$. Moreover, we can even draw the head of an arrow at the end of each of those line segments. In this way we build a field of arrows that encode, in their slopes, information about the differential equation. In this field, a particular solution is a curve $(t, x(t))$ that is everywhere *tangent* to a compatible sequence of line segments and "flows" in the direction provided by the tips of the arrows.

EXAMPLE A.10. Consider the simple ODE $\dot{x} = -kx$, $k > 0$. The general solution is $x(t) = c \cdot e^{-kt}$. To draw the direction field (also known as vector field) we draw at each point of the (t, x)-plane a little line segment with slope $-kx$ and an arrow tip on the end of such a line segment. In this way we form the "field of vectors" shown in the picture below. A particular solution is a curve that is everywhere tangent to such a field, we show two examples in the picture below.



In this plot every arrow has slope $-kx$. Notice that the arrows have been normalized in their length to aid the visualization.

We now present a few formal definitions and fundamental results regarding existence and uniqueness of solutions of ODEs $\dot{x} = f(t, x)$.

DEFINITION A.7.

- Let $(t, x) \in U \subseteq \mathbb{R} \times \mathbb{R}^n$. If U is open, and $f : U \to \mathbb{R}^n$ is continuous, then U is called *extended phase space*, f is *a time dependent vector field*, and the differential equation
$$\dot{x} = f(t, x)$$
is called *nonautonomous*.
- If $U = \mathbb{R} \times \tilde{U}$, with $\tilde{U} \subseteq \mathbb{R}^n$ open, and $f(t, x) = \tilde{f}(x)$, then the ODE is called *autonomous*.
- A differentiable function $\phi : I \to \mathbb{R}^n$, $I \subseteq \mathbb{R}$ is called *a solution to the ODE* if $\text{graph}(\phi) := \{(t, \phi(t)) \,|\, t \in I\} \subset U$ and
$$\frac{\mathrm{d}\phi}{\mathrm{d}t}\Big|_{t=\tau} = f(\tau, \phi(\tau)), \qquad \tau \in I.$$
- For $(t_0, x_0) \in U$, a map $\phi : I \to \mathbb{R}^n$ *satisfies the initial condition* (t_0, x_0) if $t_0 \in I$ and $\phi(t_0) = x_0$. Furthermore, ϕ *solves the initial value proble (IVP)* if
$$\frac{\mathrm{d}\phi}{\mathrm{d}t}\Big|_{t=\tau} = f(\tau, \phi(\tau)), \qquad \tau \in I, \qquad \phi(t_0) = x_0.$$
- The time dependent vector field $f : U \to \mathbb{R}^n$ satisfies:
  - *a global Lipschitz condition* with constant L if
$$\|f(t, x_0) - f(t, x_1)\| \leq L\|x_0 - x_1\|, \qquad (t, x_i) \in U,\ i = 0, 1,$$

– and a *a local Lipschitz condition* if each point $(\tau, x) \in U$ has a neighborhood $V \subseteq U$ such that $f|_U$ satisfies a Lipschitz condition in V, that is

$$\|f(t, x_0) - f(t, x_1)\| \leq L\|x_0 - x_1\|, \qquad (t, x_i) \in V,\ i = 0, 1,$$

for some constant $L = L(\tau, x)$.

Existence of solutions of ODEs is provided by Peano's theorem:

THEOREM A.1 (Peano). *Let $f : U \to \mathbb{R}^n$ be continuous. Then, for every $(t_0, x_0) \in U$ there exists a sufficiently small $\Delta t > 0$ and a solution $\phi : [t_0 - \Delta t, t_0 + \Delta t] \to \mathbb{R}^n$ of the ODE $\dot{x} = f(t, x)$ satisfying the initial condition $\phi(t_0) = x_0$.*

REMARK A.5. Solutions need not be unique, as already exemplified above.

REMARK A.6. Peano's theorem is sufficient: take the ODE $\dfrac{\mathrm{d}x}{\mathrm{d}t} = \dfrac{1}{x^2}$ with initial condition $x(t_0) = 0$. Indeed the function $f(t, x) = \dfrac{1}{x^2}$ is discontinuous at $x = 0$, however the ODE has solution $x(t) = (3(t - t_0))^{1/3}$ for any choice of $t_0 \in \mathbb{R}$.

Regarding uniqueness of solutions, we have the following:

DEFINITION A.8. Consider an ODE $\dot{x} = f(t, x)$. We say that a solution $\phi(t)$ to the initial value problem $\phi(t_0) = x_0$, $(t_0, x_0) \in U$, is *unique* if any two solutions $\phi_1 : I_1 \to \mathbb{R}^n$ and $\phi_2 : I_2 \to \mathbb{R}^n$ (to the same initial value problem) coincide on the interval $I = I_1 \cap I_2$.

REMARK A.7. Notice that, technically speaking, two solutions solving the IVP for different intervals I_i would be different because their domains are different. However, in principle we do not care to distinguish those because we are in principle interested in solutions to the IVP that are defined for the largest time interval possible. Hence the above definition.

THEOREM A.2 (Picard-Lindelöf). *If the time dependent vector field $f : U \to \mathbb{R}^n$ satisfies a Lipschitz condition in U, then for every $(t_0, x_0) \in U$ there exists an $\epsilon > 0$ such that the IVP $\dot{x} = f(t, x)$, $x(t_0) = x_0$ has unique solution $\phi : [t_0 - \epsilon, t_0 + \epsilon] \to \mathbb{R}^n$.*

PROOF. See Theorem 3.17 of [17]. □

The following lemma is useful in many situations.

LEMMA A.1. *If a time dependent vector field $f : U \to \mathbb{R}^n$ is continuously differentiable, then a local Lipschitz condition is satisfied on every* compact and convex *subset $V \subseteq U$ with Lipschitz constant $L := \sup_{(t,x)\in V} \|\mathrm{D}_x f(t, x)\|$.*

PROOF. See Lemma 3.14 in [17]. □

REMARK A.8. In fact, it is worth recalling that if f is a continuously differentiable vector field on $\mathbb{R}^n$, then f is (globally) Lipschitz if and only if $\sup_{x\in\mathbb{R}^n} \|\mathrm{D}f(x)\| < \infty$.

Next we have a result stating that "solutions of ODEs depend continuously on the initial conditions".

THEOREM A.3. *Under the same assumptions of Theorem A.2, each point $(t_0, x_0) \in U$ has a neighborhood $V \subset U$ and an interval $I_\epsilon = [-\epsilon, \epsilon]$ such that the family:*

$$\Phi : I_\epsilon \times V \to U$$
$$(s, (t_0, x_0)) \mapsto \phi(t_0 + s)$$

of solutions to the IVP $\dot{x} = f(t, x)$, $x(t_0) = t_0$ is a continuous mapping.

PROOF. See Theorem 3.20 of [17]. □

We now turn our attention to ODEs on manifolds. For the basics definitions and concepts regarding (differentiable) manifolds, look at appendix A of [17].

DEFINITION A.9.

- Let $f : M \to TM$ be a time independent vector field on a manifold M. A curve $\phi \in C^1(I, M)$ (the set of continuously differentiable mappings $\phi : t \mapsto \phi(t)$) is called *solution to the differential equation* $\dot{x} = f(x)$ if $\dfrac{\mathrm{d}\phi(t)}{\mathrm{d}t} = f(\phi(t))$ for all times $t \in I$.
- A vector field $f : M \to TM$ on a manifold M is called *complete* if for all $x_0 \in M$ the initial value problem $\dot{x} = f(x), \quad x(t_0) = x_0$ has a unique solution $\phi : \mathbb{R} \to M$. In other words, a vector field is called complete if each of its unique particular solutions exist for all time.

The following result gives conditions under which vector field on $\mathbb{R}^n$ are complete:

THEOREM A.4.

- *Lipschitz continuous (time independent / autonomous) vector fields $f : \mathbb{R}^n \to \mathbb{R}^n$ are complete.*
- *For time dependent vector fields: let $I \subseteq \mathbb{R}$ be an interval and let the time dependent vector field $f : I \times \mathbb{R}^n \to \mathbb{R}^n$ satisfy the time-dependent Lipschitz condition*

$$\|f(t, x_1) - f(t, x_2)\| \leq L(t)\|x_1 - x_2\|,$$

where $t \in I$ and $x_1, x_2 \in \mathbb{R}^n$, and with $L(t) > 0$ continuous. Then the initial value problem has a unique solution $\phi : I \to \mathbb{R}^n$ for all initial values $(t_0, x_0) \in I \times \mathbb{R}^n$

PROOF. See Theorem 3.23 in [17]. □

EXAMPLE A.11 (Several examples:).

- Consider $\dot{x} = x$, $x \in \mathbb{R}$. The vector field $f(x) = x$ is globally Lipschitz in $\mathbb{R}$. So, every solution $x(t) = x_0 e^t$ exists for all t.
- Consider $\dot{x} = x^2$, $x \in \mathbb{R}$. The vector field $f(x) = x^2$ is locally Lipschitz but not globally Lipschitz. The solution to the ODE $\dot{x} = x^2$ is $x(t) = \dfrac{x_0}{1 - x_0 t}$ which we see that diverges as t approaches $\dfrac{1}{x_0}$.
- (Exercise 3.25 of [17]) Consider $\dot{x} = \sin x$, $x(0) = \dfrac{\pi}{2}$. The particular solution is $x(t) = 2\cot^{-1}\left(e^{-t}\right)$, which exists and indeed is bounded for all t. Is the vector field $f(x) = \sin x$ globally Lipschitz? Is the vector field $f(x) = \sin x$ complete?
- (Lipschitz continuity is sufficient, but not necessary) Consider the vector field $f(x) = (x_1^2 + x_2^2)\begin{bmatrix} x_2 \\ -x_1 \end{bmatrix} \in \mathbb{R}^2$ and the ODE $\dot{x} = f(x)$ on $\mathbb{R}^2$. This vector field is not globally Lipschitz because its Jacobian is not bounded on the whole $\mathbb{R}^2$. However, it is not

difficult to show (use polar coordinates) that all solutions to the ODE are concentric circles. Thus, for any initial condition $x(0) \in \mathbb{R}^2$ a solution exists for all time $t \in \mathbb{R}$.

We have the following important theorem.

THEOREM A.5. *Lipschitz continuous vector fields on compact manifolds are complete.*

PROOF. See Theorem 3.27 of [17]. □

A.3.1. Linear, autonomous ODEs. In this section we recall some basic results concerning the differential equation

$$\dot{\boldsymbol{x}} = A\boldsymbol{x}, \qquad \boldsymbol{x} \in \mathbb{R}^n. \tag{139}$$

To motivate the study of linear systems, consider first the nonlinear ODE

$$\dot{\boldsymbol{x}} = f(\boldsymbol{x}), \qquad \boldsymbol{x} \in \mathbb{R}^n,$$

where f is at least twice differentiable. Suppose that $\boldsymbol{x}^* \in \mathbb{R}^n$ is an isolated equilibrium point. We can use Taylor series to obtain an approximate version of the ODE near the equilibrium point. Indeed let $\boldsymbol{y} = \boldsymbol{x} - \boldsymbol{x}^*$, thus $\boldsymbol{x} = \boldsymbol{x}^*$ corresponds to $\boldsymbol{y} = 0$. Then, by Taylor expanding near $\boldsymbol{y} = 0$ we get:

$$\dot{\boldsymbol{y}} = D_{\boldsymbol{x}} f(\boldsymbol{x}^*)\boldsymbol{y} + \cdots,$$

where $D_{\boldsymbol{x}} f(\boldsymbol{x}^*)$ is a constant matrix, and the $\cdots$ denote higher-order terms in $\boldsymbol{y}$. One would expect that the linear system $\dot{\boldsymbol{y}} = D_{\boldsymbol{x}} f(\boldsymbol{x}^*)\boldsymbol{y}$ provides some information of the dynamics of the nonlinear systems near the equilibrium point $\boldsymbol{x}^*$.

The solution of (139) is given by

$$\boldsymbol{x}(t) = \exp(tA)\boldsymbol{x}(0),$$

where $\exp(tA)$ is an $n \times n$ matrix given by:

$$\exp(tA) = I + \sum_{k=1}^{\infty} \frac{t^k}{k!} A^k.$$

EXERCISE A.1. *Show that* $\boldsymbol{x}(t) = \exp(tA)\boldsymbol{x}(0)$ *with* $\exp(tA) = I + \sum_{k=1}^{\infty} \frac{t^k}{k!} A^k$ *indeed solves* (139).

We shall denote by $M(t)$ the *fundamental matrix* of (139). The fundamental matrix has, as columns, n linearly independent solutions of (139). For example, if A has n linearly independent eigenvectors v_i, $i = 1, \ldots, n$, then each column of $M(t)$ is given by $\boldsymbol{x}_i(t) = \exp(\lambda_i t)v_i$. If there are less than n eigenvectors, then one should generate the so-called *generalized eigenvectors*.

EXERCISE A.2. *Let* $M(t) = [x_1(t), \ldots, x_n(t)]$ *where* $x_i(t) = \exp(t\lambda_i)v_i$ *and* (λ_i, v_i) *is a linearly independent eigenpair for* $i = 1, \ldots, n$. *Show that* $\exp(tA) = M(t)M(0)^{-1}$.
(Note: this result also holds when some of the v_i'*s are generalized eigenvectors).*

Notice that the linear subspaces spanned by the eigenvectors are invariant under the flow of (139). Indeed, if in particular we let v_i be a (real) eigenvector of A, we have that $\exp(tA)cv_i = c\left(I + \sum_{k=1}^{\infty} \frac{t^k}{k!} A^k\right) v_i = c\left(1 + \sum_{k=1}^{\infty} \frac{t^k}{k!} \lambda_i^k\right) v_i = c\exp(t\lambda_i)v_i \in \operatorname{span}\{v_i\}$.

EXERCISE A.3. *Prove that every (generalized) eigenspace of A is invariant under the flow of* (139).

This motivates the following definition.

DEFINITION A.10. Let $v_1, \dots, v_{n_s}$, $u_1, \dots, u_{n_u}$, and $w_1, \dots, w_{n_c}$ be the (generalized) eigenvectors associated to the n_s, n_u, n_c eigenvalues with negative, positive, zero real parts respectively. Then

$$E^s = \operatorname{span}\{v_1, \dots, v_{n_s}\}$$
$$E^u = \operatorname{span}\{u_1, \dots, u_{n_u}\}$$
$$E^c = \operatorname{span}\{w_1, \dots, w_{n_c}\},$$

are called the stable, unstable, and center subspaces respertively.

EXAMPLE A.12.

(1) Consider the linear system $\dfrac{\mathrm{d}x}{\mathrm{d}t} = \begin{bmatrix} 0 & 1 \\ 0 & -1 \end{bmatrix} x$. Thus $E^c = \operatorname{span}\left\{(1,0)^\top\right\}$, $E^s = \operatorname{span}\left\{(1,-1)^\top\right\}$, and $E^u = \emptyset$.

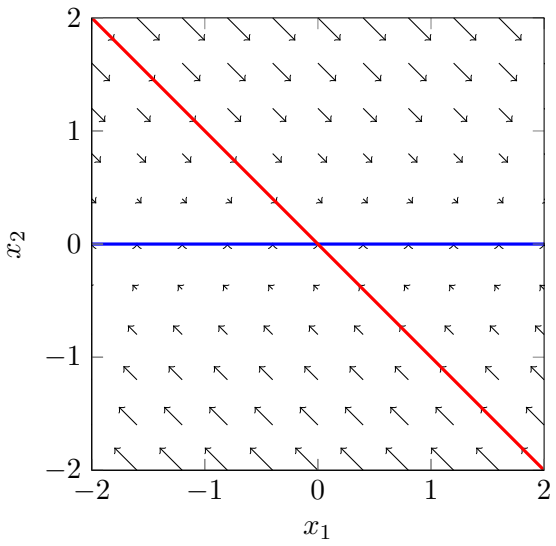


FIGURE 4. Phase portrait for this example. Notice that the center eigenspace E^c (blue) coincides with the set of equilibria $\{x_2 = 0\}$ and that it is attracting.

(2) Consider the system

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \begin{bmatrix} -1 & -1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

It follows that

$$E^s = \operatorname{span}\left\{\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}\right\}$$

$$E^u = \operatorname{span}\left\{\begin{bmatrix}0\\0\\1\end{bmatrix}\right\}$$

$$E^c = \emptyset.$$

EXERCISE A.4. *Find and compare the general solution of* $\frac{\mathrm{d}x}{\mathrm{d}t} = Ax$ *with* $A = \begin{bmatrix}0 & 0\\0 & 0\end{bmatrix}$ *and* $A = \begin{bmatrix}0 & 1\\0 & 0\end{bmatrix}$. *What do you notice?*

DEFINITION A.11. A matrix $A \in \mathbb{R}^{n\times n}$ is called *hyperbolic*, if all its eigenvalues have nonzero real part. Moreover, the linear system (139) and the equilibrium point $\boldsymbol{x}^* = 0$ are called hyperbolic if A is hyperbolic.

DEFINITION A.12. The index of a matrix A, denoted by $\operatorname{ind}(A)$, is the sum of the algebraic multiplicites of the eigenvalues of A with negative real part.

Notice that in the previous definition we have written "the equilibrium point". The uniqueness of the equilibrium point follows from the fact that hyperbolic matrices are invertible. We now present two useful results.

PROPOSITION A.1 ([15]). *If* (139) *is hyperbolic, then there exists a unique decomposition (or splitting) of* $\mathbb{R}^n$ *as* $\mathbb{R}^n = E^s \oplus E^u$, *where* E^s *and* E^u *are invariant under the flow of* (139). *Moreover, the flow restricted to* E^s *converges to the origin as* $t \to \infty$, *while the flow restricted to* E^u *converges to the origin as* $t \to -\infty$. *Notice that* $\dim E^s = \operatorname{ind}(A)$.

PROPOSITION A.2. *Let* $\dot{\boldsymbol{x}} = Ax$ *and* $\dot{\boldsymbol{x}} = Bx$ *be two hyperbolic linear systems. Then, there exists a (time-direction preserving) homeomorphism* $h : \mathbb{R}^n \to \mathbb{R}^n$ *mapping solutions of the first system to those of the second if and only if* A *and* B *have the same index. We then say that the systems are topologically equivalent.*

A.3.2. Nonlinear systems. In this section we provide a couple of important results for the smooth nonlinear system

$$\dot{\boldsymbol{x}} = f(x), \qquad \boldsymbol{x} \in \mathbb{R}^n, \tag{140}$$

and assume that $f(\boldsymbol{x}^*) = 0$. Let the linearization of (140) at $\boldsymbol{x} = \boldsymbol{x}^*$ be given by

$$\dot{\boldsymbol{x}} = A\boldsymbol{x}, \tag{141}$$

that is $A = D_x f(\boldsymbol{x}^*)$.

THEOREM A.6 (Hartman-Grobman). *If* A *is hyperbolic, then there exists a (time-direction preserving) homeomorphism* $h : U \to \mathbb{R}^n$, *defined on a neighborhood* U *of* $\boldsymbol{x}^*$, *mapping solutions of the*

nonlinear system (140) *to solutions of* (141). *In other words, if* A *is hyperbolic, then the nonlinear system* (140) *is topologically equivalent to its linearization* (141).

The next theorem generalizes Proposition A.1

THEOREM A.7 (Stable Manifold Theorem). *Consider* (140), *and let* $\boldsymbol{x}^*$ *be a hyperbolic equilibrium point. Let* $A = D_x f(\boldsymbol{x}^*)$ *have* n_s *and* n_u *eigenvalues with negative and positive real part respectively. Let* $U \subset \mathbb{R}^n$ *be a neighborhood of* $\boldsymbol{x}^*$. *There exist manifolds*

$$W^s(\boldsymbol{x}) = \left\{ \boldsymbol{x} \in U \,:\, \lim_{t\to\infty} \phi_t(\boldsymbol{x}) = \boldsymbol{x}^* \right\}$$
$$W^u(\boldsymbol{x}) = \left\{ \boldsymbol{x} \in U \,:\, \lim_{t\to-\infty} \phi_t(\boldsymbol{x}) = \boldsymbol{x}^* \right\}$$

that are smooth, invariant, and tangent to $E^s(\boldsymbol{x}^*)$ *and* $E^u(\boldsymbol{x}^*)$ *(the stable and unstable eigenspaces at* $\boldsymbol{x}^*$*) respectively. The dimension of* W^s *is* n_s *and of* W^u *is* n_u *and are called the stable and unstable manifolds of* $\boldsymbol{x}^*$.

EXAMPLE A.13. Let us consider the nonlinear system

$$\begin{aligned} \frac{\mathrm{d}x_1}{\mathrm{d}t} &= x_1 \\ \frac{\mathrm{d}x_2}{\mathrm{d}t} &= -x_2 + x_1^2. \end{aligned} \tag{142}$$

This system has a unique equilibrium point at the origin, and the linearized system is characterized by the matrix

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

Thus, it is clear that $E^s = \operatorname{span}\left\{(0,1)^\top\right\}$ and $E^u = \operatorname{span}\left\{(1,0)^\top\right\}$. Although for this example we can find the solutions analytically, let us try to find only the stable and unstable manifolds. For this, it is convenient to eliminate time by considering:

$$\frac{\mathrm{d}x_2}{\mathrm{d}x_1} = -\frac{x_2}{x_1} + x_1,$$

which has solution:

$$x_2(x_1) = \frac{x_1^2}{3} + \frac{c}{x_1}, \tag{143}$$

where c is the integration constant. We emphasize that (143) provides all solutions of (142). The unstable manifold W^u is tangent to the x_1-axis. Thus, if locally W^u is represented as a graph $x_2 = h(x_1)$ with $h(0) = \dfrac{\partial h}{\partial x_1}(0) = 0$, we can use (143) to write:

$$W^u(0) = \left\{ (x_1, x_2) \in \mathbb{R}^2 \,:\, x_2 = \frac{x^2}{3} \right\}.$$

For W^s it suffices to notice that if $x_1(0) = 0$, then $x_1(t) = 0$ for all $t > 0$, and thus, from (142), it follows that $W^s(0) = E^s$.

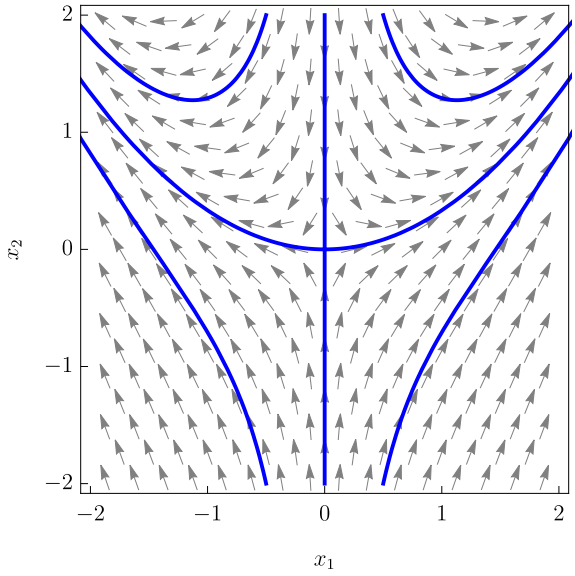


FIGURE 5. Phase portrait for this example. Notice that the unstable manifold is tangent to the x_1-axis, while the stable manifold coincides with the x_2-axis.

The next theorem generalizes Theorem A.7 to the case where the equilibrium point is nonhyperbolic.

THEOREM A.8 (Center Manifold Theorem). *Consider* (140) *with* f *a smooth vector field,* $f(x^*) = 0$ *such that* $A = D_x f(x^*)$ *has* n_s*,* n_u*, and* n_c *eigenvalues with positive, negative, and zero real parts respectively. Let the corresponding (generalized) eigenspaces be denoted by* E^s*,* E^u *and* E^c *respectively. Then, besides the stable and unstable manifolds as in Theorem A.7, there exists an invariant manifold* $W^c(x^*)$*, called* a center manifold*, tangent to* E^c *at* x^**. The center manifold is generally not unique.*

EXAMPLE A.14. Consider the planar system

$$\begin{aligned} \frac{\mathrm{d}x_1}{\mathrm{d}t} &= x_1^2 \\ \frac{\mathrm{d}x_2}{\mathrm{d}t} &= -x_2. \end{aligned} \tag{144}$$

It is clear that the origin $(x_1, x_2) = (0, 0)$ is the unique equilibrium point, and that the linearization at the origin is given by the matrix

$$A = \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}$$

The corresponding eigenspaces are $E^s = \operatorname{span}\left\{(0,1)^\top\right\}$ and $E^c = \operatorname{span}\left\{(1,0)^\top\right\}$ According to Theorem A.8 the origin possesses a (unique) 1-dimensional stable manifold tangent to the x_2-axis at the origin, and a 1-dimensional center manifold tangent to the x_1-axis at the origin. In fact $W^s = E^s$. The system (144) can be integrated, thus every solution is known analytically. Indeed, it is possible to show that every solution is given by

$$x_2(x_1) = \left(x_2(0) \exp\left(-\frac{1}{x_1(0)}\right)\right) \exp\left(\frac{1}{x_1}\right). \tag{145}$$

Notice that for all $x_1(0) < 0$ $x_2 \to 0$ as $x_1 \to 0$. Moreover,

$$\frac{\partial x_2}{\partial x_1} = -\left(x_2(0)\exp\left(-\frac{1}{x_1(0)}\right)\right)\exp\left(\frac{1}{x_1}\right)\frac{1}{x_1^2},$$

and thus $\lim_{x_1\to 0^-}\dfrac{\partial x_2}{\partial x_1} = 0$. In fact, due to the exponential term, the previous is true for any order of the derivative. This implies that any center manifold can be chosen as the union of a curve (145) for $x_1 < 0$ and the positive x_1-axis as shown in Figure

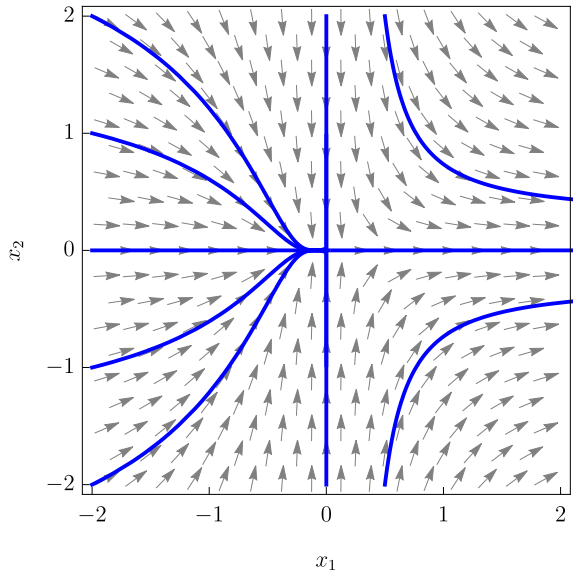


FIGURE 6. Phase portrait for this example. Notice that the center manifold is not unique (here E^c is the x_1-axis).

Bibliography

[1] kam theory: quasi-periodicity in dynamical systems.

[2] Vladimir I. Arnol'd. *Ordinary differential equations*. Springer Science & Business Media, 1992.

[3] Vladimir I. Arnol'd. *Mathematical methods of classical mechanics*, volume 60. Springer Science & Business Media, 2013.

[4] Vladimir Igorevich Arnold. *Geometrical methods in the theory of ordinary differential equations*, volume 250. Springer Science & Business Media, 2012.

[5] Nils Berglund. Perturbation theory of dynamical systems. *arXiv preprint math/0111178*, 2001.

[6] Jack Carr. *Applications of centre manifold theory*, volume 35. Springer Science & Business Media, 2012.

[7] Norman Levinson Earl A. Coddington. *Theory of Ordinary Differential Equations*. McGraw-Hill, 1987.

[8] Neil Fenichel. Geometric singular perturbation theory for ordinary differential equations. *Journal of differential equations*, 31(1):53–98, 1979.

[9] Marian Gidea, James D Meiss, Ilie Ugarcovici, and Howard Weiss. Applications of KAM theory to population dynamics. *Journal of Biological Dynamics*, 5(1):44–63, 2011.

[10] John Guckenheimer and Philip Holmes. *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*, volume 42. Springer Science & Business Media, 2013.

[11] Philip Hartman. *Ordinary differential equations*. SIAM, 2002.

[12] Michael R Herman. *Sur les courbes invariantes par les difféomorphismes de l'anneau*, volume 1. Société mathématique de France, 1983.

[13] Mark H. Holmes. *Introduction to perturbation methods*, volume 20. Springer Science & Business Media, 2012.

[14] Ale Jan Homburg and Björn Sandstede. Homoclinic and heteroclinic bifurcations in vector fields. *Handbook of dynamical systems*, 3:379–524, 2010.

[15] Welington de Melo (auth.) Jacob Palis Jr. *Geometric Theory of Dynamical Systems: An Introduction*. Springer, 1 edition, 1982.

[16] Hildeberto Jardón-Kojakhmetov and Christian Kuehn. A survey on the blow-up method for fast-slow systems. *arXiv preprint arXiv:1901.01402*, 2019.

[17] Andreas Knauf. *Mathematical physics: classical mechanics*, volume 109. Springer, 2018.

[18] Martin Krupa and Peter Szmolyan. Extending geometric singular perturbation theory to nonhyperbolic points - fold and canard points in two dimensions. *SIAM J. Math. Anal.*, 33:286–314, 2001.

[19] Christian Kuehn. *Multiple time scale dynamics*, volume 191. Springer, 2015.

[20] Peter D Lax. *Linear Algebra and Its Applications*. Pure and Applied Mathematics: A Wiley Series of Texts Monographs and Tracts. Wiley, 2nd edition edition, 2013.

[21] JD Meiss. Symplectic maps, variational principles, and transport. *Reviews of Modern Physics*, 64(3):795, 1992.

[22] Michael Spivak Michael Spivak. *A Comprehensive Introduction to Differential Geometry, Vol. 1, 3rd Edition Volume 1*. 3rd edition, 1999.

[23] E Mishchenko. *Differential equations with small parameters and relaxation oscillations*, volume 13. Springer Science & Business Media, 2013.

[24] EF Mishchenko, Yu S Kolesov, A Yu Kolesov, and N Kh Rozov. Asymptotic methods in singularly perturbed systems. *Monographs in Contemporary Mathematics (Consultants Bureau, New York, 1994)*, 1994.

[25] J Möser. On invariant curves of area-preserving mappings of an annulus. *Nachr. Akad. Wiss. Göttingen, II*, pages 1–20, 1962.

[26] Floris Takens. Constrained equations; a study of implicit differential equations and their discontinuous solutions. In *Structural stability, the theory of catastrophes, and applications in the sciences*, pages 143–234. Springer, 1976.

[27] Andrei Nikolaevich Tikhonov. Systems of differential equations containing small parameters in the derivatives. *Matematicheskii sbornik*, 73(3):575–586, 1952.

[28] Yieh-Hei Wan. Computation of the Stability Condition for the Hopf Bifurcation of Diffeomorphisms on R^2. *SIAM Journal on Applied Mathematics*, 34(1):167–175, 1978.