Lecture notes, Winter term 2018/19

Fallstudien der Mathematischen Modellbildung:

Asymptotic methods for perturbation problems

Approximation techniques in science and engineering

Stefan Possanner (*stefan.possanner@ma.tum.de*)

February 5, 2019

Technical University Munich
Department of Mathematics

Abstract

The goal of this lecture is to introduce some fundamental notions and techniques used in the asymptotic analysis of perturbation problems. Such problems are called singular if they undergo a change in their mathematical structure as the perturbation parameter ε tends to zero. A solution of the reduced problem ($\varepsilon = 0$) coincides with the limit solution of the full problem as $\varepsilon \to 0$ only if the perturbation is regular. It is the subject of asymptotic analysis to find approximate solutions of the full problem that are valid uniformly for $0 < \varepsilon \leq \varepsilon_0$, even if the perturbation is singular. Singular perturbation problems usually arise at the most critical (and interesting) regimes of physical modeling - their analysis and ultimate resolution has often lead to major advances in a specific field of science. In the first part of this course we focus on some basic principles and examples in the context of ordinary differential equations: we introduce the principle of dominant balance and discuss boundary layers, the WKB method, the method of (variational) averaging and the method of multiple scales. The guiding-center approximation of plasma physics is considered as a generic example of nonlinear perturbation theory. In the second part we extend our analysis to partial differential equations and present Prandtl's boundary layer for the Navier-Stokes equation. Moreover, we elaborate on macroscopic limits of kinetic equations in the strongly collisional regime, leading to fluid models of reduced dimensionality.

Contents

1. Introduction

A mathematical problem is deemed a perturbation problem if it is "close" to a simpler problem for which the solution is either known or can be computed with standard techniques. The closeness is usually measured in terms of a dimensionless parameter $\varepsilon \ll 1$ in the governing equations, which are typically systems of algebraic and/or differential equations with suitable initial and boundary conditions, in a way that setting $\varepsilon = 0$ yields the standard problem - henceforth called the *reduced problem*. In asymptotic analysis, the approach is to view the solutions of the governing equations as functions of ε, i.e. as a family of solutions depending on the parameter ε, and to construct an approximation to this family in the form of a series expansion in terms of simple functions of ε (typically power series in ε^n, $n \geq 0$). The big advantage is that this series can be computed term by term from simplified equations and is thus much easier to obtain than the exact solution. It occurs quite often that such an *asymptotic expansion* (AE) is the basis for a numerical investigation of the problem, which would otherwise be stiff or simply too large (in terms of degrees of freedom) to solve.

In *regular perturbation problems* the lowest order of the AE is indeed the solution of the reduced problem with $\varepsilon = 0$ across the whole domain of interest. In this case, it is straightforward to derive a system of equations with suitable initial and boundary conditions for the terms in the series, which can then be solved recursively. The lower the value of ε, the better the approximation obtained via the series expansion. We shall study regular perturbations of nonlinear ordinary differential equations (ODEs) in Chapter 3. A perturbation problem that is not regular is called *singular*. For singular problems the limiting behavior $\varepsilon \to 0$ is not captured by naive AE and the above procedure fails. In ODEs for example, singular problems occur when the derivative of the highest order is of size ε or smaller, which leads to the formation of boundary layers in certain regions of the domain as $\varepsilon \to 0$. This is because the order of the reduced problem is less than the number of initial/boundary conditions. Hence, a more subtle treatment is required to capture the correct asymptotics uniformly in the domain of interest, which is the subject of Chapter 4. Regular expansions may also fail when the domain is infinite, i.e. when small errors accumulate and become large over long times due to so-called *secular terms*. The *method of averaging* and the *method of multiple scales* can deal with secularities and will be discussed in Chapter **??**. With regard to partial differential equations (PDEs), singular problems occur when the type of the PDE changes in the reduced problem, or when the boundary conditions are such that the reduced problem is ill-posed (has no unique solution). We will present some generic examples of linear elliptic and hyperbolic equations in Chapter **??**. In particular, we shall revisit Prandtl's analysis (from 1904) of boundary layer formation around a body in a nearly inviscid

flow. This is a prime example of the power of asymptotic analysis for advancing hard physical problems, where straightforward reasoning might stall. Finally, in Chapter 5 we illustrate more examples in the context of kinetic-fluid transitions in gas dynamics and magnetized plasmas.

1.1. Regular and singular problems: algebraic examples

The difference between regular and singular perturbation problems can be readily understood by means of the following two algebraic examples.

Example 1. Consider the cubic equation

$$x_\varepsilon^3 - x_\varepsilon + \varepsilon = 0\,. \tag{1.1}$$

We view the roots x_ε as a family of solutions depending on the parameter ε. The reduced problem is obtained by setting $\varepsilon = 0$,

$$x_0^3 - x_0 = 0\,, \tag{1.2}$$

which yields the three roots $x_0 \in \{0, \pm 1\}$. In order to produce an AE of the family x_ε we assume

$$x_\varepsilon = x_0 + \varepsilon\, x_1 + \varepsilon^2\, x_2 + O(\varepsilon^3)\,. \tag{1.3}$$

We will clarify the meaning of the symbol $O(\varepsilon^3)$ later, here it is sufficient to know that it describes terms that tend to zero at least as ε^3 when $\varepsilon \to 0$ (we say these terms are "of order 3"). Inserting the AE into (1.1) and ordering terms in powers of ε yields

$$\begin{aligned} 0 &= (x_0 + \varepsilon\, x_1 + \varepsilon^2\, x_2 + \ldots)^3 - x_0 - \varepsilon\, x_1 - \varepsilon^2\, x_2 + \varepsilon + O(\varepsilon^3) \\ &= (x_0^3 - x_0) + \varepsilon\,(3x_0^2 x_1 - x_1 + 1) + \varepsilon^2 (3x_0^2 x_2 + 3x_0 x_1^2 - x_2) + O(\varepsilon^3)\,. \end{aligned} \tag{1.4}$$

Since we assume the expansion (1.3) to be valid for a finite interval $(0, \varepsilon_0]$ of ε-values, and moreover the functions ε^n are linearly independent, the coefficients in (1.4) must vanish. This leads to the system

$$\begin{aligned} x_0^3 - x_0 &= 0\,, \\ 3x_0^2 x_1 - x_1 + 1 &= 0\,, \\ 3x_0^2 x_2 + 3x_0 x_1^2 - x_2 &= 0\,, \\ &\vdots \end{aligned}$$

In the first equation we recognize the reduced problem (1.2). The other equations are linear for x_1, x_2, etc. and can be solved recursively:

$$x_1 = \frac{1}{1 - 3x_0^2}\,, \qquad x_2 = \frac{3x_0 x_1^2}{1 - 3x_0^2}\,, \qquad \ldots\,.$$

The AEs for the three roots of (1.1) thus read

$$x_{I,\varepsilon} = \varepsilon + O(\varepsilon^3)\,,$$
$$x_{II,\varepsilon} = 1 - \frac{\varepsilon}{2} - \frac{3\,\varepsilon^2}{8} + O(\varepsilon^3)\,,$$
$$x_{III,\varepsilon} = -1 - \frac{\varepsilon}{2} + \frac{3\,\varepsilon^2}{8} + O(\varepsilon^3)\,.$$

We leave it up to the reader to verify that these are indeed the first terms of the Taylor expansions of the exact roots.

Example 2. Consider now the following cubic equation:

$$\varepsilon\, x_\varepsilon^3 - x_\varepsilon + 1 = 0\,. \tag{1.5}$$

Inserting the ansatz (1.3) and equating the coefficients of ε^n to zero yields

$$-x_0 + 1 = 0\,,$$
$$x_0^3 - x_1 = 0\,,$$
$$3x_0^2 x_1 - x_2 = 0\,,$$
$$\vdots$$

such that the AE reads

$$x_\varepsilon = 1 + \varepsilon + 3\,\varepsilon^2 + O(\varepsilon^3)\,.$$

Here, we obtained only one root of the cubic equation, because the problem degenerates to a linear equation when setting $\varepsilon = 0$. **Such a qualitative change in the mathematical nature is typical for a singular perturbation problem**. What happened to the other two roots? In the present case they cannot be described with an AE of the form (1.3) because they tend to infinity as $\varepsilon \to 0$. In order to recover the correct asymptotics of these roots we need to reformulate the problem (1.5) into a regular perturbation problem. This can be done via a change of variables of the form

$$x_\varepsilon = \frac{y_\varepsilon}{\delta(\varepsilon)}\,,$$

where $y_\varepsilon = O(1)$ as $\varepsilon \to 0$ and the function $\delta(\varepsilon)$ is still to be determined. The equation for y_ε reads

$$\frac{\varepsilon}{\delta(\varepsilon)^3}\, y_\varepsilon^3 - \frac{1}{\delta(\varepsilon)}\, y_\varepsilon + 1 = 0\,. \tag{1.6}$$

To obtain a non-trivial reduced problem we require at least two leading-order terms in (1.6) to be of the same order in ε. This is called the *principle of dominant balance*, which we will encounter repeatedly in this course. Following this principle, there will be

one choice of $\delta(\varepsilon)$ which leads to meaningful results. Balancing the first two terms leads to

$$\frac{\varepsilon}{\delta(\varepsilon)^3} = \frac{1}{\delta(\varepsilon)} \quad \Longrightarrow \quad \delta(\varepsilon) = \sqrt{\varepsilon}\,.$$

With this choice of $\delta(\varepsilon)$ we obtain the regular perturbation problem

$$y_\varepsilon^3 - y_\varepsilon + \sqrt{\varepsilon} = 0\,,$$

which we can solve in the same way as (1.1) with the ansatz

$$y_\varepsilon = y_0 + \sqrt{\varepsilon}\, y_1 + \varepsilon\, y_2 + O(\varepsilon^{3/2})\,,$$

The non-zero solutions read

$$y_\varepsilon = \pm 1 - \frac{\sqrt{\varepsilon}}{2} \mp \frac{3\,\varepsilon}{8} + O(\varepsilon^{3/2})\,,$$

which yields the missing roots

$$x_\varepsilon = \pm\frac{1}{\sqrt{\varepsilon}} - \frac{1}{2} \mp \frac{3\sqrt{\varepsilon}}{8} + O(\varepsilon)\,.$$

The limit $\varepsilon \to 0$ in such an AE with $x_\varepsilon = O(\varepsilon^{-1/2})$ is called a *distinguished limit*.

There are two other possibilities for balancing two terms in (1.5). Balancing the last two terms leads to $\delta(\varepsilon) = 1$ and thus to the original problem - this is clearly a bad choice. On the other hand, balancing the first and the third term leads to $\delta(\varepsilon) = \varepsilon^{1/3}$. However, in this case the second term would be of order $O(\varepsilon^{-1/3})$ and thus dominate the other two terms, which violates the principle of dominant balance. In this example, no three-term dominant balance is possible but this can occur in other problems.

Example 3. As a last algebraic example consider the quadratic equation

$$(1-\varepsilon)\, x_\varepsilon^2 - 2\, x_\varepsilon + 1 = 0\,.$$

Trying the ansatz

$$x_\varepsilon = x_0 + \varepsilon\, x_1 + O(\varepsilon^2) \tag{1.7}$$

leads to

$$x_0^2 - 2\, x_0 + 1 = 0\,,$$
$$2x_0x_1 - x_0^2 - 2x_1 = 0\,.$$

From the first equation we obtain $x_0 = 1$ as a double root and then the second equation yields the contradiction $-1 = 0$. Hence, a solution of the form (1.7) cannot exist. The difficulty arises because $x_0 = 1$ is a repeated root of the reduced problem and thus the exact solution

$$x_\varepsilon = \frac{1 \pm \sqrt{\varepsilon}}{1-\varepsilon}$$

does not have a power series expansion in ε, but rather in $\sqrt{\varepsilon}$. An expansion of the form

$$x_\varepsilon = x_0 + \sqrt{\varepsilon}\, x_1 + \varepsilon\, x_2 + O(\varepsilon^{3/2})$$

leads to $x_0 = 1$ and $x_1^2 = 1$, which yields the correct expansion

$$x_\varepsilon = 1 \pm \sqrt{\varepsilon} + O(\varepsilon)\,.$$

1.2. The process of "non-dimensionalization"

The dimension-less parameter ε is a key factor in perturbation problems, as it allows to give a mathematical meaning to the transition between the reduced and the full problem. However, science problems are usually written in terms of dimensional quantities, i.e. variables with physical dimensions such as time, length or temperature. Most problems consist of equations that model real-world phenomena, and are thus linked at some point to experimental observations in which the model is rooted - hence the dimensional nature of the variables. The first task in the perturbative treatment of a science problem is thus to write it in non-dimensional form, thereby identifying the parameter ε. This crucial step enables the formulation of the "physical" problem as a well-defined perturbation problem. Depending on the considered problem parameters, the same model equations may lead to either a regular perturbation problem, a singular perturbation problem, or not a perturbation problem at all ($\varepsilon \approx 1$)! It is the physical information of the considered scenario, i.e. the size of the problem parameters, that in the end defines the nature of the perturbation problem.

As we will see later, the term "asymptotic" means "for ε sufficiently small". Hence, in mathematical terms, an asymptotic approximation - that is is an AE which is asymptotically equivalent to the exact solution (clarified later) - can be made arbitrarily precise by rendering ε as small as necessary. By contrast, in a given non-dimensional science problem the size of ε is fixed, usually denoting the ratio of two characteristic problem parameters. The validity of the asymptotic approximation then has to be checked for this particular value of ε. This can be done for instance if explicit expression for the remainder of the series expansion are available.

Example 4. We will now familiarize ourselves with the concept of *non-dimensionalization* (or *scaling*) by considering the damped harmonic oscillator. Let $x(t)$ denote the displacement of a mass $m > 0$ attached to spring from its equilibrium position as a function of time. If the mass is set into motion from its equilibrium position with an impulse p_0, the ensuing dynamics can be described in terms of the following initial value problem (IVP):

$$\begin{cases} m\,\dfrac{\mathrm{d}^2x}{\mathrm{d}t^2} + \beta\,\dfrac{\mathrm{d}x}{\mathrm{d}t} + k\,x = 0\,, \\ x(0) = 0\,, \qquad \dfrac{\mathrm{d}x}{\mathrm{d}t}(0) = \dfrac{p_0}{m}\,. \end{cases} \tag{1.8}$$

Here, $k > 0$ stands for the spring constant and $\beta \geq 0$ denotes the damping coefficient. Scaling the problem (1.8) starts with writing the dependent variable x and the independent variable t in terms of characteristic units of length L and of time T, respectively,

$$x(t) = L\,x'(t')\,, \qquad t = T\,t'\,.$$

Here, x' and t' are dimension-less. From the chain rule we have

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = \frac{L}{T}\,\frac{\mathrm{d}x'(t')}{\mathrm{d}t'}\,, \qquad \frac{\mathrm{d}^2x(t)}{\mathrm{d}t^2} = \frac{L}{T^2}\,\frac{\mathrm{d}^2x'(t')}{(dt')^2}\,.$$

Inserting this into (1.8) yields

$$\begin{cases} \dfrac{mL}{T^2}\,\dfrac{\mathrm{d}^2x'}{(dt')^2} + \dfrac{\beta L}{T}\,\dfrac{\mathrm{d}x'}{\mathrm{d}t'} + kL\,x' = 0\,, \\[2ex] L\,x'(0) = 0\,, \qquad \dfrac{L}{T}\,\dfrac{\mathrm{d}x'}{\mathrm{d}t'}(0) = \dfrac{p_0}{m}\,. \end{cases}$$

Now this problem is still dimensional but the dimensions (or units) have been made explicit via the pre-factors of each term. The dimension-less form is obtained by dividing by one of the pre-factors which leads for example to

$$\begin{cases} \dfrac{m}{kT^2}\,\dfrac{\mathrm{d}^2x'}{(dt')^2} + \dfrac{\beta}{kT}\,\dfrac{\mathrm{d}x'}{\mathrm{d}t'} + x' = 0\,, \\[2ex] x'(0) = 0\,, \qquad \dfrac{\mathrm{d}x'}{\mathrm{d}t'}(0) = \dfrac{p_0T}{mL}\,. \end{cases}$$

Our next task is to identify the small parameter ε. Hence we need to assign, a priori, a "size" to each of the terms in the problem. This requires additional information regarding the physical scenario we aim to consider. First of all, we identify two characteristic time scales of the problem:

$$\tau_1 = \sqrt{\frac{m}{k}} \qquad \text{and} \qquad \tau_2 = \frac{\beta}{k}\,.$$

Here, τ_1 is the period of the undamped oscillator and τ_2 is a characteristic damping time. With the choice of the time scale T we can determine which phenomena we want to resolve on a scale of order one, thus T is also called the *time scale of observation*. Setting $T = \tau_1$ will resolve the oscillatory phenomena, while setting $T = \tau_2$ will lead us to observe the damping of the oscillator on a scale of order one. The length scale L can be determined by requiring an initial velocity $\mathrm{d}x'/\mathrm{d}t'$ of order one, hence $L = p_0T/m$. Let us compare two physical scenarios:

1. Weakly damped oscillator: $\tau_2 \ll \tau_1$ and $T = \tau_1$. Setting

$$\varepsilon := \frac{\tau_2}{\tau_1}\,,$$

and omitting the primes for lighter notation, this leads to

$$P_\varepsilon^{(\mathrm{osc})} \begin{cases} \dfrac{\mathrm{d}^2 x_\varepsilon}{\mathrm{d}t^2} + \varepsilon \dfrac{\mathrm{d}x_\varepsilon}{\mathrm{d}t} + x_\varepsilon = 0\,, \\ \\ x_\varepsilon(0) = 0\,, \qquad \dfrac{\mathrm{d}x_\varepsilon}{\mathrm{d}t}(0) = 1\,. \end{cases} \tag{1.9}$$

As we will learn later this problem can be classified a regular perturbation problem as long as t is bounded, $t \in [0, T)$ with T independent of ε (more precisely, $T = O(1)$ as $\varepsilon \to 0$). Indeed, trying the AE

$$x_\varepsilon = x_0 + \varepsilon\, x_1 + O(\varepsilon^2)$$

leads to

$$P_0^{(\mathrm{osc})} \begin{cases} \dfrac{\mathrm{d}^2 x_0}{\mathrm{d}t^2} + x_0 = 0\,, \\ \\ x_0(0) = 0\,, \qquad \dfrac{\mathrm{d}x_0}{\mathrm{d}t}(0) = 1\,, \end{cases} \qquad P_1^{(\mathrm{osc})} \begin{cases} \dfrac{\mathrm{d}^2 x_1}{\mathrm{d}t^2} + x_1 = -\dfrac{\mathrm{d}x_0}{\mathrm{d}t}\,, \\ \\ x_1(0) = 0\,, \qquad \dfrac{\mathrm{d}x_1}{\mathrm{d}t}(0) = 0\,. \end{cases}$$

The reduced problem yields $x_0(t) = \sin(t)$ and then the problem $P_1^{(\mathrm{osc})}$ yields the correction $x_1(t) = -t\sin(t)/2$. Therefore, the AE reads

$$x_\varepsilon = \sin(t) - \frac{\varepsilon t}{2}\sin(t) + O(\varepsilon^2)\,.$$

This is a valid expansion as long as εt is small compared to one, hence for finite time t. Terms that blow up as $t \to \infty$ are called *secular terms* in perturbation theory. There are more sophisticated methods for (nearly-) periodic problems which can avoid secular terms in AEs (*averaging, method of multiple scales*).

2. Strongly damped oscillator: $\tau_1 \ll \tau_2$ and $T = \tau_2$. We define

$$\varepsilon := \left(\frac{\tau_1}{\tau_2}\right)^2\,,$$

which leads to

$$P_\varepsilon^{(\mathrm{dmp})} \begin{cases} \varepsilon \dfrac{\mathrm{d}^2 x_\varepsilon}{\mathrm{d}t^2} + \dfrac{\mathrm{d}x_\varepsilon}{\mathrm{d}t} + x_\varepsilon = 0\,, \\ \\ x_\varepsilon(0) = 0\,, \qquad \dfrac{\mathrm{d}x_\varepsilon}{\mathrm{d}t}(0) = 1\,. \end{cases} \tag{1.10}$$

This is a singular perturbation problem because the highest order derivative is multiplied by ε. The reduced problem reads

$$P_0^{(\mathrm{dmp})} \begin{cases} \dfrac{\mathrm{d}x_0}{\mathrm{d}t} + x_0 = 0\,, \\ \\ x_0(0) = 0\,, \qquad \dfrac{\mathrm{d}x_0}{\mathrm{d}t}(0) = 1\,. \end{cases}$$

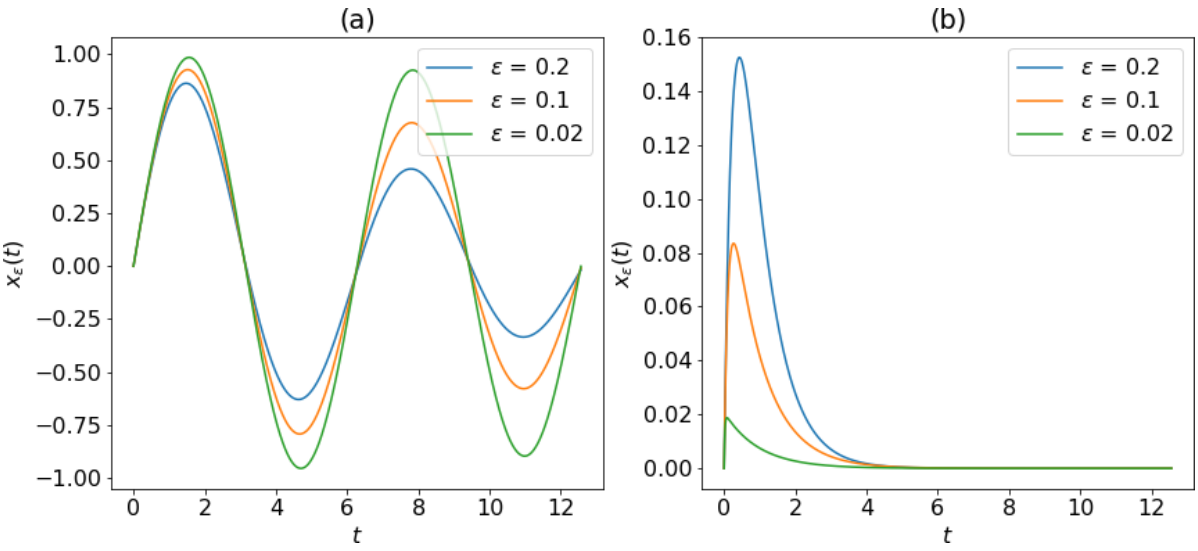

Figure 1.1.: Exact solutions for the weakly damped oscillator (1.9) in (a) and for the strongly damped oscillator (1.10) in (b), for different values of ε.

This problem is obviously ill-posed because x_0 cannot satisfy both initial conditions. Such a situation, where the order of the differential equation is decreased in the reduced problem, leads to the formation of *boundary layers*. The terminology becomes clear by looking at the exact solutions of $P_\varepsilon^{(\mathrm{dmp})}$ for different values of ε in panel (b) of Figure 1.1. Boundary layer problems can be treated with *asymptotic matching*, cf. Chapter 4.

1.3. General problem setting

In this course we are faced with finding approximate solutions to mathematical problems as a small parameter $\varepsilon > 0$ tends to zero. We write these problems symbolically as

$$P_\varepsilon[u_\varepsilon] = 0\,. \tag{1.11}$$

Here, u_ε is the solution of the problem and P_ε represents a set of model equations. In this course we focus on ordinary and partial differential equations (ODEs and PDEs). Hence, $u_\varepsilon(x)$ is defined for $x \in D \subset \mathbb{R}^n$ and we write $u_\varepsilon(x) = u(x_1, \ldots, x_n, \varepsilon)$; P_ε is some differential operator with suitable initial/boundary conditions. Often times, the problem (1.11) is too hard to solve, even with the help of numerical methods. The goal of asymptotic analysis is to find "simple" approximations for u_ε when ε is small. The term simple is subjective here; for instance, it could mean approximating the solution in terms of elementary functions or finding an approximation that can be computed on a lower dimensional subset $\overline{D} \subset D$. Setting formally $\varepsilon = 0$ leads to the *reduced problem*

$$P_0[u_0] = 0\,.$$

An immediate question arises:

Is the limit of the solution equal to the solution of the limit?

As already hinted before, there are two cases:

$$\text{regular problem:} \quad \forall\, x \in D \quad \lim_{\varepsilon\to 0} u_\varepsilon(x) = u_0(x)\,,$$

$$\text{singular problem:} \quad \exists\, x_0 \in D \;\; s.t. \;\; \lim_{\varepsilon\to 0} u_\varepsilon(x_0) \neq u_0(x_0)\,.$$

From this it is clear that a naive expansion of the form

$$u_\varepsilon = u_0 + \varepsilon u_1 + \varepsilon^2 u_2 + \dots \tag{1.12}$$

cannot be uniformly valid in D when a problem is singular. In the latter case, a viable strategy is to reformulate the equations into a regular perturbation problem and then apply an expansion of the form (1.12) to approximate the solution. This is commonly the strategy for developing asymptotic-preserving (AP) numerical schemes for stiff equations [8, 5]. If such a reformulation is not available, one distinguishes two types of problems: singular problems of *cumulative type* and singular problems of *boundary layer type*.

Singular problems of cumulative type. Theses are problems with oscillating solutions where the influence of the small parameter ε on the limit solution becomes observable only after long times of the order $t = O(\frac{1}{\varepsilon})$. The error terms are called *secular terms* and blow up as time goes to infinity, but are small for times of order one. The domain D is infinite, for instance $D = \mathbb{R}$ for ODEs. The secular terms lead to the following behavior:

$$\lim_{\varepsilon\to 0} u_\varepsilon(t) \neq u_0(t) \quad \text{for } t = O\Big(\frac{1}{\varepsilon}\Big)\,.$$

The techniques we will discuss in this course for dealing with problems of cumulative type are

- classical averaging,
- variational averaging for Lagrangian dynamical systems,
- the method of multiple scales.

Most of these techniques have been developed for studying the motion of celestial bodies and date back to the times of Poincaré ($\sim$ 1900). The technique of variational averaging [1, 10] has gained renewed attention for studying the helical motion of a charged particle in a strong magnetic field, a classical example for a problem of cumulative type.

Singular problems of boundary layer type. Many interesting phenomena in physics are characterized by a sudden change of state variables, for instance the formation of shock waves in gas dynamics or the boundary layer flow along the surface of a body. Mathematically, such problems can be described as singular perturbation problems where the domain D is *finite*. As ε tends to zero, the solution develops a jump in a very narrow

region of D, called the boundary layer. The boundary layer can be located at the edges of the domain but it does not have to be (free boundary layer problem). The main tool for treating such problems is called

- asymptotic matching.

2. Asymptotic expansions

Let us now introduce the main tools for asymptotic analysis. We shall give a precise definition of the terms appearing in expansions of the form (1.12) and generalized versions thereof. Moreover, we must clarify what is meant by an *asymptotic expansion* (AE) of u_ε. An interesting concept will be that of an asymptotic series, which in general does not converge, but nevertheless provides a good approximation to u_ε for small ε. The different notions of convergence and *asymptotic convergence* will be compared in detail.

2.1. Order functions

Definition 1. Let E be the set of real functions $\delta(\varepsilon)$ that are strictly positive and continuous on the interval $(0, \varepsilon_0]$ and such that

1. $\lim_{\varepsilon\to 0} \delta(\varepsilon)$ exists (it can be ∞),
2. $\delta_1, \delta_2 \in E \Rightarrow \delta_1\delta_2 \in E$.

A function $\delta(\varepsilon) \in E$ is called an *order function*.

The following functions are examples of order functions:

$$1,\ \varepsilon,\ 1+\varepsilon,\ \varepsilon^3,\ \frac{1}{\varepsilon},\ \frac{\varepsilon}{1+\varepsilon},\ \frac{1}{\ln(1/\varepsilon)},\ \mathrm{e}^{-1/\varepsilon}\,.$$

Note that if $\delta(\varepsilon)$ is an order function, then $1/\delta(\varepsilon)$ is too. The first condition above accepts $1/\varepsilon$, but it excludes functions with rapid variations near zero such as $1+\sin^2(1/\varepsilon)$. The second condition excludes products of such functions with viable order functions, like $\varepsilon[1+\sin^2(1/\varepsilon)]$. A comparison of order functions is accomplished via *Hardy's notation*:

$$\delta_1 \text{ is } \textit{asymptotically smaller} \text{ than } \delta_2,\ \delta_1 \prec \delta_2,\ \text{if } \lim_{\varepsilon\to 0}\frac{\delta_1}{\delta_2} = 0\,,$$

$$\delta_1 \text{ is } \textit{asymptotically equal} \text{ to } \delta_2,\ \delta_1 \sim \delta_2,\ \text{if } \lim_{\varepsilon\to 0}\frac{\delta_1}{\delta_2} = \lambda,\ 0<\lambda<\infty\,,$$

$$\delta_1 \text{ is asymptotically smaller than or equal to } \delta_2,\ \delta_1 \preceq \delta_2,\ \text{if } \lim_{\varepsilon\to 0}\frac{\delta_1}{\delta_2} = \lambda,\ 0\le\lambda<\infty\,.$$

Example 5. Using Hardy's notation we have, for $n \in \mathbb{N}_0$:

$$\varepsilon^{n+1} \prec \varepsilon^n, \qquad \mathrm{e}^{-1/\varepsilon} \prec \varepsilon^n \prec \frac{1}{\ln(1/\varepsilon)}\,,$$

$$2\varepsilon^n \sim \varepsilon^n, \qquad 2\varepsilon \sim \frac{\varepsilon}{1+\varepsilon}, \qquad 2 \sim 1+\varepsilon, \qquad \varepsilon \sim \sin(\varepsilon)\,.$$

Definition 2. A sequence of order functions δ_n is called *asymptotic sequence* if

$$\delta_{n+1} \prec \delta_n \quad \forall n\,.$$

If δ_n and γ_n are two asymptotic sequences with $\delta_n \sim \gamma_n$ for all n, than these sequences are *asymptotically equivalent*. For example,

$$\varepsilon^n, \qquad \left(\frac{\varepsilon}{1+\varepsilon}\right)^n, \qquad \sin^n(\varepsilon)\,,$$

are asymptotically equivalent. The relation $\delta_1 \sim \delta_2$ defines an equivalence relation on E, meaning it is reflexive, symmetric and transitive. Hence we can choose representatives of each class of equivalence. Usually we will work with the most convenient set of order functions like ε^n where n is an integer or ε^α where α is rational, and call these representatives *gauge functions*. The choice of gauge functions has implications on the uniqueness of asymptotic expansions discussed below.

2.2. Order of a function

Let us consider real-valued, continuous functions $u(\varepsilon)$ for $\varepsilon \in (0, \varepsilon_0]$. The magnitude of these functions as $\varepsilon \to 0$ can be compared to order functions $\delta \in E$ by means of Landau's "big-Oh" and "small-oh" notation:

Definition 3. (**Order of ε-functions.**)

(i) $u = O(\delta)$ as $\varepsilon \to 0$ if there exist positive constants k and C such that $|u(\varepsilon)| \le k\delta(\varepsilon)$ for $0 < \varepsilon < C$. (Remark that the limit $\lim_{\varepsilon\to 0} |u(\varepsilon)|/\delta(\varepsilon)$ need not exist).

(ii) $u = o(\delta)$ as $\varepsilon \to 0$ if $\lim_{\varepsilon\to 0} u(\varepsilon)/\delta(\varepsilon) = 0$.

(iii) $u = O_{\mathrm{s}}(\delta)$ if $u = O(\delta)$ and $u \neq o(\delta)$ at $x = x_0$.

If u is an order function then Hardy's and Landau's notation are equivalent. Landau's notation is however more general, because the limit in (i) need not exist, for instance we have $\sin(1/\varepsilon) = O(1)$ in Landau's notation.

In what follows we consider real-valued functions $u(x, \varepsilon) = u_\varepsilon(x)$ where $x \in D \subset \mathbb{R}^n$ and $\varepsilon \in (0, \varepsilon_0]$. For ε fixed, we suppose that $u_\varepsilon : D \to \mathbb{R}$ belongs to a normed linear space with the norm $||\cdot||$ and that $||u_\varepsilon||$ is continuous in ε. We want to compare such functions to order functions $\delta \in E$ as $\varepsilon \to 0$. This can be done either point-wise, or uniformly in some subset $D_0 \subset D$, or by using the norm $||\cdot||$.

Definition 4. (**Order, point-wise.**) Let $x_0 \in D$.

(i) $u = O(\delta)$ <u>at $x = x_0$</u> as $\varepsilon \to 0$ if there exist positive constants k and C such that $|u(x_0, \varepsilon)| \le k|\delta(\varepsilon)|$ for $0 < \varepsilon < C$.

(ii) $u = o(\delta)$ <u>at $x = x_0$</u> as $\varepsilon \to 0$ if $\lim_{\varepsilon\to 0} u(x_0, \varepsilon)/\delta(\varepsilon) = 0$.

(iii) $u = O_{\mathrm{s}}(\delta)$ <u>at $x = x_0$</u> if $u = O(\delta)$ and $u \neq o(\delta)$ at $x = x_0$.

Definition 5. (**Order, uniformly.**) Let $D_0 \subset D$.

(i) $u = O(\delta)$ uniformly in D_0 as $\varepsilon \to 0$ if there exist positive constants k and C, independent of x, such that for all $x \in D_0$, $|u(x,\varepsilon)| \leq k|\delta(\varepsilon)|$ for $0 < \varepsilon < C$.

(ii) $u = o(\delta)$ uniformly in D_0 as $\varepsilon \to 0$ if $\lim_{\varepsilon\to 0} u(x,\varepsilon)/\delta(\varepsilon) = 0$ for $x \in D_0$.

(iii) $u = O_{\mathrm{s}}(\delta)$ uniformly in D_0 if $u = O(\delta)$ and $u \neq o(\delta)$ uniformly in D_0.

Definition 6. (**Order, in norm.**) Let $D_0 \subset D$ and suppose that the restriction of u_ε to D_0 belongs to a normed linear space with norm $||u_\varepsilon||_{D_0}$.

(i) $u = O(\delta)$ in D_0 as $\varepsilon \to 0$ if $||u_\varepsilon||_{D_0} = O(\delta)$.

(ii) $u = o(\delta)$ in D_0 as $\varepsilon \to 0$ if $||u_\varepsilon||_{D_0} = o(\delta)$.

(iii) $u = O_{\mathrm{s}}(\delta)$ in D_0 as $\varepsilon \to 0$ if $u = O(\delta)$ and $u \neq o(\delta)$ in D_0.

In the last definition the order of a function depends on the chosen norm $||\cdot||_{D_0}$, which should be naturally related to the norm $||\cdot||$ for functions on D. The supremum norm is used most frequently in asymptotic analysis, that is, for u_ε continuous and bounded in D and $D_0 \subset D$,

$$||u_\varepsilon||_{D_0} = \max_{x\in D_0} |u_\varepsilon(x)| .$$

However, other norms such as L^2 can be used, depending on the type of problem one is interested in. The order of a function can then be completely different, depending on the norm chosen. Consider for example $u(x,\varepsilon) = u_\varepsilon(x) = \mathrm{e}^{-x/\varepsilon}$ on $D_0 = [0,1]$. In the supremum norm we have $u = O(1)$, whereas in the L^2-norm,

$$||u_\varepsilon||_{D_0} = \left(\int_0^1 u_\varepsilon^2(x)\,\mathrm{d}x \right)^{1/2} ,$$

we have $u = O(\sqrt{\varepsilon})$. In what follows **we shall always use the supremum norm if not stated otherwise**.

Remark 1. It is clear that Definition 5 is a special case of Defintion 6 with $||\cdot||_{D_0}$ the supremum norm.

In the analysis of singular perturbation problems the order of a function is usually not uniform in the whole domain D. This is why the notion of order in a subset $D_0 \subset D$ is important and has been stressed in the above defintions. This will become particularly apparent during the analysis of boundary layers. We close this section with an important Lemma:

Lemma 1. *Suppose $u(x,\varepsilon) = u_\varepsilon(x)$ such that $u_\varepsilon : D \to \mathbb{R}$ belongs to a normed linear space and that $||u_\varepsilon||$ is continuous in $\varepsilon \in (0,\varepsilon_0]$. Then there exists an order function $\delta \in E$ such that $u = O_{\mathrm{s}}(\delta)$ in D.*

Proof. See for example Eckhaus [7]. □

Under the conditions of Lemma 1 we can always rescale the function u such that

$$\frac{u}{\delta} = O_{\mathrm{s}}(1)\,. \tag{2.1}$$

2.3. Asymptotic expansions

We will now clarify the notions of

- asymptotic approximation (AA),
- (finite) asymptotic series,
- asymptotic expansion (AE).

These terms are often used synonymously in the literature, however some clarification is provided for example by Eckhaus [7].

Definition 7. Let $u(x,\varepsilon)$ be a function that is $O_{\mathrm{s}}(1)$ in D. A function $u_0(x,\varepsilon)$ is an *asymptotic approximation* (AA) of u if

$$u - u_0 = o(1) \text{ in } D\,.$$

Moreover, u_0 is called a *regular* AA if it is independent of ε, i.e. $u_0 = u_0(x)$.

Remark that if a regular AA of u exists it is unique, given by $u_0(x) = \lim_{\varepsilon\to 0} u(x,\varepsilon)$. AAs can be defined for functions u that are of arbitrary sharp order δ via the rescaling procedure (2.1). In general, u_0 is an asymptotic approximation of $u = O_{\mathrm{s}}(\delta)$ in D if

$$\frac{u - u_0}{\delta} = o(1) \text{ in } D\,.$$

This implies $u_0 = O_{\mathrm{s}}(\delta)$.

Example 6.

- $u_0(x,\varepsilon) = 1$ is a regular AA of $u(x,\varepsilon) = \mathrm{e}^{\varepsilon x}$ on any bounded interval $[-A, A]$ with A independent of ε, because

$$\lim_{\varepsilon\to 0}(u - u_0) = \lim_{\varepsilon\to 0}(\mathrm{e}^{\varepsilon x} - 1) = \lim_{\varepsilon\to 0}(1 + \varepsilon x + \frac{1}{2}\varepsilon^2 x^2 + O(\varepsilon^3) - 1) = 0 \text{ in } [-A, A]\,.$$

- $u_0(x,\varepsilon) = \varepsilon x + \varepsilon^2\cos(3x)$ is an AA of $u(x,\varepsilon) = \sin(\varepsilon x)$ on any bounded interval $[-A, A]$ with A independent of ε, because

$$\begin{aligned}\lim_{\varepsilon\to 0}\frac{u - u_0}{\varepsilon} &= \lim_{\varepsilon\to 0}\frac{\sin(\varepsilon x) - \varepsilon x - \varepsilon^2\cos(3x)}{\varepsilon}\\ &= \lim_{\varepsilon\to 0}\frac{\varepsilon x - \frac{1}{6}\varepsilon^3 x^3 + O(\varepsilon^5) - \varepsilon x - \varepsilon^2\cos(3x)}{\varepsilon}\\ &= \lim_{\varepsilon\to 0}(-\frac{1}{6}\varepsilon^2 x^3 - \varepsilon\cos(3x)) = 0 \text{ in } [-A, A]\,.\end{aligned}$$

An asymptotic expansion of u is constructed by successive AAs of the remainder terms. For example, let $u_0 = O_{\mathrm{s}}(1)$ be an AA of u in the sense of Definition 7. Then we define the remainder as

$$R_1 := u - u_0 = o(1)\,.$$

It follows from Lemma 1 that there exists a $\delta_1 \prec 1$ such that $R_1 = O_{\mathrm{s}}(\delta_1)$, and we can rescale via (2.1) to obtain $\overline{R}_1 := R_1/\delta_1 = O_{\mathrm{s}}(1)$. Therefore, we can write

$$u = u_0 + \delta_1 \overline{R}_1\,. \tag{2.2}$$

Suppose now we are able to find an AA of the remainder $\overline{R}_1(x,\varepsilon)$ and call it $u_1(x,\varepsilon)$. Then from Definition 7 we define the new remainder as

$$R_2 := \overline{R}_1 - u_1 = o(1)\,,$$

and there exists a δ_2 such that $R_2 = O_{\mathrm{s}}(\delta_2)$. Rescaling leads to $\overline{R}_2 := R_2/\delta_2 = O_{\mathrm{s}}(1)$ and by inserting this into (2.2) we can write

$$u = u_0 + \delta_1 u_1 + \delta_1\delta_2 \overline{R}_2\,.$$

We can repeat this procedure N times to obtain

$$u(x,\varepsilon) = \sum_{n=0}^{N} \tilde{\delta}_n(\varepsilon)\, u_n(x,\varepsilon) + o(\tilde{\delta}_N(\varepsilon))\,, \tag{2.3}$$

where $\tilde{\delta}_0 = 1$, $\tilde{\delta}_{n\geq 1}(\varepsilon) = \delta_1(\varepsilon)\cdot\ldots\cdot\delta_n(\varepsilon)$ and $u_n(x,\varepsilon) = O_{\mathrm{s}}(1)$. Remark that the product of order functions is again an order function due to the property 2 of Definition 1. The right-hand side in (2.3) is called an asymptotic expansion (AE) of the function u in D. It follows in particular that

$$\lim_{\varepsilon\to 0} \frac{||u(x,\varepsilon) - \sum_{n=0}^{N} \tilde{\delta}_n(\varepsilon)\, u_n(x,\varepsilon)||}{\tilde{\delta}_N(\varepsilon)} = 0\,.$$

Let us now formalize this result:

Definition 8. Let $(u_n^*(x,\varepsilon))_{n=0}^{\infty}$ denote a sequence of functions on $D\times(0,\varepsilon_0]$ with $u_n^* = O_{\mathrm{s}}(\delta_n)$. A (finite) series $u^{(N)}$ given by

$$u^{(N)}(x,\varepsilon) = \sum_{n=0}^{N} u_n^*(x,\varepsilon)\,, \qquad u_n^* = O_{\mathrm{s}}(\delta_n) \text{ in } D\,,$$

is called an *asymptotic series* in D if δ_n is an asymptotic sequence, hence if $\delta_{n+1} \prec \delta_n$ (see Definition 2).

By rescaling as in (2.1) one obtains $u_n^* = \delta_n u_n$ with $u_n = O_{\mathrm{s}}(1)$, hence any asymptotic series can be written as

$$u^{(N)}(x,\varepsilon) = \sum_{n=0}^{N} \delta_n(\varepsilon)\, u_n(x,\varepsilon)\,.$$

Example 7. The expressions

$$f^{(2)}(x,\varepsilon) = \varepsilon x - \frac{\varepsilon^2 x^2}{2} + \frac{\varepsilon^3 x^3}{3}\,,$$
$$g^{(2)}(x,\varepsilon) = 1 + \frac{\varepsilon^2 x^2}{2} + \frac{\varepsilon^4}{(1+\varepsilon)^4}\frac{x^4}{24}\,,$$
$$h^{(2)}(x,\varepsilon) = 1 + \varepsilon x + \sin(\varepsilon^2 x^2)\,,$$

are finite asymptotic series in any bounded interval $[-A, A]$ with A independent of ε.

Definition 9. Let $u^{(N)}$ denote a (finite) asymptotic series in D given by

$$u^{(N)}(x,\varepsilon) = \sum_{n=0}^{N} \delta_n(\varepsilon)\, u_n(x,\varepsilon)\,, \qquad u_n = O_{\mathrm{s}}(1) \text{ in } D\,,$$

then $u^{(N)}$ is an *asymptotic expansion* (AE) *to order* δ_N of u in D if

$$u - u^{(N)} = o(\delta_N) \text{ in } D\,.$$

Moreover, $u^{(N)}$ is called a *regular* AE if it is of the form

$$u^{(N)}(x,\varepsilon) = \sum_{n=0}^{N} \delta_n(\varepsilon)\, u_n(x)\,,$$

hence if the coefficient functions u_n do not depend on ε.

Example 8. For the three asymptotic series given in Example 7 we have

$$\ln(1+\varepsilon x) - f^{(2)}(x,\varepsilon) = o(\varepsilon^3) \text{ in } [-A, A]\,,$$
$$\cosh(\varepsilon x) - g^{(2)}(x,\varepsilon) = o(\varepsilon^4) \text{ in } [-A, A]\,,$$
$$\frac{1}{1-\varepsilon x} - h^{(2)}(x,\varepsilon) = o(\varepsilon^2) \text{ in } [-A, A]\,.$$

Hence, $f^{(2)}$ is an AE to order ε^3 of $\ln(1+\varepsilon x)$ in $[-A, A]$, $g^{(2)}$ is an AE to order ε^4 of $\cosh(\varepsilon x)$ in $[-A, A]$ and $h^{(2)}$ is an AE to order ε^2 of $1/(1+\varepsilon x)$ in $[-A, A]$. This follows immediately from the series expansions

$$\ln(1+x') = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{(x')^n}{n} \qquad |x'| < 1\,,$$
$$\cosh(x') = \sum_{n=0}^{\infty} \frac{(x')^{2n}}{(2n)!} \qquad x' \in \mathbb{R}\,,$$
$$\frac{1}{1-x'} = \sum_{n=0}^{\infty} (x')^n \qquad |x'| < 1\,.$$

Remark that the series expansions have a certain radius of convergence, whereas the AEs are valid in the whole interval $[-A, A]$. This is a first hint towards the difference between convergence and *asymptotic convergence*.

If the asymptotic expansion in Definition 9 holds for any positive integer N one writes

$$u \sim \sum_{n=0}^{\infty} \delta_n \, u_n \,,$$

and says the series is *asymptotically convergent* to u in D. **Be mindful however that asymptotic convergence does not imply convergence of the series in the usual sense!** Indeed, most asymptotic series do not converge, but this is also not their purpose, which is is to provide good approximations to a function as $\varepsilon \to 0$. Let us compare the two notions of convergence with respect to ε in detail. Consider $x_0 \in D$ to be fixed:

- Convergence of a series: the limit

$$\lim_{N\to\infty} \sum_{n=0}^{N} \delta_n(\varepsilon) \, u_n(x_0, \varepsilon)$$

  exists for $\varepsilon \in (0, R)$, where R is the radius of convergence.

- Asymptotic convergence to an ε-function $u(x_0, \varepsilon)$:

$$\lim_{\varepsilon\to 0} \frac{\left| u(x_0,\varepsilon) - \sum_{n=0}^{N} \delta_n(\varepsilon) \, u_n(x_0,\varepsilon) \right|}{\delta_N(\varepsilon)} = 0 \qquad \forall \; N \in \mathbb{N} \,.$$

We see that convergence is about the behavior of the series in a finite ε-region $(0, R)$ as $N \to \infty$, whereas asymptotic convergence is about the approximation of the ε-function $u(x_0, \varepsilon)$ as ε tends to zero. While the former is an *absolute* concept, the latter is a *relative concept*, always with respect to a given function u. It thus makes no sense to ask a question like *'Is this series asymptotically convergent?'* A reasonable question would be *'Is this series asymptotically convergent to u?'*

As we have already seen in Example 8, the most common examples of AEs are Taylor expansions. Suppose u is N-times differentiable at $\varepsilon = 0$, then Taylor's theorem states

$$u = \sum_{n=0}^{N} \varepsilon^n \frac{1}{n!} \frac{\mathrm{d}^n u}{\mathrm{d}\varepsilon^n}\Big|_{\varepsilon=0} + o(\varepsilon^N) \,. \tag{2.4}$$

In case that N becomes infinite the series can be convergent within a certain radius R around zero and divergent elsewhere (R can be zero). However, the finite Taylor expansion (2.4) is always a good approximation of the function u if one is sufficiently close to $\varepsilon = 0$.

Example 9. Given a function $f : \mathbb{R} \to \mathbb{R}$ which is $f = O(1)$ in $\mathbb{R}$, we have

$$\mathrm{e}^{\varepsilon f(x)} = 1 + \varepsilon f(x) + \frac{\varepsilon^2 f^2(x)}{2} + \ldots + \frac{\varepsilon^n f^n(x)}{n!} + \ldots \,.$$

This series converges for all values of ε. The AE

$$\mathrm{e}^{\varepsilon f(x)} = 1 + \varepsilon f(x) + \frac{\varepsilon^2 f^2(x)}{2} + o(\varepsilon^2) \text{ in } \mathbb{R}\,,$$

is valid near $\varepsilon = 0$ but not elsewhere.

Example 10. The order of a function $u : D \times (0, \varepsilon_0] \to \mathbb{R}$ might not be uniform in the whole domain D. Consider for instance the function

$$u(x, \varepsilon) = \sqrt{x + \varepsilon} = \sqrt{x} \cdot \sqrt{1 + \frac{\varepsilon}{x}}\,, \quad x > 0\,.$$

Taylor expansion leads to

$$u(x, \varepsilon) \sim \sqrt{x} \left(1 + \frac{\varepsilon}{2x} - \frac{\varepsilon^2}{8x^2} + \ldots + \frac{(-1)^{n-1}(2n-3)!!}{2^n n!} \frac{\varepsilon^n}{x^n} + \ldots\right)\,.$$

This is an AE of u in any left bounded interval $x \in [A, \infty)$ with $A > 0$; however it is not an AE for $x \in (0, \infty)$ because the remainder terms are not $o(\varepsilon^n)$ for $x = O(\varepsilon)$.

Example 11. Consider the error function defined by

$$\mathrm{erf}\,(t) = 1 - \frac{2}{\sqrt{\pi}} \int_t^\infty \mathrm{e}^{-s^2} \mathrm{d}s\,.$$

It can be shown ([4], page 16) that this function can be approximated by

$$\mathrm{erf}\,(t) = 1 - \frac{\mathrm{e}^{-t^2}}{\sqrt{\pi}} \left\{ \sum_{n=1}^{N} (-1)^{n-1} \frac{(2n-3)!!}{2^{n-1}} \frac{1}{t^{2n-1}} + o\Big(\frac{1}{t^{2N-1}}\Big) \right\}\,,$$

which is true for any value of $N \in \mathbb{N}$. Hence, substituting $t = 1/\varepsilon$, the following series is asymptotically convergent to the error function,

$$\mathrm{erf}\left(\frac{1}{\varepsilon}\right) \sim 1 - \frac{\mathrm{e}^{-1/\varepsilon^2}}{\sqrt{\pi}} \sum_{n=1}^{\infty} (-1)^{n-1} \frac{(2n-3)!!}{2^{n-1}}\, \varepsilon^{2n-1}\,.$$

This is an AE of erf with respect to the sequence $(\varepsilon^{2n+1}\mathrm{e}^{-1/\varepsilon^2})_{n=0}^{\infty}$ and it diverges for any value of ε. It is nevertheless a useful AE for obtaining values of the error function for large arguments by keeping only the first few terms.

Example 12. The literature on special functions is full of useful AEs of these functions. Consider for instance the Bessel function $J_0(t)$, which has the series expansion

$$J_0(t) = \sum_{n=0}^{\infty} (-1)^n \Big(\frac{t^n}{2^n n!}\Big)^2\,.$$

This AE is convergent on any bounded interval of $\mathbb{R}$. For large t, writing $t = 1/\varepsilon$, we have the well-known AE ([4], page 17)

$$J_0\Big(\frac{1}{\varepsilon}\Big) \sim \sqrt{\frac{2\varepsilon}{\pi}} \left\{ \cos\Big(\frac{1}{\varepsilon} - \frac{\pi}{4}\Big) \sum_{n=0}^{\infty} (-1)^n \frac{(4n-1)!!^2}{2^{6n}(2n)!}\, \varepsilon^{2n} \right. \\ \left. \sin\Big(\frac{1}{\varepsilon} - \frac{\pi}{4}\Big) \sum_{n=0}^{\infty} (-1)^n \frac{(4n+1)!!^2}{2^{6n+3}(2n+1)!}\, \varepsilon^{2n+1} \right\} .$$

While the convergent expansion is rather useless for for obtaining values of J_0 for large arguments, the AE is very useful. To approximate the value of $J_0(3)$ to three digits precision one needs eight terms of the convergent series but only one term of the AE.

It is clear from the computations leading to (2.3) that an asymptotic expansion (AE) can always be studied as a repeated process of asymptotic approximations (AAs). If these AAs are all regular and if we fix the δ_n to a certain set of gauge functions, for example $\delta_n = \varepsilon^n$, it follows that **the regular AE of a function u, if it exists, is unique with respect to the chosen gauge functions, because the coefficients u_n are uniquely determined by**

$$u_n(x) = \lim_{\varepsilon \to 0} \frac{u(x,\varepsilon) - \sum_{m=0}^{n-1} \delta_m(\varepsilon)\, u_m(x)}{\delta_n(\varepsilon)} .$$

The converse is however not true: a given asymptotic series $u^{(N)}$ is indeed the AE of an infinity of functions which differ by a term of $o(\delta_N)$. These statements regarding uniqueness even hold for asymptotically convergent series. Suppose for example that we choose $\delta_n = \varepsilon^n$ as gauge functions, then

$$u \sim \sum_{n=0}^{\infty} \varepsilon^n\, u_n \qquad \Leftrightarrow \qquad u + \mathrm{e}^{-1/\varepsilon} \sim \sum_{n=0}^{\infty} \varepsilon^n\, u_n \, .$$

Two functions which have the same AE with respect to a given asymptotic sequence of gauge functions are called *asymptotically equal*. Hence u is a.e. to $u + \mathrm{e}^{-1/\varepsilon}$ with respect to the sequence ε^n. It may thus very well be that

$$u(x,\varepsilon) \neq \sum_{n=0}^{\infty} \delta_n(\varepsilon)\, u_n(x,\varepsilon) \, , \qquad x \in D,\ \varepsilon \in (0,R) \, ,$$

in case that the series converges. Finally, it is clear that the AE of a function changes when the asymptotic sequence δ_n changes.

In the special case that $\delta_n = \varepsilon^n$ and the AE is regular, one calls

$$u(x,\varepsilon) = \sum_{n=0}^{N} \varepsilon^n u_n(x) + o(\varepsilon^N)$$

the *Poincaré expansion* (sometimes also *Hilbert expansion*) of u. Poincaré expansions are important tools in asymptotic analysis because of their relative simplicity; a lot can

be gained by approximating a complicated function of x and ε by a sequence of functions with the simple structure $\varepsilon^n u_n(x)$.

Another interesting question is the following: given a divergent AE that is asymptotically convergent to u, what is the optimal truncation of the series for a fixed value of ε? Optimal in this sense means with minimal error in the supremum norm. Hence we search for the optimal number N_{opt} after which to truncate the series to get the smallest error. Such questions arise frequently in physics problems, where ε is usually defined by the fraction of some fixed problem parameters.

Let us now consider some elementary operations on AEs like addition, multiplication, integration and differentiation. Let us assume expansions of Poincaré type,

$$u(x,\varepsilon) \sim \sum_{n=0}^{\infty} \delta_n(\varepsilon) u_n(x)\,, \qquad v_\varepsilon(x) \sim \sum_{n=0}^{\infty} \delta_n(\varepsilon) v_n(x) \quad \text{in } D\,.$$

Then the following is true:

- Addition:
$$u(x,\varepsilon) + v(x,\varepsilon) \sim \sum_{n=0}^{\infty} \delta_n(\varepsilon)[u_n(x) + v_n(x)]\,,$$

- Multiplication: if $\delta_n \delta_m = \delta_{n+m}$ then
$$u(x,\varepsilon)v(x,\varepsilon) \sim \sum_{n=0}^{\infty} \delta_n(\varepsilon) w_n(x)\,, \qquad w_n(x) = \sum_{m=0}^{n} u_m(x) v_{n-m}(x)\,.$$

- Integration along a path γ in D: assuming everything is integrable along γ,
$$\int_\gamma u(x,\varepsilon)\mathrm{d}\sigma \sim \sum_{n=0}^{\infty} \delta_n(\varepsilon) \int_\gamma u_n(x)\mathrm{d}\sigma\,,$$
where $d\sigma$ is the line element along $\gamma : I \subset \mathbb{R} \to D$.

- Integration with respect to ε: assuming everything is integrable w.r.t ε,
$$\int_0^\varepsilon u_{\varepsilon'}(x)\mathrm{d}\varepsilon' \sim \sum_{n=0}^{\infty} u_n(x) \int_0^\varepsilon \delta_n(\varepsilon')\mathrm{d}\varepsilon'\,.$$

- Differentiation with respect to x: if $u(x,\varepsilon)$ and $u_n(x)$ are differentiable in D,
$$\frac{\partial u(x,\varepsilon)}{\partial x_i} \sim \sum_{n=0}^{\infty} \delta_n(\varepsilon) \frac{\partial u_n(x)}{\partial x_i}\,.$$

- Differentiation with respect to ε: if $\frac{\partial u(x,\varepsilon)}{\partial \varepsilon}$ for $x \in D$ and $\frac{\mathrm{d}\delta_n(\varepsilon)}{\mathrm{d}\varepsilon}$ exist for $\varepsilon \in (0, \varepsilon_0]$, and if

$$\frac{\partial u(x,\varepsilon)}{\partial \varepsilon} \sim \sum_{n=0}^{\infty} \frac{\mathrm{d}\delta_n(\varepsilon)}{\mathrm{d}\varepsilon} u_n'(x)\,,$$

  then $u_n'(x) = u_n(x)$ for $x \in D$.

The proof of these properties is straightforward.

3. Regular perturbations

3.1. Generalities

We consider initial/boundary-value problems of the form

$$P_\varepsilon \begin{cases} L_\varepsilon[u_\varepsilon(x)] = 0\,, & x = (x_1, \ldots, x_n) \in D \subset \mathbb{R}^n\,, \\ B_\varepsilon[u_\varepsilon(x)] = 0\,, & x \in S \subset \partial D\,, \end{cases} \tag{3.1}$$

where $u_\varepsilon : D \to \mathbb{R}^m$ stands for the (vector-valued) solution of the problem, L_ε represents some differential operator and B_ε represents initial- and/or boundary conditions. The domain of the independent variables is D and the limiting conditions are prescribed on a subset S of the domain boundary ∂D. In the case of ODEs ($n = 1$), we distinguish initial-value problems (IVPs), where the boundary conditions are prescribed at one point $x_0 \in S$, and boundary-value problems (BVPs), where the conditions are prescribed at more than one point. Both L_ε and B_ε contain a small parameter ε and thus the problem P_ε is called a *perturbation problem*. The *reduced problem* is obtained by setting formally $\varepsilon = 0$ in P_ε, leading to

$$P_0 \begin{cases} L_0[u_0(x)] = 0\,, & x \in D \subset \mathbb{R}^n\,, \\ B_0[u_0(x)] = 0\,, & x \in S \subset \partial D\,. \end{cases} \tag{3.2}$$

The question we shall address is whether the solution u_0 of (3.2) is a good approximation for the solution u_ε of (3.1) in case that ε is small, and whether it is possible to improve on the approximation and to estimate the error. A first attempt to solve this problem is to assume a solution of the form

$$u_\varepsilon(x) = \sum_{n=0}^{N} \varepsilon^n u_n(x) + R_N(x, \varepsilon)\,, \tag{3.3}$$

and to try to solve for the coefficient functions u_0,..., u_N by substituting the ansatz (3.3) into P_ε and sorting in powers of ε. If this succeeds we can try to estimate the error term R_N. In case we can prove that $R_N = o(\varepsilon^N)$ in D, the solution (3.3) is a regular AE of the true solution and the perturbation problem P_ε is called *regular*.

3.2. Nonlinear initial-value problems

Many perturbation problems related to research or engineering problems can be characterised as IVPs with a perturbed vector field (right-hand side). We mention for example

dynamical systems from physics, described by Newton's equation of motion with a perturbed force, or predator-prey models from biology. The abundance of such problems in science is the reason we start with study of the regularly-perturbed IVPs as an entry point into the theory of asymptotic methods. As prerequisites we assume a basic understanding of IVPs in order to quickly proceed with the asymptotic analysis. To refresh memory, some basic material on IVPs has been gathered in appendix A.

In what follows the domain is $D \subset \mathbb{R}$ and the independent variable is denoted by $t \in D$. We search (vector-valued) functions $u_\varepsilon : D \to \mathbb{R}^m$ that satisfy the IVP

$$P_\varepsilon \begin{cases} \dfrac{\mathrm{d}u_\varepsilon}{\mathrm{d}t} = f(u_\varepsilon, t, \varepsilon)\,, \\ u_\varepsilon(t_0) = V(\varepsilon) \in \mathbb{R}^m\,, \quad t_0 \in D\,. \end{cases} \tag{3.4}$$

Here, V is the initial state of the variable u_ε and $f : \mathbb{R}^m \times D \times (0, \varepsilon_0] \to \mathbb{R}^m$ is called the "vector field" of the IVP. The vector field f is assumed to be

- N times continuously differentiable in $\mathbb{R}^m \times D \times (0, \varepsilon_0]$, with $N \geq 1$.
- Lipschitz continuous with respect to the first argument in any bounded domain $\Omega \subset \mathbb{R}^m$,
$$||f(x, t, \varepsilon) - f(y, t, \varepsilon)|| \leq L_f\, ||x - y|| \qquad x, y \in \Omega\,. \tag{3.5}$$
Here, $||\cdot||$ is some vector norm in $\mathbb{R}^m$ and L_f denotes the Lipschitz constant which does not depend on (t, ε).

Moreover, we assume that f and V have the AEs

$$\begin{aligned} f(u, t, \varepsilon) &= \sum_{n=0}^{N} \varepsilon^n f_n(u, t) + o(\varepsilon^N) \quad \text{in } \mathbb{R}^m \times D\,, \\ V(\varepsilon) &= \sum_{n=0}^{N} \varepsilon^n V_n + o(\varepsilon^N)\,. \end{aligned} \tag{3.6}$$

It follows from Theorem 4 in Appendix A that under the above assumptions, the problem P_ε has a unique solution u_ε in some maximal interval $I_\varepsilon = I_\varepsilon(V, t_0) \subset D$. For simplicity, let us restrict this solution to an interval $t_0 \leq t \leq t_0 + T_\varepsilon$ where we consider only forward propagation in time.

Our first instinct is that P_ε is a regular perturbation problem, thus we try the regular ansatz

$$u_\varepsilon(t) = \sum_{n=0}^{N} \varepsilon^n u_n(t) + R_N(t, \varepsilon)\,, \qquad R_N = o(\varepsilon^N) \text{ in } D\,. \tag{3.7}$$

Since the f_n in (3.6) are C^N-functions, we may apply Taylor's theorem to compute

$$\begin{aligned} f_n(u_\varepsilon, t) &= f_n\left(u_0 + \sum_{n=1}^{N} \varepsilon^n u_n + R_N, t\right) \\ &= f_n(u_0, t) + \sum_{k=1}^{N} \frac{1}{k!} \frac{\partial^{(k)} f_n}{\partial x^k}(u_0, t) \cdot \left(\sum_{n=1}^{N} \varepsilon^n u_n\right)^k + o(\varepsilon^N)\,. \end{aligned} \tag{3.8}$$

Here, $\partial f_n/\partial x \in \mathbb{R}^{m\times m}$ denotes the Jacobian matrix with respect to the first argument of f_n and the $\partial^{(k)} f_n/\partial x^k$ for $k > 1$ are multi-linear tensors of order $k+1$ in the same sense. The dot $\cdot$ indicates the inner product of theses tensors with k vectors, yielding again a vector in $\mathbb{R}^m$. The error term R_N has been moved into the $o(\varepsilon^N)$-terms in the second line of (3.8). Substitution of (3.6)-(3.8) into the IVP (3.4) leads to

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t}\Big(u_0 + \varepsilon u_1 + \varepsilon^2 u_2 + \dots\Big) = \\ &\quad = f_0(u_0,t) + \frac{\partial f_0}{\partial x}(u_0,t) \cdot \sum_{n=1}^{N} \varepsilon^n u_n + \frac{1}{2}\frac{\partial^2 f_0}{\partial u^2}(u_0,t) \cdot \left(\sum_{n=1}^{N} \varepsilon^n u_n\right)^2 + \dots \\ &\quad + \varepsilon\left[f_1(u_0,t) + \frac{\partial f_1}{\partial x}(u_0,t) \cdot \sum_{n=1}^{N} \varepsilon^n u_n + \frac{1}{2}\frac{\partial^2 f_1}{\partial u^2}(u_0,t) \cdot \left(\sum_{n=1}^{N} \varepsilon^n u_n\right)^2 + \dots\right] \\ &\quad + \varepsilon^2\left[f_2(u_0,t) + \frac{\partial f_2}{\partial x}(u_0,t) \cdot \sum_{n=1}^{N} \varepsilon^n u_n + \frac{1}{2}\frac{\partial^2 f_2}{\partial u^2}(u_0,t) \cdot \left(\sum_{n=1}^{N} \varepsilon^n u_n\right)^2 + \dots\right] \\ &\quad + \dots\,, \end{aligned} \tag{3.9}$$

and

$$(u_0 + \varepsilon u_1 + \varepsilon^2 u_2 + \dots)(t_0) = V_0 + \varepsilon V_1 + \varepsilon^2 V_2 + \dots$$

for the initial conditions. Sorting (3.9) in powers of ε and equating to zero yields the following equations:

$$\begin{aligned} \frac{\mathrm{d}u_0}{\mathrm{d}t} &= f_0(u_0,t)\,, \\ \frac{\mathrm{d}u_1}{\mathrm{d}t} &= \frac{\partial f_0}{\partial x}(u_0,t)\cdot u_1 + \underbrace{f_1(u_0,t)}_{=:\widetilde{f}_1(u_0,t)}\,, \\ \frac{\mathrm{d}u_2}{\mathrm{d}t} &= \frac{\partial f_0}{\partial x}(u_0,t)\cdot u_2 + \underbrace{\frac{1}{2}\frac{\partial^2 f_0}{\partial u^2}(u_0,t)\cdot u_1^2 + \frac{\partial f_1}{\partial x}(u_0,t)\cdot u_1 + f_2(u_0,t)}_{=:\widetilde{f}_2(u_0,u_1,t)}\,. \end{aligned}$$

We see the evolving pattern and find the equation for the n-th order to be

$$\frac{\mathrm{d}u_n}{\mathrm{d}t} = \frac{\partial f_0}{\partial x}(u_0,t)\cdot u_n + \widetilde{f}_n(u_0,\dots,u_{n-1},t)\,. \tag{3.10}$$

Here, $\widetilde{f}_n$ is a function of depending on lower order solutions u_k with $k \leq n-1$ only, and can thus be viewed as a source term in the equation for u_n. This leads to the following set of IVPs:

$$P_0 \begin{cases} \dfrac{\mathrm{d}u_0}{\mathrm{d}t} = f_0(u_0, t)\,, \\ u_0(t_0) = V_0\,, \end{cases} \tag{3.11}$$

$$P_1 \begin{cases} \dfrac{\mathrm{d}u_1}{\mathrm{d}t} = \dfrac{\partial f_0}{\partial x}(u_0, t) \cdot u_1 + \widetilde{f}_1(u_0, t)\,, \\ u_1(t_0) = V_1\,, \end{cases}$$

$$\vdots$$

$$P_n \begin{cases} \dfrac{\mathrm{d}u_n}{\mathrm{d}t} = \dfrac{\partial f_0}{\partial x}(u_0, t) \cdot u_n + \widetilde{f}_n(u_0, \ldots, u_{n-1}, t)\,, \\ u_n(t_0) = V_n\,, \end{cases} \tag{3.12}$$

$$\vdots$$

for $n \leq N$. (3.11) is the reduced problem; due to our assumptions on f, which carry over to f_0, the reduced problem has a unique solution in some interval $t_0 \leq t \leq t_0 + T$. The higher order problems P_n for $1 \leq n \leq N$ are all linear and are thus easily solved. **Therefore, by means of the regular ansatz (3.7), the perturbation problem (3.4) has been transformed into a system of linear equations for the u_n, $n \geq 1$, which can be solved sequentially once the solution u_0 of the reduced problem is known**. The solution of the linear problems P_n for $n \geq 1$ can be written explicitly with Duhamel's formula. First, assuming $\widetilde{f}_n = 0$, the solution to the homogeneous problem for any n reads

$$u_n(t) = \exp\left(\int_{t_0}^{t} \frac{\partial f_0}{\partial x}(u_0(s), s)\,\mathrm{d}s\right) \cdot u_n(t_0)\,.$$

Substituting the initial condition and defining the matrix

$$A(t, t_0) := \exp\left(\int_{t_0}^{t} \frac{\partial f_0}{\partial x}(u_0(s), s)\,\mathrm{d}s\right) \in \mathbb{R}^{m\times m}\,,$$

we write the solution to the homogeneous problem as

$$u_n(t) = A(t, t_0) \cdot V_n\,.$$

Duhamel's formula for the inhomogeneous problem P_n reads

$$u_n(t) = A(t, t_0) \cdot V_n + \int_{t_0}^{t} A(t, s) \cdot \widetilde{f}_n(u_0(s), \ldots, u_{n-1}(s), s)\,\mathrm{d}s\,. \tag{3.13}$$

Let us verify that this is indeed a solution of (3.12):

$$\begin{aligned}
\frac{\mathrm{d}u_n}{\mathrm{d}t} &= \frac{\mathrm{d}A(t,t_0)}{\mathrm{d}t}\cdot V_n + \frac{\mathrm{d}}{\mathrm{d}t}\int_{t_0}^{t} A(t,s)\cdot \widetilde{f}_n(u_0(s),\dots,u_{n-1}(s),s)\,\mathrm{d}s \\
&= \frac{\mathrm{d}A(t,t_0)}{\mathrm{d}t}\cdot V_n + \int_{t_0}^{t} \frac{\mathrm{d}A(t,s)}{\mathrm{d}t}\cdot \widetilde{f}_n(u_0(s),\dots,u_{n-1}(s),s)\,\mathrm{d}s \\
&\qquad + A(t,t)\cdot \widetilde{f}_n(u_0,\dots,u_{n-1},t) \\
&= \frac{\partial f_0}{\partial x}(u_0(t),t)\left[A(t,t_0)\cdot V_n + \int_{t_0}^{t} A(t,s)\cdot \widetilde{f}_n(u_0(s),\dots,u_{n-1}(s),s)\,\mathrm{d}s\right] \\
&\qquad + A(t,t)\cdot \widetilde{f}_n(u_0,\dots,u_{n-1},t) \\
&= \frac{\partial f_0}{\partial x}(u_0,t)\,u_n + \widetilde{f}_n(u_0,\dots,u_{n-1},t)\,.
\end{aligned}$$

We have thus computed the coefficient functions u_n of our ansatz (3.7) in formula (3.13).

In order to show that (3.7) is indeed an AE of the true solution u_ε we now have to prove that $R_N = o(\varepsilon^N)$. For this we aim to derive an IVP for the remainder R_N and then estimate how its size evolves using Gronwall's lemma, stated below. First, let us denote the solution we computed in terms of the u_n by

$$u^{(N)}(t,\varepsilon) = \sum_{n=0}^{N} \varepsilon^n u_n(t) \qquad t_0 \le t \le t_0 + T\,,$$

with u_0 the solution of the reduced problem and the u_n for $n \ge 1$ given in (3.13). By going backwards through the steps in (3.10)-(3.8), we find that

$$\frac{\mathrm{d}u^{(N)}}{\mathrm{d}t} = \sum_{n=0}^{N} \varepsilon^n \frac{\mathrm{d}u_n}{\mathrm{d}t} = f(u^{(N)},t,\varepsilon) + o(\varepsilon^N)\,,$$

where f is the vector field of the original IVP (3.4). Moreover, initially

$$u^{(N)}(t_0) = \sum_{n=0}^{N} \varepsilon^n V_n\,. \tag{3.14}$$

Therefore, from $R_N = u_\varepsilon - u^{(N)}$, we obtain the following IVP for the remainder:

$$\begin{cases} \dfrac{\mathrm{d}R_N}{\mathrm{d}t} = f(u^{(N)} + R_N, t, \varepsilon) - f(u^{(N)},t,\varepsilon)\,, \\ R_N(t_0) = o(\varepsilon^N)\,, \end{cases}$$

for $t_0 \le t \le t_0 + T$. Hence, $R_N(t) \in \Omega \subset \mathbb{R}^m$ bounded during this time. Integration with respect to time yields

$$R_N(t,\varepsilon) = \int_{t_0}^{t} \left[f(u^{(N)}(s,\varepsilon) + R_N(s,\varepsilon), s, \varepsilon) - f(u^{(N)}(s,\varepsilon),s,\varepsilon)\right]\mathrm{d}s + o(\varepsilon^N)\,.$$

Let us take the vector norm $||\cdot||$ in $\mathbb{R}^m$ of this result and use the fact that f is Lipschitz in Ω w.r.t the first argument, c.f. (3.5),

$$\begin{aligned} ||R_N(t,\varepsilon)|| &= || \int_{t_0}^{t} \Big[f(u^{(N)}(s,\varepsilon) + R_N(s,\varepsilon), s, \varepsilon) - f(u^{(N)}(s,\varepsilon), s, \varepsilon) \Big] \, \mathrm{d}s || + o(\varepsilon^N) \\ &\leq \int_{t_0}^{t} ||f(u^{(N)}(s,\varepsilon) + R_N(s,\varepsilon), s, \varepsilon) - f(u^{(N)}(s,\varepsilon), s, \varepsilon)|| \, \mathrm{d}s + o(\varepsilon^N) \\ &\leq L_f \int_{t_0}^{t} ||R_N(s,\varepsilon)|| \, \mathrm{d}s + o(\varepsilon^N) \,, \qquad t_0 \leq t \leq t_0 + T \,. \end{aligned} \tag{3.15}$$

The size of $||R_N||(t,\varepsilon)$ can now be estimated from the following Lemma:

Lemma 2. ***(Gronwall)*** *Suppose for $t \in [t_0, t_0 + T]$,*

$$\varphi(t) \leq a \int_{t_0}^{t} \varphi(s) \, ds + b \, (t - t_0) + c \,,$$

with $\varphi(t)$ continuous, $\varphi(t) \geq 0$ for $t \in [t_0, t_0 + T]$ and constants $a > 0$, $b, c \geq 0$, then

$$\varphi(t) \leq \Big(\frac{b}{a} + c \Big) e^{a(t - t_0)} - \frac{b}{a}$$

for $t \in [t_0, t_0 + T]$.

We now apply this Lemma to (3.15) with $a = L_f = O_{\mathrm{s}}(1)$, $b = o(\varepsilon^N)$ and $c = o(\varepsilon^N)$ to obtain

$$R_N = o(\varepsilon^N) \qquad \text{for } \; t_0 \leq t \leq t_0 + T \,,$$

with T independent of ε. Hence we have proved that $u^{(N)}$ given in (3.14) is indeed an AE of the true solution u_ε of the IVP (3.4). To summarize, we have:

Theorem 1. *Suppose that the problem P_0 in (3.11) has a unique solution $u_0(t)$ in the interval $t_0 \leq t \leq t_0 + T$. Moreover, suppose that the IVP denoted P_ε in (3.4) is such that*

i) $f \in C^N(\mathbb{R}^m \times D \times (0, \varepsilon_0])$,

ii) $f(u, t, \varepsilon)$ is Lipschitz with respect to u in any bounded domain $\Omega \subset \mathbb{R}^m$,

iii) both f and the initial condition V have AEs as in (3.6) as $\varepsilon \to 0$.

Then we have the regular AE

$$u_\varepsilon(t) = \sum_{n=0}^{N} \varepsilon^n u_n(t) + o(\varepsilon^N) \qquad \text{for } \; t_0 \leq t \leq T \,,$$

where the u_n for $n \geq 1$ satisfy the linear IVPs (3.12).

Remarks:

- Once the solution u_0 of th reduced problem P_0 is known, it is rather easy to increase the accuracy of the AE by solving just a sequence of linear IVPs. This should in general be much easier than solving the full problem P_ε, and it shows the purpose of perturbation theory: breaking down a complicated problem into a sequence of simpler problems we are able to solve.
- The AE in Theorem 1 holds only for finite time intervals $[t_0, t_0 + T]$, where T is independent of ε. If one wants to have approximations on longer time scales such as $T \sim 1/\varepsilon$, one needs to resort to other techniques such as averaging or the technique of multiple scales.

3.3. Example: The nonlinear spring

For $\varepsilon \in (0, \varepsilon_0]$, consider the IVP (Duffing equation)

$$\begin{aligned} &\frac{\mathrm{d}^2 x_\varepsilon}{\mathrm{d}t^2} + x_\varepsilon + \varepsilon x_\varepsilon^3 = 0\,, \qquad t > 0\,, \\ &x_\varepsilon(0) = 1\,, \qquad \frac{\mathrm{d}x_\varepsilon}{\mathrm{d}t}(0) = \varepsilon\,. \end{aligned}$$

We can easily recast this problem into the form of (3.4) by defining $u_\varepsilon = (u_{\varepsilon,1}, u_{\varepsilon,2})$ via $u_{\varepsilon,1} := x_\varepsilon$ and $u_{\varepsilon,2} := \frac{\mathrm{d}x_\varepsilon}{\mathrm{d}t}$, which leads to

$$P_\varepsilon \begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} u_{\varepsilon,1} \\ u_{\varepsilon,2} \end{pmatrix} = \begin{pmatrix} u_{\varepsilon,2} \\ -u_{\varepsilon,1} \end{pmatrix} + \varepsilon \begin{pmatrix} 0 \\ -u_{\varepsilon,1}^3 \end{pmatrix}, \\ \begin{pmatrix} u_{\varepsilon,1} \\ u_{\varepsilon,2} \end{pmatrix}(0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \varepsilon \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \end{cases} \tag{3.16}$$

Assuming $u_\varepsilon = u_0 + \varepsilon u_1 + \ldots$ leads to the series of problems

$$P_0 \begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} u_{0,1} \\ u_{0,2} \end{pmatrix} = \begin{pmatrix} u_{0,2} \\ -u_{0,1} \end{pmatrix}, \\ \begin{pmatrix} u_{0,1} \\ u_{0,2} \end{pmatrix}(0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \end{cases}$$

$$P_1 \begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} u_{1,1} \\ u_{1,2} \end{pmatrix} = \begin{pmatrix} u_{1,2} \\ -u_{1,1} \end{pmatrix} + \begin{pmatrix} 0 \\ -u_{0,1}^3 \end{pmatrix}, \\ \begin{pmatrix} u_{1,1} \\ u_{1,2} \end{pmatrix}(0) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \end{cases}$$

$$\vdots$$

The solution of the reduced problem is $x_0(t) = u_{0,1}(t) = \cos(t)$. The computation of the solution to the linear problem P_1 is left as training to the students in the exercise class of this course. There, it will be shown that an AE of the exact solution to P_ε , for finite time $0 \leq t \leq T$, can be given as

$$x_\varepsilon(t) = \cos(t) - \frac{3}{8}\,\varepsilon t \sin(t) + \varepsilon\Big[\sin(t) + \frac{1}{8}\cos^3(t) - \frac{1}{8}\cos(t)\Big] + o(\varepsilon)\,. \tag{3.17}$$

Remark that this is an AE only for $T < \infty$ independent of ε, due to the *secular term* $-\frac{3}{8}\,\varepsilon t \sin(t)$. This term is not $o(1)$ for $t \sim \frac{1}{\varepsilon}$, or more general for $\frac{1}{\varepsilon} \preceq t$, and it blows up for instance for $t \sim \frac{1}{\varepsilon^2}$.

3.4. The method of the strained coordinate

In the preceding section we obtained the AE (3.17) for the solution of the Duffing equation, which is valid for finite times only as $\varepsilon \to 0$. The general question is whether one can extend the validity of AEs stated in Theorem 1 to longer times, namely of the order $t \sim \frac{1}{\varepsilon}$. Indeed, there are multiple techniques to do so, some of them we shall discuss in this course on a fundamental basis: we start with the *method of the strained coordinate*, then introduce the method of *averaging* and finally briefly touch on the *technique of multiple scales*.

The method of the strained coordinate relies on a change of the independent variable $t \in D$ to a new variable τ via

$$t = (1 + \varepsilon\omega_1 + \varepsilon^2\omega_2 \ldots)\,\tau\,. \tag{3.18}$$

Here, the $\omega_i \in \mathbb{R}$ are free parameters which will be chosen appropriately in order to remove the secular terms from the AE. This method has been developed by Lindstedt and Poincaré for solving problems in celestial mechanics and is also called the *Lindstedt-Poincaré method* [4]. Let us consider again the nonlinear spring as an example. Substituting the transformation (3.18) into (3.16) yields

$$P_\varepsilon \begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}\tau}\begin{pmatrix} u^*_{\varepsilon,1} \\ u^*_{\varepsilon,2} \end{pmatrix} = (1 + \varepsilon\omega_1 + \varepsilon^2\omega_2 \ldots)\left[\begin{pmatrix} u^*_{\varepsilon,2} \\ -u^*_{\varepsilon,1} \end{pmatrix} + \varepsilon \begin{pmatrix} 0 \\ -(u^*_{\varepsilon,1})^3 \end{pmatrix}\right], \\ \begin{pmatrix} u^*_{\varepsilon,1} \\ u^*_{\varepsilon,2} \end{pmatrix}(0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \varepsilon(1 + \varepsilon\omega_1 + \varepsilon^2\omega_2 \ldots)\begin{pmatrix} 0 \\ 1 \end{pmatrix}, \end{cases}$$

where $u^*_\varepsilon(\tau) = u_\varepsilon(t)$ via the change of variables (3.18). Let us assume again a regular ansatz of the form

$$u^*_\varepsilon(\tau) = u^*_0(\tau) + \varepsilon u^*_1(\tau) + \ldots\,,$$

then substituting and sorting in powers of ε leads to the series of problems

$$P_0 \begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}\tau}\begin{pmatrix} u^*_{0,1} \\ u^*_{0,2} \end{pmatrix} = \begin{pmatrix} u^*_{0,2} \\ -u^*_{0,1} \end{pmatrix}, \\ \begin{pmatrix} u^*_{0,1} \\ u^*_{0,2} \end{pmatrix}(0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \end{cases}$$

$$P_1 \begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}\tau}\begin{pmatrix} u^*_{1,1} \\ u^*_{1,2} \end{pmatrix} = \begin{pmatrix} u^*_{1,2} \\ -u^*_{1,1} \end{pmatrix} + \begin{pmatrix} \omega_1 u^*_{0,2} \\ -(u^*_{0,1})^3 - \omega_1 u^*_{0,1} \end{pmatrix}, \\ \begin{pmatrix} u^*_{1,1} \\ u^*_{1,2} \end{pmatrix}(0) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \end{cases}$$

$$\vdots$$

From the reduced problem P_0 we obtain $u^*_{0,1}(\tau) = \cos(\tau)$ and $u^*_{0,2}(\tau) = -\sin(\tau)$. We solve the problem P_1 by using Duhamel's formula and the homogeneous solution, written in terms of the "rotation matrix" (see the exercise class)

$$\mathcal{R}(t, t_0) := \begin{pmatrix} \cos(t - t_0) & \sin(t - t_0) \\ -\sin(t - t_0) & \cos(t - t_0) \end{pmatrix},$$

as

$$\begin{pmatrix} u^*_{1,1} \\ u^*_{1,2} \end{pmatrix}(\tau) = \mathcal{R}(\tau, 0)\begin{pmatrix} 0 \\ 1 \end{pmatrix} + \int_0^\tau \mathcal{R}(\tau, s)\begin{pmatrix} \omega_1 u^*_{0,2}(s) \\ -(u^*_{0,1}(s))^3 - \omega_1 u^*_{0,1}(s) \end{pmatrix} \mathrm{d}s\,. \tag{3.19}$$

We are interested in particular in $u^*_{1,1}(\tau)$, which corresponds to the first line in eq. (3.19):

$$\begin{aligned} u^*_{1,1}(\tau) = \sin(\tau) &- \omega_1 \int_0^\tau \cos(\tau - s)\sin(s)\,\mathrm{d}s \\ &- \int_0^\tau \sin(\tau - s)\cos^3(s)\,\mathrm{d}s \\ &- \omega_1 \int_0^\tau \sin(\tau - s)\cos(s)\,\mathrm{d}s\,. \end{aligned} \tag{3.20}$$

The second integral in this equation yields the terms from the AE obtained previously in (3.17) (see the exercise class for details),

$$-\int_0^\tau \sin(\tau - s)\cos^3(s)\,\mathrm{d}s = -\frac{3}{8}\tau\sin(\tau) + \frac{1}{8}\cos^3(\tau) - \frac{1}{8}\cos(\tau)\,. \tag{3.21}$$

We still have the free parameter ω_1 in order to cancel the secular term in (3.21). For this let us compute the remaining two integrals in (3.20). We realize that

$$\int_0^\tau \sin(\tau - s)\cos(s)\,\mathrm{d}s = \int_0^\tau \sin(s)\cos(\tau - s)\,\mathrm{d}s\,,$$

and compute the remaining integral using

$$\cos(\tau - s) = \cos(\tau)\cos(s) + \sin(\tau)\sin(s)\,,$$

which leads to

$$\begin{aligned}\int_0^\tau \cos(\tau - s)\sin(s)\,\mathrm{d}s &= \cos(\tau)\int_0^\tau \cos(s)\sin(s)\,\mathrm{d}s + \sin(\tau)\int_0^\tau \sin^2(s)\,\mathrm{d}s \\ &= \frac{1}{2}\cos(\tau)\sin^2(\tau) + \sin(\tau)\frac{1}{2}\Big[-\sin(\tau)\cos(\tau) + \tau\Big] \\ &= \frac{1}{2}\tau\sin(\tau)\,.\end{aligned}$$

Therefore, we obtain

$$-2\omega_1\int_0^\tau \cos(\tau - s)\sin(s)\,\mathrm{d}s = -\omega_1\tau\sin(\tau)\,.$$

In order to cancel the secular term in (3.21) we need to choose

$$\omega_1 = -\frac{3}{8}\,.$$

The result for $u^*_{1,1}$ then becomes

$$u^*_{1,1}(\tau) = \sin(\tau) + \frac{1}{8}\cos^3(\tau) - \frac{1}{8}\cos(\tau)\,,$$

which is free of secularities. In principle, this procedure could be carried out also for $u^*_{2,1}$ and for even higher orders. Up to first order we obtained

$$x^*_\varepsilon(\tau) = u^*_{\varepsilon,1}(\tau) = \cos(\tau) + \varepsilon\Big[\sin(\tau) + \frac{1}{8}\cos^3(\tau) - \frac{1}{8}\cos(\tau)\Big] + o(\varepsilon)\,,$$

and

$$\tau = \frac{t}{(1 - \frac{3}{8}\varepsilon + o(\varepsilon))}\,.$$

Here, the expansion of the denominator yields $\omega_\varepsilon := 1 + \frac{3}{8}\varepsilon + o(\varepsilon)$, which is the angular frequency of the nonlinear spring, determined in the process. The solution of the original problem (3.16) is thus approximated as

$$x_\varepsilon(t) = x^*_\varepsilon(\omega_\varepsilon t) = \cos\left[\Big(1 + \frac{3}{8}\varepsilon\Big)t\right] + O(\varepsilon)\,. \tag{3.22}$$

The theory of averaging we discuss next will provide the rigorous justification that the AE (3.22) is indeed valid for long times of the order $t \sim \frac{1}{\varepsilon}$.

3.5. Averaging

In section 3.2 we dealt with problems of the form

$$\frac{\mathrm{d}u_\varepsilon}{\mathrm{d}t} = f(u_\varepsilon, t, \varepsilon)\,, \qquad u_\varepsilon(t_0) = V \in \mathbb{R}^m\,, \qquad t \in D \subset \mathbb{R}\,, \tag{3.23}$$

and proved that a regular AE of the solution u_ε exist for finite times $t_0 \le t \le t_0 + T$, with $T < \infty$ independent of ε, if f is sufficiently regular and can be expanded in a power series in ε (c.f. Theorem 1). In this section we deal with a method that is appropriate for constructing AEs of u_ε that are uniformly valid in a much longer time interval $t_0 \le t \le t_0 + T/\varepsilon$ - this method is called *averaging*.

3.5.1. Generic example

Before we state the fundamental theorem of averaging, let us illustrate the idea by means of a simple example. Consider a perturbed Hamiltonian system for the "action-angle variables" I and φ, given as the IVP

$$\begin{cases} \dfrac{\mathrm{d}I}{\mathrm{d}t} = \varepsilon g(\varphi)\,, & \dfrac{\mathrm{d}\varphi}{\mathrm{d}t} = \omega\,, \qquad t > 0\,, \\ I(0) = I_0\,, & \varphi(0) = \varphi_0\,. \end{cases} \tag{3.24}$$

Here, $\omega \in \mathbb{R}$ is a given frequency and the function $g(\varphi)$ is assumed continuous and 2π-periodic in φ. The solution of this system reads

$$\varphi(t) = \varphi_0 + \omega t\,, \qquad I(t) = I_0 + \varepsilon \int_0^t g(\varphi_0 + \omega s)\,\mathrm{d}s\,. \tag{3.25}$$

We estimate the integral to

$$\varepsilon \int_0^t g(\varphi_0 + \omega s)\,\mathrm{d}s \le \varepsilon t \max_{\varphi \in 2\pi} |g(\varphi)|\,,$$

and thus, for finite times $t \sim 1$, $I(t) = I_0 + O(\varepsilon)$, which is accordance with Theorem 1. However, we can do better. Let us introduce the average $\overline{g}$ and fluctuation $\widetilde{g}$ of g with respect to the angle variable φ,

$$\overline{g} := \frac{1}{2\pi} \int_0^{2\pi} g(\varphi)\,\mathrm{d}\varphi\,, \qquad \widetilde{g}(\varphi) := g(\varphi) - \overline{g}\,.$$

By construction the average of the fluctuation is zero, $\overline{\widetilde{g}} = 0$. We can now split the integral in (3.25) into two parts and write the solution of (3.24) as

$$\begin{aligned} \varphi(t) = \varphi_0 + \omega t\,, \qquad I(t) &= I_0 + \varepsilon t\,\overline{g} + \varepsilon \int_0^t \widetilde{g}(\varphi_0 + \omega s)\,\mathrm{d}s \\ &= I_0 + \varepsilon t\,\overline{g} + \frac{\varepsilon}{\omega} \int_0^{\omega t} \widetilde{g}(\varphi_0 + s')\,\mathrm{d}s' \end{aligned}$$

We see that the remaining integral is 2π periodic in φ,

$$\begin{aligned}\int_0^{\omega t+2\pi} \widetilde{g}(\varphi_0+s')\,\mathrm{d}s' &= \int_0^{\omega t} \widetilde{g}(\varphi_0+s')\,\mathrm{d}s' + \int_{\omega t}^{\omega t+2\pi} \widetilde{g}(\varphi_0+s')\,\mathrm{d}s' \\ &= \int_0^{\omega t} \widetilde{g}(\varphi_0+s')\,\mathrm{d}s' + \underbrace{\overline{\widetilde{g}}}_{=0} \\ &=: h(\omega t)\,.\end{aligned}$$

Hence the function $h(\omega t)$ is bounded for $t \in [0,\infty)$ and

$$\varphi(t) = \varphi_0 + \omega t\,, \qquad I(t) = I_0 + \varepsilon t\,\overline{g} + \frac{\varepsilon}{\omega} h(\omega t)\,.$$

It seems that the dynamics of $I(t)$ can be split into two parts:

1. a "systematic part" $I_0+\varepsilon t\,\overline{g}$, which is $O(1)$ for times $0 \le t \le \frac{T}{\varepsilon}$ with T an arbitrary constant, and whose derivative is $\varepsilon\overline{g}$,
2. an "oscillating part", given by the periodic function $\frac{\varepsilon}{\omega}h(\omega t)$, which is $O(\varepsilon)$ for $t \in [0,\infty)$.

We deduce that for times $0 \le t \le \frac{T}{\varepsilon}$, the solution of the "averaged problem"

$$\begin{cases} \dfrac{\mathrm{d}I}{\mathrm{d}t} = \varepsilon\overline{g}\,, \qquad \dfrac{\mathrm{d}\varphi}{\mathrm{d}t} = \omega\,, \qquad t>0\,, \\ I(0) = I_0\,, \qquad \varphi(0) = \varphi_0\,, \end{cases}$$

is an AE to $O(\varepsilon)$ of the solution of (3.24),

$$\varphi(t) = \varphi_0 + \omega t\,, \qquad I(t) = I_0 + \varepsilon t\,\overline{g} + O(\varepsilon)\,, \qquad 0 \le t \le \frac{T}{\varepsilon}\,.$$

Here, the secular term has been isolated and given a clear physical meaning as the systematic motion of the system, around which the true solution fluctuates.

3.5.2. K.B.M. Theorem

Let us now come from the example to the general case and state the main theorem of averaging. The averaging principle has first been proved by Krilov, Bogoliubov and Mitropolski and is thus known as the KBM-method (or K.B.M. theorem). There are several variants of the theorem, depending on the properties of the vector field f in (3.23), for instance periodicity with respect to t.

Let us consider IVPs of the form

$$P_\varepsilon \begin{cases} \dfrac{\mathrm{d}u_\varepsilon}{\mathrm{d}t} = \varepsilon f(u_\varepsilon, t)\,, \qquad t>0\,, \\ u_\varepsilon(0) = V \in \mathbb{R}^m\,, \end{cases} \tag{3.26}$$

for $\varepsilon \in (0, \varepsilon_0]$. Remark that the vector field is multiplied by ε, which is necessary for having existence and uniqueness of a solution in an interval of order $t \sim \frac{1}{\varepsilon}$ under the following assumptions on the vector field f:

i) f is T-periodic in t with T independent of ε,

ii) f is defined on $\Omega = B_r(V) \times [0, \infty)$, with the ball $B_r(V) = \{x \in \mathbb{R}^m : ||x - V|| \leq r\}$; it is continuous in Ω and continuously differentiable in $B_r(V)$,

iii) f and its Jacobian $\frac{\partial f}{\partial x} \in \mathbb{R}^{m \times m}$ have in Ω norms which are bounded by a constant M, independent of ε.

Under these assumptions we have

Theorem 2. *Let u_ε denote the solution of the problem P_ε as in (3.26), and further let u_0 stand for the solution of*

$$\overline{P}_\varepsilon \begin{cases} \dfrac{\mathrm{d}u_0}{\mathrm{d}t} = \varepsilon \overline{f}(u_0)\,, & t > 0\,, \\ u_0(0) = V \in \mathbb{R}^m\,, \end{cases}$$

where

$$\overline{f}(x) = \frac{1}{T} \int_0^T f(x,t)\,\mathrm{d}t\,,$$

and the integration is performed as if x is constant. Then

$$u_\varepsilon - u_0 = O(\varepsilon) \qquad \text{for } 0 \leq t \leq \frac{t_1}{\varepsilon}\,,$$

where $t_1 \sim 1$.

Proof. Rescaling to the macroscopic time $\tau = \varepsilon t$, the Cauchy-Lipschitz (or Picard-Lindelöf) theorem gives existence and uniqueness of u_ε in the interval $0 \leq t \leq \frac{d}{M\varepsilon}$, where $M = \max_\Omega ||f(u,t)||$, $||\cdot||$ being some vector norm in $\mathbb{R}^m$. Moreover, since $\max_{B_r(V)} ||\overline{f}(u)|| \leq M$, the unique solution u_0 exists in the same interval and stays inside the ball $B_r(V)$.

In what follows we shall denote by $\widetilde{f}(x,t) := f(x,t) - \overline{f}(x)$ the fluctuation of the vector field f. Our strategy is to make a clever ansatz for the approximation of u_ε in the form $u^{(1)} = u_0 + O(\varepsilon)$ and to show that the difference $u_\varepsilon - u^{(1)}$ is of order ε on the time scale $t \sim \frac{1}{\varepsilon}$. To achieve this we need to exploit the average and the fluctuation of the vector field f. To get an idea, let us play around with the true solution u_ε in order

to derive an ansatz involving u_0:

$$\begin{aligned} u_\varepsilon(t) &= V + \varepsilon \int_0^t f(u_\varepsilon(s), s)\,\mathrm{d}s \\ &\approx V + \varepsilon \int_0^t f(u_\varepsilon(t), s)\,\mathrm{d}s \\ &= V + \varepsilon \int_0^t \overline{f}(u_\varepsilon(t))\,\mathrm{d}s + \varepsilon \int_0^t \widetilde{f}(u_\varepsilon(t), s)\,\mathrm{d}s \\ &\approx V + \varepsilon \int_0^t \overline{f}(u_0(s))\,\mathrm{d}s + \varepsilon \int_0^t \widetilde{f}(u_0(t), s)\,\mathrm{d}s\,. \end{aligned}$$

This is what we want, because we recognize the solution of the averaged problem,

$$u_0(t) = V + \varepsilon \int_0^t \overline{f}(u_0(s))\,\mathrm{d}s\,.$$

Therefore, let us propose as an approximation to u_ε the function

$$u^{(1)}(t) := u_0(t) + \varepsilon\, h(t, u_0(t))\,,$$

where we defined

$$h(t, u_0(t)) := \int_0^t \widetilde{f}(u_0(t), s)\,\mathrm{d}s\,.$$

As usual, we now aim to derive an IVP for the error term $u_\varepsilon - u^{(1)}$ and prove that it stays small on a time scale of order $t \sim \frac{1}{\varepsilon}$. For this, note that

$$\begin{aligned} \frac{\mathrm{d}u_\varepsilon}{\mathrm{d}t} - \frac{\mathrm{d}u^{(1)}}{\mathrm{d}t} &= \varepsilon f(u_\varepsilon, t) - \varepsilon\overline{f}(u_0) - \varepsilon\frac{\partial h}{\partial t} - \varepsilon\frac{\partial h}{\partial x}\cdot\frac{\mathrm{d}u_0}{\mathrm{d}t} \\ &= \varepsilon f(u_\varepsilon, t) - \varepsilon\overline{f}(u_0) - \varepsilon\widetilde{f}(u_0, t) - \varepsilon^2\frac{\partial h}{\partial x}\cdot\overline{f}(u_0) \\ &= \varepsilon\Big[f(u_\varepsilon, t) - f(u^{(1)}, t)\Big] + \varepsilon\Big[f(u^{(1)}, t) - f(u_0, t)\Big] - \varepsilon^2\frac{\partial h}{\partial x}\cdot\overline{f}(u_0)\,. \end{aligned}$$

Integrating this equation, using the initial condition $u^{(1)}(0) = u_0(0) = V$, and taking the vector norm yields

$$\begin{aligned} ||u_\varepsilon(t) - u^{(1)}(t)|| &\le \varepsilon \int_0^t ||f(u_\varepsilon, s) - f(u^{(1)}, s)||\,\mathrm{d}s + \varepsilon \int_0^t ||f(u^{(1)}, s) - f(u_0, s)||\,\mathrm{d}s \\ &\quad - \varepsilon^2 \int_0^t \Big|\Big|\frac{\partial h}{\partial x}\cdot\overline{f}(u_0)\Big|\Big|\,\mathrm{d}s \\ &\le \varepsilon L_f \int_0^t ||u_\varepsilon - u^{(1)}||\,\mathrm{d}s + \varepsilon L_f \int_0^t ||u^{(1)} - u_0||\,\mathrm{d}s - \varepsilon^2 \int_0^t \Big|\Big|\frac{\partial h}{\partial x}\cdot\overline{f}(u_0)\Big|\Big|\,\mathrm{d}s\,. \end{aligned}$$

In the second line we used that f is Lipschitz in $B_r(V)$ because it belongs to $C^1(B_r(V))$ and denoted the Lipschitz constant by L_f, independent of (t, ε). Moreover, we know that $||u^{(1)} - u_0|| = \varepsilon||h|| = O(\varepsilon)$ uniformly in $[0, \infty)$ because h is T-periodic. We also assumed that the norms of f and its Jacobian are bounded in Ω. This yields

$$||u_\varepsilon(t) - u^{(1)}(t)|| \leq \varepsilon L_f \int_0^t ||u_\varepsilon - u^{(1)}|| \, \mathrm{d}s + \varepsilon^2 t \, C \,, \qquad C = O(1) \,.$$

We can now apply Gronwall's Lemma 2 with $a = \varepsilon L_f$, $b = \varepsilon^2 C$ and $c = 0$ to obtain

$$||u_\varepsilon(t) - u^{(1)}(t)|| \leq \frac{\varepsilon C}{L_f} \mathrm{e}^{\varepsilon L_f t} - \frac{\varepsilon C}{L_f} = O(\varepsilon) \qquad \text{for } 0 \leq t \leq \frac{t_1}{\varepsilon} \,,$$

with t_1 arbitrary of order one. Moreover, since h is bounded, also

$$u_\varepsilon - u_0 = O(\varepsilon) \qquad \text{for } 0 \leq t \leq \frac{t_1}{\varepsilon} \,.$$

□

3.5.3. The standard form

The K.B.M-Theorem considers IVPs of the form

$$P_\varepsilon \begin{cases} \dfrac{\mathrm{d}u_\varepsilon}{\mathrm{d}t} = \varepsilon f(u_\varepsilon, t) \,, & t > 0 \,, \\ u_\varepsilon(0) = V \in \mathbb{R}^m \,, \end{cases}$$

where the vector field f is multiplied by ε. This is called the *standard form* in the theory of averaging. A technique called *variation of the constants* is often useful to transform a perturbation problem into the standard form. Consider the "full" perturbation problem

$$P_\varepsilon^{\text{full}} \begin{cases} \dfrac{\mathrm{d}u_\varepsilon}{\mathrm{d}t} = f_0(u_\varepsilon, t) + \varepsilon f_1(u_\varepsilon, t) \,, & t > 0 \,, \\ u_\varepsilon(0) = V \in \mathbb{R}^m \,, \end{cases}$$

The reduced problem

$$\frac{\mathrm{d}u_0}{\mathrm{d}t} = f_0(u_0, t) \,, \qquad u_0(0) = V \in \mathbb{R}^m \,, \tag{3.27}$$

has as solution the flow map $\Phi_0(t, V)$, $\Phi_0 : \mathbb{R}_+ \times \mathbb{R}^m \to \mathbb{R}^m$, with $\Phi_0(0, V) = V$ (see Appendix A for details on the flow). Variation of constants assumes that the solution to the full problem $P_\varepsilon^{\text{full}}$ can be written as

$$u_\varepsilon(t) = \Phi_0(t, V(t)) \,.$$

Taking the time derivative yields

$$\frac{\mathrm{d}u_\varepsilon}{\mathrm{d}t} = \frac{\mathrm{d}\Phi_0(t, V)}{\mathrm{d}t} = \frac{\partial \Phi_0(t, V)}{\partial t} + \frac{\partial \Phi_0}{\partial x}(t, V) \cdot \frac{\mathrm{d}V}{\mathrm{d}t} = f_0(\Phi_0(t, V), t) + \varepsilon f_1(\Phi_0(t, V), t) \,,$$

where $\partial\Phi_0/\partial x \in \mathbb{R}^{m\times m}$ is the Jacobian of the flow. Since Φ_0 satisfies the reduced problem (3.27), in case that $\partial\Phi_0/\partial x$ is invertible, we obtain an equation for $V(t)$ in the standard form,

$$\frac{\mathrm{d}V}{\mathrm{d}t} = \varepsilon \left(\frac{\partial\Phi_0}{\partial x}(t,V)\right)^{-1} f_1(\Phi_0(t,V),t)\,.$$

This equation can be very messy, however in some cases variations of the constants is very useful. For *quasilinear problems* for instance, where $f_0(u_\varepsilon,t) = Au_\varepsilon$ in the reduced problem, with $A \in \mathbb{R}^{m\times m}$ a constant matrix, we obtain for the flow

$$\Phi_0(t,V) = \mathrm{e}^{At}\,V\,.$$

The standard form then becomes

$$\frac{\mathrm{d}V}{\mathrm{d}t} = \varepsilon\,\mathrm{e}^{-At}\,f_1(\mathrm{e}^{At}\,V,t)\,.$$

Clearly, if there is an eigenvalue of A with non-vanishing real part, then we might have problems with this equation even if f_1 is bounded. Moreover, for $f_0(u_\varepsilon,t) = Au_\varepsilon + g(t)$ in the reduced problem, with g an arbitrary continuous function, we have from Duhamel's formula the flow

$$\Phi_0(t,V) = \mathrm{e}^{At}\,V + \int_0^t \mathrm{e}^{A(t-s)}g(s)\,\mathrm{d}s\,. \tag{3.28}$$

The corresponding standard form reads

$$\frac{\mathrm{d}V}{\mathrm{d}t} = \varepsilon\,\mathrm{e}^{-At}\,f_1(\Phi_0(t,V),t)\,,$$

where Φ_0 from (3.28) is to be inserted.

3.6. The method of multiple scales

For this we have no time. In case of interest check the literature list given for the course.

4. Singular perturbations of linear ODEs

4.1. Generalities

When studying singular perturbations we are still concerned with initial/boundary value problems of the form (3.1). However, this time an ansatz of the form (3.3) will fail to produce an AE of the solution that is uniformly valid in the domain D. Instead, we will be faced with the emergence of so-called *boundary layers* in the vicinity of sub-manifolds of D. These are regions where the solution u_ε shows strong variations, which eventually lead to discontinuities as $\varepsilon \to 0$. The prototypical example of such behavior can be observed in the problem

$$\varepsilon \frac{\mathrm{d}x}{\mathrm{d}t}_{\varepsilon} + x_\varepsilon = 0\,, \qquad t > 0\,, \qquad x_\varepsilon(0) = 1\,.$$

The ε-family of solutions reads $x_\varepsilon(t) = \mathrm{e}^{-t/\varepsilon}$ and we observe the formation of a boundary layer at $x = 0$ as $\varepsilon \to 0$. A regular ansatz of the form $x_\varepsilon = x_0 + \varepsilon x_1 + \ldots$ would moreover fail to approximate the correct solution uniformly in $[0, \infty)$ as $\varepsilon \to 0$.

In the analysis of singular perturbations, the concept of a *formal approximation* plays an important role:

Definition 10. Suppose that $P_\varepsilon[u_\varepsilon] = 0$ for $\varepsilon \in (0, \varepsilon_0]$, hence u_ε denotes the ε-family of solutions to problem (3.1). A **formal approximation of order** $\delta(\varepsilon)$ of u_ε in $\tilde{D} \subset D$ is a function $v_\varepsilon : \tilde{D} \to \mathbb{R}^m$ such that $P_\varepsilon[v_\varepsilon] = O(\delta(\varepsilon))$ uniformly in $\tilde{D}$.

This definition means that the formal approximation satisfies the equations and the initial conditions in P_ε up to an error of order $O(\delta(\varepsilon))$. This does not mean that it is an AE of the solution u_ε. In order to prove that $u_\varepsilon - v_\varepsilon = o(1)$ uniformly in D we still need to show that the error term is small. This is usually done by means of *a-priori estimates* for the error, which satisfies some differential equation. The general strategy for treating perturbation problems can be summarized as follows:

1. Find a formal approximation to the solution.
2. Come up with an a-priori estimate for the error and prove that it is small uniformly in D.

In this course we will apply this strategy to a linear IVP and to a linear BVP which feature boundary layers and are thus singular perturbation problems. For such kind of problems, finding a formal approximation can be difficult and needs some more sophisticated methods than in the regular case.

4.2. The initial value problem

In this section we will formulate a theorem regarding the AEs of solutions to the following linear IVP, for $\varepsilon \in (0, \varepsilon_0]$,

$$P_\varepsilon \begin{cases} L_\varepsilon[u_\varepsilon] = \varepsilon \dfrac{\mathrm{d}^2 u_\varepsilon}{\mathrm{d}t^2} + a(t)\dfrac{\mathrm{d}u_\varepsilon}{\mathrm{d}t} + b(t)u_\varepsilon = f(t)\,, & t > 0\,, \\ u_\varepsilon(0) = \alpha\,, \qquad \dfrac{\mathrm{d}u_\varepsilon}{\mathrm{d}t}(0) = \beta\,, \end{cases} \tag{4.1}$$

where $\alpha, \beta \in \mathbb{R}$ are constants, independent of ε. Regarding the coefficient functions in L_ε we make the following assumptions:

i) $a, b, f \in C^1([0, \infty))$,

ii) $a(t) \geq a_0 > 0$ for $t \geq 0$ with a_0 independent of ε.

The second condition is particularly important in the subsequent construction of an AE. Setting $\varepsilon = 0$ in P_ε yields the reduced problem, a first-order ODE which is ill-posed because of the two initial conditions which in general cannot be fulfilled simultaneously. Even though the perturbation is singular, we can expect to extract some useful information from the reduced problem. Neglecting for the moment the second initial condition, let us consider the problem

$$\begin{cases} L_0[w_0] = a(t)\dfrac{\mathrm{d}w_0}{\mathrm{d}t} + b(t)w_0 = f(t)\,, & t > 0\,, \\ w_0(0) = \alpha\,. \end{cases} \tag{4.2}$$

The solution reads

$$w_0(t) = \alpha\, \mathrm{e}^{q(t,0)} + \int_0^t \mathrm{e}^{q(t,s)} f(s)\, \mathrm{d}s\,, \tag{4.3}$$

with

$$q(t,s) = -\int_s^t \frac{b(s')}{a(s')}\, \mathrm{d}s'\,.$$

Due to our assumptions on a, the function q is well-defined. We also see the difficulty with the second initial condition,

$$\frac{\mathrm{d}w_0}{\mathrm{d}t}(0) = -\frac{b(0)}{a(0)}\alpha + f(0) \neq \beta \qquad \text{in general}\,.$$

We need another method to be able to incorporate the second initial condition. In order to resolve the boundary-layer region near $t = 0$, we will "zoom in" and define a "microscopic variable" ξ via

$$t = \varepsilon^\nu \xi\,, \qquad \nu > 0\,.$$

In the new variable ξ the differential operator L_ε reads

$$L_\varepsilon = \varepsilon^{1-2\nu}\frac{\mathrm{d}^2}{\mathrm{d}\xi^2} + \frac{a(\varepsilon^\nu \xi)}{\varepsilon^\nu}\frac{\mathrm{d}}{\mathrm{d}\xi} + b(\varepsilon^\nu \xi)\,.$$

Much like in the algebraic toy problems from the beginning of the course, let us apply the principle of dominant balance: choosing $\nu = 1/2$ would lead to an unbalanced reduced problem, such that $\nu = 1$ is the only meaningful choice. We thus define the "microscopic operator"

$$\mathcal{L}_\varepsilon := \frac{\mathcal{L}_{-1}}{\varepsilon} + \mathcal{L}_0 \,,$$

with

$$\mathcal{L}_{-1} = \frac{\mathrm{d}^2}{\mathrm{d}\xi^2} + a(0)\frac{\mathrm{d}}{\mathrm{d}\xi} \,, \qquad \mathcal{L}_0 = \xi a'(\vartheta\varepsilon\xi)\frac{\mathrm{d}}{\mathrm{d}\xi} + b(\varepsilon\xi) \qquad \text{for some } \vartheta \in (0,1) \,.$$

Here, we used the mean value theorem (Mittelwertsatz der Differentialrechnung) to write

$$a(t) = a(0) + ta'(\vartheta t) \qquad \text{for some } \vartheta \in (0,1) \,.$$

Our plan is that the solution of $\mathcal{L}_{-1}[v] = 0$ with appropriate boundary conditions describes in some way the behavior of u_ε in the boundary layer near $t = 0$. Hence, we attempt a formal approximation of the form

$$u_\varepsilon(t) = w_0(t) + v\Big(\frac{t}{\varepsilon}\Big) + R_\varepsilon(t) = w_0(t) + v(\xi) + R_\varepsilon(t) \,, \tag{4.4}$$

where w_0 is the solution of (4.2), R_ε denotes the error term and v is the solution of the **boundary value problem**

$$\begin{cases} \mathcal{L}_{-1}[v] = \dfrac{\mathrm{d}^2 v}{\mathrm{d}\xi^2} + a(0)\dfrac{\mathrm{d}v}{\mathrm{d}\xi} = 0 \,, \qquad \xi > 0 \,, \\ \dfrac{1}{\varepsilon}\dfrac{\mathrm{d}v}{\mathrm{d}\xi}(0) = \beta - \dfrac{\mathrm{d}w_0}{\mathrm{d}t}(0) \,, \qquad \lim\limits_{\xi\to\infty} v(\xi) = 0 \,. \end{cases} \tag{4.5}$$

While the boundary condition at $t = 0$ ensures that our formal approximation (4.4) indeed satisfies the second initial condition of the original problem, the boundary condition at $\xi \to \infty$ ensures that v is a boundary layer term with contributions only in a small right neighborhood of $t = 0$. In particular, writing $\frac{\mathrm{d}v}{\mathrm{d}\xi} = r$ we obtain

$$r(\xi) = \varepsilon\left(\beta - \frac{\mathrm{d}w_0}{\mathrm{d}t}(0)\right)\mathrm{e}^{-a(0)\xi} \,,$$

which after integration in ξ yields

$$v(\xi) = \varepsilon\left(\frac{\frac{\mathrm{d}w_0}{\mathrm{d}t}(0) - \beta}{a(0)}\right)\mathrm{e}^{-a(0)\xi} + c \,,$$

where c stands for the constant of integration. The second boundary condition in (4.5) then leads to $c = 0$ and thus

$$v\Big(\frac{t}{\varepsilon}\Big) = \varepsilon\left(\frac{\frac{\mathrm{d}w_0}{\mathrm{d}t}(0) - \beta}{a(0)}\right)\mathrm{e}^{-a(0)\frac{t}{\varepsilon}} \,. \tag{4.6}$$

We can now go on to prove

Lemma 3. *Suppose that u_ε solves (4.1) under the assumptions stated, that w_0 is the solution of (4.2), and let v denote the solution of the microscopic problem (4.5). Then $w_0(t) + v(t/\varepsilon)$ is a formal approximation of order $O(\varepsilon)$ of $u_\varepsilon(t)$, i.e.*

$$L_\varepsilon[w_0 + v] = f(t) + O(\varepsilon)\,,$$

and

$$(w_0 + v)(0) = \alpha + O(\varepsilon)\,, \qquad \frac{\mathrm{d}(w_0 + v)}{\mathrm{d}t}(0) = \beta + O(\varepsilon)\,,$$

for $t \in [0, T]$ with $T < \infty$ independent of ε.

Proof. Applying the differential operator L_ε to $w_0(t) + v(t/\varepsilon)$ yields

$$\begin{aligned} L_\varepsilon(w_0 + v) &= \varepsilon\frac{\mathrm{d}^2 w_0}{\mathrm{d}t^2} + \frac{1}{\varepsilon}\frac{\mathrm{d}^2 v}{\mathrm{d}\xi^2} + a(t)\frac{\mathrm{d}w_0}{\mathrm{d}t} + \frac{a(t)}{\varepsilon}\frac{\mathrm{d}v}{\mathrm{d}\xi} + b(t)(w_0 + v) \\ &= \varepsilon\frac{\mathrm{d}^2 w_0}{\mathrm{d}t^2} + \underbrace{\frac{1}{\varepsilon}\frac{\mathrm{d}^2 v}{\mathrm{d}\xi^2} + \frac{a(0)}{\varepsilon}\frac{\mathrm{d}v}{\mathrm{d}\xi}}_{=0} - \frac{a(0)}{\varepsilon}\frac{\mathrm{d}v}{\mathrm{d}\xi} + \underbrace{a(t)\frac{\mathrm{d}w_0}{\mathrm{d}t} + b(t)w_0}_{=f(t)} + \frac{a(t)}{\varepsilon}\frac{\mathrm{d}v}{\mathrm{d}\xi} + b(t)v \\ &= f(t) + \varepsilon\frac{\mathrm{d}^2 w_0}{\mathrm{d}t^2} + \frac{a(t) - a(0)}{\varepsilon}\frac{\mathrm{d}v}{\mathrm{d}\xi} + b(t)v\,. \end{aligned}$$

It is easy to show that the last three terms in this expression are of order $O(\varepsilon)$ on any bounded segment $[0, T]$ with T independent of ε. Since a and b and their first derivatives are continuous, they are bounded in $[0, T]$; then (4.3) implies that $\frac{\mathrm{d}^2 w_0}{\mathrm{d}t^2}(t)$ is bounded in $[0, T]$ and hence $\frac{\mathrm{d}^2 w_0}{\mathrm{d}t^2} = O(1)$. Moreover, from (4.6) we obtain that $v(t/\varepsilon) = O(\varepsilon)$ uniformly in $[0, T]$ and hence $bv = O(\varepsilon)$ uniformly in $[0, T]$. For the remaining term we use the mean value theorem to write

$$\frac{a(t) - a(0)}{\varepsilon}\frac{\mathrm{d}v}{\mathrm{d}\xi}\left(\frac{t}{\varepsilon}\right) = \frac{t}{\varepsilon}a'(\vartheta t)\frac{\mathrm{d}v}{\mathrm{d}\xi}\left(\frac{t}{\varepsilon}\right) = \varepsilon\xi a'(\vartheta t)\left(\beta - \frac{\mathrm{d}w_0}{\mathrm{d}t}(0)\right)\mathrm{e}^{-a(0)\xi}\,.$$

We remark that a' is uniformly bounded in $[0, T]$ and that $\xi\mathrm{e}^{-a(0)\xi}$ is uniformly bounded in $[0, \infty)$. As a conclusion we have

$$L_\varepsilon(w_0 + v) = f(t) + O(\varepsilon)\,.$$

With regards to the initial conditions we have, using (4.5) and (4.6),

$$(w_0 + v)(0) = \alpha + \varepsilon\left(\frac{\frac{\mathrm{d}w_0}{\mathrm{d}t}(0) - \beta}{a(0)}\right), \qquad \frac{\mathrm{d}(w_0 + v)}{\mathrm{d}t} = \frac{\mathrm{d}w_0}{\mathrm{d}t} + \beta - \frac{\mathrm{d}w_0}{\mathrm{d}t} = \beta\,.$$

which shows that $w_0 + v$ is indeed a formal approximation of u_ε in $[0, T]$, because it satisfies the equations as well as the initial conditions up to order $O(\varepsilon)$. □

Using the above Lemma, we can set up an equation for the remainder $R_\varepsilon = u_\varepsilon - w_0 - v$ in (4.4):

$$\begin{cases} L_\varepsilon[R_\varepsilon] = g(t,\varepsilon)\,, & 0 < t \le T\,, \\ R_\varepsilon(0) = \alpha_1\,, & \dfrac{\mathrm{d}R_\varepsilon}{\mathrm{d}t} = 0\,, \end{cases} \tag{4.7}$$

with

$$g(t,\varepsilon) = -\varepsilon\frac{\mathrm{d}^2 w_0}{\mathrm{d}t^2} - \frac{a(t)-a(0)}{\varepsilon}\frac{\mathrm{d}v}{\mathrm{d}\xi} - b(t)v = O(\varepsilon)\,, \qquad \alpha_1 = \varepsilon\frac{\beta - \frac{\mathrm{d}w_0}{\mathrm{d}t}(0)}{a(0)} = O(\varepsilon)\,.$$

An a-priori estimate for the solution R_ε of (4.7) and its derivative $\frac{\mathrm{d}R_\varepsilon}{\mathrm{d}t}$ can be be obtained using the *method of energy integrals*. This method consists in deriving an estimate for the energy

$$W_\varepsilon := m(R_\varepsilon^2 + \varepsilon(R_\varepsilon')^2)\,,$$

where m is some generic constant. Estimating the energy rather than the solution of an equation is useful if we are interested in bounds for the first derivative of the solution, in addition to bounds for the solution itself. In the present case we observe that the "microscopic term" v is of order $O(\varepsilon)$, while its first derivative is of order $O(1)$. Hence we expect that the solution w_0 of the modified reduced problem (4.2) is an asymptotic approximation (AA) of u_ε in $[0,T]$ and that $\frac{\mathrm{d}w_0}{\mathrm{d}t} + \frac{\mathrm{d}v}{\mathrm{d}t}$ is an AA of $\frac{\mathrm{d}u_\varepsilon}{\mathrm{d}t}$ in $[0,T]$. We shall now go on to show that this is indeed the case. For the energy we have

Lemma 4. *Let a_0 denote the lower bound of the function $a(t)$ in the differential operator L_ε described in (4.1). For $\varepsilon \in (0,\varepsilon_0]$ with a suitable constant ε_0, the energy $W_\varepsilon = \frac{a_0}{2}(R_\varepsilon^2 + \varepsilon(R_\varepsilon')^2)$ corresponding to the solution of problem (4.7) satisfies*

$$W_\varepsilon \le M(||g||^2_{[0,T]} + \alpha_1^2)\,\mathrm{e}^{\frac{2M}{a_0}T}\,,$$

where $||\cdot||_{[0,T]}$ denotes the L^2-norm and with

$$M = \max_{[0,T]}\left(|-2b+a'+\varepsilon+b^2|, |a'+2+a^2|, 2, a(0)\right),$$

independent of ε.

Proof. In what follows we shall write the equation in problem (4.7) as

$$L_\varepsilon[R_\varepsilon] = \varepsilon R_\varepsilon'' + aR_\varepsilon' + bR_\varepsilon = g\,. \tag{4.8}$$

Let us assume the energy constant m to be $m = a_0/2$. At first we estimate the energy to

$$\begin{aligned} W_\varepsilon &= \frac{a_0}{2}(R_\varepsilon^2 + \varepsilon(R_\varepsilon')^2) \\ &\le (a_0-\varepsilon)R_\varepsilon^2 + \varepsilon(a_0-\varepsilon)(R_\varepsilon')^2 \\ &\le 2W_\varepsilon + \varepsilon 2R_\varepsilon R_\varepsilon' \\ &\le a(R_\varepsilon^2 + \varepsilon(R_\varepsilon')^2) + \varepsilon 2R_\varepsilon R_\varepsilon' \\ &= \int_0^t \frac{\mathrm{d}}{\mathrm{d}s}\left[a(R_\varepsilon^2 + \varepsilon(R_\varepsilon')^2) + \varepsilon 2R_\varepsilon R_\varepsilon'\right]\,\mathrm{d}s + W_\varepsilon(0)\,. \end{aligned} \tag{4.9}$$

The idea is to wright the integrand as $h_1 W_\varepsilon + h_2$ such that we can find ε-independent bounds for h_1 and h_2. The application of Gronwall's Lemma will then lead us to the desired estimate. Let us call the integrand Q and perform the differentiation:

$$Q := \left[a(R_\varepsilon^2 + \varepsilon(R_\varepsilon')^2) + \varepsilon 2 R_\varepsilon R_\varepsilon'\right]' = a' R_\varepsilon^2 + 2aR_\varepsilon R_\varepsilon' + \varepsilon a'(R_\varepsilon')^2 + \varepsilon 2 a R_\varepsilon' R_\varepsilon'' + \varepsilon 2 (R_\varepsilon')^2 + \varepsilon 2 R_\varepsilon R_\varepsilon'' \,.$$

Now we want to express the terms with the help of the equation (4.8). In order to get the terms with the second derivatives we must multiply (4.8) with $2aR_\varepsilon'$ and with $2R_\varepsilon$ and add up the results:

$$\left.\begin{aligned} \varepsilon 2 a R_\varepsilon' R_\varepsilon'' + 2a^2 (R_\varepsilon')^2 + 2ab R_\varepsilon' R_\varepsilon &= 2aR_\varepsilon' g \\ \varepsilon 2 R_\varepsilon R_\varepsilon'' + a 2 R_\varepsilon R_\varepsilon' + 2b R_\varepsilon^2 &= 2R_\varepsilon g \end{aligned}\right\} +$$

$$\begin{aligned} \Rightarrow\ & \varepsilon 2 a R_\varepsilon' R_\varepsilon'' + \varepsilon 2 R_\varepsilon R_\varepsilon'' + a 2 R_\varepsilon R_\varepsilon' \\ & = -2a^2 (R_\varepsilon')^2 - 2ab R_\varepsilon' R_\varepsilon - 2b R_\varepsilon^2 + 2aR_\varepsilon' g + 2R_\varepsilon g \,, \end{aligned}$$

$$\begin{aligned} \Rightarrow\ & \varepsilon 2 a R_\varepsilon' R_\varepsilon'' + \varepsilon 2 R_\varepsilon R_\varepsilon'' + a 2 R_\varepsilon R_\varepsilon' \underbrace{+ a' R_\varepsilon^2 + \varepsilon a'(R_\varepsilon')^2 + \varepsilon 2 (R_\varepsilon')^2}_{\text{terms 1,3,5 of } Q} \\ & = -2a^2 (R_\varepsilon')^2 - 2ab R_\varepsilon' R_\varepsilon - 2b R_\varepsilon^2 + 2aR_\varepsilon' g + 2R_\varepsilon g \\ & \quad \underbrace{+ a' R_\varepsilon^2 + \varepsilon a'(R_\varepsilon')^2 + \varepsilon 2 (R_\varepsilon')^2}_{\text{terms 1,3,5 of } Q} \,, \end{aligned}$$

$$\Rightarrow\ Q = (-2b + a')R_\varepsilon^2 + (-2a^2 + \varepsilon a' + 2\varepsilon)(R_\varepsilon')^2 - 2ab R_\varepsilon' R_\varepsilon + 2aR_\varepsilon' g + 2R_\varepsilon g \quad (4.10)$$

In order to estimate the Q-term we use

$$\begin{aligned} 2aR_\varepsilon' g &\le a^2 (R_\varepsilon')^2 + g^2 \,, \\ 2R_\varepsilon g &\le R_\varepsilon^2 + g^2 \,, \\ -2ab R_\varepsilon' R_\varepsilon &\le b^2 R_\varepsilon^2 + 2a^2 (R_\varepsilon')^2 \,. \end{aligned}$$

Inserting this into (4.10) yields

$$Q \le (-2b + a' + \varepsilon + b^2) R_\varepsilon^2 + (\varepsilon a' + 2\varepsilon + \varepsilon a^2)(R_\varepsilon')^2 + 2g^2 \,.$$

Substituting this estimate for Q as well as $W_\varepsilon(0) = a(0)\alpha_1^2$ into the integral in (4.9) leads to

$$\frac{a_0}{2}(R_\varepsilon^2 + \varepsilon(R_\varepsilon')^2) \le M \int_0^t (R_\varepsilon^2 + \varepsilon(R_\varepsilon')^2)\,\mathrm{d}s + M \int_0^t g^2\,\mathrm{d}s + M\alpha_1^2 \,, \quad (4.11)$$

where

$$M := \max_{[0,T]} \left(|-2b + a' + \varepsilon + b^2|, |a' + 2 + a^2|, 2, a(0) \right) = O(1) \quad \text{in } [0,T] \,.$$

Applying the Gronwall Lemma 2 in (4.11) with $a = 2M/a_0$, $b = 0$ and $c = M(||g||^2_{[0,t]} + \alpha_1^2)$ (this is the L^2-norm of g) yields

$$\frac{a_0}{2}(R_\varepsilon^2 + \varepsilon(R_\varepsilon')^2) \leq M(||g||^2_{[0,t]} + \alpha_1^2)\,\mathrm{e}^{\frac{2M}{a_0}t} \leq M(||g||^2_{[0,T]} + \alpha_1^2)\,\mathrm{e}^{\frac{2M}{a_0}T}\,.$$

□

With the aid of the above Lemma we immediately obtain bounds for $|R_\varepsilon|$ and for $|R_\varepsilon'|$:

$$|R_\varepsilon(t)| \leq \sqrt{\frac{2M}{a_0}}(||g||_{[0,T]} + |\alpha_1|)\,\mathrm{e}^{\frac{M}{a_0}T} \qquad t \in [0,T]\,,$$

$$|R_\varepsilon'(t)| \leq \frac{1}{\sqrt{\varepsilon}}\sqrt{\frac{2M}{a_0}}(||g||_{[0,T]} + |\alpha_1|)\,\mathrm{e}^{\frac{M}{a_0}T} \qquad t \in [0,T]\,.$$

Using that $g = O(\varepsilon)$ and $\alpha_1 = O(\varepsilon)$ in $[0,T]$ we thus have

$$|R_\varepsilon(t)| + \sqrt{\varepsilon}|R_\varepsilon'(t)| = O(\varepsilon) \qquad \text{in } [0,T]\,.$$

To summarise this section we have

Theorem 3. *Let u_ε stand for the solution of*

$$P_\varepsilon \begin{cases} \varepsilon\dfrac{\mathrm{d}^2u_\varepsilon}{\mathrm{d}t^2} + a(t)\dfrac{\mathrm{d}u_\varepsilon}{\mathrm{d}t} + b(t)u_\varepsilon = f(t)\,, & t > 0\,, \\ u_\varepsilon(0) = \alpha\,, \qquad \dfrac{\mathrm{d}u_\varepsilon}{\mathrm{d}t}(0) = \beta\,, \end{cases}$$

where $\alpha, \beta \in \mathbb{R}$ are constants, independent of ε and

i) $a, b, f \in C^1([0,\infty))$,

ii) $a(t) \geq a_0 > 0$ for $t \geq 0$ with a_0 independent of ε.

Moreover, let w_0 and v denote the solutions of the modified reduced problem (4.2) and of the microscopic problem (4.5), respectively. Under these conditions one has for $\varepsilon \to 0$ the asymptotic approximations

$$u_\varepsilon - w_0 = O(\varepsilon) \qquad \text{in } [0,T]\,,$$

and

$$\frac{\mathrm{d}u_\varepsilon}{\mathrm{d}t} - \frac{\mathrm{d}w_0}{\mathrm{d}t} - \frac{\mathrm{d}v}{\mathrm{d}t} = O(\sqrt{\varepsilon}) \qquad \text{in } [0,T]\,,$$

where $T > 0$ is arbitrary but independent of ε. Moreover,

$$\frac{\mathrm{d}u_\varepsilon}{\mathrm{d}t} - \frac{\mathrm{d}w_0}{\mathrm{d}t} = O(\sqrt{\varepsilon}) \qquad \text{in } [\delta,T]\,,$$

where $\delta > 0$ is independent of ε.

Remarks:

- If $a(t)$ is strictly negative it can be shown that u_ε diverges as $\varepsilon \to 0$ and the whole construction breaks down. Moreover, if $a(t_0) = 0$ at some point t_0 then the reduced problem becomes singular at this point and complications arise. The point t_0 is called a *turning point*.
- The boundary layer term v has no influence on the point-wise asymptotic approximation of u_ε. The term is however necessary to obtain a point-wise AA of the derivative of u_ε to order $\sqrt{\varepsilon}$ in $[0, T]$. The function v is asymptotically smaller than any power of ε away from $t = 0$.

4.3. The boundary value problem

Consult the literature list given for this course.

5. Macroscopic limits of kinetic equations

5.1. Introduction

Some of the most complex phenomena in nature can be viewed as **systems of interacting particles**: the motion of gases, fluids, plasmas, cars in traffic or the behavior of animal herds, bird swarms and school of fish, for example. It is thus of fundamental interest to develop an understanding of these systems via mathematical models. Because the number of particles is usually extremely large, the "microscopic" (in the sense of atomistic) description in terms of the exact motion of every particle is not practicable, and moreover not necessary for studying large scale properties. Essential features of such systems can be revealed by inspecting statistical averages, which leads to "macroscopic models" in terms of densities in position space (x-space). The Euler equations and the Navier-Stokes equation are examples of macroscopic models. These equations contain coefficients like viscosity and heat diffusivity, which have to be derived from the atomistic picture. In order to have a sound macroscopic theory, in accordance with microscopic models, we should be able to formulate the micro-macro transition as a perturbation problem, where the solution of the micro problem converges to the solution of the macro problem as $\varepsilon \to 0$. The formulation and study of such limits is a quite formidable task and has concerned scientists for many decades, see [6] for a short review.

A fruitful approach has been the introduction of an intermediate, or "mesoscopic" level, the so-called kinetic description. In the kinetic picture the particle system is described in terms of densities in *phase space* (*position-velocity* or x-v-space). The central quantity is the "distribution function" $f(t, x, v)$, to be understood such that $f(t, x, v)\mathrm{d}x\mathrm{d}v$ is the density of particles at the point $(x, v) \in \mathbb{R}^6$ in phase space at time t. The governing equation for f is the famous **Boltzmann equation**:

$$\begin{aligned} &\frac{\partial f}{\partial t} + v \cdot \nabla_x f + F(x,t) \cdot \nabla_v f = Q(f)\,, \\ &f|_{t=0} = f_0(x, v)\,. \end{aligned} \tag{5.1}$$

Here, $F(x, t)$ stands for an external force and $Q(f)$ is the (nonlinear) collision operator. If long-range interactions between the particles are to be included, the external force is replaced by a mean-field interaction which depends on f. We will not concern ourselves with this additional complication in this course. The Boltzmann equation (BE) has been extensively studied [2, 3] and has gained renewed attention with the advent of semiconductor physics in the computer industry, where the mesoscopic description is

the best compromise between complexity and feasibility. Its main feature is a collision operator which accounts for the short-range interaction of the particles. The collisions have the effect that the entropy of the system is steadily decreasing - the statement of the famous **H-theorem** - and the dynamics are irreversible. Today, most of the macroscopic models can be understood as limits of the Boltzmann equation in the highly-collisional regime, so-called **fluid limits**. In what follows we shall give a brief overview of the (formal) techniques used in order to obtain such limits. A discussion about the derivation of the BE from the atomistic picture can be found in the above references.

5.2. The free transport equation

Before we start with the study of fluid limits of the BE, let us consider the free transport equation (FTE), obtained from the BE in the absence of collisions:

$$\begin{aligned} &\frac{\partial f}{\partial t} + v \cdot \nabla_x f + F(x,t) \cdot \nabla_v f = 0\,, \\ &f|_{t=0} = f_0(x,v)\,. \end{aligned} \tag{5.2}$$

We call the solution (X, V) of the ODEs

$$\begin{aligned} \frac{\mathrm{d}X}{\mathrm{d}s} &= V\,, & X(t) &= x\,, \\ \frac{\mathrm{d}V}{\mathrm{d}s} &= F(X,s)\,, & V(t) &= v\,, \end{aligned}$$

the "characteristics" of the FTE. Supposing that F is sufficiently smooth, let us say $F \in C^2$ with bounded derivatives up to second order, the Cauchy-Lipschitz theorem guarantees a solution $X(s;t,x,v)$, $V(s;t,x,v)$. The application

$$\Phi_t^s : (x,v) \mapsto (X(s;t,x,v), V(s;t,x,v))$$

defines the flow of the characteristic equations. It is now easy to show that the FTE has a unique solution of the form

$$f(t,x,v) = f_0(X(0;t,x,v), V(0;t,x,v))\,.$$

This follows from the fact that f is constant along the characteristics,

$$\frac{\mathrm{d}}{\mathrm{d}s} f(s, X(s), V(s)) = 0\,. \tag{5.3}$$

Indeed,

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}s} f(s, X(s), V(s)) &= \frac{\partial f}{\partial s} + \frac{\mathrm{d}X}{\mathrm{d}s} \cdot \nabla_x f + \frac{\mathrm{d}V}{\mathrm{d}s} \cdot \nabla_v f \\ &= \frac{\partial f}{\partial s} + V \cdot \nabla_x f + F(X,s) \cdot \nabla_v f\,. \end{aligned}$$

In view of the FTE (5.2) we obtain (5.3). Integration of (5.3) between 0 and t then yields

$$f(t, X(t;t,x,v), V(t;t,x,v)) = f(t,x,v) = f(0, X(0;t,x,v), V(0;t,x,v))\,.$$

The flow Φ_t^s has the group property

$$\Phi_{t_1}^{t_3} = \Phi_{t_2}^{t_3} \circ \Phi_{t_1}^{t_2}\,,$$

and the inverse of Φ_t^s is given by Φ_s^t. Hence the dynamics of the FTE are reversible. Moreover, the flow Φ_t^s conserves phase space volume. This means that for any volume $\Omega(t)$ that is advected by the flow,

$$\Omega(t) = \{(x,v) \in \mathbb{R}^6 :\ (x,v) = \Phi_0^t(x_0,v_0) \text{ with } (x_0,v_0) \in \Omega_0 \subset \mathbb{R}^6\}\,,$$

we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\Omega(t)} \mathrm{d}^3x\,\mathrm{d}^3v = \frac{\mathrm{d}}{\mathrm{d}t}\int_{\Omega_0} \left|\det \frac{\partial \Phi_0^t}{\partial(x,v)}\right| \mathrm{d}^3x_0\,\mathrm{d}^3v_0 = 0\,.$$

This follows Liouville's theorem

$$\frac{\mathrm{d}}{\mathrm{d}t}\det\frac{\partial \Phi_0^t}{\partial(x,v)} = 0\,,$$

related to the fact the vector field $(v, F(x,t))$ of the characteristics is divergence-free.

5.3. Properties of the collision operator

The collision operator $Q(f)$ in the BE (5.1) is a nonlinear integral operator and hard to deal with in actual applications. However, there are some essential properties of Q which can be distilled to write down simpler collision operators with the same features. This approach is sufficient for understanding fluid limits. The are three essential properties of Q:

1. **Conservation of mass, momentum and energy**:

$$\int Q(f) \begin{pmatrix} 1 \\ v \\ |v|^2 \end{pmatrix} \mathrm{d}^3v = 0\,,$$

   Here and in what follows, if no boundaries are indicated with an integral sign, the integration is performed over the whole $\mathbb{R}^3$.

2. The equilibria are **Maxwellian functions**:

$$Q(f) = 0 \qquad \Leftrightarrow \qquad f = \mathcal{M} := \frac{n}{(2\pi T)^{d/2}} \exp\left[-\frac{|v-u|^2}{2T}\right]\,, \tag{5.4}$$

   where $n, T \in \mathbb{R}_+$ (strictly positive) and $u \in \mathbb{R}^3$ are arbitrary and may depend on (t,x), and d stands for the dimension of the position space, usually $1 \le d \le 3$.

3. **H-theorem**: the entropy $S(t)$ is decreasing with time,

$$\frac{\mathrm{d}}{\mathrm{d}t}S(t) = \frac{\mathrm{d}}{\mathrm{d}t}\int\int f\ln(f)\,\mathrm{d}^3x\,\mathrm{d}^3v \leq 0\,,$$

and

$$\frac{\mathrm{d}}{\mathrm{d}t}S(t) = 0 \qquad \Leftrightarrow \qquad f = \mathcal{M}\,.$$

There are two well-known simplified forms of Q that feature all the essential properties:

- The BGK (Bhatnagar–Gross–Krook) operator:

$$Q_{\mathrm{BGK}}(f) = \nu(\mathcal{M}_f - f)\,.$$

Here, $\nu > 0$, independent of v, stands for the collision frequency and the Maxwellian $\mathcal{M}_f$ is related to f via the velocity moments, hence in (5.4) we choose n, u, T as

$$n = \int f\,\mathrm{d}^3v\,, \qquad u = \frac{1}{n}\int v\,f\,\mathrm{d}^3v\,, \qquad T = \frac{1}{nd}\int |v-u|^2\,f\,\mathrm{d}^3v\,. \tag{5.5}$$

Since the Maxwellians $\mathcal{M}$ in (5.4) are Gaussians in v we have the following integrals:

$$\int \mathcal{M}\begin{pmatrix}1\\ v\\ |v|^2\end{pmatrix}\mathrm{d}^3v = \begin{pmatrix}n\\ nu\\ nT\cdot d\end{pmatrix}\,.$$

For the Maxwellian $\mathcal{M}_f$, from (5.5) it follows that

$$\int (\mathcal{M}_f - f)\begin{pmatrix}1\\ v\\ |v|^2\end{pmatrix}\mathrm{d}^3v = 0\,, \tag{5.6}$$

such that conservation of mass, momentum and energy is guaranteed by the BGK operator. The identification of the equilibria is trivial. In order to see that the H-theorem is verified, let us multiply the BE (5.1) by $\ln(f)$ and integrate over x and v to obtain

$$\frac{\partial}{\partial t}\int\int f\ln(f)\,\mathrm{d}^3x\,\mathrm{d}^3v = \nu\int\int(\mathcal{M}_f - f)\ln(f)\,\mathrm{d}^3x\,\mathrm{d}^3v\,.$$

Here the transport terms vanish due to the fact that we demand

$$\lim_{|x|\to\infty} f(t,x,v) = 0\,, \qquad \lim_{|v|\to\infty} f(t,x,v) = 0\,. \tag{5.7}$$

Substituting on the left-hand-side the definition of the entropy S we can write

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \nu\int\int(\mathcal{M}_f - f)\ln\left(\frac{f}{\mathcal{M}_f}\right)\mathrm{d}^3x\,\mathrm{d}^3v + \nu\int\int(\mathcal{M}_f - f)\ln(\mathcal{M}_f)\,\mathrm{d}^3x\,\mathrm{d}^3v\,. \tag{5.8}$$

We can write the logarithm of the Maxwellian as

$$\ln(\mathcal{M}_f) = \ln\left(\frac{n}{(2\pi T)^{d/2}}\right) - \frac{|v-u|^2}{2T} = a + b\cdot v + c|v|^2\,.$$

Then, using (5.6) we obtain that the second term on the right-hand-side of (5.8) is zero. Moreover, note that that $(x-y)(\ln(y)-\ln(x))$ is negative for all values $x \neq y$ and zero only for $x = y$, because the functions $f(x) = x$ and $f(x) = \ln(x)$ are strictly increasing. It follows that

$$\frac{\mathrm{d}S}{\mathrm{d}t} \leq 0\,, \qquad \frac{\mathrm{d}S}{\mathrm{d}t} = 0 \quad \Leftrightarrow \quad f = \mathcal{M}_f\,.$$

- The Fokker-Planck-Landau operator:

$$Q_{\mathrm{FPL}}(f) = \nu\,\nabla_v \cdot \Big[(v-u)f + T\,\nabla_v f\Big]\,.$$

Here, u and T are defined from f as in (5.5). It is straightforward to verify the conservation of mass, momentum and energy (use integration by parts in v). Moreover,

$$\nabla_v \mathcal{M}_f = -\frac{(v-u)}{T}\mathcal{M}_f$$

implies that $Q_{\mathrm{FPL}}(\mathcal{M}_f) = 0$. On the other hand,

$$Q_{\mathrm{FPL}}(f) = 0 \quad \Rightarrow \quad (v-u)f + T\,\nabla_v f = 0 \quad \Rightarrow \quad \nabla_v \ln(f) = -\frac{(v-u)}{T}\,,$$

which implies that $f = \mathcal{M}$. With regards to the H-theorem we have

$$\begin{aligned}
\frac{\mathrm{d}S}{\mathrm{d}t} &= \nu \int\int \nabla_v \cdot \Big[(v-u)f + T\,\nabla_v f\Big] \ln(f)\,\mathrm{d}^3x\mathrm{d}^3v \\
&= \nu \int\int \nabla_v \cdot \Big[(v-u)f + T\,\nabla_v f\Big] \ln\left(\frac{f}{\mathcal{M}_f}\right) \mathrm{d}^3x\mathrm{d}^3v \\
&= -\nu \int\int \Big[(v-u)f + T\,\nabla_v f\Big] \frac{\mathcal{M}_f}{f}\Big(\frac{\nabla_v f}{\mathcal{M}_f} - \frac{f}{\mathcal{M}_f^2}\nabla_v \mathcal{M}_f\Big)\,\mathrm{d}^3x \\
&= -\nu \int\int \frac{1}{Tf}\Big[(v-u)f + T\,\nabla_v f\Big]^2\,\mathrm{d}^3x\mathrm{d}^3v \\
&\leq 0\,.
\end{aligned}$$

Moreover, we see that $\frac{\mathrm{d}}{\mathrm{d}t}S = 0$ if and only if $(v-u)f + T\,\nabla_v f = 0$, which leads to the Maxwellian $\mathcal{M}_f$.

5.4. Moment equations

The velocity moments of the distribution function f are defined as

$$\begin{pmatrix} n(x,t) \\ nu(x,t) \\ w(x,t) \end{pmatrix} := \int \begin{pmatrix} 1 \\ v \\ |v^2|/2 \end{pmatrix} f(x,v,t)\,\mathrm{d}^3v\,.$$

Here, n is called the mass (density), nu stands for the momentum (density) and w denotes the energy (density). These quantities will be called "fluid variables" in the following. As an auxiliary quantity, the temperature T is defined via

$$\frac{d}{2}nT(x,t) := \frac{1}{2}\int |v-u(x,t)|^2 f(x,v,t)\,\mathrm{d}^3v\,,$$

where d is the dimension of the position space, here $d=3$. It follows that the energy can be expressed as

$$w = \frac{d}{2}nT + \frac{1}{2}n|u|^2\,,$$

which can be interpreted as the sum of an internal and a kinetic energy. The equations of motion for the fluid variables follow directly by taking moments of the Boltzmann equation:

$$\frac{\partial}{\partial t}\begin{pmatrix} n \\ nu \\ w \end{pmatrix} + \nabla_x \cdot \begin{pmatrix} nu \\ nu\otimes u + \mathbb{P} \\ wu + \mathbb{P}\cdot u + Q \end{pmatrix} = \begin{pmatrix} 0 \\ nF \\ nuF \end{pmatrix}\,.$$

Here, $(u\otimes u)_{i,j} = u_i u_j$ denotes the tensor product of two vectors, $\mathbb{P}$ stands for the pressure tensor and Q is the heat flux, defined by

$$\mathbb{P}(x,t) := \int [v-u(x,t)]\otimes[v-u(x,t)]\, f(x,v,t)\,\mathrm{d}^3v\,,$$

$$Q(x,t) := \frac{1}{2}\int |v-u(x,t)|^2[v-u(x,t)]\, f(x,v,t)\,\mathrm{d}^3v\,.$$

The moment equations are thus not closed, i.e. they still depend on the full distribution function f via the pressure tensor and the heat flux. Asymptotic closure via perturbation theory is about the computation of closure relations for $\mathbb{P}$ and Q, thereby expressing them in terms of the fluid variables n, nu and w. In order to obtain hydrodynamic equations, the perturbation problem consists of finding asymptotic expansions of f_ε as $\varepsilon\to 0$, solution of the BE

$$P_\varepsilon \begin{cases} \dfrac{\partial f_\varepsilon}{\partial t} + v\cdot\nabla_x f_\varepsilon + F(x,t)\cdot\nabla_v f_\varepsilon = \dfrac{Q(f)}{\varepsilon}\,, \\ f_\varepsilon|_{t=0} = f_{\varepsilon,\mathrm{in}}(x,v)\,, \end{cases}$$

with the boundary conditions (5.7). For the collision operator one usually chooses either the BGK or the Fokker-Planck-Landau operator discussed above. Using a Poincaré type ansatz (often called Hilbert ansatz in this context) of the form

$$f_\varepsilon = f_0 + \varepsilon f_1 + \dots$$

and using the properties of the collision operator permits to compute the coefficient functions f_0, f_1, etc. Because collisions are of order $O(1/\varepsilon)$, they dominate the other terms and the limit distribution $\lim_{\varepsilon\to 0} f_\varepsilon = f_0$ will be a Maxwellian. The equations satisfied by the moments of f_0 are the well-known Euler equations of gas dynamics. Moreover, the equations for moments of $f_0 + \varepsilon f_1$ turn out to be the compressible Navier-Stokes equations (see Hausarbeit - Topic 2).

A. Fundamentals of initial-value problems

We present here some basic notions and theorems about systems of first-order ordinary differential equations with initial conditions. For an in-depth reading on IVPswe recommend the books [9, 11, 12].

A.1. Problem setting

In what follows $J \subset \mathbb{R}$ denotes an interval on the real line and and $\Omega \subset \mathbb{R}^n$ stands for an open and connected subset of $\mathbb{R}^n$. We shall be concerned with systems of ODEs of the form

$$\frac{\mathrm{d}\mathbf{z}(t)}{\mathrm{d}t} = f(\mathbf{z}(t), t)\,, \tag{A.1}$$

where $f : \Omega \times J \to \mathbb{R}^n$ is a suitable regular function. The image of f is called the *direction field* of the system (A.1). A C^1-curve $\mathbf{z} : J \to \mathbb{R}^n$ satisfying (A.1) is tangent everywhere to the direction field. In case that f is independent of t, the system (A.1) is called *autonomous*. For autonomous systems, we call a function $E : \Omega \to \mathbb{R}$ a *first integral* if $f(\mathbf{z}) \cdot \nabla E(\mathbf{z}) = 0 \;\forall\, \mathbf{z} \in \Omega$. It is easy to see that

$$\frac{\mathrm{d}}{\mathrm{d}t} E(\mathbf{z}(t)) = 0 \quad \Longleftrightarrow \quad E \text{ is a first integral}\,.$$

The level sets of E are hypersurfaces in Ω. In case that E is a first integral, $E(\mathbf{z}(t)) = c$, the motion lies in the corresponding hypersurface.

If (A.1) is furnished with a condition $\mathbf{z}(t_0) = V \in \Omega$ for some $t_0 \in J$ we call it an *initial value problem* (IVP):

$$\begin{cases} \dfrac{\mathrm{d}\mathbf{z}(t)}{\mathrm{d}t} = f(\mathbf{z}(t), t)\,, \\ \mathbf{z}(t_0) = V\,. \end{cases} \tag{A.2}$$

A function $\mathbf{z}$ is a solution of (A.2) if and only if it satisfies the integral equation

$$\mathbf{z}(t) = V + \int_{t_0}^{t} f(\mathbf{z}(s), s)\, ds\,. \tag{A.3}$$

Hence, (A.2) and (A.3) are equivalent formulations of the IVP. By a solution of (A.2) we mean a C^1-curve $\mathbf{z} : I \to \Omega$ on some interval $I \subset J$, where $t_0 \in I$, $\mathbf{z}(t_0) = V$ and

(A.2) holds for all $t \in I$. We remark that a solution may be defined only on $I \subset J$, even if the IVP is defined on J. The *graph* of a solution is the set

$$\gamma_{\mathbf{z}} := \{(t, \mathbf{z}(t)) \in J \times \Omega : \ t \in I\} . \tag{A.4}$$

A *proper extension* of $\mathbf{z} : I \to \Omega$, solution of (A.2), is a function $\tilde{\mathbf{z}} : \tilde{I} \to \Omega$, where $I \subset \tilde{I} \subset J$, $\tilde{I} \neq I$ and $\tilde{\mathbf{z}}(t) = \mathbf{z}(t)$ for $t \in I$. A *maximal solution* of the IVP is a solution with no proper extension. The corresponding interval $\tilde{I}$ is called the *maximal interval*.

From the integral formulation (A.3) of the IVP we are inclined to search for solutions with lesser regularity, namely $\mathbf{z} \in C^0(I, \Omega)$, such that the integrand on the right-hand-side of (A.3) is a piecewise continuous function of s. The integral solution than satisfies (A.2) at points of continuity of $t \mapsto f(\mathbf{z}(t), t)$ due to the fundamental theorem of calculus. Such solutions are important in case where the function f is not continuous with respect to its second argument (sometimes arising in control problems).

A.2. Existence and uniqueness

Existence and uniqueness of solutions to the IVP (A.2) depend on the properties of the function f. Peano's existence theorem states that there is at least one solution of (A.2) on some interval $I \subset J$ if f is continuous on $\Omega \times J$. For uniqueness one needs an additional property: f must be *locally Lipschitz* with respect to its first argument, which means there exist neighborhoods $\Omega_{\mathbf{z}} \subset \Omega$ of $\mathbf{z}$ and $J_t \subset J$ of t such that

$$\frac{||f(\mathbf{z}_1, s) - f(\mathbf{z}_2, s)||}{||\mathbf{z}_1 - \mathbf{z}_2||} \leq L < \infty \qquad \forall\ \mathbf{z}_1, \mathbf{z}_2 \in \Omega_{\mathbf{z}},\ s \in J_t .$$

The constant L, which possibly depends on $\Omega_{\mathbf{z}}$ and J_t, but not on $\mathbf{z}_1, \mathbf{z}_2$, is called the *Lipschitz constant*. Remark that the limit $||\mathbf{z}_1 - \mathbf{z}_2|| \to 0$ in the above expression is finite.

Example 13. Let us view the function $f(z) = z^{1/3}$ as a mapping $f : \mathbb{R} \to \mathbb{R}$. f is continuous but not locally Lipschitz. Consider the point $z = 0$. Since $f(-1) = -1$, one has $\lim_{\varepsilon \to 0} |f(\varepsilon) - f(-\varepsilon)|/(2\varepsilon) = \lim_{\varepsilon \to 0} \varepsilon^{-2/3} = \infty$, which means that f is not locally Lipschitz at $z = 0$. Moreover, let $J = \mathbb{R} = \Omega$ and consider the scalar IVP

$$\frac{\mathrm{d}z(t)}{\mathrm{d}t} = t\, z^{1/3}(t) , \qquad z(0) = 0 . \tag{A.5}$$

We readily verify that $z = 0$ on $\mathbb{R}$ is a maximal solution, but so is the continuously differentiable function

$$z(t) = \begin{cases} (t/\sqrt{3})^3, & t \geq 0 \\ 0 \qquad\quad , & t < 0 \end{cases} .$$

Hence, there are at least two maximal solutions to the IVP (A.5). The cause for this is that $f(z) = z^{1/3}$ is not locally Lipschitz at $z = 0$.

In what follows we will call

Assumption A: f is continuous on $\Omega \times J$ and locally Lipschitz with respect to its first argument, and $t \mapsto f(\mathbf{z}(t), t)$ is jointly continuous for continuous $\mathbf{z}$.

We can now state the basic existence and uniqueness theorem:

Theorem 4. *Under the assumption* A, *the IVP* (A.2) *has a unique maximal solution for each* $(V, t_0) \in \Omega \times J$. *The associated maximal interval is denoted* $I(V, t_0) \subset J$.

This theorem is known as the *Cauchy-Lipschitz* theorem or the *Picard-Lindelöf* theorem. There are two different approaches to the proof, depending on whether one uses the continuity of f with respect to the first argument (Gronwall's lemma, c.f. Theorem 4.17 in [9] for example), or a fixed-point technique (c.f. Theorem 4.22 in [9] for example).

A.3. Transition map and local flow

Let us now introduce the notion of the *transition map* $\boldsymbol{\psi}$ of the IVP (A.2). The transition map is called the *local flow* in case that the IVP is autonomous. We define the map $\boldsymbol{\psi} : \mathrm{dom}(\boldsymbol{\psi}) \subset J \times \Omega \times J \to \Omega$ by the property that $t \mapsto \boldsymbol{\psi}(t; V, t_0)$ is the solution of the IVP (A.2). Hence the transition map can be viewed as the solution of the IVP with dependence on the initial condition. The domain of $\boldsymbol{\psi}$ is

$$\mathrm{dom}(\boldsymbol{\psi}) = \{(t, V, t_0) \in J \times \Omega \times J : \ t \in I(V, t_0)\}\,.$$

Definition 11. The system (A.2) is *autonomous* if $J = \mathbb{R}$ and if f does not depend on t.

For an interval $I \subset \mathbb{R}$ let us denote the interval shifted by t_0 via

$$I + t_0 := \{t + t_0 \in \mathbb{R} : \ t \in I\}\,.$$

We then have the following Corollary to the above theorem:

Corollary 5. *Let* $f : \Omega \times J \to \mathbb{R}^n$ *satisfy assumption* A.

1. *Let* $(V, t_0) \in \Omega \times J$ *and let* $s \in I(V, t_0)$. *Then* $I(\boldsymbol{\psi}(s; V, t_0), s) = I(V, t_0)$ *and*

$$\boldsymbol{\psi}(t; \boldsymbol{\psi}(s; V, t_0), s) = \boldsymbol{\psi}(t; V, t_0) \qquad \forall\ t \in I(V, t_0)\,. \tag{A.6}$$

2. *Assume that the system is autonomous. Then, for arbitrary* $t_0, s \in \mathbb{R}$ *and* $V \in \Omega$, $I(V, t_0) = I(V, s) - s + t_0$ *and*

$$\boldsymbol{\psi}(t + s - t_0; V, s) = \boldsymbol{\psi}(t; V, t_0) \qquad \forall\ t \in I(V, t_0)\,. \tag{A.7}$$

Proof. 1. The curve $\mathbf{z} : t \mapsto \boldsymbol{\psi}(t; \boldsymbol{\psi}(s; t_0, V), s)$ is the unique solution of the IVP

$$\frac{\mathrm{d}\mathbf{z}(t)}{\mathrm{d}t} = f(\mathbf{z}(t), t)\,, \qquad \mathbf{z}(s) = \boldsymbol{\psi}(s; V, t_0)\,. \tag{A.8}$$

However, the curve $\tilde{\mathbf{z}} : t \mapsto \boldsymbol{\psi}(t; V, t_0)$ clearly satisfies (A.8) too, which according to Theorem 4 means $\mathbf{z}(t) = \tilde{\mathbf{z}}(t)$ for all $t \in I(\boldsymbol{\psi}(s; V, t_0), s)$ and the interval is maximal. Moreover, since $\tilde{\mathbf{z}}$ is the unique solution of

$$\frac{\mathrm{d}\tilde{\mathbf{z}}(t)}{\mathrm{d}t} = f(\tilde{\mathbf{z}}(t), t)\,, \qquad \tilde{\mathbf{z}}(t_0) = V\,,$$

we obtain $I(V, t_0) = I(s, \boldsymbol{\psi}(s, t_0, V))$.

2. The curve $\mathbf{z} : t \mapsto \boldsymbol{\psi}(t; V, s)$ is the unique maximal solution of

$$\frac{\mathrm{d}\mathbf{z}(t)}{\mathrm{d}t} = f(\mathbf{z}(t))\,, \qquad \mathbf{z}(s) = V\,,$$

defined on the interval $I(V, s)$. The curve $\tilde{\mathbf{z}}(t) := \mathbf{z}(t + s - t_0)$ is thus defined on $I(V, s) - s + t_0$. It is easily verified that $\tilde{\mathbf{z}}$ satisfies

$$\frac{\mathrm{d}\tilde{\mathbf{z}}(t)}{\mathrm{d}t} = f(\tilde{\mathbf{z}}(t))\,, \qquad \tilde{\mathbf{z}}(t_0) = V\,,$$

and is hence defined on the maximal interval $I(V, t_0)$. □

The property (A.6) is the *semigroup property* of the transition map; it leads to an intuitive interpretation of $\boldsymbol{\psi}$ as a dynamical propagator. The property (A.7) mirrors the *translational invariance* of autonomous systems, that is a translation (with respect to time) of a solution is also a solution. We have another Corollary:

Corollary 6. *Let $f : \Omega \times J \to \mathbb{R}^n$ satisfy assumption* A. *Then the relation on $\Omega \times J$ defined by*

$$(V, t_0) \sim (\mathbf{D}, s) \quad \text{if} \quad s \in I(V, t_0) \quad \text{and} \quad \mathbf{D} = \boldsymbol{\psi}(s; V, t_0)$$

is an equivalence relation, and the graphs of solutions to initial conditions $(V, t_0) \in \Omega \times J$ are the equivalence classes.

Proof. We need to prove the three properties defining an equivalence relation:

1. Reflexivity: Clearly $t_0 \in I(V, t_0)$ and $\boldsymbol{\psi}(t_0; V, t_0) = V$ such that $(V, t_0) \sim (V, t_0)$.
2. Symmetry: Assume that $(V, t_0) \sim (\mathbf{D}, s)$, hence

   $$I(\mathbf{D}, s) = I(\boldsymbol{\psi}(s; V, t_0), s) = I(V, t_0) \quad \Longrightarrow \quad t_0 \in I(\mathbf{D}, s)\,,$$

   where we used Corollary 5 in the second equality. Moreover,

   $$\boldsymbol{\psi}(t_0; \mathbf{D}, s) = \boldsymbol{\psi}(t_0; \boldsymbol{\psi}(s; V, t_0), s) = \boldsymbol{\psi}(t_0; V, t_0) = V\,.$$

   It follows that $(\mathbf{D}, s) \sim (V, t_0)$, which proves symmetry.

3. Transitivity: Suppose that $(V, t_0) \sim (\mathbf{D}, s)$ and $(\mathbf{D}, s) \sim (\mathbf{E}, r)$, then from the above it follows that

$$\left.\begin{aligned} r \in I(\mathbf{D}, s) = I(V, t_0) \\ \mathbf{E} = \boldsymbol{\psi}(r; \mathbf{D}, s) = \boldsymbol{\psi}(r; \boldsymbol{\psi}(s; V, t_0), s) = \boldsymbol{\psi}(r; V, t_0) \end{aligned}\right\} \implies (V, t_0) \sim (\mathbf{E}, r)\,,$$

which proves transitivity.

By definition, if $(V, t_0) \sim (\mathbf{D}, s)$, the point $(\mathbf{D}, s) \in \Omega \times J$ lies on the graph of the solution $\mathbf{z}$ with initial condition $\mathbf{z}(t_0) = V$, so that the graphs form the equivalence classes. □

An equivalence relation partitions a set into pairwise disjoint subsets (the equivalence classes), whose union forms the whole set. Therefore, the above Corollary states that the graphs of two solutions with initial conditions at (V, t_0) and at $(\mathbf{D}, s)$ are either identical or disjoint, i.e. different graphs do not intersect (cross).

A.4. Autonomous systems

Let us now focus on autonomous systems,

$$\frac{\mathrm{d}\mathbf{z}(t)}{\mathrm{d}t} = f(\mathbf{z}(t))\,, \qquad \mathbf{z}(0) = V\,, \tag{A.9}$$

where f is locally Lipschitz on Ω. We may set the initial condition at $t_0 = 0$ in full generality because of the translation invariance of solutions, see the second part of Corollary 5. A solution is a C^1-curve $\mathbf{z} : I \to \Omega$ on some interval I containing 0, such that (A.9) holds for $t \in I$. Of course Theorem 4 applies in this case and gives existence and uniqueness of a maximally extended solution on a maximal interval $I_V = I(V, 0)$. Therefore, we may define the *local flow* $\boldsymbol{\varphi}$ of the system (A.9) as the map $\boldsymbol{\varphi} : (t, V) \mapsto \boldsymbol{\psi}(t; V, 0)$. The domain of the flow is

$$\mathrm{dom}(\boldsymbol{\varphi}) = \{(t, V) \in \mathbb{R} \times \Omega :\ t \in I_V\}\,.$$

In particular, $t \mapsto \boldsymbol{\varphi}(t, V)$ is the solution of the IVP (A.9). If $I_V = \mathbb{R}$ for some V, then the solution $t \mapsto \boldsymbol{\varphi}(t, V)$ is said to be *global*. If $I_V = \mathbb{R}$ for all $V \in \Omega$, then $\boldsymbol{\varphi}$ is called simply the *flow* of (A.9), and $\boldsymbol{\varphi}$ is said to be a *dynamical system*. The local flow satisfies

$$\begin{aligned} &\boldsymbol{\varphi}(0, V) = V \qquad \forall\ V \in \Omega\,, \\ \text{and for } s \in I_V :\quad &\boldsymbol{\varphi}(t + s, V) = \boldsymbol{\varphi}(t, \boldsymbol{\varphi}(s, V)) \quad \forall\ t \in I_V - s\,. \end{aligned} \tag{A.10a}$$

The first relation follows from the definition of $\boldsymbol{\varphi}$. To prove the second one we evoke the second result of Corollary 5 with $t_0 = 0$:

$$I_{\boldsymbol{\varphi}(s,V)} = I(\boldsymbol{\psi}(s; V, 0), 0) = I(\boldsymbol{\psi}(s; V, 0), s) - s = I(V, 0) - s = I_V - s\,.$$

Moreover, for $t \in I_V - s$ we get

$$\boldsymbol{\varphi}(t, \boldsymbol{\varphi}(s, V)) = \boldsymbol{\psi}(t; \boldsymbol{\psi}(s; V, 0), 0) = \boldsymbol{\psi}(t+s; \boldsymbol{\psi}(s; V, 0), s) = \boldsymbol{\psi}(t+s; V, 0) = \boldsymbol{\varphi}(t+s, V)\,.$$

Relation (A.10a) is termed the *group property* of the local flow. The terminology becomes clear when we regard the case where $I_V = \mathbb{R}$ for all $V \in \Omega$, so that $\boldsymbol{\varphi}$ is a dynamical system. The family of mappings $\Phi_t : V \mapsto \boldsymbol{\varphi}(t, V)$, $t \in \mathbb{R}$, forms a commutative group due to $\Phi_s \circ \Phi_t = \Phi_{s+t}$ for all $s, t \in \mathbb{R}$.

For autonomous systems, the space $\Omega \subset \mathbb{R}^n$ is called the *phase space*. In addition to the graph of the solution $\mathbf{z}$ defined in (A.4), for autonomous systems one calls the image of $\mathbf{z}$ an *orbit* (or *trajectory*) of the system. We have seen for general ODE systems that different graphs do not intersect. In the autonomous case, even more is true: different orbits do not intersect. This is because the relation $V \sim \tilde{V}$ if $\tilde{V}$ is in the orbit of V, is an equivalence relation on Ω, similar to Corollary 6. The orbits are thus equivalence classes.

Index

Bibliography

[1] A.J. Brizard and T.S. Hahm. Foundations of nonlinear gyrokinetic theory. *Rev. Mod. Phys.*, 79:421, 2007.

[2] Carlo Cercignani. *The Boltzmann Equation and Its Applications*, volume 67. Applied Mathematical Sciences, 1988.

[3] Carlo Cercignani, Mario Pulvirenti, and Reinhard Illner. *The Mathematical Theory of Dilute Gases*, volume 106. Applied Mathematical Sciences, 1994.

[4] Eduardus Marie de Jager and JF Furu. *The theory of singular perturbations*, volume 42. Elsevier, 1996.

[5] Pierre Degond. Asymptotic-preserving schemes for fluid models of plasmas. *arXiv preprint arXiv:1104.1869*, 2011.

[6] Pierre Degond, Lorenzo Pareschi, and Giovanni Russo. *Modeling and Computational Methods for Kinetic Equations* . Modeling and Simulation in Science, Engineering and Technology. Springer Science+Business Media, LLC, 2004.

[7] Wiktor Eckhaus. *Asymptotic analysis of singular perturbations*, volume 9. Elsevier, 2011.

[8] Shi Jin. Efficient asymptotic-preserving (AP) schemes for some multiscale kinetic equations. *SIAM Journal on Scientific Computing*, 21(2):441–454, 1999.

[9] H. Logemann and E.P. Ryan. *Ordinary Differential Equations: Analysis, Qualitative Theory and Control*. Undergraduate Mathematics Series. Springer, 2014.

[10] Stefan Possanner. Gyrokinetics from variational averaging: existence and error bounds. *arXiv preprint arXiv:1711.09620*, November 2017.

[11] G. Teschl. *Ordinary Differential Equations and Dynamical Systems*, volume 140 of *Graduate Studies in Mathematics*. AMS, 2012.

[12] W. Walter. *Ordinary Differential Equations*. Number 182 in Graduate Texts in Mathematics. Springer, 1991.