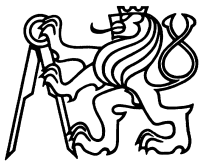
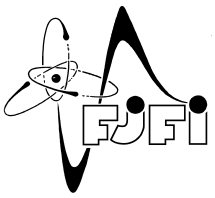


Czech Technical University in Prague
Faculty of Nuclear Sciences and Physical Engineering



Qualitative analysis of a reaction-diffusion system using weakly non-linear analysis and the WKBJ method

Kvalitativní analýza systému reakčně-difuzních rovnic pomocí slabě nelineární analýzy a WKBJ metody

Master's thesis

Author: **Bc. Juraj Kováč**

Supervisor: **doc. Ing. Václav Klika, Ph.D.**

Academic year: 2019/2020

- Zadání práce -

- Zadání práce (zadní strana) -

Acknowledgments:
I would like to express my gratitude towards the supervisor of this work, doc. Ing. Václav Klika, Ph.D., first and foremost for his patience and attentive attitude, even in the most extraordinary of circumstances, for his meaningful insights, professional approach as well as for providing me with abundance of useful sources and literature. I also wish to thank my family for a warm environment and generous support, to Róbert Babjak for his ever-helpful friendship and to God for granting me this opportunity.

Čestné prehlásenie/Declaration:
Prehlasujem, že som túto bakalársku prácu vypracoval samostatne a uviedol som všetku použitú literatúru.

Nemám závažný dôvod proti použitiu tohto diela v zmysel § 60 Zákona č. 121/2000 Sb., o práve autorskom, o právach súvisiacich s právom autorským a o zmene niektorých zákonov.

I hereby declare that I carried out this work independently and using exclusively the cited sources.

V Prahe dňa/ Prague August 18th 2020

Název práce:

Kvalitativní analýza systému reakčně-difuzních rovnic pomocí slabě nelineární analýzy a WKBJ metody

Autor: Juraj Kováč

Obor: Matematická fyzika

Druh práce: Diplomová práce

Vedoucí práce: doc. Ing. Václav Klika, Ph.D., Katedra matematiky, Fakulta jaderná a fyzikálně inženýrská, České vysoké učení technické v Praze

Abstrakt: Predstavíme základné koncepty potrebné pre hlbšie štúdium javu difúziou poháňanej (Turingovskej) nestability. Vyšetrujeme vlastnosti tohto modelu ako v okolí rovnovážnych stavov či bifurkačnch bodov, tak aj v asymptotických prípadoch. Predstavíme a predvedieme sadu nástrojov slabo nelineárnej analýzy potrebných na štúdium mechanizmov určujúcich vzor. Pokúsime sa o dôkaz aproximačného teorému pre WKBJ aproximáciu riešenia systému reakčno-difúznych rovníc.

Klíčová slova: asymptotické metódy, bifurkačná analýza, difúziou poháňaná nestabilita, WKBJ metóda, Turingov model

Title:

Qualitative analysis of a reaction-diffusion system using weakly non-linear analysis and the WKBJ method

Author: Juraj Kováč

Branch of study: Mathematical physics

Type of work: Master's thesis

Supervisor: doc. Ing. Václav Klika, Ph.D., Department of mathematics, Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague

Abstract: We will introduce basic concepts required for an advanced analysis of the phenomenon of diffusion-driven (Turing) instability. We investigate the properties of this model in the vicinity of equilibria and bifurcation points as well as in asymptotic settings. We introduce and demonstrate a set of methods of the weakly non-linear analysis relevant for studying mechanisms of pattern selection. Finally, we attempt to prove an approximation theorem for a WKBJ solution to a system of reaction-diffusion equations.

Key words: asymptotic methods, bifurcation analysis, diffusion-driven instability, the WKBJ method, Turing model

Contents

Introduction

Ever since Alan Turing's seminal paper [16] was released in the early 1950s, reaction-diffusion models have been the subject of a vast amount of research, both due to their attractive conceptual simplicity and their broad applicability, ranging from mathematical models of processes and phenomena in developmental biology and chemistry all the way to ecological and environmental models for animal populations and large-scale interactions. One of the flagship features of these models is the possibility of the transition from homogeneity to heterogeneity and from stability to instability resulting from random perturbations and spread via diffusion.

Whether in spite of all the research already undertaken and all the effort made, or due to it, the challenges still lying ahead of scientists and researchers are numerous, both of theoretical and experimental nature. The effect of domain growth, spatial heterogeneity translated into diffusion coefficients or reaction kinetics, [14] every new bit of additional complexity seems to present a new opportunity for adopting new approaches and achieving new results as well as for experimental confirmation or rejection. The fundamental questions about the underlying mechanisms and governing principles present yet another separate area of research possibilities.

In this work we largely set our focus on an analysis of the theoretical aspects of reaction-diffusion models as well as associated mathematical concepts useful for this analysis. The first two chapters concentrate on investigating the features of a near-equilibrium settings. We introduce the basic concepts of bifurcation, acquaint the reader with some basic results valid for hyperbolic fixed points, such as the Hartman-Grobman theorem. Then we dip into the non-linear approaches necessary for tackling a loss of hyperbolicity.

In the following section we introduce the terminology and basic properties and features of the reaction-diffusion equations and the concept of diffusion-driven instability. We then go on to apply these methods and approaches in the context of pattern selection, where the tools of linear analysis prove insufficient. We carry out a detailed model calculation, analyzing the the succession of steps that lead to the final result along with the underlying mathematical features.

In the last chapter we turn our focus to a different, asymptotic setting, considering the limit of large growth rates. We prove the validity of this asymptotic approach for a multicomponent system, based on the spectral properties of a typical reaction-diffusion system, yet with a restriction to the admissible types of heterogeneity present in our system of equations. In Appendix A we summarize some basic results related to spectral theory of the Lapalce operator, which provide an important theoretical background of reaction-diffusion models.

Chapter 1

Theoretical background

1.1 Bifurcation theory

1.1.1 Introduction

In this section we will examine the properties of solutions to dynamical systems in the neighbourhood of a *critical* (or *bifurcation*) *point*. Our natural motivation is that in a real physical setting we often encounter situations when the dynamics of our system of interest depends on some outer parameter, for instance pressure or temperature. These parameters are not *variables* (neither dependent nor independent) of our system in the proper sense; although they might be affected by the evolution of the system, given that their changes are smooth and sufficiently slow (or, in terms of thermodynamics, *quasistatic*) we can still manipulate them from outside according to our needs and wishes. Yet it is a common observation that with certain (sets of) values of these parameters our system might display a more complex or more interesting behaviour than with others. Furthermore, for certain values of these parameters, the dynamical system might undergo changes that make consideration of these specific parameter values more relevant and meaningful, e.g. a phase transition or a transition from stability to instability. We will call these values *critical*. Given that the "right-hand side" of the evolution equations is reasonably smooth, for parameter values sufficiently close to the critical value, the change of the system will not be too significant and we can thus expand its governing equations to a Taylor expansion. Linearization (i.e. considering only the first-order terms) is a common approach when we, for one reason or the other, wish to obtain an analytic solution. Linearization, however, invariably leads to exponential dynamics, which is usually a rather poor approximation of the long-term evolution of our system unless all the eigenvalues are negative (and thus, our system is stable). On the other hand, in a reasonably small neighbourhood of the *stable state* the dynamics of the system will be well approximated by the first few terms of this expansion. Unfortunately, a shift of the parameters further beyond these critical values would make a more complex approach necessary. We will give these ideas a more precise meaning later in this section. But first, we want to present a well-known experimental example of this type of behaviour: the Rayleigh-Bénard convection. Let us consider two horizontal plains (infinite plates) of high (infinite) thermal conductivity separated by a layer of fluid of height d in a gravitational field with acceleration g. A slight increase in temperature at the bottom plate will induce a stationary state with temperature decreasing linearly along the z-axis (i.e. with height). The temperature gradient also causes inhomogeneity in density as a consequence of thermal expansion of the lower, warmer layers of the fluid. As the temperature difference ΔT of the two plates grows, the combination of uneven density (increasing from bottom to top) and gravity becomes a potential source of instability. It turns out that the key parameter of the system is the ratio of the buoyancy force and the dissipative (viscous) force, the so-called *Rayleigh*

number,

$$R = \frac{\alpha g \Delta T d^3}{\kappa \nu},$$

where α represents the fluid's linear expansion coefficient, κ its thermal diffusivity and ν its kinematic viscosity. Theoretical analysis of the governing Oberbeck-Boussinesq equations concludes that instability will occur at the critical value $R_c = 1708$. This result is in remarkable accordance with the experimental value $R_c^{exp} = 1700 \pm 51$ [4]. Since it is proportional to the temperature difference ΔT, the temperature gradient pointing against the gravitational force can be viewed as the "driving force" of the instability.

1.1.2 The Hartman-Grobman theorem

Let us now give the introductory ideas of this section a mathematical formulation. We will consider a system of the form

$$\dot{x} = V(\mu, x), \qquad x \in \mathbb{R}^n, \mu \in \mathbb{R}, \tag{1.1}$$

where, for simplicity, μ is a *scalar* (bifurcation) parameter, i.e. the equations of motion for our system depend solely on one real parameter. Let us first make a remark concerning the generality of equation (1.1). Recall that this form also includes higher-order differential equations and systems, since any system that takes on the form

$$\frac{d^n y}{dt^n} = f\left(\mu, y, \frac{dy}{dt}, \cdots \frac{d^{n-1} y}{dt^{n-1}}\right)$$

can be transformed into

$$\frac{d}{dt}\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} x_2 \\ \vdots \\ f(\mu, x_1, \cdots, x_n) \end{pmatrix}$$

using the transformation $x_1 = y, x_2 = \frac{dy}{dt}, \ldots, x_n = \frac{d^{n-1}y}{dt^{n-1}}$ and is thus of the form (1.1). We assume that there is a *fixed point* for V in $\mathbb{R}^{n+1}$ which we shift to the origin using a linear transformation, obtaining

$$V(0, 0) = 0.$$

Hence, as suggested above, we can expand V in a Taylor series around $x = 0$. For a fixed value of μ, we have

$$\dot{x} = V(\mu, 0) + DV(\mu, 0)x + \mathcal{O}(x^2), \tag{1.2}$$

where DV denotes the Jacobi matrix of the map V with respect to x, i.e. $DV(\mu, x)_{ij} = \frac{\partial V_i}{\partial x_j}(\mu, x)$. The constant term in (1.2) vanishes for $\mu = 0$, so in a sufficiently small neighbourhood of the origin, upon neglecting higher-order terms, we obtain the linear system of ordinary differential equations (ODEs)

$$\dot{x} = DV(0, 0)x. \tag{1.3}$$

In order to make the following ideas more straightforward, we will assume the matrix $DV(0, 0)$ diagonalizable. Then its eigenvectors form a basis of $\mathbb{R}^n$ in which the solution to equation (1.3) consists of exponential modes whose stability, of course, depends on the spectrum of the matrix. Hence, one of the pivotal points of bifurcation analysis is the study of the dependence of the spectrum of $DV(\mu, 0)$ on μ in the neighbourhood of the point $\mu = 0$. But before we attempt to further develop these ideas, we want to justify this approach by a classical result that highlights the intrinsic connection between the solutions to the linearized problem (1.3) and the generally non-linear original problem (1.1). To be able to state this result - the so-called Hartman-Grobman theorem - we need to establish some of its basic notions.

We will call a fixed point (μ_0, x_0) of V *hyperbolic* if and only if (iff) all the eigenvalues λ_i of $DV(\mu_0, x_0)$ have $\mathrm{Re}\lambda_i \neq 0$. An immediate consequence of this definition is that in a hyperbolic fixed point all the modes either exponentially grow or exponentially decay with time. We further assume the continuity of all partial derivatives $\frac{\partial V}{\partial x_i}$ so that there is a *unique* correspondence between initial conditions $x(0) \in \mathbb{R}^n$ and solutions $x(t)$ to equation (1.1) for any fixed value of μ (i.e. for any initial condition, the solution to (1.1) is unique). Let us now, for purely terminological reasons, view V as a vector field whose value in any given $x \in \mathbb{R}^n$ is the tangent vector (the derivative) to the curve $x(t)$. Thanks to the 1:1 correspondence between initial conditions and integral curves we can properly define the *flow* of the vector field V as a 1-parameter family of maps $\phi_t : \mathbb{R}^n \to \mathbb{R}^n$ that, given an initial point $x(0)$, returns the (vector) value $x(t)$ of the corresponding integral curve at time t. Using this terminology, we can now proceed to formulating the Hartman-Grobman theorem:

Theorem 1.1 (Hartman-Grobman)**.** *Let $x = 0$ be a hyperbolic fixed point of equation* (1.1) *for a fixed value μ_0. Let us further assume that $\Omega \subset \mathbb{R}^n, V(\mu_0, \cdot) \in C^1(\Omega), U \subset \Omega$ is a neighbourhood of $0 \in \mathbb{R}^n$ and $I \subset \mathbb{R}$ is an interval containing 0. Let us denote by ϕ_t the flow of $\dot{x} = V(\mu_0, x)$ and by $\tilde{\phi}_t$ the flow of the corresponding linearized system $\dot{x} = DV(\mu_0, 0)x$. Then there exists a homeomorphism $\Psi : \mathbb{R}^n \to \mathbb{R}^n$ such that for all initial conditions $x_0 \in U$ and all $t \in I$*

$$\phi_t(x_0) = \Psi^{-1} \circ \tilde{\phi}_t \circ \Psi(x_0). \tag{1.4}$$

The homeomorphism Ψ can be viewed as a (continuous) change of coordinates. Hence, for a hyperbolic fixed point, the transition from the linearized system (1.3) to the non-linear equations of motion (1.1) means simply a change of coordinates! Of course, this is only true locally, as we are restricted by the neighbourhood U. In other words, the solutions to the two systems are (locally) topologically equivalent. This means that any local qualitative change to the solution of the full system (1.1) will be displayed in an analogous qualitative change to the solution of the much simpler problem (1.3). It is worth mentioning that the flow of this linear problem can be expressed explicitly using the matrix exponential map as $\tilde{\phi}_t = e^{At}$ where $A \equiv DV(\mu_0, 0)$. That again turns our attention to the spectrum of the matrix $DV(\mu_0, 0)$. By the Hartman-Grobman theorem, its spectrum can also be used to analyze the properties of the "non-linear" flow ϕ_t.

1.1.3 Invariant subspaces

Let us now consider a general (not necessarily hyperbolic) fixed point at the origin: $V(0, 0) = 0$. On the other hand, let us preserve the assumption that the matrix $A \equiv DV(0, 0)$ be diagonalizable, the reason being that if we allowed for degenerate eigenvalues, the following ideas might be invalid. In (our) case of A diagonalizable, a non-zero eigenspace E_λ is associated with every eigenvalue λ from its spectrum. In the case $\lambda \in \mathbb{R}$ this eigenspace is generated by the associated eigenvectors:

$$\lambda \in \mathbb{R} \Rightarrow E_\lambda = \{v \in \mathbb{R}^n \mid (A - \lambda I)v = 0\}.$$

The case $\lambda \in \mathbb{C}$ is more subtle. Since A is a real matrix, its eigenvectors associated with a complex eigenvalue are of the form $v = v_1 + iv_2, v_1, v_2 \in \mathbb{R}^n, v_2 \neq 0$. Vectors of this form, however, are not elements of $\mathbb{R}^n$. It is still reasonable to look for the "eigenspace" E_λ. In order to find it, let us first note that these eigenvectors satisfy $0 = \overline{(A - \lambda I)v} = (A - \bar{\lambda} I)\bar{v}$, i.e. $\bar{v} = v_1 - iv_2$ is an eigenvector of A associated with the eigenvalue $\bar{\lambda}$. Hence, if v is an eigenvector associated with either λ or $\bar{\lambda}$, by commutativity it satisfies $(A - \lambda I)(A - \bar{\lambda} I)v = 0$. Furthermore, linearity implies that this relation also holds for its real and imaginary parts v_1 and v_2, which are elements of $\mathbb{R}^n$. Vectors of this kind span the (generalized) eigenspace associated with λ, so we have

$$\lambda \in \mathbb{C} \setminus \mathbb{R} \Rightarrow E_\lambda = \left\{v \in \mathbb{R}^n \mid (A - \lambda I)(A - \bar{\lambda} I)v = 0\right\}.$$

The diagonalizability assumption implies that the eigenvectors of A span the entire space $\mathbb{R}^n$. Since the solution to the linearized problem (1.3) has the form of exponential modes, we can distinguish the subspaces E_λ based on the sign of the real part of the associated eigenvalue. Then the span of all the eigenvectors in each category defines the stable, unstable and central subspaces, respectively, in the following manner:

$$\begin{aligned} E^s &\equiv \text{span}\,\{v \in E_\lambda \mid \text{Re}\,\lambda < 0\}, \\ E^u &\equiv \text{span}\,\{v \in E_\lambda \mid \text{Re}\,\lambda > 0\}, \\ E^c &\equiv \text{span}\,\{v \in E_\lambda \mid \text{Re}\,\lambda = 0\}. \end{aligned} \tag{1.5}$$

As already mentioned, these subspaces are non-overlapping and span the whole space, i.e. $E^s \oplus E^u \oplus E^c = \mathbb{R}^n$. The simple, exponential dynamics of the system (1.3) implies an important property of each of these subspaces under this dynamics: *invariance*. To be more precise, for an initial condition $x_0 \in E^s/E^u/E^c$, the whole trajectory will be contained in the corresponding subspace, i.e. $x(t) \subset E^s/E^u/E^c$ or, equivalently, $\phi_t(x_0) \in E^s/E^u/E^c, \forall t \geq 0$. Segmenting the phase space into the stable, unstable and central subspaces thus yields an elegant a priori answer to the linear stability of the solution: the solution with initial condition $x(0) \in E^s$ is evidently linearly stable and converges to the equilibrium $(0,0)$ as $t \to +\infty$. On the other hand, $x(0) \in E^u$ implies linear instability and convergence towards $(0,0)$ for $t \to -\infty$. Stability of the equilibrium $x = 0$ for the linear system (1.3) is guaranteed given that no eigenvalue of A has a positive real part, i.e. if

$$E^s \oplus E^c = \mathbb{R}^n.$$

This equilibrium is asymptotically stable iff all eigenvalues λ of A satisfy $\text{Re}\lambda < 0$, i.e. iff

$$E^s = \mathbb{R}^n. \tag{1.6}$$

In fact, relation (1.6) tells us more about the system (1.1) than linear stability analysis can. It can be shown that if (1.6) holds, the solution $x = 0$ is asymptotically stable not only for the linear dynamics of (1.3), but also for the full, non-linear dynamics of (1.1) (see chapter 9 in [8]). There is, however, a fundamental difference between these two cases. For the linear dynamics of (1.3), asymptotic stability is clearly guaranteed given any initial condition $x(0)$ whereas for problem (1.1) relation (1.6) "only" guarantees the existence of a *neighbourhood* U of 0 for which the equilibrium $x = 0$ is attractive (by attractiveness we mean that all trajectories $x(t)$ with initial conditions $x(0) \in U$ converge to $x = 0$ as $t \to +\infty$), but gives us no knowledge of the *size* of the neighbourhood. Let us conclude this section by noting that for stability analysis of initial conditions from E^c, linear methods are insufficient and more sophisticated approaches are required.

1.1.4 Invariant manifolds

There are several ways to define the notion of *differentiable manifold*, some of them more abstract than others. Although the formal definition will not be crucial for the ideas we are about to present, for the sake of notational completeness as well as to facilitate the possibility of further contemplation of the concepts, we give here one of the less technical and hopefully more intuitive definitions:

Definition 1.2. Let $m, n, q \in \mathbb{N}, m < n$. The non-empty set $M \subset \mathbb{R}^n$ is a *differentiable manifold* of class C^q of dimension $r \equiv n - m$ iff:

(i) $(\forall x_0 \in M)(\exists H_{x_0}, \Phi : H_{x_0} \to \mathbb{R}^m \in C^q)(M \cap H_{x_0} = \{x \in H_{x_0} \mid \Phi(x) = 0\})$,

(ii) $(\forall x \in H_{x_0})(h(D\Phi(x)) = m)$.

We will hereinafter mostly omit the adjective *differentiable*, so by *manifold* we will mean *differentiable manifold*. A special case of a manifold, according to the given definition, is the space $\mathbb{R}^n$ itself. Hence, we can view the notion of manifold as a generalization the term *vector space* that also allows for non-linear structures. Accordingly, the following ideas will generalize the previous section concerning invariant subspaces. Let us now introduce the term *invariant manifold*:

Definition 1.3. Let $M \subset \mathbb{R}^n$ be a manifold in the phase space of problem (1.1). We call M an *invariant manifold* for the dynamical system (1.1) iff for any $m \in M$ the relation $\{\phi_t(m) \mid t \in \mathbb{R}\} \subset M$ holds, i.e. iff the entire solution $x(t)$ to (1.1) with initial condition $x(0) = m$ (the *trajectory through* m) is contained in M.

By definition, invariant manifolds define autonomous subsystems of the given system with 'closed' dynamics. The simplest examples of invariant manifolds might include equilibria (fixed points) and individual trajectories. Note that the invariant subspaces defined in (1.5) are invariant manifolds for the linearized dynamics of problem (1.3).

For our further considerations it is crucial to note that although the non-linear terms in equation (or the expansion of) (1.1) 'deform' the linear structure of the invariant subspaces, their qualitative properties with respect to the given dynamical system are preserved. More precisely, if $V(\mu, \cdot) \in C^{r+1}$, $r \geq 1$, then there are C^r-manofolds W^s, W^u, W^c such that $\dim E^\alpha = \dim W^\alpha$, E^α is tangent to W^α at 0 for $\alpha = u, s, c$, and for $x \in W^s(W^u)$ we have $\phi_t(x) \to 0$ for $t \to +\infty$ ($t \to -\infty$) (see chapter 27 in [1]). The defining properties of W^s and W^u are even sufficient to define them *uniquely*.

1.1.5 Basic bifurcations

Before we exploit the properties of invariant manifolds, let us introduce the theoretical frame of *bifurcation analysis*, the analysis of the dependence of our dynamical system on the parameter μ. On the example of the Rayleigh-Bénard convection we demonstrated that instability and the ensuing establishment of pattern in real systems is often incurred by exceeding a critical value of an outer parameter. By the Hartman-Grobman theorem, this qualitative change in the dynamics in the neighbourhood of a *hyperbolic* (e.g. asymptotically stable) equilibrium is reflected in the dynamics of the linearized problem as well. By the implicit function theorem, altering the parameter near a 'hyperbolic' value μ_0 (i.e. one that admits a hyperbolic fixed point) preserves the equilibrium. More precisely, the equilibrium is shifted along a unique equilibrium 'branch'. However, its qualitative properties (under linear analysis of its spectral properties) will often change dramatically.

First of all, let us now present the implicit function theorem in the formulation useful for our purposes:

Theorem 1.4. *Let $\Phi : \mathbb{R}^{r+m} \to \mathbb{R}^m$ be a C^q-map and let $q, r, m \in \mathbb{N}$. Suppose that there exists a point $(x_0, y_0) \in Dom\ \Phi$ such that*

(i) $\Phi(x_0, y_0) = 0$,

(ii) $\left|\frac{\partial \Phi_i}{\partial y_j}(x_0, y_0)\right| \neq 0$.

Then there exists a neighbourhood $H_{x_0} \subset \mathbb{R}^r$ such that the equation $\Phi(x, y) = 0$ defines a unique C^q-map $\varphi : H_{x_0} \to \mathbb{R}^m$ on H_{x_0} that satisfies

(1) $\varphi(x_0) = y_0$,

(2) $\Phi(x, \varphi(x)) = 0$ *for* $x \in H_{x_0}$.

Here we silently and shamelessly interchange the isomorphic spaces $\mathbb{R}^{r+m}$ and $\mathbb{R}^r \times \mathbb{R}^m$. For the purposes of (1.1) we set $r = 1, m = n, \Phi = V, (x_0, y_0) = (0, 0)$. Hence, in a hyperbolic equilibrium, where 0 is not an eigenvalue of $DV(0, 0)$ (which we still assume diagonalizable), the equilibrium does not vanish when the parameter μ is varied around 0 but its precise position in phase space will evolve along a unique curve $X(\mu)$ whose existence and uniqueness is guaranteed by the implicit function theorem.

The topological equivalence of the 'linear' and 'non-linear' flows in the neighbourhood of *hyperbolic* equilibria described by the Hartman-Grobman theorem 1.1 means that the relatively simple linear methods are sufficient for studying the system's dynamics as long as hyperbolicity is preserved. Therefore we will now assume that there exists a *critical* value μ_c of the bifurcation parameter where hyperbolicity is lost, i.e. some eigenvalue $\lambda(\mu)$ of the matrix $DV(\mu, 0)$ has Re $\lambda(\mu_c) = 0$. There are two qualitatively different ways this can occur:

(a) $\lambda = 0$
In this case the implicit function theorem 1.4 is no longer applicable. That usually means loss of uniqueness of the solution to equation (1.1). After passing the point $(\mu_c, X(\mu_c))$, the (previously unique) equilibrium branch $X(\mu)$ *bifurcates*, i.e. it splits into multiple branches which can differ not only in their phase-space positions, but also in stability. Their shape and other properties depend on the specific form of the map V; in the neighbourhood of the bifurcation point mostly on the lowest non-linear term of its Taylor expansion.

(b) $\lambda = \pm i\omega$
We usually refer to the case of a pair of complex conjugate eigenvalues with zero real part as *Hopf bifurcation*. It typically exhibits a time-dependent, non-linear, often oscillatory behaviour.

To gain some intuition as well as to demonstrate the features of stability analysis, we will now give an overview of several basic bifurcations. For simplicity, we will consider the case of a one-dimensional real variable x, so (1.1) now has the form

$$\dot{x} = V(\mu, x), \qquad \mu, x \in \mathbb{R}. \tag{1.7}$$

In the following we will assume, without loss of generality, that criticality in our system occurs for $\mu_c = 0$, and we again shift the equilibrium to the point $x = 0$ which leaves us with

$$V(0, 0) = 0.$$

The criticality condition for a real-valued 0 eigenvalue then yields

$$\frac{\partial V}{\partial x}(0, 0) = 0.$$

Expanding the right-hand side into a Taylor expansion, we obtain

$$\dot{x} = \frac{\partial V}{\partial \mu}(0, 0)\mu + \frac{\partial^2 V}{\partial x^2}(0, 0)\frac{x^2}{2} + \frac{\partial^2 V}{\partial x \partial \mu}(0, 0)x\mu + \frac{\partial^2 V}{\partial \mu^2}(0, 0)\frac{\mu^2}{2} + \frac{\partial^3 V}{\partial x^3}(0, 0)\frac{x^3}{6} + \cdots . \tag{1.8}$$

Depending on the coefficients of the first few terms in (1.8) we distinguish different bifurcation types. We will now describe a few of these more specifically. In what follows, unless explicitly stated otherwise, we assume the coefficients of interest to be non-zero.

(a) Saddle-node bifurcation
Neglecting the term $\frac{\partial^2 V}{\partial x \partial \mu}(0, 0)x\mu$, in a small enough neighbourhood of $(0, 0)$, we have

$$\dot{x} = \alpha\mu + \beta x^2, \qquad \alpha = \frac{\partial V}{\partial \mu}(0, 0), \beta = \frac{1}{2}\frac{\partial^2 V}{\partial x^2}(0, 0).$$

We can further simplify this expression using the rescaled variables $\tilde{x} = |\beta|x, \tilde{\mu} = |\alpha\beta|\mu$, which leads to the system (we mean *dynamical* system; for $x \in \mathbb{R}$ it consists of a single equation)

$$\dot{\tilde{x}} = \operatorname{sgn}(\alpha)\tilde{\mu} + \operatorname{sgn}(\beta)\tilde{x}^2 \tag{1.9}$$

with two distinct equilibria $\tilde{x}_\pm(\tilde{\mu}) = \pm\sqrt{-\operatorname{sgn}(\alpha)\operatorname{sgn}(\beta)\tilde{\mu}}$. These two branches merge in the critical point $\mu = 0$. We kept the signs of the coefficients α and β explicit so as to demonstrate that for $\operatorname{sgn}(\alpha)\operatorname{sgn}(\beta) = 1(-1)$ the corresponding equilibria will only exist in the left (right) half-plane of the (μ, x)-plane.
Let us now study the stability of these solutions. Since we want to consider the immediate neighbourhood of the equilibrium, we will apply *linear* stability analysis. For this purpose, let us define the perturbation $y_\pm$ to the equilibrium by introducing the shifted variable $\tilde{x} = \tilde{x}_\pm(\tilde{\mu}) + y_\pm$. Substituting this into (1.9) we obtain the evolution equation for $y_\pm$ in the form

$$\dot{y}_\pm = 2\operatorname{sgn}(\beta)\tilde{x}_\pm(\tilde{\mu})\, y_\pm + O(y_\pm^2),$$

where $O(y_\pm^2)$ this time only represents a single non-linear term $\operatorname{sgn}(\beta)y_\pm^2$. Hence, x_+ (x_-) is stable for $\beta < 0$ ($\beta > 0$) and unstable for $\beta > 0$ ($\beta < 0$). The signs of α and β determine in which half-plane of (μ, x) (i.e. to which side of $\mu = 0$) the equilibria will exist.
A characteristic feature of this type of bifurcation is that the two branches of equilibria - 1 stable and 1 unstable - approach each other and eventually merge in the critical point $\mu = 0$. Depending on the signs of the coefficients α and β the equilibrium completely vanishes for super- or subcritical values of μ. Let us conclude this part by stating that in one dimension we sometimes talk about *tangential* bifurcation rather than saddle-node bifurcation. The typical case of a *saddle-node* bifurcation is in two dimensions, where one unstable (saddle) and one stable (node) fixed point mutually annihilate in the critical point μ_c.

(b) Transcritical bifurcation
A qualitatively different situation occurs in case $\frac{\partial V}{\partial \mu}(0,0) = 0 = \frac{\partial^2 V}{\partial \mu^2}(0,0)$. This is for instance true if the existence of the equilibrium is preserved upon varying the value of μ. A perturbation of μ can change both stability and position $X(\mu)$ of the equilibrium in phase space, but as long as it 'survives', we can introduce a new variable $\bar{x} = x - X(\mu)$ which satisfies $V(\mu, \bar{x} = 0) = 0$ for all μ in an appropriate neighbourhood U_μ of $\mu_c = 0$. Hence, the coefficients of the form $\frac{\partial^n V}{\partial \mu^n}(0, \bar{x} = 0)$ vanish (within U_μ) for all $n \in \mathbb{N}$, and from expansion (1.8) we obtain the system

$$\dot{x} = x(\alpha\mu + \beta x) \equiv \tilde{V}(\mu, x), \qquad \alpha = \frac{\partial^2 V}{\partial x \partial \mu}(0,0), \beta = \frac{1}{2}\frac{\partial^2 V}{\partial x^2}(0,0), \tag{1.10}$$

which again has two stationary points $x = 0$ and $x = -\frac{\alpha}{\beta}\mu$. For the solution $x = 0$ a change in stability occurs as μ passes through 0. Let us, for instance, consider $\alpha > 0$. Then for $\mu < 0$ the system is stable, for $\mu > 0$ it is unstable, and for the critical value $\mu = 0$ the eigenvalue of the linearized system $\alpha\mu$ is 0. For the equilibrium $x = -\frac{\alpha}{\beta}\mu$ we set, analogously to the previous, $y = x + \frac{\alpha}{\beta}\mu$. By linearization we obtain

$$\dot{y} = -\alpha\mu y = \frac{\partial \tilde{V}}{\partial x}(\mu, -\frac{\alpha}{\beta}\mu)y.$$

If we for certainty assume $\alpha > 0$, then we have stability for positive values of μ and instability for μ negative. For α negative it is, of course, the other way round. Note that in a neighbourhood of the critical value $\mu_c = 0$, the system (1.10) always has one stable and one unstable equilibrium, regardless of the sign of μ. However, the stabilities of the equilibria swap as μ passes through 0.

(c) Pitchfork bifurcation
Let us now assume that, in appropriately chosen coordinates, our system has reflection symmetry, i.e. $V(\mu, -x) = -V(\mu, x)$. Then we necessarily have $V(\mu, 0) = 0$ as well as $\frac{\partial^2 V}{\partial x^2}(0,0) = 0$. These are exactly the conditions of *pitchfork bifurcation* whose normal form is

$$\dot{x} = x(\alpha\mu + \beta x^2), \qquad \alpha = \frac{\partial^2 V}{\partial x \partial \mu}(0,0), \beta = \frac{1}{6}\frac{\partial^3 V}{\partial x^3}(0,0).$$

In a sense, pitchfork bifurcation is a combination of saddle-node and transcritical bifurcations, since we have three stationary points: $x = 0, x_\pm = \pm\sqrt{-\frac{\alpha}{\beta}\mu}$. The equilibrium $x = 0$ will only exist in one half-plane (either left or right) of the phase space (μ, x), depending on the signs of the coefficients α, β. Let us remark that for the rescaled variables $\tilde{x} = |\beta|^{1/2}x, \tilde{\mu} = |\alpha|\mu$, we obtain the *normal form* for the pitchfork bifurcation as

$$\dot{\tilde{x}} = \tilde{x}(\mathrm{sgn}(\alpha)\tilde{\mu} + \mathrm{sgn}(\beta)\tilde{x}^2).$$

The stability of $x = 0$ is rather straightforward: the system is clearly stable (unstable) for $\mathrm{sgn}(\alpha\mu) = -1(+1)$. For the other two equilibria we proceed analogously as before: we introduce the perturbation $y_\pm = x - x_\pm$ of the equilibrium $x_\pm = \pm\sqrt{-\frac{\alpha}{\beta}\mu}$ which is, to first order, governed by the evolution equation $\dot{y}_\pm = 2\alpha\mu y_\pm$. Hence, both x_+ and x_- are stable for $\mathrm{sgn}(\alpha\mu) = 1$ and unstable for $\mathrm{sgn}(\alpha\mu) = -1$. Note that for (the signs of) α and β given, the stationary point $x_\pm$ only exists for either $\mu < 0$ or $\mu > 0$. Unlike for saddle-node bifurcation, here both of the equilibria $x_\pm$ are simultaneously either stable or unstable. There is a similarity to the saddle-node bifurcation, however, in that at $\mu = 0$ the two equilibria merge and annihilate.

(d) Hopf bifurcation
This kind of bifurcation only occurs for spatial dimension $n \geq 2$. Its normal form usually uses a complex variable z and has the form

$$\dot{z} = z(\lambda(\mu) + a(\mu)|z|^2), \qquad a, \lambda \in \mathbb{C}.$$

Setting $z = z_1 + iz_2$, $a(\mu) = a_1(\mu) + ia_2(\mu)$ and $\lambda(\mu) = \gamma(\mu) + i\omega(\mu)$, we obtain the system

$$\dot{z_1} = z_1\left(\gamma + a_1(z_1^2 + z_2^2)\right) - z_2\left(\omega + a_2(z_1^2 + z_2^2)\right),$$

$$\dot{z_2} = z_2\left(\gamma + a_1(z_1^2 + z_2^2)\right) + z_1\left(\omega + a_2(z_1^2 + z_2^2)\right),$$

which, in the polar coordinates $z_1 = r\cos\theta, z_2 = r\sin\theta$, using the relations

$$\dot{r} = \dot{z}_1(r, \dot{r}, \theta, \dot{\theta})\cos\theta + \dot{z}_2(r, \dot{r}, \theta, \dot{\theta})\sin\theta,$$

$$\dot{\theta}r = \dot{z}_2(r, \dot{r}, \theta, \dot{\theta})\cos\theta - \dot{z}_1(r, \dot{r}, \theta, \dot{\theta})\sin\theta,$$

can be rewritten in the form

$$\dot{r} = \gamma r + a_1 r^3,$$
$$\dot{\theta} = \omega + a_2 r^2.$$

As mentioned above, Hopf bifurcation assumes that criticality occurs via a pair of complex conjugate eigenvalues with zero real part. Let us denote the eigenvalue as $\gamma(\mu) + i\omega(\mu)$ and let us assume that

$$\gamma(0) = 0, \qquad \omega(0) \neq 0, \qquad \frac{d\gamma}{d\mu}(0) > 0. \tag{1.11}$$

These assumptions are to say that criticality in the eigenvalue $\gamma(\mu) + i\omega(\mu)$ occurs in a non-degenerate way. Let us now analyze the evolution equation for r, which is independent of θ and has two stationary points $r_0 = 0$ and $r_+(\mu) = \sqrt{-\gamma/a_1}$ [1]. The conditions (1.11) imply that in a sufficiently small neighbourhood of 0 we have $\mathrm{sgn}(\gamma) = \mathrm{sgn}(\mu)$, so the equilibrium $r = 0$ is stable for $\mu < 0$ and unstable for $\mu > 0$. Let us further assume that a depends on μ in a continuous manner. Then for $a_1(0) < 0$, the non-zero equilibrium r_+ (within a certain neighbourhood of $\mu = 0$) only exists for positive values of μ. By linearization we obtain the corresponding eigenvalue as $-2\gamma < 0$, so for any value of μ there is one stable stationary point. Hence, for $r = r_+$ the system stabilizes in a limit cycle with frequency $\omega_+(\mu) = \omega(\mu) + a_2(\mu)r_+^2$. Recovering the complex variable z, we can rewrite this in a more compact form as

$$z(t) = r_+ e^{i\omega_+ t}.$$

Since this limit cycle only exists for supercritical values $\mu > \mu_c = 0$, we often talk about *supercritical Hopf bifurcation*.
The situation is significantly different for $a_1(0) > 0$. In that case the existence of the stationary state r_+ requires $\mu < 0$, leaving the system with no stable equilibrium for $\mu > 0$ (in which case the state $r = 0$ is unstable). Unlike in the supercritical case, now the eigenvalue $-2\gamma > 0$ implies instability of the state $r = r_+$, making a non-trivial limit cycle impossible. In this case we talk about *subcritical Hopf bifurcation*.

1.2 Reaction-Diffusion equations

1.2.1 The basic setup

Reaction diffusion (RD) equations and pattern formation are of special interest within this work. Both the choice and the extent of different mathematical concepts under study, such as bifurcation theory or others that will follow, are motivated by RD equations. Let us therefore present the general form of a RD equation and make some basic observations. A reaction diffusion equation is a partial differential equation of the form

$$\frac{\partial \mathbf{c}}{\partial t} = D\Delta\mathbf{c} + \mathbf{f}(\mathbf{c}), \tag{1.12}$$

where $\mathbf{c} : \Omega \times \mathbb{R} \to \mathbb{R}^n_+$, with $\Omega \subset \mathbb{R}^m$, is the concentration vector of the underlying n chemical substances[2], $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$ describes their reaction kinetics, Δ is the Laplace operator representing diffusion (applied componentwise), and D is the diagonal $n \times n$ matrix of (positive) diffusion coefficients. This concept, where the diffusion flux is proportional to the (negatively taken) gradient of concentration, is usually referred to as *Fickian diffusion*, the purely diffusive equation of motion $\frac{\partial \mathbf{c}}{\partial t} = D\Delta\mathbf{c}$ being known as *Fick's second law*. As already indicated, the admissible phase space for the vector of concentrations is the positive quadrant of $\mathbb{R}^n$. Note that we apply the differential operator Δ *before* the diffusion matrix, thus assuming no spatial dependence of the diffusion coefficients for now; otherwise, the diffusion term would have to be replaced by $\nabla \cdot (D\nabla\mathbf{c})$, where $\nabla\mathbf{c}$ is a tensor and both ∇ operators are again applied accordingly (we can view $\nabla\mathbf{c}$ as well as $D\nabla\mathbf{c}$ as $n \times m$ matrices, then the divergence operator acts on each line of this matrix separately). More generalizations are possible; there could for instance be a temporal dependence in $\mathbf{f}$, indicating changes of the outer setting (e.g. manipulating the temperature of a chemical system) resulting in changes of the reaction rates. A partial differential problem of the form (1.12)

[1] We obviously allow neither for imaginary nor for negative values of the radial variable r.

[2] In other contexts $\mathbf{c}$ can also represent other quantities, e.g. concentrations of animal species (i.e. a continuous model of occurrence) in ecological problems.

obviously requires boundary conditions in order to be well-posed. In the context of *embryogenesis* or - more generally - *spontaneous* pattern formation we usually consider the Neumann boundary conditions

$$(\boldsymbol{n} \cdot \nabla)\mathbf{c}(\boldsymbol{r}, t) = 0 \text{ for } \boldsymbol{r} \in \partial\Omega,$$

which arise naturally in Fickian diffusion as zero boundary flux. Here $\boldsymbol{n}$ represents the outward normal to the boundary $\partial\Omega$.

Let us now assume that there exists a homogeneous steady state $\mathbf{c}_*$ satisfying

$$\mathbf{f}(\mathbf{c}_*) = 0. \tag{1.13}$$

This is a very reasonable assumption since reaction kinetics of chemical systems is usually described by polynomial or rational functions, thus providing us with *equilibrium* states given by the (constant) solutions (roots) of equation (1.13). Probably the most fundamental observation, originally made by Alan Turing in his well-known paper [16], is that given a 'favourable' reaction kinetics f, reaction and diffusion alone can account for the transition from the homogeneous equilibrium to an inhomogeneous steady state as well as for a large variety of patterns and phenomena observed in nature such as spots, stripes, wave-like phenomena etc. This observation is based on linear stability analysis, which provides a sufficient description of (at least the qualitative features of) the system in states close to the equilibrium $\mathbf{c}_*$. Not only does this follow from intuition, it is also underlined by the Hartman-Grobman theorem 1.1 for hyperbolic equilibria.

Let us state the stability argument in a more precise manner. Since we wish to apply *linear* stability analysis, we need to linearize equation (1.12) around a fixed point $\mathbf{c}_*$. This procedure, upon introducing the perturbation $\mathbf{w} \equiv \mathbf{c} - \mathbf{c}_*$, yields

$$\frac{\partial \mathbf{w}}{\partial t} = D\Delta\mathbf{w} + D\mathbf{f}(\mathbf{w} = \mathbf{0})\mathbf{w} + O\left(\|\mathbf{w}\|^2\right).$$

We can see that after neglecting the $O\left(\|\mathbf{w}\|^2\right)$ terms the linear stability of the solution to the linear equation

$$\frac{\partial \mathbf{w}}{\partial t} \equiv \partial_t \mathbf{w} = D\Delta\mathbf{w} + D\mathbf{f}(\mathbf{0})\mathbf{w} \tag{1.14}$$

will depend on the spectrum of the Laplace operator as well as that of the Jacobi matrix $D\mathbf{f}(\mathbf{0})$. In order to be able to address the issue of linear stability more clearly, we wish to exploit the fact that the Laplace operator $-\Delta$ on a bounded domain Ω with the Neumann (or, more generally, Robin) boundary condition is self-adjoint, non-negative and if the boundary $\partial\Omega$ satisfies relatively mildly-restrictive regularity conditions (e.g. $\partial\Omega$ of class C^2 suffices), it has a purely discrete spectrum so its eigenfunctions form an orthonormal basis of the Hilbert space $L^2(\Omega)$ (see Appendix A). Hence, we can discuss the stability of the individual modes which are given as the solutions of

$$\begin{aligned} \Delta w_k &= -k^2 w_k, \\ (\boldsymbol{n} \cdot \nabla)\, w_k &= 0 \text{ on } \partial\Omega, \end{aligned}$$

where we denoted the non-negative eigenvalue by k^2; k is usually referred to as the *wavenumber*. Unless stated otherwise, we will consider the case of Neumann boundary conditions for which $k = 0$ is always in the spectrum. Recalling the completeness of the set of eigenfunctions of the Laplace operator in $L^2(\Omega)$, we know that at any given time t_0 the solution $\mathbf{w}$ to (1.14) (more precisely, each of its components) lies in the span of the functions $\{w_k\}_{k=0}^{\infty}$, with the uniquely given coefficients $\{\mathbf{a}_k\}_{k=0}^{\infty}$ allowing for temporal dependence. Thus, in general we have

$$\mathbf{w}(\boldsymbol{r}, t) = \sum_{k=0}^{\infty} \mathbf{a}_k(t) w_k(\boldsymbol{r}), \qquad \mathbf{a}_k : \mathbb{R} \to \mathbb{R}^n.$$

Inserting this separation of variables into (1.14), we obtain a (Fourier) expansion in the orthonormal basis $\{w_k\}_{k=0}^{\infty}$ on both sides so requiring their equality yields the coefficient equation

$$\dot{\mathbf{a}}_k = (A - k^2 D)\mathbf{a}_k, \tag{1.15}$$

where we denoted $A \equiv D\mathbf{f}(\mathbf{0})$. This is a system of linear ODEs with constant coefficients whose solution has the form

$$\mathbf{a}_k(t) = \mathbf{v}e^{\lambda_k t}$$

for some $\mathbf{v} \in \mathbb{R}^n$ determined by the initial conditions. Its linear stability clearly depends on the sign of (the real part of) the eigenvalue λ_k.

1.2.2 Diffusion-driven instability

With this said, let us now present the idea of *diffusion-driven* (or *Turing*) *instability*, hereinafter sometimes shortened to DDI. As the term reveals, Turing's idea was that diffusion could become the driving force of pattern formation in that it could cause instability of a homogeneous steady state if an inhomogeneous perturbation kicked in. For this to occur (or at least be admissible), several requirements must be fulfilled. Firstly, *stability* of the homogeneous steady state (1.13) is required. A *homogeneous* state eliminates any effect of diffusion so that the linearized reaction diffusion equation (1.14) gives $\partial_t \mathbf{w} = A\mathbf{w}$ and the equation of motion (1.15) becomes

$$\dot{\mathbf{a}}_k = A\mathbf{a}_k.$$

Solving this linear system leads to the well-known problem

$$|\lambda I - A| = 0.$$

The (asymptotic) stability of the homogeneous equilibrium, in accordance with the observations made in section 1.1.3, requires that all eigenvalues of the matrix A be negative (have a negative real part). It is a matter of elementary algebra to show that in the case of a two-component RD system, where A is a 2×2 matrix, its eigenvalues are given as

$$\lambda_{1,2} = \frac{1}{2}\left[\operatorname{Tr} A \pm \sqrt{(\operatorname{Tr} A)^2 - 4|A|}\right],$$

so requiring their (real parts') negativeness is equivalent to requiring both

$$\operatorname{Tr} A < 0 \text{ and } |A| > 0. \tag{1.16}$$

We can use these conditions to gain some intuition for the concept of diffusion-driven instability. A good mental setting for this is provided by predator-prey models. Let us in this case denote $\mathbf{c} - \mathbf{c}_* = \begin{pmatrix} u - u_* \\ v - v_* \end{pmatrix}$ and $\mathbf{f} = \begin{pmatrix} f(u,v) \\ g(u,v) \end{pmatrix}$, where $\mathbf{c}_* = \begin{pmatrix} u_* \\ v_* \end{pmatrix}$ is the homogeneous equilibrium. Then the Jacobi matrix $A = D\mathbf{f}(\mathbf{c}_*)$ takes on the form

$$A = \begin{pmatrix} f_u & f_v \\ g_u & g_v \end{pmatrix}_{(u_*, v_*)},$$

where the indices u and v again represent the partial derivatives with respect to the corresponding variable. The condition on its trace (hereinafter we always consider derivatives of the reaction kinetics to be evaluated at the equilibrium unless stated otherwise) then becomes $f_u + g_v < 0$. In other words, at least

one of the substances u, v needs to be self-inhibiting in the sense that its 'diagonal' partial derivative be negative; such a substance has a stabilizing effect on itself near the equilibrium. This is self-evident in the predator-prey context.

Let us consider a region with an even distribution of two species, a predator and its prey, in an equilibrium state. If we, for simplicity, consider food to be the only factor affecting the occurrence of either species, then an equilibrium state means that the number of the prey is just sufficient to support the predators whose number, on the other hand, is just right to prevent a growth in the prey numbers. Imagine now that we do not allow for diffusion-like phenomena (a local increase in concentration in one place at the cost of a decline in its vicinity or vice versa) due to migration, i.e. we only admit uniform distributions of the species. Then it seems intuitive that the equilibrium state would be unstable if a slight (homogeneous) increase in the number of either of the species stimulated further growth. For instance, if the food capacity of the region for the prey was (hypothetically) infinite, any increase in its number would also enhance its reproduction; a greater number of the prey could in turn support more predators and this spiral could continue forever, resulting in an ever-growing number of both species. That is certainly not what stability looks like. Note that nature has taken care of this in its very principles since an increase (decrease) in the occurrence of predators at an equilibrium state means more (less) competition, so the predator count will usually (but not always) be self-inhibiting rather than self-activating in the vicinity of an equilibrium state. But let us ignore this for a little while longer, for instance assuming that the sudden increase in prey concentration was large enough to neutralise the effect of competition (for food, at least). One could still be a little uneasy about the fact that we ignored the cross-dependence: the effect of predation. It seems reasonable that if the predator reproduced faster than the prey, even an infinite food capacity would not mean a boundless growth of the prey concentration: the predator-to-prey ratio would sooner or later increase beyond some threshold value, causing a reduction of the prey. Insufficient food supplies would then result in a drop in the predator numbers and one could naively hope that the equilibrium state might be restored. This would, however, lead to a *limit cycle* rather than stability given that the non-linear terms *do* admit cyclic behaviour. We can observe such dynamics even if the predator is indeed self-inhibiting. An example of such dynamics is provided by the Lotka-Volterra model (see chapter 3.1 in [13]). Cyclic behaviour will be discussed a little bit more later in this section. Had we wished to avoid this discussion, instead of a predator-prey model we could have considered a parasite-host interaction or a chemical reaction, where two self-activating reactants intuitively seem more inclined to a chain reaction (e.g. nuclear fission) than towards stability. Based on all these examples, it (hopefully) appears intuitively clear that the self-inhibiting effects need to prevail for two substances/species to establish a stable equilibrium, in accordance with the first condition in (1.16).

Assuming that the steady state is indeed stable, we wish to see how diffusion can be the cause of instability and result in the onset of pattern formation. Let us therefore consider an inhomogeneous perturbation of the equilibrium $\mathbf{c}_*$. Then diffusion becomes a factor and we need to investigate the stability of the solutions to (1.15). The problem is again linear and the corresponding eigenvalues are given by the characteristic equation

$$|\lambda I - A + k^2 D| = 0. \tag{1.17}$$

For diffusion-driven instability to occur, it is necessary that for some k we have $\Re\lambda\left(k^2\right) > 0$. An immediate observation is that the substances must not diffuse at equal rates; if D was a (positive) multiple of the identity matrix $D = dI$, we would just be looking for eigenvalues of A of the form $\tilde{\lambda} = \lambda + dk^2$. Since all eigenvalues of A are negative, so will be λ and we would have stability. Hence, the substances (e.g. chemicals or species) need to diffuse at different rates. For further intuition we again study the case of a two-component system. In that case we usually non-dimensionalize the equations in such a way that

the diffusion matrix has the form $D = \begin{pmatrix} 1 & 0 \\ 0 & d \end{pmatrix}$, where d is the *ratio* of the diffusion coefficients. Then the equation (1.17) becomes

$$\begin{vmatrix} \lambda - f_u + k^2 & -f_v \\ -g_u & \lambda - g_v + dk^2 \end{vmatrix} = 0$$

and the eigenvalues are thus given as the solutions of

$$\begin{aligned} \lambda^2 + \lambda\left[k^2(1+d) - (f_u + g_v)\right] + (f_u g_v - f_v g_u) - k^2(g_v + d f_u) + dk^4 = \\ = \lambda^2 + \lambda \underbrace{\left[k^2(1+d) - \operatorname{Tr} A\right]}_{\equiv l(k^2)} + \underbrace{|A| - k^2(g_v + d f_u) + dk^4}_{\equiv h(k^2)} = 0. \end{aligned}$$

As already mentioned, for diffusion-driven instability it is essential that this equation have negative solutions for some k^2 from the spectrum of the (negatively taken) Laplace operator. The eigenvalues are clearly given as

$$\lambda(k^2)_{1,2} = \frac{1}{2}\left[-l(k^2) \pm \sqrt{(l(k^2))^2 - 4h(k^2)}\right]. \tag{1.18}$$

The relation $\lambda(k^2)$ is usually referred to as the *dispersion relation*. Let us first note that by (1.16) the coefficient of the linear term $l(k^2)$ is always positive. Hence, if an eigenvalue of the form (1.18) is to have a positive real part (and thus provide us with an unstable mode), it has to be real and the term $h(k^2)$ needs to attain negativity. By (1.16) we have $|A| > 0$, so it is clear that the negativity of h necessarily requires that $(d f_u + g_v)$ be positive. This in turn requires $d \neq 1$ as $f_u + g_v = \operatorname{Tr} A < 0$, which is hardly surprising; we have already argued that diffusion coefficients cannot all be equal for DDI to occur. But it tells us more: it tells us that f_u and g_v need to have opposite signs around equilibrium. It might seem intuitive for some to think that two self-inhibiting chemicals or species would likely mean 'too much damping' for any instability to take place. We leave such considerations, along with the algebraic 'hints' that there might be some substance in such view, up to the reader.

Let us now turn our focus back to the conditions for the occurrence of instability. As h is a quadratic function in k^2 with a positive 'quadratic' coefficient d, a necessary and sufficient condition for its negativity (anywhere) is that its 'bottom', attained at the point $k^2_{min} = \frac{d f_u + g_v}{2d}$, be in the negative half-plane. Inserting this value into h, we obtain the requirement that

$$h_{min} \equiv h(k^2_{min}) = \left[|A| - \frac{(d f_u + g_v)^2}{4d}\right] < 0 \Leftrightarrow |A| < \frac{(d f_u + g_v)^2}{4d}. \tag{1.19}$$

We have thus dissipated the blur surrounding the term 'favourable kinetics' used at the start of this section: for a two-component system it is one that admits a uniform equilibrium and whose derivatives at equilibrium satisfy

$$f_u + g_v = \operatorname{Tr} A < 0, \qquad d f_u + g_v > 0, \qquad 0 < f_u g_v - f_v g_u = |A| < \frac{(d f_u + g_v)^2}{4d}. \tag{1.20}$$

More generally, **favourable kinetics has to admit a uniform stable equilibrium** which becomes **unstable under** inhomogeneous perturbations due to the effects of **diffusion**.

Let us now make a few more observations based on the foregoing analysis. We will focus on a two-component system since the analysis there is most straightforward and so is our intuition. As we know, f_u and g_v need to have opposite signs for the system to admit diffusion driven instability. The positiveness of $|A|$ then means that the same has to apply to the derivatives describing cross-dependence, namely f_v and g_u. Again, it seems intuitive that any 'favourable' kinetics must have two positive and

two negative derivatives at equilibrium as the pro-growth and anti-growth factors need to be balanced for the system to establish an inhomogeneous steady state. There are obviously two ways this can occur: either $f_v > 0$ and $g_u < 0$ or $f_v < 0$ and $g_u > 0$. These two possibilities correspond to qualitatively different situations. Let us, for certainty, assume that $f_u > 0$ and $g_v < 0$. That also means we have to have $d > 1$. In the predator-prey context, as long as we still consider food to be the decisive element, the latter scenario with $g_u > 0$ means that u represents prey in a region sufficiently abundant in vegetation ($f_u > 0$) whereas v stands for a predator competing for food (and/or, quite commonly, territory; since $g_v < 0$). In this case an inhomogeneous steady state thus requires that the dispersion of the predator be faster than that of the prey. A local (random) increase in prey density is autocatalytic but also stimulates local growth of predator numbers. Due to faster dispersion, predators will then spread into neighbouring areas (before the prey does) where increased predator density will cause a drop in both predator and prey count ($f_v, g_v < 0$) so the pattern we expect should display a correlation between high (and low) densities of both the predator and its prey. Note that if it were the prey that dispersed faster, no such pattern would be possible; instead we could expect to see a 'wave' of increased prey density spreading from our 'source' area, stimulating a wave of predator abundance that would (possibly) restore the equilibrium.

On the other hand, the former case with $f_v > 0$ makes v count the prey while u counts the predator. We can assume that an increase in the prey numbers results in food shortages as $g_v < 0$, whereas higher concentration of predators has a positive effect on their hunting and/or reproductive effectiveness ($f_u > 0$). This time the prey needs to disperse faster for anything 'interesting' to happen. Now a locally higher prey density will provoke a local rise in predator numbers as well ($f_v > 0$), but will eventually drop naturally due to insufficient food supplies as well as predation ($g_v, g_u < 0$). Given that these effects are strong enough (note that the impact of cross-dependence has to 'prevail' since $|f_v g_u| > |f_u g_v|$ by (1.16)), it will drop well below the equilibrium level, inducing net influx of prey from the vicinity of the 'outbreak'. If the flow of the prey is fast enough, it will provide the higher number of predators with sufficient amount of food so that the latter can persist at their above-equilibrium concentration. On the contrary, predator count will drop in the neighbouring areas due to the outflow of prey, which in turn enables the prey to achieve higher density there. The flux of the prey from this high density areas into low prey/high predator density territories takes care of the self-inhibiting effect of increased prey numbers (e.g. local food shortages). Hence, we expect that high predator density should correlate with low prey concentration and vice versa should a permanent pattern be established. Note that in both cases the self-inhibiting species had to diffuse faster than the self-activating one. This is one of the well-known features of DDI and is often referred to as *short-range activation, long-range inhibition*.

Another notable point is that, beside the necessary condition for instability given in (1.19), we can in fact be quite precise about the *frequency scope* of the instability. Since $h(k^2)$ is a quadratic function with positive highest-order coefficient d, the range of unstable wavenumbers k is simply the region between the two roots of the equation $h(k^2) = 0$, i.e.

$$\begin{aligned} k_{min}^2 \equiv \frac{1}{2d}\left[df_u + g_v - \sqrt{(df_u + g_v)^2 - 4d|A|}\right] < k^2 \\ < \frac{1}{2d}\left[df_u + g_v + \sqrt{(df_u + g_v)^2 - 4d|A|}\right] \equiv k_{max}^2. \end{aligned} \tag{1.21}$$

This condition is equivalent to $\Re\lambda(k^2) > 0$ so it earmarks exactly the modes that grow exponentially in a first-order approximation, in contrast to modes with wavenumbers from outside the range determined in (1.21), which are exponentially damped. The bad news is that if an inhomogeneous steady state is to be established, an exponentially growing (and hence unbounded) solution is obviously not admissible,

and additional analysis has to be employed in order to make sure that the full, non-linear dynamics decelerates the initially exponential behaviour and restricts it to a bounded domain in the phase space (for a two-component system the positive quadrant of the (u, v)-plane). More precisely, we wish to find an *invariant region*, i.e. a closed subset Σ of the phase space with the property that if all the initial and boundary values lie in Σ, then so does the whole solution to the RD problem (1.12). The good news is that, according to a result by J. Smoller (see chapter 14 in [15]), in our case of a diagonal diffusion matrix, the same invariant region that contains the uniform (diffusionless) integral curves given as the solutions to

$$\frac{d\mathbf{c}}{dt} = \mathbf{f}(\mathbf{c}) \tag{1.22}$$

will also be an invariant region for the full RD system (1.12). The property that a subset of the phase space should have to be an invariant region is quite intuitive: whenever a solution (an integral curve) to the evolution equations (with given initial and boundary values) arrives at its boundary, it has to be turned back inwards by the dynamics of the system. In more technical terms, 'inwards' means a negative projection on the outward normal $\mathbf{n}$ to the boundary $\partial\Sigma$. The question is *what* should have a negative projection on $\mathbf{n}$ if we want to make sure our solution does not cross the boundaries of Σ. The obvious answer is the *tangent vector* to our integral curve $\mathbf{c}(t)$. Using Smoller's observation that (the absence of) diffusion does not play any role here, we can think of it as being given by (1.22) as $\mathbf{f}(\mathbf{c})$. Thus, the condition we need to verify in order to make sure that Σ is indeed an invariant region is

$$\mathbf{n} \cdot \mathbf{f}(\mathbf{c}) < 0 \text{ on } \partial\Sigma.$$

As long as this is true, any integral curve that arrives at the boundary of Σ will be 'repelled' back into its interior. Perhaps we could have written $\mathbf{n}(\mathbf{c})$ to emphasize that both factors vary with phase space position. It is conceivable that $\mathbf{n}$ may not even exist everywhere on the boundary since we did not require that $\partial\Sigma$ be smooth. In fact, probably the most natural choice of rectangular sets (with piecewise smooth boundary) of the form $\Sigma = \cap_{i=1}^{n}\{\mathbf{c} \mid a_i \leq c_i \leq b_i\}$ is not at all uncommon. Note that in a two-dimensional phase space (two-component RD system) in the absence of diffusion (i.e. if no inhomogeneity is present), by the Poincaré-Bendixson theorem our solution will either converge to a fixed point or, should this be unstable (or in case there are no fixed points at all within Σ), it will end up in a limit cycle. Unfortunately, we lose this strong a priori knowledge of the possible dynamics of the system once inhomogeneity kicks in.

The transition from stability to instability can occur in different ways. In our analysis, experimental as well as theoretical, we can therefore work with different bifurcation parameters. For instance, it is clear from (1.17) that the diffusion matrix will always have an impact on stability/instability of a RD system. Hence, in the two-component case, one of the possible bifurcation parameters is the (relative) diffusion coefficient d. It is quite straightforward that the threshold (critical) value $d_c(> 1)$, at which a transition towards instability should be expected, can be obtained by requiring equality in (1.19). It is thus given as the (larger) root d_c^+ of

$$f_u^2 d_c^2 + 2(2f_v g_u - f_u g_v)d_c + g_v^2 = 0 \Rightarrow$$
$$d_c^{\pm} = \frac{1}{f_u^2}\left[(f_u g_v - 2f_v g_u) \pm 2\sqrt{f_v^2 g_u^2 - f_u f_v g_u g_v}\right].$$

Note that by (1.16) both roots are always real and positive. However, using a Taylor expansion for the square root in the variable $\frac{f_u g_v}{f_v g_u}(< 1)$ and the conditions (1.20) it can be shown that d_c^- is not eligible. Consequently, the critical value for d is indeed uniquely given by d_c^+. The associated critical wavenumber

will be given as

$$k_c^2 = \frac{d_c f_u + g_v}{2d_c} = \sqrt{\frac{|A|}{d_c}}.$$

Other possible bifurcation parameters could include the size of our domain Ω. For example, in the case of a line of length L and a rectangle $a \times b$, the spatial eigenvalues are given as $k_n = \frac{n\pi}{L}$ and $k_{l,n} = \pi\sqrt{\frac{l^2}{b^2} + \frac{n^2}{a^2}}$, respectively. [10] For L ($\max\{a, b\}$) small enough, the second inequality in (1.21) could fail even for the smallest integers, leaving the system with no unstable modes. Recall that the (uniform) modes corresponding to the wavenumbers $k_0 = 0 = k_{0,0}$ are always stable by (1.16) (or simply by the observation that $k_{min}^2 > 0$). The length L or $\max\{a, b\}$ could thus be viewed as a bifurcation parameter, with the critical values corresponding to the case $\frac{\pi}{L} = k_1 = k_{max}$ on the line and $\pi\frac{l}{b} = k_{1,0} = k_{max}$ (for $a < b$) on the rectangle. For these values (or just beyond them) the first unstable modes, corresponding to these unstable wavenumbers, should appear in the system. If the dimensions of the domains are smaller, Turing instability is impossible.

So far we have simplified our discussion by silently assuming that instability would occur in the system whenever $h(k^2) < 0$ was satisfied for some k or, equivalently, whenever (1.19) was true. Another possible equivalent formulation is that the unstable region given by (1.21) be non-empty. There are several reasons why that is a significant simplification. Firstly, as we argue in Appendix A, the spectrum of the Laplace operator on a bounded domain is purely discrete. Thus the condition $k_{min} < k_{max}$ may not be sufficient. To provide the system with an unstable mode, at least *one of the* discrete *eigenvalues* must lie in the 'region of instability' between k_{min} and k_{max}. Secondly, the unstable mode needs to 'be present' in the system. We mentioned already that from a conceptual point of view inhomogeneity is usually thought of as appearing in the system via random perturbations, especially in biological/embryogenetical context. From a mathematical point of view, these are inscribed in the system by means of non-uniform initial conditions. In first-order approximation the solution to RD equations is given as a superposition of spatial modes (eigenfunctions of the Laplace operator) with exponential dynamics. Thus at $t = 0$ this superposition has to reduce to the decomposition of the initial condition in the Laplace eigenbasis. This should very much remind us of the procedure applied when solving (the somehow closely related) wave equation, e.g. on a string. The dynamics is known in advance (in that case it is harmonic) but to determine whether a specific mode does appear in the solution one needs to investigate whether it was 'present' in the initial conditions using a Fourier decomposition. Hence, if the range of unstable eigenvalues does not match the perturbation, any 'algebraic conformity' of the system could be fruitless and we would observe no instability, at least as far as first-order effects are concerned. This is also one of the possible explanations for the specific patterns associated with specific species, should these indeed come into being via a reaction diffusion mechanism. The mechanism could be the same, yet a different course of the respective species' developmental processes could mean a different 'starting point', resulting in different patterns. Let us note, however, that unlike in chemistry or other disciplines, there has only been limited amount of experimental proof for the presence of reaction diffusion mechanisms in pattern formation in developmental biology. On the theoretical level, on the other hand, the Turing model is ready to take credit for a large variety of pattern formation, ubiquitous in biology and beyond.

Chapter 2

Weakly non-linear analysis. Pattern selection

The term weakly non-linear analysis is usually used to refer to a set of methods and approaches that largely use the assumptions and results of linear analysis on the one hand but go beyond it and take into consideration the lowest-order non-linear effects to improve these results on the other hand. We would like to use this chapter to introduce one of these methods and apply it in the biologically very relevant context of pattern selection. As it turns out, immediately beyond criticality (as it was described in chapter 1.1), which is where instability enters the system, the selection of pattern is governed precisely by those lowest-order non-linear effects that linear analysis fails to consider.

Conditions (1.16) guarantee stability of the system (1.12) with respect to homogeneous perturbations of the steady state. The scope of unstable wavenumbers is given by the two inequalities in (1.21). As long as the lowest wavenumber k_1 satisfies $k_1 > k_{max}$, all the modes are obviously stable. Criticality, or loss of hyperbolicity, occurs once k_1, due to a modification of the system's parameters, reaches the borderline of instability, $k_1 = k_{max}$, or exceeds it slightly. Throughout this chapter, our focus will aim at this very situation. In this case the temporal eigenvalue of the first mode, $\tilde{\lambda} \equiv \lambda(k_1)$, becomes 'marginally' positive, which is why we will denote it as $\tilde{\lambda} = \delta\lambda$ for a small positive parameter δ. On a sphere, where the spatial eigenvalues are given as $k_l = \frac{l(l+1)}{R^2}$, this can result from increasing the radius R slightly above a critical value R_c. As we have stated in the earlier chapters, as soon as hyperbolicity is lost, the methods of linear analysis become insufficient. We therefore adopt here a weakly non-linear approach motivated by the article by B. Ermentrout. [7]

Let us consider a system of the general form

$$\mathcal{L}u = \tilde{\lambda}\mathcal{B}u + Q(u,u) + C(u,u,u) + \dots \tag{2.1}$$

for a linear operator $\mathcal{L}$ acting on a Hilbert space $\mathcal{H}$ and with $\mathcal{B}, Q$, and C a linear, bilinear and trilinear form acting on $\mathbb{R}^n$, respectively, while the dots represent higher-order terms. By 'acting on $\mathbb{R}^n$' we mean that if $\mathcal{H}$ is a function space, for instance $\mathcal{H} = \mathbb{R}^n \otimes L^2(\Omega)$ and $u_i \in \mathcal{H}$ are of the form $u_i = \mathbf{a}_i f_i(x)$ for some $\mathbf{a}_i \in \mathbb{R}^n$ and $f_i \in L^2(\Omega)$, then $\mathcal{B}, Q$, and C satisfy

$$\begin{aligned} &(\mathcal{B}u_i)(x) = (B\mathbf{a}_i)f_i(x), \quad Q(u_i,u_j)(x) = Q(\mathbf{a}_i,\mathbf{a}_j)f_i f_j(x), \\ &C(u_i,u_j,u_k)(x) = C(\mathbf{a}_i,\mathbf{a}_j,\mathbf{a}_k)f_i f_j f_k(x) \end{aligned} \tag{2.2}$$

for some linear map (matrix) $B : \mathbb{R}^n \to \mathbb{R}^n$, some bilinear map $Q : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ and some trilinear map $C : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$. For our purposes, $\tilde{\lambda}$ in (2.1) represents the (exponential) growth rate of

the eigenmodes of $A - k_1 D$ and, since we consider a state near criticality, we set $\tilde{\lambda} = \delta\lambda$. For reaction-diffusion systems (1.12) we also have

$$\mathcal{L} = -D^1 F(u_*, v_*) - D\Delta, \quad Q = D^2 F(u_*, v_*), \quad C = D^3 F(u_* v_*),$$

where the diagonal matrix of diffusion coefficients D is not to be confused with the expressions of the form $D^m F$, which symbolically represent corresponding terms of the Taylor expansion of $F = \begin{pmatrix} f \\ g \end{pmatrix}$. Multicomponent Turing systems could, of course, be rewritten in an analogous fashion. For any number of components (greater than one) it is convenient to introduce a new set of variables $u_1 \equiv u - u_*, u_2 \equiv v - v_*$ and the 'tensor-like' notation for the stability matrix $A \equiv D^1 F(0,0) = (\nabla F(0,0))^T$ (now A acts as a vector in $\mathbb{R}^2 \otimes \mathbb{R}^2$ and Ax can be rewritten as $\nabla F(0,0) \cdot x$, with $\cdot$ the scalar product on $\mathbb{R}^2$) and for the Hessian matrix $D^2 F \equiv H$ of F whose components are elements of $\mathbb{R}^2$ (more generally, for n components they are elements of $\mathbb{R}^n$) of the form

$$H_{ij} = \frac{\partial^2 F}{\partial u_i \partial u_j}(0,0) \equiv \begin{pmatrix} \frac{\partial^2 f}{\partial u_i \partial u_j}(0,0) \\ \frac{\partial^2 g}{\partial u_i \partial u_j}(0,0) \end{pmatrix}$$

(here we avoided using different symbols for $F(u,v)$ and $F(u_1, u_2)$). With this notation, H formally lives in $\mathbb{R}^2 \otimes \mathbb{R}^2 \otimes \mathbb{R}^2$ and we can rewrite the second-order terms from the Taylor expansion of F as $Q(w,w) = w^T H w$ for $w \equiv \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$. Hence, the condition (2.2) for Q is manifestly satisfied. Let us also write the explicit form of the cubic term

$$D^3 F = \frac{1}{3!} \sum_{l=0}^{3} \binom{3}{l} \frac{\partial^3 F}{\partial^l u_1 \partial^{3-l} u_2} u_1^l u_2^{3-l},$$

which is the highest-order term that we are going to use explicitly.

The criticality condition $k_1 = k_{max}$ implies that the matrix $A - k_1 D$ has a zero eigenvalue, which means $\det(A - k_1 D) = 0$. [1] It also means that this matrix has a non-zero eigenvector associated with the eigenvalue 0; we shall denote it $\mathbf{e}$. For later purposes, we also note that by the invariance of determinant under transposition, $A^T - k_1 D$ also has a non-zero eigenvector $\mathbf{f} \in \ker\left(A^T - k_1 D\right)$, which we assume can be chosen such that $|\mathbf{e} \cdot \mathbf{f}| = 1$. [2] By the Fredholm alternative, this is equivalent to assuming that $(A - k_1 D)\,\mathbf{x} = \mathbf{e}$ has no solution in $\mathbb{R}^n$. It follows from criticality that all the modes corresponding to wavenumbers $k_n > k_1, n \neq 1$, are stable. Hence, any admissible pattern will be given as a combination of the modes corresponding to k_1; let us denote these as $\{w_{11}, w_{12}, \ldots w_{1m}\}$. Possible symmetries of the pattern will thus depend on the coefficients of these modes in the solution. For instance, in a rectangle $a \times b$, where the lowest eigenmodes are given as $\cos\left(\frac{\pi}{a}x\right), \cos\left(\frac{\pi}{b}y\right)$, the pattern will be either striped, if only one of these modes is represented non-trivially in the solution, or spotted, if both are. We would therefore like to know how to investigate the properties and stability of either case with respect to outer characteristics (parameters) of the system.

Let us consider a unit sphere. As mentioned before, the spectrum of the Laplace operator is given as $k_l = \frac{l(l+1)}{R^2}$ and the corresponding eigenfunctions are the spherical harmonics

$$\tilde{Y}_{lm}(\vartheta, \varphi) = (-1)^m \sqrt{\frac{2l+1}{4\pi} \frac{(l-|m|)!}{(l+|m|)!}} P_l^{|m|}(\cos\vartheta) e^{im\varphi},$$

[1] The opposing signs result from the fact that $k_1 > 0$ is an eigenvalue of $-\Delta$.

[2] We are going to choose the sign of $\mathbf{e} \cdot \mathbf{f}$ later.

where the associated Legendre functions are given as $P_l^m(x) = (1-x^2)^{m/2}\frac{d^m P_l}{dx^m}$, with $P_l(x) = \sum_{k=0}^{\lfloor l/2 \rfloor}(-1)^k \frac{(2l-2k)!}{2^l k!(l-k)!(l-2k)!}x^{l-2k}$ being the Legendre polynomials (cf. [10]). For the rest of the calculation we will prefer to work with the real basis

$$\begin{aligned} Y_{lm} &\equiv (-1)^m \sqrt{2}\Re\left(\tilde{Y}_{lm}\right) \quad m > 0, \\ Y_{l0} &\equiv \tilde{Y}_{l0}, \\ Y_{lm} &\equiv (-1)^m \sqrt{2}\Im\left(\tilde{Y}_{l|m|}\right) \quad m < 0. \end{aligned}$$

We will usually omit the normalization constants and hence by Y_{lm} we will mean the following: $Y_{11} \equiv \sin\theta\cos\varphi, Y_{1-1} \equiv \sin\theta\sin\varphi, Y_{10} \equiv \cos\theta$ and analogously for $l > 1$. The near-critical pattern will thus be given as a combination of these three eigenmodes corresponding to k_1. Each of these functions clearly has a rotational symmetry with respect to plains given (in Cartesian coordinates) as $x_i = const, i \in \{1,2,3\}$. It is easy to show that

$$s\sin\theta\cos\varphi + t\sin\theta\sin\varphi + r\cos\theta = sY_{11} + tY_{1-1} + rY_{10} \equiv Y_1,$$

has a rotational symmetry with respect to rotations around an axis given in Cartesian coordinates by the vector (s,t,r). We will provide the proof at the end of this chapter. Note that the norm of each of these three functions is equal to $\sqrt{\frac{4\pi}{3}}$, so neglecting it is equivalent to computing $\tilde{s} = \sqrt{\frac{4\pi}{3}}s, \tilde{r} = \sqrt{\frac{4\pi}{3}}r, \tilde{t} = \sqrt{\frac{4\pi}{3}}t$ instead of s,r,t. Before we proceed to the actual solution, we make one more assumption: for simplicity we assume that the critical value is $k_1 = 1$, which means $k_2 = 3, k_3 = 6, \ldots, k_l = \frac{l(l+1)}{2}$.

As mentioned before, we assume $\tilde{\lambda} \equiv \delta\lambda$ small and look for a solution to (2.1) in the form

$$w(\theta,\varphi) = \delta^\alpha w_1 + \delta^{2\alpha}w_2 + \delta^{3\alpha}w_3 + \ldots. \tag{2.3}$$

Substituting this expansion into (2.1), we obtain

$$\begin{aligned} \mathcal{L}\left(\delta^\alpha w_1 + \delta^{2\alpha}w_2 + \delta^{3\alpha}w_3 + \ldots\right) = \quad & \lambda\mathcal{B}\left(\delta^{\alpha+1}w_1 + \delta^{2\alpha+1}w_2 + \delta^{3\alpha+1}w_3 + \ldots\right) + \\ & + \delta^{2\alpha}Q(w_1,w_1) + 2\delta^{3\alpha}Q(w_1,w_2) + \ldots + \\ & + \delta^{3\alpha}C(w_1,w_1,w_1) + \ldots, \end{aligned} \tag{2.4}$$

so for the lowest order δ^α we obtain $\mathcal{L}w_1 = 0$. At this moment it is important to realize that if we decompose w into the eigenbasis Y_{lm}, $\mathcal{L}$ will act on each subspace corresponding to a fixed wavenumber k_l as just a matrix, namely as $L_l \equiv -A + k_l D$. Recalling that these matrices have non-zero determinant for $l \neq 1$ (including $L_0 = -A$, by (1.16)), we conclude that any non-trivial solution for w_1 must lie in the subspace corresponding to k_1, so the most general solution is

$$w_1 = \mathbf{e}(sY_{11} + tY_{1-1} + rY_{10}) = \mathbf{e}Y_1.$$

We have not yet discussed the parameter α. Its value becomes relevant at the next order. More precisely, it plays a decisive role when we ask what order is the second lowest in δ. The lowest order in the first term on the right-hand side is $\delta^{\alpha+1}$, in the remainder of the equation (apart from the δ^α-term) it is $\delta^{2\alpha}$. Let us be naive for a while and try to combine all these terms by setting $\alpha = 1$. That would lead to

$$\mathcal{L}w_2 = \lambda\mathcal{B}w_1 + Q(w_1,w_1). \tag{2.5}$$

At first glance, this equation looks good. We already know what w_1 is, so we just need to 'invert' the operator on the left-hand side and apply it to the explicitly given function on the right-hand side. This, as we

will see, is not possible. For a solution for w_2 to exist, the right-hand side must necessarily lie in Ran $\mathcal{L}$. From the Fredholm alternative we have $\ker \mathcal{L}^* = (\text{Ran}\,\mathcal{L})^\perp$. Hence, the solvability condition requires that the right-hand side be orthogonal to $\ker \mathcal{L}^* = \text{span}\{\mathbf{f}Y_{11}, \mathbf{f}Y_{10}, \mathbf{f}Y_{1-1}\}$. Recalling that Y_{lm} and Y_{jk} are orthogonal for $(l,m) \neq (j,k)$, we see that it would be useful to express the term $Q(w_1, w_1) = \mathbf{e}^T H\mathbf{e}(Y_1)^2$ in terms of the eigenfunctions Y_{lm}. There are at least two ways of doing this: one can either apply all sorts of trigonometric identities to express the underlying functions in the way they are represented in the spherical harmonics Y_{lm} (i.e as constant multiples of $P_l^{|m|}(\cos\theta)\sin(m\varphi)$ or $P_l^{|m|}(\cos\theta)\cos(m\varphi)$) or, if one only needs to know the projections of a given function $f(\theta,\varphi)$ on Y_{lm} for a specific set of values of (l,m) (which is our case here), it is often easier to compute these projections via the scalar product[3]

$$\langle f, Y_{lm}\rangle = \int_0^{2\pi} d\varphi \int_0^{\pi} d\theta \sin\theta Y_{lm}(\theta,\varphi) f(\theta,\varphi).$$

However, with the latter procedure one should remember to reintroduce the normalizing constants into the calculation to obtain the correct results. After some tedious algebra we obtain

$$\begin{aligned}(Y_1)^2 = &\left[\tfrac{s^2+r^2+t^2}{3} + \tfrac{s^2-t^2}{2}\sin^2\theta\cos 2\varphi + \left(\tfrac{r^2}{3} - \tfrac{s^2+t^2}{6}\right)(3\cos^2\theta - 1) + \right.\\ &\left. + rs\sin\theta\cos\theta\cos\varphi + rt\sin\theta\cos\theta\sin\varphi + \tfrac{st}{2}\sin^2\theta\sin 2\varphi\right] \equiv\\ \equiv &\left[\tfrac{s^2+r^2+t^2}{3}Y_{00} + \tfrac{s^2-t^2}{2}Y_{22} + \left(\tfrac{r^2}{3} - \tfrac{s^2+t^2}{6}\right)Y_{20} + rsY_{21} + rtY_{2-1} + \tfrac{st}{2}Y_{2-2}\right].\end{aligned}$$

While the second term on the right-hand side of (2.5) is thus clearly orthogonal to $\ker \mathcal{L}^*$, the first term $\lambda\mathcal{B}w_1$ is generally not unless we assume $\mathbf{f}\cdot\mathcal{B}\mathbf{e} = 0$,[4] so equation (2.5) is not guaranteed to have a solution. The remedy is at hand, though: we conclude that $\alpha \neq 1$, which leads to the correct (for now, at least, solvable) equation

$$\mathcal{L}w_2 = \mathbf{e}^T H\mathbf{e}\Big[\frac{s^2+r^2+t^2}{3}Y_{00} + \frac{s^2-t^2}{2}Y_{22} + \left(\frac{r^2}{3} - \frac{s^2+t^2}{6}\right)Y_{20} + rsY_{21} + rtY_{2-1} + \frac{st}{2}Y_{2-2}\Big]. \tag{2.6}$$

A foresighted reader senses that we are going to need an explicit expression for w_2 later, so we need to invert this relation. This, however, is not that difficult now that we have expressed the right-hand side as a combination of the eigenfunctions Y_{lm}. Having done this bit of algebra, we recall that $\mathcal{L}$ acts on each subspace of the form $W_l \equiv \text{span}\{Y_{lm}\}_{m=-l}^{l}$ as a (for $l \neq 1$ invertible!) matrix L_l, since we have

$$\mathcal{L}(\mathbf{v}Y_{lm}) = (-A - D\Delta)(\mathbf{v}Y_{lm}) = (-A\mathbf{v})Y_{lm} + k_l(D\mathbf{v})Y_{lm} \equiv (L_l\mathbf{v})Y_{lm}.$$

All of these subspaces are clearly invariant under $\mathcal{L}$. Using all these observations and introducing an abridged notation $Q(\mathbf{e},\mathbf{e}) = \mathbf{e}^T H\mathbf{e} \equiv \mathbf{q}$, we can invert equation (2.6) into

$$w_2 = \frac{s^2+r^2+t^2}{3}L_0^{-1}\mathbf{q} + \left[\frac{s^2-t^2}{2}Y_{22} + \left(\frac{r^2}{3} - \frac{s^2+t^2}{6}\right)Y_{20} + rsY_{21} + rtY_{2-1} + \frac{st}{2}Y_{2-2}\right]L_2^{-1}\mathbf{q}.$$

Now we can proceed to the next order in δ. Since we ruled out $\alpha = 1$, the 'next best guess' is to associate the $\delta^{\alpha+1}$-term with the $\delta^{3\alpha}$-terms, i.e. to put $\alpha = 1/2$. In fact, this is the best guess. Not only is this choice algebraically admissible, as we will see in the following lines, it also makes sense physically. The dispersion relation (1.18) gives an equation for the temporal eigenvalue λ that is quadratic in the wavenumber k; it is therefore logical that if λ scales as δ, the spatial modes should scale as $\delta^{1/2}$. For

[3] Here we make use of the fact that for a Hilbert space $\mathcal{H}$ and an orthonormal basis $\{u_i\}$ of $\mathcal{H}$ one can (uniquely) express any $x \in \mathcal{H}$ as $x = \sum_i \langle x, u_i\rangle u_i$.

[4] Such an assumption would be very restrictive, since for example $\mathcal{B} = -\mathbb{I}$ is perfectly plausible for our cause.

convenience we can (and often do) introduce a new parameter $\varepsilon = \delta^{1/2}$ instead of δ and write $\tilde{\lambda} = \varepsilon^2\lambda$. Hence, to the next order we have

$$\mathcal{L}w_3 = \lambda\mathcal{B}w_1 + 2Q(w_1, w_2) + C(w_1, w_1, w_1) \equiv w_{rhs}.$$

To be able to obtain the sought-after equations for s, t and r, we need to apply the solvability condition one more time. In order to make the long and tedious calculation easier to follow, we again demonstrate how the quantities and coefficients that will appear in the scalar product came about. Again, we would like to express w_{rhs} in terms of Y_{lm}, particularly for $l = 1$. The second, quadratic term in w_{rhs}, contains products of the form $Y_1 Y_{2m}$. We demonstrate their decomposition for $m = 0$:

$$\begin{aligned} Y_{20}Y_1 &= 3(s\cos^2\theta\sin\theta\cos\varphi + r\cos^3\theta + t\cos^2\theta\sin\theta\sin\varphi) - Y_1 \\ &= -\frac{2}{5}sY_{11} + \frac{4}{5}rY_{10} - \frac{2}{5}tY_{1-1} + \sum_{l\neq 1}\sum_{|m|\leq l} a_{lm}Y_{lm}, \end{aligned}$$

where we do not bother to write down (or even compute) explicitly the coefficients of Y_{lm} for $l \neq 1$. This way, upon again abbreviating our notation by introducing

$$\mathbf{p}_i \equiv Q\left(\mathbf{e}, L_i^{-1}Q(\mathbf{e}, \mathbf{e})\right) = \mathbf{e}^T H L_i^{-1}\mathbf{q},$$

we obtain

$$\begin{aligned} 2Q(w_1, w_2) =& \mathbf{p}_0\Big[\frac{2}{3}(s^2 + t^2 + r^2)\Big(sY_{11} + tY_{1-1} + rY_{10}\Big)\Big] + \\ &+\mathbf{p}_2\Big[(s^2 - t^2)\Big(\frac{2}{5}sY_{11} - \frac{2}{5}tY_{1-1} + \ldots\Big) + \\ &+\frac{2r^2 - s^2 - t^2}{3}\Big(-\frac{2}{5}sY_{11} - \frac{2}{5}tY_{1-1} + \frac{4}{5}rY_{10} + \ldots\Big) + \\ &+2sr\Big(\frac{1}{5}rY_{11} + \frac{1}{5}sY_{10} + \ldots\Big) + 2rt\Big(\frac{1}{5}rY_{1-1} + \frac{1}{5}tY_{10} + \ldots\Big) + st\Big(\frac{2}{5}tY_{11} + \frac{2}{5}sY_{1-1} + \ldots\Big)\Big], \end{aligned}$$

where the dots again stand for (constant) multiples of Y_{lm} with $l \neq 1$. Once we have decomposed the term $C(w_1, w_1, w_1) = C(\mathbf{e}, \mathbf{e}, \mathbf{e})(Y_1)^3$ in an analogous fashion, we introduce the (scalar) quantities

$$\begin{aligned} \alpha_0 &\equiv \mathbf{f}\cdot\mathbf{p}_0 = \mathbf{f}\cdot Q\left(\mathbf{e}, L_0^{-1}Q(\mathbf{e}, \mathbf{e})\right), \\ \alpha_2 &\equiv \mathbf{f}\cdot\mathbf{p}_2 = \mathbf{f}\cdot Q\left(\mathbf{e}, L_2^{-1}Q(\mathbf{e}, \mathbf{e})\right), \\ \beta &\equiv \mathbf{f}\cdot C(\mathbf{e}, \mathbf{e}, \mathbf{e}), \\ \mu &\equiv \mathbf{f}\cdot\mathcal{B}\mathbf{e}, \end{aligned}$$

and apply the solvability condition to be able to write

$$\begin{aligned} 0 \stackrel{!}{=} \langle\mathbf{f}Y_{11}, w_{rhs}\rangle &= \alpha_0\frac{2}{3}s(s^2 + t^2 + r^2) + \alpha_2\left[\frac{2}{5}s(s^2 - t^2) - \frac{4}{5}s\left(\frac{r^2}{3} - \frac{s^2 + t^2}{6}\right) + \frac{2}{5}sr^2 + \frac{2}{5}st^2\right] + \\ &+\beta\left[\frac{3}{4}s^3 + \frac{3}{5}s(t^2 + r^2)\right] + \mu\lambda s = \\ &= s\left(\mu\lambda + \left[\frac{2}{3}\alpha_0 + \frac{8}{15}\alpha_2 + \frac{3}{4}\beta\right]s^2 + \left[\frac{2}{3}\alpha_0 + \frac{2}{15}\alpha_2 + \frac{3}{5}\beta\right]r^2 + \left[\frac{2}{3}\alpha_0 + \frac{2}{15}\alpha_2 + \frac{3}{5}\beta\right]t^2\right). \end{aligned}$$

Expressing the scalar products $\langle\mathbf{f}Y_{10}, w_{rhs}\rangle$ and $\langle\mathbf{f}Y_{1-1}, w_{rhs}\rangle$ in a similar manner and finally setting

$$z_1 \equiv \frac{2}{3}\alpha_0 + \frac{8}{15}\alpha_2 + \frac{3}{5}\beta, \qquad z_2 \equiv \frac{2}{3}\alpha_0 + \frac{2}{15}\alpha_2 + \frac{3}{5}\beta,$$

we find the coveted equations for s, t, r to be

$$\begin{aligned} s\left(\mu\lambda + z_1 s^2 + z_2\left(r^2 + t^2\right)\right) &= 0, \\ t\left(\mu\lambda + z_1 t^2 + z_2\left(r^2 + s^2\right)\right) &= 0, \\ r\left(\mu\lambda + z_1 r^2 + z_2\left(s^2 + t^2\right)\right) &= 0. \end{aligned} \tag{2.7}$$

Since the functions Y_{11}, Y_{1-1}, Y_{10} correspond to the three Cartesian coordinates on the unit sphere, it comes as no surprise that the equations are symmetric with respect to any permutation of $\{s, t, r\}$. As we mentioned at the beginning of this chapter, we would like to investigate possible (stable) stationary states and thus possible axes of symmetry of the solution to a given system, ı.e for fixed values of z_1, z_2. First of all, we need evolution equations for s, t and r. It could seem that we replaced the temporal derivative in (2.1) (if the system is to represent an evolution equation at all) by the term $\tilde{\lambda}\mathcal{B}u$, assuming exponential dynamics of the form $u = \tilde{u}e^{\tilde{\lambda}\tau}$. If that is true, we are left with no natural way of obtaining evolution equations for (s, t, r). The second part of the statement is indeed true, but its first part is only partially true. In fact, by setting $\tilde{\lambda} = \delta\lambda$ we effectively introduced a new time scale $T = \delta\tau$ to the problem. Hence, the actual assumption we made was $u(\tau, \mathbf{x}) = \tilde{u}(T(\tau), \mathbf{x})e^{\tilde{\lambda}\tau}$. For the 'fast' time scale τ we assumed the exponential dynamics that follows from linearization, while preserving the 'slow' time scale T for the amplitudes. Applying the chain rule, we then have

$$\frac{\partial u}{\partial \tau} = \tilde{\lambda}\tilde{u}(T(\tau), \mathbf{x})e^{\tilde{\lambda}\tau} + \delta\frac{\partial \tilde{u}}{\partial T}(T(\tau), \mathbf{x})e^{\tilde{\lambda}\tau}.$$

The first term has been with us the whole time. The lowest order in δ in the second term, which is

$$\delta^{\alpha+1}\frac{\partial w_1}{\partial T} = \delta^{\alpha+1}\mathbf{e}\left(\dot{s}Y_{11} + \dot{t}Y_{1-1} + \dot{r}Y_{10}\right),$$

with dot representing derivatives with respect to T, gives us precisely what we are looking for, namely the evolution equations for the amplitudes $s(T), t(T), r(T)$

$$\begin{pmatrix} \dot{s} \\ \dot{t} \\ \dot{r} \end{pmatrix} = \begin{pmatrix} s\left(\mu\lambda + z_1 s^2 + z_2\left(r^2 + t^2\right)\right) \\ t\left(\mu\lambda + z_1 t^2 + z_2\left(r^2 + s^2\right)\right) \\ r\left(\mu\lambda + z_1 r^2 + z_2\left(s^2 + t^2\right)\right) \end{pmatrix} \equiv V(r, t, s).$$

Note that there is nothing indecent about this: had we assumed the dependence on T in the amplitudes from the very beginning, and more generally the expansion (2.3) as

$$w(T, \theta, \varphi) = \delta^{\alpha}w_1(T, \theta, \varphi) + \delta^{2\alpha}w_2(T, \theta, \varphi) + \delta^{3\alpha}w_3(T, \theta, \varphi) + \ldots,$$

then instead of (2.7), the orthogonality condition would have presented us directly with the system

$$\begin{aligned} \dot{s} - s\left(\mu\lambda + z_1 s^2 + z_2\left(r^2 + t^2\right)\right) &= 0, \\ \dot{t} - t\left(\mu\lambda + z_1 t^2 + z_2\left(r^2 + s^2\right)\right) &= 0, \\ \dot{r} - r\left(\mu\lambda + z_1 r^2 + z_2\left(s^2 + t^2\right)\right) &= 0. \end{aligned}$$

This way we just saved ourselves some unnecessary work with terms including derivatives with respect to T. The stability analysis then proceeds completely in compliance with chapter 1.1: stability or instability of a stationary state is determined by the eigenvalues of its stability matrix. For instance, for the

equilibrium $\begin{pmatrix} s_* \\ t_* \\ r_* \end{pmatrix} = \begin{pmatrix} \sqrt{\frac{-\mu\lambda}{z_1}} \\ 0 \\ 0 \end{pmatrix} \equiv \boldsymbol{s}_*$, we must have
$\operatorname{sgn}\mu = -\operatorname{sgn} z_1$ and the stability matrix

$$DV(s_*, t_*, r_*) = \begin{pmatrix} 2z_1 s_*^2 & 0 & 0 \\ 0 & \mu\lambda + z_2 s_*^2 & 0 \\ 0 & 0 & \mu\lambda + z_2 s_*^2 \end{pmatrix}\Bigg|_{\boldsymbol{s}_*} = \begin{pmatrix} -2\mu\lambda & 0 & 0 \\ 0 & \mu\lambda\left(1 - \frac{z_2}{z_1}\right) & 0 \\ 0 & 0 & \mu\lambda\left(1 - \frac{z_2}{z_1}\right) \end{pmatrix}$$

is diagonal so we can read the stability conditions directly as

$$\mu > 0 \wedge z_2 < z_1 < 0.$$

This is the right moment to specify the sign of $\mathbf{e} \cdot \mathbf{f}$ (to choose the sign of $\mathbf{f}$ such that we indeed have $\mu > 0$.

Note that this calculation is more useful for problems with a weaker level of symmetry than we have. In the case of a unit sphere no proper pattern selection is done since all possible patterns (i.e. each of the three modes and all their linear combinations) have the same (rotational) symmetry. It is just the axis of this symmetry that varies for different stationary states. Let us prove this proposition properly. We want to show that the function $f(\theta, \varphi) \equiv sY_{11} + tY_{1-1} + rY_{10}$ has rotational symmetry (here $s, t, r \in \mathbb{R}$). Having rotational symmetry is equivalent to being constant on an infinite set of circles (on the unit sphere) whose plains are mutually parallel (e.g. all circles with a vertical axis). But the equation $f = M = const$ can be rewritten in Cartesian coordinates as $sx + ty + rz = M$, which clearly defines a plain

$$W \equiv \{(x, y, z) \in \mathbb{R}^3 \mid sx + ty + rz = M\}.$$

Moreover, these plains are obviously parallel for various values of M as they have empty intersections. Since the intersection of a plain and a sphere is a circle (or a single point, as long as it is non-empty), this proves rotational symmetry of f. If we want to determine the axis of these rotations, we just need to pick one of these plains and find a line $\gamma(z) \equiv (Az, Bz, z)$, parametrized for example by its z-coordinate, that is perpendicular to two linearly independent vectors from that plain. Choosing the plain with $M = 0$ and the vectors $v_1 \equiv (-\frac{t}{s}, 1, 0)$ and $v_2 = (-\frac{r}{s}, 0, 1)$ reveals that $\gamma(t) = (\frac{s}{r}z, \frac{t}{r}z, z)$ which proves that the vector (s, t, r) is indeed the axis of the rotational symmetry of f as we stated earlier.

Chapter 3

The WKBJ approximation for a system of ODEs

The WKBJ method, developed independently by the physicists Gregor Wentzel, Hendrik A. Kramers and Léon N. Brillouin in 1926 as well as by the mathematician Harold Jeffreys specifically for second-order equations three years earlier, is a powerful method for obtaining approximate solutions to linear differential equations whose highest-order derivative is multiplied by a small parameter. Unlike with some other methods, the product of a WKBJ approximation is an *asymptotic series*. The term is in a sense a generalization of the notion of convergent series: every convergent power series, e.g. a convergent Taylor series, is also an asymptotic series but the converse is not true. In fact, when we talk about asymptotic series, we usually mean that the series in question is indeed divergent. The key step that gives a meaning to such an approximation is *truncation*. As with expansions of any kind, we usually only use a finite (and rather small) number of terms. However, in case of a Taylor series of a well-behaved function, truncation occurs at the expense of precision; the more terms we take into consideration, the better precision we obtain. This is obviously not true for a divergent asymptotic expansion. Here we not only truncate for practical reasons, we also truncate it for the sake of precision. As we will argue later in this chapter, that precision can be astounding.

3.1 Asymptotic series

Let us now formulate the idea of an asymptotic expansion more precisely. An *asymptotic sequence* at x_0 is a sequence of functions $\{\varphi_n\}_{n=0}^{\infty}$ such that $\varphi_{n+1} \ll \varphi_n$ for $x \to x_0$ or, equivalently, a sequence satisfying for all $n \in \mathbb{N}$

$$(\forall \varepsilon > 0)(\exists H_{x_0})(\forall x \in H_{x_0})(|\varphi_{n+1}(x)| \leq \varepsilon |\varphi_n(x)|), \tag{3.1}$$

where x_0 need not be finite. The 'little-o notation' $\varphi_{n+1}(x) = o(\varphi_n(x))$ is also common. Note that $\varphi_n(x) = (x - x_0)^n$ is an asymptotic sequence for x_0 finite and $\varphi_n(x) = x^{-n}$ is one for $x_0 = \pm\infty$. Once we have an asymptotic sequence, an *asymptotic series* can be constructed as $\sum_n a_n \varphi_n(x)$ using any sequence $\{a_n\}_{n=0}^{\infty}$. Apart from 'pathological' cases like $\varphi_{2n+1}(x) \equiv 0$, $\varphi_{2n}(x) = f_n(x)$, this definition of an asymptotic series includes most convergent series and, as argued above, all convergent power series since all power series satisfy condition (3.1). However, one of the generalizations that the notion of asymptotic series provides us is the possibility of using expansions with non-integer powers of x. Series of the form $\sum_n a_n x^{\alpha n}$ (i.e.

$\varphi_n(x) = x^{\alpha n}$) comply with the requirement on asymptotic series in (3.1) for $x \to 0$[1] and $\alpha > 0$, and the same applies to $\sum_n a_n x^{-\alpha n}$ for $|x| \to \infty, \alpha > 0$.

Having established the notion of an asymptotic series, let us now clarify what we mean by an *asymptotic expansion of a function*. We say that the series $\sum_{n=0}^{\infty} a_n \varphi_n(x)$ is asymptotic to $y(x)$ and write

$$y(x) \sim \sum_{n=0}^{\infty} a_n \varphi_n(x) \tag{3.2}$$

iff

$$y(x) - \sum_{n=0}^{N} a_n \varphi_n(x) \ll \varphi_N(x) \text{ for } x \to x_0 \text{ and any } N. \tag{3.3}$$

The validity of this statement *for any N* makes this equivalent to the statement that $y(x) - \sum_{n=0}^{N} a_n \varphi_n(x) \sim a_{N+1}\varphi_{N+1}(x)$, given that $a_{N+1} \neq 0$, or, more generally, equivalent to

$$y(x) - \sum_{n=0}^{N} a_n \varphi_n(x) \sim a_M \varphi_M(x) \text{ for } x \to x_0, \tag{3.4}$$

where $M > N$ is the next integer with $a_M \neq 0$.

It is convenient to define asymptotic sequence as an infinite sequence and asymptotic expansion as an infinite sum. This concept is indeed sufficient for many purposes, for example in case we apply the Frobenius method in order to solve a differential equation in the form of a series (in either integer and non-integer powers of the independent variable). However, in asymptotic analysis we may sometimes wish to use a finite set of functions to approximate a function that is given implicitly, most commonly as a solution to a differential equation. One of the examples of situations when this occurs is if we use the method of *dominant balance* to investigate the behaviour of a function to a certain order, very often the leading order. We will study this method more thoroughly later in this chapter. What we wish to point out at this moment is that in some cases we may require the validity of conditions (3.1) and (3.3) only for a finite range $n \in \{1, 2, \ldots, N\}, N \in \mathbb{N}$.

Let us now study some basic properties of asymptotic expansions, starting with uniqueness. Based on the given definitions, the natural suspicion is that, unlike a Taylor series, an asymptotic series of a function is not determined *uniquely*. This is indeed the case since we have been given no guidelines for the explicit choice of the asymptotic sequence $\{\varphi_n\}_{n=0}^{\infty}$. Still a weaker uniqueness property is safeguarded by this definition. It is not too difficult to see that given an asymptotic sequence $\{\varphi_n\}_{n=0}^{\infty}$, its *coefficients* $\{a_n\}_{n=0}^{\infty}$ are uniquely given by the relations

$$a_0 = \lim_{x \to x_0} \frac{y(x)}{\varphi_0(x)}, \qquad a_{N+1} = \lim_{x \to x_0} \frac{y(x) - \sum_{n=0}^{N} a_n \varphi_n(x)}{\varphi_{N+1}(x)}.$$

Another straightforward observation is that the converse relationship is not unique either, i.e. a given asymptotic expansion does not define a unique function. For instance, if $\varphi_n(x) = x^n$ and $y(x) \sim \sum_{n=0}^{\infty} a_n x^n$, it suffices to modify y by a term that decreases faster than any power of x around 0 and the asymptotic expansion will remain unchanged. It is thus clear that the function $\tilde{y}(x) \equiv y(x) + C \exp(-\frac{1}{x^2})$ provides an example of a function that shares the same asymptotic expansion with y around 0. Let us also investigate the properties of the relation $\sim$ of asymptotic equality under integration and differentiation. It follows directly from (3.1) that if $\phi(x) \ll \psi(x)$ around x_0 and ψ is locally integrable on a (finite)

[1]Recall that for expansions around a finite x_0 we can assume $x_0 = 0$ without loss of generality and thus write $\sum_n a_n x^{\alpha n}$ instead of $\sum_n a_n (x - x_0)^{\alpha n}$.

neighbourhood of x_0, then

$$\Phi(x) \equiv \int_{x_0}^{x} \phi(s)ds \ll \int_{x_0}^{x} \psi(s)ds \equiv \Psi(x) \text{ for } x \to x_0.$$

As a consequence, the relation 'be asymptotically equal to' is invariant under integration in the sense that

$$f(x) \sim g(x) \text{ for } x \to x_0 \Rightarrow \int_{x_0}^{x} f(s)ds \sim \int_{x_0}^{x} g(s)ds \text{ for } x \to x_0$$

since the integral of the correction $f(x) - g(x)$ will remain much smaller than the integral of the approximation $g(x)$ itself due to $|f(x)-g(x)| \ll |g(x)|, x \to x_0$. This fact is often used for estimation/evaluation of integrals of solutions to differential equations. Observe that (3.1) also provides an estimate of the relative error of such asymptotic integration. If we choose a neighbourhood such that $|f(x) - g(x)| \leq \varepsilon_0|g(x)|$, then within this neighbourhood we clearly have

$$\left| \int_{x_0}^{x} f(s)ds - \int_{x_0}^{x} g(s)ds \right| \leq \varepsilon_0 \int_{x_0}^{x} |g(x)|.$$

However, this kind of invariance does not hold for differentiation since the relation $\sim$ may neglect terms/corrections which are small in magnitude and vary on a small scale, i.e quickly. Hence, for the sake of differentiation it may be necessary to increase our 'resolution', or, in other words, restrict ourselves to a smaller portion of the interval in question.

We have not yet discussed the truncation procedure. For the sake of this discussion we are going to need to qualify possible singularities of our problem of interest in the usual manner. Let us for now focus on the one-dimensional case and consider the linear differential equation

$$Lu \equiv u^{(n)} + p_{n-1}(x)u^{(n-1)} + \ldots + p_1(x)u' + p_0(x)u = 0. \tag{3.5}$$

Definition 3.1. A point $x_0 \in \mathbb{R}$ is called an *ordinary point* of the equation (3.5) iff all the coefficients p_k, considered as functions of the complex variable $z = x + iy$, are analytic in a neighbourhood of the point $z = x_0$. Otherwise it is called a *singular point*. These we divide further into two categories. A point x_0 is called a *regular singular point* iff the functions $(x - x_0)p_{n-1}, (x - x_0)^2 p_{n-2}, \ldots, (x - x_0)^n p_0$ are analytic at $z = x_0$. Otherwise x_0 is called an *irregular singular point*.

The crucial observation for our cause is that if x_0 is a regular singular point, problem (3.5) always has a solution of the form

$$y(x) = (x - x_0)^s A(x), \tag{3.6}$$

where $A(x)$ is analytic at x_0 and s is given by the indicial equation (see chapter 4 in [17]). In case all the roots of the indicial equation are non-negative, it follows that this solution is well-behaved at x_0.

Now that we have established this classification, we can turn our focus back to the question of truncation. It is important here to keep in mind that asymptotic methods are by definition methods of *local* approximation. But for practical use it is usually preferable to approximate a function on an interval rather than just in the infinitesimal neighbourhood of a single point x_0. There might be points of extraordinary importance, though, mostly points where boundary or initial conditions are specified. It is natural to use these as starting points of the approximation. Whether we are using an infinite set of functions $\{\varphi_n\}$ (e.g. we are looking for a series expansion of the solution) or a finite one, we require the strict validity of (3.1) around these points. But in order to attain maximal precision elsewhere we need to analyze the validity of condition (3.1) for all points in the interval. The asymptotic estimate in (3.4) then provides a straightforward suggestion on how to truncate the sum in (3.2) in the optimal manner: we simply need

to minimize the magnitude of $a_n\varphi_n(x)$.[2] Note that for any convergent expansion we necessarily have $a_n\varphi_n(x) \to 0$ for $n \to \infty$ within the entire domain of convergence, so this procedure will invariably instruct us to include as many terms as possible. This is one of the differences between convergent and non-convergent asymptotic expansions as well as one of the reasons why we usually restrict our understanding of the term *asymptotic series* to non-convergent expansions. In that case the dependence on x becomes much more relevant in the sense that it affects the number of terms we need to include to optimize the precision of our approximation of $y(x)$. This can be readily illustrated in case we are using a method that yields a recurrence relation for $\{a_n\}$, e.g. the Frobenius method for ordinary differential equations. Let us assume a simple first-order recurrence of the form $a_{n+1} = \phi(n)a_n$. The magnitude of the terms in the expansion obviously decreases as long as the ratio of consecutive terms in the expansion satisfies

$$\left|\frac{a_{n+1}\varphi_{n+1}(x)}{a_n\varphi_n(x)}\right| = \left|\phi(n)\frac{\varphi_{n+1}(x)}{\varphi_n(x)}\right| \leq 1. \tag{3.7}$$

This inequality can be used to find a relation, implicit or explicit, between n and x. It becomes very easy in the case of a power series, i.e. with $\varphi_n(x) = x^n$, and the same applies to the more general case $\varphi_n(x) = x^{n+s}$. Then condition (3.7) simplifies to $|\phi(n)x| \leq 1$ which means we need to change the number of terms (from n to either $n+1$ or $n-1$, depending on ϕ) in the expansion of $y(x)$ at $|x| = \left|\frac{1}{\phi(n)}\right|$. If our interval of interest is unbounded, useful asymptotic relations between x and n (the optimal number of terms in the expansion) for $x \to \infty$ can be extracted easily. Analogous truncation rules can be obtained for $\varphi_n(x) = x^{-n}$ (for approximations around x_0 infinite), or more generally for $\varphi_n(x) = x^{\alpha n}$. Observe that inequality (3.7) and the ensuing '$1/\phi$-rule' has practical applications for convergent series as well. There it reveals at which n the magnitude of the terms starts to decrease so in this sense it determines the *minimal* number of terms necessary for a reasonable approximation of $y(x)$.

The Frobenius method provides one of the common ways of acquiring an asymptotic expansion. It is also instrumental for demonstrating the truncation procedure. Therefore we now give an example where we apply this method and find an asymptotic expansion of a solution to a differential equation.

Example 3.1. *Consider the differential equation* $y'' + y' - \frac{y'}{x} + \frac{y}{x^2} = 0$ *with the boundary condition* $y(0) = 0$. *The point* $x = 0$ *is a regular singular point so a solution of the form* (3.6) *ought to exist. Therefore we assume the form* $y(x) = x^s \sum_{n=0}^{\infty} a_n x^n$ *and substitute into the equation, obtaining for the lowest power* x^{s-2} *the indicial equation*

$$[s(s-1) - s + 1]\, a_0 = 0.$$

Applying the requirement of a non-trivial solution $a_0 \neq 0$, *the indicial equation has the solutions* $s_1 = 1/2, s_2 = 2$. *For a general power of x we have*

$$[(n+s+1)(n+s) - (n+s+1) + 1]\, a_{n+1} + (n+s)a_n = 0$$

which yields the recurrent relation $a_{n+1} = -\frac{1}{(n+s)}a_n$. *As expected, the series* $\sum_n a_n x^n$ *is convergent around* $x = 0$, *it even has an infinite radius of convergence. Since the difference* $s_2 - s_1$ *is non-integer, the 2 solutions for s in fact determine the two linearly independent solutions to the given equation. Both of them clearly satisfy the boundary condition. As seen from analyzing the ratio* $\frac{a_{n+1}x^{n+s+1}}{a_n x^{n+s}}$, *the magnitude of the terms in the expansion continues to grow as long as* $n + s < x$, *so the number of terms required for a decent approximation grows linearly with x. One might also notice that the solution corresponding to*

[2] Recalling that the relation '$f(x) \sim g(x)$ for $x \to x_0$' can be read as 'f is equal to g up to terms much smaller/decreasing much faster than g in the vicinity of x_0', we can see that the symbol $\sim$ in (3.4) does not induce any looseness into this argument.

$s_2 = 2$ can be easily re-expressed as

$$y_2(x) = a_0 x \sum_{n=0}^{\infty} \frac{(-1)^n}{(n+1)!} x^{n+1} = -a_0 x(e^{-x} - 1),$$

where a_0 is to be determined from the initial conditions.

Let us now give an example of an asymptotic series around $x = \infty$.

Example 3.2. *Find an asymptotic series for the solution to the equation $y' - y\left(\frac{1}{x} + \frac{1}{\sqrt{x}}\right) = 0$ with boundary condition $y(\infty) = 1$. Note that due to the $\frac{y}{\sqrt{x}}$-term, $x = \infty$ is an irregular singular point. This term also means that we need to look for an expansion in half-integer powers of x, ι.e. $y(x) = x^s \sum_{n=0}^{\infty} a_n x^{-\frac{n}{2}}$. The boundary condition implies that the only admissible value of s for an asymptotic approximation around infinity is 0 as well as that $a_0 = 1$. Substituting now into the equation, we arrive at*

$$-\frac{n}{2} a_n - a_n + a_{n+1} = 0,$$

yielding, as expected, a divergent series with coefficients given by $a_{n+1} = \frac{n+2}{2}$. Looking for the error-minimizing truncation relation we obtain the condition

$$\frac{n+1}{2\sqrt{x}} \leq 1 \Leftrightarrow x \geq \frac{(n+1)^2}{4}, \tag{3.8}$$

so the optimal number of terms is proportionate to $\sqrt{x}$. For $x < 4$ we are instructed to keep only the first two terms so we can expect the approximation to be poor for such small values of x. If we are, however, interested in asymptotic behaviour of the solution around 0, we can assume the asymptotic relation $y \sim x^\alpha$ for $x \to 0$ and substitute into the equation, finding $\alpha = 1$. The solution hence tends towards 0 linearly in the vicinity of $x = 0$.

3.2 The WKBJ approximation in 1 dimension

The WKBJ approximation is a singular perturbation method for solving differential equations. As mentioned above, it is usually used for solving equations whose highest-order derivative is assumed to have an infinitesimal coefficient ε and its applicability is restricted to *linear* equations. Although this is a significant restriction, the method still finds many applications in different fields, the most common representatives being quantum mechanics and bifurcation problems of virtually any kind. The reason, as we argued in chapter 1.1, is that in sufficiently small neighbourhoods of bifurcation points, first-order expansions of the equation's coefficients can offer a reasonable approximation to the actual equation.

The WKBJ approximation of a solution $y(x)$ to a linear differential equation

$$\varepsilon y^{(n)} + p_{n-1}(x) y^{(n-1)} + \ldots + p_1(x) y' + p_0(x) y = 0$$

is an expansion of the form

$$y(x) \sim \exp\left(\frac{1}{\delta} \sum_{n=0}^{\infty} \delta^n S_n(x)\right), \qquad \delta \to 0_+. \tag{3.9}$$

In practical calculations, first-order expansions with an explicit amplitude function of the form

$$y(x) \sim A(x) \exp\left(\frac{S(x)}{\delta}\right), \qquad \delta \to 0_+,$$

are a very common special case.

For now we will adhere to the more general expansion (3.9). The exponential form of this expansion means that, in practice, we are looking for an asymptotic expansion to $\ln y$. It also suggests that WKBJ approximations are exceptionally well-suited for describing dissipative (exponentially damped) and dispersive ('pseudoharmonic') phenomena. The form is also convenient for describing the solution's behaviour near singular (or otherwise exceptional) points; the logarithmic scale can be very useful for this purpose. For the sum in (3.9) to offer a proper asymptotic expansion (to $\ln y$), we formally impose the natural requirement that the former terms of the expansion dominate the latter terms, i.e.

$$\begin{aligned} S_1(x) &\ll \frac{1}{\delta} S_0(x), \qquad \delta \to 0_+, \\ &\vdots \\ S_{n+1}(x) &\ll \frac{1}{\delta} S_n(x), \qquad \delta \to 0_+. \end{aligned}$$

Note that the presence of the parameter δ substitutes for the necessity to specify the point around which these asymptotic relations should hold. These enables us to acquire an expansion that is asymptotically valid on the entirety of the interval in question, especially if it is bounded. However, the parameter δ does not represent an additional 'degree of freedom'. It is intrinsically related to the 'original' parameter ε and its dependence on the latter needs to be determined using the techniques of dominant balance. This approach will often reveal that δ is proportional to ε and the most common choice is thus $\delta = \varepsilon$. Note, however, that non-linear relations between δ and ε are also possible.

Let us now derive the governing equations for S in the case that is most relevant to quantum mechanics as well as to our cause. For this purpose we first express the asymptotic forms of the first and second derivatives of y:

$$\begin{aligned} y'(x) &\sim \left(\frac{1}{\delta} \sum_{n=0}^{\infty} \delta^n S_n'(x) \right) \exp\left(\frac{1}{\delta} \sum_{n=0}^{\infty} \delta^n S_n(x) \right), \qquad \delta \to 0_+, \\ y''(x) &\sim \left[\frac{1}{\delta} \sum_{n=0}^{\infty} \delta^n S_n''(x) + \frac{1}{\delta^2} \left(\sum_{n=0}^{\infty} \delta^n S_n'(x) \right)^2 \right] \exp\left(\frac{1}{\delta} \sum_{n=0}^{\infty} \delta^n S_n(x) \right), \qquad \delta \to 0_+. \end{aligned} \tag{3.10}$$

Now consider the second-order (Schrödinger) differential equation

$$\varepsilon^2 y'' + Q(x) y = 0 \tag{3.11}$$

with $Q(x) \neq 0$. Note that we chose to denote the infinitesimal coefficient of y'' as ε^2 rather than ε for convenience. First, we determine the dependence $\delta(\varepsilon)$ using the method of dominant balance. Upon substituting (3.10) into (3.11) and cancelling the non-zero exponential factor we have

$$\frac{\varepsilon^2}{\delta^2} \left(S_0' \right)^2 + \frac{\varepsilon^2}{\delta} \left(2 S_0' S_1' + S_0'' \right) + \ldots = -Q(x).$$

Since we want to consider δ infinitesimal, the largest term (in δ) on the left-hand side is obviously the one of order δ^{-2}, i.e. $\frac{\varepsilon^2}{\delta^2} \left(S_0' \right)^2$. By dominant balance, this term must be of the same $O(1)$-order as the largest (and only) term on the right-hand side. Hence, we have $\delta = k\varepsilon$ and, as mentioned above, we usually make the convenient choice $\delta = \varepsilon$. Since there are no higher-order terms in δ on the right-hand

side, all other orders of δ on the left-hand side must vanish, too, yielding

$$\begin{aligned} S_0'^2 &= -Q(x), \\ 2S_0'S_1' + S_0'' &= 0, \\ S_1'^2 + 2S_0'S_2' + S_1'' &= 0, \\ &\vdots \\ \sum_{j=0}^{n} S_j'S_{n-j}' + S_{n-1}'' &= 0. \end{aligned} \tag{3.12}$$

Recurrently solving the first two equations, we obtain

$$\begin{aligned} S_0(x) &= \pm \int_a^x \sqrt{-Q(s)}ds, \\ S_1(x) &= -\frac{1}{4}\ln|Q(x)|, \end{aligned}$$

which clarifies why we assumed $Q(x) \neq 0$. Note that this assumption, together with the continuous differentiability (and hence continuity) of Q, also guarantees that taking the absolute value of Q in the logarithm does not give rise to any singularities in $S_1'(x)$. The exponential form of (3.9) allows us to merge all the constants of integration of S_0 and S_1 (and all the other functions S_i in higher-order approximations) into one. Observing that the two possible signs of S_0 lead to two linearly independent solutions, we arrive at the leading-order approximation of the solution to (3.11) in the form

$$y(x) \sim c_1 Q(x)^{-1/4} \exp\left(\frac{1}{\varepsilon}\int_a^x \sqrt{-Q(s)}ds\right) + c_2 Q(x)^{-1/4} \exp\left(-\frac{1}{\varepsilon}\int_a^x \sqrt{-Q(s)}ds\right). \tag{3.13}$$

3.3 WKBJ approximation for a system of ODEs

We would now like to prove that the first-order WKBJ approximation of a form similar to (3.13) is also valid for a system of several ordinary differential equations (ODEs). Most of the following considerations will be valid for any finite number of equations. However, in order to be more specific as well as to relate the general problem to a two-component reaction-diffusion system, we will often give explicit forms of various relevant quantities valid specifically for a system of two equations. Let us reveal in advance that the proof given in this section is only valid for systems of ODEs whose matrix has a specific form. The procedure that we will follow is inspired by a lecture given by P. Siegl of Queen's University Belfast at a student conference on pure and applied mathematics, where a similar proof was given for a single equation, and motivated by the asymptotic approach by Klika et al. [11]

Let us again consider a second-order system of differential equations of the form

$$\varepsilon^2 \boldsymbol{y}'' + \tilde{Q}(x)\boldsymbol{y} = 0. \tag{3.14}$$

For a linearized reaction-diffusion system (1.14) in a single spatial dimension, e.g. for $x \in [0, L]$, $\tilde{Q}$ takes on the form

$$\tilde{Q}(x) = D^{-1}(J(x) - \lambda\mathbb{I}), \tag{3.15}$$

where λ again represents the growth rate. Note that, unlike in chapter 1.2, here we allow for explicit spatial dependence of the reaction kinetics and thus of the stability matrix J. This adds substantial difficulty to the problem, as now even the linearized equations cannot generally be solved exactly. The parameter

$\varepsilon^2 = \frac{D_1 T}{L^2}$, which is given as the ratio of the characteristic temporal scales of the reaction kinetics and diffusion, arises from the non-dimensionalization procedure. In the context of developmental biology, specifically for the morphogen pair nodal and lefty, it is of the order 10^{-4}. [11] It is given here explicitly mainly in order to relate (3.14) to (3.11) as an asymptotic setting and we will soon have $\tilde{Q}$ absorb it as we will be interested in asymptotic behaviour for the limit of large growth rates $\lambda \to +\infty$ rather than $\varepsilon \to 0$, although the effect of the latter limit would be similar. Let us therefore define the matrix $Q(x) \equiv \frac{1}{\varepsilon^2}\tilde{Q}(x)$, which for reaction-diffusion equations is thus given as $Q(x) = \frac{1}{\varepsilon^2}\left[D^{-1}\left(J(x) - \lambda\mathbb{I}\right)\right]$, and rewrite (3.14) as

$$\boldsymbol{y}'' + Q(x)\boldsymbol{y} = 0. \tag{3.16}$$

The procedure will be as follows: we will find an exact solutions u, v for the specific case $q(x) = x\mathbb{I}$, rewrite them in terms of Q rather than x and find that for a general $Q(x)$ they do not satisfy (3.16), but instead obtain additional linear correction terms. We will then be able to use these terms to find an implicit integral form of the solution and apply the limit $\lambda \to +\infty$ to show that the correction of the solution vanishes in this limit.

Let us now define $\mu_j(x)$ to be the eigenvalues of $Q(x)$. For a two-component reaction-diffusion system that means

$$\mu_\pm(x) = \frac{1}{2}\left[\operatorname{Tr} Q(x) \pm \sqrt{(\operatorname{Tr} Q(x))^2 - 4\det Q(x)}\right].$$

From the relations

$$\begin{aligned}\operatorname{Tr} Q(x) &= \operatorname{Tr}(D^{-1}J(x)) - \lambda\left(1 + \frac{1}{d}\right)\\ \det Q(x) &= \frac{1}{d}\det(J(x) - \lambda\mathbb{I}) = \frac{1}{2d}\left[\lambda^2 - \operatorname{Tr} J(x)\lambda + \det J(x)\right]\end{aligned}$$

and the conditions (1.16) for $J(x)$ we can conclude that $\det Q(x) > 0$ for $\lambda \geq 0$, and hence that $\mu_\pm(x) \to -\infty$ linearly with λ for $\lambda \to +\infty$ in every $x \in [0, L]$. This is intuitively clear, since we can view λ as shifting the spectrum of $D^{-1}J(x)$ to the left (although not 'uniformly' due to the fact that we subtract λD^{-1} rather than $\lambda\mathbb{I}$). For later use, we also note that

$$\mu'_\pm(x) = \frac{1}{2}\left[\operatorname{Tr}(D^{-1}J(x))' \pm \frac{\operatorname{Tr} Q(x)\operatorname{Tr}(D^{-1}J(x))' - \lambda\frac{2}{d}(\operatorname{Tr} J(x))' + 2(\det J(x))'}{\sqrt{(\operatorname{Tr} Q(x))^2 - 4\det Q(x)}}\right]$$

is of order unity (which means bounded) for $\lambda \to +\infty$, and so is $\mu''_\pm(x)$. We then go on to define

$$\xi_\pm(x) \equiv \int_0^x \sqrt{-\mu_\pm(s)}ds,$$

which for λ large enough is real and positive, and becomes unbounded with growing λ for every $x > 0$. Having made all these observations that will come in handy soon, we can define the two functions

$$\begin{aligned}u_\pm(x) &\equiv \sqrt{\frac{\xi_\pm(x)}{\xi'_\pm(x)}}K_{1/3}\left(\xi_\pm(x)\right),\\ v_\pm(x) &\equiv \sqrt{\frac{\xi_\pm(x)}{\xi'_\pm(x)}}I_{1/3}\left(\xi_\pm(x)\right),\end{aligned} \tag{3.17}$$

with $K_\alpha(x), I_\alpha(x)$ being the modified Bessel functions defined as the two linearly independent solutions to $x^2y'' + xy' - (x^2 + \alpha^2) = 0$. As stated above, for $Q(x) = x\mathbb{I}$ we have two independent Airy equations

and u, v are their exact solutions. Even more importantly, using the recurrent relations

$$\begin{aligned} K_n'(x) &= \frac{n}{x}K_n(x) - K_{n+1}(x), \\ I_n'(x) &= \frac{n}{x}I_n(x) + I_{n+1}(x), \\ &= I_{n-1}(x) - \frac{n}{x}I_n(x), \end{aligned} \tag{3.18}$$

along with the symmetric property $K_{-n}(x) = K_n(x)$ (cf. Theorems 4.15 and 4.16 in [3]), and after some exhausting algebra we arrive at the following property that holds for both u and v:

$$\psi_\pm'' = \left[(\xi_\pm')^2 - T_\pm\right]\psi_\pm = 0, \tag{3.19}$$

where

$$T_\pm(x) = \frac{1}{4}\left[\frac{5}{9}\left(\frac{\xi_\pm'}{\xi_\pm}\right)^2 + 2\frac{\xi_\pm'\xi_\pm'''}{\xi_\pm'^2} - 3\left(\frac{\xi_\pm''}{\xi_\pm'}\right)^2\right](x) = \frac{1}{4}\left[\frac{5}{9}\left(\frac{\xi_\pm'}{\xi_\pm}\right)^2 + \frac{(\xi_\pm'^2)''}{\xi_\pm'^2} - \frac{5}{4}\frac{\left((\xi_\pm'^2)'\right)^2}{\xi_\pm'^4}\right](x) =$$
$$= \frac{1}{4}\left[-\frac{5}{9}\frac{\mu_\pm}{\xi_\pm^2} + \frac{\mu_\pm''}{\mu_\pm} - \frac{5}{4}\frac{\mu_\pm'^2}{\mu_\pm^2}\right](x).$$

Since u and v are the two linearly independent solutions to (3.19), with $K_{1/3}(x)$ regular at $x \to +\infty$ and $I_{1/3}(x)$ regular for $x \to 0$, their Wronskian $W[u, v](x)$ can be shown to be constant. In fact, u is the exponentially decaying solution while v is the exponetially growing solution and $W[u, v](x)$ is equal to 1. For a more detailed account of the properties of the modified Bessel functions and Airy functions, we refer the reader to the aforementioned book by Bell or the well-known book of special functions by Abramowitz and Stegun, especially chapters 9.6 and 10.4.[2] We can now proceed by defining another function $G(x, s) = u(x)v(s) - v(x)u(s)$.[3] This function can now be used as part of the kernel of an integral operator to find the exact solution to (3.16). More precisely, it is just a matter of direct differentiation and applying the property (3.19) of u and v to show that the implicitly defined function

$$y_\pm = u_\pm + \int_x^L G_\pm(x, s)T_\pm(s)y_\pm(s)ds \tag{3.20}$$

[4] satisfies

$$y_\pm''(x) = \left[-T_\pm + \xi_\pm'^2\right]y_\pm(x) + T(x)y_\pm(x)W[u, v](x) = \xi_\pm'^2 y_\pm(x).$$

We would love to avail ourselves of this fact and use the eigenvectors $\boldsymbol{p}_\pm$ of Q, which are defined by $Q(x)\boldsymbol{p}_\pm = \mu_\pm(x)\boldsymbol{p}_\pm$, to diagonalize Q in order to obtain the solution to (3.16) in the form

$$\boldsymbol{y}_\pm(x) = \boldsymbol{p}_\pm y_\pm(x). \tag{3.21}$$

However, this is generally not possible, since $\boldsymbol{p}_\pm$ might depend on x. Still there are spatially dependent matrices whose eigenvectors do not have spatial dependence. The necessary and sufficient condition for this to be true is that Q have the form

$$Q(x) = q(x)M, \tag{3.22}$$

with q a scalar function and M a constant matrix. This is obviously a significant restriction. It is clearly not necessary in order for this procedure to yield a correct solution. In fact, it can be used for any

[3] This is obviously not a proper Green's function. However the property that we are about to show 'justifies' this notation.

[4] We omit v from the non-integral part of $y_\pm$ since ξ is unbounded for any $x > 0$ when $\lambda \to +\infty$ and v, in turn, is unbounded for an unbounded argument.

spatially dependent diagonal matrix (as long as ξ is well-defined), since then the system consists of several mutually independent equations for which this form of solution works properly. But from now on we are going to adhere to the restriction for Q given in (3.22).

A matrix Q of the form (3.15) cannot generally be written in the form (3.22). We are therefore going to reproduce the properties of such Q in the assumptions on the scalar function $q_\lambda(x)$ and the matrix M. Note that now $\mu_\pm$ is given as

$$\mu_\pm(x) = \frac{q_\lambda(x)}{2}\left[\operatorname{Tr} M \pm \sqrt{(\operatorname{Tr} M)^2 - 4\det M}\right].$$

We are thus going to assume $\det M > 0$, $\operatorname{Tr} M < 0$ and $q_\lambda(x) \to +\infty$ as $\lambda \to +\infty$ for every $x \in [0, L]$. With these assumptions, the key property

$$\mu_\pm(x) \to -\infty \text{ for } \lambda \to +\infty, \forall x(0, L],$$

is preserved. We could also add various obscure sets of assumptions to secure that the function $T_\pm(x)$ be bounded (in both x and λ). While we are going to assume q sufficiently smooth so as to allow for a continuous second derivative (in x), we will simplify the rest by merely requiring that the boundedness of T in both x and λ holds true as this is clearly the case for the case of reaction-diffusion equations with Q given by (3.15).

Let us continue where we left off. Now (3.21) clearly satisfies (3.16) as we have

$$\boldsymbol{y}''_\pm + Q(x)\boldsymbol{y}_\pm = \xi'^2_\pm y_\pm(x)\boldsymbol{p}_\pm + \mu_\pm(x)y_\pm(x)\boldsymbol{p}_\pm = 0.$$

We would like to show that the correction to u given by the integral becomes arbitrarily small for λ large. However, this requires some subtle handling as it is not generally true for (3.20); but before we take care of that, we take a small detour that will bring us closer to completing the proof. As things stand, the integral in (3.20) features the unknown function y, which is very inconvenient but can be taken care of in a rather elegant manner. Let us introduce the integral operator

$$(A_\pm f)(x) \equiv \int_x^L G_\pm(x, s)T_\pm(s)f(s)ds.$$

Then (3.20) can be rewritten $(\mathbb{I} - A)y = u$. If we could prove that $\|A\| < 1$, we would be able to define the inverse to the left-hand side via the Neumann series. Let us for now hope that this can be done. Then we would have $y = (\mathbb{I} - A)^{-1}u$ and the correction to u in y can thus be written as

$$y - u = (\mathbb{I} - A)^{-1}u - u = \left[(\mathbb{I} - A)^{-1} - \mathbb{I}\right]u = \left(\sum_{n=1}^{\infty} A^n\right)u,$$

which implies

$$\|y - u\|_{L^\infty(\Omega)} \le \sum_{n=1}^{\infty} \|A\|_\infty^n \|u\|_{L^\infty(\Omega)} = \frac{\|A\|_\infty}{1 - \|A\|_\infty}\|u\|_{L^\infty(\Omega)}. \tag{3.23}$$

In this estimate we denoted by $\|\cdot\|_\infty$ the operator norm induce by $\|\cdot\|_{L^\infty(\Omega)}$ and set $\Omega = (0, L)$. In this way we eliminated the unknown function y from our estimate of the correction $r(x) \equiv y(x) - u(x)$. Note that both of the norms on the right-hand side depend on λ. That means that if we could prove the right-hand side to approach 0 with increasing λ, the proof would be complete.

To show that, we need to take certain precautions with respect to the function $G(x, s) = u(x)v(s) - v(x)u(s)$. By the definition of A, we are only interested in the function for $s \ge x$. We stated earlier that u is the decaying and v the increasing solution, both exponentially. This fact can be expressed more precisely,

namely in terms of the suitable weighing functions $w_1(x) \equiv (-\mu(x))^{1/4}$ and $w_2(x) \equiv \exp\left(\int_0^x \sqrt{-\mu(s)}ds\right)$. Note that the functions $\nu_1(x) \equiv \frac{1}{w_1(x)w_2(x)}$ and $\nu_2(x) \equiv \frac{w_2(x)}{w_1(x)}$ correspond to the two modes in (3.13), either both harmonic or one exponentially growing and the other one decaying. For these two functions we have the estimates

$$|u(x)| \leq \frac{M}{w_1(x)w_2(x)}, \qquad |v(x)| \leq \frac{Mw_2(x)}{w_1(x)}, \tag{3.24}$$

which implies $|u(x)v(x)| \leq \frac{M}{w_1(x)^2} = \frac{M}{\sqrt{-\mu(x)}}$. That way we see that since ξ is increasing in x, the second term in $G(x, s)$ can easily be shown to be bounded. Not so the first one, which is why we need to be smart one last time. Instead of estimating $\|Af\|$, let us investigate $\|\tilde{Af}\|$, where $\tilde{f} \equiv w_1 w_2 f$. Then, upon dropping the subscript $\pm$ and expanding the integrand by $\frac{w_1^2(s)w_2^2(s)}{w_1^2(s)w_2^2(s)}$ we can rewrite the action of the modified operator $\tilde{A}$ as

$$\tilde{Af}(x) = \int_x^L \frac{\tilde{u}(x)\tilde{v}(s) - \tilde{v}(x)\tilde{u}(s)}{w_2^2(s)} \cdot \frac{T(s)}{w_1^2(s)}\tilde{f}(s)ds. \tag{3.25}$$

Since T was assumed bounded with respect to λ, the second fraction approaches 0 for λ large due to $\sqrt{-\mu(s)}$ in the denominator. Since L is finite, all that is left is to show that the first fraction is bounded. Since $s \geq x$, for the second term we have

$$\left|\frac{\tilde{v}(x)\tilde{u}(s)}{w_2^2(s)}\right| \leq \left|u(x)v(x)\frac{w_1(s)w_2(x)w_1(x)}{w_2(s)}\right| \leq M\left|\frac{w_1(s)w_2(x)}{w_1(x)w_2(s)}\right| \leq M\left|\frac{w_1(s)}{w_1(x)}\right|,$$

which is bounded by the boundedness of μ with respect to x. Using (3.24) and the fact that $\left|\frac{v(s)}{w_2(s)}\right| \leq \frac{M}{|w_1(s)|}$, for the first term in (3.25) we obtain

$$\left|\frac{\tilde{u}(x)\tilde{v}(s)}{w_2^2(s)}\right| \leq M\left|\frac{w_2(x)w_1(x)u(x)}{w_2(s)}\right| \leq \frac{M^2}{w_2(s)}, \tag{3.26}$$

which vanishes for every $s > 0$ if λ grows boundlessly, which completes the proof. The result is thus what we expected: the greater the growth, the better description is provided by the asymptotic methods. The fact that the proof was to a big extent based on the spectral properties of the reaction diffusion problem suggests that this asymptotic approach is justified in such settings. However, the significant restriction (3.22) reduces the room for drawing fundamental conclusions with respect to such problems.

Conclusion

This work was intended to provide an overview of the basic concepts and frameworks for reaction-diffusion equations and dynamical systems in general, as well as to further develop the methods of theoretical analysis of these and related models in mathematical biology and beyond. First we present a set of tools applicable to a large amount of evolution models, with a focus on those useful in near-equilibrium states. We conclude this part by listing several basic types of bifurcations and compare their properties.

After (re-)acquainting the reader with the most significant notions of reaction-diffusion models, their main features and conditions, as well as attempting to offer an intuitive understanding of these features and concepts, we dare a step 'beyond the linear' and present and apply selected methods of weakly non-linear analysis, such as separation of scales and solvability analysis, a term especially popular in Chapter 2.

In Chpater 3 we present the notion of asymptotic analysis, an approach fundamentally different from those adopted in the earlier chapters. We analyse the key properties of reaction-diffusion models with respect to such approaches, and successfully attempt to prove an approximation theorem for multicomponent systems, although one that assumes a simplified form of spatial dependence in the reaction kinetics. This result makes us hopeful that further (and less restrictive) generalizations might be possible.

Bibliography

[1] R. Abraham and J. Robbin. *Transversal mappings and flows*. An appendix by Al Kelley. W. A. Benjamin, Inc., New York-Amsterdam, 1967.

[2] M. Abramowitz and I.A. Stegun. *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*, chapter 9.6,10.4. Applied mathematics series. Dover Publications, 1964.

[3] W.W. Bell. *Special functions for scientists and engineers*. Van Nostrand, 1968.

[4] S. Chandrasekhar. *Hydrodynamic and hydromagnetic stability*. Dover Books on Physics. Dover Publications, 2013.

[5] E. B. Davies. *Spectral Theory and Differential Operators*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 1995.

[6] D.E. Edmunds and W.D. Evans. *Spectral Theory and Differential Operators*. Oxford Monographs of Modern Lang and Lit Series. Clarendon Press, 1987.

[7] B. Ermentrout and J. E. Enderby. Stripes or spots? nonlinear effects in bifurcation of reaction-diffusion equations on the square. *Proc. Royal Soc. A*, 434(1891):413–417, 1991.

[8] M. W. Hirsch and S. Smale. *Differential equations, dynamical systems, and linear algebra*. Number 60 in Pure and applied mathematics. Acad. Press, 1974.

[9] T. Kato. *Perturbation Theory for Linear Operators*. Classics in Mathematics. Springer Berlin Heidelberg, 1995.

[10] J. Kováč. Turingův model prostorového uspořádání a vliv geometrie, bachelor thesis. `https://physics.fjfi.cvut.cz/publications/mf/2017/bp_mf_17_kovac.pdf`, 2017. accessed 8 April 2020.

[11] A. Krause, V. Klika, T. Woolley, and E. Gaffney. From one pattern into another: analysis of turing patterns in heterogeneous domains via wkbj. *Journal of The Royal Society Interface*, 17:20190621, 01 2020.

[12] D. Krejčiřík. Geometrical aspects of spectral theory. `http://nsa.fjfi.cvut.cz/david/other/gspec19.pdf`. accessed 8 April 2020.

[13] J.D. Murray. *Mathematical Biology: I. An Introduction*. Interdisciplinary Applied Mathematics. Springer New York, 2007.

[14] S. T. A. Pickett and M. L. Cadenasso. Landscape ecology: Spatial heterogeneity in ecological systems. *Science*, 269(5222):331–334, 1995.

[15] J. Smoller. *Shock Waves and Reaction—Diffusion Equations*. Springer-Verlag, Berlin-Heidelberg-New York, 1983.

[16] A. M. Turing. The chemical basis of morphogenesis. *Phil. Trans. R. Soc. Lond. B 1952 237, 37-72*, 1952.

[17] R.B. White. *Asymptotic Analysis of Differential Equations*. Imperial College Press, 2010.

Appendix A

The spectrum of the Laplace operator

In this section we will briefly summarize a few useful results regarding the spectral properties of the Laplace operator, which come in handy when setting up the theoretical framework of reaction-diffusion equations. We will only refer the reader to external sources where we find it necessary, i.e. in case we feel we are well beyond the bounds of 'general knowledge', which, for our purposes, is thus defined by the (absence of) bibliographic references hereinafter.

First of all, we need to establish what we mean by the term *spectrum*. We will give the conventional, commonly used definition. But before we do that, let us introduce some basic notation. In what follows, by $\mathcal{L}(V, W)$ we will mean the vector space of linear (not necessarily bounded) maps from one vector space V to another, W; if $W = V$, we will just write $\mathcal{L}(V)$ (the same for bounded operators, self-adjoint operators etc.). In the case of Banach spaces X and Y over the field of complex numbers $\mathbb{C}$, the subspace of bounded operators will be denoted as $\mathcal{B}(X, Y)$. $\mathcal{H}$ will always denote a Hilbert space over $\mathbb{C}$. Recall that *closedness* and *closability* of an operator $A \in \mathcal{L}(X, Y)$ are defined using the graph $\Gamma(A) \equiv \{(x, Ax) \mid x \in D(A)\}$ and that the former is equivalent to completeness of the domain $D(A)$ with respect to the graph norm

$$\|x\|_A \equiv \sqrt{\|x\|_X^2 + \|Ax\|_Y^2}. \tag{A.1}$$

An immediate observation is that for $\mathrm{Ran}(A)$ to be a closed subspace A must necessarily be closed.

Now we can proceed to defining the *resolvent set* for a closed[1] operator $A \in \mathcal{L}(X)$ as

$$\rho(A) \equiv \{z \in \mathbb{C} \mid (A - zI) \text{ is injective and } (A - zI)^{-1} \in \mathcal{B}(X)\}$$

and the *spectrum* as its complement

$$\sigma(A) \equiv \mathbb{C} \setminus \rho(A).$$

Recall that in general spectrum consists of three disjoint subsets: point spectrum σ_p (eigenvalues, i.e. $(A - \lambda I)$ is not injective), continuous spectrum σ_c ($(A - \lambda I)^{-1}$ is defined densely but not everywhere) and residual spectrum σ_r ($(A - \lambda I)^{-1}$ is not densely defined) and that for a densely defined self-adjoint (i.e. A coincides with A^*, including domains) operator (we denote the subspace of such operators $\mathcal{L}_{\mathrm{sa}}$) on a Hilbert space we have $\sigma_r(A) = \emptyset, \sigma(A) \subset \mathbb{R}$ and a different disjoint decomposition is possible, namely $\sigma(A) = \sigma_d(A) \cup \sigma_{ess}(A)$, where the *discrete spectrum* $\sigma_d \subset \sigma_p$ consists solely of isolated eigenvalues of finite multiplicity; its complement σ_{ess} is called the *essential spectrum*.

[1] If A is not closed, then neither is $(A - \lambda I)$ or its range, which coincides with the domain of its inverse (if this exists). However, the domain of a bounded operator is the whole space and thus always closed. Consequently, $(A - \lambda I)^{-1}$ cannot be a bounded operator, meaning $\sigma(A) = \mathbb{C}$.

The procedure we will follow from now on is basically adopted from the lecture notes [12]. Our first goal is to show that the (negatively taken) Laplace operator on an L^2-space on a bounded domain[2] $\Omega \subset \mathbb{R}^m$ with Neumann boundary conditions is (or rather 'can be made') self-adjoint and positive so that we can then apply one of the spectral theorems applicable to such operators. To this end, let us note that the action of the Laplace operator with respect to the scalar product on L^2

$$(\varphi, -\Delta\psi) = -\int_\Omega \overline{\varphi}\Delta\psi = -\int_{\partial\Omega} \overline{\varphi}\underbrace{(\boldsymbol{n}\cdot\nabla)\psi}_{=0 \text{ on } \partial\Omega} + \int_\Omega \overline{\nabla\varphi}\cdot\nabla\psi = \int_\Omega \overline{\nabla\varphi}\cdot\nabla\psi,$$

$$\text{in the special case } \varphi = \psi: \qquad (\psi, -\Delta\psi) = \int_\Omega |\nabla\psi|^2 = \|\nabla\psi\|^2$$

is symmetric and positive (in the sense that $(\psi, -\Delta\psi) \geq 0$; $\boldsymbol{n}$ again stands for the outward normal to $\partial\Omega$). Observe as well that the right-hand side relaxes the smoothness assumptions on ψ required for the left-hand side to make good sense, and that it can be used to define a positive quadratic[3] form $a'[\psi] = (\psi, -\Delta\psi) = \|\nabla\psi\|^2$; we will make good use of this fact later on. Due to the 'pleasurable' properties of the right-hand side we will restrict ourselves to this *distributional* (or *weak*) Laplace operator that generalizes the 'classical' action of the operator on (at least twice) continuously differentiable functions. As the reader most probably expects, this leads us to the notion of *weak derivatives* and *Sobolev spaces*.

Definition A.1. Let $\alpha = (\alpha_1, \ldots, \alpha_m) \in \mathbb{N}_0^m$ be a multiindex of order $|\alpha| = \alpha_1 + \ldots + \alpha_m$. We say that the function $\psi \in L^p(\Omega)$[4] has a *weak partial derivative*

$$D^\alpha\psi \equiv \frac{\partial^{|\alpha|}\psi}{\partial x_1^{\alpha_1}\partial x_2^{\alpha_2}\cdots\partial x_m^{\alpha_m}}$$

iff there exists a function $\varphi \in L^1_{loc}(\Omega)$ such that

$$\forall f \in C_0^\infty(\Omega): \qquad \int_\Omega \varphi f = (-1)^{|\alpha|}\int_\Omega \psi D^\alpha f,$$

where $D^\alpha f$ is understood in the classical sense. In the positive case we write $\varphi = D^\alpha\psi$.

It is a straightforward task to show that the weak derivative indeed has some of the properties that we expect, namely that if the derivative exists in the classical sense, it will coincide with the weak derivative, and that if a function has (all the) weak derivatives of a given order, say k, then it also has (all) weak derivatives of orders $l < k$. Hence, we can define spaces of functions with a certain level of *weak smoothness* in an intuitive fashion, just as we do in the classical case. Here the condition of continuity of the derivatives of a certain order is naturally replaced by the condition of integrability. In this manner we arrive at the well-known definition of *Sobolev spaces* $W^{k,p}$:

Definition A.2. Let $k \in \mathbb{N}, p \in [1, \infty)$.[5] The Sobolev space $W^{k,p}(\Omega)$ is the linear function space $\{\psi \in L^p(\Omega) \mid D^\alpha\psi \in L^p(\Omega), \forall\alpha \text{ s.t. } 0 < |\alpha| \leq k\}$ equipped with the norm

$$\|\psi\|_{W^{k,p}(\Omega)} \equiv \left(\sum_{0\leq|\alpha|\leq k} \|D^\alpha\psi\|^p_{L^p(\Omega)}\right)^{1/p}. \tag{A.2}$$

[2]The case of Ω being a bounded domain is of interest but the following considerations are applicable to any bounded open set.

[3]Recall the 1:1 correspondence between sesquilinear and quadratic forms following from the polarization identities.

[4]More generally, $\psi \in L^1_{loc}(\Omega)$ can be used in the definition of weak derivatives. For our purposes, a definition for L^p-functions is sufficient.

[5]The distinct case $p = +\infty$ is usually included in the definition of Sobolev spaces but is not relevant to our considerations.

The essential observation from a function-theoretical point of view is that Sobolev spaces are complete (and thus Banach) with respect to the norm (A.2). A by-product of this observation is that for $p = 2$ they are Hilbert spaces (usually denoted H^k) with the inner product

$$(\varphi, \psi)_{H^k(\Omega)} = \sum_{0 \leq |\alpha| \leq k} (D^\alpha \varphi, D^\alpha \psi)_{L^2(\Omega)}$$

following from the polarization identities. For a more precise and general account of Sobolev spaces see e.g. chapter V in [6].

The main obstacle on the way towards a self-adjoint Laplace operator, which we have avoided addressing so far, is its domain. Classically we require that functions in its domain be at least of class $C^2(\Omega) \cap C^1\left(\overline{\Omega}\right)$. This is rather restrictive and results in the fact that, despite being densely defined and symmetric, $A \equiv -\Delta$ (defined on such a subspace of $L^2(\Omega)$) does not coincide with A^* but rather $A \subset A^*$.[6] The intuitive understanding might be that, since $D(A)$ is not 'broad' enough, the requirement on an element $\phi \in \mathcal{H} \equiv L^2(\Omega)$ to belong to $D(A^*)$ is 'too weak/mild', and consequently $D(A^*)$ is too 'broad' to coincide with $D(A)$, instead yielding $D(A) \subset D(A^*)$. The common approach then is to look for a *self-adjoint extension* of A, intuitively meaning that we wish to extend $D(A)$ (and hence confine $D(A^*)$) so as to 'meet in the middle' (while preserving the symmetry of A, of course). In a quantum mechanical or function-theoretical setting this is a well-studied (which by no means implies easy-to-solve) problem, and there are several (hopeful) ways to do this. One popular approach is to exploit the correspondence between linear operators and sesquilinear forms; the reader is probably familiar with the 1:1 correspondence between *bounded* sesquilinear forms and *bounded* linear operators, which can serve as a good example for the fact that there is indeed a close relationship between these two notions on a Hilbert space. The idea is as follows: a densely defined, symmetric and bounded-from-below operator A' defines a form a' of the same properties. This form may or may not be closed but is always closable; let us denote its closure by a. This form in turn defines our coveted self-adjoint (which includes densely defined) and bounded-from-below extension A of A'. The details and the representation theorems underlying this procedure, known as *Friedrichs extension*, can be found for instance in chapter IV.2 of [6] or chapter VI.2 of [9].

The key outcome is that the search for a self-adjoint 'version' of the Laplace operator leads to the Sobolev space $W^{1,2}(\Omega)$ as its 'natural habitat'. This is not too surprising: since we wished to extend $D(-\Delta)$, the generalizations provided by the notions of distributional Laplacian and weak derivatives do seem a hopeful way to achieve this. In this context it might be interesting to notice that the norm $\|\psi\|_{W^{1,2}} = \sqrt{\|\psi\|_{L^2}^2 + \|\nabla\psi\|_{L^2}^2}$ is exactly the 'graph norm' (A.1) of the quadratic form $a[\psi] = \|\nabla\psi\|_{L^2}^2$ associated with the generalized (distributional) Laplace operator (viewed as a map $a : L^2(\Omega) \to \mathbb{C}$). Using all of this, one can show that the operator $-\Delta_N^\Omega$ defined by

$$\begin{aligned} D(-\Delta_N^\Omega) \equiv & \{\psi \in W^{1,2}(\Omega) \mid \exists \eta \in L^2(\Omega) \text{ s.t.} \forall \phi \in W^{1,2}(\Omega) : (\nabla\phi, \nabla\psi) = (\phi, \eta)\} \\ = & \{\psi \in W^{1,2}(\Omega) \mid \Delta\psi \in L^2(\Omega) \wedge (\boldsymbol{n} \cdot \nabla)\psi = 0 \text{ on } \partial\Omega\} \\ -\Delta_N^\Omega \psi \equiv & \eta, \end{aligned}$$

where $\Delta\psi$ is understood in the distributional sense, is indeed self-adjoint and bounded from below. Note that the boundary condition together with the requirement $\Delta\psi \in L^2(\Omega)$ define a closed subspace of $W^{1,2}(\Omega)$ due to the completeness of L^p-spaces.

We are now left with the task to argue that the eigenvectors of this operator form an orthonormal basis of the space $L^2(\Omega)$. This can be shown to be equivalent to the operator of interest having a purely discrete

[6] Recall that the domain of A^* is given as $D(A^*) \equiv \{\phi \in \mathcal{H} \mid \exists \eta \in \mathcal{H} \text{ s.t.} \forall \psi \in D(A) : (A\psi, \phi) = (\psi, \eta)\}$ and that in the positive case we write $A^*\phi = \eta$.

spectrum, i.e. $\sigma_{ess} = \emptyset$, as well as to the operator having a compact resolvent (see Corollary 4.2.3 in [5]). This is in turn equivalent to the domain $D(A)$ (which in our case is a closed subspace of $W^{1,2}$) being compactly embedded in $\mathcal{H} = L^2(\Omega)$ by the proof of Theorem B.7 of [12]. It is therefore sufficient to prove that the embedding $W^{1,2}(\Omega) \hookrightarrow L^2(\Omega)$ is compact. This will be done via a detour to rectangular domains, where the eigenvalue problem $-\Delta_N^{\Omega_{rec}}\psi = \lambda\psi$ [7] can be solved directly (and relatively easily) using separation of variables (see e.g. [10]) and the explicitly expressed harmonic eigenfunctions are well-known to form a complete set in the corresponding L^2-space (as follows from the Stone-Weierstrass theorem). This is equivalent to the embedding $W^{1,2}(\Omega_{rec}) \hookrightarrow L^2(\Omega_{rec})$ being compact, which we wish to avail ourselves of as the central element in the chain

$$W^{1,2}(\Omega) \hookrightarrow W^{1,2}(\Omega_{rec}) \hookrightarrow L^2(\Omega_{rec}) \hookrightarrow L^2(\Omega), \tag{A.3}$$

where $\Omega \subset \Omega_{rec}$. However, before we triumphantly proclaim the composition of embeddings (A.3) to be our coveted compact embedding, we need to resolve a subtle yet important technical issue. Since compact operators form a two-sided ideal in *bounded* operators, we need to make sure that both the first (the exttension to Ω_{rec}) and the third (the restriction back to Ω are bounded). Whereas this is straightforward to see with the latter, it is not so easy with the former. Note that this is no issue with the Dirichlet Laplacian, whose 'natural domain' is provided by $W_0^{1,2}(\Omega) \equiv \overline{C_0^\infty(\Omega)}^{\|\cdot\|_{W^{1,2}(\Omega)}}$.[12] Since all functions $\psi \in W_0^{1,2}(\Omega)$ vanish on the boundary $\partial\Omega$, they can be easily extended (by 0) to Ω_{rec} (or any other open superset of Ω) 'in a bounded manner'. This is not the case for Neumann boundary conditions, where some (although not much) regularity must be required of $\partial\Omega$ for this to be true. Let us first formulate what we will mean by the *extension property*:

Definition A.3. Let $\Omega \subset \mathbb{R}^m$ be an open subset. Then Ω has the extension property iff there exists a *bounded* operator $E \in \mathcal{B}\left(W^{1,2}(\Omega), W^{1,2}(\mathbb{R}^m)\right)$ such that $\forall\psi \in W^{1,2}(\Omega) : E\psi|_\Omega = \psi$.

It is clear that if such a bounded extension operator exists, then via a composition with a restriction (which is always bounded) to an open set $\tilde{\Omega} \supset \Omega$ we can obtain a bounded extension to any open superset of Ω. We would therefore love to have a sufficient condition that would guarantee that Ω indeed has the extension property. Fortunately, this is the case thanks to Theorem 4.12. of [6]. It states that if Ω has a *minimally smooth boundary*, it always has the extension property. First of all, let us note that this requirement does not just serve the cause of the proof; there are indeed examples of bounded domains that do not admit a bounded extension operator (see e.g. Example 4.1 in [12]). Secondly, let us remark that the notion of minimally smooth boundary has a rather technical definition and we will not provide it in its full form as it is not instrumental for our cause. Instead, we will list several classes of open sets that comply with the requirement to demonstrate its 'mildness'. As far as bounded, open sets are concerned, all sets with boundary of class $C^{0,1}$ (i.e. 1-Hölder continuous)[8] have the extension property. This includes all (bounded, open) sets with a Lipschitz-continuous boundary (in the sense of the footnote), with a boundary of class C^1 (or higher) or, interestingly enough, all (bounded, open) convex subsets of $\mathbb{R}^m$ (by Theorem 4.2. in [6]). For all these classes of open sets Ω the extension $W^{1,2}(\Omega) \hookrightarrow W^{1,2}(\Omega_{rec})$ is bounded and hence the composition of embeddings (A.3) is compact. Thus, the spectrum of the Laplace operator is purely discrete and its eigenvectors form an orthonormal basis of $L^2(\Omega)$. This is true not only for Neumann boundary conditions but even for the more general Robin boundary conditions given as

[7] The additional sub- and superscript on $-\Delta$ emphasize the choice of boundary conditions and of the domain Ω rather than the use of the distributional Laplacian. Here the Laplace operator allows for a purely classical approach.

[8] α-Hölder continuous functions are real/complex-valued functions satisfying $|f(x) - f(y)| \leq C\|x - y\|^\alpha$. Roughly speaking, a boundary is of class $C^{0,1}$ if it admits local coordinates $(x_1, \ldots, x_m)$ in which the boundary points are locally given by a map $x_m = \Phi(x_1, \ldots, x_{m-1})$, which is of class $C^{0,1}$. For details consult chapter V.4.1 in [6].

$$\alpha\psi + (\boldsymbol{n} \cdot \nabla)\psi = 0 \text{ on } \partial\Omega,$$

with $\alpha \in L^\infty(\partial\Omega)$. [12]