Computer-Assisted Bifurcation Diagram Validation and Applications in Materials Science

Thomas Wanner
Department of Mathematical Sciences
George Mason University
Fairfax, VA 22030, USA

November 27, 2015

Abstract

Most differential equation models in the applied sciences involve parameters which often cannot be determined with absolute certainty. It is therefore important to study such models for whole parameter ranges, and in particular, to detect parameter values at which the system behavior changes qualitatively. A first step towards accomplishing this goal has to be the understanding of the set of equilibria or stationary solutions of the model. In this lecture, we demonstrate how rigorous computational techniques can be used to validate bifurcation diagrams, both in finite- and certain infinite-dimensional problems. We focus particularly on the verification of branches, as well as on saddle-node and symmetry-breaking pitchfork bifurcations. Throughout, our approach will be applied to two examples from materials science. On the one hand, we consider lattice dynamical systems, specifically the discrete Allen-Cahn equation. Such systems have been proposed as more realistic models due to the existing underlying discreteness, and unlike their continuum counterparts, lattice models can account for phenomena such as pinning. As our second example, we consider the diblock copolymer model, which models microphase separation.

AMS subject classifications: 35B40, 35B41, 35K55, 60F10, 60H15, 74N99

Keywords: Bifurcation diagram, pitchfork bifurcations, saddle-node bifurcations, symmetry, computer assisted-proof, numerical implicit function theorem

Contents

1 Introduction

Differential equations are one of the central tools in the applied sciences. They model dynamical processes, and their study can lead to a deeper understanding of the underlying laws of nature. During the modeling process, one usually tries to incorporate all the essential dynamical variables into the final differential equation, while input quantities which are considered constant enter as parameters. In this way, most applied models can be described abstractly as a parameter-dependent differential equation of the form

$$\dot{u} = f(\lambda, u) \,, \tag{1}$$

where the smooth right-hand side f depends on a possibly vector-valued parameter λ, and on the vector u of dynamical variables. While in theory the study of (1) involves fixing the "correct" parameter value λ and then trying to understand the resulting dynamical behavior, in practice the parameter λ often cannot be determined with absolute certainty. It is therefore important to study (1) for whole parameter ranges, and in particular, to detect parameter values at which the system behavior changes significantly.

The first natural step to performing such a parameter-dependent study is of course to restrict one's attention to simpler building blocks for the dynamics, and then build more structure around them through other techniques of the qualitative theory of differential equations. The most basic building blocks are solutions of (1) which are constant in time, and these *equilibria*, or *stationary solutions*, satisfy the algebraic equation

$$f(\lambda, u) = 0 \,, \tag{2}$$

which is obtained from (1) by setting the left-hand side derivative equal to zero. If, for a given function f we plot all solution pairs (λ, u) of (2) in a diagram, we obtain what is usually

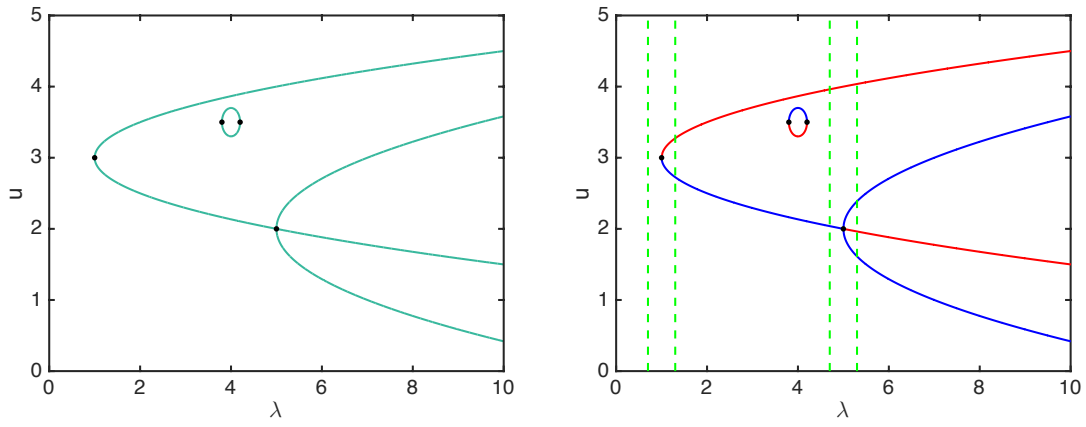

Figure 1: Sample bifurcation diagram for a scalar differential equation of the form (1). The left panel shows the set of equilibrium solutions, i.e., all solution pairs (λ, u) satisfying (2), with the nonlinearity given by (3). Black dots indicate the bifurcation points in the diagram, which from left to right correspond to three saddle-node and one pitchfork bifurcation. The right panel includes stability information for the associated dynamics of (1). Red equilibrium points are unstable, while blue ones correspond to stable solutions. The vertical dashed green lines indicate parameter values that are referenced in the text.

referred to as a *bifurcation diagram*. Consider for example the function $f : \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(\lambda, u) \;=\; \left(1 + 4(u-3)^2 - \lambda\right)\left(5 + 2(u-2)^2 - \lambda\right)\left((\lambda - 4)^2 + \left(u - \frac{7}{2}\right)^2 - \frac{1}{25}\right) , \qquad (3)$$

whose zero set is partially shown in the left panel of Figure 1. From this bifurcation diagram, one can readily deduce the number of equilibrium solutions for each parameter value λ^*, by intersecting the green zero set with the vertical line $\lambda = \lambda^*$. Note that in general, if (λ^*, u^*) is a point on the bifurcation diagram and if $\lambda \approx \lambda^*$, then there exists a unique value u close to u^* such that (λ, u) is also a solution of (2). This is no longer true if we are at an equilibrium bifurcation point, which we define informally as follows.

Definition 1.1 (Bifurcation Point)**.** *A solution pair (λ^*, u^*) of (2) is called* (equilibrium) bifurcation point*, if as we increase the parameter λ from slightly to the left of λ^* to slightly to the right of λ^*, the number of solutions (λ, u) of (2) with $u \approx u^*$ changes.*

One can easily see that this definition is satisfied exactly at the four black dots marked in the left panel of Figure 1. For the two solutions at $(\lambda^*, u^*) = (1, 3)$ and $(\lambda^*, u^*) = (3.8, 3.5)$, the number of solutions changes from 0 over 1 to 2 with increasing λ, while for $(\lambda^*, u^*) = (4.2, 3.5)$ it changes from 2 over 1 to 0. In the following, we will call bifurcation points of this type *saddle-node bifurcation points*, and they are characterized by the creation or destruction of a pair of equilibria as λ increases. In contrast, the bifurcation point at $(\lambda^*, u^*) = (5, 2)$ is called a *pitchfork bifurcation point*, since the associated jump from one to three equilibrium solutions (or vice versa) resembles the farm tool bearing the same name.

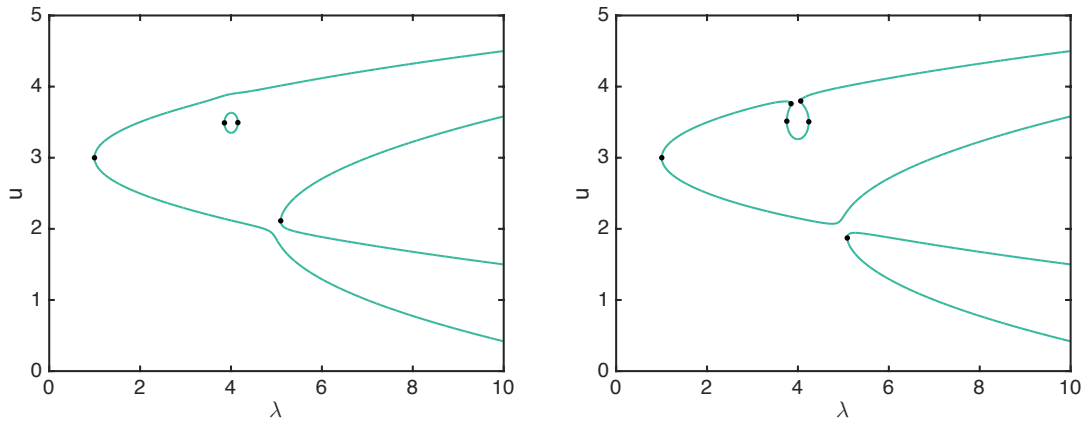

Figure 2: Bifurcation diagram changes under perturbations of f. The left and right panels show the equilibrium bifurcation diagrams which correspond to subtracting or adding, respectively, the constant value $1/5$ to the function f defined in (3), which generates the left panel of Figure 1. Note that large parts of the diagram persist qualitatively, but that the pitchfork bifurcation point breaks up, and that the island close to $(4, 3.5)$ merges with a nearby branch. In both cases, new saddle-node bifurcation points are created.

While the bifurcation diagram in the left panel of Figure 1 provides information on the structure of the equilibrium set of (1) with nonlinearity (3), it does not include information on stability properties of these stationary states. If available, such information can be incorporated into the diagram as shown in the right panel of Figure 1. In the new diagram, the branches are colored according to the stability properties of their equilibrium solutions, with blue corresponding to asymptotically stable equilibria and red to unstable ones. Such a more detailed diagram indicates for example that for $\lambda = 0.7$ all solutions of (1) are strictly increasing and converge to $+\infty$, while for $\lambda = 1.3$ only solutions starting at initial values larger than that of the unstable equilibrium exhibit this behavior. The remaining ones will be attracted by the second created stationary state. Similarly, while for $\lambda = 4.7$ all solutions originating for example in the interval $[0, 3]$ will be attracted by the same asymptotically stable state, for $\lambda = 5.3$ there are two newly created possible stable states.

Needless to say, the sample function f given in (3) was conceived in such a way that its bifurcation structure can easily be inferred directly from the form of the equations, since it is defined as a product of terms which define simple geometric shapes. Note, however, that if instead of f we consider the shifted nonlinearities $f \pm 1/5$, then numerical methods have to be used to determine the bifurcation structure of (2). The resulting bifurcation diagrams for the functions $f - 1/5$ and $f + 1/5$ are shown in the left and right panels of Figure 2, respectively. It is worth noting that while these small shifts in f result in only minor changes in the overall global structure of the zero set, the observed local changes can be quite significant. For example, in both diagrams the pitchfork bifurcation is no longer present, and it breaks up in two different ways. Similarly, while the left panel still contains the zero loop close to $(4, 3.5)$, in the right diagram this loop has merged with one of the other zero branches. In all of these changes, new saddle-node bifurcations are introduced. In fact, both bifurcation diagrams in

Figure 2 contain only saddle-node bifurcation points, and no other variety. We will return to this important observation later in the lecture.

From a mathematical perspective, the study of bifurcation diagrams has a long history, and many excellent introductions to the subject exist. See for example [8, 12, 30, 54], as well as the references therein. These books contain many different results which ensure the existence of a wealth of bifurcation points (going well beyond our limited definition above), and such results can be achieved through both topological and analytical approaches. Of fundamental importance in this context is the implicit function theorem, which we only state in a simple finite-dimensional setting.

Theorem 1.2 (Implicit Function Theorem)**.** *Let $f : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ denote a smooth function, and assume that $f(\lambda^*, u^*) = 0$ for a pair $(\lambda^*, u^*) \in \mathbb{R} \times \mathbb{R}^n$. Furthermore, suppose that the Jacobian matrix $D_u f(\lambda^*, u^*) \in \mathbb{R}^{n \times n}$ of the function f with respect to u at (λ^*, u^*) is invertible. Then there exists a neighborhood $U \subset \mathbb{R}^n$ of u^*, a constant $\delta > 0$, as well as a smooth function $g : [\lambda^* - \delta, \lambda^* + \delta] \to \mathbb{R}^n$ such that the following holds:*

$$(\lambda, u) \in [\lambda^* - \delta, \lambda^* + \delta] \times U \ \text{ satisfies } \ f(\lambda, u) = 0 \qquad \text{if and only if} \qquad u = g(\lambda) \, .$$

In other words, all solutions of (2) in the set $[\lambda^* - \delta, \lambda^* + \delta] \times U$ lie on the graph of g.

This result shows that as long as the Jacobian matrix $D_u f(\lambda^*, u^*)$ is invertible, small variations in the parameter λ do not lead to changes in the number of solutions (λ, u) with $u \approx u^*$, i.e., one obtains the following direct corollary.

Corollary 1.3 (Necessary Condition for Bifurcation)**.** *In the situation of Theorem 1.2, if the pair (λ^*, u^*) is a solution of (2) with invertible Jacobian matrix $D_u f(\lambda^*, u^*) \in \mathbb{R}^{n \times n}$, then (λ^*, u^*) cannot be an equilibrium bifurcation point. In other words, a necessary condition for the point (λ^*, u^*) to be a bifurcation point is that $D_u f(\lambda^*, u^*)$ is not invertible.*

The importance of both the implicit function theorem and its simple corollary cannot be overemphasized. While Theorem 1.2 allows one in theory to establish branch segments of equilibria, Corollary 1.3 provides an algebraic criterion for where bifurcation points can occur. We would like to point out, however, that Corollary 1.3 is usually not sufficient to ensure the actual occurrence of a bifurcation point. For this, one needs to establish additional conditions, which vary with the type of the bifurcation point.

Despite its importance, the applicability of Theorem 1.2 is limited in many practical applications. For example, the bifurcation diagrams in Figure 2 were computed numerically, and we actually do not have an explicit formula for any of the equilibrium solutions for the shifted functions $f \pm 1/5$. The goal of this lecture is to demonstrate that in such situations, rigorous computational techniques can be used to verify bifurcation diagrams through a combination of interval arithmetic and theoretical fixed point arguments. This necessarily has to address both the validation of branches, and the location and verification of bifurcation points. In order to keep the lecture as self-contained as possible, we restrict our attention to saddle-node bifurcation points and pitchfork bifurcations induced through symmetry-breaking.

In more detail, the remainder of these notes is organized as follows. In Section 2 we introduce two sample models from materials science which will serve as the main examples.

While Section 2.1 is devoted to the lattice dynamical system given by the discrete Allen-Cahn equation, Section 2.2 introduces the diblock copolymer model for phase separation. These models cover the case of high-dimensional ordinary differential equations, as well as parabolic partial differential equations with gradient structure. The theoretical foundations for computer-assisted validation of bifurcation diagrams are the subject of Section 3. We begin in Section 3.1 to formulate a numerical version of the implicit function theorem, which will then be used in Section 3.2 to validate branch segments by employing a tangent-based predictor-corrector method. In the subsequent Section 3.3, it is demonstrated that the folds around saddle-node bifurcation points can be resolved using a suitable coordinate transformation. Throughout the section, the results will be applied to our two basic examples. Finally, Section 4 turns towards direct methods for verifying bifurcation points. The case of saddle-node bifurcations is the subject of Section 4.1, while symmetry-breaking pitchfork bifurcations are addressed in Section 4.2. For related results we refer the reader to [2, 16, 21, 39, 40].

2 Two Examples from Materials Science

Throughout this lecture, we will illustrate the various results in the context of two concrete examples. The first of these is a high-dimensional ordinary differential equation, which arises as a discrete model for the motion of an interface between two phases of a binary alloy. The second example is concerned with microphase separation in diblock copolymers, and the resulting model is a parabolic partial differential equation.

2.1 The Discrete Allen-Cahn Equation

One of the fundamental models for phase separation in alloys is the Allen-Cahn equation, which is a parabolic partial differential equation of the form

$$u_t = \Delta u + \lambda f(u) \quad \text{in} \quad \Omega\,, \qquad \text{subject to} \quad \frac{\partial u}{\partial \nu} = 0 \quad \text{on} \ \partial\Omega\,. \tag{4}$$

In this equation, the domain $\Omega \subset \mathbb{R}^d$ is bounded with a suitably smooth boundary $\partial\Omega$, the strictly positive real number λ is a parameter, and the vector ν denotes the outward unit normal vector to the domain Ω at points on its boundary $\partial\Omega$. Furthermore, the nonlinearity f is given by the cubic polynomial $f(u) = u - u^3$. For any solution u of (4) the value $u(t, x)$ denotes the local alloy composition at time t and spatial location $x \in \Omega$, and values of u close to ± 1 represent two different pure phases, or alloy components, of the underlying material; values between -1 and $+1$ correspond to mixtures of the phases or components, with $u = 0$ implying an equal split. One can show that if we choose an initial condition $u(0, \cdot) : \Omega \to \mathbb{R}$ which is a small random perturbation of the stationary state $u \equiv 0$, then as time increases one first observes instantaneous phase separation, followed by convergence of the solution to either the equilibrium $u \equiv +1$ or $u \equiv -1$ which is facilitated by the motion of interfaces between the two phases.

While the Allen-Cahn model does describe the interface motion fairly accurately, the above-mentioned long-term behavior is an idealization. In experiments, it is observed that

interfaces can become stuck through a phenomenon called *pinning*, and this cannot be observed in (4). To remedy this shortcoming, discretized versions of the Allen-Cahn equation have been proposed in the literature, see for example [9, 13, 14, 15, 20, 22, 23, 24, 26, 34], as well as the references therein. For the purposes of this lecture, we consider the *discrete Allen-Cahn equation*

$$\dot{u}_k = u_{k+1} - 2u_k + u_{k-1} + \lambda f(u_k) \qquad \text{for all} \qquad k = 1, \ldots, n \,, \tag{5}$$

where we set $u_0 = u_1$ and $u_{n+1} = u_n$, and the nonlinearity is again given by $f(u) = u - u^3$. Notice that the discrete Allen-Cahn equation is a coupled system of ordinary differential equations for the n unknown real-valued functions $u_1, \ldots, u_n$ which depend on time t. In fact, equation (5) can be viewed as an approximation to the Allen-Cahn model (4) on the one-dimensional domain $\Omega = (0, 1)$ in the following sense. Consider the spatial discretization points

$$x_k = \frac{2k-1}{2n} \quad \text{for} \quad k = 1, \ldots, n \,,$$

and use the finite difference approximation of the one-dimensional Laplacian $\Delta u = u''$ given by

$$u''(x_k) \;\approx\; \frac{u(x_{k+1}) - 2u(x_k) + u(x_{k-1})}{\Delta x^2} \quad \text{with} \quad \Delta x = \frac{1}{n} \,.$$

Then the solution u_k of (5) can be interpreted as an approximation of $u(\cdot, x_k)$ if we replace λ by λ/n^2 and rescale time appropriately. Note in particular that the definition of the discretization points leads to the correct approximation of the boundary conditions.

Despite the above motivation via finite difference approximations, the discrete Allen-Cahn equation (5) has been proposed as a model in its own right, and we will consider it as such. For large values of the parameter λ one can easily see that (5) has an abundance of equilibrium solutions. For this, consider any vector $\bar{u} \in \mathbb{R}^n$ with $\bar{u}_k \in \{-1, 0, 1\}$ for $k = 1, \ldots, n$. Then the implicit function theorem can be used to show that for large enough $\lambda > 0$ there exists an equilibrium of the discrete Allen-Cahn equation near $\bar{u}$. In other words, if λ is sufficiently large, the system (5) has at least 3^n stationary states, and we call solutions arising in this way *mosaic solutions*. While some of these equilibria have counterparts in the regular Allen-Cahn model (4) on the domain $\Omega = (0, 1)$, the vast majority does not — and it is exactly this discrepancy that is connected to the phenomenon of pinning.

The significance of the mosaic equilibria for pinning becomes evident when we turn our attention to their stability. One can show that if the vector $\bar{u} \in \{-1, 0, 1\}^n$ contains one or more zero components, then the corresponding equilibrium is unstable, while the remaining 2^n vectors $\bar{u} \in \{\pm 1\}^n$ give all rise to asymptotically stable stationary states, i.e., interface motion will get stuck upon entering the domains of attraction of these states.

But how are all of these attracting states created? At which parameter values λ can they be observed first? From a numerical point of view this question can easily be answered using the path-following software AUTO [18]. In the left panel of Figure 3 we plot a number of equilibrium branches which lead to mosaic solutions, and all of them seem to be created via saddle-node bifurcation points. It is worth noting that the λ-coordinates of the saddle-node points are staggered in an intriguing way, and it was conjectured in [22] that the precise

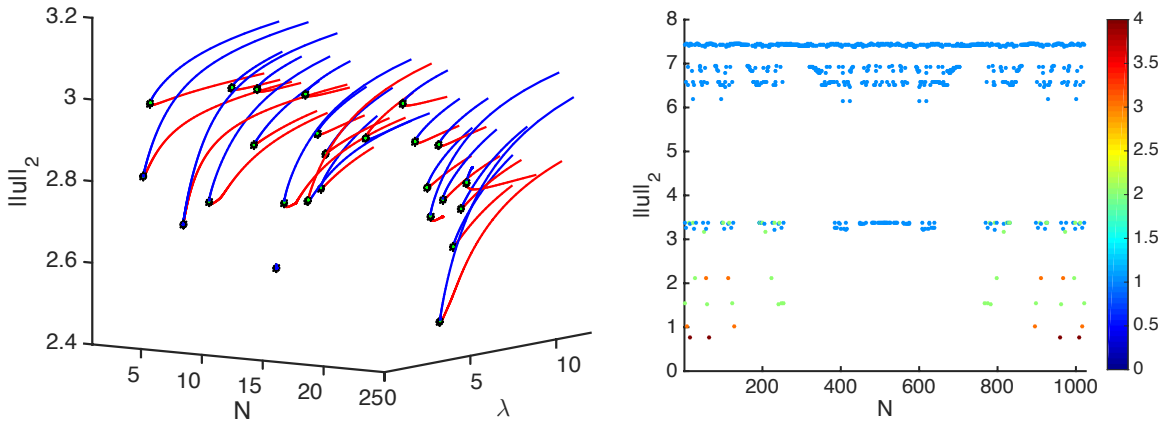

Figure 3: Mosaic equilibrium solutions of the discrete Allen-Cahn model (5) for $n = 10$. The left image shows sample branches of such stationary states, which are created through saddle-node bifurcations. In the right image, we show all 2^n saddle-node bifurcation points which lead to mosaic solutions close to vectors in $\{\pm 1\}^n$. These points are color-coded by the smallest grain size in the solution, as encoded in the colorbar.

location is correlated with the *grain size* of the mosaic solution, which is defined as the length of the smallest consecutive string of nonzero components of the same sign in the vector $\bar{u}$. We would like to point out that Grant [22] could partially answer this question for the special case of piece-wise linear nonlinearities f, which are qualitatively of the same form as the standard cubic polynomial $f(u) = u - u^3$. However, his methods do not seem to be applicable to this polynomial nonlinearity.

As we will demonstrate later on, rigorous computational techniques can be used to derive computer-assisted proofs linking the grain size of a mosaic solution to the specific location of the saddle-node bifurcation points. This leads to the right panel in Figure 3, which contains all 2^n saddle-node bifurcation points connected to mosaic solutions close to vectors in $\{\pm 1\}^n$. These points are color-coded by the smallest grain size in the solution, as encoded in the colorbar, and they clearly show the relation between the staggered creation of the solutions and their grain size. While the detailed results are contained in [47], many of the results described later in this lecture are taken from this paper.

2.2 The Diblock Copolymer Model

Our second example is concerned with diblock copolymers, which can be found in products such as upholstery foams, adhesive tapes, and asphalt additives [4, 5]. Their building blocks are macromolecules which are obtained by joining two chemically distinct polymer blocks, each of which is a linear chain of identical monomers. The two involved monomer types are usually thermodynamically incompatible, and this typically leads to phase separation within the material which is constrained by the covalently bonded chains — thereby resulting in a phenomenon called *microphase separation*, where the two blocks separate on a mesoscopic scale. One of the basic models for microphase separation in diblock copolymers has been

described in [3, 37, 38]. Consider a material constrained to the bounded domain $\Omega \subset \mathbb{R}^d$. Then the model proposes a free energy functional for the relative macroscopic monomer concentration u, i.e., for the difference between the two monomer volume fractions, given by

$$E_{\lambda,\sigma}[u] = \int_\Omega \left(\frac{1}{2\lambda} |\nabla u|^2 + F(u) \right) dx + \frac{\sigma}{2} \int_\Omega \left| (-\Delta)^{-1/2} (u(x) - \mu) \right|^2 dx \, . \tag{6}$$

In this formula, the number $\mu = \int_\Omega u \, dx / |\Omega|$ denotes the difference of the averaged monomer masses, and the nonlinear function F is a double-well potential with global minima at ± 1. Throughout this paper, we consider the specific potential $F(u) = (u^2 - 1)^2/4$. Note that the energy functional (6) is the standard van der Waals free energy, but with an additional nonlocal term which involves the square root of the inverse Laplacian on a space with zero total average. There are different ways in which one can associate gradient-like dynamics to (6), and in its standard form the diblock copolymer model considers the equation

$$\begin{aligned} u_t &= -\Delta \left(\Delta u + \lambda f(u) \right) - \lambda \sigma (u - \mu) \quad \text{in } \Omega \, , \\ \mu &= \frac{1}{|\Omega|} \int_\Omega u(x) \, dx \, , \quad \text{and} \quad \frac{\partial u}{\partial \nu} = \frac{\partial \Delta u}{\partial \nu} = 0 \quad \text{on } \partial\Omega \, , \end{aligned} \tag{7}$$

where $f(u) = -F'(u) = u - u^3$, and ν denotes again the unit outward normal vector to the boundary $\partial\Omega$. In this formulation, the value of u describes the local material composition in the following way, which is reminiscent of our discussion in the previous section. Values of $u(t, x)$ close to $+1$ are interpreted as only monomer A being present at point $x \in \Omega$ and at time $t \geq 0$, and the value -1 indicates that only monomer B is present; values in between correspond to mixtures of the two monomers, with zero representing an equal mixture. The parameter μ denotes the average mass of the mixture, and the two remaining parameters $\lambda > 0$ and $\sigma \geq 0$ are dimensionless interaction lengths. Informally, $\lambda > 0$ being large corresponds to short range repulsions being strong, and this induces a strong tendency to separate. On the other hand, σ being large represents strong long range chain elasticity forces, which leads to a strong tendency for holding together. We would like to point out that for $\sigma = 0$ the diblock copolymer model (7) reduces to the celebrated Cahn-Hilliard equation [10], which serves as a basic model for the phase separation phenomena spinodal decomposition [32, 33, 45, 46, 49] and nucleation [6, 7, 17].

The derivation of the diblock copolymer model (7) guarantees that along any solution u the energy $E_{\lambda,\sigma}[u]$ is strictly decreasing, unless the solution is an equilibrium. In fact, one can show that the right-hand side of (7) is the negative gradient of (6) in the H^{-1}-topology. Since the energy clearly is bounded below by zero, one would therefore expect that any solution will eventually converge to an equilibrium of the diblock copolymer model, and generically, this limit should be an (at least local) minimizer of the energy. But what can be said about this limiting stationary state? For this, we consider the question of pattern formation during the material production. In the initial polymer melt, both monomers will produce a homogeneous mixture throughout the container, i.e., the initial state $u(0, \cdot) \approx \mu$ will be a small random perturbation of the constant equilibrium μ. If the parameter λ is large enough, one can show that this stationary state is unstable, thereby triggering spontaneous phase separation.

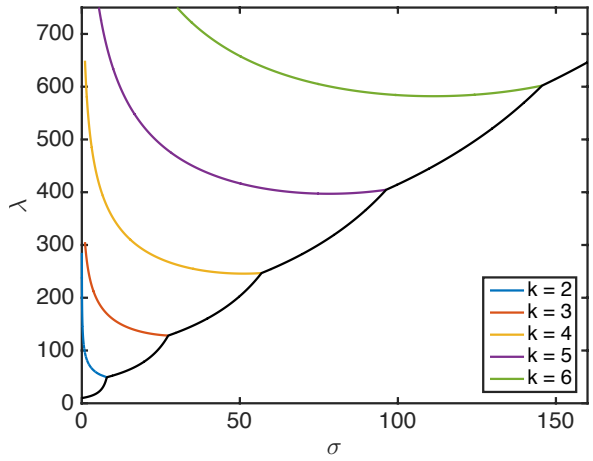

Figure 4: Structure of typical long-term limits for solutions of (7) on $\Omega = (0, 1)$ which start close to the homogeneous state $\mu = 0$. The image partitions the σ-λ-parameter quadrant into regions with different behaviors of the long-term limit. Below the black lower envelope the homogeneous equilibrium is stable, i.e., no phase separation occurs and all solutions converge to the constant state. For parameter combinations in the remaining region, if the parameter point lies below the plotted curve for $k + 1$ and above the one for k, the typical long-term limit is periodic and has exactly k zeros.

In the recent paper [27], the above question was studied through a combination of numerical and rigorous bifurcation-theoretic techniques for the special case $\mu = 0$ and for the one-dimensional domain $\Omega = (0, 1)$. In this paper, it was shown that while the diblock copolymer model exhibits a high level of *multistability*, i.e., the energy functional $E_{\lambda,\sigma}$ typically has many local minimizers, solutions of (7) which originate close to μ converge with high probability to limiting states with the same geometry. These limiting states seem to be periodic solutions with k zeros, whose periodicity depends only on the specific choice of the parameters λ and σ. This is illustrated further in Figure 4, which partitions the σ-λ-parameter quadrant into regions with different behaviors of the long-term limit. Below the black lower envelope one can show that the homogeneous state $\mu = 0$ is stable, i.e., no phase separation occurs. For parameter combinations in the remaining regions, the number k of zeros of the typical periodic long-term limit is one less than the value associated with the plotted curve which bounds the region from above. In other words, below the blue curve the limits have one zero, between the blue and orange curves they have two, and so on. We would like to emphasize that these results were obtained through numerical Monte Carlo type simulations, and that different long-term limits can be observed even at the same parameter combination. However, *most solutions* (by a large margin) starting close to the homogeneous state exhibit the above behavior, and the transitions between preferred limits do occur along the well-defined curves shown in Figure 4. For more details, we refer the reader to [27, 51].

To explain the origin of the delineating curves in Figure 4 further, we need to turn our attention to the equilibrium solutions of the diblock copolymer model, where again we only consider the one-dimensional domain $\Omega = (0, 1)$ and total mass $\mu = 0$. Such solutions satisfy

the nonlinear elliptic problem

$$-\Delta\left(\Delta u + \lambda f(u)\right) - \lambda\sigma(u-\mu) = 0 \quad \text{in } \Omega\,, \tag{8}$$

subject to the mass constraint $\mu = \int_\Omega u(x)\,dx/|\Omega|$ and homogeneous Neumann boundary conditions for both u and Δu. One can easily see that for all $\lambda > 0$ and $\sigma \geq 0$ this equation has the solution $u \equiv \mu$, and we will call this constant equilibrium the *trivial solution* of (8).

For the special case $\sigma = 0$ the nonlinear problem (8) reduces to the stationary Cahn-Hilliard equation, whose complete equilibrium set has been derived rigorously in [25]. In this paper it is shown that for $\mu = 0$ and as λ increases from zero, new nontrivial solutions appear close to the trivial solution $u \equiv 0$ through an infinite series of pitchfork bifurcations. These bifurcations occur at the parameter values $\lambda = k^2\pi^2$, where $k \in \mathbb{N}$, and for every k-value, the two new nontrivial solutions qualitatively resemble the function $\varphi_k(x) = \sqrt{2}\cos k\pi x$. The nontrivial solutions grow in amplitude with increasing λ, and ultimately converge to a maximum norm close to one as $\lambda \to \infty$. This behavior is depicted in the top left bifurcation diagram of Figure 5, where the horizontal axis corresponds to the parameter λ, and the vertical axis measures the $L^2(0,1)$-norm of an equilibrium solution u. Each colored point in the image corresponds to at least one solution of (8) for $\sigma = 0$ and mass $\mu = 0$. Points on the horizontal line $\|u\|_{L^2(0,1)} = 0$ are clearly the trivial solutions, while points on the emanating curves are nontrivial ones. Moreover, each point with $\|u\|_{L^2(0,1)} > 0$ gives rise to two solutions, since for the chosen parameter values, if u is a solution of (8), then so is $-u$. Thus, the bifurcation diagram shown in the top left of Figure 5 indicates that for every $k^2\pi^2 < \lambda \leq (k+1)^2\pi^2$, the Cahn-Hilliard model with mass $\mu = 0$ has exactly $2k+1$ equilibrium solutions.

How do these rigorous results change for the diblock copolymer case $\sigma > 0$? At first glance, one would assume that any changes should be minor, since the diblock copolymer model arises from the Cahn-Hilliard model through a regular perturbation at $\sigma = 0$. Yet, nothing could be further from the truth. In the remaining images of Figure 5, we show bifurcation diagrams for increasing values of σ, all of which were computed numerically using AUTO [18], see also [27]. As in the Cahn-Hilliard case, every point on a curve corresponds to at least one equilibrium solution. In fact, due to symmetries in the equation, some points correspond to as many as four solutions. These diagrams indicate that the bifurcation structure of the diblock copolymer model for $\sigma > 0$ is considerably more complicated than the one in the Cahn-Hilliard setting, with many secondary bifurcations.

One of the main results in [27] is the insight that the basic qualitative changes in the bifurcation diagrams of (8) for $\mu = 0$ as σ increases from zero can be uncovered by a local bifurcation analysis. As σ increases, the branches emanating from the trivial solution line move to the right, but at varying speeds. For example, the 1-branch, i.e., the branch on which solutions have the same shape as $\varphi_1(x) = \sqrt{2}\cos\pi x$, moves quickly to the right with increasing σ, interacts with all subsequent k-branches for $k \geq 2$, and then moves off to infinity and disappears for $\sigma \geq \pi^2$. Next, the 2-branch moves to the right and interacts with the k-branches for $k \geq 3$, before it moves off to infinity and ceases to exist for $\sigma \geq 2^2\pi^2$, and so forth. All of the local branch interactions can be analyzed rigorously using a Lyapunov-Schmidt reduction at two-dimensional kernels [27], and this implies that every local interaction involves a secondary bifurcation point on one of the participating branches. One of these secondary

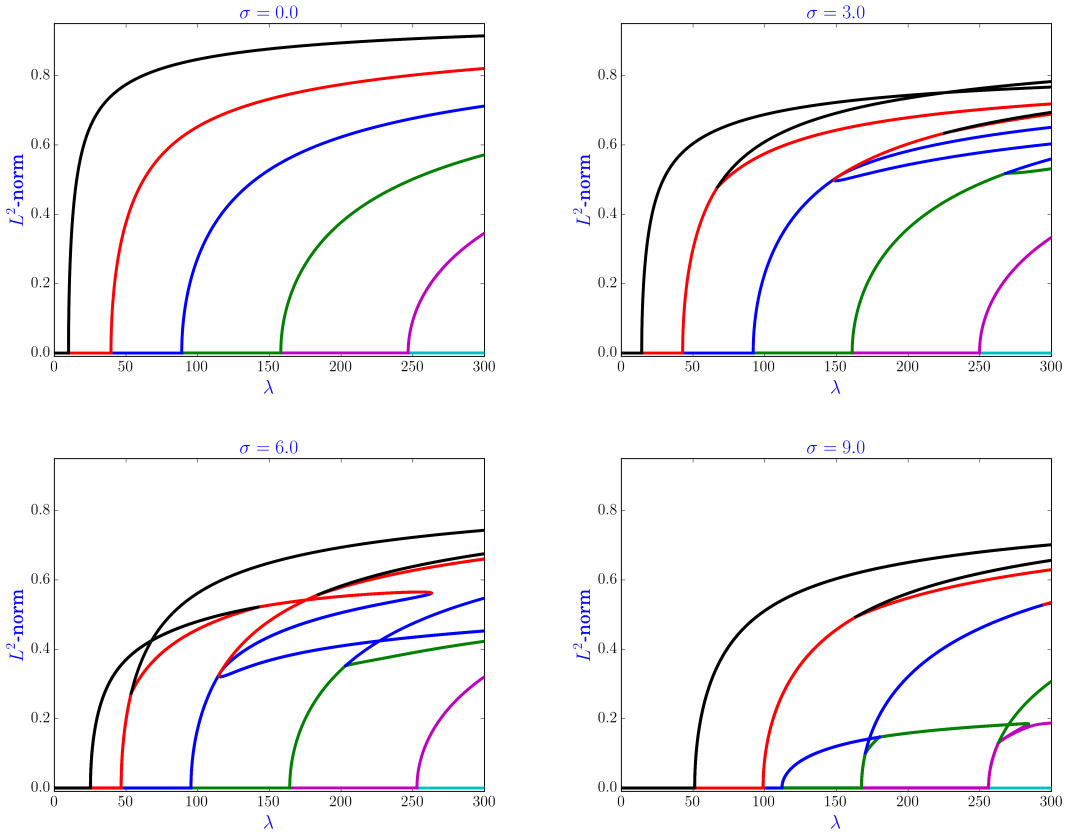

Figure 5: Sample equilibrium bifurcation diagrams for the diblock copolymer model (7) on the one-dimensional domain $\Omega = (0, 1)$ and for total mass $\mu = 0$. From top left to bottom right the bifurcation diagrams are for $\sigma = 0$, 3, 6, and 9, respectively. In each diagram, the vertical axis measures the $L^2(0, 1)$-norm of the solutions, and the horizontal axis uses the parameter λ. The solution branches are color-coded by the Morse index of the solutions, and black, red, blue, green, magenta, and cyan correspond to indices 0, 1, 2, 3, 4, and 5, respectively.

bifurcation points can be seen in the lower left image of Figure 5 for $\lambda \approx 50$. It represents a pitchfork bifurcation which creates two index one solutions, and which renders the 2-branch equilibrium asymptotically stable. In other words, the secondary bifurcation point turns an index one equilibrium into a local minimizer of the energy (6).

It turns out that these secondary bifurcation points lie at the heart of the long-term asymptotic behavior of solutions u of (7) which originate close to the homogeneous state μ. To explain this, consider again the secondary pitchfork bifurcation in the lower left diagram of Figure 5. The parameter combination $\lambda \approx 50$ and $\sigma = 6$ at which this bifurcation point can be observed lies precisely on the blue curve in Figure 4. In fact, every point on the blue curve corresponds to a parameter combination at which the 2-branch has a pitchfork bifurcation. Similarly, the remaining colored curves with labels $k \geq 3$ in Figure 5 represent the locations

of pitchfork bifurcation points on the k-branch emanating from the trivial solution line. As in the case $k = 2$, all of these bifurcation points are stabilizing in the sense that they render the equilibrium on the k-branch asymptotically stable.

In view of the above numerical results it is natural to wonder whether any of these statements can be proven. Unfortunately, even the equilibrium structure of (7) for $\sigma > 0$ lies beyond the reach of current analytical techniques, since the ones used in [25] no longer apply. The situation is even more bleak in the case of the pitchfork bifurcation points. It is possible, however, to employ rigorous computational techniques to obtain computer-assisted proofs. In this way, branches of equilibrium solutions could be established in [50], and the location of bifurcation points can be rigorously verified using the techniques of [31]. Some of these results will be described in the following.

3 Branches and the Numerical Implicit Function Theorem

Beginning with this section we present the necessary theoretical foundations for validating bifurcation diagrams. Our approach for this will be somewhat different from the radii polynomial approach taken by some of the other lecturers, but is of course related. Rather than setting up explicit nonlinear systems for various problems and verifying a contraction condition, we return to the introduction of this lecture. As mentioned there, the implicit function theorem is one of the fundamental tools for understanding bifurcations, even though its applicability in the original form is somewhat limited. Based on this, we will formulate a *numerical implicit function theorem* in Section 3.1 below, which removes the assumption of a precisely known zero by the knowledge of an approximative zero. This theorem is then applied in Section 3.2 to validate branch segments of equilibrium solutions for nonlinear equations, based on a tangent-based predictor-corrector approach. Finally, Section 3.3 will explain how the numerical implicit function theorem can be used to rigorously follow branches around saddle-node bifurcation points. Much of the presentation in Section 3 reflects the one of [47], and we refer the reader to this paper for proofs of the involved theorems.

3.1 A Numerical Implicit Function Theorem

In its original formulation, the implicit function theorem guarantees the existence of a branch of solutions of a nonlinear parameter-dependent equation, provided we have

- an exact solution of the problem at a specific parameter value, as well as
- an invertibility condition for the linearization of the nonlinear problem at this solution.

These assumptions certainly seem reasonable, but in practice, they are usually hard to verify. Especially for high-dimensional nonlinear problems, one cannot expect to have an interesting precise solution available. Generally, the best we can hope for is a good numerical approximation. But is this enough to still conclude the existence of a nearby branch of solutions?

To answer this question, one has to take a closer look at the proof of the implicit function theorem. While there are many different approaches for doing this, see for example the

excellent account in [29], we focus on the following technique which is based on the contraction mapping principle. To fix notation, we consider parameter-dependent problems of the form

$$\mathcal{G}(\alpha, x) = 0 \,, \tag{9}$$

where $\mathcal{G} : \mathcal{P} \times \mathcal{X} \to \mathcal{Y}$ is a Fréchet differentiable nonlinear mapping between two Banach spaces $\mathcal{X}$ and $\mathcal{Y}$, and the parameter α is taken from a Banach space $\mathcal{P}$. The norms on these Banach spaces are denoted by $\|\cdot\|_{\mathcal{P}}$, $\|\cdot\|_{\mathcal{X}}$, and $\|\cdot\|_{\mathcal{Y}}$, respectively.[1] Suppose further that for a pair $(\alpha^*, x^*) \in \mathcal{P} \times \mathcal{X}$ we have $\mathcal{G}(\alpha^*, x^*) = 0$, and that the Fréchet derivative $D_x\mathcal{G}(\alpha^*, x^*)$ is invertible. In order to prove the implicit function theorem, one then considers the mapping

$$\mathcal{T} : \mathcal{P} \times \mathcal{X} \to \mathcal{X} \quad \text{ defined by } \quad \mathcal{T}(\alpha, x) = x - D_x\mathcal{G}(\alpha^*, x^*)^{-1}\mathcal{G}(\alpha, x) \,.$$

It can easily be seen that the fixed point equation $\mathcal{T}(\alpha, x) = x$ is equivalent to the problem (9), due to the assumed invertibility of $D_x\mathcal{G}(\alpha^*, x^*)$. Furthermore, one can compute the Fréchet derivative of $\mathcal{T}$ with respect to x at the pair (α^*, x^*) as

$$D_x\mathcal{T}(\alpha^*, x^*) \;=\; I - D_x\mathcal{G}(\alpha^*, x^*)^{-1} D_x\mathcal{G}(\alpha^*, x^*) \;=\; I - I \;=\; 0 \,.$$

This implies that in a small neighborhood of the pair (α^*, x^*) the operator norm of $D_x\mathcal{T}(\alpha, x)$ is bounded by a constant $c < 1$, and together with the mean value theorem and $\mathcal{G}(\alpha^*, x^*) = 0$ this shows that $\mathcal{T}$ is a uniform contraction in this neighborhood. An application of the uniform contraction mapping principle then finally furnishes the branch of solutions.

What happens in the above procedure if we only know $\mathcal{G}(\alpha^*, x^*) \approx 0$? If we are close enough to an actual solution of (9) at which the Fréchet derivative is invertible, then one would expect $D_x\mathcal{G}(\alpha^*, x^*)$ to be invertible as well — and this shows that the fixed point problem $\mathcal{T}(\alpha, x) = x$ is still equivalent to (9). Moreover, since $D_x\mathcal{T}(\alpha^*, x^*) = 0$ remains true, one can show that $\mathcal{T}$ satisfies a Lipschitz condition with constant $c < 1$ near (α^*, x^*). Thus, if we can find a neighborhood which is mapped into itself by $\mathcal{T}$, the assertion of the implicit function theorem should be valid as before.

These heuristic arguments can be made precise, and this has been accomplished in detail in the papers [47, 50], which in turn are slight extensions of the approach by Plum [41]. To formulate the corresponding result, assume we are given a pair $(\alpha^*, x^*) \in \mathcal{P} \times \mathcal{X}$ which satisfies the following four hypotheses.

(H1) The residual of the nonlinear operator $\mathcal{G}$ at the pair (α^*, x^*) is small, i.e., there exists a constant $\varrho > 0$ such that

$$\|\mathcal{G}(\alpha^*, x^*)\|_{\mathcal{Y}} \leq \varrho \,. \tag{10}$$

In other words, the pair (α^*, x^*) is an approximate solution of the nonlinear problem (9).

(H2) The Fréchet derivative $D_x\mathcal{G}(\alpha^*, x^*) \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$, where $\mathcal{L}(\mathcal{X}, \mathcal{Y})$ denotes the Banach space of all bounded linear operators from $\mathcal{X}$ into $\mathcal{Y}$, is one-to-one and onto, and its inverse $D_x\mathcal{G}(\alpha^*, x^*)^{-1} : \mathcal{Y} \to \mathcal{X}$ is bounded and satisfies

$$\left\| D_x\mathcal{G}(\alpha^*, x^*)^{-1} \right\|_{\mathcal{L}(\mathcal{Y}, \mathcal{X})} \leq K \,, \tag{11}$$

[1] For a first reading, the reader may certainly assume that all of these spaces equal $\mathbb{R}$, and that the norm is just the regular absolute value. In this setting, Fréchet differentiability is then normal differentiability of a function $\mathcal{G} : \mathbb{R}^2 \to \mathbb{R}$, and the derivative $D_x\mathcal{G}(\alpha^*, x^*)$ is just the partial derivative of $\mathcal{G}$ with respect to x.

where $\|\cdot\|_{\mathcal{L}(\mathcal{Y},\mathcal{X})}$ denotes the operator norm in $\mathcal{L}(\mathcal{Y},\mathcal{X})$. In other words, the operator $D_x\mathcal{G}(\alpha^*,x^*)$ is invertible and not very close to being singular.

(H3) For (α,x) close to (α^*,x^*), the Fréchet derivative $D_x\mathcal{G}(\alpha,x)$ is locally Lipschitz continuous in the following sense. There exist positive real constants L_1, L_2, ℓ_x, and $\ell_\alpha \geq 0$ such that for all pairs $(\alpha,x) \in \mathcal{P}\times\mathcal{X}$ with $\|x-x^*\|_{\mathcal{X}} \leq \ell_x$ and $\|\alpha-\alpha^*\|_{\mathcal{P}} \leq \ell_\alpha$ we have

$$\|D_x\mathcal{G}(\alpha,x) - D_x\mathcal{G}(\alpha^*,x^*)\|_{\mathcal{L}(\mathcal{X},\mathcal{Y})} \leq L_1\|x-x^*\|_{\mathcal{X}} + L_2\|\alpha-\alpha^*\|_{\mathcal{P}} \ , \tag{12}$$

where $\|\cdot\|_{\mathcal{L}(\mathcal{X},\mathcal{Y})}$ denotes the operator norm in $\mathcal{L}(\mathcal{X},\mathcal{Y})$.

(H4) For α close to α^*, the Fréchet derivative $D_\alpha\mathcal{G}(\alpha,x^*)$ satisfies a Lipschitz-type bound. More precisely, there exist positive real constants L_3 and L_4, such that for all $\alpha\in\mathcal{P}$ with $\|\alpha-\alpha^*\|_{\mathcal{P}} \leq \ell_\alpha$ one has

$$\|D_\alpha\mathcal{G}(\alpha,x^*)\|_{\mathcal{L}(\mathcal{P},\mathcal{Y})} \leq L_3 + L_4\|\alpha-\alpha^*\|_{\mathcal{P}} \ , \tag{13}$$

where ℓ_α is chosen as in (H3).

Notice that all of the constants ϱ, K, L_k, as well as both ℓ_x and ℓ_α, depend on the choice of the pair (α^*,x^*). Furthermore, in specific applications one has to be able to determine explicit values for these constants. This is usually accomplished through a combination of rigorous estimates and interval computations.

The above hypotheses specify in detail what is needed to prove the existence of a curve (in the special case $\mathcal{P}=\mathbb{R}$) of actual solutions of the nonlinear problem (9) if all that is available is an approximation. We would like to point out that while the first two hypotheses (H1) and (H2) put definite constraints on the admissible approximations (α^*,x^*), the remaining two hypotheses are usually satisfied and can be verified easily. The next result shows that under certain conditions on the constants ϱ, K, and L_1, one can always find a "branch" of solutions parameterized by the parameter α.

Theorem 3.1 (Numerical Implicit Function Theorem)**.** *Let $\mathcal{P}$, $\mathcal{X}$, and $\mathcal{Y}$ be Banach spaces, suppose that the nonlinear parameter-dependent operator $\mathcal{G}:\mathcal{P}\times\mathcal{X}\to\mathcal{Y}$ is Fréchet differentiable, and assume that $(\alpha^*,x^*)\in\mathcal{P}\times\mathcal{X}$ satisfies hypotheses (H1), (H2), (H3), and (H4). Finally, suppose that*

$$4K^2\varrho L_1 < 1 \qquad \text{and} \qquad 2K\varrho < \ell_x \ . \tag{14}$$

Then there exist pairs of constants (δ_α,δ_x) with $0\leq\delta_\alpha\leq\ell_\alpha$ and $0<\delta_x\leq\ell_x$, as well as

$$2KL_1\delta_x + 2KL_2\delta_\alpha \leq 1 \qquad \text{and} \qquad 2K\varrho + 2KL_3\delta_\alpha + 2KL_4\delta_\alpha^2 \leq \delta_x \ , \tag{15}$$

and for each such pair the following holds. For every $\alpha\in\mathcal{P}$ with $\|\alpha-\alpha^*\|_{\mathcal{P}}\leq\delta_\alpha$ there exists a uniquely determined element $x(\alpha)\in\mathcal{X}$ with $\|x(\alpha)-x^*\|_{\mathcal{X}}\leq\delta_x$ such that $\mathcal{G}(\alpha,x(\alpha))=0$. In other words, if we define

$$\mathcal{B}^{\mathcal{X}}_{\delta_x} = \{\xi\in\mathcal{X} \ : \ \|\xi-x^*\|_{\mathcal{X}}\leq\delta_x\} \qquad \text{and} \qquad \mathcal{B}^{\mathcal{P}}_{\delta_\alpha} = \{p\in\mathcal{P} \ : \ \|p-\alpha^*\|_{\mathcal{P}}\leq\delta_\alpha\} \ ,$$

then all solutions of the nonlinear problem $\mathcal{G}(\alpha,x)=0$ in the set $\mathcal{B}^{\mathcal{P}}_{\delta_\alpha}\times\mathcal{B}^{\mathcal{X}}_{\delta_x}$ lie on the graph of the function $\alpha\mapsto x(\alpha)$. In addition, for all pairs $(\alpha,x)\in\mathcal{B}^{\mathcal{P}}_{\delta_\alpha}\times\mathcal{B}^{\mathcal{X}}_{\delta_x}$ the Fréchet derivative $D_x\mathcal{G}(\alpha,x)\in\mathcal{L}(\mathcal{X},\mathcal{Y})$ is a bounded invertible linear operator, whose inverse is in $\mathcal{L}(\mathcal{Y},\mathcal{X})$.

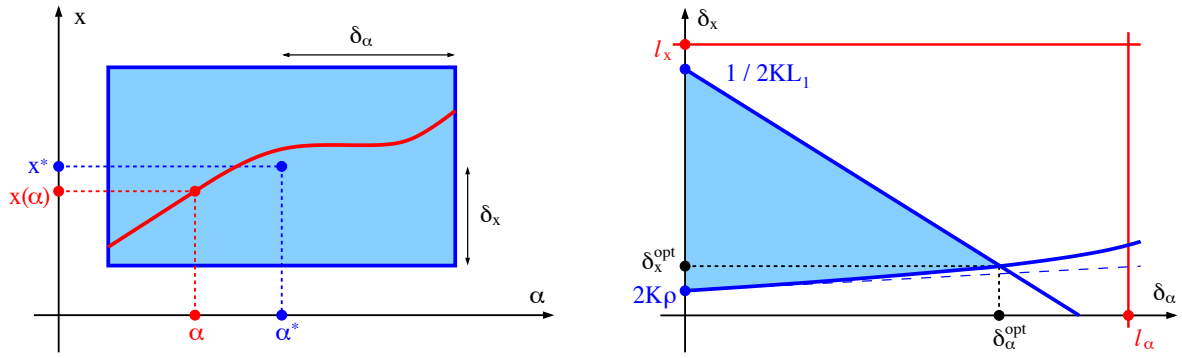

Figure 6: Visualization of the numerical implicit function theorem. The left image illustrates the general statement of the theorem, i.e., in a window around the approximative solution (α^*, x^*) all solutions of $\mathcal{G}(\alpha, x) = 0$ line on the curve $\alpha \mapsto x(\alpha)$. The size of the window is determined by the pair $(\delta_\alpha, \delta_x)$, and the shape of the admissible region from which these pairs can be chosen is illustrated in the right image.

The statement of the numerical implicit function theorem is visualized in the left image of Figure 6, and the right panel in this figure depicts the admissible region of all pairs $(\delta_\alpha, \delta_x)$ for which the result holds. For the proof of Theorem 3.1 we refer the reader to [47].

Theorem 3.1 crystallizes the main assumptions necessary for the establishment of solutions to nonlinear parameter-dependent problems in four simple hypotheses. The first of these, (H1) just states that the initial numerical approximation (α^*, x^*) has to be sufficiently accurate in the sense of having a small residual, and hypothesis (H2) ensures that the linearization of the nonlinear problem at the approximative solution is sufficiently non-singular. These are clearly the two main assumptions, and they relax the conditions of an exact solution with an invertible Fréchet derivative in the standard implicit function theorem.

Beyond (H1) and (H2), the two assumptions (H3) and (H4) are mostly technical in nature. They are always satisfied if $\mathcal{G}$ is twice continuously differentiable. Note, however, that in order to apply Theorem 3.1 we need explicit values for the constants L_1 through L_4, and they in turn depend on the choice of the constants ℓ_α and ℓ_x which determine the overall maximal "window size" for the result. Usually, these estimates can be obtained using standard mean value theorem arguments, and in many situations it is not necessary to obtain them in their sharpest form. In particular in the setting of partial differential equations this means that standard estimates involving Sobolev embeddings often suffice, and these can be formulated to hold for a wide variety of nonlinearities.

Once the constants in the above hypotheses have been established, the numerical implicit function theorem can be applied as long as (14) holds. While the second inequality in this assumption is usually satisfied and merely a reflection of a reasonable choice of ℓ_x, the first constraint is essential. Considering that the Lipschitz constant L_1 is what it is for a given nonlinearity, the estimate implies that as long as the residual ϱ is sufficiently small, we can always find true solutions near (α^*, x^*). In practice, of course, this is limited by the inherent accuracy limitations of specific implementations of computer arithmetic. We would like to

point out, however, that of the four Lipschitz constants in (H3) and (H4) only L_1 enters the essential constraint. The role of the constants L_2, L_3, and L_4 is secondary, and merely affects the size of the solution window indicated in the left panel of Figure 6 with respect to α. This will be discussed in substantial detail in the next two sections.

As a first application of Theorem 3.1 we consider the question of verifying the existence of solutions of (9) at a specific parameter value α^*, or equivalently, we consider the case of a *parameter-independent nonlinear problem*, since the parameter-dependent case will be treated in considerable detail in the next two sections. Notice that if the parameter α^* is kept fixed, we can use $\ell_\alpha = 0$ in hypotheses (H3) and (H4). In fact, one can readily see that in this case (H4) is completely unnecessary, and in (H3) one only has to find a Lipschitz constant L_1 such that

$$\|D_x\mathcal{G}(\alpha^*, x) - D_x\mathcal{G}(\alpha^*, x^*)\|_{\mathcal{L}(\mathcal{X},\mathcal{Y})} \leq L_1 \|x - x^*\|_{\mathcal{X}} \tag{16}$$

is satisfied whenever $\|x - x^*\|_{\mathcal{X}} \leq \ell_x$. Moreover, for the verification of hypothesis (H2) we need the following important auxiliary result, whose proof can be found in [47], and which is a simple consequence of the Neumann series argument in [28, Sections I.4.4 and III.3.2].

Lemma 3.2 (Linear Operator Norm Estimate)**.** *Consider two Banach spaces $\mathcal{X}$ and $\mathcal{Y}$, and let $\mathcal{A} \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ denote a bounded linear operator. In addition, let $\mathcal{B} \in \mathcal{L}(\mathcal{Y}, \mathcal{X})$ denote a bounded linear operator which is one-to-one and onto, and assume that*

$$\|I - \mathcal{B}\mathcal{A}\|_{\mathcal{L}(\mathcal{X},\mathcal{X})} \;\leq\; \varrho_1 < 1\,, \qquad \text{as well as} \qquad \|\mathcal{B}\|_{\mathcal{L}(\mathcal{Y},\mathcal{X})} \;\leq\; \varrho_2\,, \tag{17}$$

for two constants $0 \leq \varrho_1 < 1$ and $\varrho_2 > 0$. Then the linear operator $\mathcal{A}$ is one-to-one and onto as well, and its bounded inverse $\mathcal{A}^{-1}$ satisfies

$$\left\|\mathcal{A}^{-1}\right\|_{\mathcal{L}(\mathcal{Y},\mathcal{X})} \;\leq\; \frac{\varrho_2}{1 - \varrho_1}\,. \tag{18}$$

We now turn our attention to the existence of mosaic solutions in the discrete Allen-Cahn equation (5). It will be demonstrated in the following example how the parameter-independent version of the numerical implicit function theorem can be used to this effect, without any numerical assistance. More precisely, the example will show that (5) has equilibrium solutions close to any vector $\bar{u} \in \{-1, 0, 1\}^n$ for sufficiently large $\lambda > 0$. In fact, we will obtain explicit lower bounds on λ and on the size of the neighborhood in which these stationary states exist.

Example 3.3 (Mosaic Solutions for the Discrete Allen-Cahn Equation)**.** Consider the equilibrium equation associated with the discrete Allen-Cahn equation (5), rewritten in the form

$$\mathcal{G}(u) = \frac{1}{\lambda}Au + f(u) = 0 \qquad \text{with} \qquad \lambda \in \mathbb{R}^+ \quad \text{and} \quad u \in \mathbb{R}^n\,,$$

where the matrix $A \in \mathbb{R}^{n\times n}$ is given by

$$A = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 1 & -2 & 1 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & 1 & -2 & 1 \\ 0 & \cdots & 0 & 1 & -1 \end{bmatrix}. \tag{19}$$

We consider the specific nonlinearity $f(u) = u - u^3$, use the notation $u = (u_1, \dots, u_n)^t$ for the components of the vector u, and we define $f(u) = (f(u_1), \dots, f(u_n))^t$ componentwise. In addition, we adopt the abbreviation $\operatorname{diag}(u)$ for the diagonal matrix in $\mathbb{R}^{n\times n}$ with diagonal entries $u_1, \dots, u_n$. Throughout these notes, we employ the maximum norm to measure the length of vectors in $\mathbb{R}^n$, as well as the induced matrix norm, i.e., for all $v \in \mathbb{R}^n$ and $B \in \mathbb{R}^{n\times n}$ we let

$$\|v\| = \max_{i=1,\dots,n} |v_i| \qquad \text{and} \qquad \|B\| = \max_{i=1,\dots,n} \sum_{j=1}^{n} |B_{i,j}| \ , \tag{20}$$

and drop the usual subscript ∞ for the sake of brevity.

Consider now an arbitrary vector $u^* \in \{-1, 0, 1\}^n$. Our goal is to use Theorem 3.1 to find an equilibrium solution of (5) for sufficiently large $\lambda > 0$, i.e., we need to to establish hypotheses (H1) and (H2), as well as (H3) in the reduced form (16). To begin with, notice that due to the form of f and choice of u^* one has $f(u^*) = 0 \in \mathbb{R}^n$, and together with

$$\|\mathcal{G}(u^*)\| = \left\| \frac{1}{\lambda} A u^* + f(u^*) \right\| \leq \frac{\|A\| \|u^*\|}{\lambda} \leq \frac{4}{\lambda}$$

we see that (H1) holds with $\varrho = 4/\lambda$. As for (H2), one can easily compute the Jacobian matrix $D_u\mathcal{G}(u^*)$ as

$$D_u\mathcal{G}(u^*) = \frac{1}{\lambda} A + \operatorname{diag} f'(u^*) \ ,$$

where the diagonal of the matrix $\operatorname{diag} f'(u^*)$ is a vector in $\{-2, 1\}^n$, since $f'(u) = 1 - 3u^2$. If we let $\mathcal{B} \in \mathbb{R}^{n\times n}$ denote the inverse of $\operatorname{diag} f'(u^*)$, then one obtains with $\|\mathcal{B}\| \leq 1$ the estimate

$$\|I - \mathcal{B}\, D_u\mathcal{G}(u^*)\| = \left\| I - \frac{1}{\lambda}\mathcal{B}A - \mathcal{B} \operatorname{diag} f'(u^*) \right\| = \left\| \frac{1}{\lambda}\mathcal{B}A \right\| \leq \frac{\|\mathcal{B}\|\|A\|}{\lambda} \leq \frac{4}{\lambda} \ ,$$

which together with Lemma 3.2 shows that (H2) is satisfied with $K = \lambda/(\lambda - 4)$, for $\lambda > 4$. Finally, if we let $\ell_u > 0$ be arbitrary, then an application of the standard mean value theorem furnishes for all $u \in \mathbb{R}^n$ with $\|u - u^*\| \leq \ell_u$ the estimate

$$\|D_u\mathcal{G}(u) - D_u\mathcal{G}(u^*)\| \leq \max_{|\xi| \leq \|u^*\| + \ell_u} \left| f''(\xi) \right| \|u - u^*\| \leq 6\,(1 + \ell_u)\, \|u - u^*\| \ ,$$

which establishes hypothesis (H3) with $L_1 = 6(1 + \ell_u)$.

We now restrict ourselves to finding solutions in a 0.1-neighborhood of u^*, i.e., we consider the case $\ell_u = 0.1$, which implies $L_1 = 6.6$. Then the conditions of the numerical implicit function theorem in (14) are equivalent to

$$\frac{105.6\,\lambda}{(\lambda - 4)^2} < 1 \qquad \text{and} \qquad \frac{8}{\lambda - 4} < 0.1 \ , \qquad \text{which hold for all } \ \lambda > 113.459 \ .$$

Thus, Theorem 3.1 implies that for all $\lambda \geq 114$, there exists a unique equilibrium solution u_λ of (5) which satisfies

$$\|u_\lambda - u^*\| \ \leq \ 2K\varrho \ = \ \frac{8}{\lambda - 4} \ ,$$

and this solution is unique within a neighborhood of radius

$$\frac{1}{2KL_1} \;=\; \frac{\lambda-4}{13.2\,\lambda} \;\to\; \frac{1}{13.2} \;\approx\; 0.07575 \qquad \text{for} \qquad \lambda\to\infty\,.$$

While the above estimates are clearly not best possible, they do demonstrate that the numerical implicit function theorem can be applied to obtain explicit existence proofs for equilibrium solutions. We would also like to point out that the above approach can easily be extended to show that the solutions of (5) for $\lambda \geq 114$ do in fact form a branch of solutions, which is a continuous function of the paremeter λ. $\diamond$

Our second application of the parameter-independent version of Theorem 3.1 is not directly connected to the verification of bifurcation diagrams. It is, however, essential if we are interested in obtaining stability information for the involved equilibrium solutions. The following example demonstrates how eigenvalue/eigenvector pairs can be validated for matrices. Thus, the presented method can be used to determine the index of stationary states for the discrete Allen-Cahn equation.

Example 3.4 (Verification of Simple Matrix Eigenvalues)**.** Let $\mathcal{M} \in \mathbb{C}^{n\times n}$ be an arbitrary matrix, and suppose we would like to validate an eigenvalue $\eta \in \mathbb{C}$ and an associated eigenvector $v \in \mathbb{C}^n$. If one picks a fixed suitable normalization vector $\hat{v} \in \mathbb{C}^n \setminus \{0\}$, then such a pair has to solve the system

$$\mathcal{M}v - \eta v = 0 \quad \text{ and } \quad \hat{v}^t v - 1 = 0\,, \tag{21}$$

where the second equation is added to isolate the eigenvector. In fact, one can show that if η is a simple eigenvalue of $\mathcal{M}$, and if v is an eigenvector with $\hat{v}^t v = 1$, then the pair (v,η) is an isolated solution of (21). Such solutions can be validated using the numerical implicit function theorem by considering the nonlinear function $\mathcal{G} : \mathbb{C}^n \times \mathbb{C} \to \mathbb{C}^n \times \mathbb{C}$ defined by

$$\mathcal{G}(x) = \big(\mathcal{M}v - \eta v\,,\;\; \hat{v}^t v - 1\big)\;, \quad \text{ where } \quad x = (v,\eta) \in \mathbb{C}^n \times \mathbb{C}\,,$$

whose Jacobian matrix is given by

$$D_x\mathcal{G}(x) = \left(\begin{array}{c|c} \mathcal{M} - \eta I & -v \\ \hline \hat{v}^t & 0 \end{array}\right)\;.$$

Now let $x^* = (v^*,\eta^*)$ denote an approximate eigenvector/eigenvalue pair. Then one can show that the estimate $\|D_x\mathcal{G}(x) - D_x\mathcal{G}(x^*)\| \leq 2\|x - x^*\|$ holds with respect to the maximum norm in $\mathbb{C}^n$ and the induced matrix norm in $\mathbb{C}^{n\times n}$, i.e., hypothesis (H3) is satisfied with $L_1 = 2$ and $\ell_x = \infty$, as well as $\ell_\alpha = 0$. In order to satisfy hypotheses (H1) and (H2), assume that we have determined a small constant $\varrho > 0$ such that

$$\|\mathcal{M}v^* - \eta^* v^*\| \leq \varrho \qquad \text{and} \qquad \big|\hat{v}^t v^* - 1\big| \leq \varrho\,, \tag{22}$$

as well as a constant $K > 0$ with

$$\left\| \left(\begin{array}{c|c} \mathcal{M} - \eta^* I & -v^* \\ \hline \hat{v}^t & 0 \end{array}\right)^{-1} \right\| \leq K\,. \tag{23}$$

If we finally assume that the inequality $8K^2\varrho < 1$ holds, then for every $2K\varrho \leq \delta \leq 1/(4K)$ there exists a unique pair $(v, \eta) \in \mathbb{C}^n \times \mathbb{C}$ which solves the system (21) and which satisfies both $|\eta - \eta^*| \leq \delta$ and $\|v - v^*\| \leq \delta$. In other words, the maximum norm ball of radius $2K\varrho$ centered at the pair (v^*, η^*) contains a unique normalized eigenvector/eigenvalue pair for the matrix B, and this pair is unique up to distance $1/(4K)$.

In practical applications, the constant ϱ in (22) can easily be determined using interval arithmetic. As for the constant K in (23), one can still apply Lemma 3.2: If we denote the extended matrix in (23) by $\mathcal{A}$, then one can use a *numerically computed inverse* $\mathcal{B}$ of $\mathcal{A}$ as the second matrix required in the lemma. Even if $\mathcal{A}$ is an interval matrix, interval computations can be used to find rigorous bounds ϱ_1 and ϱ_2 using the numerical inverse $\mathcal{B}$ alone, and this typically leads to tight bounds K. See [47] for more details.

While the above procedure is easy to implement and to apply, two warnings are in order. First of all, note that the approach cannot be used to locate eigenvalues of multiplicity two and higher, since in such cases the pair (v, η) is never an isolated solution of (21). Secondly, while the above method proves the existence of an eigenvalue η, it does not provide an interval around η in which this eigenvalue is unique. The uniqueness assertion of the numerical implicit function theorem involves the pair (v, η), i.e., in principle there could be another eigenvalue pair $(\tilde{v}, \tilde{\eta})$ which is reasonably far away, but for which $\eta \approx \tilde{\eta}$. $\diamond$

At first glance, the above method seems to be all one would need to rigorously determine the index of an equilibrium of the discrete Allen-Cahn equation, which is given by the number of positive eigenvalues of the linearization at the stationary state. Note, however, that if the dimension n in (5) is large and if the index of the considered equilibrium small, one would have to validate a large number of eigenvalues, most of which are negative and therefore irrelevant for determining the index. Even worse, the involved negative eigenvalues could form clusters, thereby leading to eigenvalues which are no longer simple in a numerical sense, and then the above method can easily fail to validate all n eigenvalues. In [47] we have described another approach based on rigorous eigenvalue exclusion and deflation which avoids these pitfalls. We refer the reader to this paper for more details.

3.2 Verification of Branch Segments

Just as in the classical case, the numerical implicit function theorem is a general result that is meant to be applied in a variety of contexts, and in this and the next section we present two of them. As a first application, we revisit the question of validating equilibrium branches. In its original form, Theorem 3.1 is ill-suited to prove the existence of branches of stationary states, since the targeted solution curve will usually not be aligned with the direction of the parameter axis.

As was already mentioned in previous lectures, efficient methods for following branches of solutions have to take the direction of the solution curve into account. This can either be accomplished by pseudo-arclength continuation, or by selecting numerical approximations of solutions at the endpoints of the considered parameter interval and then considering the line segment determined by them. The latter method can be implemented using the radii polynomial approach, which requires, however, explicit estimates along the parametrization of the line segment of approximative solutions.

In the present section, we will take a slightly different approach, which uses Theorem 3.1 as a proof-theoretic tool rather than the main continuation mechanism. In order to avoid confusion, we therefore change our notation describing the underlying parameter-dependent problems. From now on, we consider equilibrium problems of the form

$$\mathcal{F}(\lambda, u) = 0 \,, \tag{24}$$

where $\mathcal{F} : \mathbb{R} \times \mathcal{U} \to \mathcal{V}$ is a Fréchet differentiable nonlinear operator between two Banach spaces $\mathcal{U}$ and $\mathcal{V}$. The norms on these Banach spaces are denoted by $\|\cdot\|_{\mathcal{U}}$ and $\|\cdot\|_{\mathcal{V}}$, respectively. In all of the following results, we will study (24) in the neighborhood of a pair $(\lambda^*, u^*) \in \mathbb{R} \times \mathcal{U}$ which is an approximate solution. Our main goals will be two-fold:

- First of all, we would like to specify the line segment of approximative solutions in the form $u^* + (\lambda - \lambda^*)v^*$, for λ in an interval around λ^*, where $v^* \in \mathcal{U}^*$ denotes an appropriately chosen direction. Our result should explicitly state the size of the admissible λ-interval, and it should contain a criterion which measures the suitability of the "tangent" direction v^*. In this way, we do not have to provide two numerical approximations to validate the branch segment, but we can find the longest possible branch segment from a starting point and a direction alone. This is illustrated in the left panel of Figure 7.

- Secondly, we would like to avoid having to perform complicated nonlinearity estimates along the parametrized line segment of approximative solutions. Rather, we want to be able to base the result on simple standard nonlinearity results, which allow for relatively easy changes in the function $\mathcal{F}$ without the necessity for repeated involved nonlinearity estimates.

A closer look at the assumptions of the numerical implicit function theorem reveals that the main nonlinearity assumptions involve Lipschitz estimates for the first-order Fréchet derivatives. Thus, we assume that the function $\mathcal{F}$ satisfies the following.

(A) For (λ, u) close to (λ^*, u^*), the two Fréchet derivatives $D_u\mathcal{F}(\lambda, u)$ and $D_\lambda\mathcal{F}(\lambda, u)$ are locally Lipschitz continuous in the following sense. There exist real constants $M_k \geq 0$ for $k = 1, \ldots, 4$, as well as $d_u > 0$ and $d_\lambda > 0$, such that for all pairs $(\lambda, u) \in \mathbb{R} \times \mathcal{U}$ with $\|u - u^*\|_{\mathcal{U}} \leq d_u$ and $|\lambda - \lambda^*| \leq d_\lambda$ we have

$$\|D_u\mathcal{F}(\lambda, u) - D_u\mathcal{F}(\lambda^*, u^*)\|_{\mathcal{L}(\mathcal{U},\mathcal{V})} \;\leq\; M_1 \|u - u^*\|_{\mathcal{U}} + M_2 |\lambda - \lambda^*| \,, \tag{25}$$

$$\|D_\lambda\mathcal{F}(\lambda, u) - D_\lambda\mathcal{F}(\lambda^*, u^*)\|_{\mathcal{V}} \;\leq\; M_3 \|u - u^*\|_{\mathcal{U}} + M_4 |\lambda - \lambda^*| \,, \tag{26}$$

where $\|\cdot\|_{\mathcal{L}(\mathcal{U},\mathcal{V})}$ denotes the operator norm in $\mathcal{L}(\mathcal{U}, \mathcal{V})$, and as usual we identify $\mathcal{V}$ with $\mathcal{L}(\mathbb{R}, \mathcal{V})$.

As we will see later on in the context of both the discrete Allen-Cahn equation and the diblock copolymer problem, estimates such as the ones required for (A) are usually easy to come by. Specifically in the case of partial differential equations, they can be achieved using standard Sobolev embedding techniques. Note in particular that these estimates are

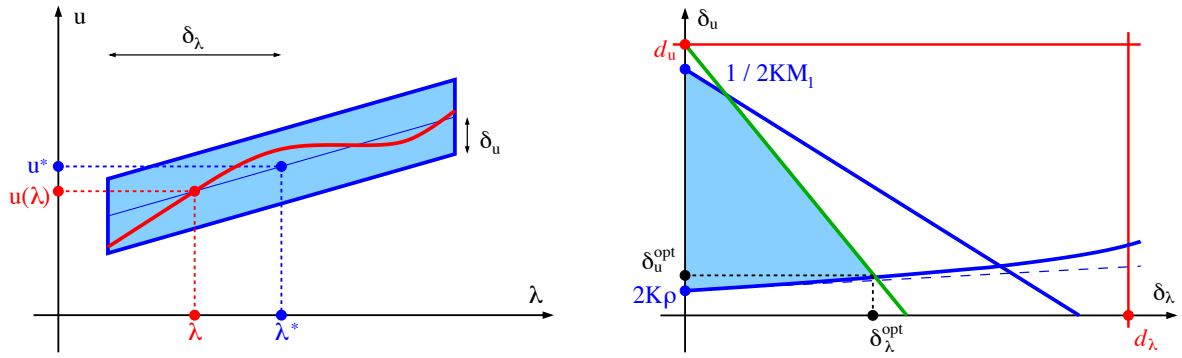

Figure 7: Visualization of the branch segment verification theorem. The left image illustrates the general statement of the theorem, i.e., in a slanted window around the approximative solution (λ^*, u^*) all solutions of $\mathcal{F}(\lambda, u) = 0$ line on the curve $\lambda \mapsto u(\lambda)$. The size of the window is determined by the pair $(\delta_\lambda, \delta_u)$, and the shape of the admissible region from which these pairs can be chosen is illustrated in the right panel. Notice that if $\|v^*\|_{\mathcal{U}} \gg 0$, then the last constraint in (29), which is indicated in green, might further restrict the admissible region. In practice, this usually does not happen.

completely independent of the particular numerical approach which is used for finding an approximative solution pair (λ^*, u^*).

Based on the first goal stated earlier, we are trying to validate solutions (λ, u) of (24) close to the approximate solution $u^* + (\lambda - \lambda^*)v^*$. It is therefore natural to introduce new variables $\alpha \in \mathbb{R}$ and $x \in \mathcal{U}$ and consider the new nonlinear problem

$$\mathcal{G}(\alpha, x) = \mathcal{F}\left(\lambda^* + \alpha,\, u^* + \alpha v^* + x\right) = 0\,,$$

where $\mathcal{G} : \mathbb{R} \times \mathcal{U} \to \mathcal{V}$. Since $\mathcal{G}(0,0) = \mathcal{F}(\lambda^*, u^*)$, the new problem has an approximate zero at $(0,0)$, as long as $\mathcal{F}$ has an approximate zero at (λ^*, u^*). Moreover, the partial Fréchet derivatives of $\mathcal{G}$ with respect to x and α are given by

$$\begin{aligned} D_x\mathcal{G}(\alpha, x) &= D_u\mathcal{F}\left(\lambda^* + \alpha, u^* + \alpha v^* + x\right)\,, \\ D_\alpha\mathcal{G}(\alpha, x) &= D_\lambda\mathcal{F}\left(\lambda^* + \alpha, u^* + \alpha v^* + x\right) + D_u\mathcal{F}\left(\lambda^* + \alpha, u^* + \alpha v^* + x\right)[v^*]\,, \end{aligned}$$

and this shows that $D_x\mathcal{G}(0,0) = D_u\mathcal{F}(\lambda^*, u^*)$. This in turn implies that if we assume the existence of constants ϱ and K with

$$\|\mathcal{F}(\lambda^*, u^*)\|_{\mathcal{V}} \le \varrho \qquad \text{and} \qquad \left\|D_u\mathcal{F}(\lambda^*, u^*)^{-1}\right\|_{\mathcal{L}(\mathcal{V},\mathcal{U})} \le K\,,$$

then $\mathcal{G}$ satisfies hypotheses (H1) and (H2) at the point $(\alpha^*, x^*) = (0,0)$, with the same constants. Thus, for an application of the numerical implicit function theorem we only have to establish hypotheses (H3) and (H4). For the former, one can easily see that the definition of $\mathcal{G}$ and (A) yield the estimate

$$\|D_x\mathcal{G}(\alpha, x) - D_x\mathcal{G}(0,0)\|_{\mathcal{L}(\mathcal{U},\mathcal{V})} \le M_1\|x\|_{\mathcal{U}} + \left(M_1\|v^*\|_{\mathcal{U}} + M_2\right)|\alpha|\,,$$

and similarly one obtains

$$\begin{aligned}\|D_\alpha \mathcal{G}(\alpha,0)\|_{\mathcal{V}} &\leq \|D_\alpha \mathcal{G}(0,0)\|_{\mathcal{V}} + \|D_\alpha \mathcal{G}(\alpha,0) - D_\alpha \mathcal{G}(0,0)\|_{\mathcal{V}} \\ &\leq \|D_\lambda \mathcal{F}(\lambda^*,u^*) + D_u \mathcal{F}(\lambda^*,u^*)[v^*]\|_{\mathcal{V}} \\ &\qquad + |\alpha| \cdot \left(M_1 \|v^*\|_{\mathcal{U}}^2 + (M_2+M_3)\|v^*\|_{\mathcal{U}} + M_4 \right) .\end{aligned}$$

In other words, all the assumptions of Theorem 3.1 are satisfied, and the constants in hypotheses (H3) and (H4) can easily be computed from the constants in (A) and the choice of the continuation direction v^*. Note, however, that since (A) contains restrictions on the deviations $\|u-u^*\|_{\mathcal{U}}$ and $|\lambda-\lambda^*|$, these will lead to an additional constraint on the window sizes δ_α and δ_x in Theorem 3.1. All of this is summarized in the following result, whose detailed proof can be found in [47].

Theorem 3.5 (Branch Segment Verification)**.** *Let $\mathcal{U}$ and $\mathcal{V}$ be Banach spaces, and suppose that the nonlinear parameter-dependent operator $\mathcal{F}: \mathbb{R}\times\mathcal{U} \to \mathcal{V}$ is both Fréchet differentiable and satisfies (A). Assume that $(\lambda^*,u^*) \in \mathbb{R}\times\mathcal{U}$ satisfies the estimates*

$$\|\mathcal{F}(\lambda^*,u^*)\|_{\mathcal{V}} \leq \varrho \qquad \text{and} \qquad \left\|D_u\mathcal{F}(\lambda^*,u^*)^{-1}\right\|_{\mathcal{L}(\mathcal{V},\mathcal{U})} \leq K \, , \tag{27}$$

for some positive constants ϱ and K, and let $v^* \in \mathcal{U}$ be given with

$$\|D_\lambda \mathcal{F}(\lambda^*,u^*) + D_u\mathcal{F}(\lambda^*,u^*)[v^*]\|_{\mathcal{V}} \leq \sigma \, , \tag{28}$$

for some constant $\sigma \geq 0$. Finally, suppose that

$$4K^2\varrho M_1 < 1 \qquad \text{and} \qquad 2K\varrho < d_u \, .$$

Then there exist pairs of constants $(\delta_\lambda,\delta_u)$ which satisfy

$$0<\delta_\lambda \leq d_\lambda \, , \qquad 0<\delta_u \leq d_u \, , \qquad \text{and} \qquad \delta_\lambda \|v^*\|_{\mathcal{U}} + \delta_u \leq d_u \, , \tag{29}$$

as well as

$$2KM_1\delta_u + 2K\left(M_1\|v^*\|_{\mathcal{U}} + M_2\right)\delta_\lambda \leq 1$$

and

$$2K\varrho + 2K\sigma\delta_\lambda + 2K\left(M_1\|v^*\|_{\mathcal{U}}^2 + (M_2+M_3)\|v^*\|_{\mathcal{U}} + M_4\right)\delta_\lambda^2 \leq \delta_u \, ,$$

and for each such pair the following holds. For every $\lambda \in \mathbb{R}$ with $|\lambda-\lambda^*| \leq \delta_\lambda$ there exists a uniquely determined element $u(\lambda) \in \mathcal{U}$ with $\|u(\lambda) - (u^* + (\lambda-\lambda^*)v^*)\|_{\mathcal{U}} \leq \delta_u$ for which the nonlinear equation $\mathcal{F}(\lambda,u(\lambda)) = 0$ is satisfied. In other words, all solutions of the nonlinear problem $\mathcal{F}(\lambda,u) = 0$ in the slanted set

$$\{(\lambda,u) \in \mathbb{R}\times\mathcal{U} \, : \, |\lambda-\lambda^*| \leq \delta_\lambda \quad \text{and} \quad \|u - (u^* + (\lambda-\lambda^*)\,v^*)\|_{\mathcal{U}} \leq \delta_u\}$$

lie on the branch $\lambda \mapsto u(\lambda)$. This statement is illustrated in the left panel of Figure 7, while the right panel depicts the admissible region of all pairs $(\delta_\lambda,\delta_u)$ for which the result holds.

Our above discussion shows that Theorem 3.5 is more or less a direct corollary of the numerical implicit function theorem. It is therefore natural to wonder why we even formulated this new result. In fact, one can easily see that if we choose $v^* = 0$, then the branch segment verification theorem reduces to Theorem 3.1 for the special case $\mathcal{P} = \mathbb{R}$. In this situation, the value of $L_3 = \sigma$ in (28) is basically given by the norm of $D_\lambda \mathcal{F}(\lambda^*, u^*)$, which in turn could be quite large. Note, however, that the value of L_3 in the numerical implicit function theorem affects the slope of the dashed blue line in the right panel of Figure 6, and large values of this slope lead to extremely small values of $\delta_\alpha^{\text{opt}}$. In other words, if L_3 is large, the parameter width of the validation window is extremely small.

Consider now the situation of Theorem 3.5, and choose the direction vector v^* close to the tangent direction of the solution branch. Then the expression on the left-hand side of (28) is extremely small, i.e., the slope of the dashed blue line in the right panel of Figure 7 will be close to zero, which leads to considerably larger values of $\delta_\lambda^{\text{opt}}$. This comes at the price of the additional constraint in (29), which is indicated in green in the right panel of Figure 7. Usually, however, this constraint does not impose any additional restriction — and even if it does, one can easily determine the new value of the optimal λ-interval. In other words, Theorem 3.5 fulfills our second goal stated at the beginning of the section. The value of σ is a means of quantifying the quality of the continuation direction: Values of σ close to zero lead to the largest window sizes in the left panel of Figure 7, larger values of σ to smaller parameter validation intervals. But in either case, the branch segment validation theorem provides explicit formulas for the size of the branch segment that can be validated.

We close this section by outlining how the branch segment verification theorem can be applied to the one-dimensional diblock copolymer model.

Example 3.6 (Branch Segments for the Diblock Copolymer Model)**.** We now return to the diblock copolymer equation (7) which was introduced in Section 2.2. To keep the presentation simple, we only consider the one-dimensional domain $\Omega = (0, 1)$. Moreover, we slightly rewrite the equation by incorporating the total mass value μ into the equation through the transformation $u \mapsto \mu + u$. In this way, the mass constraint for u can be kept fixed at zero, and (8) is equivalent to the nonlinear operator equation

$$\mathcal{F}(\lambda, u) \;=\; -\left(u_{xx} + \lambda f_\mu(u)\right)_{xx} - \lambda\sigma u \;=\; 0\,, \tag{30}$$

where due to the above transformation we usually choose $f_\mu(u) = (\mu+u)-(\mu+u)^3$. Note that even though both μ and σ are parameters of the model, we only emphasize the dependence on λ in the definition of F, as this is our main bifurcation parameter.

To apply Theorem 3.5, one first needs to define suitable Banach spaces for the operator $\mathcal{F}$. As usual in partial differential equations, we consider spaces which already incorporate the boundary conditions, and for the purposes of this lecture we consider function spaces related to the standard Sobolev spaces $H^k(0,1) = W^{k,2}(0,1)$, see for example [1]. Particularly important is the space $H^1(0,1)$ of all weakly differentiable $L^2(0,1)$-functions whose weak derivative u' is square-integrable, equipped with the standard norm

$$\|u\|_{H^1(0,1)} = \sqrt{\|u\|^2_{L^2(0,1)} + \|u'\|^2_{L^2(0,1)}}\,.$$

Due to the zero mass constraint for u, we specifically consider subspaces $\mathcal{H}^k$ of the Sobolev spaces $H^k(0,1)$, which are defined as follows. Note first that the eigenvalues and eigenfunctions of the negative Laplacian $-u_{xx}$ on the one-dimensional domain $\Omega = (0,1)$ and subject to homogeneous Neumann boundary conditions, are given by the constant function $\varphi_0(x) = 1$ with eigenvalue $\kappa_0 = 0$, together with the sequence

$$\kappa_\ell = \ell^2\pi^2 \quad \text{and} \quad \varphi_\ell(x) = \sqrt{2}\cos\ell\pi x \quad \text{for} \quad \ell \in \mathbb{N}\,.$$

Then we define

$$\mathcal{H}^k = \left\{ u = \sum_{\ell=1}^{\infty} \alpha_\ell\varphi_\ell \;:\; \|u\|_{\mathcal{H}^k} < \infty \right\}\,, \quad \text{where} \quad \|u\|_{\mathcal{H}^k} = \sqrt{\sum_{\ell=1}^{\infty} \kappa_\ell^k \alpha_\ell^2}\,.$$

In the definition of $\mathcal{H}^k$, the constant eigenfunction φ_0 is omitted due to the zero mass constraint. For $k \geq 0$ the coefficients α_ℓ are given by $\alpha_\ell = (u, \varphi_\ell)_{L^2(0,1)}$, since the eigenfunctions φ_ℓ form a complete orthonormal set in $L^2(0,1)$. Furthermore, one can easily show that the spaces $\mathcal{H}^k$ are Hilbert spaces, see [50].

With the help of these spaces, we can now reformulate the nonlinear problem (30) in the proper functional-analytic setting. For this, let $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}_0^+$ be arbitrary, but fixed, real constants. Then we consider the nonlinear operator $\mathcal{F}$ defined by

$$\begin{aligned} \mathcal{F} : \mathbb{R}\times\mathcal{U} \to \mathcal{V} \qquad &\text{with} \qquad \mathcal{F}(\lambda,u) = -\left(u_{xx} + \lambda f_\mu(u)\right)_{xx} - \lambda\sigma u \\ &\text{and} \qquad \mathcal{U} = \mathcal{H}^1\,, \quad \mathcal{V} = \mathcal{H}^{-3}\,. \end{aligned} \tag{31}$$

It follows from standard results that this operator is Fréchet differentiable, and straightforward partial differential equations estimates based on Sobolev's embedding theorem yield the following. Let $d_u > 0$ and $d_\lambda > 0$ be arbitrary constants, and let $(\lambda^*, u^*) \in \mathbb{R}\times\mathcal{U}$. Then for all $(\lambda, u) \in \mathbb{R}\times\mathcal{U}$ with $\|u - u^*\|_{\mathcal{U}} \leq d_u$ and $|\lambda - \lambda^*| \leq d_\lambda$ we have

$$\begin{aligned} \|D_u\mathcal{F}(\lambda,u) - D_u\mathcal{F}(\lambda^*,u^*)\|_{\mathcal{L}(\mathcal{U},\mathcal{V})} &\leq M_1\|u-u^*\|_{\mathcal{U}} + M_2|\lambda-\lambda^*|\,, \\ \|D_\lambda\mathcal{F}(\lambda,u) - D_\lambda\mathcal{F}(\lambda^*,u^*)\|_{\mathcal{V}} &\leq M_3\|u-u^*\|_{\mathcal{U}}\,. \end{aligned}$$

Using the two constants $C_a = \sqrt{(e^2+1)/(e^2-1)} \approx 1.14588$ and $C_c = \sqrt{1+\pi^2}/\pi \approx 1.04944$, see for example [50, Lemma 3.1], the Lipschitz constants are given by

$$\begin{aligned} M_1 &= \frac{2\sqrt{2}C_aC_c^2}{\pi^2}\left(f_{\max}^{(2)} + f_{\max}^{(3)}C_aC_cd_u + f_{\max}^{(3)}C_aC_c\|u^*\|_{\mathcal{H}^1}\right)(|\lambda^*| + d_\lambda)\,, \\ M_2 &= \frac{\sigma}{\pi^4} + \frac{2C_aC_c}{\pi^2}\left\|f_\mu'(u^*)\right\|_{H^1(0,1)}\,, \quad \text{and} \\ M_3 &= \frac{\sigma}{\pi^4} + \frac{\sqrt{2}C_c}{\pi^2}\left(f_{\max}^{(1)} + f_{\max}^{(2)}C_aC_cd_u + f_{\max}^{(2)}C_aC_c\|u^*\|_{\mathcal{H}^1}\right)\,, \end{aligned}$$

where we use the abbreviations

$$f_{\max}^{(p)} = \max\left\{\left|f_\mu^{(p)}(\xi)\right| \;:\; |\xi| \leq \|u^*\|_\infty + C_aC_cd_u\right\}\,. \tag{32}$$

Thus, the nonlinear operator $\mathcal{F}$ defined in (31) satisfies assumption (A) with constants as above, as well as with $M_4 = 0$. We would like to point out that the above estimates hold for arbitrary nonlinearities f_μ, only the estimates in (32) have to be adjusted. For simple nonlinearities such as $f_\mu(u) = (\mu + u) - (\mu + u)^3$ this formula can easily be evaluated, and if $\|u^*\|_\infty$ is not easily available one can further use the estimate $\|u^*\|_\infty \leq C_a C_c \|u^*\|_{\mathcal{H}^1}$ to extend the range for ξ in (32).

With the completion of the general nonlinearity estimates in (A), one is now in a position to apply the branch segment validation theorem. This can be accomplished as follows.

- Suppose we have constructed an approximation u^* of a solution of (30) at the parameter value λ^* in the form

$$u^* = \sum_{\ell=1}^{N} \alpha_\ell^* \varphi_\ell \quad \text{for some} \quad N \in \mathbb{N} \quad \text{and} \quad \alpha_1^*, \ldots, \alpha_N^* \in \mathbb{R} \,.$$

  In [50], for example, this was achieved by using a spectral discretization of the diblock copolymer model in combination with AUTO [18]. Then for polynomial nonlinearities both the first estimate in (27) and the one in (28) can be established easily using interval arithmetic.

- Finding a constant K such that the second estimate in (27) holds is more involved, and lies beyond the scope of this lecture. We refer the reader to [50] for more details, and only provide a very brief outline. The main idea is to use the eigenvalue exclusion methods developed in [52] to construct an interval $(-\tau, \tau)$ in which an underlying self-adjoint linear operator has no eigenvalues, and then use $K = 1/\tau$. The main difficulty in using this approach for the linearization $D_u\mathcal{F}(\lambda^*, u^*)$ is the fact that it is not self-adjoint in the space $\mathcal{H}^1$. This can, however, be addressed by showing that $(-\Delta)^{-2} D_u\mathcal{F}(\lambda^*, u^*)$ is self-adjoint in $\mathcal{H}^1$, and then considering a suitably transformed eigenvalue problem.

For a much more detailed account of the diblock copolymer case, particularly concerning the computation of the inverse operator norm bound K, we refer the reader to [50]. Notice that this paper only uses the numerical implicit function theorem in its original form, i.e., for the special case $v^* = 0$. This, however, can easily be extended using the estimates above. ◇

3.3 Resolving Saddle-Node Bifurcations

Through iterative applications, the branch segment validation theorem can be used to validate long pieces of solution branches. This is illustrated in the left image of Figure 8, where for each piece of the red solution curve of the equation $\mathcal{F}(\lambda, u) = 0$ one uses a possibly different direction vector v^* for the application of Theorem 3.5. For the upper half of the branch, the validation starts at the right-most box, and the black dot at its center indicates the first approximative solution. Subsequent applications move to the left, and in these cases the new solution approximations are located at the left sides of the respective previous boxes — still indicated by black dots. This leads to overlapping regions, where for the sake of clarity we only show the left halves of the later validation boxes. The lower half of the branch is also validated from right to left. Notice that in all of these applications of Theorem 3.5,

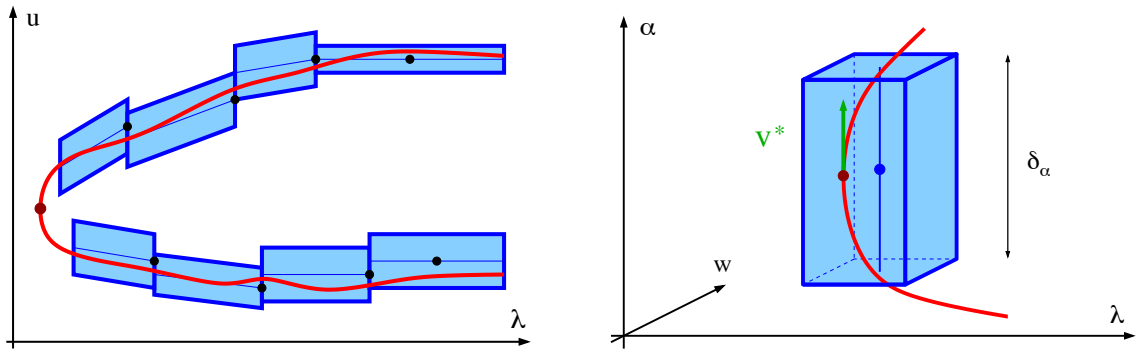

Figure 8: Visualization of the saddle-node branch verification theorem. The left panel illustrates how successive applications of Theorem 3.5 can be used to validate solution branches away from saddle-node bifurcation points. The branch close to the bifurcation point can be validated with the help of Theorem 3.7, whose setup is illustrated in the right panel. The vector v^* is an approximation for the kernel element of the Fréchet derivative $D_u\mathcal{F}$ at the saddle-node bifurcation point. The α-axis is parallel to v^*, and the w-axis represents a complementary subspace in $\mathcal{U}$.

the λ-extent of the validation neighborhood is automatically provided. Yet, the image also shows something else. If the solution branch contains a saddle-node bifurcation point, such as the one indicated by a dark red dot, Theorem 3.5 can in principle not be used to validate the complete branch — locally near the bifurcation point the solution curve is not a graph over λ. As will be demonstrated in the following, even though Theorem 3.5 does not apply near saddle-node bifurcation points, one can still use the numerical implicit function theorem to validate the solution branch in a neighborhood of such bifurcations. This time, however, Theorem 3.1 has to be applied in combination with a suitable change of variables. We would like to point out that this type of validation can also be achieved using pseudo-arclength continuation, yet at the expense of a more involved implementation.

To motivate the change of variables used in our approach, we assume for the moment that (λ_0, u_0) is a saddle-node bifurcation point for (24). Then the Fréchet derivative $D_u\mathcal{F}(\lambda_0, u_0)$ has to have a nontrivial kernel — and generically one would expect that this kernel is one-dimensional and spanned by some nonzero element $v_0 \in \mathcal{U}$. In addition, at the saddle-node bifurcation point the solution curve should be parallel to the vector v_0. Now let $\mathcal{U}^\perp \subset \mathcal{U}$ denote a subspace which is complementary to the kernel of $D_u\mathcal{F}(\lambda_0, u_0)$. Based on this discussion one can expect that if we decompose $u \in \mathcal{U}$ as

$$u = u_0 + \alpha v_0 + w \quad \text{with} \quad \alpha \in \mathbb{R} \quad \text{and} \quad w \in \mathcal{U}^\perp ,$$

then locally near the assumed saddle-node bifurcation point (λ_0, u_0) the solution curve of (24) can be parameterized by α.

In practice, the saddle-node bifurcation point is not known exactly, i.e., one only has an approximation $(\lambda^*, u^*) \approx (\lambda_0, u_0)$ which satisfies $\mathcal{F}(\lambda^*, u^*) \approx 0$. This usually implies that the Fréchet derivative $D_u\mathcal{F}(\lambda^*, u^*)$ at the approximative bifurcation point is invertible, so at first glance there does not seem to be a natural candidate for the tangent direction v^*. Yet,

one can pick any nonzero element $v^* \in \mathcal{U}$ for which $D_u\mathcal{F}(\lambda^*, u^*)[v^*] \approx 0$, and it is possible to construct the complementary space $\mathcal{U}^\perp$ in the following way. Choose a bounded linear one-to-one operator of the form $w \mapsto \mathcal{H}w$ where the range $R(\mathcal{H})$ satisfies $\mathcal{U} = \text{span}[v^*] \oplus R(\mathcal{H})$. Using these preparations, the numerical implicit function theorem can then be used to validate the branch. This leads to the following theorem.

Theorem 3.7 (Saddle-Node Branch Verification)**.** *Let $\mathcal{U}$ and $\mathcal{V}$ be Banach spaces, and suppose that the nonlinear parameter-dependent operator $\mathcal{F} : \mathbb{R} \times \mathcal{U} \to \mathcal{V}$ is both Fréchet differentiable and satisfies (A). Assume that $(\lambda^*, u^*) \in \mathbb{R} \times \mathcal{U}$ and $v^* \in \mathcal{U} \setminus \{0\}$ satisfy the estimates*

$$\|\mathcal{F}(\lambda^*, u^*)\|_{\mathcal{V}} \leq \varrho \qquad \text{and} \qquad \|D_u\mathcal{F}(\lambda^*, u^*)[v^*]\|_{\mathcal{V}} \leq \sigma \,, \tag{33}$$

for some nonnegative constants $\varrho \geq 0$ and $\sigma \geq 0$. In addition, let $\mathcal{W}$ denote another Banach space with norm $\|\cdot\|_{\mathcal{W}}$, and let $\mathcal{H} : \mathcal{W} \to \mathcal{U}$ be an injective bounded linear operator such that

$$R(\mathcal{H}) \;\; \textit{is closed} \qquad \textit{and} \qquad \mathcal{U} = \text{span}[v^*] \oplus R(\mathcal{H}) \,,$$

and assume that there exists a constant $K > 0$ such that

$$K \left\| D_\lambda\mathcal{F}(\lambda^*, u^*)\bar{\lambda} + D_u\mathcal{F}(\lambda^*, u^*)\,[\mathcal{H}\bar{w}] \right\|_{\mathcal{V}} \geq \max\left\{ \left|\bar{\lambda}\right| \,,\, \|\bar{w}\|_{\mathcal{W}} \right\} \tag{34}$$

for all $\bar{\lambda} \in \mathbb{R}$ and $\bar{w} \in \mathcal{W}$. Finally, define

$$L_1 = M_1 \left\|\mathcal{H}\right\|^2_{\mathcal{L}(\mathcal{W},\mathcal{U})} + (M_2 + M_3) \left\|\mathcal{H}\right\|_{\mathcal{L}(\mathcal{W},\mathcal{U})} + M_4$$

and suppose that

$$4K^2\varrho L_1 < 1 \qquad \textit{and} \qquad 2K\varrho < \ell_x = \min\left\{ d_\lambda \,,\, \frac{d_u}{\|\mathcal{H}\|_{\mathcal{L}(\mathcal{W},\mathcal{U})}} \right\} \,. \tag{35}$$

If in addition to the definition of ℓ_x in (35) we also define $\ell_\alpha = d_u/\|v^*\|_{\mathcal{U}}$, then there exist pairs of constants $(\delta_\alpha, \delta_x)$ which satisfy

$$0 < \delta_\alpha \leq \ell_\alpha \,, \qquad 0 < \delta_x \leq \ell_x \,, \qquad \textit{and} \qquad \delta_\alpha \left\|v^*\right\|_{\mathcal{U}} + \delta_x \left\|\mathcal{H}\right\|_{\mathcal{L}(\mathcal{W},\mathcal{U})} \leq d_u \,, \tag{36}$$

as well as

$$2KL_1\delta_x + 2K\left(M_1 \left\|\mathcal{H}\right\|_{\mathcal{L}(\mathcal{W},\mathcal{U})} + M_3 \right) \left\|v^*\right\|_{\mathcal{U}} \delta_\alpha \leq 1 \tag{37}$$

and

$$2K\varrho + 2K\sigma\delta_\alpha + 2KM_1 \left\|v^*\right\|^2_{\mathcal{U}} \delta^2_\alpha \leq \delta_x \,, \tag{38}$$

and for each such pair the following holds. For every $\alpha \in \mathbb{R}$ with $|\alpha| \leq \delta_\alpha$ there exist uniquely determined elements $\lambda(\alpha) \in \mathbb{R}$ and $w(\alpha) \in \mathcal{W}$ with

$$|\lambda(\alpha) - \lambda^*| \leq \delta_x \qquad \textit{and} \qquad \|w(\alpha)\|_{\mathcal{W}} \leq \delta_x \,,$$

for which the nonlinear equation $\mathcal{F}(\lambda(\alpha), u^* + \alpha v^* + \mathcal{H}w(\alpha)) = 0$ is satisfied. In other words, all solutions of the nonlinear problem $\mathcal{F}(\lambda, u) = 0$ in the set

$$\{(\lambda, u) \in \mathbb{R} \times \mathcal{U} \;:\; |\lambda - \lambda^*| \leq \delta_x \textit{ and } u = u^* + \alpha v^* + \mathcal{H}w \;\; \textit{with} \;\; |\alpha| \leq \delta_\alpha \,,\;\; \|w\|_{\mathcal{W}} \leq \delta_x\}$$

lie on the branch $\alpha \mapsto (\lambda(\alpha), u(\alpha))$, where we use the abbreviation $u(\alpha) = u^* + \alpha v^* + \mathcal{H}w(\alpha)$. This statement is illustrated in the right panel of Figure 8.

The detailed proof of Theorem 3.7 can be found in [47], in the following we only present a brief sketch of the basic setup. Define $\mathcal{P} = \mathbb{R}$ and $\mathcal{Y} = \mathcal{V}$, and consider the product Banach space $\mathcal{X} = \mathbb{R} \times \mathcal{W}$ with norm $\|(\lambda, w)\|_{\mathcal{X}} = \max\{|\lambda|\,,\, \|w\|_{\mathcal{W}}\}$. Moreover, define the nonlinear operator $\mathcal{G} : \mathcal{P} \times \mathcal{X} \to \mathcal{Y}$ as

$$\mathcal{G}(\alpha, x) = \mathcal{F}\left(\lambda,\, u^* + \alpha v^* + \mathcal{H}w\right)\,, \qquad \text{where} \qquad x = (\lambda, w) \in \mathcal{X} = \mathbb{R} \times \mathcal{W}\,.$$

The goal of Theorem 3.7 is to establish the existence of a branch of solutions of $\mathcal{G}(\alpha, x) = 0$ close to the point $(\alpha^*, x^*) = (0, (\lambda^*, 0))$. Since $\mathcal{G}(\alpha^*, x^*) = \mathcal{F}(\lambda^*, u^*)$, hypothesis (H1) follows from the first estimate in (33). Moreover, the new mapping $\mathcal{G}$ is Fréchet differentiable at every point (α, x), where $x = (\lambda, w) \in \mathbb{R} \times \mathcal{W}$, and both

$$D_\alpha \mathcal{G}(\alpha, x) = D_u \mathcal{F}\left(\lambda, u^* + \alpha v^* + \mathcal{H}w\right)[v^*]$$

and

$$D_x \mathcal{G}(\alpha, x)\left[\left(\bar\lambda, \bar w\right)\right] = D_\lambda \mathcal{F}\left(\lambda, u^* + \alpha v^* + \mathcal{H}w\right)\bar\lambda + D_u \mathcal{F}\left(\lambda, u^* + \alpha v^* + \mathcal{H}w\right)[\mathcal{H}\bar w]$$

are satisfied, where $\bar x = (\bar\lambda, \bar w) \in \mathcal{X} = \mathbb{R} \times \mathcal{W}$ is arbitrary. At the specific pair (α^*, x^*) these formulas simplify to

$$D_\alpha \mathcal{G}(\alpha^*, x^*) = D_u \mathcal{F}\left(\lambda^*, u^*\right)[v^*]$$

and

$$D_x \mathcal{G}(\alpha^*, x^*)\left[\left(\bar\lambda, \bar w\right)\right] = D_\lambda \mathcal{F}\left(\lambda^*, u^*\right)\bar\lambda + D_u \mathcal{F}\left(\lambda^*, u^*\right)[\mathcal{H}\bar w]\,,$$

and hypothesis (H2) follows readily from (34). The remaining two hypotheses (H3) and (H4) can easily be obtained from assumption (A), and as in the case of the branch segment validation theorem, this leads to the additional last constraint in (36), as well as a modified (37).

At first glance, the form of Theorem 3.7 might be somewhat overwhelming. However, its underlying assumptions are straightforward to verify. As before, one only needs a parameter-dependent mapping $\mathcal{F}$ satisfying (A), as well as bounds of the form

$$\|\mathcal{F}(\lambda^*, u^*)\|_{\mathcal{V}} \le \varrho \qquad \text{and} \qquad \|D_u \mathcal{F}(\lambda^*, u^*)[v^*]\|_{\mathcal{V}} \le \sigma\,,$$

where usually one ensures $\|v^*\|_{\mathcal{U}} \approx 1$ and tries to make the constants ϱ and σ as small as possible. The latter should always be possible close to a saddle-node bifurcation point. Beyond these two estimates, one also needs the invertibility condition (34), which leads to the inverse bound K. Theorem 3.7 then applies as long as the estimate

$$4K^2\varrho\left(M_1 \|\mathcal{H}\|^2_{\mathcal{L}(\mathcal{W},\mathcal{U})} + (M_2 + M_3)\|\mathcal{H}\|_{\mathcal{L}(\mathcal{W},\mathcal{U})} + M_4\right) < 1$$

holds. In other words, the above basic assumptions always guarantee the existence of a small solution branch. In practice, one would like this branch to be as large as possible, and for this one needs to be able to choose the constant δ_α as large as possible. The crucial constraint in this context is (38), which contains the small constant σ in the linear δ_α-term. As discussed in the last section, larger branch pieces can be obtained for smaller values of σ, and this in

turn can be achieved by employing accurate approximations of the kernel function at the saddle-node bifurcation point.

We close this section by outlining how Theorems 3.5 and 3.7 can be applied to the one-dimensional discrete Allen-Cahn equation. In combination with Example 3.3, this allows us to rigorously prove the existence of global solution branches which are generated through saddle-node bifurcation points.

Example 3.8 (Branch Validation for the Discrete Allen-Cahn Equation)**.** Consider the discrete Allen-Cahn equation defined in (5), but for the specific nonlinearity $f(u) = u - u^3$. In contrast to Example 3.3, we now write the equation for equilibrium solutions of (5) in the form

$$\mathcal{F}(\lambda, u) = Au + \lambda f(u) = 0 \qquad \text{with} \qquad \lambda \in \mathbb{R} \quad \text{and} \quad u \in \mathbb{R}^n \,, \tag{39}$$

where $A \in \mathbb{R}^{n\times n}$ was defined in (19), and as before, we use the notation $u = (u_1, \ldots, u_n)^t$ for the components of the vector u and define $f(u) = (f(u_1), \ldots, f(u_n))^t$ componentwise. Moreover, we consider the Banach space $\mathbb{R}^n$ equipped with the maximum norm, and on the set of matrices in $\mathbb{R}^{n\times n}$ we use the induced matrix norm, see (20).

In order to apply Theorems 3.5 and 3.7 to the discrete Allen-Cahn equilibrium equation, the fundamental first step is the establishment of assumption (A). Due to the simple form of the nonlinearity $f(u)$, this can easily be done using the standard mean value theorem. If we assume that $\lambda^* \in \mathbb{R}$ and $u^* \in \mathbb{R}^n$ are arbitrary, but fixed, and if $d_u > 0$ is a given constant, then for all $\lambda \in \mathbb{R}$ and $u \in \mathbb{R}^n$ with $\|u - u^*\| \le d_u$ one obtains the estimates

$$\begin{aligned} \|D_u\mathcal{F}(\lambda, u) - D_u\mathcal{F}(\lambda^*, u^*)\| &\le M_1 \|u - u^*\| + M_2 \,|\lambda - \lambda^*| \,, \\ \|D_\lambda\mathcal{F}(\lambda, u) - D_\lambda\mathcal{F}(\lambda^*, u^*)\| &\le M_3 \|u - u^*\| + M_4 \,|\lambda - \lambda^*| \,, \end{aligned}$$

where

$$M_1 = |\lambda^*| \max_{|\xi|\le\|u^*\|+d_u} \left|f''(\xi)\right| \,, \quad M_2 = M_3 = \max_{|\xi|\le\|u^*\|+d_u} \left|f'(\xi)\right| \,, \quad \text{and} \quad M_4 = 0 \,. \tag{40}$$

In other words, the nonlinear mapping $\mathcal{F}$ defined in (39) satisfies (A) with d_u as above and $d_\lambda = \infty$. We would like to point out that these formulas for the constants M_k hold for arbitrary smooth nonlinearities f and can easily be evaluated for specific choices. In our situation, for example, one can choose

$$M_1 = 6\,|\lambda^*|\,(\|u^*\| + d_u) \qquad \text{and} \qquad M_2 = M_3 = 1 + 3\,(\|u^*\| + d_u)^2 \tag{41}$$

for the three nonzero constants by further bounding the right-hand sides in (40) from above using $f'(u) = 1 - 3u^2$ and $f''(u) = -6u$. Despite the fact that these new constants are not optimal, they suffice for our applications and will be used in the remainder of the example.

Applying Theorems 3.5 and 3.7 to the discrete Allen-Cahn equation is now straightforward. For this, assume we have found a numerical approximation (λ^*, u^*) to a solution of $\mathcal{F}(\lambda, u) = 0$. We only need to establish the estimates in (27) and (28), or in (33) and (34), respectively. This can be done in a computational way, by taking into account all occurring roundoff errors through the use of interval arithmetic [36]. For our application, we use the MATLAB toolbox INTLAB [43] in the following way.

- **Computing** ϱ: Using the numerical approximation (λ^*, u^*) one can use interval arithmetic to find an enclosing interval for the value of $\|\mathcal{F}(\lambda^*, u^*)\|$. If we denote the right endpoint of this interval by ϱ, then the first estimates in both (27) and (33) hold. We would like to emphasize that if one starts with a good enough solution approximation, then the constant ϱ will be only slightly larger than machine precision.

- **Computing** σ: To establish (28), let v^* denote a numerical approximation of the solution v of the linear system $D_u\mathcal{F}(\lambda^*, u^*)[v] = -D_\lambda\mathcal{F}(\lambda^*, u^*)$, which in the discrete Allen-Cahn case is given by

$$Av + \lambda^* \operatorname{diag}(f'(u^*))v = -f(u^*) \; .$$

  Proceeding as in the first point, one can then compute a constant σ such that (28) is satisfied. The same method can be used to establish the second estimate in (33), if the vector v^* is chosen as an approximation to the kernel function of the linearization at the saddle-node bifurcation point. This can be done by letting v^* be a normalized numerical approximation to an eigenvector of $D_u\mathcal{F}(\lambda^*, u^*)$ which corresponds to the eigenvalue closest to zero.

- **Computing** K: For the second estimate in (27), we need to determine a rigorous upper bound on the norm of the inverse of $D_u\mathcal{F}(\lambda^*, u^*) = A + \lambda^* \operatorname{diag}(f'(u^*))$. Passing to interval enclosures for both λ^* and u^*, INTLAB can be used to compute an interval matrix which contains the true linearization $D_u\mathcal{F}(\lambda^*, u^*)$. In other words, if we let Ξ denote the set of all matrices whose entries are contained in the respective interval entries of this interval matrix, then $D_u\mathcal{F}(\lambda^*, u^*) \in \Xi$. Following [44], let $B \in \mathbb{R}^{n\times n}$ be the numerically computed inverse of an arbitrary matrix in Ξ, for example the midpoint of Ξ. Then interval computations provide rigorous bounds $0 < \varrho_1 < 1$ and $\varrho_2 > 0$ with

$$\|I - BC\| \leq \varrho_1 \quad \text{for all} \quad C \in \Xi \; , \qquad \text{as well as} \qquad \|B\| \leq \varrho_2 \; ,$$

  and Lemma 3.2 implies that every $C \in \Xi$ is invertible with $\|C^{-1}\| \leq \varrho_2/(1-\varrho_1)$. In other words, if we choose K as the right endpoint of the interval enclosure of $\varrho_2/(1-\varrho_1)$, then the second estimate in (27) holds.

  Assumption (34) can be treated similarly, but we first have to choose the operator $\mathcal{H}$. For this, define $\mathcal{W} = \mathbb{R}^{n-1}$ and let $k \in \{1, \ldots, n\}$ denote the index for which

$$|v_k^*| \; = \; \|v^*\| \; = \; \max_{\ell=1,\ldots,n} |v_\ell^*| \; > \; 0 \; .$$

  Now let $\mathcal{H} \in \mathbb{R}^{n\times(n-1)}$ denote the matrix which is obtained from the identity matrix in $\mathbb{R}^{n\times n}$ by removing the k-th column. Then the range of $\mathcal{H}$ is just the orthogonal complement of the k-th standard unit vector in $\mathbb{R}^n$, and due to $v_k^* \neq 0$ this immediately implies $\mathbb{R}^n = \operatorname{span}[v^*] \oplus R(\mathcal{H})$. Moreover, the induced matrix norm of $\mathcal{H}$ is exactly 1 and $\mathcal{H}$ has a trivial kernel. Then as above we can find a constant K such that

$$\left\| \left(\; D_\lambda\mathcal{F}(\lambda^*, u^*) \; \middle| \; D_u\mathcal{F}(\lambda^*, u^*)\mathcal{H} \; \right)^{-1} \right\| \leq K \; ,$$

  and this yields the inequality in (34).

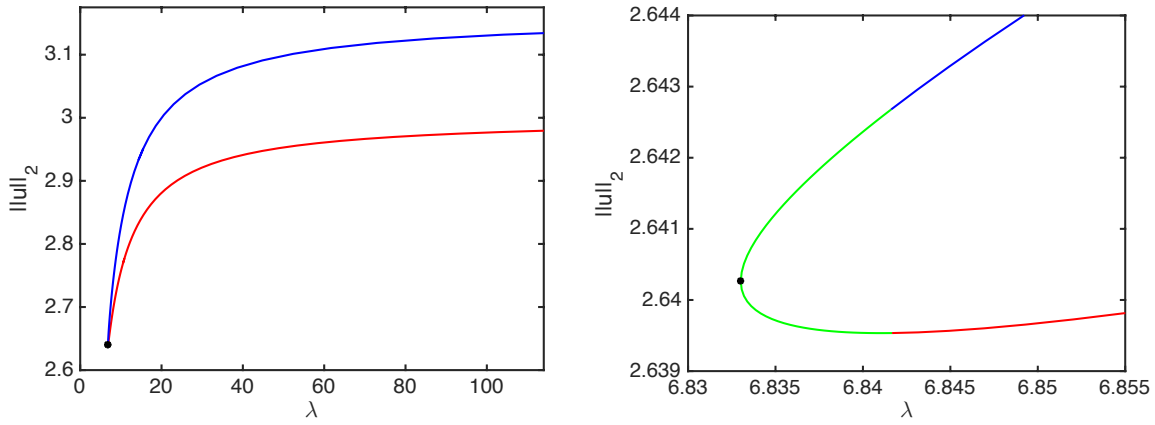

Figure 9: Rigorously validated branch of equilibrium solutions for the discrete Allen-Cahn equation (5) with $n = 10$. While the left image shows the complete validated branch, the right image depicts the neighborhood of the saddle-node bifurcation point. The numerical approximation (λ^*, u^*) of the saddle-node bifurcation point is indicated by a black dot in both images. The green part of the branch was validated using Theorem 3.7, the red and blue parts were validated by Theorem 3.5.

As a specific application, consider the case $n = 10$ in (39) and the mosaic solution pattern prescribed by the vector

$$\bar{u}_m = (1, -1, 1, -1, 1, -1, -1, -1, -1, -1)^t \in \mathbb{R}^{10}. \tag{42}$$

It was shown in Example 3.3 that for all $\lambda \geq 114$ there exist equilibrium solutions of (5) close to $\bar{u}_m$, and in fact, these solutions lie on a smooth branch. How does this branch continue for smaller λ-values? Regular numerical path-following computations using AUTO [18] indicate the existence of a saddle-node bifurcation point close to $\lambda = 6.8330142962$, and these computations readily provide a numerical approximation $(\lambda^*, u^*) \in \mathbb{R} \times \mathbb{R}^{10}$ of the bifurcation point. At this pair, the Jacobian matrix $D_u\mathcal{F}(\lambda^*, u^*)$ has a simple eigenvalue close to zero, and we let v^* denote the approximation of an associated eigenvector with $\|v^*\| = 1$. We then applied Theorem 3.7 iteratively 25 times in the direction v^*, and 25 times in the direction of $-v^*$. In all cases, the theorem succeeded to validate branch segments, whose union is shown in green in the right image of Figure 9. Starting at each of the endpoints, we then used Theorem 3.5 iteratively to follow the two halves of the branch until the parameter λ exceeded $\lambda = 114$. Also in this case, validation succeeded in every iteration, and the resulting validated branches are shown in red and blue in the two images of Figure 9. For the validation, we needed 83 and 64 iterations, respectively. In combination with Example 3.3, these computations prove the existence of a saddle-node bifurcation point which leads to a branch of equilibrium solutions connecting the mosaic state near $\bar{u}_m$ to the mosaic equilibrium near $\bar{u} = (1, -1, 1, -1, 0, -1, -1, -1, -1, -1)^t$. In fact, the method can be used to validate almost all of the 2^n branches which involve mosaic solutions $\bar{u} \in \{\pm 1\}^n$. This is discussed further in [47], where in addition the stability of all involved equilibrium solutions is determined using rigorous computations. $\diamond$

4 Direct Localization of Bifurcation Points

The results of the last section can be used to obtain computer-assisted proofs for the existence of solution branches of parameter-dependent nonlinear problems of the form $\mathcal{F}(\lambda, u) = 0$. While Theorem 3.5 can deal with branch segments which are functions of the parameter λ, Theorem 3.7 can treat branches in the neighborhood of a saddle-node bifurcation point. Combined, both results allow for the verification of a wide variety of bifurcation diagrams, namely those which only exhibit saddle-node bifurcation points. As we mentioned earlier, these are in a generic sense "almost all" bifurcation diagrams, which can be seen by means of an application of Sard's theorem [12, Section 2.10]. Typically, small perturbations of bifurcation diagrams such as the one shown in the left panel of Figure 1 lead to diagrams as in Figure 2, where for example all pitchfork bifurcations have been broken up into separate branch pieces involving only saddle-node bifurcations.

Based on the above discussion, one might be tempted to think that Theorems 3.5 and 3.7 suffice in the majority of all cases. This, however, is not the case. On the one hand, notice that while the two theorems can be used to detect saddle-node bifurcation points *indirectly*, one usually does not get a rigorous statement as to *where exactly* the saddle-node bifurcation point occurs. This can be seen in the right image of Figure 9. The fact that the λ-values at the endpoints of the green segment are strictly larger than the ones for other points on this segment implies that there has to be a saddle-node point, but Theorem 3.7 provides no insight as to where it could be. It would therefore be nice to have a rigorous computational technique available which can locate saddle-node bifurcation points *directly*.

To complicate things further, there are quite a number of situations where one does in fact have to worry about bifurcation points other than the simple saddle-node case. As we will see later on, certain structures in a nonlinear problem can make pitchfork bifurcations robust against classes of perturbations, and one therefore has to be able to detect such bifurcations directly as well. In the following two sections, it will be demonstrated how this direct verification of bifurcation points can be achieved in a computer-assisted framework.

4.1 Saddle-Node Bifurcation Points

It was mentioned in the introduction that a pair (λ_0, u_0) cannot be a bifurcation point if it satisfies the assumptions of the implicit function theorem, as stated in Theorem 1.2. As a consequence, we obtain Corollary 1.3 which provides a *necessary* condition for bifurcation. While in the introduction these two results were only formulated in the finite-dimensional setting, they remain valid in a general Banach space framework, see for example [12]. Unfortunately, however, the necessary noninvertibility condition in Corollary 1.3 is in general not sufficient for bifurcation. In this, the necessary condition is comparable to the concept of a critical point in calculus, which can be used to locate potential extreme values of a function. Being a critical point is only a necessary requirement, and one needs additional conditions to actually ensure that a critical point is a local maximum or minimum. We would like to point out that in the case of extreme values, the necessary condition is formulated using the first derivative of the considered function, while the sufficient condition involves the second derivative as well. This will be similar in the case of saddle-node bifurcation points.

Before being able to use computer-assisted methods to directly locate saddle-node bifurcations, we first need to understand how they can be guaranteed, i.e., we need to recall the sufficient conditions mentioned above. For this, we return to a general nonlinear problem of the form

$$\mathcal{F}(\lambda, u) = 0 \,, \tag{43}$$

where $\mathcal{F} : \mathbb{R} \times \mathcal{U} \to \mathcal{V}$ is a smooth nonlinear mapping, and $\mathcal{U}$ and $\mathcal{V}$ are real Banach spaces. Throughout this section, we assume the following.

Assumption 4.1 (Fredholm Assumption for the Linearization)**.** *Let $\mathcal{U}$ and $\mathcal{V}$ denote real Banach spaces, and assume that the parameter-dependent nonlinear operator $\mathcal{F} : \mathbb{R} \times \mathcal{U} \to \mathcal{V}$ is sufficiently smooth. Suppose that the pair $(\lambda_0, u_0) \in \mathbb{R} \times \mathcal{U}$ is a solution of the nonlinear problem (43), i.e., suppose that the identity $\mathcal{F}(\lambda_0, u_0) = 0$ holds. Finally, assume that the Fréchet derivative $L = D_u\mathcal{F}(\lambda_0, u_0) \in \mathcal{L}(\mathcal{U}, \mathcal{V})$ is a Fredholm operator of index zero.*

Recall that a bounded linear operator L is a Fredholm operator if it has a finite-dimensional nullspace $N(L)$, if its range $R(L)$ has finite codimension, and if the range is closed. In this case, its index is the difference between the dimension of the nullspace and the codimension of the range. In fact, it was shown in [19, 42] that the closedness assumption on the range is not necessary, as it automatically follows from the remaining assumptions. Moreover, we would like to point out that if $\mathcal{U}$ and $\mathcal{V}$ are finite-dimensional and of the same dimension, then Assumption 4.1 is always true.

Assumption 4.1 describes the basic framework that is convenient for the study of bifurcation problems. In addition, we will only consider possible bifurcation points with a one-dimensional nullspace. This leads to the following second assumption.

Assumption 4.2 (One-Dimensional Nullspace)**.** *Suppose that Assumption 4.1 holds. Assume further that the linearization $L = D_u\mathcal{F}(\lambda_0, u_0)$ has a one-dimensional kernel. Since L has index zero, its range has then codimension one. In this case, we have*

$$N(L) = \operatorname{span}[\varphi_0] \qquad \text{and} \qquad R(L) = N\left(\psi_0^*\right)$$

for some nonzero elements $\varphi_0 \in \mathcal{U}$ and $\psi_0^* \in \mathcal{V}^*$, where $\mathcal{V}^*$ denotes the dual space of $\mathcal{V}$. Finally, let $\tilde{\mathcal{U}} \subset \mathcal{U}$ and $\tilde{\mathcal{V}} \subset \mathcal{V}$ denote closed subspaces such that

$$\mathcal{U} = N(L) \oplus \tilde{\mathcal{U}} \qquad \text{and} \qquad \mathcal{V} = \tilde{\mathcal{V}} \oplus R(L) \,.$$

The continuous linear projector $P : \mathcal{V} \to \mathcal{V}$ is defined via $R(P) = \tilde{\mathcal{V}}$ and $N(P) = R(L)$, while the projector $Q : \mathcal{U} \to \mathcal{U}$ is defined via $R(Q) = N(L)$ and $N(Q) = \tilde{\mathcal{U}}$. Notice that both P and Q have rank one.

In the above assumption, the introduction of an eigenfunction φ_0 which spans the one-dimensional nullspace is natural, and φ_0 will clearly play a prominent role in the sufficient condition for a saddle-node bifurcation point. Note, however, that it is also necessary to have a characterization of the range of L at hand, since in our situation the linearization will no longer be onto at the bifurcation point. The most convenient way to do this is via the element ψ_0^* in Assumption 4.2. In the finite-dimensional case $\mathcal{V} = \mathbb{R}^d$, the dual space $\mathcal{V}^*$ is

isomorphic to $\mathcal{V}$, i.e., the element ψ_0^* is just a vector in $\mathbb{R}^d$. Moreover, the condition $v \in R(L)$ is then equivalent to the dot product of ψ_0^* and v being zero. In fact, in this case, one can choose any vector which does not lie in $R(L)$ as ψ_0^*.

After these preparations, we now turn our attention to the first and simplest type of bifurcation — the saddle-node bifurcation. As mentioned earlier, this is the only bifurcation point which can be observed in generic systems. For this, assume that we have a solution (λ_0, u_0) of the problem (43), and that Assumption 4.2 holds. Then the condition

$$\psi_0^* D_\lambda \mathcal{F}(\lambda_0, u_0) \neq 0 \tag{44}$$

is generically satisfied, and in this case one often refers to (λ_0, u_0) as a *simple saddle-node bifurcation point*.

In order to understand the implications of condition (44), we briefly return to Theorem 3.7. Since at the moment we assume to have a precise solution (λ_0, u_0) of problem (43) at hand, and since we defined φ_0 as an element in the kernel of $D_u\mathcal{F}(\lambda_0, u_0)$, the estimates in (33) are satisfied with $\varrho = 0$ and $\sigma = 0$. If, in addition, we choose $\mathcal{H} = I - Q$ and $\mathcal{W} = \tilde{\mathcal{U}}$, with the projection Q and the subspace $\tilde{\mathcal{U}}$ from Assumption 4.2, then $\mathcal{U} = \operatorname{span}[\varphi_0] \oplus R(\mathcal{H})$ is satisfied as well. Moreover, condition (44) implies that the vector $D_\lambda\mathcal{F}(\lambda_0, u_0)$ is not contained in the range of $D_u\mathcal{F}(\lambda_0, u_0)$, which in combination with the Fredholm property yields that the linear operator

$$\mathbb{R} \times \tilde{\mathcal{U}} \ \ni \ (\bar{\lambda}, \bar{w}) \quad \mapsto \quad D_\lambda\mathcal{F}(\lambda_0, u_0)\bar{\lambda} + D_u\mathcal{F}(\lambda_0, u_0)\,[\mathcal{H}\bar{w}] \ \in \ \mathcal{V}$$

is one-to-one and onto, i.e., there exists a constant $K > 0$ such that (34) is satisfied. Finally, notice that due to $\varrho = 0$ both inequalities in (35) are trivially valid, and therefore we can apply the saddle-node branch verification theorem. This shows that there exists a small branch of solutions of (43) which lies in the direction of the line $\alpha \mapsto u_0 + \alpha\varphi_0$ and which is uniquely determined in a neighborhood of (λ_0, u_0).

Drawn in a bifurcation diagram with axes λ and u, the just-established solution branch emanates vertically from the point (λ_0, u_0). Note, however, that this does not automatically imply that the point (λ_0, u_0) is a bifurcation point, as the simple one-dimensional example $\mathcal{F}(\lambda, u) = \lambda - u^3$ shows. For this, we need an additional condition involving the second derivative of $\mathcal{F}$, which is contained in the following *sufficient condition* for a saddle-node bifurcation point, whose proof can be found in [31].

Proposition 4.3 (Sufficient Condition for Saddle-Node Bifurcation)**.** *Suppose that Assumptions 4.1 and 4.2 are satisfied, and that (44) holds. If in addition the generic condition*

$$\psi_0^* D_{uu}\mathcal{F}(\lambda_0, u_0)[\varphi_0, \varphi_0] \neq 0 \tag{45}$$

is true, then the nonlinear problem (43) undergoes a saddle-node bifurcation at (λ_0, u_0). Furthermore, if we have

$$\frac{\psi_0^* D_{uu}\mathcal{F}(\lambda_0, u_0)[\varphi_0, \varphi_0]}{\psi_0^* D_\lambda\mathcal{F}(\lambda_0, u_0)} \ > \ 0 \ , \tag{46}$$

then the bifurcating solutions exist for $\lambda < \lambda_0$ close to the bifurcation point, if the ratio is negative then they exist for $\lambda > \lambda_0$. These two cases are usually referred to as a subcritical *or a* supercritical saddle-node bifurcation*, respectively.*

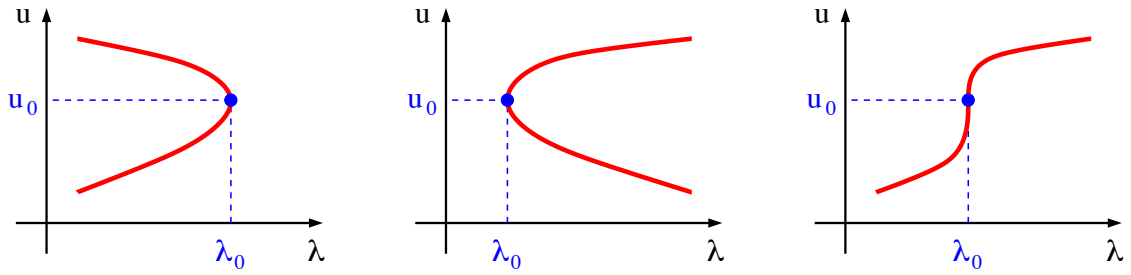

Figure 10: Visualization of the sufficient condition for a saddle-node bifurcation given in Proposition 4.3. The three panels show possible shapes of the solution branch of (43) near the solution (λ_0, u_0) as long as (44) is satisfied. While the first two panels are for the cases of strictly positive and negative values, respectively, of the fraction in (46), the third panel shows one of several cases which are possible if $\psi_0^* D_{uu}\mathcal{F}(\lambda_0, u_0)[\varphi_0, \varphi_0] = 0$. In other words, in this inconclusive case (λ_0, u_0) might or might not be a bifurcation point.

The statement of the proposition is illustrated in Figure 10. The left two panels show sub- and supercritical saddle-node bifurcation points, respectively, which are guaranteed as long as (45) is satisfied. If the latter condition is violated, the proposition still guarantees a solution branch with vertical tangent at (λ_0, u_0), but its precise shape can not be inferred from second derivative information alone. One possible geometry is shown in the right-most panel of the figure, in which (λ_0, u_0) is not a bifurcation point at all.

While the above result provides generic conditions that guarantee a saddle-node bifurcation, They are not yet amenable to a rigorous computational approach which involves the theory developed so far. For this, we need to reformulate Proposition 4.3 as a nonlinear system in which isolated zeros corresponds to saddle-node bifurcation points. For this, we follow the approach in [35] and consider a suitable extended system. More precisely, we supplement the nonlinear parameter-dependent equation (43) by a second one which forces the existence of a kernel function v of the appropriate Fréchet derivative, together with a normalizing condition on v. This leads to the extended system

$$\begin{aligned} \mathcal{F}(\lambda, u) &= 0 \,, \\ D_u\mathcal{F}(\lambda, u)[v] &= 0 \,, \\ \ell(v) - 1 &= 0 \,, \end{aligned} \tag{47}$$

where $\ell \in \mathcal{U}^*$ is a fixed element of the dual space of $\mathcal{U}$. We abbreviate this system as

$$\mathcal{G}(\lambda, u, v) = (0, 0, 0) \,,$$

where

$$\mathcal{G} \;:\; \begin{cases} \mathbb{R} \times \mathcal{U} \times \mathcal{U} & \to & \mathbb{R} \times \mathcal{V} \times \mathcal{V} \\ (\lambda, u, v) & \mapsto & (\ell(v) - 1,\ \mathcal{F}(\lambda, u),\ D_u\mathcal{F}(\lambda, u)[v]) \end{cases} \tag{48}$$

Then the following result is analogous to [35, 48]. Since in these papers only operators acting on one Banach space are considered, we present the straightforward extension to mappings between different Banach spaces. The proof of the result can be found in [31].

Theorem 4.4 (Saddle-Node Bifurcations via Extended Systems)**.** *Suppose that Assumption 4.1 is satisfied. Then the following two statements hold.*

(a) If the nonlinear operator $\mathcal{F}$ satisfies Assumption 4.2, as well as both conditions (44) and (45), and if $\ell \in \mathcal{U}^*$ is any functional such that $\ell(\varphi_0) = 1$, then the Fréchet derivative $D_{(\lambda,u,v)}\mathcal{G}(\lambda_0, u_0, \varphi_0)$ is invertible, i.e., the solution $(\lambda_0, u_0, \varphi_0)$ of the extended system (47) is an isolated non-degenerate zero of the mapping $\mathcal{G}$.

(b) Conversely, if there exists an $\ell \in \mathcal{U}^*$ and a $\varphi_0 \in \mathcal{U}$ such that $\mathcal{G}(\lambda_0, u_0, \varphi_0) = (0, 0, 0)$, and if the Fréchet derivative $D_{(\lambda,u,v)}\mathcal{G}(\lambda_0, u_0, \varphi_0)$ is invertible, then the nonlinear operator $\mathcal{F}$ satisfies Assumption 4.2, as well as both conditions (44) and (45).

In other words, the nonlinear problem (43) undergoes a saddle-node bifurcation at (λ_0, u_0) in the sense of Lemma 4.3, if and only if the triple $(\lambda_0, u_0, \varphi_0)$ is a non-degenerate zero of the nonlinear map $\mathcal{G}$ which defines the extended system (47).

With Theorem 4.4 we have placed the verification of saddle-node bifurcation points firmly into the framework outlined in Section 3. By applying the parameter-independent version of the numerical implicit function theorem to the extended system (47), one can easily establish the existence of a saddle-node bifurcation point. Moreover, if the nonlinear problem under consideration depends on a second parameter, one can even establish branches of saddle-node bifurcation points using Theorem 3.5, or follow such branches around folds using Theorem 3.7. We close this section with a brief application to the discrete Allen-Cahn equation.

Example 4.5 (Saddle-Node Bifurcations in the Discrete Allen-Cahn Equation)**.** We again consider the discrete Allen-Cahn equation in the form (39), i.e., the nonlinear function $\mathcal{F}$ is given by $\mathcal{F}(\lambda, u) = Au + \lambda f(u)$ for $\lambda \in \mathbb{R}$ and $u \in \mathbb{R}^n$, and we let $f(u) = u - u^3$. Our goal is to apply the parameter-independent version of Theorem 3.1 to the associated extended system (47), where throughout the example, we use the maximum norm for vectors, and the induced matrix norm, see (20).

The numerical implicit function theorem can easily be applied. For this, suppose we have found a numerical approximation $(\lambda^*, u^*) \in \mathbb{R} \times \mathbb{R}^n$ to a saddle-node bifurcation point, and assume that $v^* \in \mathbb{R}^n \setminus \{0\}$ is an approximation of a kernel function at the bifurcation point. Let $\varrho > 0$ be a constant with

$$\|\mathcal{F}(\lambda^*, u^*)\| \leq \varrho \,, \qquad \|D_u\mathcal{F}(\lambda^*, u^*)[v^*]\| \leq \varrho \,, \qquad \text{and} \qquad \left|\hat{v}^t v^* - 1\right| \leq \varrho \,,$$

where the vector $\hat{v} \in \mathbb{R}^n$ is a fixed normalization vector. Furthermore, let $d_u > 0$, introduce the abbreviation

$$f^{(p)}_{\max} = \max\left\{\left|f^{(p)}(\xi)\right| \,:\, |\xi| \leq \|u^*\| + d_u\right\} \,, \tag{49}$$

and define

$$L = 2f^{(1)}_{\max} + f^{(2)}_{\max}\left(2\,|\lambda^*| + 2\,\|v^*\| + d_u\right) + f^{(3)}_{\max}\,|\lambda^*|\,\|v^*\| \,. \tag{50}$$

Finally, let $K > 0$ be such that

$$\left\| \left(\begin{array}{c|c|c} A + \lambda^* \operatorname{diag}(f'(u^*)) & 0 & f(u^*) \\ \hline \lambda^* \operatorname{diag}(f''(u^*))\operatorname{diag}(v^*) & A + \lambda^* \operatorname{diag}(f'(u^*)) & \operatorname{diag}(f'(u^*))v^* \\ \hline 0 & 2(v^*)^t & 0 \end{array}\right)^{-1} \right\| \leq K \,,$$

and suppose that

$$4K^2\varrho L < 1 \qquad \text{and} \qquad 2K\varrho < d_u \ .$$

Then for every constant δ which satisfies $2K\varrho \le \delta \le \min\{1/(2KL),\, d_u\}$ there exists a unique triple $(\lambda_0, u_0, v_0) \in \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n$ which solves the extended system (47) and which satisfies the estimates

$$|\lambda_0 - \lambda^*| \le \delta \ , \qquad \|u_0 - u^*\| \le \delta \ , \qquad \text{and} \qquad \|v_0 - v^*\| \le \delta \ .$$

In other words, the maximum norm ball of radius $2K\varrho$ centered at the pair (λ^*, u^*) contains a unique saddle-node bifurcation point for $\mathcal{F}(\lambda, u) = 0$.

The statements of the last paragraph follow immediately from the parameter-independent version of the numerical implicit function theorem. One only has to apply the result to the nonlinear mapping $\mathcal{G} : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}$ defined by

$$\mathcal{G}(x) = \big(\mathcal{F}\,(\lambda, u) \ , \ \ D_u\mathcal{F}\,(\lambda, u)\,[v]\,, \ \ \hat{v}^t v - 1\big) \ , \qquad \text{where} \qquad x = (u, v, \lambda) \ .$$

Then one can easily see that its Jacobian matrix is given by

$$D_x\mathcal{G}(x) = \left(\begin{array}{c|c|c} D_u\mathcal{F}\,(\lambda, u) & 0 & D_\lambda\mathcal{F}\,(\lambda, u) \\ \hline D_{uu}\mathcal{F}\,(\lambda, u)\,[v, \cdot] & D_u\mathcal{F}\,(\lambda, u) & D_{\lambda u}\mathcal{F}\,(\lambda, u)\,[v] \\ \hline 0 & \hat{v}^t & 0 \end{array}\right) \ ,$$

which for the discrete Allen-Cahn equation takes the concrete form

$$D_x\mathcal{G}(x) = \left(\begin{array}{c|c|c} A + \lambda\,\mathrm{diag}(f'(u)) & 0 & f(u) \\ \hline \lambda\,\mathrm{diag}(f''(u))\,\mathrm{diag}(v) & A + \lambda\,\mathrm{diag}(f'(u)) & \mathrm{diag}(f'(u))v \\ \hline 0 & \hat{v}^t & 0 \end{array}\right) \ .$$

Standard mean value theorem arguments imply that for all $x \in \mathbb{R}^{2n+1}$ with $\|x - x^*\| \le d_u$ we have

$$\|D_x\mathcal{G}(x) - D_x\mathcal{G}(x^*)\| \ \le \ L\,\|x - x^*\| \ ,$$

with L as defined in (50). In other words, hypothesis (H3) holds with $\ell_\alpha = 0$ and $\ell_x = d_u$, and with Lipschitz constants $L_1 = L$ and $L_2 = 0$. Since we are only interested in the parameter-independent version of Theorem 3.1, hypothesis (H4) does not have to be verified, and the above result follows. For more details, we refer the reader to [47].

Using this result, one can easily establish the existence of the saddle-node bifurcation points which were described in Section 2.1. As a numerical example, we return to the mosaic solution close to the sign vector $\bar{u}_m$ defined in (42). As we saw in Example 3.8, the associated solution branch is created via a saddle-node bifurcation close to $\lambda^* = 6.8330142962$, and a numerical approximation to the vector u^* can be found in [47], see also the left panel in Figure 11. With these approximations, which were produced by AUTO using error tolerances of the order 10^{-8}, the above result can be used to rigorously establish the existence of a saddle-node bifurcation point within a maximum norm ball of radius $\delta_{\min} = 2.95 \cdot 10^{-7}$, and this bifurcation point is unique within a ball of radius $\delta_{\max} = 2.84 \cdot 10^{-4}$. Tighter bounds can be achieved if one first refines the solution approximation through a few Newton iterations

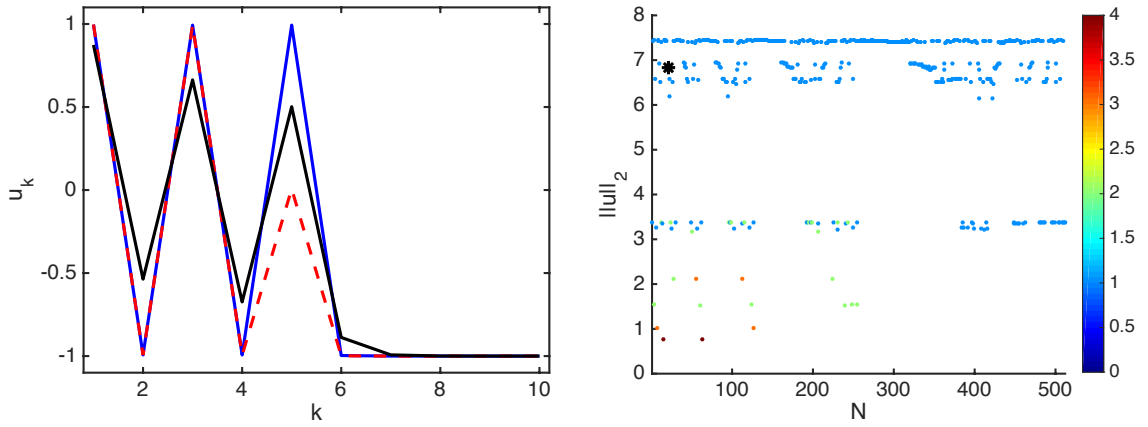

Figure 11: Verification of saddle-node bifurcation points in the discrete Allen-Cahn model (5) for $n = 10$. The left image shows the numerical approximation u^* of a saddle-node point in black, while the blue and red curves depict the equilibrium solutions on the generated solution branches at $\lambda = 300$. The location of this saddle-node bifurcation point is indicated in the right panel as a black star. The horizontal axis in this image depends on the integer N, which encodes the ± 1-pattern of the resulting mosaic solution via the binary representation of N. The bifurcation point in the left panel corresponds to $N = 21$.

of the extended system (47). This leads to an improved bifurcation point approximation less than 10^{-9} away from the pair computed with AUTO, and another application of Theorem 3.1 now furnishes $\delta_{\min} = 3.68 \cdot 10^{-14}$ and $\delta_{\max} = 2.84 \cdot 10^{-4}$. This change is mostly due to the resulting smaller value of the residual ϱ, which drops from about 10^{-8} to 10^{-15}, while the constants $K \approx 11.8$ and $L \approx 148.6$ remain basically unchanged.

This example computation is illustrated in Figure 11. In the left panel of the figure, the saddle-node bifurcation point approximation $u^* \in \mathbb{R}^{10}$ is shown in black, while the solutions on the created branches at $\lambda = 300$ are shown in blue and red. The right image shows the location of the bifurcation point as a black star. In this image, the vertical axis denotes the λ-value at which a saddle-node bifurcation occurs, while the integer N on the horizontal axis encodes the ± 1-pattern of the resulting mosaic solution through the binary representation of N. The above $\bar{u}_m$ corresponds to $N = 21$. We would like to point out that all of the remaining dots in the right panel of Figure 11 have been validated using the above approach. For illustration purposes, this panel only contains half of the N-values contained in the right image of Figure 3. Yet, validation was performed on all points.

All of these sample computations focused on the specific nonlinearity $f(u) = u - u^3$. Notice, however, that the formula for L in (50) is based on the general constants in (49), and these can easily be adjusted for other nonlinearities f. For more examples, we refer the reader to [47]. ◇

Due to time and space constraints, we refrain from detailed applications to the diblock copolymer model in the context of saddle-node bifurcations. These can be found in [31], and will briefly be mentioned at the end of the next section.

4.2 Symmetry-Breaking Pitchfork Bifurcations

It was mentioned in the introduction that in general generic systems one would only expect saddle-node bifurcation points. This is due to the fact that in view of Sard's theorem, there always exist arbitrarily small perturbations of a nonlinear function $\mathcal{F}(\lambda, u)$ such that zero becomes a regular value. This implies that at any zero (λ_0, u_0) of $\mathcal{F}$ the total derivative has full rank, and this in turn yields that the implicit function theorem can be applied *in some direction*. In other words, the solution set of $\mathcal{F}(\lambda, u) = 0$ is comprised of smooth curves which do not intersect.

In reality, however, the situation is more complicated. The above argument requires that we allow for *arbitrary perturbations*. But what if the model under consideration has certain internal structures which should be part of any realistic perturbation? In such cases, Sard's theorem would not apply, and it is therefore feasible to assume that for certain models, it might not be possible to break up pitchfork bifurcations via *realistic perturbations*. One situation which is frequently encountered in applications involves *inherent symmetries* in models. Such symmetries can come in a great variety of forms, and we refer the reader to [11] for an in-depth account. For the purposes of this lecture, we assume that symmetry occurs in the following simple specific form.

Definition 4.6 (Symmetry in the Form of $\mathbb{Z}_2$-Equivariance)**.** *In the situation of Assumptions 4.1 and 4.2, suppose there exist bounded linear operators $S_{\mathcal{U}} \in \mathcal{L}(\mathcal{U},\mathcal{U})$ and $S_{\mathcal{V}} \in \mathcal{L}(\mathcal{V},\mathcal{V})$ such that*

$$S_{\mathcal{U}} \neq I\ , \qquad S_{\mathcal{V}} \neq I\ , \qquad S_{\mathcal{U}}^2 = I\ , \qquad S_{\mathcal{V}}^2 = I\ ,$$

as well as

$$\mathcal{F}\left(\lambda, S_{\mathcal{U}} u\right) \;=\; S_{\mathcal{V}} \mathcal{F}\left(\lambda, u\right) \quad \text{for all} \quad \lambda \in \mathbb{R} \quad \text{and} \quad u \in \mathcal{U}\ . \tag{51}$$

Then we say that $\mathcal{F}$ is $\mathbb{Z}_2$-equivariant*. Based on the symmetry operators $S_{\mathcal{U}}$ and $S_{\mathcal{V}}$ one can also decompose the underlying Banach spaces into the* symmetric elements *and the* antisymmetric elements*. More precisely, for the Banach space $\mathcal{U}$ we define*

$$\mathcal{U}_s = \{u \in \mathcal{U}\ :\ S_{\mathcal{U}} u = u\} \qquad \text{and} \qquad \mathcal{U}_a = \{u \in \mathcal{U}\ :\ S_{\mathcal{U}} u = -u\}\ ,$$

and analogously one can define the subspaces $\mathcal{V}_s$ and $\mathcal{V}_a$ of $\mathcal{V}$. One can easily see that

$$\mathcal{U} = \mathcal{U}_s \oplus \mathcal{U}_a \qquad \text{as well as} \qquad \mathcal{V} = \mathcal{V}_s \oplus \mathcal{V}_a\ ,$$

since $u = (u + S_{\mathcal{U}} u)/2 + (u - S_{\mathcal{U}} u)/2$ yields a decomposition of an arbitrary element u into a symmetric and an antisymmetric one.

For the nonlinear equation (43), equivariance has immediate consequences regarding invariance of subspaces, both for the equation and for certain derivatives of $\mathcal{F}$. For example, by differentiating the identity (51) with respect to u one immediately obtains

$$D_u \mathcal{F}\left(\lambda, S_{\mathcal{U}} u\right)\left[S_{\mathcal{U}} \bar{u}\right] = S_{\mathcal{V}} D_u \mathcal{F}\left(\lambda, u\right)\left[\bar{u}\right] \tag{52}$$

for all $\lambda \in \mathbb{R}$ and arbitrary $u, \bar{u} \in \mathcal{U}$. If one assumes in addition that $u \in \mathcal{U}_s$, then (52) implies the $\mathbb{Z}_2$-equivariance of the Fréchet derivative $D_u \mathcal{F}(\lambda, u)$. The following lemma collects a number of similar properties, which shed more light on the effects of $\mathbb{Z}_2$-equivariance. The straightforward proof can be found in [53].

Lemma 4.7 (General Consequences of $\mathbb{Z}_2$-Equivariance)**.** *Suppose that Assumptions 4.1 and 4.2 hold, and that (43) is $\mathbb{Z}_2$-equivariant as in Definition 4.6. Then the following statements are true for all $\lambda \in \mathbb{R}$ and $u \in \mathcal{U}_s$:*

(a) We have both $\mathcal{F}(\lambda, u) \in \mathcal{V}_s$ and $D_\lambda \mathcal{F}(\lambda, u) \in \mathcal{V}_s$.

(b) Both the inclusions $D_u \mathcal{F}(\lambda, u)[\mathcal{U}_s] \subset \mathcal{V}_s$ and $D_u \mathcal{F}(\lambda, u)[\mathcal{U}_a] \subset \mathcal{V}_a$ are satisfied, as well as $D_{\lambda u} \mathcal{F}(\lambda, u)[\mathcal{U}_s] \subset \mathcal{V}_s$ and $D_{\lambda u} \mathcal{F}(\lambda, u)[\mathcal{U}_a] \subset \mathcal{V}_a$.

(c) The inclusions $D_{uu} \mathcal{F}(\lambda, u)[\mathcal{U}_s, \mathcal{U}_s] \subset \mathcal{V}_s$ and $D_{uu} \mathcal{F}(\lambda, u)[\mathcal{U}_a, \mathcal{U}_a] \subset \mathcal{V}_s$ are satisfied, as well as $D_{uu} \mathcal{F}(\lambda, u)[\mathcal{U}_s, \mathcal{U}_a] \subset \mathcal{V}_a$.

The above lemma applies to the derivatives of $\mathcal{F}$ at any pair (λ, u), as long as $u \in \mathcal{U}_s$ is a symmetric element. If we consider specifically a pair (λ_0, u_0) as in Assumption 4.2, even more can be said. As the following result shows, both the eigenfunction φ_0 and the element ψ_0^* necessarily have to be either symmetric or antisymmetric. In addition, the complementary spaces $\tilde{\mathcal{U}}$ and $\tilde{\mathcal{V}}$ in Assumption 4.2 can always be chosen in such a way that they respect the symmetry operations as well.

Lemma 4.8 ($\mathbb{Z}_2$-Equivariance and One-Dimensional Kernels)**.** *Suppose that Assumptions 4.1 and 4.2 hold, that (43) is $\mathbb{Z}_2$-equivariant as in Definition 4.6, and that $u_0 \in \mathcal{U}_s$. Then the following statements are true:*

(a) The eigenfunction φ_0 which spans the nullspace $N(L)$ of $L = D_u \mathcal{F}(\lambda_0, u_0)$ is either an element of $\mathcal{U}_s$ or of $\mathcal{U}_a$, i.e., we have $S_{\mathcal{U}} \varphi_0 = \varepsilon_{\mathcal{U}} \varphi_0$ for some $\varepsilon_{\mathcal{U}} \in \{\pm 1\}$.

(b) The element ψ_0^* which characterizes $R(L)$ is either symmetric or antisymmetric with respect to the equivariance $S_{\mathcal{V}}^*$, i.e., we have $S_{\mathcal{V}}^* \psi_0^* = \varepsilon_{\mathcal{V}} \psi_0^*$ for some $\varepsilon_{\mathcal{V}} \in \{\pm 1\}$.

(c) The projections P and Q in Assumption 4.2 can be chosen in such a way that they commute with the symmetry actions, i.e., such that $S_{\mathcal{V}} P = P S_{\mathcal{V}}$ and $S_{\mathcal{U}} Q = Q S_{\mathcal{U}}$.

At first glance, it seems somewhat surprising that the kernel function φ_0 has to be either a symmetric or an antisymmetric element. Notice, however, that due to Lemma 4.7*(b)* we have both $L\mathcal{U}_s \subset \mathcal{V}_s$ and $L\mathcal{U}_a \subset \mathcal{V}_a$, i.e., the splitting into symmetric and antisymmetric elements is respected by L. If we now consider

$$\varphi_0 = \varphi_{0,s} + \varphi_{0,a} \in \mathcal{U}_s \oplus \mathcal{U}_a \quad \text{with} \quad \varphi_{0,s} = \frac{\varphi_0 + S_{\mathcal{U}} \varphi_0}{2} \quad \text{and} \quad \varphi_{0,a} = \frac{\varphi_0 - S_{\mathcal{U}} \varphi_0}{2} \,,$$

then the splitting furnishes $0 = L\varphi_0 = L\varphi_{0,s} + L\varphi_{0,a}$, and therefore $L\varphi_{0,s} = -L\varphi_{0,a}$, which in turn implies that both $L\varphi_{0,s}$ and $L\varphi_{0,a}$ are contained in $\mathcal{U}_s \cap \mathcal{U}_a = \{0\}$. Together, this shows that both $\varphi_{0,s}$ and $\varphi_{0,a}$ are in $N(L)$, hence multiples of φ_0. Since one of the two elements has to be nonzero, this yields that either $\varphi_0 \in \mathcal{U}_s$ or $\varphi_0 \in \mathcal{U}_a$, which is equivalent to $S_{\mathcal{U}} \varphi_0 = \varphi_0$ or $S_{\mathcal{U}} \varphi_0 = -\varphi_0$, respectively. In other words, statement *(a)* is true. The remaining parts of Lemma 4.8 can be established using similar arguments, see also [31].

After these preparations we now return to the study of pitchfork bifurcations. We will see below that such bifurcation points frequently occur in nonlinear problems which are $\mathbb{Z}_2$-equivariant. Similar to our approach in the last section, we first need to formulate a sufficient

condition which guarantees the existence of a pitchfork bifurcation, and then as a second step reformulate this condition as a zero finding problem to make it amenable to the numerical implicit function theorem. For the first step, we make use of a classical technique in nonlinear analysis, which reduces the solution of a possibly infinite-dimensional algebraic system to the solution of a finite- and usually low-dimensional equation. This method is summarized in the following proposition.

Proposition 4.9 (Lyapunov-Schmidt Reduction)**.** *In the situation of Assumptions 4.1 and 4.2 there exist a neighborhood Λ_0 of λ_0, a neighborhood V_0 of $v_0 = Qu_0 \in N(L)$, a smooth function $W : \Lambda_0 \times V_0 \to \tilde{\mathcal{U}}$, as well as a smooth real-valued function b which is defined in a neighborhood of the point $(\lambda_0, 0) \in \mathbb{R}^2$ such that the following hold:*

(a) If (λ, α) is sufficiently close to the point $(\lambda_0, 0) \in \mathbb{R}^2$ and satisfies $b(\lambda, \alpha) = 0$, then

$$\mathcal{F}(\lambda, u) = 0 \quad \text{for} \quad u = v_0 + \alpha\varphi_0 + W\left(\lambda, v_0 + \alpha\varphi_0\right) .$$

(b) Conversely, if (λ, u) is close enough to (λ_0, u_0) and solves $\mathcal{F}(\lambda, u) = 0$, then for α defined via $\alpha\varphi_0 = Q(u - u_0)$ we have $b(\lambda, \alpha) = 0$ and $u = Qu + W(\lambda, Qu)$. Notice that in this case we have $Qu = v_0 + \alpha\varphi_0$.

In other words, the solution set of the

$$\text{bifurcation equation} \qquad b(\lambda, \alpha) = 0$$

in a neighborhood of the point $(\lambda_0, 0) \in \mathbb{R}^2$ is in one-to-one correspondence to the solution set of $\mathcal{F}(\lambda, u) = 0$ in a neighborhood of (λ_0, u_0).

Even though we do not prove the result in detail in these lecture notes, the basic idea can easily be conveyed. Using the projections P and Q from Assumption 4.2, one can see that solving the nonlinear problem (43) is equivalent to solving the pair of equations

$$P\mathcal{F}(\lambda, v + w) = 0 \qquad \text{and} \qquad (I - P)\mathcal{F}(\lambda, v + w) = 0 , \tag{53}$$

where $v = Qu$ and $w = (I-Q)u$. Using the abbreviation $G(\lambda, v, w) = (I-P)\mathcal{F}(\lambda, v+w)$, our assumptions imply that $G : \mathbb{R} \times N(L) \times \tilde{\mathcal{U}} \to R(L)$. Now set $v_0 = Qu_0$ and $w_0 = (I - Q)u_0$. Then we have $G(\lambda_0, v_0, w_0) = 0$, and one can show that $D_w G(\lambda_0, v_0, w_0) \in \mathcal{L}(\tilde{\mathcal{U}}, R(L))$ is a continuous isomorphism. The implicit function theorem can then be used to solve the second equation in (53) locally for w as a function of λ and v. This guarantees the function W as in the formulation of the theorem. Plugging W into the first equation (53) shows that locally near (λ_0, u_0) a pair (λ, u) solves the nonlinear problem $\mathcal{F}(\lambda, u) = 0$ if and only if the pair $(\lambda, v) = (\lambda, Qu)$ satisfies $P\mathcal{F}(\lambda, v + W(\lambda, v)) = 0$. If we finally set

$$b(\lambda, \alpha) = \psi_0^* \left(P\mathcal{F}\left(\lambda, v_0 + \alpha\varphi_0 + W(\lambda, v_0 + \alpha\varphi_0)\right)\right) , \tag{54}$$

then Proposition 4.9 follows due to the choices of ψ_0^* and P.

In our setting of a linearization which is a Fredholm operator of index zero and which has a one-dimensional kernel at the pair (λ_0, u_0), the bifurcation equation $b(\lambda, \alpha) = 0$ is

particularly simple. It is a single equation in two real unknowns. In fact, one can easily derive a Taylor expansion for the bifurcation function $b(\lambda, \alpha)$ near the point $(\lambda_0, 0)$ in the form

$$\begin{aligned} b(\lambda_0 + \nu, \alpha) &= \nu \cdot D_\lambda b(\lambda_0, 0) + \frac{\nu^2}{2} \cdot D_{\lambda\lambda} b(\lambda_0, 0) + \alpha\nu \cdot D_{\lambda\alpha} b(\lambda_0, 0) + \frac{\alpha^2}{2} \cdot D_{\alpha\alpha} b(\lambda_0, 0) \\ &\quad + \frac{\nu^3}{6} \cdot D_{\lambda\lambda\lambda} b(\lambda_0, 0) + \frac{\alpha\nu^2}{2} \cdot D_{\lambda\lambda\alpha} b(\lambda_0, 0) + \frac{\alpha^2\nu}{2} \cdot D_{\lambda\alpha\alpha} b(\lambda_0, 0) \\ &\quad + \frac{\alpha^3}{6} \cdot D_{\alpha\alpha\alpha} b(\lambda_0, 0) + R(\nu, \alpha) \, , \end{aligned}$$

where $R(\nu, \alpha) = O(\|(\nu, \alpha)\|^4)$. Despite the fact that the function W usually is not known explicitly, one can explicitly compute the derivatives of b at the pair $(\lambda_0, 0)$, see [31] for more details. For example, one can show that $D_\lambda b(\lambda_0, 0) = \psi_0^* D_\lambda F(\lambda_0, u_0)$, and this can immediately be used to give an alternative proof of Proposition 4.3.

For our study of pitchfork bifurcations, we need a refinement of Proposition 4.9 which takes $\mathbb{Z}_2$-equivariance into account. This leads to the following result, whose proof can be found in [11, 31].

Proposition 4.10 (Equivariant Lyapunov-Schmidt Reduction)**.** *Suppose that all the assumptions of Proposition 4.9 are satisfied, and that $\mathcal{F}$ is $\mathbb{Z}_2$-equivariant as in Definition 4.6. Finally, assume that $u_0 \in \mathcal{U}_s$, and suppose that the projections P and Q are chosen as in Lemma 4.8(c). Then the following hold:*

(a) The nullspace $N(L)$ of $L = D_u \mathcal{F}(\lambda_0, u_0)$ is invariant under $S_\mathcal{U}$, and the range $R(L)$ is invariant under $S_\mathcal{V}$.

(b) The mapping W appearing in the the Lyapunov-Schmidt reduction in Proposition 4.9 is $\mathbb{Z}_2$-equivariant, i.e., we have

$$W(\lambda, S_\mathcal{U} v) = S_\mathcal{U} W(\lambda, v) \quad \text{for all} \quad \lambda \in \mathbb{R} \ \text{ and } \ v \in N(L) \, .$$

(c) The bifurcation function $b(\lambda, \alpha)$ satisfies

$$b(\lambda, \varepsilon_\mathcal{U} \alpha) = \varepsilon_\mathcal{V} b(\lambda, \alpha) \quad \text{for all} \quad (\lambda, \alpha) \ \text{ close to } \ (\lambda_0, 0) \, ,$$

where $\varepsilon_\mathcal{U}$ and $\varepsilon_\mathcal{V}$ encode the symmetry properties of φ_0 and ψ_0^* as in Lemma 4.8.

While in general it does not seem to be the case that $\varepsilon_\mathcal{U}$ and $\varepsilon_\mathcal{V}$ have to be related, there are special situations when they have to coincide, see for example the discussion in [31]. Of particular interest for us is the case $\varepsilon_\mathcal{U} = \varepsilon_\mathcal{V} = -1$, since then Proposition 4.10*(c)* implies that the bifurcation function b defined in (54) is an odd function with respect to α. This in turn automatically guarantees that there is a trivial solution line $(\lambda, \alpha) = (\lambda, 0)$ for the bifurcation equation $b(\lambda, \alpha) = 0$, for all λ close to λ_0. In fact, in this situation one generically can observe pitchfork bifurcations, which are called *symmetry-breaking pitchfork bifurcations*. A sufficient condition for their existence is the subject of the next result.

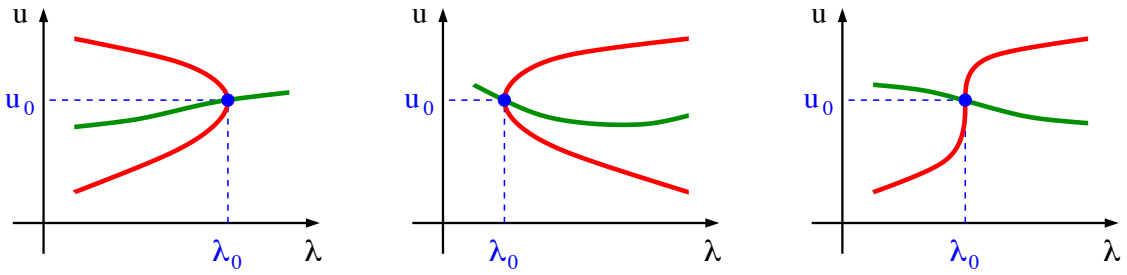

Figure 12: Visualization of the sufficient condition for a symmetry-breaking pitchfork bifurcation given in Proposition 4.11. The three panels show possible shapes of the solution branches of (43) near the solution (λ_0, u_0). While the first two panels are for the cases $\gamma > 0$ and $\gamma < 0$, respectively, the third panel shows one of several cases which are possible in the inconclusive case $\gamma = 0$. In all images, the branch of symmetric solutions is shown in green, and the symmetry-breaking branch is shown in red.

Proposition 4.11 (Sufficient Condition for Symmetry-Breaking Pitchfork Bifurcation)**.** *Suppose that Assumptions 4.1 and 4.2 hold, and that (43) is $\mathbb{Z}_2$-equivariant as in Definition 4.6. Furthermore, assume that*

$$S_{\mathcal{U}} u_0 = u_0 \,, \qquad S_{\mathcal{U}} \varphi_0 = -\varphi_0 \,, \qquad \text{as well as} \qquad S^*_{\mathcal{V}} \psi^*_0 = -\psi^*_0 \,, \tag{55}$$

that the projections P and Q are chosen as in Lemma 4.8(c), and that

$$\psi^*_0 D_{\lambda u}\mathcal{F}(\lambda_0, u_0)[\varphi_0] + \psi^*_0 D_{uu}\mathcal{F}(\lambda_0, u_0)[\varphi_0, \xi_0] \neq 0 \,, \tag{56}$$

where $\xi_0 \in \mathcal{U}_s$ solves the equation

$$D_u\mathcal{F}(\lambda_0, u_0)[\xi_0] + (I - P) D_\lambda \mathcal{F}(\lambda_0, u_0) = 0 \,. \tag{57}$$

Then the nonlinear problem (43) undergoes a pitchfork bifurcation at (λ_0, u_0). Locally at this point the solution set of (43) consists of a smooth solution curve parameterized by λ, together with a parabolic curve which is tangent to φ_0 at (λ_0, u_0). Consider the ratio

$$\gamma = \frac{\psi^*_0 D_{uuu}\mathcal{F}(\lambda_0, u_0)[\varphi_0, \varphi_0, \varphi_0] + 3\psi^*_0 D_{uu}\mathcal{F}(\lambda_0, u_0)[\varphi_0, \zeta_0]}{\psi^*_0 D_{\lambda u}\mathcal{F}(\lambda_0, u_0)[\varphi_0] + \psi^*_0 D_{uu}\mathcal{F}(\lambda_0, u_0)[\varphi_0, \xi_0]} \,,$$

where ξ_0 was defined in (57) and $\zeta_0 \in \mathcal{U}_s$ is defined by

$$D_u\mathcal{F}(\lambda_0, u_0)[\zeta_0] + (I - P) D_{uu}\mathcal{F}(\lambda_0, u_0)[\varphi_0, \varphi_0] = 0 \,.$$

If the ratio γ is positive, then the solutions on the parabolic branch exist for $\lambda < \lambda_0$ close to the bifurcation point, if γ is negative then they exist for $\lambda > \lambda_0$. If $\gamma = 0$, either half of the parabolic branch could lie on either side of λ_0.

This sufficient condition is illustrated in Figure 12. From left to right the images show the three cases outlined in the proposition. While the first two panels are for $\gamma > 0$ and $\gamma < 0$,

respectively, the last panel is for the inconclusive situation $\gamma = 0$. In all images, the green curve represents the primary branch of symmetric solutions, which correspond to the trivial solution of the bifurcation equation $b(\lambda, \alpha) = 0$. In contrast, solutions on the red bifurcating branch break this symmetry, since in a small neighborhood of the bifurcation point (λ_0, u_0) they are close to the line $\alpha \mapsto u_0 + \alpha\varphi_0$, which consists of linear combinations of the symmetric solution u_0 and the antisymmetric eigenfunction φ_0.

The detailed proof of Proposition 4.11 can be found in [31], but its main idea can easily be described. We have already seen earlier that due to (55), the bifurcation equation $b(\lambda, \alpha) = 0$ guaranteed by Propositions 4.9 and 4.10 has the trivial solution $b(\lambda, 0) \equiv 0$. If we then define a function r in a neighborhood of $(\lambda_0, 0)$ by setting

$$r(\lambda, \alpha) = \begin{cases} \dfrac{b(\lambda, \alpha)}{\alpha} & \text{for} \quad \alpha \neq 0 \,, \\ \\ D_\alpha b(\lambda, 0) & \text{for} \quad \alpha = 0 \,, \end{cases}$$

one can show that close to $(\lambda_0, 0)$ the mapping r is smooth with $b(\lambda, \alpha) = \alpha \cdot r(\lambda, \alpha)$. While the condition $\alpha = 0$ corresponds to the primary branch of symmetric solutions, the bifurcating branch has to solve the equation $r(\lambda, \alpha) = 0$. From the expansion of the bifurcation function b stated earlier one can see that

$$\begin{aligned} r(\lambda_0, 0) &= D_\alpha b(\lambda_0, 0) = 0 \,, \quad \text{and} \\ D_\lambda r(\lambda_0, 0) &= D_{\lambda\alpha} b(\lambda_0, 0) = \psi_0^* D_{\lambda u} F(\lambda_0, u_0)[\varphi_0] + \psi_0^* D_{uu} F(\lambda_0, u_0)[\varphi_0, \xi_0] \neq 0 \,, \end{aligned}$$

where ξ_0 was defined in (57), and the inequality in the second line follows from (56). An application of the implicit function theorem then implies that the equation $r(\lambda, \alpha) = 0$ can be solved locally for λ as a function of α, and this gives rise to the red branch in Figure 12. For the tangency statements of the proposition, we refer the reader to [31].

Finally we are in a position to return to computer-assisted proofs. As the last step, we need to reformulate the conditions of Proposition 4.11 in such a way that the numerical implicit function theorem can be used to infer the existence of symmetry-breaking pitchfork bifurcations from a numerical approximation. Unfortunately, however, we can clearly no longer use the extended system (47), since an isolated zero of the system implies the existence of a saddle-node bifurcation point. While it is still true that (47) has a zero at a pitchfork bifurcation point, the Jacobian matrix of the mapping $\mathcal{G}$ defined in (48) at this zero is no longer invertible.

Luckily, the prevalent role of symmetries in the formation of pitchfork bifurcations allows for a straightforward adjustment. For this, we still consider the extended system as defined in (47). This time, however, the map $\mathcal{G}$ induced by the system is restricted in domain and range to certain symmetry-induced subspaces. More precisely, consider $\mathcal{G}$ defined as

$$\mathcal{G} \;:\; \begin{cases} \mathbb{R} \times \mathcal{U}_s \times \mathcal{U} & \to & \mathbb{R} \times \mathcal{V}_s \times \mathcal{V} \\ (\lambda, u, v) & \mapsto & (\ell(v) - 1,\ \mathcal{F}(\lambda, u),\ D_u \mathcal{F}(\lambda, u)[v]) \end{cases} \tag{58}$$

where the nonlinear operator part of the map is restricted to the subspace $\mathcal{U}_s$ of symmetric elements of $\mathcal{U}$. Notice that the restriction to $\mathcal{V}_s$ in the image is justified by Lemma 4.7*(a)*.

Then the following result holds, which is a slight extension of an analogous result in [48] to the case of different Banach spaces in domain and range. For the proof, see [31].

Theorem 4.12 (Symmetry-Breaking Pitchfork Bifurcations via Extended Systems)**.** *Suppose that Assumption 4.1 is satisfied. Then the following two statements hold.*

(a) Suppose that all assumptions of Proposition 4.11 are satisfied, and let $\ell \in \mathcal{U}^*$ be a functional such that $\ell(\varphi_0) = 1$. Then the Fréchet derivative $D_{(\lambda,u,v)}\mathcal{G}(\lambda_0, u_0, \varphi_0)$ of the mapping defined in (58) is invertible, i.e., the solution $(\lambda_0, u_0, \varphi_0)$ of the extended system $\mathcal{G}(\lambda, u, \varphi) = (0, 0, 0)$ is an isolated non-degenerate zero.

(b) Conversely, if there exist an $\ell \in \mathcal{U}^*$ and a $\varphi_0 \in \mathcal{U}_a$ such that $(\lambda_0, u_0, \varphi_0)$ is an isolated non-degenerate zero of the equation $\mathcal{G}(\lambda, u, \varphi) = (0, 0, 0)$, and if the last identity in (55) holds for a nontrivial element $\psi_0^* \in N(L^*)$, then the full system $\mathcal{F}$ satisfies all assumptions of Proposition 4.11. In other words, the point (λ_0, u_0) is a symmetry-breaking pitchfork bifurcation point for (43).

As in the case of saddle-node bifurcations, Theorem 4.12 allows us to establish the existence of symmetry-breaking pitchfork bifurcations using the framework from Section 3. By applying the parameter-independent version of Theorem 3.1 to the map $\mathcal{G}$, single bifurcation points can be verified, and branches of pitchfork bifurcations can be validated using Theorems 3.5 and 3.7, as long as the problem under consideration depends on an additional parameter. We close this section with a brief application to the diblock copolymer model.

Example 4.13 (Bifurcations in the Diblock Copolymer Model)**.** As we saw in Section 2.2, equilibrium solutions of the diblock copolymer model have to satisfy the nonlinear elliptic problem (8). Consider now the special case of total mass $\mu = 0$ and $\Omega = (0, 1)$. Then finding equilibrium solutions of the evolution equation is equivalent to solving the nonlinear operator equation

$$\mathcal{F}(\lambda, u) \;=\; -\left(u_{xx} + \lambda f(u)\right)_{xx} - \lambda\sigma u \;\;=\;\; 0\,,$$

where $\mathcal{F} : \mathbb{R} \times \mathcal{U} \to \mathcal{V}$ for suitable Banach spaces $\mathcal{U}$ and $\mathcal{V}$. In the interest of space and time, we only mention that these spaces can be defined using Fourier cosine series, where the domain $\mathcal{U}$ incorporates the boundary conditions and the mass constraint $\int_0^1 u\,dx = 0$. See [31] for more details.

Now assume that the pair $(\lambda_0, u_0) \in \mathbb{R} \times \mathcal{U}$ is arbitrary. Then the Fréchet derivative of the nonlinear operator $\mathcal{F}$ with respect to u is given by

$$Lv \;=\; D_u\mathcal{F}(\lambda_0, u_0)[v] \;=\; -\left(v_{xx} + \lambda_0 f'(u_0)v\right)_{xx} - \lambda_0\sigma v\,,$$

and the associated adjoint operator is of the form $L^*w = -w_{xxxx} - \lambda_0 f'(u_0)w_{xx} - \lambda_0\sigma w$. One can then show that the operator L is a Fredholm operator of index one. Furthermore, if we assume that $N(L)$ is one-dimensional, and if ψ_0 denotes a solution of the linear elliptic problem $L^*w = 0$ subject to homogeneous Neumann boundary conditions for both w and w_{xx}, then the range $R(L)$ of L can be characterized as

$$R(L) = N(\psi_0^*)\,, \qquad \text{where} \qquad \psi_0^*(w) = \int_0^1 \psi_0(x)w(x)\,dx\,.$$

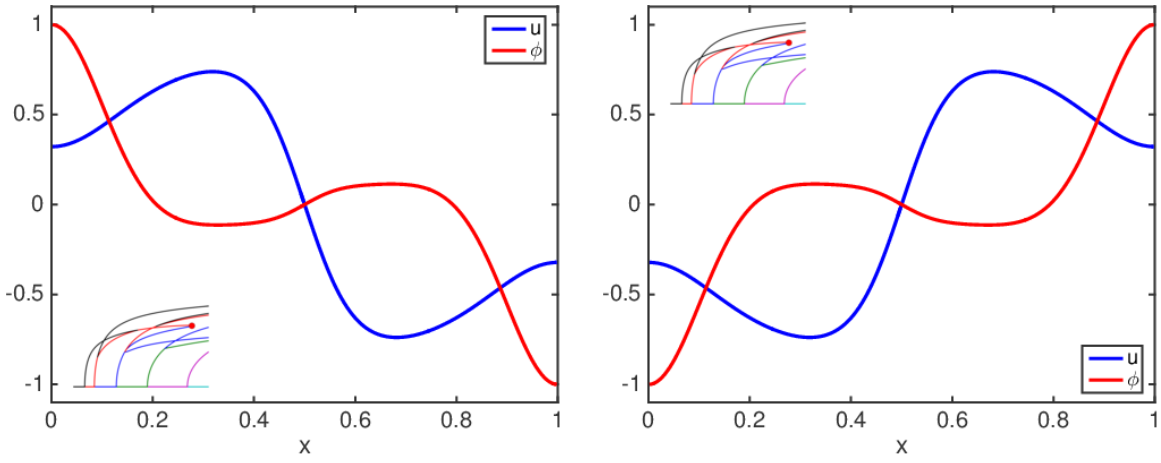

Figure 13: Saddle-node bifurcation points for the diblock copolymer model with $\sigma = 6$, total mass $\mu = 0$, and on the one-dimensional domain $\Omega = (0, 1)$. The two blue functions are approximations to saddle-node bifurcation points at the same parameter value $\lambda^* \approx 262.9$, and with the same $L^2(0, 1)$-norm close to 0.562. Using Theorem 4.4 one can prove that actual saddle-node bifurcation points are nearby. The red curves are approximations to kernel functions at the bifurcation points.

In other words, Assumptions 4.1 and 4.2 are satisfied.

We can now apply the characterizations given in Theorems 4.4 and 4.12 to rigorously establish bifurcation points for the diblock copolymer model close to numerical approximations. This was done in [31] using the radii polynomial approach, which was prominently displayed in most of the other lectures. Rather than going through the specific estimates, we therefore only survey a couple of results:

- In the context of saddle-node bifurcation points, Figure 13 shows two numerical approximations u^* of solutions of the nonlinear elliptic equation $\mathcal{F}(\lambda^*, u) = 0$ at the same parameter value $\lambda^* \approx 262.9$. In both of these cases, Theorem 4.4 can be applied successfully, and it furnishes actual saddle-node points in a small neighborhood.

- Similarly, symmetry-breaking pitchfork bifurcation points can be validated using Theorem 4.12. For this, consider Figure 14, which shows four approximations u^* to equilibrium solutions in blue. In each image, the location of the approximating stationary state in the bifurcation diagram is indicated in the overlayed diagram, and an approximation to the normalized kernel function of the Fréchet derivative of $\mathcal{F}$ with respect to u at the point u^* is drawn in red. In order to apply the above result, we need to specify suitable symmetry operations. Consider for example the symmetry operators $S_{\mathcal{U}}$ and $S_{\mathcal{V}}$ defined by the same formula

$$(Su)(x) = u(1 - x) \qquad \text{for} \qquad x \in \Omega = (0, 1) \, .$$

  Then $\mathcal{F}$ is $\mathbb{Z}_2$-equivariant, and the solution approximations in the right column of Figure 14 are symmetric with respect to S, while the corresponding kernel functions

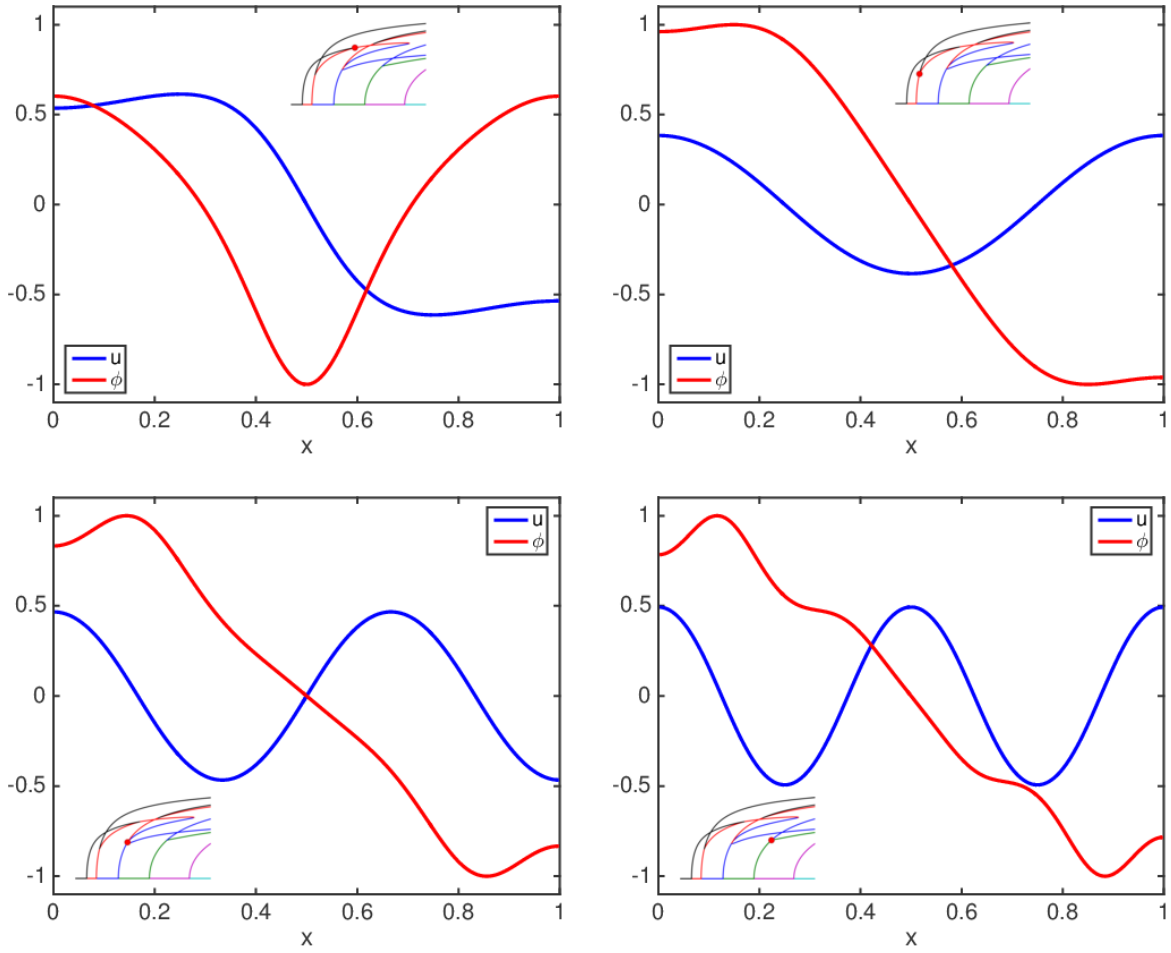

Figure 14: Symmetry-breaking pitchfork bifurcation points for the diblock copolymer model with $\sigma = 6$, total mass $\mu = 0$, and on the one-dimensional domain $\Omega = (0, 1)$. In each panel, an approximation to an equilibrium solution is shown in blue, an approximation to the normalized kernel function of the Fréchet derivative of $\mathcal{F}$ with respect to u at this point is drawn in red, and the location of the solution in the bifurcation diagram is indicated in the overlayed diagram. Using Theorem 4.12 one can show that close to the functions shown in the top row, as well as in the bottom right image, there are true pitchfork bifurcation points. Note, however, that the situation shown in the lower left image is not covered by our approach.

are antisymmetric. Moreover, one can show that the functions in the nullspace of the adjoint operator are antisymmetric as well. In both cases, Theorem 4.12 can be applied, and this guarantees symmetry-breaking pitchfork bifurcation points which lie on the second and fourth bifurcation branches from the trivial solution curve shown in the lower left image of Figure 5.

But what about the pitchfork bifurcation points on the first and third branches? These

cases are depicted in the left column of Figure 14, and the solutions on the branch are no longer symmetric with respect to the above symmetry. Yet, if instead we define a new symmetry operator via

$$(Su)(x) = -u(1-x) \qquad \text{for} \qquad x \in \Omega = (0,1)\ ,$$

then $\mathcal{F}$ remains $\mathbb{Z}_2$-equivariant, and the blue functions in the left column of Figure 14 are again symmetric elements. Since the kernel function shown in the upper left image of the figure is antisymmetric with respect to this new symmetry, one can apply Theorem 4.12, and this guarantees an actual bifurcation point nearby. Note, however, that in the lower left panel of Figure 14 the kernel function is not antisymmetric, so our result no longer applies. In fact, this bifurcation is not a $\mathbb{Z}_2$-symmetry-breaking bifurcation, but rather a symmetry-breaking bifurcation with a different equivariance group.

For more details and explicit bounds we refer the reader to [31]. ◇

Acknowledgments

The author was partially supported by NSF grants DMS-1114923 and DMS-1407087.

References

[1] R. A. Adams. *Sobolev Spaces*. Academic Press, San Diego – London, 1978.

[2] G. Arioli and H. Koch. Computer-assisted methods for the study of stationary solutions in dissipative systems, applied to the Kuramoto-Sivashinski equation. *Archive for Rational Mechanics and Analysis*, 197(3):1033–1051, 2010.

[3] M. Bahiana and Y. Oono. Cell dynamical system approach to block copolymers. *Physical Review A*, 41:6763–6771, 1990.

[4] F. S. Bates. Block copolymer thermodynamics: Theory and experiment. *Annual Review of Physical Chemistry*, 41:525–557, 1990.

[5] F. S. Bates and G. H. Fredrickson. Block copolymers–designer soft materials. *Physics Today*, 52:32–39, 1999.

[6] P. W. Bates and P. C. Fife. The dynamics of nucleation for the Cahn-Hilliard equation. *SIAM Journal on Applied Mathematics*, 53(4):990–1008, 1993.

[7] D. Blömker, B. Gawron, and T. Wanner. Nucleation in the one-dimensional stochastic Cahn-Hilliard model. *Discrete and Continuous Dynamical Systems, Series A*, 27(1):25–52, 2010.

[8] R. F. Brown. *A Topological Introduction to Nonlinear Analysis*. Birkhäuser, Boston, 1993.

[9] J. W. Cahn, S.-N. Chow, and E. S. Van Vleck. Spatially discrete nonlinear diffusion equations. *The Rocky Mountain Journal of Mathematics*, 25(1):87–118, 1995.

[10] J. W. Cahn and J. E. Hilliard. Free energy of a nonuniform system I. Interfacial free energy. *Journal of Chemical Physics*, 28:258–267, 1958.

[11] P. Chossat and R. Lauterbach. *Methods in Equivariant Bifurcations and Dynamical Systems*, volume 15 of *Advanced Series in Nonlinear Dynamics*. World Scientific, 2000.

[12] S.-N. Chow and J. K. Hale. *Methods of Bifurcation Theory*. Springer-Verlag, New York – Berlin – Heidelberg, 1982.

[13] S.-N. Chow and J. Mallet-Paret. Pattern formation and spatial chaos in lattice dynamical systems. I. *IEEE Transactions on Circuits and Systems. I. Fundamental Theory and Applications*, 42(10):746–751, 1995.

[14] S.-N. Chow, J. Mallet-Paret, and E. S. Van Vleck. Dynamics of lattice differential equations. *International Journal of Bifurcation and Chaos in Applied Sciences and Engineering*, 6(9):1605–1621, 1996.

[15] S.-N. Chow, J. Mallet-Paret, and E. S. Van Vleck. Pattern formation and spatial chaos in spatially discrete evolution equations. *Random & Computational Dynamics*, 4(2-3):109–178, 1996.

[16] S. Day, J.-P. Lessard, and K. Mischaikow. Validated continuation for equilibria of PDEs. *SIAM Journal on Numerical Analysis*, 45(4), 2007.

[17] J. P. Desi, H. Edrees, J. Price, E. Sander, and T. Wanner. The dynamics of nucleation in stochastic Cahn-Morral systems. *SIAM Journal on Applied Dynamical Systems*, 10(2):707–743, 2011.

[18] E. Doedel. AUTO: a program for the automatic bifurcation analysis of autonomous systems. In *Proceedings of the Tenth Manitoba Conference on Numerical Mathematics and Computing, Vol. I (Winnipeg, Man., 1980)*, volume 30, pages 265–284, 1981.

[19] D. E. Edmunds and W. D. Evans. *Spectral Theory and Differential Operators*. Oxford University Press, Oxford – New York, 1987.

[20] C. M. Elliott and A. M. Stuart. The global dynamics of discrete semilinear parabolic equations. *SIAM Journal on Numerical Analysis*, 30(6):1622–1663, 1993.

[21] M. Gameiro, J.-P. Lessard, and K. Mischaikow. Validated continuation over large parameter ranges for equilibria of PDEs. *Mathematics and Computers in Simulation*, 79(4):1368–1382, 2008.

[22] C. P. Grant. Grain sizes in the discrete Allen-Cahn and Cahn-Hilliard equations. *Discrete and Continuous Dynamical Systems*, 7(1):127–146, 2001.

[23] C. P. Grant. Superabundance of stationary solutions for the discrete Allen-Cahn equation. *Dynamics of Continuous, Discrete & Impulsive Systems. Series B. Applications & Algorithms*, 8(1):71–91, 2001.

[24] C. P. Grant and E. S. Van Vleck. Slowly-migrating transition layers for the discrete Allen-Cahn and Cahn-Hilliard equations. *Nonlinearity*, 8:861–876, 1995.

[25] M. Grinfeld and A. Novick-Cohen. Counting stationary solutions of the Cahn-Hilliard equation by transversality arguments. *Proceedings of the Royal Society of Edinburgh*, 125A:351–370, 1995.

[26] J. K. Hale. Numerical dynamics. *Contemporary Mathematics*, 172:1–30, 1994.

[27] I. Johnson, E. Sander, and T. Wanner. Branch interactions and long-term dynamics for the diblock copolymer model in one dimension. *Discrete and Continuous Dynamical Systems. Series A*, 33(8):3671–3705, 2013.

[28] T. Kato. *Perturbation Theory for Linear Operators*. Springer-Verlag, Berlin – Heidelberg – New York, 1984.

[29] S. G. Krantz and H. R. Parks. *The Implicit Function Theorem*. Birkhäuser/Springer, New York, 2013.

[30] Y. A. Kuznetsov. *Elements of Applied Bifurcation Theory*. Springer-Verlag, New York, second edition, 1998.

[31] J.-P. Lessard, E. Sander, and T. Wanner. Rigorous continuation of bifurcation points in the diblock copolymer equation. *In preparation*, 2015.

[32] S. Maier-Paape and T. Wanner. Spinodal decomposition for the Cahn-Hilliard equation in higher dimensions. Part I: Probability and wavelength estimate. *Communications in Mathematical Physics*, 195(2):435–464, 1998.

[33] S. Maier-Paape and T. Wanner. Spinodal decomposition for the Cahn-Hilliard equation in higher dimensions: Nonlinear dynamics. *Archive for Rational Mechanics and Analysis*, 151(3):187–219, 2000.

[34] J. Mallet-Paret and S.-N. Chow. Pattern formation and spatial chaos in lattice dynamical systems. II. *IEEE Transactions on Circuits and Systems. I. Fundamental Theory and Applications*, 42(10):752–756, 1995.

[35] G. Moore and A. Spence. The calculation of turning points of nonlinear equations. *SIAM Journal on Numerical Analysis*, 17(4):567–576, 1980.

[36] R. E. Moore, R. B. Kearfott, and M. J. Cloud. *Introduction to Interval Analysis*. SIAM, Philadelphia, 2009.

[37] Y. Nishiura and I. Ohnishi. Some mathematical aspects of the micro-phase separation in diblock copolymers. *Physica D*, 84(1-2):31–39, 1995.

[38] T. Ohta and K. Kawasaki. Equilibrium morphology of block copolymer melts. *Macromolecules*, 19:2621–2632, 1986.

[39] M. Plum. Existence and enclosure results for continua of solutions of parameter-dependent nonlinear boundary value problems. *Journal of Computational and Applied Mathematics*, 60(1-2):187–200, 1995.

[40] M. Plum. Enclosures for two-point boundary value problems near bifurcation points. In *Scientific computing and validated numerics (Wuppertal, 1995)*, volume 90 of *Mathematical Research*, pages 265–279. Akademie Verlag, Berlin, 1996.

[41] M. Plum. Computer-assisted proofs for semilinear elliptic boundary value problems. *Japan Journal of Industrial and Applied Mathematics*, 26(2-3):419–442, 2009.

[42] A. G. Ramm. A simple proof of the Fredholm alternative and a characterization of the Fredholm operators. *American Mathematical Monthly*, 108(9):855–860, 2001.

[43] S. M. Rump. INTLAB - INTerval LABoratory. In T. Csendes, editor, *Developments in Reliable Computing*, pages 77–104. Kluwer Academic Publishers, Dordrecht, 1999. http://www.ti3.tuhh.de/rump/.

[44] S. M. Rump. Verification methods: rigorous results using floating-point arithmetic. *Acta Numerica*, 19:287–449, 2010.

[45] E. Sander and T. Wanner. Monte Carlo simulations for spinodal decomposition. *Journal of Statistical Physics*, 95(5–6):925–948, 1999.

[46] E. Sander and T. Wanner. Unexpectedly linear behavior for the Cahn-Hilliard equation. *SIAM Journal on Applied Mathematics*, 60(6):2182–2202, 2000.

[47] E. Sander and T. Wanner. Rigorous bifurcation methods and applications to lattice dynamical systems. *In preparation*, 2015.

[48] A. Spence and B. Werner. Nonsimple turning points and cusps. *IMA Journal of Numerical Analysis*, 2(4):413–427, 1982.

[49] T. Wanner. Maximum norms of random sums and transient pattern formation. *Transactions of the American Mathematical Society*, 356(6):2251–2279, 2004.

[50] T. Wanner. Computer-assisted equilibrium validation for the diblock copolymer model. *Discrete and Continuous Dynamical Systems, Series A*, 2015. To appear.

[51] T. Wanner. Topological analysis of the diblock copolymer equation. In K. Obuse, Y. Nishiura, and N. Yoshinaga, editors, *Mathematical Challenge to a New Phase of Materials Science*, Springer Proceedings in Mathematics & Statistics. Springer-Verlag, 2015. To appear.

[52] Y. Watanabe, K. Nagatou, M. Plum, and M. T. Nakao. Verified computations of eigenvalue exclosures for eigenvalue problems in Hilbert spaces. *SIAM Journal on Numerical Analysis*, 52(2):975–992, 2014.

[53] B. Werner and A. Spence. The computation of symmetry-breaking bifurcation points. *SIAM Journal on Numerical Analysis*, 21(2):388–399, 1984.

[54] E. Zeidler. *Nonlinear Functional Analysis and its Applications. I: Fixed-Point Theorems*. Springer-Verlag, New York – Berlin – Heidelberg, 1986.