SIAM J. Applied Dynamical Systems
Vol. 15, No. 3, pp. 1690–1733

Validated Saddle-Node Bifurcations and Applications to Lattice Dynamical Systems*

Evelyn Sander† and Thomas Wanner†

Abstract. The use of rigorous verification methods is a powerful tool which permits progress in the analysis of dynamical processes that is not possible using purely analytical techniques. In this paper we develop a set of tools for branch validation, which allows for the rigorous verification of branch behavior, bifurcation, and solution index on branches generated through a saddle-node bifurcation. While the presented methodology can be applied in a variety of settings, we illustrate the use of these tools in the context of materials science. In particular, lattice models have been proposed as a more realistic reflection of the behavior of materials than traditional continuum models. For example, unlike their continuum counterparts, lattice models can account for phenomena such as pinning, and a significant body of work has been developed to study traveling waves. However, in a variety of other contexts such as bifurcation theory, questions about lattice dynamical systems are significantly harder to answer than those for continuum models. In the present paper, we show that computer-assisted proof techniques can be used to answer some of these questions. We apply these tools to the discrete Allen–Cahn equation, giving us results on the existence of branches of mosaic solutions and their robustness as it relates to grain size. We also demonstrate that there are situations in which classical continuation methods can fail to identify the correct branching behavior.

Key words. lattice dynamical system, bifurcation diagram, saddle-node bifurcations, computer-assisted proof, constructive implicit function theorem

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

DOI. 10.1137/16M1061011

1. Introduction.

In the study of high-dimensional systems of ordinary differential equations varying with a parameter, it is often the case that numerical continuation methods yield interesting results, but rigorously proving the observed behavior is intractable. Much recent work therefore seeks to address this problem through the use of rigorous computational methods, which provide computer-assisted proofs of dynamical questions through a combination of fixed point arguments and interval arithmetic; see, for example, [2, 10, 13, 25, 26, 27, 29, 33, 34] and the references therein. We develop a method using a constructive version of the implicit function theorem and use it to develop rigorous validation methods for proving the existence of branches of equilibrium solutions containing a saddle-node bifurcation, the location of the bifurcation point, and the index of the solutions on the branch.

This paper fits into a rich and thriving literature, stretching back for decades, of rigorous

*Received by the editors February 11, 2015; accepted for publication (in revised form) June 20, 2016; published electronically September 13, 2016.
http://www.siam.org/journals/siads/15-3/M106101.html
Funding: This work was partially supported by NSF grant DMS-1407087. The second author was also partially supported by NSF grant DMS-1114923.
†Department of Mathematical Sciences, George Mason University, Fairfax, VA 22030 (esander@gmu.edu, twanner@gmu.edu).

methods for validation of dynamical structures. Our approach is closely related to previous work with some key differences. We orient our development around the most fundamental tool in bifurcation theory, the implicit function theorem [19]. This result forms the basis of most bifurcation theory results, yet in its original formulation it is unsuitable for computer-assisted proofs. On the one hand, the classical implicit function theorem requires knowledge of an exact solution of a parameter-dependent nonlinear algebraic problem, which is usually not readily available in concrete nonlinear situations. On the other hand, it only establishes the existence of a small uniquely determined branch of solutions as a function of the parameter, without providing detailed information on the exact size of this branch segment. To overcome these issues, we present a *constructive implicit function theorem* in Theorem 1 which gives quantitative size information on the guaranteed solution branch and which only requires an approximative solution of the underlying nonlinear problem as input. A constructive implicit function theorem was formulated in Chierchia [5] for the Banach space of continuous functions. Our version of the result is a direct extension of a result of Plum [27] and is also related to other works by the same author [25, 26]; a preliminary form of our current formulation first appeared in a recent paper by Wanner [33]. All of these results fall into the category of Newton–Kantorovitch theorems, which are also referred to as a posteriori implicit function theorems. See [3] for statements and a brief history of such theorems in the context of numerical bifurcation theory (without validation). Our methods make use of contraction mapping arguments such as was done in the classical book of Ortega and Rheinboldt [24]. In the context of validation methods, this type of approach can be found, for example, in the work of Yamamoto and coworkers [36, 37].

The present paper is devoted to a new approach to computer-assisted bifurcation diagram validation. While our approach is certainly related to the many works described above, it differs in some essential points, which we now highlight as follows:

- We provide general results which are valid in arbitrary Banach spaces, and thereby isolate four main hypotheses which have to be established for a computer-assisted proof. These assumptions do not depend on specific solution approximations such as the finite-element approach developed in [25, 26, 27].
- By separating out the main assumptions, different techniques can be used to establish each of them. In contrast, in the radii-polynomial approach developed in [10, 13] all ingredients are incorporated into one statement involving interconnected estimates, and this can make it harder to accommodate modifications in the problem setup. In addition, two of our four main assumptions consist of Lipschitz-type nonlinearity estimates. These estimates can be derived using standard mean value theorem type results, i.e., in a coordinate-free and high-level setting. For example, in the context of partial differential equations, standard Sobolev embedding theorems can be employed. This also implies that our approach can deal equally easily with any kind of nonlinearity; i.e., there is no special advantage to considering polynomial nonlinearities.
- In the context of branch segment verification, our approach establishes a validated branch segment based on one solution approximation and a branch direction. It is not necessary to perform estimates along a parameterized curve of solution approximations as in [13]. Moreover, we obtain an explicit criterion for the appropriateness of the validation direction, as well as explicit bounds on the size of the validated branch. In

this way, our method can easily be used in an iterative way to validate large branch pieces, all starting from one solution approximation, and it is not necessary to find a sequence of numerical approximations along the branch.

- We provide two results which allow for branch validation away from and in a neighborhood of saddle-node bifurcation points, respectively. In this way, all branches which contain only bifurcation points of this type can be validated. Our method does not require the implementation of more involved approaches, such as pseudo-arclength continuation. We emphasize that our validation of saddle-node bifurcation points uses the exact same framework of the constructive implicit function theorem as the continuation method. In contrast, the methods in [10, 13] only consider continuation of branches away from bifurcations and do not address bifurcation points at all. We are not the first to validate bifurcation points. Reference [39] validates bifurcation points using dynamical systems/Conley index-type arguments. The method relies on having coordinate-specific information. Reference [2] is much more closely related to the ideas of the current paper in that it contains validations of saddle-node and pitchfork bifurcations via contraction mapping principle based arguments. However, the method requires the use of a Fourier series, and it is restricted to solutions with analyticity. The framework developed in this paper allows us to work with solutions with much lower regularity.
- Our modular approach allows us to identify exactly which steps provide the bottlenecks which have to be approached with careful analysis, and which steps give sufficiently good results to allow the computer's interval arithmetic and automatic differentiation algorithms to take care of things. In this way, we are able to make changes in our setup without performing a new set of computations. For example, in our application, we have done the majority of our calculations in the case of dimension $n = 10$, but we are able to change to $n = 100$ with just a few adjustments to the code. In contrast to [10, 13], we are able to compute continuous branches of equilibria through our framework rather than computing individual points on branches and using a separate argument to justify that these points fit together in a continuous branch. We are able to compute saddle-node bifurcation points using the same framework as we use to compute nonbifurcating branches. Furthermore, we have built this modular framework so that it can be easily adapted to apply to a wide variety of dynamical contexts, including different types of bifurcations, infinite-dimensional settings, and a variety of different types of Banach spaces. Depending on the specific context, the bottlenecks will change. By thinking through each hypothesis separately, our framework allows for the discovery of which analytical estimates need to be made sharp, and which will not significantly affect the results if they are done automatically.
- We have chosen to restrict the current paper to a finite-dimensional application so as to avoid some of the technical parts of implementing the method in an infinite-dimensional context. A preliminary version of this method has been applied in an infinite-dimensional context in [33, 34], in a case where the norm embeddings worked out nicely. However, making this method more general for infinite-dimensional problems requires constructive versions of Sobolev embedding theorems. This is a topic of ongoing research.

Based on these observations, we believe that the approach developed in the current paper can serve as the backbone of a more modular approach to bifurcation diagram validation.

Throughout this paper, our approach is illustrated in the context of a specific lattice dynamical system. More precisely, we apply our method to the spatially discrete Allen–Cahn equation, a system for the study of phase boundaries in crystalline solids. We are able to validate branches of equilibrium solutions containing a saddle-node bifurcation, find the index of solutions in the branches, and establish the location of bifurcation points. We use our method to validate a statement on the robustness of solutions as a function of their minimum grain size. We would like to point out that in our study of the discrete Allen–Cahn model, we consider examples in high dimensions, as well as different nonlinearities, to illustrate the relative simplicity of our modular approach.

The paper proceeds as follows. In section 2, we present the constructive implicit function theorem, an adaptation of the classical implicit function theorem which is ideally suited for validation methods. We show how the constructive implicit function theorem is used to validate branches of solutions with saddle-node bifurcations. We would like to point out that the results of this section are formulated in general Banach spaces and can easily be applied in the infinite-dimensional context as well. See, for example, [33, 34]. For the present paper, we decided to demonstrate the applicability of the developed methods in the finite-dimensional setting, since in this way we do not have to deal with the additional issues arising in the context of partial differential equations. However, we are planning to extend this work in the future; see also [32]. Following the setup of this framework, in section 3 we apply our abstract results on branches of solutions to the discrete Allen–Cahn equation with dimension $n = 10$. Our results apply to solutions known as mosaic solutions for which the resulting branch has a saddle-node bifurcation. In particular, for the stable mosaic solutions, we validate the existence and uniqueness of the entire branch (unbounded in λ) for almost 94% of the $1024 = 2^{10}$ mosaic solutions. We validate branches of solutions, their bifurcations points, and the number of unstable eigenvalues after bifurcation. This allows us to establish a relationship between the geometry of a stable mosaic solution and the location of its bifurcation. We demonstrate the flexibility of our method by validating, without much extra work, solution branches for a different nonlinearity for $n = 100$. In particular, we give precise implementation details on how these abstract methods are applied to the lattice Allen–Cahn equation in order to validate the branch, bifurcation point, and index of solutions.

2. Computer-assisted bifurcation diagram validation.

We now turn our attention to presenting the theoretical underpinnings of our numerical validation approach. In the interest of future applications, all of these results are presented in the setting of general Banach spaces. We begin by proving a constructive implicit function theorem, which applies when the usual assumptions of the implicit function theorem are satisfied only approximately. This result is then used to establish the existence of solution branches of parameter-dependent problems. After this, we show how the constructive implicit function theorem can be used to enclose the solution branch in a neighborhood of saddle-node bifurcation points.

2.1. A constructive implicit function theorem.

In classical bifurcation theory, the implicit function theorem is arguably the most versatile and important tool. While it provides a necessary condition for bifurcation, it also lies at the heart of a number of results which

establish certain bifurcation points. See, for example, the extensive discussions in [6, 38]. Despite its importance, the implicit function theorem is an inherently theoretical result, since its application requires the a priori knowledge of an exact solution of a nonlinear problem. In most applied problems such exact solutions are out of reach, and the best one can hope for are numerically computed approximations.

In the current section, we demonstrate that in many cases it suffices to know a solution approximation in order to still deduce a result which captures the essence of the implicit function theorem. The result is formulated for parameter-dependent problems of the form

$$\mathcal{G}(\alpha, x) = 0 \,, \tag{1}$$

where $\mathcal{G} : \mathcal{P} \times \mathcal{X} \to \mathcal{Y}$ is a Fréchet differentiable nonlinear operator 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. Our approach is based on the work of Plum [27] and the recent paper [33]; we extend this work to provide a more applicable result. Iteratively applied, our result leads to a versatile and fast rigorous continuation method.

We now list our principal hypotheses under which we are able to prove an implicit function theorem. The 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 (1) if all that is available is an approximation.

(H1) Assume that we are given a pair $(\alpha^*, x^*) \in \mathcal{P} \times \mathcal{X}$ which is an approximate solution of the nonlinear problem (1). More precisely, 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

$$\|\mathcal{G}(\alpha^*, x^*)\|_{\mathcal{Y}} \leq \varrho \,. \tag{2}$$

(H2) We assume that the operator $D_x\mathcal{G}(\alpha^*, x^*)$ is invertible and not very close to being singular. That is, 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

$$\left\|D_x\mathcal{G}(\alpha^*, x^*)^{-1}\right\|_{\mathcal{L}(\mathcal{Y},\mathcal{X})} \leq K \,, \tag{3}$$

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

(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

$$\|D_x\mathcal{G}(\alpha, x) - D_x\mathcal{G}(\alpha^*, x^*)\|_{\mathcal{L}(\mathcal{X},\mathcal{Y})} \leq L_1 \|x - x^*\|_{\mathcal{X}} + L_2 \|\alpha - \alpha^*\|_{\mathcal{P}} \,. \tag{4}$$

(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

$$\|D_\alpha\mathcal{G}(\alpha, x^*)\|_{\mathcal{L}(\mathcal{P},\mathcal{Y})} \leq L_3 + L_4 \|\alpha - \alpha^*\|_{\mathcal{P}} \,, \tag{5}$$

where ℓ_α is the constant that was chosen in (H3).

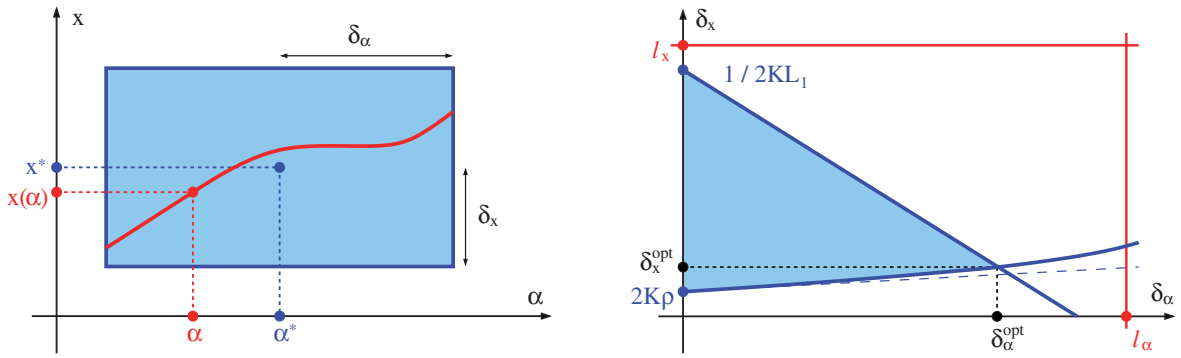

Figure 1. *Visualization of the constructive implicit function theorem. The left image illustrates the general statement of the theorem; i.e., in a window around the approximative solution (α^*, x^*) all solutions of $\mathcal{G}(\alpha, x) = 0$ lie on the curve $\alpha \mapsto x(\alpha)$. The size of the window is determined by the pair $(\delta_\alpha, \delta_x)$, and the shape of the admissible region from which the pair can be chosen is illustrated in the right image. The point $(\delta_\alpha^{\text{opt}}, \delta_x^{\text{opt}})$ determines the largest parameter jump which still satisfies the theorem. The dashed blue line shows the slope of the lower blue curve at $\delta_\alpha = 0$.*

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^*). However, 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. Furthermore, in specific applications one has to be able to determine explicit values for these constants. This is accomplished through a combination of rigorous estimates and interval computations.

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 1 (constructive 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)–(H4). *Finally, suppose that*

$$4K^2\varrho L_1 < 1 \qquad \text{and} \qquad 2K\varrho < \ell_x \,. \tag{6}$$

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

$$2KL_1\delta_x + 2KL_2\delta_\alpha \le 1 \qquad \text{and} \qquad 2K\varrho + 2KL_3\delta_\alpha + 2KL_4\delta_\alpha^2 \le \delta_x \,, \tag{7}$$

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

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

then all solutions of the nonlinear problem $\mathcal{G}(\alpha, x) = 0$ in the set $\mathcal{B}_{\delta_\alpha}^{\mathcal{P}} \times \mathcal{B}_{\delta_x}^{\mathcal{X}}$ lie on the graph of the function $\alpha \mapsto x(\alpha)$. This statement is visualized in the left image of Figure 1, and the right panel in this figure depicts the admissible region of all pairs $(\delta_\alpha, \delta_x)$ for which the result holds. In addition, the following two statements are satisfied:

- *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* $\mathcal{L}(\mathcal{Y}, \mathcal{X})$.
- *If the mapping* $\mathcal{G} : \mathcal{P} \times \mathcal{X} \to \mathcal{Y}$ *is* k*-times continuously Fréchet differentiable, then so is the solution function* $\alpha \mapsto x(\alpha)$.

The constructive implicit function theorem will be crucial for all of the following results in this paper, and therefore prior to giving the proof, we spend a few paragraphs discussing its assumptions and statements with an eye towards practical and efficient implementation. The central assumptions are the two inequalities $4K^2\varrho L_1 < 1$ and $2K\varrho < \ell_x$ stated in (6), which in some sense quantify that the approximation (α^*, x^*) is indeed close to an actual solution of (1). Of these estimates, the second is usually minor and automatically satisfied. The first estimate, however, relates the size of the residual ϱ of the approximation, the bound K on the operator norm of the inverse of the Fréchet derivative at the approximation, as well as the Lipschitz constant L_1 for the Fréchet derivative at the approximation. Since the latter two constants will have values which are essentially beyond our control, assumption (6) states that for small enough values of the residual, the implicit function theorem applies.

Theorem 1 guarantees that under assumption (6) one can always find bounds $\delta_x > 0$ and $\delta_\alpha \geq 0$ which satisfy the inequalities given in (7); see also the right panel in Figure 1. For the special case $\delta_\alpha = 0$ the theorem establishes the existence of a unique solution at the fixed parameter value α^* which is contained in the neighborhood $\mathcal{B}_{\delta_x}$ of x^* for radius values δ_x satisfying

$$2K\varrho \;<\; \delta_x \;<\; \min\left\{\frac{1}{2KL_1}\,,\;\; \ell_x\right\}\,. \tag{8}$$

In other words, the lower bound in (8) is a measure of the approximation error of (α^*, x^*), while the upper bound captures the level of isolation of the actual solution.

If, on the one hand, we pick an admissible pair $(\delta_\alpha, \delta_x)$ with $\delta_\alpha > 0$, then we can establish the existence of a manifold of solutions over an α-ball of radius δ_α. Not surprisingly, this leads to a reduced range of possible values of δ_x. The largest possible α-range for the solution graph occurs at the intersection point of the blue parabola and the blue solid line in the right image of Figure 1, which can be computed explicitly as

$$\delta_\alpha^{\mathrm{opt}} = \frac{2(C - 2K\varrho)}{D + 2KL_3 + \sqrt{(D + 2KL_3)^2 + 8KL_4(C - 2K\varrho)}} \quad \text{and} \quad \delta_x^{\mathrm{opt}} = C - D\delta_\alpha^{\mathrm{opt}}\,, \tag{9}$$

with the constants C and D given by

$$C = \frac{1}{2KL_1} \quad \text{and} \quad D = \frac{L_2}{L_1}\,.$$

For this admissible pair, the possible δ_x-range shrinks to the singleton δ_x^{opt}, and the parameter range over which solutions can be established has the maximal radius $\delta_\alpha^{\mathrm{opt}}$. Notice that of the constants we have control over, the location of this optimal pair is affected most significantly by the constant L_3 in hypothesis (H4). This constant influences the slope of the dashed blue line in the right image of Figure 1, which is the tangent line to the blue parabola at $\delta_\alpha = 0$. If

this slope is large, the optimal pair moves upwards, thereby decreasing the optimal α-range. If, on the other hand, L_3 is close to zero, the tangent is basically horizontal, and the α-range is larger. This will be crucial for the results of the following sections. Finally, note that the theorem's assumptions require sufficiently large values for ℓ_x and ℓ_α. In practice, the required values occur without significant extra hypotheses.

With this preliminary discussion completed, we now proceed with the proof of the constructive implicit function theorem.

Proof. The proof follows along the lines of [33, Theorem 2.2], with some slight but crucial extensions. Consider the parameter-dependent Newton-like operator defined by

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

and let $\delta_\alpha \geq 0$, $\delta_x > 0$, $\mathcal{B}^{\mathcal{X}}_{\delta_x}$, and $\mathcal{B}^{\mathcal{P}}_{\delta_\alpha}$ be as in the formulation of the theorem. Then hypotheses (H2) and (H3) imply for all $\alpha \in \mathcal{B}^{\mathcal{P}}_{\delta_\alpha}$ and arbitrary $x \in \mathcal{B}^{\mathcal{X}}_{\delta_x}$ the estimate

$$\begin{aligned}
\|D_x\mathcal{T}(\alpha, x)\|_{\mathcal{L}(\mathcal{X},\mathcal{X})} &= \left\|D_x\mathcal{G}(\alpha^*, x^*)^{-1}\left(D_x\mathcal{G}(\alpha^*, x^*) - D_x\mathcal{G}(\alpha, x)\right)\right\|_{\mathcal{L}(\mathcal{X},\mathcal{X})} \\
&\leq K\,\|D_x\mathcal{G}(\alpha, x) - D_x\mathcal{G}(\alpha^*, x^*)\|_{\mathcal{L}(\mathcal{X},\mathcal{Y})} \\
&\leq KL_1\,\|x - x^*\|_{\mathcal{X}} + KL_2\,\|\alpha - \alpha^*\|_{\mathcal{P}} \\
&\leq KL_1\delta_x + KL_2\delta_\alpha \;\leq\; \frac{1}{2}\,,
\end{aligned}$$

where we have also used the first estimate in (7). Furthermore, an application of the mean value theorem [38, Theorem 4.A] yields

$$\|\mathcal{T}(\alpha, x) - \mathcal{T}(\alpha, \bar{x})\|_{\mathcal{X}} \leq \frac{1}{2}\,\|x - \bar{x}\|_{\mathcal{X}} \quad \text{for all} \quad \alpha \in \mathcal{B}^{\mathcal{P}}_{\delta_\alpha}\,,\;\; x, \bar{x} \in \mathcal{B}^{\mathcal{X}}_{\delta_x}\,. \tag{10}$$

Now let $\alpha \in \mathcal{B}^{\mathcal{P}}_{\delta_\alpha}$ be arbitrary. Then due to (H1), (H2), and (H4) we have

$$\begin{aligned}
\|\mathcal{T}(\alpha, x^*) - x^*\|_{\mathcal{X}} &\leq \left\|D_x\mathcal{G}(\alpha^*, x^*)^{-1}\right\|_{\mathcal{L}(\mathcal{Y},\mathcal{X})}\,\|\mathcal{G}(\alpha, x^*)\|_{\mathcal{Y}} \\
&\leq K\,\|\mathcal{G}(\alpha, x^*) - \mathcal{G}(\alpha^*, x^*)\|_{\mathcal{Y}} + K\,\|\mathcal{G}(\alpha^*, x^*)\|_{\mathcal{Y}} \\
&\leq K\,\|\alpha - \alpha^*\|_{\mathcal{P}} \sup_{\|\tilde{\alpha} - \alpha^*\|_{\mathcal{P}} \leq \delta_\alpha} \|D_\alpha\mathcal{G}(\tilde{\alpha}, x^*)\|_{\mathcal{L}(\mathcal{P},\mathcal{Y})} + K\varrho \\
&\leq K\delta_\alpha\,(L_3 + L_4\delta_\alpha) + K\varrho\,,
\end{aligned}$$

and in combination with (10) one obtains for all $(\alpha, x) \in \mathcal{B}^{\mathcal{P}}_{\delta_\alpha} \times \mathcal{B}^{\mathcal{X}}_{\delta_x}$ the estimate

$$\begin{aligned}
\|\mathcal{T}(\alpha, x) - x^*\|_{\mathcal{X}} &\leq \|\mathcal{T}(\alpha, x) - \mathcal{T}(\alpha, x^*)\|_{\mathcal{X}} + \|\mathcal{T}(\alpha, x^*) - x^*\|_{\mathcal{X}} \\
&\leq \frac{1}{2}\,\|x - x^*\|_{\mathcal{X}} + K\delta_\alpha\,(L_3 + L_4\delta_\alpha) + K\varrho \\
&\leq \frac{\delta_x}{2} + KL_3\delta_\alpha + KL_4\delta_\alpha^2 + K\varrho\,,
\end{aligned}$$

and the second estimate in (7) finally implies

$$\|\mathcal{T}(\alpha, x) - x^*\|_{\mathcal{X}} \leq \delta_x \quad \text{for all} \quad \alpha \in \mathcal{B}^{\mathcal{P}}_{\delta_\alpha}\ , \ \ x \in \mathcal{B}^{\mathcal{X}}_{\delta_x}\ .$$

Together with (10) this establishes $\mathcal{T} : \mathcal{B}^{\mathcal{P}}_{\delta_\alpha} \times \mathcal{B}^{\mathcal{X}}_{\delta_x} \to \mathcal{B}^{\mathcal{X}}_{\delta_x}$ as a uniform contraction, and the statement of the theorem follows from the uniform contraction mapping principle.

In order to complete the proof of the theorem we need only show that there actually exist admissible pairs $(\delta_\alpha, \delta_x)$ with $0 \leq \delta_\alpha \leq \ell_\alpha$ and $0 < \delta_x \leq \ell_x$, which also satisfy both estimates in (7). For this, we refer the reader to the right image of Figure 1. One can easily see that the region bounded by the inequalities in (7) in combination with $\delta_\alpha \geq 0$ is qualitatively of the form depicted in blue in the image. This is due to the fact that the first inequality in (6) implies $1/(2KL_1) > 2K\varrho$; i.e., we obtain a nontrivial interval of admissible δ_x values for $\delta_\alpha = 0$. Moreover, the additional constraint $\delta_x \leq \ell_x$ cannot destroy this, since the second inequality in (6) gives $2K\varrho < \ell_x$. Finally, since both inequalities in (6) are strict, this implies that one can also find admissible pairs $(\delta_\alpha, \delta_x)$ with $\delta_\alpha > 0$.

It only remains to verify the last two statements of the theorem. For the first of these, let $\alpha \in \mathcal{B}^{\mathcal{P}}_{\delta_\alpha}$ and $x \in \mathcal{B}^{\mathcal{X}}_{\delta_x}$ be arbitrary. It was shown earlier in the proof that then

$$\|D_x\mathcal{T}(\alpha, x)\|_{\mathcal{L}(\mathcal{X},\mathcal{X})} = \big\|I - D_x\mathcal{G}(\alpha^*, x^*)^{-1} D_x\mathcal{G}(\alpha, x)\big\|_{\mathcal{L}(\mathcal{X},\mathcal{X})} \leq \frac{1}{2} < 1\ ,$$

and a standard Neumann series argument implies that $D_x\mathcal{G}(\alpha^*, x^*)^{-1} D_x\mathcal{G}(\alpha, x) \in \mathcal{L}(\mathcal{X}, \mathcal{X})$ is one-to-one and onto, with continuous inverse; see also Lemma 4 below. Together with (H2), this completes the proof of the first statement. For the second statement, note that if $\mathcal{G}$ is k-times continuously Fréchet differentiable, then so is the uniform contraction $\mathcal{T}$. The statement now follows from the uniform contraction principle [6, Theorem 2.2]. This completes the proof of the theorem. ∎

For the sake of later reference, we close this section with the following remarks and lemma.

Remark 2 (additional window constraints). In the formulation of Theorem 1 we concentrated on the simplest possible external restrictions for the validation constants δ_α and δ_x, namely $\delta_\alpha \leq \ell_\alpha$ and $\delta_x \leq \ell_x$. These conditions are a reflection of the fact that the estimates assumed in (H3) and (H4) usually only hold locally around the point (α^*, x^*), and we chose to concentrate on a rectangular window with side lengths ℓ_α and ℓ_x, as shown in the left panel of Figure 1. In this simple form, the constructive implicit function theorem is (theoretically) ideally suited for a rigorous version of parameter continuation, in which the parameter is iteratively varied by a fixed value, followed by finding an approximate solution at the new parameter value. However, from a numerical standpoint it is known to be more flexible and more efficient to be able to use arclength continuation, in which both the parameter and the function are iteratively varied. For example, this allows for a continuation method that is able to continue at a saddle-node bifurcation point. In order to develop a rigorous method which has these same speedups and flexibilities, in some of the applications below, it will be more convenient to consider windows which are not parallel to the α- and x-axes. In these cases, hypotheses (H3) and (H4) will be satisfied as long as we have $\|x - x^*\|_{\mathcal{X}} \leq \delta_x$ and $\|\alpha - \alpha^*\|_{\mathcal{P}} \leq \delta_x$, where the thresholds δ_α and δ_x satisfy (one or more) additional estimates

of the form

$$C_1\delta_x + C_2\delta_\alpha \le C_3 \tag{11}$$

for certain positive constants C_1, C_2, and C_3. One can easily see that if we add the constraints (11), the proof of Theorem 1 remains unaffected. All it does is potentially restrict the set of admissible pairs $(\delta_\alpha, \delta_x)$, which can now be found by intersecting the blue region in the right panel of Figure 1 with the half-spaces defined in (11).

Remark 3 (parameter-independent constructive implicit function theorem). In certain situations, it is desirable to apply Theorem 1 in a parameter-independent version; i.e., one is only interested in establishing a solution for fixed α^*. By choosing $\delta_\alpha = 0$, it can readily be seen from the proof of the theorem that in this case hypothesis (H4) is not necessary, and that (H3) only has to be established with $\ell_x > 0$ and $\ell_\alpha = 0$. In other words, only Lipschitz constant L_1 has to be determined.

In order to verify hypothesis (H2), we need to be able to estimate the norm of the inverse of a linear operator. The following lemma gives a useful estimate of this norm. Its proof is based on a standard Neuman series argument, but the result is quite general and can be applied in a variety of settings; see also [23, Proposition 2.4.2] and [30, Lemma 2.1].

Lemma 4 (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*

$$\|I - \mathcal{B}\mathcal{A}\|_{\mathcal{L}(\mathcal{X},\mathcal{X})} \le \varrho_1 < 1 \quad \text{as well as} \quad \|\mathcal{B}\|_{\mathcal{L}(\mathcal{Y},\mathcal{X})} \le \varrho_2 \tag{12}$$

for two constants $0 \le \varrho_1 < 1$ 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

$$\left\|\mathcal{A}^{-1}\right\|_{\mathcal{L}(\mathcal{Y},\mathcal{X})} \le \frac{\varrho_2}{1-\varrho_1}\,. \tag{13}$$

Proof. According to (12), the Neumann series $\mathcal{C} = \sum_{k=0}^{\infty}(I - \mathcal{B}\mathcal{A})^k$ converges in $\mathcal{L}(\mathcal{X}, \mathcal{X})$, and the operator $\mathcal{C}$ satisfies $\|\mathcal{C}\|_{\mathcal{L}(\mathcal{X},\mathcal{X})} \le 1/(1-\varrho_1)$; see [18, sections I.4.4 and III.3.2]. This implies the identities

$$(\mathcal{B}\mathcal{A})\,\mathcal{C} = (I - (I - \mathcal{B}\mathcal{A}))\,\mathcal{C} = I = \mathcal{C}\,(I - (I - \mathcal{B}\mathcal{A})) = \mathcal{C}\,(\mathcal{B}\mathcal{A})\ ,$$

which show that the operators $\mathcal{C}$ and $\mathcal{B}\mathcal{A}$ are one-to-one and onto, as well as $(\mathcal{B}\mathcal{A})^{-1} = \mathcal{C}$. But then $\mathcal{A} = \mathcal{B}^{-1}\mathcal{C}^{-1}$ is one-to-one and onto with $\mathcal{A}^{-1} = \mathcal{C}\mathcal{B}$, and together with (12) this finally implies

$$\left\|\mathcal{A}^{-1}\right\|_{\mathcal{L}(\mathcal{Y},\mathcal{X})} = \|\mathcal{C}\mathcal{B}\|_{\mathcal{L}(\mathcal{Y},\mathcal{X})} \le \|\mathcal{C}\|_{\mathcal{L}(\mathcal{X},\mathcal{X})}\,\|\mathcal{B}\|_{\mathcal{L}(\mathcal{Y},\mathcal{X})} \le \frac{\varrho_2}{1-\varrho_1}\,.$$

This completes the proof of the lemma. ∎

2.2. Validation of branches of equilibria. Beginning with this section we apply the constructive implicit function theorem in a variety of contexts, whose combination allows us to establish the results on the lattice Allen–Cahn equation in section 3. As Theorem 1 is now considered a tool which will be applied in several different ways, we change our notation. From now on, we will study parameter-dependent problems of the form

$$\mathcal{F}(\lambda, u) = 0\,, \tag{14}$$

where $\mathcal{F} : \mathbb{R} \times \mathcal{U} \to \mathcal{V}$ is a Fréchet differentiable parameter-dependent 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 (14) in the neighborhood of a pair $(\lambda^*, u^*) \in \mathbb{R} \times \mathcal{U}$ which is an approximate solution.

The goal of the present section is to establish the existence of solution branch pieces for the nonlinear problem (14). While this question has already been addressed by the constructive implicit function theorem, this result has limited direct applicability in its stated simple form. To see this, recall that Theorem 1 guarantees the existence of positive constants δ_α and δ_x such that 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, the solutions $x(\alpha)$ are constructed in balls of radius δ_x and with center x^*, and the latter is kept fixed even as the parameter α is varied. In practice, this approach will lead to severe restrictions on the size of the α-ball on which Theorem 1 applies. This is of course due to the fact that the graph of the solution map $\alpha \mapsto x(\alpha)$ usually is not parallel to the x-axis; i.e., depending on the tangent directions of the map, restrictions on δ_x will affect and severely limit the size of the radius δ_α. In particular, in the constructive implicit function theorem in the case $\mathcal{P} = \mathbb{R}$, the value of L_3 is basically given by the norm of $D_\lambda \mathcal{F}(\lambda^*, u^*)$, which could be quite large. As we pointed out earlier, the value of L_3 affects the slope of the dashed blue line in the right panel of Figure 1, and thus would lead to extremely small values of $\delta_\alpha^{\text{opt}}$.

This shortcoming of Theorem 1 is addressed by the following theorem for the special case $\mathcal{P} = \mathbb{R}$, i.e., for branches of solutions. In the new result, it is not enough to specify just the approximative solution (λ^*, u^*), but one also has to provide a direction v^* which (ideally) approximates the actual solution curve. If one chooses v^* close to the tangent direction of the solution branch and focusing on a parallelogram of possible equilibria, as shown in the left panel of Figure 2, then with respect to the direction of v^*, the constant L_3 will be extremely close to zero. This implies that the same will be true for the slope of the dashed blue line in the right panel of Figure 2, which leads to considerably larger values of $\delta_\lambda^{\text{opt}}$. This comes at the price of an additional constraint, stated in (20), which is indicated in green in the right panel of Figure 2. In applications, this constraint does not usually impose any additional condition—and even if it does, one can easily determine the new value of the optimal λ-interval using Remark 1. Even though this leads to a slightly more complicated formulation of the theorem, the later numerical payoff will be significant.

While the method presented in this section bears a close resemblance to an arclength continuation method in that our rigorous predictor step is chosen in the approximate tangent direction to the solution curve, the method is not a true arclength continuation, since the rigorous corrector step does not use an extended system. Unlike the speedup for the standard numerical method, in practice we find that the computational time of our rigorous method

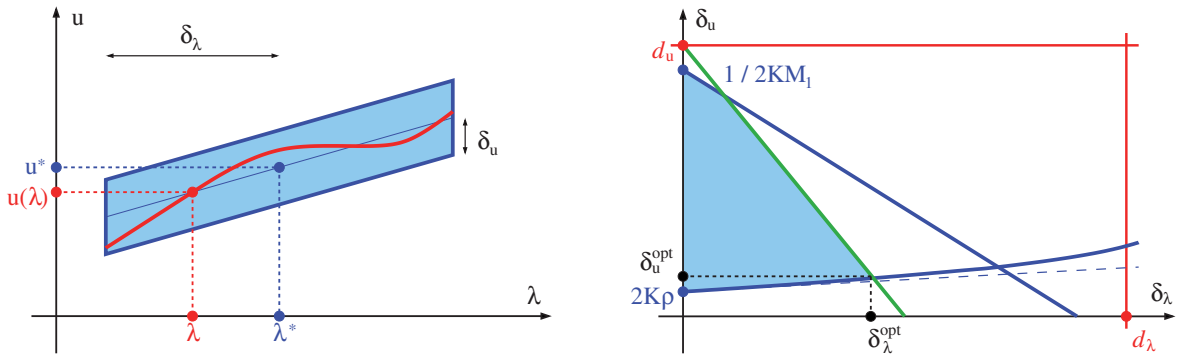

Figure 2. *Visualization of the regular branch validation theorem. The left image illustrates the general statement of the theorem; i.e., in a slanted window around the approximative solution (λ^*, u^*) all solutions of $\mathcal{F}(\lambda, u) = 0$ lie on the curve $\lambda \mapsto u(\lambda)$. The 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 image. Notice that if $\|v^*\|_{\mathcal{U}} \gg 0$, then the last constraint in (20), which is indicated in green, might further restrict the admissible region. In practice, this usually does not happen.*

is lengthened if we have to create rigorous estimates in a new coordinate system on each step. In the next section, we use the constructive implicit function theorem to give a rigorous continuation method in the region of a saddle-node bifurcation. Combining the methods in this and the next section then provides a rigorous validation method for the continuation of solution branches containing a fold.

In order to carry out the above procedure, we will use the constructive implicit function theorem as a tool and apply it to problem (14) after a suitable transformation. For this, we need to make the following assumption on the function $\mathcal{F}$:

(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

$$\|D_u\mathcal{F}(\lambda, u) - D_u\mathcal{F}(\lambda^*, u^*)\|_{\mathcal{L}(\mathcal{U},\mathcal{V})} \leq M_1 \|u - u^*\|_{\mathcal{U}} + M_2 |\lambda - \lambda^*| \ , \tag{15}$$

$$\|D_\lambda\mathcal{F}(\lambda, u) - D_\lambda\mathcal{F}(\lambda^*, u^*)\|_{\mathcal{V}} \leq M_3 \|u - u^*\|_{\mathcal{U}} + M_4 |\lambda - \lambda^*| \ , \tag{16}$$

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 $\mathcal{L}(\mathbb{R}, \mathcal{V})$.

Assumption (A) is a quantified Lipschitz continuity requirement for the Fréchet derivatives of $\mathcal{F}$. It will be used in Theorem 5 to guarantee (H3) from the first equation in (A), and (H4) from the second equation. In this way, we will be able to derive the constants L_k in the constructive implicit function theorem directly from the constants in (A).

Theorem 5 (regular branch segment validation). *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*

$$\|\mathcal{F}(\lambda^*, u^*)\|_{\mathcal{V}} \leq \varrho \qquad \text{and} \qquad \left\|D_u\mathcal{F}(\lambda^*, u^*)^{-1}\right\|_{\mathcal{L}(\mathcal{V},\mathcal{U})} \leq K \tag{17}$$

for some positive constants ϱ and K, which will be used to show (H1) *and* (H2)*, and let $v^* \in \mathcal{U}$ be given with*

$$\|D_\lambda \mathcal{F}(\lambda^*, u^*) + D_u \mathcal{F}(\lambda^*, u^*)[v^*]\|_{\mathcal{V}} \leq \sigma \tag{18}$$

for some constant $\sigma \geq 0$, which will indicate the slant of the box containing the solution branch. Finally, assume that the estimates

$$4K^2 \varrho M_1 < 1 \qquad \text{and} \qquad 2K\varrho < d_u \tag{19}$$

hold, which will be used to show (6)*. Then there exist pairs of constants $(\delta_\lambda, \delta_u)$ which satisfy*

$$0 < \delta_\lambda \leq d_\lambda \,, \qquad 0 < \delta_u \leq d_u \,, \qquad \text{and} \qquad \delta_\lambda \|v^*\|_{\mathcal{U}} + \delta_u \leq d_u \,, \tag{20}$$

and satisfy the two inequalities

$$2K M_1 \delta_u + 2K \left(M_1 \|v^*\|_{\mathcal{U}} + M_2\right) \delta_\lambda \leq 1 \tag{21}$$

and

$$2K\varrho + 2K\sigma\delta_\lambda + 2K \left(M_1 \|v^*\|_{\mathcal{U}}^2 + (M_2 + M_3) \|v^*\|_{\mathcal{U}} + M_4\right) \delta_\lambda^2 \leq \delta_u \,, \tag{22}$$

which will be used to show (7)*, and for each such pair the following holds. For every parameter value $\lambda \in \mathbb{R}$ with $|\lambda - \lambda^*| \leq \delta_\lambda$ there exists a uniquely determined element $u(\lambda) \in \mathcal{U}$ which satisfies $\|u(\lambda) - (u^* + (\lambda - \lambda^*)v^*)\|_{\mathcal{U}} \leq \delta_u$, and for which the nonlinear equation $\mathcal{F}(\lambda, u(\lambda)) = 0$ holds. 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 $\lambda \mapsto u(\lambda)$. In addition, if the mapping $\mathcal{F} : \mathbb{R} \times \mathcal{U} \to \mathcal{V}$ is k-times continuously Fréchet differentiable, then so is the solution function $\lambda \mapsto u(\lambda)$. This statement is illustrated in the left image of Figure 2, while the right image depicts the admissible region of all pairs $(\delta_\lambda, \delta_u)$ for which the result holds.

Prior to giving the proof of this theorem, we note that if we choose $v^* = 0$, then Theorem 5 reduces to the constructive implicit function theorem. However, by choosing $v^* \neq 0$, we can gain improvements on the optimal δ_α, but at the expense of an additional constraint determined by (18). As mentioned earlier, this often does not lead to any actual restriction. When it does lead to an additional constraint, one can easily find it by using (9) with the constants $C = d_u$ and $D = \|v^*\|_{\mathcal{U}}$ to compute the intersection point with the parabolic curve given by (22), as long as the constants L_k are chosen as in Theorem 5. We now proceed with the proof of the theorem.

Proof. In order to prove the theorem, we reduce its more refined approach to the constructive implicit function theorem. For this, define $\mathcal{P} = \mathbb{R}$, $\mathcal{X} = \mathcal{U}$, and $\mathcal{Y} = \mathcal{V}$, and consider the nonlinear operator $\mathcal{G} : \mathcal{P} \times \mathcal{X} \to \mathcal{Y}$ given by

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

Then $\mathcal{G}$ is Fréchet differentiable with $D_x\mathcal{G}(\alpha,x) = D_u\mathcal{F}\left(\lambda^*+\alpha, u^*+\alpha v^*+x\right)$, and one can easily see that $D_\alpha\mathcal{G}(\alpha,x) = D_\lambda\mathcal{F}\left(\lambda^*+\alpha, u^*+\alpha v^*+x\right) + D_u\mathcal{F}\left(\lambda^*+\alpha, u^*+\alpha v^*+x\right)[v^*]$. These formulas imply that at the point $(\alpha^*, x^*) = (0,0)$ we have $\mathcal{G}(0,0) = \mathcal{F}(\lambda^*, u^*)$, as well as both

$$D_x\mathcal{G}(0,0) = D_u\mathcal{F}\left(\lambda^*, u^*\right) \quad \text{and} \quad D_\alpha\mathcal{G}(0,0) = D_\lambda\mathcal{F}\left(\lambda^*, u^*\right) + D_u\mathcal{F}\left(\lambda^*, u^*\right)[v^*]\,,$$

and therefore the estimates in (17) imply hypotheses (H1) and (H2). We now verify hypotheses (H3) and (H4). For this, suppose that the constants $0 < \delta_\lambda \le d_\lambda$ and $0 < \delta_u \le d_u$ satisfy (20), and let $\alpha \in \mathbb{R}$ be arbitrary with $|\alpha| \le \delta_\lambda$. Then $\|\alpha v^*\|_{\mathcal{U}} = |\alpha|\|v^*\|_{\mathcal{U}} \le d_u$ implies, together with (A), the estimate

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

Similarly, if we assume that α and x satisfy both $|\alpha| \le \delta_\lambda$ and $\|x\|_{\mathcal{U}} \le \delta_u$, then we obtain the inequality $\|\alpha v^* + x\|_{\mathcal{U}} \le |\alpha|\|v^*\|_{\mathcal{U}} + \|x\|_{\mathcal{U}} \le d_u$, and assumption (A) yields the estimate

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

In other words, if we define $\ell_\alpha = d_\lambda$ and $\ell_x = d_u$, and if we assume in addition that δ_λ and δ_u satisfy (20), then both hypotheses (H3) and (H4) are satisfied with $L_1 = M_1$ and $L_3 = \sigma$, and satisfy both

$$L_2 = M_1\|v^*\|_{\mathcal{U}} + M_2 \quad \text{and} \quad L_4 = M_1\|v^*\|_{\mathcal{U}}^2 + (M_2+M_3)\|v^*\|_{\mathcal{U}} + M_4\,.$$

Now an application of Theorem 1 completes the proof if we take into account Remark 2. Notice in particular that while the last constraint in (20) might make the set of admissible pairs $(\delta_\lambda, \delta_u)$ smaller, it will not render it empty; see also Figure 2. ∎

To close this section we briefly comment on how the branch segment validation theorem can be applied successively to establish large branch segments.

Remark 6 (linking branch segments). While Theorem 5 is able to validate significantly larger branch segments than the original version of the constructive implicit function theorem, in nearly all applications it has to be used successively to cover large branches. For this, one has to make sure that the small branch segments are properly linked.

Suppose that we have applied the branch segment validation theorem to an approximative solution $(\lambda_k^*, u_k^*) \in \mathbb{R}\times\mathcal{U}$, using the direction $v_k^* \in \mathcal{U}$. Theorem 5 then provides validation constants $\delta_{u,k}^{\min}$, $\delta_{u,k}^{\text{opt}}$, and $\delta_{\lambda,k}^{\text{opt}}$ such that the following holds:

- At $\lambda = \lambda_k^*$, there exists a unique solution of (14) which has distance at most $\delta_{u,k}^{\min}$ from the solution approximation u_k^*. In fact, we have $\delta_{u,k}^{\min} = 2K_k\varrho_k$, where ϱ_k and K_k denote the constants from (H1) and (H2) at the pair (λ_k^*, u_k^*).
- For every parameter λ with $|\lambda - \lambda_k^*| \le \delta_{\lambda,k}^{\mathrm{opt}}$, the problem (14) has a unique solution u which satisfies $\|u - (u_k^* + (\lambda - \lambda_k^*)v_k^*)\|_{\mathcal{U}} \le \delta_{u,k}^{\mathrm{opt}}$. In other words, there is an explicitly given slanted cylinder which contains a uniquely determined solution branch.

Assume now that we want to extend this validated branch piece to the right. While $u_k^* + \delta_{\lambda,k}^{\mathrm{opt}} v_k^*$ is an approximation of the solution of (14) for $\lambda = \lambda_{k+1}^* = \lambda_k^* + \delta_{\lambda,k}^{\mathrm{opt}}$, it is usually not good enough for an efficient application of Theorem 5. We therefore first determine a new numerical approximation $(\lambda_{k+1}^*, u_{k+1}^*)$ by applying a few steps of Newton's method to the operator $\mathcal{F}(\lambda_{k+1}^*, \cdot)$ starting at the above approximation $u_k^* + \delta_{\lambda,k}^{\mathrm{opt}} v_k^*$. If the branch segment validation theorem can be applied successfully at the new point $(\lambda_{k+1}^*, u_{k+1}^*)$, we obtain new validation constants $\delta_{u,k+1}^{\min}$, $\delta_{u,k+1}^{\mathrm{opt}}$, and $\delta_{\lambda,k+1}^{\mathrm{opt}}$ as above. Assume now that

$$\delta_{u,k+1}^{\min} + \left\| u_{k+1}^* - \left(u_k^* + \delta_{\lambda,k}^{\mathrm{opt}} v_k^* \right) \right\|_{\mathcal{U}} \le \delta_{u,k}^{\mathrm{opt}} \tag{23}$$

is satisfied. Then a straightforward application of the triangle inequality implies that the branch segment validated near the approximation $(\lambda_{k+1}^*, u_{k+1}^*)$ is connected with the segment validated near (λ_k^*, u_k^*). If, on the other hand, the inequality (23) is violated, then we cannot link the branches. This usually is an indication that one is too close to a bifurcation point. The above procedure is illustrated in the left image of Figure 3.

2.3. Resolving the neighborhood of a saddle-node bifurcation. The regular branch segment validation theorem from the last section can be used to validate long pieces of solution branches. For this, all one has to do is apply it iteratively, thereby covering a larger range of λ values. This is illustrated in the left image of Figure 3, 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 validation. In each validation region, the approximative solution at its center is indicated by a blue dot. However, the image also shows something else. If the solution branch contains a saddle-node bifurcation point, such as the one indicated by a black dot, Theorem 5 cannot in principle be used to validate the complete branch—locally near the bifurcation point the solution curve is not a graph over λ.

In the present section, we demonstrate that even though Theorem 5 does not apply near saddle-node bifurcations, we can still use the constructive implicit function theorem to validate the branch in a neighborhood of the bifurcation point. This time, however, Theorem 1 has to be applied in combination with a suitable change of variables. To motivate the latter, assume for the moment that (λ_0, u_0) is a saddle-node bifurcation point for (14). Then the Fréchet derivative $D_u\mathcal{F}(\lambda_0, u_0)$ has to have a nontrivial kernel. In the generic case, one would expect that this kernel is one-dimensional, spanned by some nonzero element $v_0 \in \mathcal{U}$. Moreover, at the saddle-node bifurcation, the solution curve should be tangent to v_0. It therefore seems reasonable to expect that if $\mathcal{U}^\perp \subset \mathcal{U}$ denotes a subspace which is complementary to the kernel of $D_u\mathcal{F}(\lambda_0, u_0)$, and if we decompose $u \in \mathcal{U}$ as

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

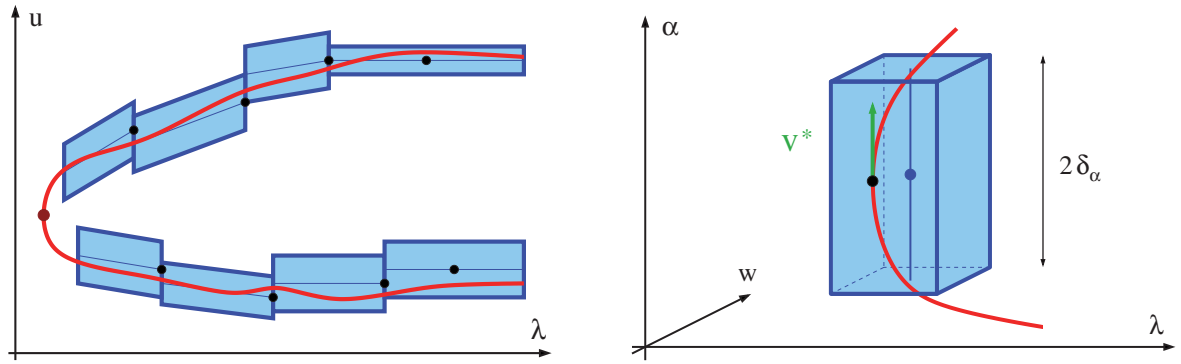

Figure 3. *Visualization of the saddle-node branch validation theorem. The left image illustrates how successive applications of Theorem 5 can be used to validate solution branches away from saddle-node bifurcation points. In order to validate the branch close to the bifurcation point, one has to employ Theorem 7, whose setup is illustrated in the right image. 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 $\mathcal{U}$.*

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

In practice, of course, we do not know the saddle-node bifurcation point exactly; i.e., we only have an approximation $(\lambda^*, u^*) \approx (\lambda_0, u_0)$ which satisfies $\mathcal{F}(\lambda^*, u^*) \approx 0$. Generically, this implies that the Fréchet derivative $D_u\mathcal{F}(\lambda^*, u^*)$ at the approximative bifurcation point is actually invertible, so there is no natural candidate for the tangent direction v^*. Rather, we will pick any nonzero element $v^* \in \mathcal{U}$ for which $D_u\mathcal{F}(\lambda^*, u^*)[v^*] \approx 0$. Similarly, the construction of the complementary space $\mathcal{U}^\perp$ has to be done in an approximative way, and we choose to model it as a bounded linear one-to-one operator of the form $w \mapsto \mathcal{H}w$, where the range $R(\mathcal{H})$ satisfies $\mathcal{U} = \operatorname{span}[v^*] \oplus R(\mathcal{H})$. Using these preparations, it is then indeed possible to apply the constructive implicit function theorem to validate the branch. This leads to the following theorem, which is illustrated in the right image of Figure 3.

Theorem 7 (saddle-node branch validation). *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*

$$\|\mathcal{F}(\lambda^*, u^*)\|_{\mathcal{V}} \le \varrho \qquad \text{and} \qquad \|D_u\mathcal{F}(\lambda^*, u^*)[v^*]\|_{\mathcal{V}} \le \sigma \tag{24}$$

for some nonnegative constants $\varrho \ge 0$ and $\sigma \ge 0$, which are needed to show (H1) *and a verticality condition similar to the slant condition in* (18). *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 with the properties*

$$R(\mathcal{H}) \text{ is closed} \qquad \text{and} \qquad \mathcal{U} = \operatorname{span}[v^*] \oplus R(\mathcal{H}) \,, \tag{25}$$

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

$$K \left\| D_\lambda\mathcal{F}(\lambda^*, u^*)\bar{\lambda} + D_u\mathcal{F}(\lambda^*, u^*)\,[\mathcal{H}\bar{w}] \right\|_{\mathcal{V}} \ge \max\left\{ |\bar{\lambda}| \,,\, \|\bar{w}\|_{\mathcal{W}} \right\} \tag{26}$$

holds for all $\bar{\lambda} \in \mathbb{R}$ *and* $\bar{w} \in \mathcal{W}$*, which will be used to show* (H2)*. Finally, define the constant* L_1 *as*

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

and suppose that

$$4K^2 \varrho L_1 < 1 \qquad \text{and} \qquad 2K\varrho < \ell_x = \min\left\{ d_\lambda \,,\; \frac{d_u}{\left\|\mathcal{H}\right\|_{\mathcal{L}(\mathcal{W},\mathcal{U})}} \right\} \tag{27}$$

hold, which will be used to establish (H3) *and* (6)*. If, in addition to the definition of* ℓ_x *in* (27)*, we also define* $\ell_\alpha = d_u / \|v^*\|_{\mathcal{U}}$*, then there exist pairs of constants* $(\delta_\alpha, \delta_x)$ *which satisfy*

$$0 < \delta_\alpha \le \ell_\alpha \,, \qquad 0 < \delta_x \le \ell_x \,, \qquad \text{and} \qquad \delta_\alpha \left\|v^*\right\|_{\mathcal{U}} + \delta_x \left\|\mathcal{H}\right\|_{\mathcal{L}(\mathcal{W},\mathcal{U})} \le d_u \,, \tag{28}$$

and satisfy the two conditions

$$2KL_1\delta_x + 2K\left(M_1 \left\|\mathcal{H}\right\|_{\mathcal{L}(\mathcal{W},\mathcal{U})} + M_3\right) \left\|v^*\right\|_{\mathcal{U}} \delta_\alpha \le 1 \tag{29}$$

and

$$2K\varrho + 2K\sigma\delta_\alpha + 2KM_1 \left\|v^*\right\|^2_{\mathcal{U}} \delta_\alpha^2 \le \delta_x \,, \tag{30}$$

which are needed to satisfy (7)*, and for each such pair the following holds. For every* $\alpha \in \mathbb{R}$ *with* $|\alpha| \le \delta_\alpha$ *there exist uniquely determined elements* $\lambda(\alpha) \in \mathbb{R}$ *and* $w(\alpha) \in \mathcal{W}$ *with*

$$|\lambda(\alpha) - \lambda^*| \le \delta_x \qquad \text{and} \qquad \left\|w(\alpha)\right\|_{\mathcal{W}} \le \delta_x \,,$$

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 vertical set*

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

lie on the branch $\alpha \mapsto (\lambda(\alpha), u(\alpha))$*, where* $u(\alpha) = u^* + \alpha v^* + \mathcal{H}w(\alpha)$*. In addition, if the mapping* $\mathcal{F} : \mathbb{R} \times \mathcal{U} \to \mathcal{V}$ *is* k*-times continuously Fréchet differentiable, then so is the solution function* $\alpha \mapsto (\lambda(\alpha), u(\alpha))$*. The statement of the theorem is illustrated in the right image of Figure* 3*.*

At first glance, the formulation of Theorem 7 might seem overwhelming. However, its underlying assumptions are straightforward to verify. All we need is a parameter-dependent mapping $\mathcal{F}$ which satisfies (A), and satisfy bounds of the form

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

where usually we will have $\|v^*\|_{\mathcal{U}} \approx 1$. In these bounds, our goal has to be the smallness of both ϱ and σ, as it will ensure that we are close to a candidate for a saddle-node bifurcation. Beyond these two estimates, one also needs the invertibility condition (26), which leads to the inverse bound K. Based on this information, the theorem usually applies as long as

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

which is similar to Theorem 5. In other words, the basic assumptions will always guarantee a small solution branch. In practice, of course, we would like this branch to be as large as possible, and for this we would like to choose the constant δ_α as large as possible. The crucial constraint in this context is (30), and it contains the small constant σ in the linear δ_α-term. As discussed at the end of section 2.2, larger branch pieces are obtained for smaller values of σ, and this in turn can be achieved by producing accurate approximations of the kernel function at the saddle-node bifurcation point. We now proceed with the proof of the theorem.

Proof. As in the regular branch segment validation theorem, we reduce the current result to the constructive implicit function theorem. 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 $\|(\lambda, w)\|_{\mathcal{X}} = \max\{|\lambda|\,, \|w\|_{\mathcal{W}}\}$. In addition, define the nonlinear operator $\mathcal{G} : \mathcal{P} \times \mathcal{X} \to \mathcal{Y}$ via

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

Our goal is to establish the existence of a branch of solutions of the problem $\mathcal{G}(\alpha, x) = 0$ 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 (24). Moreover, one can easily verify that $\mathcal{G}$ is Fréchet differentiable at every pair (α, x), where $x = (\lambda, w) \in \mathbb{R} \times \mathcal{W}$, and we have both

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

and

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

for all elements $\bar x = (\bar\lambda, \bar w) \in \mathcal{X} = \mathbb{R} \times \mathcal{W}$. At the specific pair (α^*, x^*) defined above, 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 (26). We now verify hypotheses (H3) and (H4). For this, assume that ℓ_α and ℓ_x are defined as in the formulation of the theorem, and consider a pair $(\delta_\alpha, \delta_x)$ satisfying (28). Furthermore, let $\alpha \in \mathbb{R}$, $\lambda \in \mathbb{R}$, and $w \in \mathcal{W}$ be arbitrary with

$$|\alpha| \le \delta_\alpha\,, \quad |\lambda - \lambda^*| \le \delta_x\,, \qquad \text{and} \qquad \|w\|_{\mathcal{W}} \le \delta_x\,.$$

These last estimates imply both $|\lambda - \lambda^*| \le d_\lambda$ and $\|\alpha v^* + \mathcal{H}w\|_{\mathcal{U}} \le d_u$, and therefore we can apply assumption (A) to obtain the estimates

$$\begin{aligned}
&\|D_u \mathcal{F}\left(\lambda, u^* + \alpha v^* + \mathcal{H}w\right)\mathcal{H} - D_u \mathcal{F}\left(\lambda^*, u^*\right)\mathcal{H}\|_{\mathcal{L}(\mathcal{W},\mathcal{V})} \\
&\qquad \le\; M_1 \|v^*\|_{\mathcal{U}} \|\mathcal{H}\|_{\mathcal{L}(\mathcal{W},\mathcal{U})} |\alpha| + M_1 \|\mathcal{H}\|^2_{\mathcal{L}(\mathcal{W},\mathcal{U})} \|w\|_{\mathcal{W}} + M_2 \|\mathcal{H}\|_{\mathcal{L}(\mathcal{W},\mathcal{U})} |\lambda - \lambda^*|\,, \\
&\|D_\lambda \mathcal{F}\left(\lambda, u^* + \alpha v^* + \mathcal{H}w\right) - D_\lambda \mathcal{F}\left(\lambda^*, u^*\right)\|_{\mathcal{V}} \\
&\qquad \le\; M_3 \|v^*\|_{\mathcal{U}} |\alpha| + M_3 \|\mathcal{H}\|_{\mathcal{L}(\mathcal{W},\mathcal{U})} \|w\|_{\mathcal{W}} + M_4 |\lambda - \lambda^*|\,.
\end{aligned}$$

Assume now that $(\bar{\lambda}, \bar{w}) \in \mathbb{R} \times \mathcal{W}$ is given with $|\bar{\lambda}| \leq 1$ and $\|\bar{w}\|_{\mathcal{W}} \leq 1$. Then the last two estimates imply

$$\begin{aligned}
&\left\| \left(D_x \mathcal{G}(\alpha, x) - D_x \mathcal{G}\left(\alpha^*, x^*\right) \right) \left[\left(\bar{\lambda}, \bar{w}\right) \right] \right\|_{\mathcal{V}} \\
&\quad \leq \left\| D_u \mathcal{F}\left(\lambda, u^* + \alpha v^* + \mathcal{H} w\right) \mathcal{H} - D_u \mathcal{F}\left(\lambda^*, u^*\right) \mathcal{H} \right\|_{\mathcal{L}(\mathcal{W}, \mathcal{V})} \|\bar{w}\|_{\mathcal{W}} \\
&\qquad + \left\| D_\lambda \mathcal{F}\left(\lambda, u^* + \alpha v^* + \mathcal{H} w\right) - D_\lambda \mathcal{F}\left(\lambda^*, u^*\right) \right\|_{\mathcal{V}} \left|\bar{\lambda}\right| \\
&\quad \leq M_1 \|v^*\|_{\mathcal{U}} \|\mathcal{H}\|_{\mathcal{L}(\mathcal{W}, \mathcal{U})} |\alpha| + M_1 \|\mathcal{H}\|^2_{\mathcal{L}(\mathcal{W}, \mathcal{U})} \|w\|_{\mathcal{W}} + M_2 \|\mathcal{H}\|_{\mathcal{L}(\mathcal{W}, \mathcal{U})} |\lambda - \lambda^*| \\
&\qquad + M_3 \|v^*\|_{\mathcal{U}} |\alpha| + M_3 \|\mathcal{H}\|_{\mathcal{L}(\mathcal{W}, \mathcal{U})} \|w\|_{\mathcal{W}} + M_4 |\lambda - \lambda^*| \\
&\quad \leq |\alpha| \left(M_1 \|\mathcal{H}\|_{\mathcal{L}(\mathcal{W}, \mathcal{U})} + M_3 \right) \|v^*\|_{\mathcal{U}} \\
&\qquad + \|x - x^*\|_{\mathcal{X}} \left(M_1 \|\mathcal{H}\|^2_{\mathcal{L}(\mathcal{W}, \mathcal{U})} + (M_2 + M_3) \|\mathcal{H}\|_{\mathcal{L}(\mathcal{W}, \mathcal{U})} + M_4 \right) ,
\end{aligned}$$

as well as

$$\begin{aligned}
\|D_\alpha \mathcal{G}(\alpha, x^*)\|_{\mathcal{V}} &= \|D_u \mathcal{F}\left(\lambda^*, u^* + \alpha v^*\right) [v^*]\|_{\mathcal{V}} \\
&\leq \|D_u \mathcal{F}\left(\lambda^*, u^*\right) [v^*]\|_{\mathcal{V}} + \|D_u \mathcal{F}\left(\lambda^*, u^* + \alpha v^*\right) [v^*] - D_u \mathcal{F}\left(\lambda^*, u^*\right) [v^*]\|_{\mathcal{V}} \\
&\leq \sigma + \|D_u \mathcal{F}\left(\lambda^*, u^* + \alpha v^*\right) - D_u \mathcal{F}\left(\lambda^*, u^*\right)\|_{\mathcal{L}(\mathcal{U}, \mathcal{V})} \|v^*\|_{\mathcal{U}} \\
&\leq \sigma + |\alpha| M_1 \|v^*\|^2_{\mathcal{U}} \ .
\end{aligned}$$

In other words, with the above definitions of ℓ_α and ℓ_x, and if we assume in addition that δ_α and δ_x satisfy (28), both hypotheses (H3) and (H4) are satisfied with L_1 as defined in the formulation of the theorem, $L_2 = (M_1 \|\mathcal{H}\|_{\mathcal{L}(\mathcal{W}, \mathcal{U})} + M_3) \|v^*\|_{\mathcal{U}}$, as well as with $L_3 = \sigma$ and $L_4 = M_1 \|v^*\|^2_{\mathcal{U}}$. Now an application of Theorem 1 completes the proof if we take into account Remark 2. Notice in particular that while the last constraint in (28) might make the set of admissible pairs $(\delta_\alpha, \delta_x)$ smaller, it will still be nonempty. ∎

To close this section we briefly comment on how the above theorem can be applied successively to cover larger branch segments close to a saddle-node bifurcation point, and how one can switch from applications of Theorem 7 to using Theorem 5.

Remark 8 (linking branch segments near saddle-node bifurcations). Just as with Theorem 5, one can apply the saddle-node branch validation theorem successively to cover larger branch pieces. While the basic procedure described in Remark 6 still applies, the specifics are a little more involved due to the nontrivial coordinate change used in Theorem 7.

Similarly to Remark 6, suppose we have applied Theorem 7 at an approximative solution $(\lambda_k^*, u_k^*) \in \mathbb{R} \times \mathcal{U}$, using the kernel element approximation $v_k^* \in \mathcal{U}$. Then the theorem provides validation constants $\delta_{x,k}^{\min}$, $\delta_{x,k}^{\text{opt}}$, and $\delta_{\alpha,k}^{\text{opt}}$, which cover the new parameter α and unknown $x = (\lambda, w)$ in the nonlinear problem

$$\mathcal{G}(\alpha, x) \ = \ \mathcal{F}\left(\lambda, u_k^* + \alpha v_k^* + \mathcal{H} w\right) \ = \ 0 \, .$$

Assume we want to extend the solution branch in the direction of positive α. For this, choose a step size $0 < \delta_{\alpha,k}^{\text{step}} < \delta_{\alpha,k}^{\text{opt}}$, and compute an improved numerical approximation for the transformed problem $\mathcal{G}(\delta_{\alpha,k}^{\text{step}}, x) = 0$ using Newton's method starting at $x = 0$. If the resulting approximation is denoted by $\bar{x}_k = (\bar{\lambda}_k, \bar{w}_k)$, we can determine the updated point in the λ-u-coordinate system as

$$\lambda_{k+1}^* = \bar{\lambda}_k \qquad \text{and} \qquad u_{k+1}^* \approx u_k^* + \delta_{\alpha,k}^{\text{step}} v_k^* + \mathcal{H}\bar{w}_k \,. \tag{31}$$

We would like to point out that the second identity is in fact only an approximation, due to the inevitability of roundoff errors in computing its right-hand side. Write this new solution approximation in the form $u_{k+1}^* = u_k^* + \alpha_{k+1} v_k^* + \mathcal{H} w_{k+1}$, and assume that both

$$\left|\alpha_{k+1} - \delta_{\alpha,k}^{\text{step}}\right| \leq \delta_{\alpha,k}^{\text{opt}} - \delta_{\alpha,k}^{\text{step}} \tag{32}$$

and

$$\delta_{x,k+1}^{\min} + \left\|\bar{x}_k - \left(\delta_{\alpha,k}^{\text{step}}, 0\right)\right\|_{\mathcal{X}} + \|w_{k+1} - \bar{w}_k\|_{\mathcal{W}} < \delta_{x,k}^{\text{opt}} \tag{33}$$

hold, where $\delta_{x,k+1}^{\min}$ is the corresponding validation constant from the application of Theorem 7 to the new pair $(\lambda_{k+1}^*, u_{k+1}^*)$. Then a straightforward application of the triangle inequality shows that the two validated branch segments connect. We would like to point out that on the right-hand side of (32), one can usually choose $\delta_{\alpha,k}^{\text{step}} = 0.9\,\delta_{\alpha,k}^{\text{opt}}$, which should suffice to account for the deviation in the α value generated by the evaluation of the rightmost expression in (31). Moreover, the three terms on the left-hand side of (33) account for the error tolerance of the next validation, the approximation error incurred by the Newton refinement, and the roundoff error generated by the evaluation of the rightmost expression in (31), respectively.

Remark 9 (transitioning from saddle-node to regular branches). We close this section by noting that a similar technique can be used to make sure that the validated branch segments connect if one switches from successive applications of Theorem 7 to applying Theorem 5. For this, assume that we are given an approximative solution $(\lambda^*, u^*) \in \mathbb{R} \times \mathcal{U}$ such that

- Theorem 7 can be successfully applied using the kernel element approximation $v^* \in \mathcal{U}$, leading to the validation constants $\delta_x^{\min}$, δ_x^{opt}, and $\delta_\alpha^{\text{opt}}$; and
- Theorem 5 can be successfully applied using the direction $v^* = 0$, leading to the validation constants $\delta_u^{\min}$, δ_u^{opt}, and $\delta_\lambda^{\text{opt}}$.

According to the first result, applied in the parameter-independent version, there exists a solution $(\bar{\alpha}, \bar{x})$ of $\mathcal{G}(\alpha, x) = 0$ with $\bar{\alpha} = 0$ and $\|\bar{x}\|_{\mathcal{X}} \leq \delta_x^{\min}$, where $\bar{x} = (\bar{\lambda}, \bar{w})$. This implies both

$$\left|\bar{\lambda} - \lambda^*\right| \leq \delta_x^{\min} \qquad \text{and} \qquad \|\bar{w} - 0\|_{\mathcal{W}} \leq \delta_x^{\min} \,.$$

Then one can easily see that this solution lies on the branch segment established by Theorem 5, as long as the inequalities

$$\delta_x^{\min} \leq \delta_\lambda^{\text{opt}} \qquad \text{and} \qquad \|\mathcal{H}\|_{\mathcal{L}(\mathcal{W},\mathcal{U})}\, \delta_x^{\min} \leq \delta_u^{\text{opt}}$$

are satisfied. This can easily be checked using interval arithmetic, and it shows that the two validated branches are indeed connected.

3. Application to the lattice Allen–Cahn equation. One of the fundamental models for the motion of interfaces in two-phase materials is the second-order partial differential equation

$$u_t = \Delta u + \lambda f(u) \tag{34}$$

due to Allen and Cahn [1]. While this model has been successful at explaining a variety of phenomena, it cannot explain the pinning of fronts, i.e., fronts which get stuck and stop moving as time increases. In order to remedy this shortcoming, a spatially discrete analogue of the Allen–Cahn equation has been proposed. In its simplest form, this model is given by the system of ordinary equations

$$\dot{u}_k = u_{k+1} - 2u_k + u_{k-1} + \lambda f(u_k) \qquad \text{for all} \qquad k = 1, \dots, n \ , \tag{35}$$

where we set $u_0 = u_1$ and $u_{n+1} = u_n$, and the nonlinearity is defined as

$$f(u) = \left(1 - u^2\right)(u - \mu) \ . \tag{36}$$

In our analysis, we will keep the parameter $\mu \in (-1, 1)$ fixed and consider only λ as a bifurcation parameter. While the more celebrated continuum Allen–Cahn model (34) has been studied without ceasing since it was first presented in 1979 [1], the discrete version has a much shorter, but also sparser, history, in part due to the intractability of many basic results due to the discrete nature of the system. See [4, 7, 8, 9, 12, 14, 15, 16, 17, 20, 31]. (Note that compared to the formulations in these papers, we have rescaled time, and our bifurcation parameter $\lambda > 0$ is such that $1/\sqrt{\lambda}$ is the interaction length.) In these previous works, it was shown that for $u \in \mathbb{R}^n$, for sufficiently large λ, there are exactly 3^n equilibria, known as *mosaic solutions*, such that for each index k we have either $u_k \approx 0$ or $u_k \approx -1$ or $u_k \approx +1$. Of these mosaic solutions, 2^n equilibria are stable, exactly when for every index k one has $u_k \approx \pm 1$. However, this does not explain the way in which the bifurcation diagram changes as a function of λ, and much work has gone into trying to understand the structure of the bifurcation diagram for this equation and how it compares to the limiting case of the continuum Allen–Cahn partial differential equation. In order to make rigorous headway, these papers are forced to consider a simplified version of the nonlinearity f. In contrast, we are able to study the bifurcation structure numerically and rigorously verify our results, all without having to simplify the nonlinear term f.

In this section, we demonstrate how the abstract results of section 2 can be used to obtain validation of paths of mosaic solutions as the parameter λ is varied. These results are presented in sections 3.1 through 3.3. This validation is accomplished by applying Theorems 5 and 7 to the discrete Allen–Cahn equation, which in return leads to the numerical validation of branches of mosaic solutions as the parameter λ is varied. We would like to point out that this application mainly rests on establishing rigorous bounds on the Jacobian matrix of the right-hand side of (35), which—as well as the remaining computations—can be obtained using interval arithmetic. After the branches have been established, we address two more topics. On the one hand, we demonstrate in section 3.5 how the constructive implicit function theorem can be used to rigorously verify the actual saddle-node bifurcation points on the above-mentioned solution curves, since their existence is only implied indirectly in section 3.3.

On the other hand, section 3.6 is devoted to rigorously establishing the index of equilibrium solutions together with their unstable eigendirections. This is accomplished using a combination of eigenvalue exclusion and eigenfunction-eigenvalue validation. All of these results are established using the constructive implicit function theorem. The discussions in the present section provide an example of how the abstract results of section 2 can easily be applied in the context of high-dimensional ordinary differential equations, and lead to rigorous results which are beyond the scope of classical methods.

3.1. Classification of mosaic solutions. Consider the discrete Allen–Cahn equation (35) introduced above. For sufficiently large values of the parameter λ there are 3^n mosaic solutions, 2^n of which are stable. Each stable mosaic solution lies on a smooth curve of mosaic solutions in the product $\mathbb{R} \times \mathbb{R}^n$ of the parameter and the phase space. Whereas all of these solutions exist for large λ, for small positive λ values, only very few mosaic solutions remain. Indeed, most solution branches undergo a unique saddle-node bifurcation. To specify their notation, let $u(\lambda)$ represent a stable mosaic solution branch for sufficiently large values of λ. Let the

$$\text{bifurcation value } \lambda_0(u) > 0$$

be the parameter value at which this branch first undergoes a bifurcation. If the branch contains a unique saddle-node bifurcation, it contains two equilibrium solutions for $\lambda > \lambda_0(u)$. One of these is the stable mosaic solution $u(\lambda)$, while the other is an unstable mosaic solution $w = w(\lambda)$. The fact that $w = (w_1, \ldots, w_n)$ is an unstable mosaic solution implies that for large values of λ, we have $w_k \approx \pm 1$ or $w_k \approx 0$, and the fact that w is unstable implies $w_k \approx 0$ for at least one value of the subscript k.

The goal of this paper is rigorous computer-assisted validation, but no validation is possible without a good initial numerical approximation. Therefore, in this section we describe how our numerical approximations were obtained, and we then give a detailed account of their subsequent validation. In the course of this, we will see that sometimes numerical computations can prove to be somewhat misleading.

As mentioned previously, for sufficiently large λ and $n = 10$, any stable mosaic solution satisfies $u_k \approx \pm 1$ for each index k and can therefore be uniquely described using a binary representation of an integer between 0 and 1023. In particular, for fixed $0 \le N \le 1023$, we use the binary representation $N = \sum_{k=0}^{9} (a_k\, 2^k)$ with $a_k \in \{0, 1\}$. Let $g = (g_1, \ldots, g_{10}) \in \mathbb{R}^{10}$, where

$$g_k = \begin{cases} +1 & \text{if} \quad a_{k-1} = 1 \\ -1 & \text{if} \quad a_{k-1} = 0 \end{cases} \qquad \text{for} \quad k = 1, \ldots, 10\,. \tag{37}$$

Since the solutions are completely symmetric, we need only consider the solutions with N ranging from 0 to 511—the other half of the solutions are equivalent. More precisely, the behavior of the solution associated with N is identical to the behavior of the solution associated with $1023 - N$. In fact, it is also unnecessary to consider the case $N = 0$, since it corresponds to an identically constant solution, which is an equilibrium for all values of λ. Thus, we restrict our attention to the integers $1 \le N \le 511$.

For $\lambda = 300$, we use Newton's method with initial guess g. Without exception, this method yielded a numerical approximation for the stable mosaic solution u corresponding

Table 1

Solution validation. *For each solution of the discrete Allen–Cahn equation with dimension $n = 10$, we specify the solution type and whether the branch and bifurcation point are validated. We find that all nonvalidated solutions have either even or odd symmetry. Most nonvalidated branches are validated in one direction but not in the other. We observe numerically that this is due to a secondary pitchfork bifurcation along the branch in one direction.*

Type	Solution number N	Total	Valid
Saddle-node	All others	422	Yes
Number index jump of 2	11, 12, 20, 23, 24, 36, 40, 43, 46, 68, 80, 81, 84, 92, 94, 95, 96, 97, 105, 116, 136, 144, 160, 161, 163, 165, 168, 174, 175, 184, 185, 188, 190, 191, 192, 213, 214, 232, 235, 244, 339, 346, 362, 363, 395, 406, 422, 423, 428, 464, 468, 471, 487, 488, 491, 500	56	Yes
Number index jump of 4	361, 421	2	Yes
Even symmetry	48, 120, 180, 204, 252, 258, 378, 390, 462, 510	10	No
Odd symmetry	47, 87, 103, 155, 171, 211, 227, 285, 301, 341, 357, 409, 425, 465, 481	15	No
No saddle-node	Odd symbol: 31; Even symbol: 72, 132, 306, 330, 438	6	No

to N. Starting with this numerical approximation, we employ numerical continuation using AUTO [11] to locate a bifurcation point for each of the equilibria. In all but six cases, AUTO locates a numerical saddle-node bifurcation point. In the remaining six cases, the branch only contained a numerical pitchfork bifurcation, and since the method in this paper is designed for validation of saddle-node bifurcations, we do not attempt validation. With all other cases, we then apply our validation method to get a rigorously validated bifurcation point, branch, and index for the corresponding approximate solution. We show later in the paper that if we can validate the bifurcation branch up to $\lambda = 114$, then the solution will persist without bifurcations for all $\lambda > 114$ as well. Therefore, by rigorously validating our solution up to the parameter value $\lambda = 114$, we are able to validate the entire branch of solutions for all λ values. This validation is successful for the entire branch in 480 out of the 511 cases.

Table 1 displays a full list of which N values correspond to stable mosaic solutions for which we have validated the full branch of solutions. The solutions that do not validate are precisely the solutions with either even or odd symmetry. These symmetric solutions include the six solutions which do not have any numerical saddle-node bifurcation point. In the other 25 symmetric cases, there is a saddle-node bifurcation point, but there is also a secondary bifurcation, which from the numerics appears to be a pitchfork bifurcation. In most cases the solution branch validates in one direction, where there is no secondary bifurcation, but it does not validate in the other direction. In addition to validation information, in Table 1 we list 58 solutions whose branches could be validated, but for which AUTO computations indicated even index jumps along the branch. More precisely, for 56 of these solutions the numerical AUTO branch computation included a jump in index from 0 to 2, and for two solutions the computed branch includes a jump from 0 to 4. In these 58 cases, the AUTO computation proceeds without any bifurcation being detected. In the next paragraph, we describe how our validation allows us to uncover more details of these 58 index jumps.

For the most part, obtaining the above results is merely a matter of taking the numerical approximations given by AUTO and feeding them into the validation code, which will be

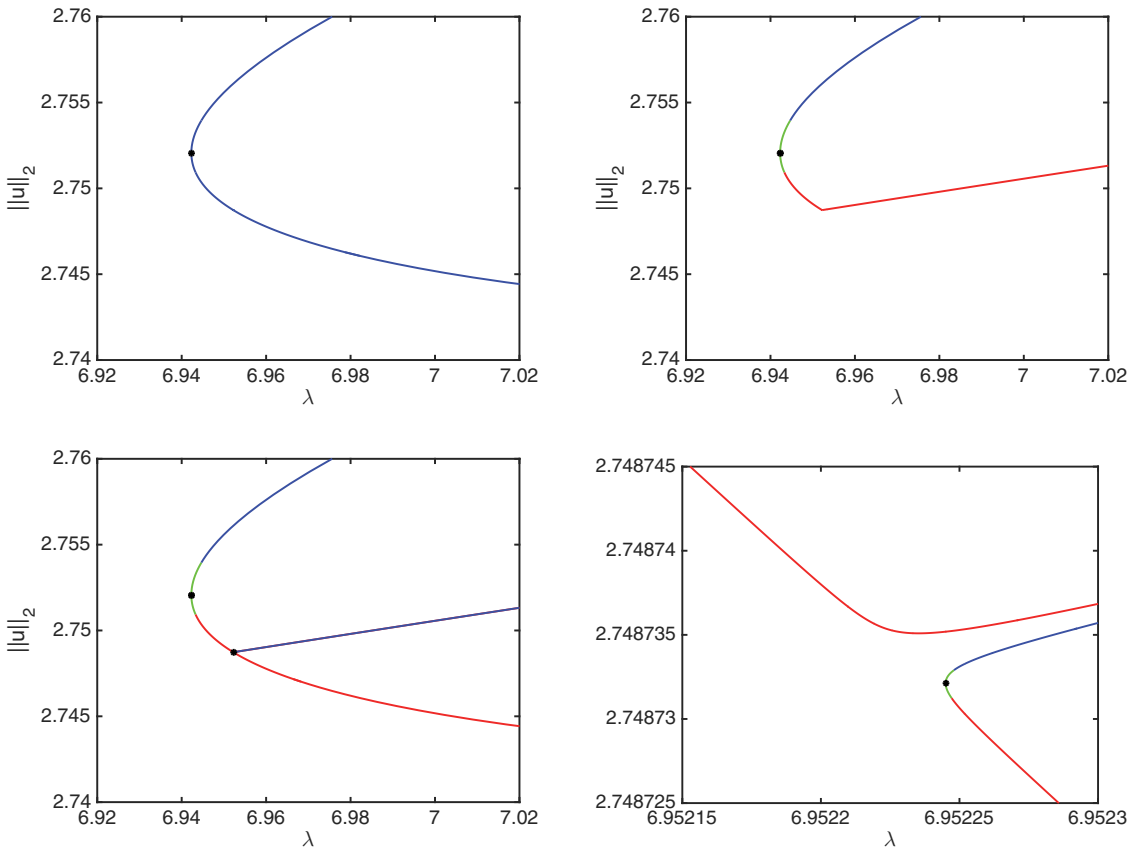

Figure 4. *Validated local bifurcation structure for $N = 40$. The top left iamge shows the result of an* AUTO *continuation of the branch corresponding to the $N = 40$ mosaic solution. In contrast, the validated branch, shown in the top right image, shows significantly different behavior. The two images in the bottom row give the complete validated picture. The original branch does indeed perform a sharp right turn, since nearby there is a second saddle-node bifurcation. The lower right image is a blowup of the lower left one.*

described in more detail in the following sections. However, in a number of cases the results are surprising at first sight. Consider for example the mosaic solution encoded by $N = 40$. The original AUTO simulation produces the branch shown in the top left image of Figure 4. The computation found one saddle-node bifurcation point and gives no other indications for bifurcations. We then used our validation code to establish the existence of the saddle-node point and followed the two half-branches out to $\lambda = 114$ using computer-assisted validation. Very much to our surprise, this leads to the image depicted in the upper right of Figure 4. As it turns out, the complete bifurcation diagram is as shown in the two images in the bottom row of the figure. Some distance away from the primary saddle-node bifurcation point, there is a second, disconnected branch generated by a distinct second saddle-node point. At this new bifurcation point, the branches are close enough so that AUTO jumps directly from one to the other, without any indication of the associated index change from 0 to 2. Similar behavior can be observed in the case $N = 144$, which is shown in Figure 5. In this case, AUTO does find two saddle-node points close to each other, but jumps again from one branch to the other.

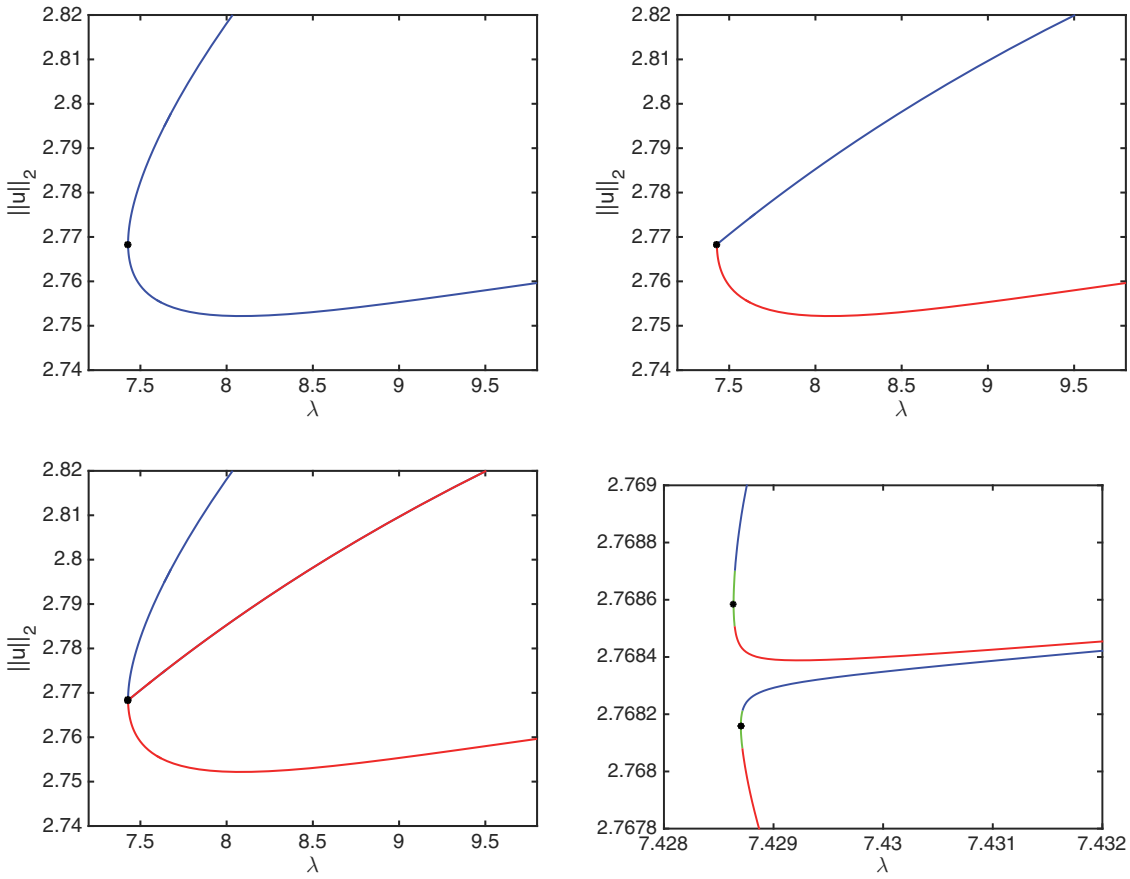

Figure 5. *Validated local bifurcation structure for $N = 144$. The top left image shows the result of an* AUTO *continuation of the branch corresponding to the $N = 144$ mosaic solution. As was seen in the previous figure, the validated branch, shown in the top right image, shows significantly different behavior. The two images in the bottom row give the complete validated picture, which consists of two separate saddle-node bifurcation points.*

While the correct behavior can probably be computed by changing the AUTO parameters, this is to be expected, since AUTO is designed to use the smoothness of the curve in order to follow the curve, meaning that close to the bifurcation, this problem will inevitably occur, whereas the actual branches are disjoint and smooth on a small level; from afar the first branch almost appears to have a point of nondifferentiability. On the other hand, our numerical validation code computes the appropriate step size along the branch, and therefore automatically takes care of the branch following in these near-bifurcation cases, either by faithfully following the branch or, if too close to the bifurcation, by returning a flag to the user that the branch cannot be validated. While we have given only two examples in the figures and our description, this same explanation is true of all 58 branches with the index jump. We also observe that without exception, in the cases where AUTO has an index jump of two or four, the branch that it jumps on has an unstable solution $\tilde{w}$ with the property that $\tilde{w}_k \approx 0$ for *two* or *four*, respectively, different values of k.

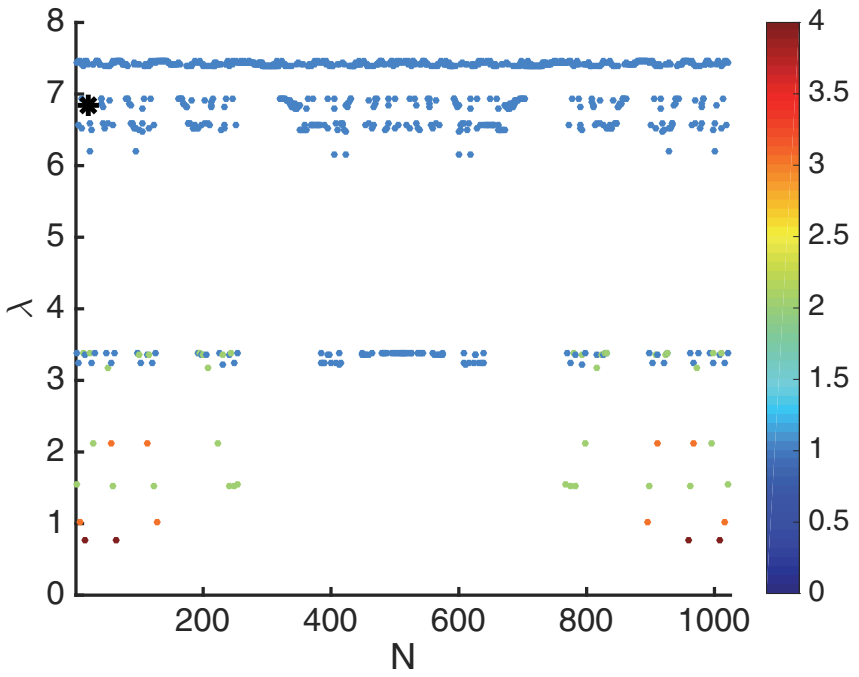

Figure 6. *The bifurcation parameter for dimension $n = 10$ for each of the 960 solutions with a saddle-node bifurcation, plotted with respect to the solution number N. The color bar indicates the size of the smallest grain. The black star at $N = 21$ indicates the solution branch shown in Figure 9.*

3.2. Robustness and bifurcation of equilibria. In this section, we present a result for the discrete Allen–Cahn equation for the specific dimension $n = 10$, which is similar in spirit to [15]. In particular, we characterize the branches of all the stable mosaic solutions which have a unique saddle-node bifurcation and can be validated by the methods discussed in the previous section and appearing in Table 1.

Recall from our description in the previous section that the stable mosaic solutions are of the form $u = (u_1, \ldots, u_{10})$ with $u_k \approx \pm 1$, giving us a convenient way of indexing these solutions through the one-to-one and onto correspondence between the stable mosaic solutions and the 10-digit binary numbers. Figure 6 shows the bifurcation parameter $\lambda_0(u)$ for each of the stable mosaic solutions as a function of the integer N associated with the mosaic solution u. Figure 7 shows the actual solutions along a branch for two different branches, associated with different binary integers N. These values do not appear random but seem to follow a pattern. For example, there are several clear gaps in the λ_0 values. In fact, there appear to be four distinct types of behavior. A large-λ_0 set with $6 < \lambda_0 < 8$, a medium-λ_0 set with $3 < \lambda_0 < 3.5$, a small-λ_0 set with $1.0 < \lambda < 2.5$, and four solutions in an extremely small-λ_0 set with $\lambda_0 \approx 0.778$. We are able to relate the geometry of these stable mosaic solutions to the location of the corresponding bifurcation point $\lambda_0(u)$ for the rigorously validated solutions listed in Table 1.

Before presenting our results, we need to introduce some notation. Consider two stable

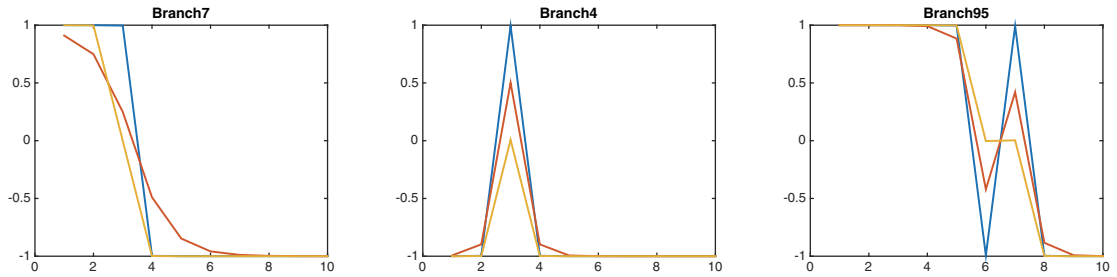

Figure 7. *Left: For dimension $n = 10$, the stable mosaic solution u corresponding to $N = 7$ computed for $\lambda = 300$ is shown in blue. The solution consists of values close to -1 and $+1$ encoded by the binary representation of $7 = 1 \cdot 2^0 + 1 \cdot 2^1 + 1 \cdot 2^2$, and the minimum grain size is 3 and the number of grains is 2. The solution at bifurcation value λ_0 is shown in red. The unstable solution w on the same branch at $\lambda = 300$ is shown in yellow, with $w_k \approx \pm 1$ for $k \neq 3$, and $w_3 \approx 0$. Middle: Stable, bifurcation, and unstable solutions for the branch which corresponds to $N = 4 = 1 \cdot 2^2$ with the same color scheme as in the left image. The stable solution u has minimum grain size of 1, and the number of grains is 3. The unstable solution w has the almost zero $w_3 \approx 0$. Right: Solutions for $N = 95$. The stable solution has again a minimum grain size of 1. The unstable solution $\tilde{w}$ that* AUTO *jumps on has two almost zero entries, which are adjacent to each other. Adjacent double almost zeros with $\tilde{w}_{k^*-1}$ and $\tilde{w}_{k^*+2}$ of opposite sign occur 24 times, all with very close bifurcation values $6.1 < \lambda_0 < 6.2$. Such a case, but with $\tilde{w}_{k^*-1}$ and $\tilde{w}_{k^*+2}$ of the same sign, occurs six times, all with very close bifurcation values near 3.4.*

mosaic solutions u and v. We say that u is *more robust* than v if $\lambda_0(u) < \lambda_0(v)$. Define a *grain* of a solution u to be a largest set of consecutive sites k such that u_k always has the same sign. The *grain size* is the number of sites within the grain. Grant [15] shows that in the case of a particular piecewise linear set-valued nonlinearity, the stable equilibria with smaller sized grains are less robust than equilibria possessing only larger sized grains. Similarly, we consider the structure of robust and nonrobust equilibria in the case of a smooth nonlinearity. We get the following result.

Theorem 10 (grain size and robustness). *Let u denote a stable mosaic solution of the discrete Allen–Cahn equation (35) with $\mu = 0$ and $n = 10$, and let $\lambda_0(u)$ denote the parameter value of the associated saddle-node bifurcation. If $\lambda_0 < 3$, then the size of the smallest grain is greater than 1 and the number of grains is equal to 2 or 3. If $\lambda_0 > 3.5$, then the size of the smallest grain is equal to 1 and the number of grains is at least 3.*

We now proceed to make some further observations that are based on a combination of our rigorous validation and the numerics done in AUTO. To make this clear, we label these results observations rather than theorems. At this point, we have only developed our validation methods for saddle-node points. In future work, we plan to also develop validation for pitchfork methods, at which point it will be possible to reconsider these statements using validation techniques.

Observation 11 (further robustness and grain size). *Assume the situation of the above theorem. For all but four solutions, $\lambda_0 < 2.5$ if and only if the smallest grain size is at least 2 and the number of grains is at most 3. Furthermore, each of the four exceptional solutions has an index jump of 2 in the* AUTO *simulation, as denoted in Table 1, and the corresponding unstable solution has adjacent almost zeros. We believe that the fact that the branch is close*

to another solution branch alters the behavior of solutions. For all four of these exceptional solutions, the bifurcation point λ_0 *is almost identical, with a value between* 3.3 *and* 3.5.

We pointed out in the previous section that exactly the 58 validated solutions with numerical index jumps of 2 or 4 had the property that for the AUTO-computed unstable mosaic solution $\tilde{w}$ one has $\tilde{w}_k \approx 0$ for two or four values of k. In addition, for 30 of these 58 cases, the mosaic solution $\tilde{w}$ had two adjacent almost zeros occur at indices k^* and k^*+1. For example, all four of the exceptional branches in the previous observation have $\tilde{w}$ with adjacent almost zeros. We now make an observation that applies to the 450 validated branches without adjacent almost zeros.

Let w be the unstable mosaic solution in the same branch as stable mosaic solution u. We know that there exists k^* such that $w_{k^*} \approx 0$. We say that this zero is *contained in a transition layer* if $w_{k^*-1} \approx \pm 1$ and $w_{k^*+1} \approx \pm 1$ are of opposite sign, or if either $k^* = 1$ or $k^* = 10$.

Observation 12 (robustness and transition layers). *Let* u *denote a stable mosaic solution of the discrete Allen–Cahn equation* (35) *with* $\mu = 0$ *and* $n = 10$*, and let* $\lambda_0(u)$ *denote the parameter value of the associated saddle-node bifurcation. Furthermore, let* w *denote an unstable solution on the same branch for large values of* λ*. Then a nonadjacent almost zero of* w *is contained in a transition layer if and only if* $\lambda_0 < 3.5$*.*

We now proceed with an observation about the 30 cases with adjacent almost zeros. It is in a sense also about transition layers, but the conclusions are quite different in these cases. In 6 of the 30 cases, the components $\tilde{w}_{k^*-1}$ and $\tilde{w}_{k^*+2}$ close to ± 1 are of the same sign, and in 24 cases they have opposite signs; see, for example, the branch 95 solution depicted in Figure 7.

Observation 13 (adjacent almost zeros). *Assume the situation of the last two observations and the theorem. For the* 30 *branches with an associated* $\tilde{w}$ *with adjacent almost zeros, there are two possible sets of behavior. In the first case, a branch of solutions has bifurcation value* $6.1 < \lambda_0 < 6.2$*, and the unstable solution* w *has two adjacent almost zeros* $\tilde{w}_{k^*} \approx 0$ *and* $\tilde{w}_{k^*+1} \approx 0$*, and* $\tilde{w}_{k^*-1}$ *and* $\tilde{w}_{k^*+2}$ *are close to* ± 1 *and have opposite signs. Among all* 480 *validated solutions, these* 24 *are the only solutions with* λ_0 *in this parameter range. In the second case,* $3.3 < \lambda_0 < 3.5$*, where* $\tilde{w}_{k^*}, \tilde{w}_{k^*+1} \approx 0$*, and* $\tilde{w}_{k^*-1}$ *and* $\tilde{w}_{k^*+2}$ *have the same sign. In this second case, the bifurcation points are not isolated. That is, there are other validated solutions out of the full* 480 *with a bifurcation point in this same parameter range.*

The above results apply to just one specific example which demonstrates that for a fixed smooth nonlinearity and fixed dimension n, computer-assisted proofs can be used to make rigorous statements about the relation between grain sizes and bifurcations in lattice dynamical systems. We would like to point out, however, that this result was selected as a proof of concept. Nevertheless, the methods described in the following sections are extremely flexible. One can easily exchange the nonlinearity f, and choose considerably larger dimensions n. For example, we validated branches of solutions for two mosaic solutions in dimension $n = 100$ with nonlinearity $f(u) = \sin(\pi u)/\pi$. This merely required a small change of the validation code—one has to exchange the function definition and adapt certain nonlinearity estimates for f; see (39). After these changes, validation succeeded as easily as in the previous situation. In Figure 8, we show two solutions at different locations along branches of the new model, along with the validated saddle-node points and branches.

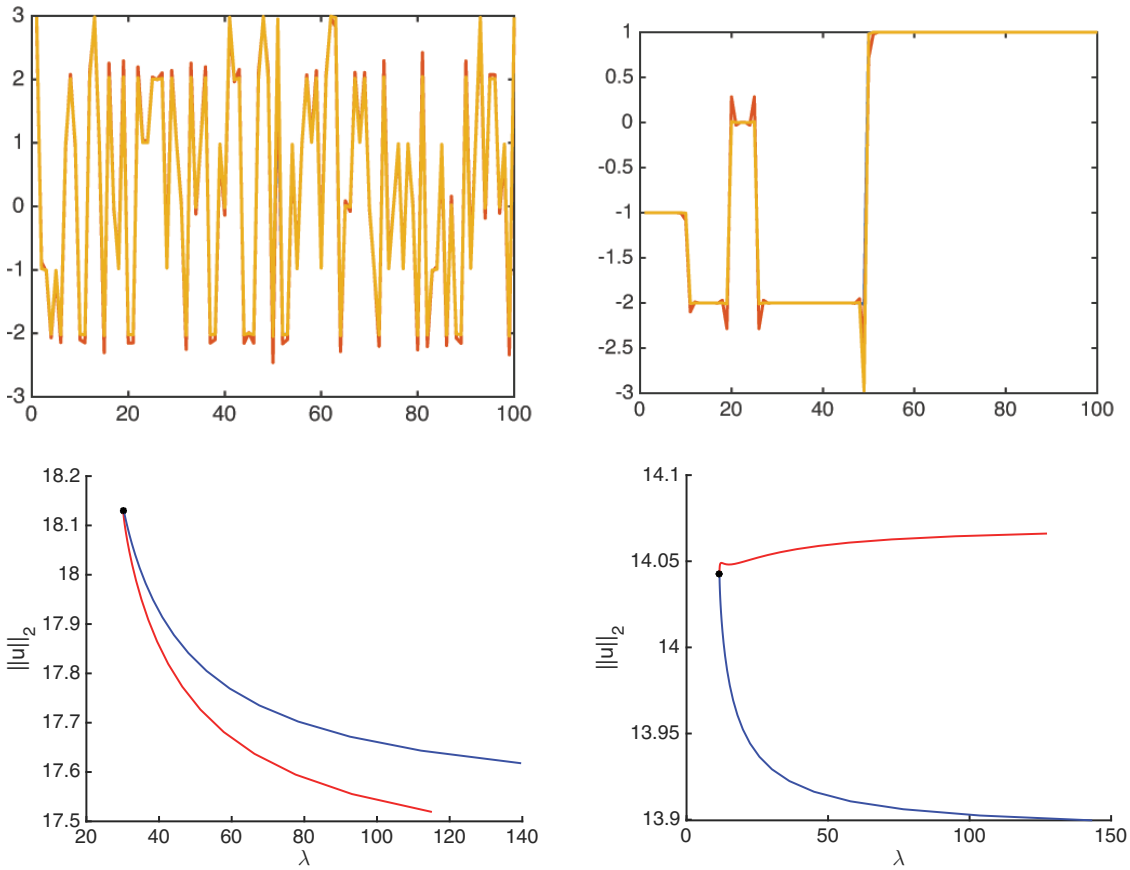

Figure 8. *The top two images show mosaic solutions for the* 100*-dimensional discrete Allen–Cahn equation. In blue are the stable mosaic solutions at* $\lambda = 200$*. In red are solutions on the same branch at the bifurcation point, and in yellow are the resulting unstable mosaic solutions on the same branch when the solutions once again reach the parameter value* $\lambda = 200$*. Note that the blue stable solutions are barely visible since they only vary at a few places from the other two solutions along the branch. The bottom two images show the validated branches of the corresponding solutions shown in the top row.*

3.3. Branch validation for mosaic solutions. Now that we have established our discrete Allen–Cahn validation results in the previous two sections, we consider the details of how we validate branches for the discrete Allen–Cahn model (35) with nonlinearity $f(u) = (1 - u^2)(u - \mu)$ given in (36), where $\mu \in (-1, 1)$ is a fixed value and λ is our bifurcation parameter. For the purposes of this section, it will be convenient to rewrite the equation satisfied by equilibrium solutions in the form

$$\mathcal{F}(\lambda, u) = Au + \lambda f(u) = 0 \qquad \text{with} \qquad \lambda \in \mathbb{R} \quad \text{and} \quad u \in \mathbb{R}^n\,, \tag{38}$$

where from the form of the discrete Laplacian, the matrix $A \in \mathbb{R}^{n\times n}$ is given by

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

As before, we 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. Finally, we use the abbreviation $\mathrm{diag}(u)$ for the diagonal matrix in $\mathbb{R}^{n\times n}$ with diagonal entries $u_1, \dots, u_n$. Throughout this section, we use the maximum norm in $\mathbb{R}^n$ to measure the length of vectors, 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 denote

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

dropping the usual subscript ∞ for the sake of brevity.

In order to apply Theorems 5 and 7 from section 2 we need only establish the validity of assumption (A). This is the subject of the following lemma.

Lemma 14. *Let $\mu \in (-1, 1)$ be a fixed value, consider the nonlinearity $f(u) = (1-u^2)(u-\mu)$ from (36), let $n \in \mathbb{N}$, and define the nonlinear operator $\mathcal{F} : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ as in (38). In addition, let $\lambda^* \in \mathbb{R}$ and $u^* \in \mathbb{R}^n$ be arbitrary, but fixed, and let $d_u > 0$ be a given constant. Then for all $\lambda \in \mathbb{R}$ and $u \in \mathbb{R}^n$ with $\|u - u^*\| \le d_u$, we have*

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

where

$$M_1 = |\lambda^*| \max_{|\xi|\le\|u^*\|+d_u} \left|f''(\xi)\right| , \quad M_2 = M_3 = \max_{|\xi|\le\|u^*\|+d_u} \left|f'(\xi)\right| , \quad \text{and} \quad M_4 = 0 . \tag{39}$$

In other words, the nonlinear mapping $\mathcal{F}$ satisfies (A) from section 2.2 with d_u as above and $d_\lambda = \infty$. Recall that in all of the above estimates, we use the maximum norm for vectors as well as the induced matrix norm.

Note that with this lemma, we are in a position to apply the branch validation results from the previous section to the specific application of the discrete Allen–Cahn equation. However, we have stated the lemma such that it is by no means limited to this particular case. Indeed, the estimates in Lemma 14 hold for *arbitrary* matrices $A \in \mathbb{R}^{n\times n}$ and nonlinearities $f : \mathbb{R} \to \mathbb{R}$. For our specific choice of $f(u) = (1 - u^2)(u - \mu) = u - u^3 + \mu u^2 - \mu$ one can easily see that in the notation of the above proof, we have

$$f^{(1)}_{\max} \le 1 + 2|\mu| \left(\|u^*\| + d_u\right) + 3 \left(\|u^*\| + d_u\right)^2 \quad \text{and} \quad f^{(2)}_{\max} \le 2|\mu| + 6 \left(\|u^*\| + d_u\right) , \tag{40}$$

since $f'(u) = 1-3u^2+2\mu u$ and $f''(u) = 2\mu-6u$. Similarly, for our second choice of nonlinearity $f(u) = \sin(\pi u)/\pi + \mu$, we get $f^{(1)}_{\max} = 1$ and $f^{(2)}_{\max} = \pi$. We now proceed with the proof of the lemma.

Proof. For $p \in \mathbb{N}$ we use the abbreviations

$$f^{(p)}_{\max} = \max\left\{ \left| f^{(p)}(\xi) \right| \ : \ |\xi| \le \|u^*\| + d_u \right\} . \tag{41}$$

Let $u \in \mathbb{R}^n$ be arbitrary with $\|u - u^*\| \le d_u$. Then $|f'(u_k)| \le f^{(1)}_{\max}$ for each $k = 1, \ldots, n$, and the mean value theorem further implies for some ξ_k between u_k and u^*_k the estimate

$$\left| f'(u_k) - f'(u^*_k) \right| = \left| f''(\xi_k) \right| |u_k - u^*_k| \le f^{(2)}_{\max} \|u - u^*\| \ .$$

One can easily see that the mapping $\mathcal{F}$ defined in (38) is smooth, and that its Jacobian matrix with respect to u is given by $D_u\mathcal{F}(\lambda, u) = A + \lambda \operatorname{diag}(f'(u))$. Now the above estimates yield

$$\begin{aligned} \|D_u\mathcal{F}(\lambda, u) - D_u\mathcal{F}(\lambda^*, u^*)\| &\le \left\| \lambda \operatorname{diag}(f'(u)) - \lambda^* \operatorname{diag}(f'(u^*)) \right\| \\ &\le |\lambda - \lambda^*| \left\| \operatorname{diag}(f'(u)) \right\| + |\lambda^*| \left\| \operatorname{diag}(f'(u) - f'(u^*)) \right\| \\ &\le |\lambda - \lambda^*| \, f^{(1)}_{\max} + \|u - u^*\| \, |\lambda^*| \, f^{(2)}_{\max} \ . \end{aligned}$$

We now turn our attention towards the derivative $D_\lambda\mathcal{F}(\lambda, u) = f(u)$. Using the mean value theorem one more time, one obtains $|f(u_k) - f(u^*_k)| \le f^{(1)}_{\max}|u_k - u^*_k|$, and this finally implies

$$\|D_\lambda\mathcal{F}(\lambda, u) - D_\lambda\mathcal{F}(\lambda^*, u^*)\| \ = \ \|f(u) - f(u^*)\| \ \le \ f^{(1)}_{\max} \|u - u^*\| \ .$$

This completes the proof of the lemma. ∎

Using the above lemma, both the regular branch segment validation theorem and the saddle-node branch validation theorem can easily be applied, and this will be described in the remainder of this section.

Regular branch validation. Applying Theorem 5 to the discrete Allen–Cahn equation is now straightforward. For this, assume we have found a numerical approximation (λ^*, u^*) to a solution of $\mathcal{F}(\lambda, u) = 0$. The constants $M_1, \ldots, M_4$ are known from (39) and (40), and we need only establish the estimates in (17) and (18). This will be done in a computational way by taking into account all occurring roundoff errors through the use of interval arithmetic [22]. For our applications, we use the MATLAB toolbox INTLAB [28] in the following way:

- In order to satisfy the left side of (17), 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, we have

  $$\|\mathcal{F}(\lambda^*, u^*)\| \le \varrho \ .$$

  Usually if one starts with a good enough solution approximation, the constant ϱ will be only slightly larger than machine precision.

- Next, for (18) one has to decide on a tangent direction $v^* \in \mathbb{R}^n$ for the branch continuation. While in principle one could use $v^* = 0$, significantly larger branch pieces can be validated if we let v^* denote a numerical approximation of the solution v of the tridiagonal linear system $D_u\mathcal{F}(\lambda^*, u^*)[v] = -D_\lambda\mathcal{F}(\lambda^*, u^*)$, which for the discrete Allen–Cahn equation gives $Av + \lambda^* \operatorname{diag}(f'(u^*))v = -f(u^*)$. Proceeding as in the first point, one can then compute a constant σ such that

$$\|D_\lambda\mathcal{F}(\lambda^*, u^*) + D_u\mathcal{F}(\lambda^*, u^*)[v^*]\| \le \sigma \,,$$

  and also the constant σ will usually be close to machine precision.
- Finally, to satisfy the right side of (17), we need to determine a rigorous upper bound on the induced ∞-norm of the inverse of the matrix $D_u\mathcal{F}(\lambda^*, u^*) = A+\lambda^* \operatorname{diag}(f'(u^*))$. If we pass 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 the inclusion $D_u\mathcal{F}(\lambda^*, u^*) \in \Xi$ holds.
  We now proceed as in [33] to establish the invertibility of all matrices in Ξ, and to determine a common upper bound K on the ∞-norms of the inverses of all of these matrices. Following [29, 30], let $B \in \mathbb{R}^{n\times n}$ be the numerically computed inverse of an arbitrary matrix in Ξ, for example, of the linearization $D_u\mathcal{F}(\lambda^*, u^*)$. Then interval computations can be used to find rigorous bounds $0 < \varrho_1 < 1$ and $\varrho_2 > 0$ with

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

  and Lemma 4 implies that every matrix in Ξ is invertible with

$$\left\|C^{-1}\right\| \le \frac{\varrho_2}{1-\varrho_1} \quad \text{for all} \quad C \in \Xi \,.$$

  If we let K be the right endpoint of the interval enclosure of $\varrho_2/(1-\varrho_1)$, then we finally have

$$\left\|D_u\mathcal{F}(\lambda^*, u^*)^{-1}\right\| \le K \,.$$

  This time, the constant K will not necessarily be small. In fact, its size determines whether or not validation is possible.

For each of the results in sections 3.1 and 3.2, the above three steps were performed iteratively to cover the large portions of solution branches away from the saddle-node bifurcation points.

Saddle-node branch validation. In order to describe the solution branch close to the saddle-node bifurcation point, we need to apply Theorem 7. While in principle this is not more complicated than the procedure above, a few modifications are necessary. These are outlined in the following. Our starting point now is a numerical approximation (λ^*, u^*) to a saddle-node bifurcation point for $\mathcal{F}(\lambda, u) = 0$, and we still can use the constants $M_1, \ldots, M_4$ from (39) and (40). As for the remaining assumptions of Theorem 7, we proceed as follows:

- As in the first step of the regular branch validation, in order to ensure the left side of (24), we use the numerical approximation (λ^*, u^*) and interval arithmetic to rigorously compute a small constant $\varrho > 0$ with $\|\mathcal{F}(\lambda^*, u^*)\| \le \varrho$.

- For the right side of (24) in the application of Theorem 7, the vector v^* now has to be an approximation to the kernel function of the linearization at the saddle-node bifurcation point. Note that usually the matrix $D_u\mathcal{F}(\lambda^*, u^*)$ will be invertible, since (λ^*, u^*) is only an approximation. Therefore, we choose v^* as a numerical approximation to an eigenvector of $D_u\mathcal{F}(\lambda^*, u^*)$ which corresponds to the eigenvalue closest to zero. By normalizing v^*, and then using interval arithmetic to find an enclosure for its maximum norm, we obtain a rigorous bound for $\|v^*\|$, which is only slightly larger than 1. Furthermore, proceeding as before one can compute a constant σ such that $\|D_u\mathcal{F}(\lambda^*, u^*)[v^*]\| \leq \sigma$, and the constant σ will be close to machine precision.
- To verify (25), we need to choose the operator $\mathcal{H}$. In our application, we use $\mathcal{W} = \mathbb{R}^{n-1}$ and construct the matrix $\mathcal{H} \in \mathbb{R}^{n\times(n-1)}$ as follows. Let $k \in \{1, \ldots, n\}$ denote the index for which we have

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

  and let $\mathcal{H}$ denote the matrix which is obtained from the identity matrix in $\mathbb{R}^{n\times n}$ by removing the kth column. Then the range of $\mathcal{H}$ is just the orthogonal complement of the kth standard unit vector in $\mathbb{R}^n$, and due to $v_k^* \neq 0$ this immediately implies that $R(\mathcal{H})$ is closed and provides the necessary identity $\mathbb{R}^n = \operatorname{span}[v^*] \oplus R(\mathcal{H})$. Furthermore, since all of the columns of $\mathcal{H}$ are different standard unit vectors in $\mathbb{R}^n$, the ∞-norm of $\mathcal{H}$ is exactly 1, and the matrix $\mathcal{H}$ has trivial kernel.
- In order to verify the inequality in (26), we first determine a constant K such that for all $\bar\lambda \in \mathbb{R}$ and $\bar w \in \mathcal{W}$ one has

$$K \left\| \left(\; D_\lambda\mathcal{F}(\lambda^*, u^*) \;\middle|\; D_u\mathcal{F}(\lambda^*, u^*)\mathcal{H} \; \right) (\bar\lambda, \bar w)^t \right\| \geq \left\| (\bar\lambda, \bar w)^t \right\|$$

  or, equivalently,

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

  In other words, we bound the induced ∞-norm of the inverse of a matrix whose first column is given by the vector $D_\lambda\mathcal{F}(\lambda^*, u^*) = f(u^*)$, and whose remaining columns are determined by the matrix product $D_u\mathcal{F}(\lambda^*, u^*)\mathcal{H} = (A + \lambda^* \operatorname{diag}(f'(u^*)))\mathcal{H}$. This can be accomplished as above using Lemma 4, and it establishes (26).

Theorem 7 can be applied iteratively, and in this way one can rigorously verify the remaining part of the equilibrium branch close to the saddle-node bifurcation point.

We now demonstrate the applicability of the above approach in a specific example. Consider the discrete Allen–Cahn equation (38) with $\mu = 0$ and $n = 10$. Then regular numerical path-following computations using AUTO [11] indicate the existence of a saddle-node

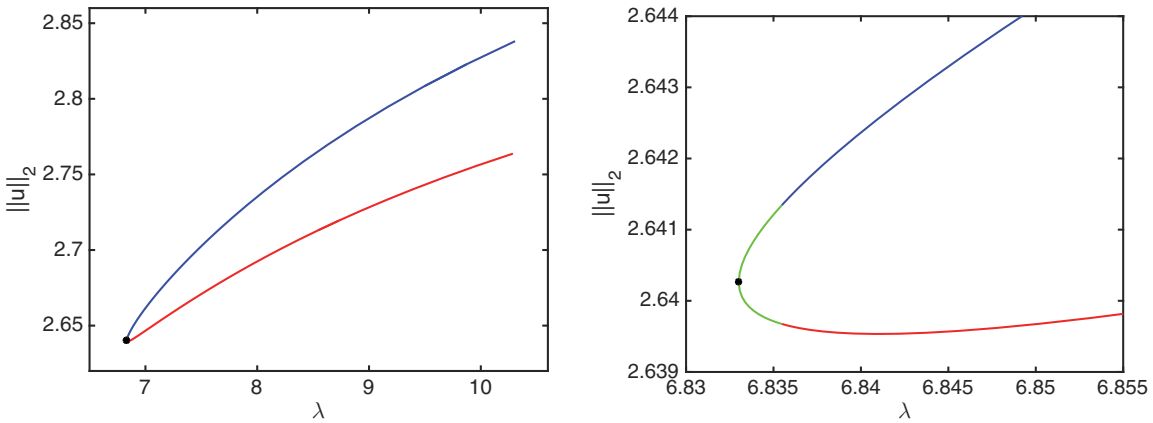

Figure 9. *Rigorously validated branch of equilibrium solutions for the discrete Allen–Cahn equation (38) with $\mu = 0$ and $n = 10$. While the left image shows the validated branch, the right image depicts the neighborhood of the saddle-node bifurcation point. The numerical approximation (λ^*, u^*) given in (42) 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 7, and the red and blue parts were validated by Theorem 5. This solution branch is denoted by a black star in Figure 6.*

bifurcation point close to the pair $(\lambda^*, u^*) \in \mathbb{R} \times \mathbb{R}^{10}$ given by

$$\lambda^* = 6.8330142962 \qquad \text{and} \qquad u^* = \begin{pmatrix} 0.87393566065 \\ -0.53677750139 \\ 0.66351174478 \\ -0.67399510947 \\ 0.50181564902 \\ -0.88782171323 \\ -0.99274089040 \\ -0.99953444484 \\ -0.99997015142 \\ -0.99999796478 \end{pmatrix}. \tag{42}$$

Note that this solution is associated with $N = 21$, depicted by a black star in Figure 6. 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 7 iteratively 10 times in the direction v^* and 10 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 5 iteratively to follow the two halves of the branch until the parameter λ exceeded $\lambda = 10$. 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 55 and 50 iterations, respectively.

3.4. Validating branches for large parameter values. The branch validation in the previous section is based on an iterative approach, meaning that by its nature we can only ever

use it to validate a finite portion of the solution branch, i.e., up to some finite λ value. While previous work has shown that these mosaic solutions exist for all sufficiently large values of λ, it does not provide explicit bounds as to what "sufficiently large" means in practice. In this section we show that the parameter-independent version of the constructive implicit function theorem can be used without any numerical assistance to furnish such bounds. More precisely, it will be shown below that for $\mu = 0$ all branches of mosaic solutions exist as long as $\lambda \geq 114$, and we will obtain explicit bounds on the size of the neighborhood in which these stationary states exist. In view of the previous section, this implies that if we can use Theorems 5 and 7 to validate the branches over intervals of the form $[\lambda_0, 114]$, then we have in fact validated the entire branches for $\lambda \in [\lambda_0, \infty)$.

In order to apply the constructive implicit function theorem directly, we consider again the equilibrium equation associated with the discrete Allen–Cahn equation (35), yet rewritten in the form

$$\mathcal{G}(u) = \frac{1}{\lambda} A u + f(u) = 0 \qquad \text{with} \qquad \lambda \in \mathbb{R}^+ \quad \text{and} \quad u \in \mathbb{R}^n \,, \tag{43}$$

where the matrix $A \in \mathbb{R}^{n\times n}$ and the nonlinearity $f : \mathbb{R}^n \to \mathbb{R}^n$ are defined as before. In addition, let $u^* \in \{-1, 0, 1\}^n$ be arbitrary and $\mu = 0$. Our goal is to use Theorem 1 to find an equilibrium solution of (43) for sufficiently large $\lambda > 0$, i.e., we need to establish hypotheses (H1) and (H2), as well as (H3) in the reduced form discussed in Remark 3. 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 we see that, together with

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

(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 entries of the matrix $\operatorname{diag} f'(u^*)$ are contained in the set $\{1, -2\}^n$, since for $\mu = 0$ we have $f'(u) = 1 - 3u^2$. 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 4 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 constructive implicit function theorem in (6) 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 1 implies that for all $\lambda \geq 114$, there exists a unique equilibrium solution u_λ of (43) 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 constructive implicit function theorem can be applied to obtain explicit existence proofs for equilibrium solutions. In fact, one can easily see that Theorem 1 does in fact establish smooth branches of solutions for all $\lambda \geq 114$.

3.5. Locating saddle-node bifurcation points. The approach described in the previous two sections allows us to validate complete solution branches for the discrete Allen–Cahn equation, even at saddle-node bifurcation points. In this way, one can deduce the existence of such bifurcation points indirectly. However, in many situations it is desirable to provide a computer-assisted existence proof for the bifurcation point itself, including error tolerances. The purely numerical aspects of this have been addressed in [21]. In this paper it is shown that every nondegenerate solution (λ_0, u_0, v_0) of the extended system

$$\begin{aligned} \mathcal{F}(\lambda_0, u_0) &= 0\ , \\ D_u\mathcal{F}(\lambda_0, u_0)\,[v_0] &= 0\ , \\ \hat{v}^t v_0 - 1 &= 0 \end{aligned} \tag{44}$$

gives rise to a saddle-node bifurcation for the problem $\mathcal{F}(\lambda, u) = 0$ at the point (λ_0, u_0), where the nonzero vector $\hat{v}$ is used for normalizing the kernel element v_0. In the next result, we demonstrate that one can use the constructive implicit function theorem to rigorously solve the extended system (44). In particular, we let $\mathcal{G}$ denote the nonlinear mapping associated with the extended system $\mathcal{G} : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}$ defined by

$$\mathcal{G}(x) = \big(\mathcal{F}(\lambda, u)\ ,\ \ D_u\mathcal{F}(\lambda, u)\,[v]\,,\ \ \hat{v}^t v - 1\big)\ , \qquad \text{where} \qquad x = (u, v, \lambda). \tag{45}$$

We view x as the unknown variable vector, implying that the function $\mathcal{G}$ is parameter-independent. Then we can use the parameter-independent version of the constructive implicit function theorem to prove the existence of a zero of $\mathcal{G}$, as formulated in the following result.

Proposition 15 (saddle-node bifurcation point validation). *Let $\mathcal{F} : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ be given by $\mathcal{F}(\lambda, u) = Au + \lambda f(u)$ as in (38). Let $\mu \in (-1, 1)$ be a fixed constant, consider the specific nonlinearity $f(u) = (1 - u^2)(u - \mu)$, and let $\hat{v} \in \mathbb{R}^n$ be a fixed normalization vector. In addition, let $\lambda^* \in \mathbb{R}$, $u^* \in \mathbb{R}^n$, and $v^* \in \mathbb{R}^n$ be such that for some constant $\varrho > 0$ we have*

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

Let $d_u > 0$, define $f^{(p)}_{\max}$ as in (41), and let

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

Finally, let $K > 0$ be such that

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

and suppose that

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

Then for every constant δ which satisfies

$$2K\varrho \;\leq\; \delta \;\leq\; \min\left\{\frac{1}{2KL}\,,\; d_u\right\}$$

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 (44) and satisfies the estimates

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

In other words, the maximum norm ball of radius $2K\varrho$ centered at the triple (λ^*, u^*, v^*) contains a unique saddle-node bifurcation point for $\mathcal{F}(\lambda, u) = 0$, and this bifurcation point is unique up to distance $\min\{1/(2KL), d_u\}$.

Proof. Let $\mathcal{F}$ be as in the statement of the theorem. Let $\mathcal{G}$ be the extended system defined in (45). Then one can see that its Jacobian matrix is given by

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

Specifically for the discrete Allen–Cahn equation, this Jacobian takes the concrete form

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

The assumptions of the proposition directly imply hypotheses (H1) and (H2) of the constructive implicit function theorem. We now show that hypothesis (H3) holds with $\ell_\alpha = 0$

and $\ell_x = d_u$. In order to bound the norm of the difference $D_x\mathcal{G}(x) - D_x\mathcal{G}(x^*)$ we first recall from the proof of Lemma 14 that for $\|u - u^*\| \le d_u$ the estimates

$$\left\|\lambda \operatorname{diag}(f'(u)) - \lambda^* \operatorname{diag}(f'(u^*))\right\| \le f^{(1)}_{\max} |\lambda - \lambda^*| + f^{(2)}_{\max} |\lambda^*| \, \|u - u^*\| \qquad \text{and}$$
$$\|f(u) - f(u^*)\| \le f^{(1)}_{\max} \|u - u^*\|$$

hold, where $f^{(p)}_{\max}$ was defined in (41). Similarly, one can show that as long as $\|u - u^*\| \le d_u$ we have

$$\begin{aligned}
&\left\|\lambda \operatorname{diag}(f''(u)) \operatorname{diag}(v) - \lambda^* \operatorname{diag}(f''(u^*)) \operatorname{diag}(v^*)\right\| \\
&\quad \le \left\|\operatorname{diag}(f''(u)) \operatorname{diag}(\lambda v - \lambda^* v^*)\right\| + \left\|\lambda^* \operatorname{diag}(f''(u) - f''(u^*)) \operatorname{diag}(v^*)\right\| \\
&\quad \le f^{(2)}_{\max} \|\lambda v - \lambda^* v^*\| + f^{(3)}_{\max} |\lambda^*| \, \|v^*\| \, \|u - u^*\| \\
&\quad \le f^{(2)}_{\max} \|v\| \, |\lambda - \lambda^*| + f^{(2)}_{\max} |\lambda^*| \, \|v - v^*\| + f^{(3)}_{\max} |\lambda^*| \, \|v^*\| \, \|u - u^*\|
\end{aligned}$$

as well as

$$\begin{aligned}
\left\|\operatorname{diag}(f'(u))v - \operatorname{diag}(f'(u^*))v^*\right\| &\le \left\|\operatorname{diag}(f'(u)) \, (v - v^*)\right\| + \left\|\operatorname{diag}(f'(u) - f'(u^*))v^*\right\| \\
&\le f^{(1)}_{\max} \|v - v^*\| + f^{(2)}_{\max} \|v^*\| \, \|u - u^*\| \ .
\end{aligned}$$

Now let $\bar{x} = (\bar{u}, \bar{v}, \bar{\lambda})$ be arbitrary with maximum norm $\|\bar{x}\| \le 1$. These estimates imply both

$$\begin{aligned}
&\|(D_u\mathcal{F}(\lambda, u) - D_u\mathcal{F}(\lambda^*, u^*)) \, \bar{u}\| + \| (D_\lambda\mathcal{F}(\lambda, u) - D_\lambda\mathcal{F}(\lambda^*, u^*)) \, \bar{\lambda}\| \\
&\quad \le \left\|\lambda \operatorname{diag}(f'(u)) - \lambda^* \operatorname{diag}(f'(u^*))\right\| \|\bar{u}\| + \|f(u) - f(u^*)\| \, |\bar{\lambda}| \\
&\quad \le f^{(1)}_{\max} |\lambda - \lambda^*| + f^{(2)}_{\max} |\lambda^*| \, \|u - u^*\| + f^{(1)}_{\max} \|u - u^*\| \\
&\quad \le \left(2f^{(1)}_{\max} + f^{(2)}_{\max} |\lambda^*|\right) \max\{|\lambda - \lambda^*|, \|u - u^*\|\} \ \le \ \left(2f^{(1)}_{\max} + f^{(2)}_{\max} |\lambda^*|\right) \|x - x^*\|
\end{aligned}$$

and

$$\begin{aligned}
&\|D_{uu}\mathcal{F}(\lambda, u)[v, \bar{u}] - D_{uu}\mathcal{F}(\lambda^*, u^*)[v^*, \bar{u}]\| + \|D_u\mathcal{F}(\lambda, u)[\bar{v}] - D_u\mathcal{F}(\lambda^*, u^*)[\bar{v}]\| \\
&\qquad + \left\|D_{\lambda u}\mathcal{F}(\lambda, u)[v]\bar{\lambda} + D_{\lambda u}\mathcal{F}(\lambda^*, u^*)[v^*]\bar{\lambda}\right\| \\
&\quad \le \left(f^{(2)}_{\max} \|v\| \, |\lambda - \lambda^*| + f^{(2)}_{\max} |\lambda^*| \, \|v - v^*\| + f^{(3)}_{\max} |\lambda^*| \, \|v^*\| \, \|u - u^*\|\right) \|\bar{u}\| \\
&\qquad + \left(f^{(1)}_{\max} |\lambda - \lambda^*| + f^{(2)}_{\max} |\lambda^*| \, \|u - u^*\|\right) \|\bar{v}\| \\
&\qquad + \left(f^{(1)}_{\max} \|v - v^*\| + f^{(2)}_{\max} \|v^*\| \, \|u - u^*\|\right) |\bar{\lambda}| \\
&\quad \le \left(2f^{(1)}_{\max} + f^{(2)}_{\max} (2|\lambda^*| + 2\|v^*\| + d_u) + f^{(3)}_{\max} |\lambda^*| \, \|v^*\|\right) \|x - x^*\| \ ,
\end{aligned}$$

and since the operator norm $\|D_x\mathcal{G}(x) - D_x\mathcal{G}(x^*)\|$ equals the maximum of the right-hand sides of the last two estimates, one finally obtains

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

with L as defined in the formulation of the theorem. 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 1, hypothesis (H4) does not have to be verified according to Remark 3, and the result follows. ∎

Using the above result, one can establish the existence of the saddle-node bifurcation points which were described in sections 3.1 and 3.2. One can proceed exactly as outlined in section 3.3 using the interval arithmetic toolbox INTLAB.

As a numerical example, we return to the pair (λ^*, u^*) defined in (42), which approximates a saddle-node bifurcation point. Using Proposition 15 one can then rigorously establish the existence of the saddle-node 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 of the extended system (44). This leads to an improved bifurcation point approximation less than 10^{-9} away from the pair in (42), and another application of Proposition 15 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.

3.6. Eigenvalue exclusion and index validation. As our final topic, we turn our attention to rigorously determining the index of an equilibrium solution, i.e., to computing the number of positive eigenvalues of the linearization at the equilibrium. Once we have found a numerical approximation of the equilibrium and established a true solution in a small neighborhood, one can use the resulting region to find an interval matrix which contains the linearization—and our task is to count the number of all positive eigenvalues for all matrices described by the interval matrix. To do this, we first focus on a specific matrix $B \in \mathbb{R}^{n\times n}$ and on the problem of validating an eigenvalue $\eta \in \mathbb{R}$ and associated eigenvector $v \in \mathbb{R}^n$. If we pick a fixed suitable normalization vector $\hat{v} \in \mathbb{R}^n \setminus \{0\}$, then such a pair has to solve the system

$$Bv - \eta v = 0 \quad \text{and} \quad \hat{v}^t v - 1 = 0 \;, \tag{46}$$

where the second equation is added to isolate the solution. In fact, one can show that if η is a simple eigenvalue, and if v is an eigenvector with $\hat{v}^t v = 1$, then the pair (v, η) is the only solution of (46). Such isolated solutions can be validated using the constructive implicit function theorem applied to the parameter-independent function $\mathcal{G}(v, \eta) = (Bv - \eta v, \hat{v}^t v - 1)$ as follows.

Proposition 16 (eigenvalue and eigenvector validation). *Let $B \in \mathbb{C}^{n\times n}$ be a given matrix. Furthermore, let $\eta^* \in \mathbb{C}$ and $v^* \in \mathbb{C}^n$ be given in such a way that for some $\varrho > 0$ we have*

$$\|Bv^* - \eta^* v^*\| \leq \varrho \qquad \text{and} \qquad \left|\hat{v}^t v^* - 1\right| \leq \varrho \;,$$

let $K > 0$ be such that

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

and 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 (46) and which satisfies both of the estimates $|\eta - \eta^*| \leq \delta$ 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 $1/(4K)$. The analogous result is valid for real eigenvalues of real matrices.

Proof. Let $x^* = (v^*, \eta^*)$ and consider the mapping $\mathcal{G} : \mathbb{C}^n \times \mathbb{C} \to \mathbb{C}^n \times \mathbb{C}$ defined by

$$\mathcal{G}(x) = \big(Bv - \eta v \, , \ \hat{v}^t v - 1 \big) \ , \qquad \text{where} \qquad x = (v, \eta) \ ,$$

whose Jacobian matrix is given by

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

Now let $\bar{x} = (\bar{v}, \bar{\eta}) \in \mathbb{C}^n \times \mathbb{C}$ be arbitrary with maximum norm $\|\bar{x}\| \leq 1$. Then one obtains

$$\begin{aligned} \|(D_x \mathcal{G}(x) - D_x \mathcal{G}(x^*))\, \bar{x}\| &\leq |\eta - \eta^*| \, \|\bar{v}\| + \|v - v^*\| \, |\bar{\eta}| \\ &\leq |\eta - \eta^*| + \|v - v^*\| \ \leq \ 2 \, \|x - x^*\| \ ; \end{aligned}$$

i.e., we have $\|D_x \mathcal{G}(x) - D_x \mathcal{G}(x^*)\| \leq 2\|x - x^*\|$. This implies that hypothesis (H3) is satisfied with $L_1 = 2$ and $\ell_u = \infty$, as well as $\ell_\alpha = 0$. Since the assumptions of the proposition clearly imply hypotheses (H1) and (H2), the result now follows in view of Remark 3. ∎

Note that the above proposition 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 (46).

In principle, we can now use numerical approximations for all positive eigenvalues of the linearization of (38) at an equilibrium and validate them one by one. However, this does not yet imply the index of the equilibrium. We also have to make sure that there are no other positive eigenvalues. For this, we need the following lemma.

Lemma 17 (eigenvalue exclusion). *Let $B \in \mathbb{C}^{n \times n}$ be a given matrix. Furthermore, let $\eta \in \mathbb{C}$ be arbitrary, and suppose that the induced ∞-norm of the inverse of the matrix $B - \eta I$ satisfies*

$$\left\| (B - \eta I)^{-1} \right\| \leq K$$

for some $K > 0$. Then the matrix B has no eigenvalue in the open disk with radius $1/K$ centered at the point η in the complex plane.

Proof. Let $\kappa \in \mathbb{C}$ denote an arbitrary eigenvalue of B, and let $v \in \mathbb{C} \setminus \{0\}$ be an associated eigenvector. Then the norm bound on $(B - \eta I)^{-1}$ implies the estimate $\|v\| \leq K \|(B - \eta I)v\|$, which in turn leads to

$$|\kappa - \eta| \ = \ \frac{\|(\kappa - \eta)v\|}{\|v\|} \ = \ \frac{\|(B - \eta I)v\|}{\|v\|} \ \geq \ \frac{1}{K} \ ,$$

and this completes the proof of the lemma. ∎

At first glance, one might be tempted to think that the above two results suffice to establish the index of an equilibrium solution. After all, all positive eigenvalues have to be contained in the interval $(0, \|B\|]$, and after validating the eigenvalues which determine the index using Proposition 16, we just have to exclude the remainder of the interval using Lemma 17. Note, however, that while Proposition 16 proves the existence of an eigenvalue η, it does not provide an interval around η in which this eigenvalue is unique. The uniqueness assertion 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 $\eta \approx \tilde{\eta}$.

Fortunately, it is possible to obtain an interval of uniqueness around a validated eigenvalue by investing a little more work. For this, we make use of the eigenvalue deflation technique described in [35, pp. 596ff].

Lemma 18 (eigenvalue deflation). *Let $B \in \mathbb{C}^{n\times n}$ be a given complex matrix with eigenvalues $\eta_1, \ldots, \eta_k \in \mathbb{C}$ and associated eigenvectors $v_1, \ldots, v_k \in \mathbb{C}^n$. Furthermore, let $w \in \mathbb{C}^n$ denote a vector such that $w^t v_1 = 1$. Then the matrix*

$$C = B - v_1 w^t B$$

has the eigenvalue 0 with eigenvector v_1, and for $\ell = 2, \ldots, k$ it still has the eigenvalue η_ℓ, but now with associated eigenvector $v_\ell - (w^t v_\ell) v_1$.

Proof. One can easily see that $Cv_1 = \eta_1 v_1 - v_1 w^t \eta_1 v_1 = \eta_1 v_1 - \eta_1 v_1 \left(w^t v_1\right) = 0$, and for arbitrary $\ell = 2, \ldots, k$ we further have

$$C\left(v_\ell - (w^t v_\ell) v_1\right) \;=\; Cv_\ell \;=\; \eta_\ell v_\ell - v_1 w^t \eta_\ell v_\ell \;=\; \eta_\ell \left(v_\ell - \left(w^t v_\ell\right) v_1\right) \;.$$

This completes the proof of the lemma. ∎

With the above results, we can now proceed as follows to validate the index of an equilibrium solution of the discrete Allen–Cahn equation:

- First, we need to compute an interval enclosure (λ^*, u^*) for an equilibrium solution of (38). Using this enclosure, one can then compute an interval matrix B which contains the true linearization at the equilibrium.
- We then find numerically all positive eigenvalues for one of the matrices enclosed by B. We also find all of the associated eigenvectors. These can then be validated for the interval matrix B using Proposition 16.
- Now let η^* be an interval which contains a verified eigenvalue for all matrices in B, and let v^* denote an interval vector which contains the associated eigenvector. We now determine an interval matrix C which contains all matrices of the form $B - v^* w^t B$, where w is an interval vector such that $w^t v^*$ contains 1. Using Lemma 17, one can then determine an interval around η^* in which none of the matrices enclosed by C has an eigenvalue. According to Lemma 18, this means that all of the matrices in B have exactly one eigenvalue in this interval—namely the one enclosed by η^*.
- As a last step, Lemma 17 is applied iteratively to exclude eigenvalues in the remaining intervals of $[0, \|B\|]$.

Notice that this index validation has to be performed only for one solution on each verified solution branch in the bifurcation diagram. Due to the last assertion in the constructive

implicit function theorem, the linearization of (38) at any stationary state on a verified branch piece is invertible; i.e., since these linearizations are symmetric, the index can only change at bifurcation points. We would like to also point out that while the above procedure might seem somewhat involved at first sight, in practice it usually is pretty straightforward. Most solutions of interest will have fairly small index, even if the dimension n of the problem is large, and by using the above approach we need only validate a handful of eigenvalues. In fact, in many cases some of the negative eigenvalues might be fairly close to each other and form clusters, and in such a situation they would be more difficult to validate.

To demonstrate the above approach, we consider two specific stationary solutions on the branch of equilibria shown in Figure 9. The branch originates from a saddle-node bifurcation point close to the pair (λ^*, u^*) defined in (42). For the solution on the blue branch piece at parameter value $\lambda = 10$, the numerical index computation reveals no positive eigenvalues. Using an upper bound on the maximum norm of the linearization at the equilibrium point, one can see that in order to prove the asymptotic stability of the equilibrium, it suffices to exclude eigenvalues in the interval $[0, 24]$. This can be achieved using three iterative applications of Lemma 17. As a second solution, we consider the solution on the red branch piece in Figure 9 at parameter value $\lambda = 10$. This time, standard numerical routines find one possible positive eigenvalue at $\eta \approx 6.411297$. Using Proposition 16 this eigenvalue can be validated, and both eigenvalue and eigenvector can be determined up to an error bound of $4 \cdot 10^{-14}$. Using Lemma 18 it can be shown that the eigenvalue is unique in the interval $(6.9 \cdot 10^{-13}, 12.822)$, and the maximum norm of the linearization at the equilibrium point is again bounded by 24. Eigenvalues in the remaining intervals $[0, 6.9 \cdot 10^{-13}]$ and $[12.822, 24]$ can then be excluded using one and two iterations of Lemma 17. This shows that all equilibria on the blue branch piece in Figure 9 are asymptotically stable, while those on the red piece have index one.

The above two examples are typical in many situations, since frequently the index of solutions of interest is low. Nevertheless, the approach also works extremely well in more degenerate situations. To demonstrate this, we determined a numerical approximation of an equilibrium solution at the parameter value $\lambda = 50$ through a Newton iteration starting at the point $(1, -1, 1, -1, 0, -1, 0, -1, 0, -1)^t$. This leads to a nearby equilibrium solution which can be validated as above. In this case, the linearization of the discrete Allen–Cahn equation has three positive eigenvalues $\eta_1 > \eta_2 > \eta_3 > 0$, all of which can be verified using Proposition 16. Moreover, Lemma 18 produces the following uniqueness intervals:

$$\begin{aligned}
\eta_3 &\approx 47.75322 \in (47.74738\,,\ 47.75907)\,,\\
\eta_2 &\approx 47.76341 \in (47.75807\,,\ 47.76875)\,,\\
\eta_1 &\approx 47.77367 \in (47.76776\,,\ 47.77958)\,.
\end{aligned}$$

Notice that both the first and the third interval overlap with the second, so one only has to exclude eigenvalues in the two intervals $[0, 47.74738]$ and $[47.77958, 98.87545]$, where the last number is a bound on the induced maximum norm of the linearization. Using Lemma 17, these intervals can be shown to be devoid of eigenvalues in 3 and 14 iterations, respectively. Note that our method is precise enough to even resolve these clustered single eigenvalues.

Acknowledgments. We thank the referees for their careful reading of our paper and their constructive suggestions for improvements.

REFERENCES

[1] S. M. Allen and J. W. Cahn, *A microscopic theory for antiphase boundary motion and its application to antiphase domain coarsening*, Acta Metallurgica, 27 (1979), pp. 1085–1095.

[2] G. Arioli and H. Koch, *Computer-assisted methods for the study of stationary solutions in dissipative systems, applied to the Kuramoto-Sivashinski equation*, Arch. Ration. Mech. Anal., 197 (2010), pp. 1033–1051.

[3] C. Beentjes, *Computing Bifurcation Diagrams with Deflation*, MSc thesis, University of Oxford, Oxford, UK, 2015.

[4] J. W. Cahn, S.-N. Chow, and E. S. Van Vleck, *Spatially discrete nonlinear diffusion equations*, Rocky Mountain J. Math., 25 (1995), pp. 87–118.

[5] L. Chierchia, *KAM lectures*, in Dynamical Systems. Part I, Pubbl. Cent. Ric. Mat. Ennio Giorgi, Scuola Norm. Sup., Pisa, Italy, 2003, pp. 1–55.

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

[7] S.-N. Chow and J. Mallet-Paret, *Pattern formation and spatial chaos in lattice dynamical systems* I., IEEE Trans. Circuits and Systems. I. Fund. Theory Appl., 42 (1995), pp. 746–751.

[8] S.-N. Chow, J. Mallet-Paret, and E. S. Van Vleck, *Dynamics of lattice differential equations*, Intern. J. Bifur. Chaos Appl. Sci. Engrg., 6 (1996), pp. 1605–1621.

[9] S.-N. Chow, J. Mallet-Paret, and E. S. Van Vleck, *Pattern formation and spatial chaos in spatially discrete evolution equations*, Random Comput. Dynam., 4 (1996), pp. 109–178.

[10] S. Day, J.-P. Lessard, and K. Mischaikow, *Validated continuation for equilibria of PDEs*, SIAM J. Numer. Anal., 45 (2007), pp. 1398–1424, doi:10.1137/050645968.

[11] 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), Vol. 30, 1981, pp. 265–284.

[12] C. M. Elliott and A. M. Stuart, *The global dynamics of discrete semilinear parabolic equations*, SIAM J. Numer. Anal., 30 (1993), pp. 1622–1663, doi:10.1137/0730084.

[13] M. Gameiro, J.-P. Lessard, and K. Mischaikow, *Validated continuation over large parameter ranges for equilibria of PDEs*, Math. Comput. Simul., 79 (2008), pp. 1368–1382.

[14] C. P. Grant, *Grain sizes in the discrete Allen-Cahn and Cahn-Hilliard equations*, Discrete Contin. Dyn. Syst., 7 (2001), pp. 127–146.

[15] C. P. Grant, *Superabundance of stationary solutions for the discrete Allen-Cahn equation*, Dyn. Contin. Discrete Impulsive Syst. Ser. B Appl. Algorithms, 8 (2001), pp. 71–91.

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

[17] J. K. Hale, *Numerical dynamics*, Contemp. Math., 172 (1994), pp. 1–30.

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

[19] S. G. Krantz and H. R. Parks, *The Implicit Function Theorem. History, Theory, and Applications*, Reprint of the 2003 edition, Modern Birkhäuser Classics, Birkhäuser/Springer, New York, 2013.

[20] J. Mallet-Paret and S.-N. Chow, *Pattern formation and spatial chaos in lattice dynamical systems.* II, IEEE Trans. Circuits Systems. I. Fund. Theory Appl., 42 (1995), pp. 752–756.

[21] G. Moore and A. Spence, *The calculation of turning points of nonlinear equations*, SIAM J. Numer. Anal., 17 (1980), pp. 567–576.

[22] R. E. Moore, R. B. Kearfott, and M. J. Cloud, *Introduction to Interval Analysis*, SIAM, Philadelphia, 2009, doi:10.1137/1.9780898717716.

[23] A. Neumaier, *Introduction to Numerical Analysis*, Cambridge University Press, Cambridge, UK, 2001.

[24] J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Classics Appl. Math. 30, SIAM, Philadelphia, 2000, doi:10.1137/1.9780898719468.

[25] M. Plum, *Existence and enclosure results for continua of solutions of parameter-dependent nonlinear boundary value problems*, J. Comput. Appl. Math., 60 (1995), pp. 187–200.

[26] M. Plum, *Enclosures for two-point boundary value problems near bifurcation points*, in Scientific Computing and Validated Numerics (Wuppertal, 1995), Math. Res. 90, Akademie Verlag, Berlin, 1996, pp. 265–279.

[27] M. Plum, *Computer-assisted proofs for semilinear elliptic boundary value problems*, Japan J. Indust. Appl. Math., 26 (2009), pp. 419–442.
[28] S. M. Rump, *INTLAB—INTerval LABoratory*, in Developments in Reliable Computing, T. Csendes, ed., Kluwer Academic Publishers, Dordrecht, 1999, pp. 77–104, http://www.ti3.tuhh.de/rump/.
[29] S. M. Rump, *Verification methods: Rigorous results using floating-point arithmetic*, Acta Numer., 19 (2010), pp. 287–449.
[30] S. M. Rump, *Verified bounds for singular values, in particular for the spectral norm of a matrix and its inverse*, BIT, 51 (2011), pp. 367–384.
[31] A. Scheel and E. S. Van Vleck, *Lattice differential equations embedded into reaction-diffusion systems*, Proc. Roy. Soc. Edinburgh Sect. A, 139 (2009), pp. 193–207.
[32] T. Wanner, *Computer-assisted bifurcation diagram validation and applications in materials science*, AMS short course lecture notes, submitted.
[33] T. Wanner, *Computer-assisted equilibrium validation for the diblock copolymer model*, Discrete Contin. Dyn. Syst. Ser. A, to appear.
[34] T. Wanner, *Topological analysis of the diblock copolymer equation*, in Mathematical Challenges in a New Phase of Materials Science, Y. Nishiura and M. Kotani, eds., Springer Proc. Math. Statist. 116, Springer, Japan, 2016, pp. 27–51.
[35] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Monogr. Numer. Anal., Oxford Science Publications, The Clarendon Press, Oxford University Press, New York, 1988.
[36] N. Yamamoto, *A numerical verification method for solutions of boundary value problems with local uniqueness by Banach's fixed-point theorem*, SIAM J. Numer. Anal., 35 (1998), pp. 2004–2013, doi:10.1137/S0036142996304498.
[37] N. Yamamoto, M. T. Nakao, and Y. Watanabe, *A theorem for numerical verification on local uniqueness of solutions to fixed-point equations*, Numer. Funct. Anal. Optim., 32 (2011), pp. 1190–1204.
[38] E. Zeidler, *Nonlinear Functional Analysis and Its Applications.* I: *Fixed-Point Theorems*, Springer-Verlag, New York, Berlin, Heidelberg, 1986.
[39] P. Zgliczyński, *Steady state bifurcations for the Kuramoto-Sivashinsky equation: A computer assisted proof*, J. Comput. Dynam., 2 (2015), pp. 95–142.