Data-driven continuation of patterns and their bifurcations

Wenjun Zhao  
Division of Applied Mathematics  
Brown University  
Providence, RI 02912, USA

Samuel Maffa  
Broad Institute of MIT and Harvard  
Cambridge, MA 02142, USA

Björn Sandstede  
Division of Applied Mathematics  
Brown University  
Providence, RI 02912, USA

April 25, 2024

Abstract

Patterns and nonlinear waves, such as spots, stripes, and rotating spirals, arise prominently in many natural processes and in reaction-diffusion models. Our goal is to compute boundaries between parameter regions with different prevailing patterns and waves. We accomplish this by evolving randomized initial data to full patterns and evaluate feature functions, such as the number of connected components or their area distribution, on their sublevel sets. The resulting probability measure on the feature space, which we refer to as pattern statistics, can then be compared at different parameter values using the Wasserstein distance. We show that arclength predictor-corrector continuation can be used to trace out transition and bifurcation curves in parameter space by maximizing the distance of the pattern statistics. The utility of this approach is demonstrated through a range of examples involving homogeneous states, spots, stripes, and spiral waves.

1 Introduction

Reaction-diffusion models, and the natural processes modelled by them, exhibit a wide range of domain-filling patterns, such as spots and stripes, and nonlinear waves, such as spiral waves [16, 27, 38]. In many applications, it is important to determine the specific region in parameter space where a given pattern or nonlinear wave is prevalent and to understand what bifurcations or transitions occur when crossing the boundaries of these regions. In this paper, we develop a framework based on pattern statistics to characterize and distinguish different patterns, such as stripes and spots, or dynamical behaviors, such as rotating, meandering, or turbulent spiral waves. We then describe how these pattern statistics can be used to compute the hyper-surfaces in parameter space that separate regions with distinct prevalent patterns or waves via continuation algorithms. We focus exclusively on the case of two-dimensional parameter spaces, so that the hyper-surfaces separating different regions are curves.

Before discussing our framework, we briefly review two existing approaches that have been designed to compute curves that separate regions with different patterns. The first approach consists of dividing parameter space into a grid with a small spacing between adjacent nodes. At each grid node, direct numerical simulations starting from random initial conditions will likely produce the pattern that is prevalent at these parameter values. Comparing the resulting patterns by eye will then provide a coarse-grained bifurcation diagram. This method is relatively
Our proposed scheme tries to merge the two approaches we reviewed above. Direct numerical simulations are inexpensive and normally relatively easy to implement. To use direct simulations within a continuation framework, we need to be able to characterize and differentiate the patterns we try to distinguish. One option is to use topological data analysis as was done in [35] for domain-filling hexagon patterns with penta-hepta defects and in [15, 32] for simulated zebrafish patterns arising in agent-based models. More generally, we can evolve a random initial condition until a fixed time that is chosen so large that patterns have emerged. We then consider an appropriate sublevel set of the solution and evaluate a feature function on the sublevel set. Examples of feature functions are the number of connected components of the sublevel set, their area distribution, or the distribution of their roundness scores (which is the ratio of area and perimeter of a connected component, thus reflecting its elongation); see Figure 1 for an illustration. In practice, sublevel sets are approximated by $\alpha$-shapes, which are polygonal approximations of the boundary of the sublevel sets [20]. We consider feature functions that map into a metric space $\mathcal{Z}$ and conduct several simulations starting from randomized initial data to obtain an empirical probability measure on $\mathcal{Z}$, which we refer to as the pattern statistics. We quantify the difference between two pattern statistics via the Wasserstein distance. We choose feature functions tailored to specific bifurcations or transitions: for instance, the number of connected components distinguishes spots and stripes, and the area of the region traced out by the tip of a spiral waves differentiates rotating from meandering spirals. Maximizing the

Figure 1: The panels illustrate (i) a color plot of the solution of the underlying PDE model at time that is so large that spots and stripes have emerged, (ii) the corresponding sublevel set, (iii) the associated $\alpha$-shape (a polygonal approximation of the boundary of the sublevel set), and (iv) the distribution (histogram) of three different feature functions evaluated on each of ten simulations that start from different random initial data, namely the areas of the connected components of the $\alpha$-shape (top), their perimeters (center), and the roundness score, which is the fraction of area over perimeter and therefore measures of how elongated each connected component is (bottom).
Figure 2: Shown is a segment of the transition curve that separates the regions in parameter space where, respectively, spots and stripes are prevalent in the Brusselator model. The feature function is given by the roundness scores of the connected components. Sample pattern statistics (given by histograms of feature evaluations of direct simulations with randomized initial data) and sample patterns are included in the insets. The colored disks correspond to parameter values where the pattern statistics was computed during continuation, with colors indicating the expectation of the roundness score.

Wasserstein distance between the pattern statistics computed at two nearby points in parameter space allows us to compute the corresponding transition curve using predictor-corrector arclength continuation; see Figure 2 for a result for the Brusselator model. The proposed methodology is purely data-driven and enables automated and efficient bifurcation tracing with limited prior knowledge of the underlying system.

Our manuscript is organized as follows. In the next section, we will introduce a probabilistic framework that serves as a theoretical foundation for pattern statistics and their computation. In particular, we will show that the pipeline from choosing randomized initial data, evolving these for a fixed time $T$, mapping to the sublevel set $u^{-1}((−\infty, c])$ for a fixed value of $c$, and evaluating a feature function on the resulting set with values in a compact metric space $Z$ will, under appropriate assumptions, yield a probability measure on the feature space $Z$, our pattern statistics, that depends continuously on parameters in the Wasserstein metric. Afterwards, we discuss how we can use pattern statistics to compute bifurcation and transition curves based on predictor-corrector algorithms that utilize bisection and quadratic interpolation. Finally, we will demonstrate the utility and applicability of the proposed continuation framework through a range of examples involving homogeneous states, spots, stripes, and spiral waves. Specifically, we will show that the approach can be used to trace out

- curves corresponding to stationary interfaces between homogeneous states, spots, and stripe patterns,
- fold bifurcations of spots and stripes,
- boundaries of snaking regions, and
- transitions between rigidly-rotating, meandering, drifting, period-doubled, and turbulent spiral waves.

We will also outline how our approach via $\alpha$-shapes can be used to compute spiral waves via the freezing method [11, 45] without the need to implement solvers for algebraic-differential systems. Finally, we will describe other potential applications and extensions to multi-parameter continuation in the discussion section.

Acknowledgments. Sandstede was partially supported by the NSF under grants DMS-2038039 and DMS-2106566.
2 Pattern statistics: A probabilistic framework

Reaction-diffusion models

We consider nonlinear reaction-diffusion systems on bounded square domains with periodic boundary conditions of the form
\[ \frac{\partial U}{\partial t} = \mathcal{D}U + \mathcal{N}(U, p), \quad x \in D := (\mathbb{R}/2\pi \mathbb{Z})^2, \quad U \in \mathbb{R}^d, \quad p \in \mathcal{P} \subset \mathbb{R}^2. \tag{2.1} \]
We note that our approach applies also to other bounded domains, but we will not consider these for simplicity.

Spaces of probability measures

We begin by briefly reviewing push-forward measures and spaces of probability measures equipped with the 2-Wasserstein distance. If \((\Omega, \sigma)\) and \((\tilde{\Omega}, \tilde{\sigma})\) are measurable spaces, \(F: \Omega \to \tilde{\Omega}\) is measurable, and \(\mu\) is a probability measure on \((\Omega, \sigma)\), then the push-forward \(F_{\#}\mu\), defined by \(F_{\#}\mu(B) := \mu(F^{-1}(B))\) for all \(B \in \tilde{\sigma}\), is a probability measure on \((\tilde{\Omega}, \tilde{\sigma})\). From now on, let \((Z, d_Z)\) be a compact metric space equipped with the Borel \(\sigma\)-algebra. If \(F: \Omega \to Z\) is measurable and \(g \in C^0(\Omega, \mathbb{R})\), then we have
\[ \int_Z g \, dF_{\#}\mu = \int_{\Omega} g \circ F \, d\mu. \tag{2.2} \]

We denote by \(\text{Prob}(Z)\) the space of probability measures on \(Z\) equipped with the Borel \(\sigma\)-algebra. We say that a sequence \((\mu_n)_{n \in \mathbb{N}}\) of measures in \(\text{Prob}(Z)\) converges weakly to \(\mu \in \text{Prob}(Z)\), denoted by \(\mu_n \rightharpoonup \mu\), if for each \(g \in C^0(\Omega, \mathbb{R})\) we have
\[ \int_Z g \, d\mu_n \to \int_Z g \, d\mu \quad \text{as } n \to \infty. \tag{2.3} \]

Given \(\mu, \nu \in \text{Prob}(Z)\), we denote by \(\Pi(\mu, \nu) := \{ \pi \in \text{Prob}(Z \times Z) : (P_1)_{\#}\pi = \mu, (P_2)_{\#}\pi = \nu\}\) the space of couplings of \((\mu, \nu)\), that is, the space of probability measures on \(Z \times Z\) with marginals \(\mu\) and \(\nu\), where \(P_j : Z \times Z \to Z\) projects onto the \(j\)th component. We equip \(\text{Prob}(Z)\) with the 2-Wasserstein distance defined by
\[ d_W(\mu, \nu) := \left( \inf_{\pi \in \Pi(\mu, \nu)} \int_{Z \times Z} d_Z(x, y)^2 \, d\pi(x, y) \right)^{\frac{1}{2}}, \quad \mu, \nu \in \text{Prob}(Z). \tag{2.4} \]

The function \(d_W\) is a metric on \(\text{Prob}(Z)\) that metrizes the weak convergence of measures on \(Z\) \([47, \text{Corollary } 6.13]\) (that is, \(\mu_n \rightharpoonup \mu\) if and only if \(d_W(\mu_n, \mu) \to 0\)), and the metric space \((\text{Prob}(Z), d_W)\) is compact \([2, \text{Theorem } 5.1.3, \text{Equation (5.1.20)}, \text{and Proposition } 7.1.5]\). For empirical measures of the form \(\mu = \frac{1}{N} \sum_{n=1}^{N} \delta_{z_n}\) and \(\nu = \frac{1}{N} \sum_{n=1}^{N} \delta_{\tilde{z}_n}\), with \(z_n, \tilde{z}_n \in Z\), the Wasserstein distance becomes the discrete optimal-transport problem
\[ d_W(\mu, \nu) = \left( \inf \left\{ \sum_{m,n=1}^{N} \Gamma_{mn} d_Z(z_m, \tilde{z}_n) : \Gamma \in [0, \infty)^{N \times N}, \sum_{j=1}^{N} \Gamma_{jn} = \sum_{j=1}^{N} \Gamma_{mj} = \frac{1}{N} \forall m, n \right\} \right)^{\frac{1}{2}}, \tag{2.5} \]

which for the case \(Z = I \subset \mathbb{R}\) gives
\[ d_W(\mu, \nu) = \left( \sum_{n=1}^{N} |z_j - \tilde{z}_j|^2 \right)^{\frac{1}{2}} \quad \text{assuming the ordering } z_1 \leq \ldots \leq z_N, \tilde{z}_1 \leq \ldots \leq \tilde{z}_N; \tag{2.6} \]

see \([37, \text{Remark } 2.28]\). Finally, when \(Z = I \subset \mathbb{R}\), the expectation
\[ E: \text{Prob}(I) \to I, \quad \mu \mapsto E(\mu) := \int_I z \, d\mu(z) \tag{2.7} \]
is continuous in the Wasserstein distance.
We focus on feature functions that operate on closed sublevel sets of, say, the first component $U_1$ of $U \in C^2(D, \mathbb{R}^d)$, that is on sets of the form $\{x \in D : U_1(x) \leq c\}$ for some threshold $c$; we note that the same results apply also to sets of the form $\{x \in D : U_1(x) \geq c\}$. To formalize our notation, choose a threshold $c \in \mathbb{R}$, let $\mathcal{X} := C^2(D, \mathbb{R})$, and define

$$\mathcal{X}_{\text{reg}} := \{u \in C^2(D, \mathbb{R}): u^{-1}(c) \neq \emptyset, \text{ and } \nabla u(x) \neq 0 \text{ for all } x \in u^{-1}(c)\}$$

(2.8)

to be the set of functions in $C^2(D, \mathbb{R})$ that have $c$ as a regular value and attain this value. Since $D$ is compact, we know that $\mathcal{X}_{\text{reg}}$ is open in $\mathcal{X}$. Throughout, we refer to sublevel sets of the form $u^{-1}((-\infty, c])$ for a function $u \in \mathcal{X}_{\text{reg}}$ as a pattern.

For completeness, we first state a result on the existence of $C^2$ tubular neighborhoods of compact one-dimensional compact $C^2$ submanifolds (or 1-manifolds, for short) of the torus. Note that each 1-manifold inside the torus is the finite disjoint union of one-dimensional manifolds that are each $C^2$-diffeomorphic to a circle.

**Lemma 2.1** Let $B$ be a one-dimensional compact $C^2$ submanifold of $D$, then there exists an open neighborhood $V$ of $B$ in $D$ and a $C^2$-diffeomorphism $\theta : B \times (-1, 1) \to V$, $(b, y) \mapsto \theta(b, y)$ so that $\theta|_{B \times \{0\}}$ is a $C^2$-diffeomorphism onto $B$. The map $\theta$ is referred to as a tubular neighborhood of $B$ in $D$.

**Proof.** From [26, Theorem 3.6 in Chapter 2], we know that there is a $C^\infty$ manifold pair $(D^\infty, B^\infty)$ with $D^\infty \subset \mathbb{R}^3$ and a $C^2$-diffeomorphism $\theta_1 : (D^\infty, B^\infty) \to (D, B)$. Next, [26, Theorem 5.2 in Chapter 4] shows that $B^\infty$ has a $C^\infty$ tubular neighborhood $\theta_2$ in $D^\infty$ so that $\theta_2 : B^\infty \times (-1, 1) \to D^\infty$ is a $C^\infty$ diffeomorphism onto an open neighborhood of $B^\infty$ in $D^\infty$ and $\theta_2|_{B^\infty \times \{0\}}$ is a diffeomorphism onto $B^\infty$. Hence, $\theta := \theta_2 \circ \theta_1$ is a $C^2$ tubular neighborhood of $B$ in $D$. ■

For $u \in \mathcal{X}_{\text{reg}}$, we denote by $A(u) := u^{-1}((-\infty, c])$ the sublevel set of $u$. Our next result shows that $A(u)$ is a $C^2$-submanifold with boundary $\partial A(u) = u^{-1}(c)$ and describes how $A(u)$ changes as $u$ varies in $\mathcal{X}_{\text{reg}}$.

**Lemma 2.2** For each $u_0 \in \mathcal{X}_{\text{reg}}$, the sublevel set $A_0 := A(u_0) = u_0^{-1}((-\infty, c])$ is a $C^2$ submanifold of $D$ with boundary $\partial A_0 = u_0^{-1}(c)$. Furthermore, there are open neighborhoods $V \subset D$ and $U \subset C^2(D, \mathbb{R})$ of $\partial A_0$ and $u_0$, respectively, and a $C^2$ map $\tau : A_0 \times U \to D$ so that for each $u \in U$ the map $\tau(\cdot, u) : A_0 \to D$ is a $C^2$-diffeomorphism from $(A_0, \partial A_0)$ onto $(A(u), \partial A(u))$ with $\tau(a, u) = a$ for all $a \in A_0 \setminus V$ and $\tau(a, u_0) = a$ for all $a \in A_0$.

**Proof.** Since $u_0 \in \mathcal{X}_{\text{reg}}$, we have $u_0^{-1}(c) \neq \emptyset$ and $\nabla u_0(b) \neq 0$ for each $b \in u_0^{-1}(c)$; hence, $u_0^{-1}(c)$ is a nonempty $C^2$ 1-manifold in $D$. Furthermore, $u_0(a)$ assumes all values near $c$ for appropriate values of $a \in D$ near $b \in u_0^{-1}(c)$, and we conclude that $u_0^{-1}((-\infty, c])$ is a $C^2$ submanifold with boundary $\partial A_0 = u_0^{-1}(c)$ as claimed.

Next, let $I = (-1, 1)$ and denote by $\theta : \partial A_0 \times I \to V$ a $C^2$ tubular neighborhood of $\partial A_0$ in $D$. Since $\nabla u_0(b) \neq 0$ for all $b \in u_0^{-1}(c)$, we can choose $\theta$ so that $\theta(\partial A_0 \times (-1, 0)) = A_0 \cap V$.

The evaluation map $\text{ev} : \partial A_0 \times I \times \mathbb{R} \times C^2(D, \mathbb{R}) \to \mathbb{R}$ given by $\text{ev}(b, y, \alpha, u) := u(h(b, y)) - \alpha \in \mathbb{R}$ is $C^2$ by [14, Theorem 10.10 in Chapter 2] or [33, Corollary 11.7]. By definition, we have $\text{ev}(b, 0, c, u_0) = 0$ with $\text{ev}_a(b, 0, c, u_0) = -1$ for all $b \in \partial A_0$, and there is a $\delta_0 > 0$ so that $|\text{ev}_a(b, 0, c, u_0)| \geq \delta_0$ for $b \in \partial A_0$ since $\partial A_0$ is compact and $\nabla u_0(b) \neq 0$ for all $b \in \partial A_0$. Hence, we can apply the implicit function theorem to the equation $\text{ev}(b, y, \alpha, u) = 0$ near each point $(b, 0, c, u_0)$ with $b \in \partial A_0$ to obtain the existence of a neighborhood $U$ of $u_0$ in $C^2(D, \mathbb{R})$, open intervals $I, J \subset \mathbb{R}$ with $0 \in I$ and $c \in J$, and unique $C^2$ maps $\psi_0 : \partial A_0 \times I \times U \to I$ and $\psi_1 : \partial A_0 \times I \times U \to J$ so that $u(h(b, y)) = \alpha$ for $(b, y, \alpha, u) \in \partial A_0 \times I \times J \times U$ and if only if $y = \psi_0(b, \alpha, u)$ or, equivalently, $\alpha = \psi_1(b, y, u)$. In particular, the maps $\psi_0(b \cdot, u)$ and $\psi_1(b \cdot, u)$ are inverses of each other for each fixed $(b, u)$, and we have $\psi_0(\cdot, c, u_0) \equiv 0$ and $\psi_1(\cdot, 0, u_0) \equiv c$. 

5
Let $\chi: J \to [0, 1]$ be a $C^\infty$ cutoff function with $\chi(0) = 0$ for $\alpha$ near $c$ and $\chi(\alpha) = 1$ for $\alpha$ near $\partial J$. We define 

$$
\tilde{\tau}: \partial A_0 \times I \times U \to \partial A_0 \times I,
(b, y, u) \mapsto (b, \psi_1(b, y, u_0), u) =: (b, \alpha, u) \mapsto (b, \psi_0(b, \alpha, u + \chi(\alpha)(u_0 - u)))
$$

so that $\tilde{\tau}$ is a $C^2$ diffeomorphism with $\tilde{\tau}(\partial A_0 \times \{0\}, u) = \theta^{-1}(\partial A(u))$ and $\tilde{\tau}(\partial A_0 \times (-1, 0], u) = \theta^{-1}(A(u) \cap V)$. For each $u$, we define the diffeomorphism $\tau(\cdot, u)$ by $\tau(\cdot, u) := \theta \circ \tilde{\tau}(\cdot, u) \circ \theta^{-1}$ on $A_0 \cap V$ and extend it to $A_0 \setminus V$ by the identity.

Next, we will use the characterization of $A(u) = u^{-1}((-\infty, c])$ provided in Lemma 2.2 to analyze feature functions.

**Feature functions**

The intuition from Lemma 2.2 is that each function that operates continuously on finite disjoint unions of connected two-dimensional $C^2$ manifolds with boundaries inside the torus $D$ defines a feature function. Our goal is to formalize this notion, give a few examples of feature functions that will be used later to distinguish spatially homogeneous states, stripe patterns, and spot patterns, and prove that these feature functions satisfy our formal definition.

We first formalize our notion of feature functions. Let $Z$ be a compact metric space. We say that a function $f: X_{reg} \to Z$ is a *feature function* provided $Z$ is a compact metric space and $f$ is continuous. Since $X_{reg}$ is open in $C^2(D, \mathbb{R})$, we can extend each feature function $f: X_{reg} \to Z$ to a measurable function $f: C^2(D, \mathbb{R}) \to Z$ by mapping $C^2(D, \mathbb{R}) \setminus X_{reg}$ to an arbitrary fixed element in $Z$.

We now introduce several functions that map $X_{reg}$ into appropriate compact metric spaces $Z$. For each two-dimensional $C^2$ submanifold $A$ with boundary of the torus $D$, we denote by $\mu_{\text{Leb}}(A)$ its Lebesgue measure. To ensure compactness of the range of some of the feature functions we introduce, we choose an $m \gg 1$ and cap some of the quantities below at $m$. We will again use the notation $A(u) := u^{-1}((-\infty, c])$ for elements $u \in X_{reg}$.

With this notation, we define the following functions:

1. $f_{\text{Conn}}(u) := \min \{ \beta(A(u)), m \} \in Z_{\text{Conn}} := \mathbb{N} \cap [0, m]$ is the zeroth Betti number $\beta(A(u))$, that is, the number of connected components of $A(u)$ (capped at $m$).

2. $f_{\text{Leb}}(u) := \mu_{\text{Leb}}(A(u)) \in Z_{\text{Leb}} := [0, \mu_{\text{Leb}}(D)]$ is the Lebesgue measure of $A(u)$.

3. $f_{\text{AreaDistr}}(u) := \frac{1}{\beta(A(u))} \sum_{j=1}^{\beta(A(u))} \delta_{\mu_{\text{Leb}}(A_j(u))}$ is the probability measure with atoms on the areas of the connected components $A_j(u)$ of $A(u)$, where $Z_{\text{AreaDistr}} := \text{Prob}(0, \mu_{\text{Leb}}(D))]$ is equipped with the 2-Wasserstein distance.

4. $f_{\text{RoundDistr}}(u) := \frac{1}{\beta(A(u))} \sum_{j=1}^{\beta(A(u))} \delta_{g(A_j(u))}$ is the probability measure with atoms on the roundness scores $g(A_j(u)) := \min \left\{ \frac{4\pi \mu_{\text{Leb}}(A_j(u))}{|\text{Perimeter of } A_j(u)|^2}, m \right\} \in [0, m]$ of the connected components $A_j(u)$ of $A(u)$ (capped at $m$), where $Z_{\text{RoundDistr}} := \text{Prob}(0, m]$ is equipped with the 2-Wasserstein distance.

For each connected $C^2$ submanifold $A$ with boundary in $D$, the individual roundness score $g(A)$ will be close to zero when $A$ is an elongated stripe, while it will be close to one when $A$ is close to a regular disk. The next lemma shows that the functions defined in (1)-(4) are indeed feature functions.

**Lemma 2.3** The functions defined in (1)-(4) are continuous from $X_{reg}$ into their respective ranges and therefore define feature functions.

**Proof.** We established in Lemma 2.2 that $(A(u), \partial A(u))$ is diffeomorphic to $(A(u_0), \partial A(u_0))$ for all $u$ close to $u_0$ in $X_{reg}$ via the diffeomorphism $\tau(\cdot, u)$ with $\tau(\cdot, u_0) = \text{id}$. In particular, the number $f_{\text{Conn}}(u)$ of connected components of $A(u)$ is locally constant in $X_{reg}$, and therefore continuous. Similarly, we can combine the diffeomorphism
If $Z = \text{Prob}(I)$ for a bounded interval $I \subset \mathbb{R}$, and $f : \mathcal{X}_{\text{reg}} \to Z$ is a feature function, then its composition $E \circ f : \mathcal{X}_{\text{reg}} \to I$ with the expectation defined in (2.7) is also a feature function. In particular, we can define the feature functions $f_{E, \text{AreaDist}} := E \circ f_{\text{AreaDist}}$ and $f_{E, \text{RoundDist}} := E \circ f_{\text{RoundDist}}$ which map a pattern to the expectation of, respectively, the area and roundness score distributions of its connected components.

Randomization of initial data

Our next step is to construct randomized initial data that generate ensembles of solutions on which we can evaluate a given feature function. Recall that we pose the reaction-diffusion system (2.1) on the torus $(\mathbb{R}/2\pi \mathbb{Z})^2$. From now on, we use the notation $H^2(D) := H^2(D, \mathbb{R}^d)$ and $\ell^2(\mathbb{Z}^2) := \ell^2(\mathbb{Z}^2, \mathbb{R}^d)$. We represent initial conditions $U \in H^2(D)$ via their Fourier series. Let $e_k(x) = \frac{1}{\sqrt{2\pi}} e^{ikx}$ with $k \in \mathbb{Z}^2$ be the standard orthonormal Fourier-series basis of $H^2(D)$.

First, we choose two functions $U_1, U_2 \in H^2(D)$: the base function $U_2$ will represent the mean of the randomized initial conditions, while $U_1$ will be used for the actual randomization. We write $U_1 = \sum_{k \in \mathbb{Z}^2} a_k e_k(x) \in H^2(D)$ so that $(a_k)_{k \in \mathbb{Z}^2} \in \ell^2(\mathbb{Z}^2)$ satisfies $|U_1|_{H^2(D)} = |(a_k)|_{\ell^2(\mathbb{Z}^2)}$.

Following [12, 36], we now randomize the fixed function $U_1$. We choose a probability space $(\Omega, \sigma, \mu_1)$ and a sequence of independent, zero-mean $\mathbb{R}^d$-valued random variables $(b_k)_{k \in \mathbb{Z}^2}$ with the property that there is a constant $C_0 > 0$ so that

$$\int_{\Omega} |b_k(\omega)|^2 \, d\mu_1 \leq C_0 \quad \text{for all } k \in \mathbb{Z}^2. \quad (2.9)$$

We define

$$\xi : \Omega \to \ell^2(\mathbb{Z}^2), \quad \omega \mapsto (a_k \odot b_k(\omega))_{k \in \mathbb{Z}^2}, \quad (2.10)$$

where $\odot$ denotes elementwise multiplication of vectors in $\mathbb{R}^d$ (that is, for $a, b \in \mathbb{R}^d$ we set $a \odot b := (a_j b_j)_{j=1,\ldots,d} \in \mathbb{R}^d$). Our first result shows that $\xi$ is a well-defined measurable function and therefore defines an $\ell^2$-valued random variable.

**Lemma 2.4** The map $\xi$ defined in (2.10) is measurable with $\xi \in L^2(\Omega, \ell^2(\mathbb{Z}^2))$.

**Proof.** For each $N \geq 1$, we define $\xi_N : \Omega \to \ell^2$ by $[\xi_N(\omega)]_k = \mathbf{1}_{|k| \leq N} a_k \odot b_k(\omega)$ so that $[\xi_N(\omega)]_k = 0$ for all $|k| > N$. These functions are measurable with $\xi_N \in L^2(\Omega, \mathcal{L}^2)$ for each $N$. Using (2.9), we have for each $M \geq N$ that

$$\int_{\Omega} |\xi_M(\omega) - \xi_N(\omega)|^2 \, d\mu_1 \leq \sum_{k=N+1}^{M} \int_{\Omega} |a_k \odot b_k(\omega)|^2 \, d\mu_1 \leq \sum_{k=N+1}^{M} |a_k|^2 \int_{\Omega} |b_k(\omega)|^2 \, d\mu_1 \leq C_0 \sum_{k=N+1}^{M} |a_k|^2 \to 0$$

as $M, N \to \infty$. Hence, $\xi_N \in L^2(\Omega, \mathbb{L}^2)$ is a Cauchy sequence, and we conclude that there is a function $\xi \in L^2(\Omega, \ell^2)$ so that $\xi_N \to \xi$ in $L^2$ as $N \to \infty$. Furthermore, there is a subsequence $(N_n)_n$ so that $\xi_{N_n}(\omega) \to \xi(\omega)$ as $n \to \infty$ for almost every $\omega \in \Omega$, which shows that $\xi$ is given by (2.10) almost everywhere. 

We now use the linear isomorphism $\iota : \ell^2(\mathbb{Z}^2) \to H^2(D)$ provided by the Fourier-series expansion to define the map

$$\iota \circ \xi : \Omega \to H^2(D), \quad \omega \mapsto U^\omega := U_0 + U_1^\omega \quad \text{with } U_1^\omega := \sum_{k \in \mathbb{Z}^2} (a_k \odot b_k(\omega))e_k(\omega). \quad (2.11)$$
Note that \( \iota \circ \xi \) is measurable and defines a \( H^2(D) \)-valued random variable that lies in \( L^2(\Omega, H^2(D)) \).

Finally, we remark that some of the transition curves we discuss below can be computed efficiently using a single deterministic initial condition \( U_b \) instead of a randomization of initial data. These cases still fall into the framework discussed above upon defining \( \Omega \) to consist of a single point and \( \xi \) to map this point to \( U_b \).

**Ensembles of solutions**

Next, we propagate the randomized initial conditions forward in time to create ensembles of solutions that are parametrized by \( \omega \in \Omega \). We fix a time \( T > 0 \) and note that the semiflow \( \Phi_T(U_b, p) \) of the reaction-diffusion system (2.1) is a smooth map from \( H^2(D) \times \mathcal{P} \) into \( C^2(D) \). We denote by \( P_1: C^2(D) \to \mathcal{X} = C^2(D, \mathbb{R}) \), \( U = (U_1, \ldots, U_d) \mapsto U_1 \) the continuous projection onto the first component. The ensemble function

\[
E_p := P_1 \circ \Phi_T(\cdot, p) \circ \iota: \ell^2(\mathbb{Z}^2) \to \mathcal{X}
\]

then maps initial data in Fourier space to the first component of the solution evaluated at time \( T \), and we see that the composite map

\[
E_p \circ \xi: \Omega \to \mathcal{X} = C^2(D, \mathbb{R}), \quad \omega \mapsto P_1 \Phi_T(U^{\omega'}, p)
\]

is measurable for each \( p \in \mathcal{P} \).

Recall that feature functions are assumed to be continuous only on the subset \( \mathcal{X}_{\text{reg}} \) of functions in \( C^2(D, \mathbb{R}) \) that attain \( c \) as a regular value. We will therefore assume that our randomization (consisting of our choices of \( U_b, U_t \in H^2(D) \) and the random variables \( (b_k)_{k \in \mathbb{Z}^2} \)), the constant \( c \) appearing in the definition (2.8) of \( \mathcal{X}_{\text{reg}} \), and the time \( T > 0 \) can be chosen so that \( E_p \circ \xi \) maps almost surely into \( \mathcal{X}_{\text{reg}} \) for each \( p \in \mathcal{P} \).

We formalize this equivalently as follows. Since \( E_p \circ \xi \) is measurable for each \( p \), the push-forward \( \mu_\mathcal{X}(p) := (E_p \circ \xi)\# \mu_\Omega \) is a well-defined probability measure on \( \mathcal{X} = C^2(D, \mathbb{R}) \) equipped with the Borel \( \sigma \)-algebra. We assume that the following hypothesis is met:

**Hypothesis (H1)** Assume that \( (b_k)_{k \in \mathbb{Z}^2} \) are independent, zero-mean \( \mathbb{R}^d \)-valued random variables that satisfy (2.9). Furthermore, assume that these random variables, the functions \( U_b, U_t \in H^2(D) \), and the constants \( c \in \mathbb{R} \) and \( T > 0 \) are such that \( \mu_\mathcal{X}(p)(\mathcal{X}_{\text{reg}}) = 1 \) for each \( p \in \mathcal{P} \).

Our hypothesis essentially assumes that the set of \( \omega \) for which the sublevel set of the solution at time \( T \) is not a manifold (and instead undergoes a bifurcation) has measure zero. For fixed \( p \), these bifurcations should occur at most along codimension-one sets, so the hypothesis should be satisfied for generic systems (and generic choices of the quantities mentioned in (H1)). We will assume from now on that (H1) is met.

**Pattern statistics**

For each given feature function \( f: \mathcal{X}_{\text{reg}} \to \mathcal{Z} \), where \( \mathcal{Z} \) is a compact metric space, the composition

\[
f \circ E_p \circ \xi: \Omega \to \mathcal{Z}
\]

is measurable. We define the *pattern statistics* to be the map

\[
\mu_f: \mathcal{P} \to \text{Prob}(\mathcal{Z}), \quad p \mapsto \mu_f(p) := (f \circ E_p \circ \xi)\# \mu_\Omega
\]

that associates to each \( p \in \mathcal{P} \) the push-forward probability measure of \( \mu_\Omega \) under \( f \circ E_p \circ \xi \). Thus, \( \mu_f(p) \) is the distribution of features of the ensemble of patterns generated by the randomized initial conditions at the parameter value \( p \in \mathcal{P} \). If the feature space \( \mathcal{Z} \) is a compact interval in \( \mathbb{R} \), we can also define the *feature mean*

\[
E_f(p) \in \mathbb{R} \quad p \mapsto E_f(p) := E(\mu_f(p)) = \int_{\mathcal{Z}} z \, d\mu_f(z; p) \in \mathcal{Z}.
\]
Our next result shows that \( \mu_f \) and \( E_f \) are continuous when we equip \( \mathcal{P} \) and \( \mathbb{R} \) with the standard Euclidean metric and the space \( \text{Prob}(Z) \) with the 2-Wasserstein metric \( d_W \) defined in (2.4).

Lemma 2.5 Assume that Hypothesis (H1) is met, then the map \( \mu_f : (\mathcal{P}, d_{\text{Eucl}}) \to (\text{Prob}(Z), d_W) \) is continuous. Furthermore, if \( Z \subset \mathbb{R} \) is a compact interval, then the map \( E_f : \mathcal{P} \to Z \) is continuous.

Proof. We focus first on continuity of \( \mu_f \). Using the results we quoted in our Digression, it suffices to show that \( \mu_f(q) \to \mu_f(p) \) at \( q \to p \) in \( \mathcal{P} \). Using the definition

\[
F(\omega;p) := (f \circ \xi_p \circ \xi)(\omega)
\]

we need to prove that for each fixed choice of \( g \in C^0(Z, \mathbb{R}) \) we have

\[
\int_Z g(z) \, d\mu_f(z; p) = \int_Z g(F(\omega; q)) \, d\mu_\Omega \to \int_Z g(F(\omega; p)) \, d\mu_\Omega = \int_Z g(z) \, d\mu_f(z; p) \quad \text{as} \quad q \to p \in \mathcal{P}. \tag{2.15}
\]

Thus, fix \( g \in C^0(Z, \mathbb{R}) \) and let \( m := \max_{z \in Z} |g(z)| \). Pick \( \epsilon > 0 \) and define \( X_\delta := \chi \setminus \chi_{\text{reg}} \). Since \( \mu_X(p)(X_\delta) = 0 \) by Hypothesis (H1), there is a \( \delta > 0 \) so that \( \mu_X(p)(U_\delta(X_\delta)) \leq \frac{\epsilon}{4m} \), where \( U_\delta(A) \) denotes the \( \delta \)-neighborhood of a set \( A \) in \( \chi \). Hence, by definition, \( \Omega_\delta := (\xi_p \circ \xi)^{-1}(U_\delta(X_\delta)) \) satisfies \( \mu_\Omega(\Omega_\delta) \leq \frac{\epsilon}{4m} \). We set \( \Omega_\delta^\circ := \Omega \setminus \Omega_\delta \) and claim that for each fixed \( \omega \in \Omega_\delta^\circ \), the function \( g(F(\omega; q)) \) is continuous in \( q \) for all \( q \in \mathcal{P} \) near \( p \): this claim is true since \( (\xi_p \circ \xi)(\omega) \) is continuous in \( q \), the image lies in \( \chi_{\text{reg}} \) for all \( q \) near \( p \) since \( \omega \in \Omega_\delta \), and \( g \circ f \) is, by definition, continuous on \( \chi_{\text{reg}} \). Lebesgue’s dominated convergence theorem therefore implies that there is a \( \tilde{\delta} > 0 \) so that

\[
\int_{\Omega_\delta^\circ} |g(F(\omega; q)) - g(F(\omega; p))| \, d\mu_\Omega < \frac{\epsilon}{2} \quad \text{for all} \quad q \in \mathcal{P} \quad \text{with} \quad |q - p| < \tilde{\delta}.
\]

On \( \Omega_\delta \), we have

\[
\int_{\Omega_\delta} |g(F(\omega; q)) - g(F(\omega; p))| \, d\mu_\Omega \leq 2m \mu_{\Omega}(\Omega_\delta) < \frac{\epsilon}{2} \quad \text{for all} \quad q \in \mathcal{P}.
\]

We conclude that

\[
\int_{\Omega} |g(F(\omega; q)) - g(F(\omega; p))| \, d\mu_\Omega < \epsilon \quad \text{for all} \quad q \in \mathcal{P} \quad \text{with} \quad |q - p| < \tilde{\delta},
\]

which establishes weak convergence of \( \mu_f(q) \). Finally, if \( Z \) is an interval in \( \mathbb{R} \), the expression for \( E_f(p) \) from (2.14) coincides with the left-hand side of (2.15) with \( g(z) = z \), and continuity of \( E_f(p) \) therefore follows from the arguments for weak convergence given above. \qed

Empirical measures

In practice, given a reaction-diffusion model (2.1) and a feature function \( f \), we will not be able to compute the resulting pattern statistics analytically. Instead, we will approximate the pattern statistics numerically using empirical measures. In (2.10), we defined the function \( \xi : \Omega \to \ell^2(\mathbb{Z}^2) \) that provided the randomization of initial conditions. The empirical measure will be based on a sequence of random variables with the same distribution as \( \xi \).

Hypothesis (H2) Assume that \( (\xi^n)_{n \in \mathbb{N}} \) is a sequence of independent, identically distributed random variables \( \xi^n : \Omega \to \ell^2(\mathbb{Z}^2) \) with \( \xi^n \in L^2(\Omega, \ell^2(\mathbb{Z}^2)) \) that satisfy \( \xi^n_{\#} \mu_\Omega = \xi_{\#} \mu_\Omega \) for all \( n \).

We set \( \xi^{n,\omega} := \xi^n(\omega) \). For each fixed \( \omega \in \Omega \) and each \( N \geq 1 \), we then define the empirical measure

\[
\mu_{\xi^2}^{N,\omega} := \frac{1}{N} \sum_{n=1}^{N} \delta_{\xi^{n,\omega}}
\]
on the space $\ell^2(\mathbb{Z}^2)$. We can think of each empirical measure $\mu_{f,\omega}^N$ as arising from drawing $N$ independent
samples from the space $\Omega$ with the measure $\mu_\Omega$ and constructing the resulting push-forward measure on $\ell^2(\mathbb{Z}^2)$
under the map $\xi$. For each fixed $N \in \mathbb{N}$ and $\omega \in \Omega$, we can then define the empirical pattern statistics via

$$\mu_{f,\omega}^N : \mathcal{P} \rightarrow \operatorname{Prob}(Z), \quad p \mapsto \mu_{f,\omega}^N(p) := (f \circ E_p)_\# \mu_{x,\omega}^N = \frac{1}{N} \sum_{n=1}^N \delta((f_\circ E_p)(\xi^n,\omega)). \quad (2.16)$$

As before, if the feature space $Z$ is a compact interval in $\mathbb{R}$, the empirical feature mean $E_{f,\omega}^N(p) \in \mathbb{R}$ is given by

$$E_{f,\omega}^N : \mathcal{P} \rightarrow \mathbb{R}, \quad p \mapsto E_{f,\omega}^N(p) := \int_Z z \, d\mu_{f,\omega}^N(z;p) = \frac{1}{N} \sum_{n=1}^N (f \circ E_p)(\xi^n,\omega) \in Z. \quad (2.17)$$

Our next result shows that for almost every $\omega \in \Omega$ these statistics converge to the full statistics as $N \to \infty$.

**Lemma 2.6** Assume Hypotheses (H1)-(H2) are met, then for each $p \in \mathcal{P}$ the sets

$$\Omega_1 := \{ \omega : d_W(\mu_{f,\omega}^N(p), \mu_f(p)) \rightarrow 0 \text{ as } N \to \infty \} \quad \text{and} \quad \Omega_2 := \{ \omega : \left| E_{f,\omega}^N(p) - E_f(p) \right| \rightarrow 0 \text{ as } N \to \infty \}$$

satisfy $\mu_\Omega(\Omega_j) = 1$ for $j = 1, 2$. Furthermore, for almost every $\omega \in \Omega$ we have that $\mu_{f,\omega}^N(p)$ and $E_{f,\omega}^N(p)$ are continuous in $p$ for each $N \in \mathbb{N}$.

**Proof.** Weak convergence of empirical measures for almost every $\omega$ was established in [46, Theorem 3], which yields convergence in the 2-Wasserstein distance by [47, Corollary 6.13]. As in the proof of Lemma 2.5, weak convergence of $\mu_{f,\omega}^N(p)$ implies convergence of $E_{f,\omega}^N(p)$. It remains to prove continuity in $p$. We use again the notation $X_\omega := X \setminus X_{\omega,\text{reg}}$. Since the random variables $\xi^n$ have the same distribution as $\xi$, the set

$$\Omega_3 := \{ \omega \in \Omega : E_p(\xi^n,\omega) \in X_\omega \text{ for some } n \geq 1 \} = \bigcup_{n \geq 1} \{ \omega \in \Omega : E_p(\xi^n,\omega) \in X_\omega \} \text{ has measure zero by (H1)}$$

has measure zero for each fixed $p \in \mathcal{P}$. For each $\omega \in \Omega_3$, we can now proceed as in Lemma 2.5 to prove continuity in $p$ for each $N \geq 1$. \qed

**Objective functions**

Next, we compare the pattern statistics for different parameter values using the Wasserstein distance. We define the objective function $G_f$ via

$$G_f : \mathcal{P} \times \mathcal{P} \rightarrow \mathbb{R}, \quad (p, q) \mapsto G_f(p, q) := d_W(\mu_f(p), \mu_f(q)) \quad (2.18)$$

and the empirical objective function $G_f^N$ via

$$G_f^N : \mathcal{P} \times \mathcal{P} \rightarrow \mathbb{R}, \quad (p, q) \mapsto G_f^N(p, q) := d_W(\mu_{f,\omega}^N(p), \mu_{f,\omega}^N(q)). \quad (2.19)$$

In the empirical objective function, we use the same number of samples in both arguments as this allows us to use the discrete optimal-transport formulation (2.5) for the Wasserstein distance. In contrast, we will typically evaluate the two empirical measures in the argument at different elements $\omega, \tilde{\omega} \in \Omega$. If $Z = [0, m]$, we will also use the objective functions

$$G_{E_f} : \mathcal{P} \times \mathcal{P} \rightarrow \mathbb{R}, \quad (p, q) \mapsto G_{E_f}(p, q) := |E_f(p) - E_f(q)| \quad (2.20)$$

$$G_{E_f}^N : \mathcal{P} \times \mathcal{P} \rightarrow \mathbb{R}, \quad (p, q) \mapsto G_{E_f}^N(p, q) := \frac{1}{N} \sum_{n=1}^N (f \circ E_p)(\xi^n,\omega) - \frac{1}{N} \sum_{n=1}^N (f \circ E_q)(\xi^n,\tilde{\omega}) \quad (2.21)$$

that compare the feature means at different parameter values. Lemmas 2.5 and 2.6 show that the objective functions depend continuously on $p$ (almost surely in the case of empirical objective functions) and that $G_f^N$ and $G_{E_f}^N$ converge to $G_f$ and $G_{E_f}$, respectively, as $N \to \infty$ almost surely.
Bags and empirical pattern statistics

For the case where \( Z = \text{Prob}([0, m]) \), we can also use an empirical pattern statistics that aggregates feature values of connected pattern components across samples. In this case, the empirical pattern statistics

\[
\mu^{N,\omega}_f(p) = \frac{1}{N} \sum_{n=1}^{N} \delta_{(f \circ E_p)(\xi^n,\omega)} \in \text{Prob}(Z)
\]

has support on the elements \((f \circ E_p)(\xi^n,\omega) \in Z = \text{Prob}([0, m])\). We assume that these elements are finite sums of \(\delta\)-functions so that

\[
(f \circ E_p)(\xi^n,\omega) = \frac{1}{m_p} \sum_{j=1}^{m_p} \delta_{a_j^n,\omega} \in Z = \text{Prob}([0, m]), \quad a_j^n,\omega \in [0, m]
\]

(2.21)
as is the case for the feature functions \(f_{\text{AreaDistr}} \) and \(f_{\text{RoundDistr}} \) that we defined earlier. We use the collection (the "bag") of feature values \(a_{j,p}^n,\omega \in [0, m] \) across the probability measures \((f \circ E_p)(\xi^n,\omega) \) on \([0, m] \) to define the probability measure

\[
\mu^{N,\omega}_{f,\text{bag}}(p) := \frac{1}{N_{\text{bag}}(p)} \sum_{n=1}^{N} \sum_{j=1}^{m_p} \delta_{a_j^n,\omega} \in Z = \text{Prob}([0, m]), \quad N_{\text{bag}}(p) := \sum_{n=1}^{N} m_p^n
\]

(2.22)
on \([0, m] \). The empirical measure \(\mu^{N,\omega}_{f,\text{bag}}(p) \) can therefore be thought of as the distribution of the feature values of the individual connected components collected from the multiset of the \(N \) pattern samples. The objective function

\[
G^N_{f,\text{bag}}(p, q) = d_W \left( \mu^{N,\omega}_{f,\text{bag}}(p), \mu^{N,\omega}_{f,\text{bag}}(q) \right)
\]

compares the resulting empirical measures using the Wasserstein distance on \( Z = \text{Prob}([0, m]) \).

Note that empirical pattern statistics \(\mu^{N,\omega}_{f,\text{bag}}(p) \) cannot, to our knowledge, be interpreted as sampling from an underlying distribution on \( \text{Prob}(Z) \). Furthermore, even though the sequence \(\mu^{N,\omega}_{f,\text{bag}}(p) \) will have convergent subsequences as \(N \to \infty \) due to compactness of \( Z \), it is not clear whether the sequence itself converges and what the weak limit would be.

Illustration and intuition

We illustrate the concepts we introduced above through a very simple example. Consider a reaction-diffusion system at a single parameter value and assume that the system generates only two distinct patterns \(A\) and \(B\). We assume that \(A\) has three connected components each with area \(a\), while \(B\) is connected with area 1. We encounter the set \(A\) with probability \(\rho \in [0, 1]\) and the set \(B\) with probability \(1 - \rho\). We set \(I := [0, m]\) and focus on the feature function \(f_A := f_{\text{AreaDistr}}\) that associates to a pattern \(C\) the probability measure \(f(C) = \frac{1}{\beta(C)} \sum_{j=1}^{\beta(C)} \delta_{|C_j|}\) in \(Z = \text{Prob}(I)\) with atoms on the areas \(|C_j|\) of the connected components \(C_j\) of the pattern \(C\), and its expectation \(f_{E_A} := f_{E_{\text{AreaDistr}}} = \mathbb{E}(f_{\text{AreaDistr}}) = \frac{1}{\beta(C)} \sum_{j=1}^{\beta(C)} |C_j| \) in \(I\).

We have \(f_A(A) = \frac{1}{3} \sum_{j=1}^{3} \delta_a = \delta_a\) and \(f_A(B) = \delta_1\) as well as \(f_{E_A}(A) = \rho\) and \(f_{E_A}(B) = 1\). The pattern statistics \(\mu_{f_A}\) associated with the feature \(f_A\) is given by

\[
\mu_{f_A} = \rho \delta_{f_A(A)} + (1 - \rho) \delta_{f_A(B)} = \rho \delta_a + (1 - \rho) \delta_1 \in \text{Prob}(\text{Prob}(I)).
\]

Furthermore, the feature mean of \(f_{E_A}\) is given by \(E_{f_{E_A}} = \rho f_{E_A}(A) + (1 - \rho) f_{E_A}(B) = \rho a + (1 - \rho)\), which provides the expectation of the average area of a single connected component in each individual pattern sample.

Next, we consider the empirical measure \(\mu^N_{f_A,\text{bag}}\) defined in (2.22) for the area distribution \(f_A\) from \(N\) sampled patterns with \(N \gg 1\). We obtain

\[
\mu^N_{f_A,\text{bag}} = \frac{1}{1 + 2\rho} (3\rho \delta_a + (1 - \rho) \delta_1) \in \text{Prob}(I), \quad E(\mu^N_{f_A,\text{bag}}) = \frac{1}{1 + 2\rho} (3\rho a + (1 - \rho)) = \frac{1 + \rho(3a - 1)}{1 + 2\rho} \in I
\]
for the empirical measure and its expectation. The expectation $E(\mu^N_{f, \text{bag}})$ gives the expected area of individual connected components in the multiset (the "bag") of all connected components across $N$ samples of the patterns $A$ and $B$. Note that $E(\mu^N_{f, \text{bag}})$ and $E_{\mu^N}$ are not equal, since the former corresponds to the expected area in the multiset of all connected components across all patterns, while the latter captures the expected area for the connected components in a typical pattern.

Spatial discretization

We now discuss the numerical computation of pattern statistics, focusing first on the spatial discretization of the reaction-diffusion system (2.1) and then on the numerical approximation of patterns and sublevel sets using $\alpha$-shapes. To discretize (2.1) in space, we fix $K$ and approximate solutions of (2.1) via the orthogonal projection $\mathcal{Q}_K$ onto the closed subspace $H^2_K(D) := \{ U = \sum_{k \in \mathbb{Z}^2, |k| \leq K} a_k e_k \}$ of $H^2(D)$ so that (2.1) becomes

$$U_t = \Delta U + \mathcal{Q}_K N(U, p), \quad U \in H^2_K(D). \quad (2.23)$$

The solutions (2.23) are of the form $U_K(t) = \sum_{k \in \mathbb{Z}^2, |k| \leq K} a_k(t)e_k \in H^2_K(D) \subset H^2(D)$, and we write $\Phi^K_T(U_K(0), p) := U_K(T)$. We can use the randomization of initial conditions we introduced before projected by $\mathcal{Q}_K$ onto the set of Fourier coefficients corresponding to wavenumbers $k$ with $|k| \leq K$. The remaining part of the framework remains unchanged since $\Phi^K_T(U, p)$ still lies in $H^2(D)$. For each fixed $U_0 \in H^2(D)$, we have $\mathcal{Q}_K(\mathcal{Q}_K U_0, p) \to \Phi_T(U_0, p)$ in $H^2(D)$ as $K \to \infty$ (see [22, 24]). We note that choosing our randomized initial data from $H^2(D)$ instead if $H^2(D)$ would also ensure regularity in $C^2(D)$. Using these convergence properties, we can proceed analogously to Lemma 2.5 to show that the pattern statistics associated with (2.23) converges weakly, and hence also in the Wasserstein distance, to the statistics associated with (2.1) as $K \to \infty$. In particular, discretization of the reaction-diffusion system allows us to faithfully approximate the pattern statistics of the full system.

Computation of feature functions via $\alpha$-shapes

Next, we discuss the numerical computation of sublevel sets. We choose $M \geq 1$ and define the finite lattice $D_M$ consisting of $M^2$ equally spaced points in the domain $D = (\mathbb{R}/2\pi \mathbb{Z})^2$. For $u \in \mathcal{X}_{\text{reg}}$, the sublevel set $A := \{ x \in D : u(x) \leq c \}$ is a 2-manifold with boundary of class $C^2$. The pattern belonging to $u$ on $D_M$ is then given by the discrete set $A_M := A \cap D_M = \{ x \in D_M : u(x) \leq c \}$ of points $x$ on the lattice $D_M$ for which $u(x) \leq c$. We will use $\alpha$-shapes to approximate the number of connected components of $A$, the length $\lambda(\partial A)$ of its boundary, and its area $\mu_{\text{Leb}}(A)$.

Alpha-shapes are defined as follows (see also Figure 3 for an illustration). For fixed $\alpha > 0$, the $\alpha$-shape $S_\alpha(A_M)$ of $A_M$ is a disjoint union of polygons, whose edges are defined as follows [20]. Two elements $x_i, x_j \in A_M$ form an edge in $S_\alpha(A_M)$ if and only if there is an open ball $U_\alpha$ of radius $\alpha$ so that $x_i, x_j \in \partial U_\alpha$ and $A_M \cap U_\alpha = \emptyset$. 

Figure 3: We illustrate the definition of $\alpha$-shapes and their dependence on the radius $\alpha$. Panel (i) shows the original data set. The associated $\alpha$-shapes for radii $\alpha_1$ and $\alpha_2$ with $\alpha_1 < \alpha_2$ are shown as polygons in panels (ii) and (iii), respectively. The circles $\partial U_{\alpha_1,2}$ determine which data points are connected by edges.
As shown in [20], there is a unique connected component of $\mathbb{R}^2 \setminus S_\alpha(A_M)$ that contains $A_M$, and we refer to this component as the interior face $\hat{S}_\alpha(A_M)$ of $S_\alpha(A_M)$. In preparation for the next lemma, we say that a set $A$ satisfies the r-rolling condition if for each $x \in A$ there is an open ball $U_r$ of radius $r$ so that $x \in \partial U_r$ and $A \cap U_r = \emptyset$. It was shown in [48, Theorem 1] that if $A$ is a compact 2-manifold of class $C^2$, then there is an $r > 0$ so that $A$ and $A^c$ satisfy the r-rolling condition: we denote this radius by $r(A)$.

**Lemma 2.7** Let $A$ be a 2-manifold with $C^2$ boundary in $D$, and fix $\alpha \in (0, r(A))$, then there are constants $C_0, M_0$ that depend only on $(\alpha, r(A))$ so that the following is true for all $M \geq M_0$. The set $A$ and the interior face $\hat{S}_\alpha(A_M)$ have the same number $\beta(A)$ of connected components, denoted by $A_j$ and $\hat{S}_\alpha^j(A_M)$, respectively, and these can be labeled so that $A_j \cap D_M \subset \hat{S}_\alpha^j(A_M)$ for $j = 1, \ldots, \beta(A)$. Furthermore, we have

$$
\left| \mu_{\text{Leb}}(A) - \mu_{\text{Leb}}(\hat{S}_\alpha(A_M)) \right| + \max_{j=1,\ldots,\beta(A)} |\lambda(\partial A_j) - \lambda(S_\alpha^j(A_M))| \leq \frac{C_0}{M} \quad (2.24)
$$

for each $M \geq M_0$.

**Proof.** The lemma follows from the results in [3]. While that paper focused primarily on the case of independent samples from the uniform distribution in $A$, many of the results are for the deterministic case, and we will now show how they can be used to prove our claims. We will be very brief and refer to [3] for details and notation.

First, we will show that the set $H_{ij,t} \cap G_{ij}$ considered in [3, Proposition 5] is not just of small measure but is indeed empty in our situation provided $t \geq 8\pi C_1/M$, where $C_1 = C_1(\alpha, r(A))$ denote the constant defined in [3, Lemma 5]. Assume that $x_1, x_2$ form an edge in $S_\alpha(A_M)$ so that $x_1, x_2 \in U_\alpha(z)$ for $z \notin S$, where $U_\alpha(z)$ denotes the open ball of radius $\alpha$ centered at $z$. If $\alpha - d(z, S) \geq 8\pi/M$, then we can follow the arguments in [3, Lemma 5] to show that $U_\alpha(z) \cap A_M \neq \emptyset$, since each square of length $4\pi/M$ inside $A$ necessarily contains an element in $A_M$. In particular, if $t \geq 8\pi C_1/M$ and $0 < d(z, S) < \alpha - t/C_1$, then we have $8\pi/M \leq t/C_1 < \alpha - d(z, S)$ and therefore $U_\alpha(z) \cap A_M \neq \emptyset$, which shows that the set $H_{ij,t} \cap G_{ij}$ considered in [3, Proposition 5] is empty. The arguments in [3, §4] now show that $|\lambda(\partial A) - \lambda(S_\alpha(A_M))| \leq C_0 t$ for all $t$ with $t \geq 8\pi C_1/M$. Choosing $t = 8\pi C_1/M$ completes the proof of the perimeter estimate. Using the fact that [3, Propositions 1 and 5] are true in our situation provided $t \geq 8\pi C_1/M$, the statements in [3, Propositions 5 and 6] also hold and show that the distance between $\partial A$ and $S_\alpha(A_M)$ is bounded by $C_2/M$, where $C_2$ depends only on $(\alpha, r(A))$. This fact can now be used to establish the estimate for the difference of the areas of $A$ and $\hat{S}_\alpha(A_M)$ and to show the statements about their connected components. □

In summary, the pattern statistics of (2.1) is accurately approximated under spatial discretization provided $K \gg 1$ is large enough. Similarly, the features of each fixed sublevel set are approximated at order $O(1/M)$ when we compute them using the $\alpha$-shapes on the lattice $D_M$ with $M^2$ points: we remark that we do not have uniformity in $A = u^{-1}((-\infty, c])$, since $u$ may be arbitrarily close to $\mathcal{X} \setminus \mathcal{X}_{\text{reg}}$ where $c$ is no longer a regular value.

3 Tracing bifurcations using pattern statistics

We build on the framework outlined in §2 to trace out curves in the two-dimensional parameter space $\mathcal{P} \subset \mathbb{R}^2$ that separate regions in parameter space with different prevailing patterns. If $f$ is a feature function that can distinguish the patterns we are interested in, we will argue in this section that the objective function $G_f(p, q) = d_W(\mu_f(p), \mu_f(q))$, which measures the difference between the associated pattern statistics $\mu_f$ at different parameter values, can be used to characterize and compute transition curves using predictor-correct continuation.
Bifurcation functions

We illustrate the concepts by focusing first on the feature function \( f = f_{\text{Conn}} \) that counts the connected components in a given pattern. This feature function will distinguish spots and stripes. Assume that \( \gamma : \mathbb{R} \rightarrow \mathcal{P} \in C^1 \) is a curve so that its trace \( \Gamma = \{ \gamma(t) : t \in \mathbb{R} \} \subset \mathcal{P} \) separates regions of parameter space where each of these patterns prevails; see Figure 4(i) for an illustration. As indicated in Figure 4(ii), the expectation \( E(\mu_f) \) changes most rapidly near \( \gamma(s) \). Alternatively, we can choose a small offset \( 0 < h \ll 1 \) and consider the rate of change of \( E(\mu_f(p)) \) measured by the slope

\[
\frac{1}{2h} |E(\mu_f(p + hn(s))) - E(\mu_f(p - hn(s)))|
\]

of the secant of the graph of \( E(\mu_f(p)) \): this slope will be largest at \( p = \gamma(s) \), reflecting the fact that \( E(\mu_f(p)) \) changes most rapidly near \( \gamma(s) \). Thus, we can compute \( \gamma(s) \) via

\[
\gamma(s) = \arg \max_{p \in L(s)} \frac{1}{2h} |E(\mu_f(p + hn(s))) - E(\mu_f(p - hn(s)))|.
\]

In general, we are interested in the pattern statistics \( \mu_f(p) \), and the discussion above shows that we need to identify points \( p \) where \( \mu_f(p) \) changes most rapidly. To formalize this, we use the objective function

\[
G_f : \mathcal{P} \times \mathcal{P} \rightarrow \mathbb{R}^+, \quad (p, q) \mapsto G_f(p, q) = d_W(\mu_f(p), \mu_f(q))
\]

defined in (2.18), which measures the 2-Wasserstein distance between the pattern statistics evaluated at the two parameter values \( p \) and \( q \) (in practice, we would use the empirical objective function \( G^N_f \) defined in (2.19)).

Denote by \( n(s) \) the unit vector normal to the tangent vector \( \gamma'(s) \) at a point \( \gamma(s) \in \Gamma \). For fixed \( s \in \mathbb{R} \), let \( L(s) := \gamma(s) + \mathbb{R} n(s) \) denote the line segment perpendicular to the curve \( \Gamma \) at the point \( \gamma(s) \). Given a small offset \( 0 < h \ll 1 \), we then expect that the bifurcation function

\[
g(p; h) := \frac{1}{2h} G_f(p - hn(s), p + hn(s)), \quad p \in L(s), \tag{3.1}
\]

which measures the rate of change of \( \mu_f(p) \) at \( p \), is maximized at \( p = \gamma(s) \). In particular, given the line segment \( L(s) \), we can find the intersection \( \gamma(s) = \Gamma \cap L(s) \) by maximizing \( g(p; h) \) along \( L(s) \) so that

\[
\gamma(s) = \arg \max_{p \in L(s)} g(p; h). \tag{3.2}
\]

We now describe how we can use the arg-max formulation (3.2) to approximate \( \Gamma \) numerically.

Predictor-corrector continuation of pattern statistics

We outline our approach of using the bifurcation function \( g(p) \) defined in (3.1) to trace out a curve in parameter space \( \mathcal{P} \) that separates regions with different prevailing patterns. We will postpone a discussion of how the
We assume that we found two parameter values $p_0, p_1 \in \mathcal{P}$ with $|p_1 - p_0| = s$ so that $p_0, p_1$ lie on the transition curve $\Gamma$ we want to compute. We usually accomplish this by direct simulations near a selected starting point in $\mathcal{P}$.

**Initialization** We assume that we found two parameter values $p_0, p_1 \in \mathcal{P}$ with $|p_1 - p_0| = s$ so that $p_0, p_1$ lie on the transition curve $\Gamma$ we want to compute. We usually accomplish this by direct simulations near a selected starting point in $\mathcal{P}$.

**Predictor-corrector steps** We assume that we have successively computed the distinct points $p_0, \ldots, p_{m-1} \in \mathcal{P}$ on the curve $\Gamma$ for some $m \geq 2$ and now outline how we determine the next point $p_m$ on the curve $\Gamma$. We define

$$ t_m := \frac{p_{m-1} - p_{m-2}}{|p_{m-1} - p_{m-2}|}, \quad n_m := t_m, \quad p_m^* := p_{m-1} + st_m, \quad L_m := \{ p \in \mathcal{P} : p = p_m^* + zn_m, z \in \mathbb{R} \} $$

and refer to Figure 5(i) for an illustration of the underlying geometry. Our goal is to find $p_m$ as the solution to

$$ p_m := \arg \max_{p \in L_m} g_m(p; h), \quad g_m(p; h) := \frac{1}{2h}G_f(p - hn_m, p + hn_m). \quad (3.3) $$

Using the linear parametrization $P_m(z) := p_m^* + zn_m$ of $L_m$ by $z \in \mathbb{R}$, we rewrite (3.3) equivalently as

$$ p_m := P_m \left( \arg \max_{z \in \mathbb{R}} g_m(z; h) \right), \quad g_m(z; h) := \frac{1}{2h}G_f(p_m^* + (z-h)n_m, p_m^* + (z+h)n_m). \quad (3.4) $$

We showed in Lemmas 2.5 and 2.6 that $g_m(z; h)$ is continuous in $z$. Even though we do not know whether this function is differentiable, we will assume in step 3. below that $g_m(z; h)$ has a limit $g_m(z)$ as $h \to 0$ and use this limit to motivate the use of quadratic interpolation to find the maximum of $g_m(z; h)$. We approximate the solution of (3.4) in the following three steps and refer to Figure 5 for an illustration of these steps:

1. **Bisection:** Evaluate $g^+ := g_m(h; h)$ and $g^- := g_m(-h; h)$.
2. **Refinement:** Set $\sigma := \text{sign}(g^+ - g^-)$, and evaluate $g_{\sigma/2} := g_m(\frac{h\sigma}{2}; \frac{h}{2})$ and $g_{3\sigma/2} := g_m(\frac{3h\sigma}{2}; \frac{h}{2})$.
3. **Interpolation:** Determine the quadratic function $g_m^{(2)}(z) = a_2z^2 + a_1z + a_0$ that passes through the points $g-\sigma$, $g_{\sigma/2}$, and $g_{3\sigma/2}$ defined in step 2. for $z = -\sigma h$, $z = \frac{\sigma h}{2}$, and $z = \frac{3\sigma h}{2}$, respectively. If $a_2 < 0$ and $z^*_m = \frac{-a_1}{2a_2} \in (\min\{-\sigma h, \frac{3\sigma h}{2}\}, \max\{-\sigma h, \frac{3\sigma h}{2}\})$, then we accept $p_{m+1} = p_m^* + z^*_m n_m$. 

Figure 5: Panel (i) outlines the geometry for the predictor-corrector step, where we omitted the subscript $m$ in the notation. Panel (ii) indicates at which points we evaluate the bifurcation function $g(z)$ (labeled by disks) and how this translates into evaluations of the pattern statistics $\mu_f(p)$ (labeled by crosses). The points in green correspond to the initial point in step 1., while the points in blue correspond to the refinement in step 2. for the case $\sigma = 1.$

function $g(p)$ can be evaluated to the next section and instead focus here on predictor-corrector continuation, assuming that we can evaluate $g(p)$ numerically. The main issues we need to tackle are (i) that $g(p)$ will be continuous but not necessarily differentiable and (ii) that the definition of $g$ as a distance requires the computation of the pattern statistics at two distinct points.

We will assume that the feature function $f$ can distinguish the patterns we are interested in. Our algorithm depends on the choice of the stepsize $s$ along the curve we want to compute and the offset $h$ in the computation of $g$. We assume that $h$ and $s$ have been chosen so that $0 < h, s \ll 1$. 

Initialization We assume that we found two parameter values $p_0, p_1 \in \mathcal{P}$ with $|p_1 - p_0| = s$ so that $p_0, p_1$ lie on the transition curve $\Gamma$ we want to compute. We usually accomplish this by direct simulations near a selected starting point in $\mathcal{P}$.
We emphasize that step 1 requires the evaluation of the pattern statistics \( \mu_f(p) \) at the points \( p = p_m^* + jhn_m \) for \( j = 0, \pm 1 \). To complete step 2, assuming for simplicity that \( \sigma = 1 \), we need to evaluate \( \mu_f(p) \) only at the additional point \( p = p_m^* + 2hn_m \). We note that it is possible to replace the quadratic interpolation with additional bisection refinements: we found that quadratic interpolation produces better results in our numerical case studies.

We refer to Figure 2 for a practical implementation of the algorithm described above: the colored disks in the center panel correspond to the parameter values at which the pattern statistics was computed for predictor-corrector step as outlined above.

### Implementation

The implementation of the continuation algorithm described above requires the following choices:

1. **Feature function**: We choose a feature function \( f \) that can differentiate between the patterns we want to distinguish.

2. **Spatial discretization**: We choose the number \( K \) of Fourier modes so that we can resolve the nonlinearity and the expected patterns at the expected wavelength. Alternatively, and this is how our numerical computations were conducted, we can use finite differences to solve the PDE model for a sufficiently small spatial stepsize that resolves the patterns we are interested in and use the resulting grid also for \( \alpha \)-shapes.

3. **Initial data**: We need to choose the deterministic part \( U_b \) and the randomized part \( U_t \) of our initial data in (2.11). We choose \( U_b \) to ensure that we reach the patterns we are interested in: for domain-filling patterns, \( U_b \) is typically an unstable homogeneous rest state. We choose \( U_t = \sum_{|k| \leq K} a_k \epsilon_k \) with \( a_k \neq 0 \) for randomized data so that all wave numbers in the system can be activated (\( U_t \) is zero for deterministic initial data).

4. **Randomization**: We sample the random variables \( \{b_k(\omega)\}_{|k| \leq K} \) from a uniform distribution. The number \( N \) of samples, which are used to calculate the empirical measure \( \mu_i^{N, \omega} \) (or the empirical feature mean \( \mu_i^{N, \omega} \) if applicable), can be adjusted using, for instance, a small-sample paired t-test (which tests the null hypothesis that the mean of the difference of feature samples is zero) to ensure that there is a statistically significant difference between the two empirical features in the argument of the bifurcation function \( g \).

5. **Integration time**: The integration time \( T > 0 \) is chosen so that we reach the relevant pattern regime from the initial data \( U_b + U_t^\omega \) within the time interval \([0, T]\). It is possible to adapt \( T \) during continuation, for instance by choosing shorter or longer values and comparing the resulting feature values.

6. **Sublevel sets**: We evaluate the feature function on the sublevel sets \( U_j^{-1}((-\infty, c]) \) of the \( j \)-th component of the solution \( U \) to (2.1). We choose the index \( 1 \leq j \leq d \) and the threshold \( c \in \mathbb{R} \) so that the corresponding sublevel sets best reflect the patterns we are interested in.

7. **\( \alpha \)-shapes**: The evaluation of the feature functions we consider require the computation of the \( \alpha \)-shapes of the sublevel sets. We need to pick the radius \( \alpha \) of the \( \alpha \)-shape and the number \( M^2 \) of lattice points on which we evaluate the solution to approximate the sublevel set. We usually choose \( M := K \) and set \( \alpha = 10/M \). We note that we can adapt \( M \) and \( \alpha \) by comparing the resulting pattern statistics using the Wasserstein metric to ensure that they do not change upon increasing \( M \) or varying \( \alpha \).

8. **Predictor-corrector steps**: We need to choose the arclength stepsize \( s \) and the parameter offset \( h \). We normally pick \( h := s \) and note that the stepsize \( s \) can be adapted based on the successive changes of the angle of the secants, which are indicative of the curvature of the curve \( \Gamma \).

### 4 Results

In this section, we summarize the results of computations of bifurcation and transition curves for several common PDE models that utilize the predictor-corrector approach introduced in §3 to trace out curves based on feature
functions and pattern statistics. Details on the model systems to which we apply our approach can be found in the appendix. Our code is publicly available [49], and we will therefore not discuss the numerical schemes we used in detail.

**Boundaries of coexistence regions**

First, we consider Turing bifurcations, along which a homogeneous rest state destabilizes and patterned states emerge, and transition curves between spot and stripe patterns. For both scenarios, we use randomized perturbations of the homogeneous rest state to create an ensemble of initial conditions. The sublevel sets of the resulting solutions are computed using either $U_1^{-1}(c(0.7), \infty))$ or $U_1^{-1}(\infty, c(0.3))$ with $c(s) := s \max(U_1) + (1 - s) \min(U_1)$ depending on the type of spots we are interested in. To compute the pattern statistics, we use the bagged empirical probability measure $\mu_{RoundDistr, \text{bag}}^N$, defined in (2.22), for the roundness-score feature function $f_{RoundDistr}$ of $\alpha$-shapes.

Figure 6 shows our results for the curves that correspond to Turing bifurcations and transitions from spots to stripes in the Brusselator, Swift–Hohenberg, Gray–Scott, and Schnakenberg models. The figure also contains

Figure 6: Shown are curves that separate transitions from homogeneous states to spots or stripes (black circles), spots to stripes (purple triangles), and stripes to re-entrant spots (blue circles) for the (i) Brusselator, (ii) Swift–Hohenberg, (iii) Gray–Scott, and (iv) Schnakenberg models. The tiles show the results of direct numerical simulations for parameters set at the center of each tile to illustrate and validate the computed transition curves.
the results of typical direct numerical simulations as tiles to facilitate comparison of direct simulations with
the transition curves computed using continuation. Figure 7 shows comparisons of the analytical expressions of
Turing bifurcation curves for the Brusselator and Schnakenberg models with the curves we computed numerically
as well as a comparison of the transition curve between spots and stripes in the Swift–Hohenberg equation with
the associated Maxwell curve. The Swift–Hohenberg equation is variational, and the Maxwell curve corresponds
to parameter values where spots and stripes, for the wave number that destabilizes first at the Turing bifurcation,
have the same PDE energy and Lagrangian, and should therefore co-exist along this curve. The results illustrated
in Figures 6 and 7 show very good agreement between the curves computed using our approach and comparisons
with direction numerical simulations and analytical bifurcation curves. We note that the discrepancy between
the Maxwell curve and the transition curve for the Swift–Hohenberg equation is likely due to the fact that the
wave number selected by spots and stripes deviates from the wave number selected at the Turing bifurcation
curve $\mu = 0$.

**Fold bifurcation curves**

Next, we consider fold bifurcations of spots and stripes in the planar Swift–Hohenberg equation posed on a
square domain. For each of these two cases, we choose a single deterministic initial condition that consists of
spots (or stripes) in the left half of the square domain and the homogeneous rest state in the right half of the
domain. We apply the empirical measure $\mu_{\text{RoundDistr,bag}}^N$ to the $\alpha$-shapes in the multiset of solution profiles to
trace out the bifurcation curves.

As shown in Figure 8(i), this approach accurately traces the fold bifurcation curve of spots that emerges from
the origin in parameter space. Figure 8(ii) shows the results for stripes. For $0 \leq \nu \leq \nu_* := \sqrt{27/38}$, stripes
bifurcate supercritically from the homogeneous rest state along the Turing bifurcation curve $\mu = 0$. As $\nu$ crosses
$\nu = \nu_*$, the bifurcation to stripe patterns along the Turing curve $\mu = 0$ becomes subcritical, and a genuine fold
bifurcation curve of stripe patterns emerges at $(\mu, \nu) = (0, \nu_*)$ and reaches into the region $\mu < 0$. Since our
prepared initial condition consists of stripes and the homogeneous rest state in the left and right halves of the
square domain, our continuation framework first traces out the Turing curve $\mu = 0$ before it picks up the fold
bifurcation curve of stripes at $\nu = \nu_*$. 

Figure 7: In panels (i) and (ii), we compare the analytical Turing bifurcation curves, which correspond to transitions
from homogeneous states to spots or stripes, to the curves found using our numerical algorithm for the Brusselator
(left) and Schnakenberg (center) models. In panel (iii), we compare the numerically computed curve that delineates
transitions from spots to stripes in the Swift–Hohenberg model to the Maxwell curve along which spots and stripes
have the same PDE energy and Lagrangian, where we computed the Maxwell curve using AUTO [18].
Figure 8: We illustrate the computation of fold bifurcations of spots and stripes in the planar Swift–Hohenberg equation. Panel (i) contains the fold curve of spots, which emanates from the origin. Panel (ii) shows results for stripes: Below the line $\nu = \nu_* := \sqrt{27/38} \approx 0.843$, the curve reflects the supercritical bifurcation of stripes along $\mu = 0$ into the region $\mu > 0$ where the rest state $U = 0$ is unstable. At $\nu = \nu_*$ (shown as the dotted horizontal line), this bifurcation becomes subcritical, and an additional bifurcation curve emerges at $(\mu, \nu) = (0, \nu_*)$ that corresponds to fold bifurcations of stripes.

Figure 9: We compare the bifurcations curves for the (i) Brusselator, (ii) Gray–Scott, and (iii) Swift-Hohenberg models obtained through the pattern statistics $\mu_N^f$ for the number $f = f_{\text{Conn}} \in \mathbb{N}$ of connected components in each pattern and the roundness score distribution $f = f_{\text{RoundDist}} \in \text{Prob}([0,m])$ with the curves obtained using the bagged empirical roundness score $\mu_{N_{\text{RoundDist}}},\text{bag}$.

Comparison of feature functions

So far, we used the bagged empirical pattern statistics $\mu_{N_{\text{RoundDist}}},\text{bag}$ derived from the roundness score distribution. We illustrate now how the bifurcation curves obtained from different feature functions compare to each other. Figure 9 contains a comparison of the curves obtained from $\mu_{N_{\text{RoundDist}}},\text{bag}$ in the preceding sections with the curves obtained using the empirical measures $\mu_N^f$ for the number $f = f_{\text{Conn}} \in \mathbb{N}$ of connected components and the roundness score distribution $f = f_{\text{RoundDist}} \in \text{Prob}([0,m])$. Overall, these curves agree well though the curve associated with the roundness score distribution shows more variability compared to the other two curves.

Boundaries of snaking regions

The one-dimensional Swift–Hohenberg equation exhibits stationary solutions whose spatial profiles consist of a spatially localized periodic plateau; see Figure 10(ii) for a sample profile. These localized roll solutions exist in an open region in parameter space and, for each fixed parameter value $(\mu, \nu)$ inside this region, there are countable many localized roll solutions that differ by the length $L \in \ell \mathbb{N}$ of the spatially periodic plateaus where $\ell > 0$. The boundaries of the existence regions are delineated by bifurcation curves that correspond to folds of localized roll solutions (and we note that there are infinitely many of these fold bifurcation curves, namely one for each
Figure 10: Panel (i) contains the numerically computed curves that delineate the snaking region in the one-dimensional Swift-Hohenberg equation together with fold bifurcation curves of localized roll patterns (a sample profile is shown in panel (ii)) that were computed in AUTO. We also included direct simulations starting from localized roll patterns: in the upper left region, the roll plateau expands in time, while it shrinks to zero in the lower right region in parameter space.

Detection of bifurcations of spiral waves

Next, we describe how our approach via feature functions can be used to trace out bifurcation curves of spiral waves. We refer to [42, §12 and Figure 12.8] and the references therein for background and more details on the bifurcations we consider here. Throughout, we use a single deterministic rigidly-rotating spiral-wave profile as the initial condition. We characterize spiral waves through the shape of their tip trajectories. The location $x(t) \in \mathbb{R}^2$ of the tip of a spiral wave $U(x, t) \in \mathbb{R}^2$ can be defined, for instance, through the requirement that $U(x(t), t) = \bar{U}$ for some fixed $\bar{U} \in \mathbb{R}^2$ (if $U \in \mathbb{R}^d$ with $d > 2$, we define the tip position using two of the $d$ components of $U$). The tip location $x$ can be computed numerically using Newton’s method applied to the equation $U(x(t), t) = \bar{U}$ at each time point $t$ during a direct numerical simulation. Having computed the time-dependent tip location $x(t)$, we define the tip trajectory $T$ by $T := \{x(t) : t \in [0, T]\}$. The tip trajectory of a rigidly-rotating spiral wave is a circle, and we focus first on bifurcations where the shape of the spiral tip trajectory ceases to be a circle. The feature functions we choose to trace bifurcation curves will depend on the specific bifurcation scenario we are interested in, and we now discuss these choices in detail and refer to Figure 11 for illustrations.

At retracting-wave bifurcations, the temporal frequency of a rigidly-rotating spiral wave approaches zero, and the tip trajectory of the spiral wave changes from a circle to a semi-infinite line along which the spiral wave retracts to the domain boundary. We could therefore use the length $\tilde{l}_{\text{retract}}(S_\alpha(T)) := \text{Perimeter}(S_\alpha(T))$ of the $\alpha$-shape $S_\alpha(T)$ of the tip trajectory as the feature function that distinguishes rigidly-rotating from retracting spiral waves. Alternatively, and this is the feature function we used in our numerical computations, we can use the Boolean function

$$f_{\text{retract}}(u) := \begin{cases} 1 & \text{max}(u) - \text{min}(u) < 0.01 \\ 0 & \text{otherwise,} \end{cases}$$

which detects transitions between the spiral wave ($f_{\text{retract}}(u) = 0$) and a homogeneous rest state ($f_{\text{retract}}(u) = 1$).
Figure 11: Panel (i) shows the tip trajectories of a rigidly-rotating spiral (circle) and a retracting spiral (line) with initial tip position indicated by a filled circle: the arc length of the tip trajectory can be used to distinguish these cases. Panel (ii) illustrates the tip trajectories of rigidly-rotating and meandering spiral waves: here, the area of the face of the $\alpha$-shape can be used to distinguish these trajectories. Panel (iii) shows the transition from inwardly meandering to drifting to outwardly meandering spiral waves: with the points $x_b$ and $x_{b+1}$ in the feature function $f_{\text{drift}}$ defined in (4.1) as indicated, the function $f_{\text{drift}}$ will be zero for the drifting tip trajectory in the center, negative for the trajectory on the left and positive for the trajectory on the right.

Meandering instabilities correspond to Hopf bifurcations of rigidly-rotating spiral waves at which the tip trajectory acquires a second temporal frequency and changes from a circle to an epicycloid. In particular, the $\alpha$-shape $S_\alpha(T)$ of the spiral tip trajectory changes from a circle with vanishing area to an annulus with strictly positive area. This motivates the feature function

$$f_{\text{meander}}(S_\alpha(T)) := \tanh \left( \frac{4\mu_{\text{Leb}}(S_\alpha(T))}{d_1d_2} \right)$$

to detect transitions from rigidly-rotating to meandering spiral waves, where $d_j$ denotes the length of the interval $P_j S_\alpha(T)$ for the projection $P_j$ of $\mathbb{R}^2$ onto the $j$th component for $j = 1, 2$.

Drifting spiral waves arise from meandering spiral waves when the tip trajectory becomes unbounded; see Figure 11 for a sketch. Drifting spirals exist along curves in parameter space, and we continue these curves as follows. We compute the $\alpha$-shape $S_\alpha(T)$ of the spiral tip trajectory for a large value of $\alpha$, so that the $\alpha$-shape is close to the convex hull of $T$. Denote by $\{x_j\}_{j=1,...,s}$ the vertices of the polygon $S_\alpha(T)$, which we order so that $x_j$ and $x_{j+1}$ are adjacent on $S_\alpha(T)$, where we set $x_{s+1} := x_1$. Let $1 \leq b \leq s$ denote the index for which $|x_{b+1} - x_b|$ is largest. We then define the feature function

$$f_{\text{drift}}(S_\alpha(T)) := \frac{1}{s} \sum_{j=0}^{s} \text{sgn}(x_j - x_b, (x_{b+1} - x_b) \perp)$$

(4.1)

to continue drifting spiral waves.

Spiral-wave turbulence arises when a large-scale spiral wave breaks up into many small spiral-wave segments. In this case, each spiral-wave segment has its own tip, and the equation $U_1(x, t) - \bar{U} = 0$ that defines the tip will therefore have many solutions. In our algorithm, we apply Newton’s method to the tip equation starting with initial data on a grid that spans the domain and collect the resulting tip positions in the set $X_t$ for each time $t$ during the direct simulation. Our feature function is then given by

$$f_{\text{turbulence}}(U_1) := \begin{cases} 
1 & \text{diam}(X_{t_{\text{max}}}) > 10 \\
0 & \text{otherwise}.
\end{cases} \quad X_t := \{ x : U(x, t) = \bar{U} \}, \quad t_{\text{max}} := \arg \max_t \# \{ x \in X_t \},$$

which tests whether multiple tips occur as some time point during the simulation. Though the feature function $f_{\text{turbulence}}$ is not continuous, it allows us to delineate transitions to turbulence efficiently and accurately.

We apply our continuation framework with the feature functions described above to the Barkley and the Bär-Eiswirth models, which exhibit retracting-wave instabilities, transitions to meandering and drifting spirals, and
Figure 12: Shown are retracting-wave bifurcation curves (R), curves corresponding to bifurcations from rigidly-rotating to meandering spirals (M), existence curves of drifting spiral waves (D), and transition curves from regular spiral waves to spiral-wave turbulence (T) for the Barkley model in panel (i) and the Bär–Eiswirth model in panel (ii). The tiles show the tip trajectories of spiral waves in the different regions of parameter space obtained from direct numerical simulations. Shown in gray dotted lines are the bifurcation curves obtained previously in [7, Figure 1] for the Barkley model and in [6, Figure 4] for the Bär–Eiswirth model (reproduced via digitizing the original images).

the emergence of spiral-wave turbulence. The results of our numerical continuation are shown in Figure 12, where we also include the results of direct numerical simulations as well as the bifurcation curves obtained in [7, Figure 1] and [6, Figure 4] for comparison.

Period-doubling bifurcations of spiral waves lead to spiral waves with broken spiral-arm segments as shown in Figure 13. These bifurcations cannot be detected easily by tip trajectories (while it is known that period-doubled spiral waves will drift, the drift speed is typically very small), and we therefore rely on a feature function that is based on the \( \alpha \)-shape of the full spiral profile. For a fixed value \( 1 \leq k \leq d \), we select the function

\[
 f_{pd}(A) := \frac{1}{\beta(A)} \sum_{j=1}^{\beta(A)} \mu_{Leb}(A_j), \quad A := U_k^{-1}([0.1 \min U_k + 0.9 \max U_k, \infty])
\]

that reflects the sum of the areas of the \( \beta(A) \) connected components of the set \( U_k^{-1}([0.1 \min U_k + 0.9 \max U_k, \infty]) \), which consists of points \( x \) for which \( U_k(x) \) is close to \( \max U_k \). Figure 13 shows the results of the continuation of period-doubling bifurcations of spiral waves in the three-component Rössler system for which we take \( k = 3 \).

The comparison with direct numerical simulations indicates that our feature functions traces out the bifurcation curve accurately.

**Freezing method for spiral waves**

The freezing method was developed in [11, 45] to compute relative equilibria of reaction-diffusion systems via direct numerical simulations by adding algebraic constraints that keep the solution profile frozen at a specific location in space. The advantage of this approach is that it can be used to compute traveling waves by fixing their position and, at the same time, calculating their accumulated position (and velocity) without a prior knowledge of their speed; the disadvantage is that the method requires the use of solvers for algebraic-differential systems in order to account for the algebraic constraints that fix the location of the wave.

Here, we apply this approach to relative equilibria and relative periodic orbits in the form of rigidly-rotating as well as meandering and drifting spiral waves. The advantage of using the freezing method is that the spiral waves cannot leave the domain (which will happen for drifting spiral waves) and that this method directly generates
the frequency and tip position of the underlying spiral wave. The approach we propose here relies on α-shapes and does not require the use of solvers for algebraic-differential systems.

Given the initial condition \( U(x, 0) \) of the spiral-wave solution we plan to compute using the freezing method, we define \( A(0) := S_\alpha(U_1^{-1}((-\infty, c]), 0)) \). We then integrate the underlying PDE from \( t = 0 \) to \( t = t_1 \) and set \( A(t_1) := S_\alpha(U_1^{-1}((-\infty, c]), t_1)) \). Next, we solve

\[
(\varphi_1, q_1) := \arg \max_{(\varphi, q)} \text{Area}(\text{Overlap}(R_\varphi(A(0) + q), A(t_1))), \quad R_\varphi := \begin{pmatrix} \cos \varphi & \sin \varphi \\ -\sin \varphi & \cos \varphi \end{pmatrix}
\]

by rotating and shifting the α-shape at \( t = t_1 \) so that it aligns best with the α-shape at \( t = 0 \). After the optimal transformation is obtained, we apply the inverse of the alignment transformation to the spatial grid at time \( t = 0 \) to obtain the grid \( x^\text{grid} := R_\varphi^T x^\text{grid} - q_1 \) on which the aligned solution at time \( t = t_1 \) is defined. We use linear interpolation to define the solution at time \( t = t_1 \) on the original grid \( x^\text{grid} \) and repeat the process by solving

\[
(\varphi_{n+1}, q_{n+1}) := \arg \max_{(\varphi, q)} \text{Area}(\text{Overlap}(R_\varphi(A(t_n) + q), A(t_{n+1}))),
\]

where \( n \) is the iteration count.

Figure 13: Shown is the period-doubling curve of spiral waves for the Rössler system using the feature function \( f_{pd} \) with \( k = 3 \). The tiles in the left panel correspond to the results of direct numerical simulations; the two insets in the right panel illustrate the difference between rigidly-rotating and period-doubled spiral waves.

Figure 14: Shown are the cumulative angles \( \theta(t) \) (top row) and the tip positions \( p(t) \) (bottom row) of rigidly-rotating, meandering, and drifting spiral waves in panels (i), (ii), and (iii), respectively. The spiral waves were computed using the implementation of the freezing method via α-shapes for the Barkley model posed on a disk.
the PDE starting at time \( t = t_1 \) and solving until \( t = t_2 \), proceeding as above with \( t_1 \) replaced by \( t_2 \). At time \( t = t_n \), we collected angles \( (\varphi_j)_{j=1, \ldots, n} \) and translations \( (q_j)_{j=1, \ldots, n} \), which we need to integrate in time to obtain the cumulative angles \( \theta_n \) and tip positions \( p_n \). This can be done by composing the inverses of the alignments sequentially, and we obtain

\[
\theta_n = \sum_{j=1}^{n} \varphi_j, \quad p_n = R(\varphi_n)q_n + R(\varphi_{n-1})R(\varphi_n)q_{n-1} + \ldots + R \left( \sum_{j=1}^{n} \varphi_j \right) q_1.
\]

We apply this approach to rigidly-rotating, meandering, and drifting spiral waves in the Barkley model posed on a disk of radius 30. The results are shown in Figure 14.

5 Discussion

In this paper, we outlined a framework that allowed us to trace out bifurcation and transition curves of patterned states via feature functions and pattern statistics. The advantage of the approach discussed here is that we can use randomized initial conditions and compare the values of measure-valued feature functions using Wasserstein distances. The main disadvantages are (i) the lack of differentiability of the bifurcation functions and (ii) the current restriction to two-dimensional parameter spaces. We note that the lack of smoothness is inherent to our framework since we rely on Wasserstein distance in spaces of probability measures for which no differentiable structure exists. Extensions to higher-dimensional parameter spaces might be possible using multi-dimensional continuation algorithms as outlined in [1, 25] and the references in these papers.

There are several numerical parameters that affect the computational algorithm, including the feature function itself, the randomization, the spatial step sizes, the integration time, the threshold for the sublevel sets, and the radius used in the calculation of \( \alpha \)-shapes. We did not systematically explore the dependence of our results on these parameters, and we also did not explore systematic ways to calibrate and optimize them.

Finally, we briefly discuss extensions of the framework we introduced in this paper. First, it should be possible to use feature functions and pattern statistics also to infer and identify parameters in simulations. If a single feature function is not able to distinguish patterns across the entire parameter space, it might be possible to use ”majority votes” of several feature functions to correctly classify patterns and infer parameter values. Another potential application is to evaluate feature functions on solution trajectories to infer time dynamics from data similar to how we used tip trajectories to classify spiral-wave dynamics. Finally, pattern statistics could be useful to compare and fit models via parameter optimization; see [15] for initial work that pursues this idea.

Appendix A. Models

Barkley and Bär–Eiswirth models: Both models are examples of the Fitzhugh–Nagumo equation

\[
\begin{align*}
    u_t &= \Delta u + \frac{1}{\epsilon} u(1-u)\left( u - \frac{v + a}{b} \right), \\
    v_t &= g(u, v).
\end{align*}
\]

Simulations for both models are carried out using Barkley’s code ezspiral [9]. The Barkley model [7, 8] is characterized by the nonlinearity \( g(u, v) = u - v \). We use the spatial domain \([0, 100] \times [0, 100]\) discretized with 501 points in each direction, choose \( \epsilon = 0.02 \), and vary \((a, b)\). The Bär–Eiswirth model [6] has the nonlinearity

\[
g(u, v) = h(u) - v, \quad h(u) = \begin{cases} 
    0 & u < \frac{1}{4} \\
    1 - 6.75u(u - 1)^2 & u \in \left[\frac{1}{4}, 1\right] \\
    1 & u > 1.
\end{cases}
\]

We use the domain \([0, 50] \times [0, 50]\) with 501 mesh points in each direction, choose \( a = 0.84 \), and vary \((b, \epsilon)\).
**Brusselator:** The Brusselator model [39] is defined by

\[
    u_t = D_1 \Delta u + a - (b + 1)u + vu^2, \quad v_t = D_2 \Delta v + bu - vu^2.
\]

The homogeneous rest state \((u, v) = (a, b/a)\) undergoes a Turing bifurcation along the curve \(b = (1 + a\sqrt{D_1/D_2})^2\). We set \(D_1 = 4\) and \(D_2 = 32\), pose the equation on the domain \([0, 50] \times [0, 50]\) discretized with 50 mesh points in each direction, and evolve the system in time with step size \(dt = 0.005\) until \(T = 100\).

**Gray–Scott:** The Gray–Scott model [23] is given by

\[
    u_t = D_1 \Delta u + u^2v - F(1-u), \quad v_t = D_2 \Delta v + uv^2 - (F + k)v.
\]

We pose this system on the domain \([0, 2.5] \times [0, 2.5]\) with diffusion constants \(D_1 = 2 \times 10^{-5}\) and \(D_2 = 10^{-5}\). We use a uniform spatial grid \(200 \times 200\) mesh points and evolve the system with temporal step size \(dt = 1\) until \(T = 10,000\). For the initial data, we perturb from the homogeneous rest state \((u, v) = (1, 0)\) by creating ten random spots of 0.0625 in the domain.

**Rössler:** The Rössler model [21, 41] is given by

\[
    u_t = 0.4 \Delta u - v - w, \quad v_t = 0.4 \Delta v + u + av, \quad w_t = 0.4 \Delta w + uv - cw + 0.2.
\]

We use the domain \([0, 250] \times [0, 250]\) discretized with 526 mesh points in each direction.

**Schnakenberg:** The Schnakenberg model [43] is given by

\[
    u_t = D_1 \Delta u - u + u^2v + \frac{b - a}{2}, \quad v_t = D_2 \Delta v - u^2v + \frac{b + a}{2}.
\]

The homogeneous rest state \((u, v) = (b, \frac{a+b}{2})\) undergoes a Turing bifurcation when \(a = b^2 \sqrt{\frac{D_1}{D_2}} \left(2 + b \sqrt{\frac{D_1}{D_2}}\right)\). We set \(D_1 = 0.005\) and \(D_2 = 1\), pose the equation on the spatial domain \([0, 4] \times [0, 4]\), and discretize with 60 mesh points in each spatial dimension. Initial conditions are given by the homogeneous rest state with a small amount of uniform noise added. Solutions are evolved with time step \(dt = 0.001\) until \(T = 500\).

**Swift–Hohenberg:** The Swift–Hohenberg equation [44] is given by

\[
    u_t = -(1 + \Delta)^2 u + \mu u + \nu u^2 - u^3.
\]

The homogeneous rest state \(u = 0\) undergoes a Turing bifurcation at \(\mu = 0\). We pose this equation on both \(x \in \mathbb{R}\) (to investigate localized roll solutions) and \(x \in \mathbb{R}^2\) (to study spots and stripes). When \(x \in \mathbb{R}\), we chose the domain \([-32\pi, 32\pi]\) with a uniform grid of 512 points. Solutions were computed using the time step \(dt = 0.1\) and evolved until \(T = 50\). To delineate the existence region of localized roll solutions, we chose the initial condition \(u(x, 0) = (\tanh(x + 4\pi) - \tanh(x - 4\pi)) \cos(x)\). For \(x \in \mathbb{R}^2\), we used the domain \([0, 16\pi] \times [0, 16\pi]\) with 128 mesh points in each direction. Solutions are solved in time using an ETD scheme with time step 0.1 and evolved until \(T = 500\). Randomized initial data are sampled from \(U(0, 1.5)\), while we use the prepared initial conditions

\[
    u_{\text{spots}}(x, 0) = \frac{1}{6} \left( \cos(x_1) + \cos \left( \frac{x_1 + \sqrt{3}x_2}{2} \right) + \cos \left( \frac{x_1 - \sqrt{3}x_2}{2} \right) \right), \quad u_{\text{stripes}}(x, 0) = 0.2 \cos(x_1)
\]

for spots and stripes, respectively.
References


