On the Foundations of Vision Modeling III. Pattern-Theoretic Analysis of Hopf and Turing's Reaction-Diffusion Patterns

JIANHONG SHEN AND YOON MO JUNG *

Abstract. After Turing's ingenious work on the chemical basis of morphogenesis fifty years ago, reaction-diffusion patterns have been extensively studied in terms of modelling and analysis of pattern formations (both in chemistry and biology), pattern growing in complex laboratory environments, and novel applications in computer graphics. But one of the most fundamental elements has still been missing in the literature. That is, what do we mean exactly by (reaction-diffusion) *patterns*? When presented to human vision and visual system, the patterns usually look deceptively simple and are often tagged by household names like *spots* or *stripes*. But are such split-second pattern identification and classification equally simple for a computer vision system? The answer does not seem to be confirmative, just as in the case of face recognition, one of the greatest challenges in contemporary A.I. and computer vision research.

Inspired and fuelled by the recent advancement in mathematical image and vision analysis (Miva), as well as modern *pattern theory*, the current paper develops both statistical and geometrical tools and frameworks for identifying, classifying, and characterizing common reaction-diffusion patterns and pattern formations. In essence, it presents a data mining theory for the scientific simulations of reaction-diffusion patterns.

Key words. Reaction-diffusion, Gray-Scott model, Turing and Hopf instabilities, spots, stripes, nets, entropy, maturity, skewness, kurtosis, geometric statistics, total variation, geometric measures, total curvatures, zebra curves singularities.

1. Introduction and Motivation.

1.1. Turing and Hopf Reaction-Diffusion Patterns. Fifty years ago in his seminal paper [60], Turing surprised the entire scientific community by revealing the foundation for morphogen pattern formation in a *diffusive* chemical reaction environment. By intuition, as a smoothing and filtering process broadly applied even in image and vision analysis [48, 1], diffusion would wipe out any inhomogeneous spatial patterns like the stripes on zebra skins [42]. Turing, however, brilliantly identified Mother Nature's behavior as the counterintuitive consequence caused by the *non-commutativity* of linear operators [46], the very same mechanism underlying Heisenberg's uncertainty principle in quantum mechanics [26].

Consider a planar two-species model as an example. Locally near a *stable* uniform kinetic equilibrium, let D denote the positive diffusivity matrix, and K the stable kinetic matrix with negative real parts of its two eigenvalues. Then the one-parameter family of (linearized) reaction-diffusion operators:

$$L_\mu = -\mu D + K, \qquad \mu \geq 0, \tag{1.1}$$

could lose stability for some positive μ, i.e., at least one eigenvalue of L_μ could cross the imaginary axis from the left. (The parameter μ is proportional to the eigenvalues of $-\nabla^2$, the negative Laplacian.) This could never happen if reaction commutes with diffusion: $[K, D] = KD - DK = 0$. For a non-commutative

*Jianhong (Jackie) Shen and Yoon Mo Jung are with the School of Mathematics, University of Minnesota, 206 Church Street, S.E., Minneapolis, MN 55455, USA. Emails: {jhshen, ymjung}@math.umn.edu. This project is partially supported by USA-NSF's Program of Applied Mathematics under grant number DMS-0202565.

reaction-diffusion pair such as

$$K = \begin{bmatrix} 2 & 7 \\ -1 & -3 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 1 & 0 \\ 0 & 10 \end{bmatrix}, \quad \text{one has} \;\; L_1 = \begin{bmatrix} 1 & 7 \\ -1 & -8 \end{bmatrix},$$

and $\det L_1 = -1$. As a result, one eigenvalue becomes positive and stirs up Turing instability.

Such linear Turing instability, once stabilized by nonlinear kinetics, could sustain stable spatial or temporal patterns such as zebra stripes. In the current paper, we shall only work with planar two-species reaction-diffusion models in the common form of

$$\begin{cases} u_t = D_1 \Delta u + f(u, v) \\ v_t = D_2 \Delta v + g(u, v), \end{cases} \tag{1.2}$$

where f and g denote the nonlinear kinetics.

Reaction-diffusion patterns also frequently arise when Hopf instability occurs, i.e., when a focusing spiral loses its stability and a stable limit cycle is born from bifurcation. In fact, many simulation evidences (e.g., Roy Williams' comprehensive computational results obtained on Caltech's supercomputing facilities [63]) seem to show that reaction-diffusion patterns are more easily observed on the region of Hopf instability.

After Turing's pioneering work, exactly four decades elapsed before a remarkable experiment design by Castets et al. [3] in 1990 eventually confirmed Turing patterns in laboratory. The longtime resisting difficulty had lied in the design of reaction-diffusion systems like (1.2) in a controlled and analyzable manner in actual complex environments.

Coincidentally, almost at the same time, computer scientists started to pick up the reaction-diffusion mechanism for graphic designs. Around 1991, Turk [61] and Witkin and Kass [64] artistically applied reaction-diffusion systems like (1.2) to generate a variety of animal skin patterns in graphic applications. Like flow simulations based on the Navier-Stokes equation, such biochemical simulations of life phenomena surely hold a bright future.

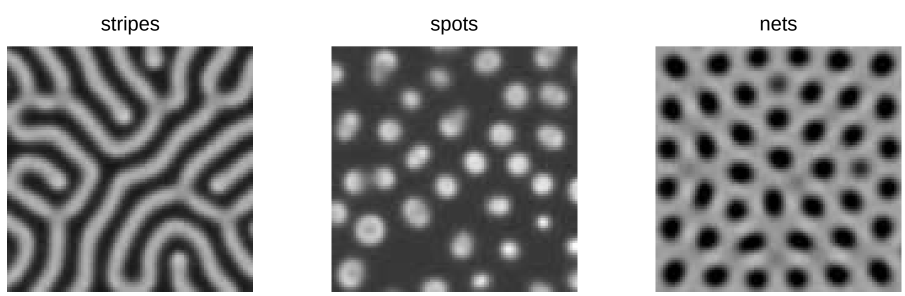

FIG. 1.1. *Some typical examples of reaction-diffusion patterns.*

1.2. Quest for the Meaning of "Pattern". Despite all the afore mentioned remarkable historical works on the theoretical modelling, laboratory growing, or graphic simulations of reaction-diffusion patterns, some key elements have still been surprisingly missing, or blindly taken for granted for a long time. That is, what do we mean precisely by reaction-diffusion *patterns*? how to identify and classify them? and how to characterize their detailed properties?

Human vision and intelligence have become so superior after millions of years' evolution, that centuries after centuries, their swift and effortless recognition of complex patterns has deceived most scientists. In the entire reaction-diffusion literature, for example, it has become household to identify and call some common patterns by "spots," "hexagons," or "stripes." But just as in the case of face or voice recognition, such identification and classification cannot be so trivial as they appear since our central nerve systems do carry out massive computation before reaching these succinct yet quite accurate descriptions.

As David Mumford put it in his most recent perspective article for ICM 2002 [37], *pattern theory*, a booming interdisciplinary field, is really about the mathematics of *perception*. Its mission is to reveal and model all the key elements and processes of human perception and intelligence, which are involved in a split-second action of pattern recognition. To achieve this goal, many mathematicians are intensively working on proper mathematical and statistical frameworks for effectively learning, classifying, analyzing and synthesizing general complex patterns (see for example, Grenander [25], Mumford [37], and Poggio and Smale [49]).

1.3. Pattern Analysis Meets Pattern Formation. On the other hand, the trend is growing stronger that pattern analysis theory should be tightly coupled with pattern formation research. This is mainly driven by the digital and information technologies, and the thirst of being more automatic and independent of human involvement in many areas of data formation and analysis (such as in numerous astronomical and medical applications). Take the simulation of Navier-Stokes fluids for example. Many researchers are striving to design efficient algorithms for automatically reporting and analyzing important flow features and patterns such as vortex sheets (see, for example, Thompson et al. [59] and the references therein) from massive 4-D (space plus time) simulation data. Another remarkable example is Molecular Dynamics. Someday in the future when the complete folding processes of proteins or DNA chains are successfully simulated, could a computer win the Nobel Prize by automatically recognizing some crucial yet unknown patterns (as analogous to the double-helix pattern discovered by James Watson and Francis Crick [62] based on the analysis of Rosalind Franklin's X-ray images of DNA)?

It is guided by the very same spirit that the current work has been carried out. We attempt to apply novel tools recently developed in image and vision analysis, as well as in general pattern theory, to analyze and characterize generic Turing and Hopf reaction-diffusion patterns.

Since reaction-diffusion patterns are generated from the evolutionary partial differential equations, our tools are mainly *statistical* as well as based on *geometric measures*. Statistical tools such as histograms, invariant moments, and the entropy, allow to efficiently differentiate and classify major spatial patterns. On the other hand, tools from geometric measures lead to robust schemes for computing important geometric features of each class of patterns, such as the average radius of a spots pattern, and the total length of a stripes pattern, etc.

1.4. Contributions and Organization. We now highlight what we believe the most significant contributions of the current work.

(1) Our approach systematically combines many important tools developed in modern mathematical image and vision analysis (Miva), and could foster further broad applications of Miva techniques in numerous data mining and analysis problems arising from physical, chemical, or biological simulations.

(2) For the first time, generic reaction-diffusion patterns such as *spots* and *stripes* are analyzed both statistically and geometrically. Many novel notions and tools are developed based on the intrinsic characteristics of these patterns.

(3) Although mainly focused on the patterns from the Gray-Scott model [47], most statistical and geometric tools developed here apply to more general reaction-diffusion models such as CIMA [29, 30] and Brusselator [50]. This is achieved by developing the corresponding models for generic patterns that are independent of the biochemical kinetics.

(4) We introduce the notion of *entropy* from thermodynamics and statistical mechanics to monitor the dynamic behavior of pattern formations. Statistical analysis of typical pattern data leads to the exciting empirical discovery that on average the entropy steadily increases, and converges when the target pattern is stabilized or *matures*.

While this inspirational connection to the *Second Law of Thermodynamics* awaits further explanation in the future, it enables us to introduce the important notion of "maturity" for developing patterns. The determination of maturity solves a long standing problem in the literature of reaction-diffusion simulations, namely, the problem of *stopping time*. To the best knowledge of the authors, no rigorous metric has ever been introduced before us for automatically stopping the evolution when patterns mature.

(5) The geometric features identified by the current work could make a substantial contribution to the solution of another long standing problem in biochemical system identification, namely, to test modelling hypotheses and identify biochemical parameters such as diffusion coefficients.

Consider an idealized leopard for an oversimplified explanation. Suppose its spotty skin pattern is well modelled by the reaction-diffusion system (1.2). Let d and N denote the average radius and average spots density (per square inch, say) separately. Suppose that predictive relations are available in the form of (e.g., Turing's formula for intrinsic chemical wavelength [60]):

$$d = \phi(D_1, D_2) \quad \text{and} \quad N = \psi(D_1, D_2),$$

where D's stand for the diffusivity coefficients. (For the purpose of illustration, all the kinetic parameters involved are assumed to be known.) Then the measurement of d and N could help us identify the two diffusivity coefficients D_1 and D_2, which are often extremely difficult or costly to obtain for an on-going chemical reaction or a living leopard.

From the surface features and patterns to the underlying driving-force parameters, this type of inverse problems well justify the efforts and values of pattern analysis in physical, biological, or chemical systems.

The organization goes as follows. In Section 2, the working reaction-diffusion model of the current paper - the Gray-Scott system, is analyzed in terms of both Hopf and Turing instabilities. To our best knowledge, the existence of Turing instability is shown here for the first time in the literature.

Various statistical tools are developed in Section 3 for the monitoring of pattern formations and for the identification and classfication of matured patterns. One of the most significant empirical discoveries is that pattern formations seem to obey the *Second Law of Thermodynamics*, namely, the entropy increases monotonically. Based on entropies, we are able to introduce the key quantities of pattern *maturity* and the *dynamic activity* of matured patterns. On the other hand, direct as well as geometry based histograms and their statistics successfully characterize and classify typical reaction-diffusion patterns. We also explain why most of the key statistical features are intrinsic and independent of the reaction-diffusion systems that generate them.

Pattern analysis based on geometric measure theory is presented in Section 4. Important geometric measures such as total areas, total variations, and total curvatures are employed to compute the key geometric features of common patterns, which include (a) the average radius of a (matured) spots pattern, (b) the total number of spots in a given reaction-diffusion domain, (c) the average width of a stripes pattern, (d) the total length of *all* the stripes in a stripes pattern, and (e) the number of singular points (i.e. endpoints and Y-shaped branching points) in a stripes pattern. The estimation errors of the leading term approximations are shown to be intrinsically controlled by the signal-to-noise ratios (SNR, for spots and nets patterns) and the aspect ratios (for stripes patterns).

Section 5 contains a brief conclusion.

2. Hopf and Turing Patterns via the Gray-Scott Model.

2.1. Gray-Scott Model. There are a number of well known planar two-species reaction-diffusion models in the literature, including the Brusselator model [50], and the CIMA (chlorite-iodide-malonic-acid) reaction [29, 30]. Throughout the current paper, the major example to be focused on is the Gray-Scott model, which was first studied phenomenologically in John Pearson's *Science* paper [47], and has also been comprehensively computed using Caltech's supercomputing facilities by Roy Williams [63]. Both the kinetics and reported patterns of the Brusselator model are very similar.

The Gray-Scott model is a variant of the autocatalytic Selkov model of glycolysis and addresses the following hypothetical reactions involving two chemical morphogens:

$$U + 2V \to 3V \quad \text{and} \quad V \to P. \tag{2.1}$$

Here the product P is an inert precipitate, U is fed into the system by a constant feeding rate, and V removed by the same rate. After a step of partial dimensionless reduction, the reaction-diffusion equation for the concentrations u and v becomes:

$$\begin{cases} u_t = D_1 \Delta u - uv^2 + F(1-u) \\ v_t = D_2 \Delta v + uv^2 - (F+k)v, \end{cases} \tag{2.2}$$

where F corresponds to the feeding and extraction rate, and k the rate constant for the second reaction. Following Pearson [47], we shall also simulate the model with diffusivity coefficients:

$$D_1 = 2 \times 10^{-5} \quad \text{and} \quad D_2 = 10^{-5}.$$

We state two reasons that have motivated this specific but clever choice. First, as clear in the next section, near the unique kinetic equilibrium that nurtures pattern formation, u acts as the inhibitor and v the activator. Then the choice is consistent with the famous mechanism of Turing pattern formation known as *long-range inhibition and short-range self-activation* [42, 45]. Secondly, for the numerical simulation of the system, the well known CFL (Courant-Friedrichs-Levy) stability condition requires (for 2-D diffusion equations):

$$\frac{\Delta t}{(\Delta x)^2}\max(D_1, D_2) \leq \frac{1}{4}.$$

Since the spatial grid size Δx is often in the order of $O(10^{-2})$, the diffusivity order $O(10^{-5})$ is therefore the highest possible that allows fast numerical marching in the order of $\Delta t = O(1)$.

It shall be assumed that the reaction-diffusion domain Ω is a Cartesian rectangle, and no feeding or extraction is carried out along its boundary. That is, the Neumann conditions are imposed: $\partial u/\partial \boldsymbol{n} = \partial v/\partial \boldsymbol{n} = 0$ along $\partial\Omega$.

Discussed next are the equilibrium analysis and Hopf and Turing instabilities. Related work has also appeared in [31]. The two major contributions here are: (a) A new pair of normalized parameters (α, δ) is introduced based on the raw kinetic parameters (F, k). The former leads to simpler solution representations and much better phase plane resolution; (b) We show for the first time that Turing instability indeed occurs in the Gray-Scott model, which has never been reported either theoretically or computationally in the previous works [47, 31] due to nonlinear complexity.

2.2. Kinetic Equilibria. At kinetic equilibrium,

$$\begin{cases} 0 = -uv^2 + F(1-u) \\ 0 = +uv^2 - (F+k)v. \end{cases} \tag{2.3}$$

It is easy to see that $(u, v) = (1, 0)$ is the trivial equilibrium. Assume that $v \neq 0$. Define as in [31]

$$\delta = \frac{F+k}{F} = 1 + \frac{k}{F} > 1. \tag{2.4}$$

Then the addition of the two equilibrium equations lead to $u + \delta v = 1$, and the second equation gives

$$u(\delta v) = \alpha, \quad \text{with} \;\; \alpha = \delta(F+k) = F\delta^2 > 0. \tag{2.5}$$

Thus nontrivial equilibrium is subject to the canonical quadratic system:

$$u + \delta v = 1 \quad \text{and} \quad u(\delta v) = \alpha, \tag{2.6}$$

from which u and δv are seen to be symmetric. Denote the two nontrivial equilibria by

$$\left(u_+ = \frac{1}{2} - \sqrt{\frac{1}{4} - \alpha}, \quad \delta v_+ = \frac{1}{2} + \sqrt{\frac{1}{4} - \alpha} \right), \quad \text{and} \tag{2.7}$$

$$\left(u_- = \frac{1}{2} + \sqrt{\frac{1}{4} - \alpha}, \quad \delta v_- = \frac{1}{2} - \sqrt{\frac{1}{4} - \alpha} \right), \tag{2.8}$$

provided that α is less than the critical value $\alpha_c = 1/4$, or equivalently,

$$k < k_c = \frac{\sqrt{F}}{2} - F = \frac{1}{16} - (\sqrt{F} - \frac{1}{4})^2, \tag{2.9}$$

which is an upsidedown parabola $k_c = k_c(f)$ with respect to $f = \sqrt{F}$ (see Figure 2.1).

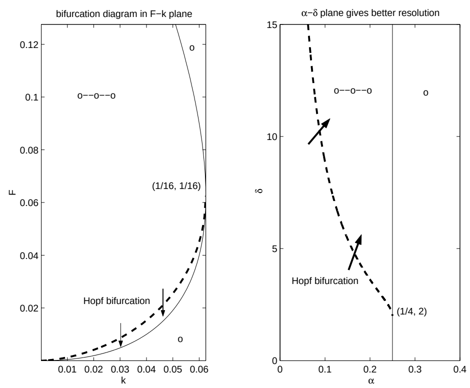

FIG. 2.1. *Bifurcation diagrams in both the $k-F$ plane and $\alpha-\delta$ plane. The solid lines indicate the onset of equilibrium bifurcation (with the little circles symbolically representing the number of equilibria). The dashed ones indicate the onset of Hopf bifurcation (with the arrows pointing toward the birth of stable limit cycles). It is clear that the $\alpha - \delta$ plane better resolves (or differentiates) these two different types of bifurcation.*

2.3. Kinetic Stability and Hopf Bifurcation. The trivial kinetic equilibrium $(1, 0)$ is always a stable node since by dropping the second order terms $uv^2 = O(v^2)$, the kinetic system decouples to

$$u_t = F(1-u) \quad \text{and} \quad v_t = -(F+k)v.$$

For the other two nontrivial ones (2.7) and (2.8), notice that the kinetic Jacobian is

$$K = K(u, v) = \begin{bmatrix} -v^2 - F & -2uv \\ v^2 & 2uv - (F+k) \end{bmatrix}.$$

With the definitions of α and δ, and the equilibrium conditions

$$u(\delta v) = \alpha \quad \text{and} \quad (\delta v)^2 - (\delta v) + \alpha = 0,$$

the kinetic matrix simplifies to

$$\delta^2 K = \begin{bmatrix} -\delta v & -2\alpha\delta \\ \delta v - \alpha & \alpha\delta \end{bmatrix}, \quad \text{with} \quad \mathrm{trace}(\delta^2 K) = \alpha\delta - \delta v \quad \text{and} \quad \det(\delta^2 K) = \alpha\delta(\delta v - 2\alpha). \tag{2.10}$$

Thus the signs of $\mathrm{trace}(K)$ and $\det(K)$ are the same as

$$(\alpha\delta - \delta v) \quad \text{and} \quad (\delta v - 2\alpha). \tag{2.11}$$

At (u_+, v_+) in (2.7), $\delta v_+ - 2\alpha > 0$ since

$$\delta v_+ > \frac{1}{2} \quad \text{and} \quad 2\alpha < \frac{1}{2}.$$

Thus $\det(K_+) > 0$ with $K_+ = K(u_+, v_+)$. Setting the trace to 0 yields $\alpha\delta - \delta v_+ = 0$ which reveals the onset of Hopf bifurcation:

$$\delta_c = \delta_c(\alpha) = \frac{\delta v_+}{\alpha} = \frac{1}{\alpha}\left(\frac{1}{2} + \sqrt{\frac{1}{4} - \alpha}\,\right).$$

Thus for any fixed $\alpha < \alpha_c = \frac{1}{4}$, as δ increasingly crosses $\delta_c(\alpha)$, $\mathrm{trace}(K_+)$ changes from negative sign to positive. As a result, (u_+, v_+) experiences a Hopf bifurcation (see Figure 2.1 and Figure 2.2).

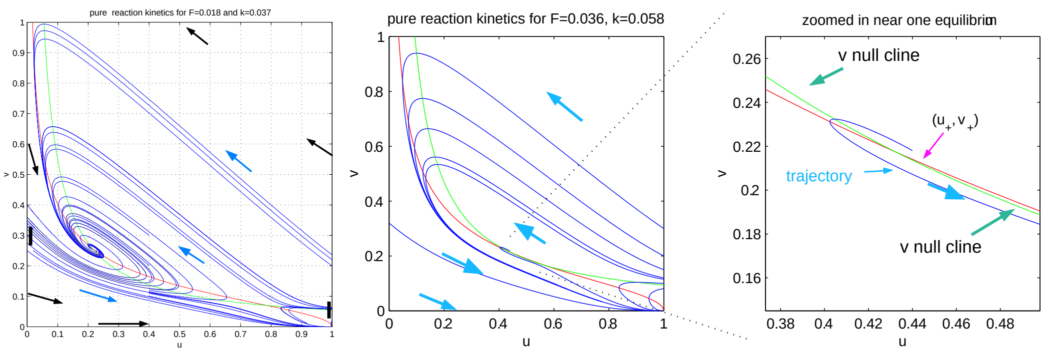

FIG. 2.2. *Stream lines of the kinetic reaction system.*

At (u_-, v_-), notice that

$$\delta v_- = \frac{\alpha}{u_-} < \frac{\alpha}{1/2} = 2\alpha.$$

implying that $\det K(u_-, v_-) < 0$, and (u_-, v_-) is a saddle point.

Figure 2.1 shows the kinetic bifurcation diagrams in both the $k - F$ parameter plane and the $\alpha - \delta$ plane. While the two bifurcations stay too close in the $k - F$ plane, they are well unfolded and resolved in the $\alpha - \delta$ plane. Figure 2.2 displays the streamlines of the kinetic system in two different sets of parameters. The leftmost one clearly outlines the stable spiral (u_+, v_+) before Hopf bifurcation, while the middle one depicts the streamlines when (u_+, v_+) loses stability. The rightmost is a zoomed-in version of the middle one.

2.4. Existence of Turing Instability. Turing instability of the Gray-Scott model has not been reported in any of the previous works [47, 31], largely due to the complexities caused by nonlinearity, including the way the kinetic parameters are involved in the equilibria.

To overcome the difficulty, we take a *demand-and-supply* based unconventional approach, instead of the traditional exhaustive analysis. That is, we pre-select a set of eigenmodes for potential Turing instability (i.e., the demand), and then investigate whether indeed exist reaction-diffusion parameters (i.e., the supply) to meet the instability requirement.

For numerical convenience, assume the reaction-domain is $\Omega = [0, 2] \times [0, 2]$, and the diffusion coefficients are $D_1 = 2 \times 10^{-5} = 2D_2$, as explained in Section 2.1. Then the eigenmodes of the Laplacian (with the Nuemann boundary condition) on the domain are tensor products:

$$\phi_{n,m}(x, y) = \cos(n\pi \frac{x}{2}) \ \cos(m\pi \frac{y}{2}), \quad n, m = 0, 1, 2, \cdots,$$

corresponding to eigenvalues

$$\mu_{n,m} = (n^2 + m^2)(\frac{\pi}{2})^2, \quad n, m \geq 0.$$

Suppose it is demanded that all the following eigenmodes lose stability in the reaction-diffusion:

$$12 \leq \frac{\mu_{n,m}}{\pi^2} \leq 25, \tag{2.12}$$

which will at least include the following eigenmodes:

$$\begin{array}{ccccc} \cdot & \cdot & (4,6) & (4,7) & (4,8) \\ \cdot & (5,5) & (5,6) & (5,7) & (5,8) \\ (6,4) & (6,5) & (6,6) & (6,7) & (6,8) \\ (7,4) & (7,5) & (7,6) & (7,7) & \cdot \\ (8,4) & (8,5) & (8,6) & \cdot & \cdot \end{array} \tag{2.13}$$

The motivation is clear: when satisfied, it will roughly lead to <u>several</u> standing waves in both x and y directions, which is good for observation in simulations - neither too dense nor too loose. Since $\pi^2 = 9.8796\ldots$, the instability demand (2.12) is contained in the more workable interval

$$100 \leq \mu \leq 250. \tag{2.14}$$

Now consider the combined reaction-diffusion system when (u_+, v_+) is a *stable* equilibrium, i.e., when

$$\alpha < \alpha_c = \frac{1}{4} \quad \text{and} \quad \delta < \delta_c(\alpha) = \frac{1}{\alpha}\left(\frac{1}{2} + \sqrt{\frac{1}{4} - \alpha}\ \right), \tag{2.15}$$

as previously derived. In particular, observe that

$$\delta_c(\alpha) > \frac{1}{2\alpha} > 2 \ \text{ for } \ \alpha < \alpha_c, \quad \text{and} \quad \delta_c(\alpha_c) = 2.$$

Let $L_+ = K_+ - \mu D$ denote the linearized reaction-diffusion at (u_+, v_+), where as in (2.10),

$$\delta^2 K_+ = \begin{bmatrix} -w & -2\alpha\delta \\ w - \alpha & \alpha\delta \end{bmatrix}, \quad D = \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix}, \quad w = \delta v_+. \tag{2.16}$$

Notice that the signs of the kinetic matrix K_+ are in the form of $[-, -; +, +]$, the type in the reaction-diffusion literature known as *cross* activator-and-inhibitor [42, 45]. Define $\lambda = \delta^2\mu$ and

$$L_\lambda = \delta^2 L_+ = \begin{bmatrix} -w - \lambda D_1 & -2\alpha\delta \\ w - \alpha & \alpha\delta - \lambda D_2 \end{bmatrix} \tag{2.17}$$

Then Turing instability occurs if and only if $\det(L_\lambda) > 0$.

THEOREM 2.1. *Suppose that* $D_1 = 2 \times D_2 = 2 \times 10^{-5}$*, Then Turing instability indeed occurs over the range (2.14) for any* $(\alpha, \delta = 2)$ *with* $\alpha = \frac{1}{4} - \varepsilon^2$ *and* $0 < \varepsilon \leq 0.003$.

A careful check of the proof below confirms that by only using a three-digit decimal number, 0.003 is the best value for the upper limit. Such narrow concentration of Turing's instability near the critical point $(\alpha = \frac{1}{4}, \delta = 2)$ in the parameter plane is well supported by Roy Williams' comprehensive computational results [63].

Proof. When $\delta = 2$, over the range $100 \leq \mu \leq 250$, we have $\lambda = \delta^2 \mu \in [400, 1000]$. Define $\eta = \lambda D_2 \in [0.004, 0.01]$. Then under the specified conditions,

$$0.1 < \alpha < \frac{1}{4} \quad \text{and} \quad \eta < 0.1\alpha. \tag{2.18}$$

Also define $q = w - \alpha\delta$. Then

$$q = \frac{1}{2} + \sqrt{\frac{1}{4} - \alpha} - 2\alpha = \varepsilon + 2\varepsilon^2 < 0.0031. \tag{2.19}$$

Finally, by the definition of L_λ,

$$\begin{aligned}
\det L_\lambda &= \alpha\delta(w - 2\alpha) - \lambda(\alpha\delta D_1 - w D_2) + \lambda^2 D_1 D_2 \\
&= (\alpha\delta + \lambda D_2) q - \alpha^2 \delta (2 - \delta) - \alpha\delta\lambda D_2 + 2(\lambda D_2)^2 \\
&= (2\alpha + \eta) q - 2\alpha\eta + 2\eta^2 \\
&< 2.1\alpha q - 2\alpha\eta + 0.2\alpha\eta \qquad \text{by (2.18)} \\
&= (2.1 q - 1.8\eta)\alpha \\
&\leq (2.1 \times 0.0031 - 1.8 \times 0.004)\alpha < 0,
\end{aligned}$$

which completes the proof. □

3. Statistical Analysis of Reaction-Diffusion Patterns. Our approach has been profoundly influenced by the recent progress in image, vision, and pattern analysis.

To analyze complex natural images such as wooden textures, satellite images, tissues or more general medical images, many statistical tools including statistical mechanics have been introduced and extended by image analysis experts. Spatial image patterns are often treated as Gibbs canonical ensembles or Markov random fields, whose short-range energies are defined through spatial filters and parametric or non-parametric learning. This framework was pioneered by Geman and Geman, Mumford, Zhu, and their collaborators [21, 65, 66, 40].

For reaction-diffusion patterns, the statistical approach is intrinsically justified. Although most such patterns are either spatially or temporally varying from the deterministic point of view, statistically they do seem to remain stationary. Take some spots patterns for example. The sizes and locations of the spots often look quite arbitrary. Furthermore, some spots are constantly created while others being annihilated. But overall the spots clearly exhibit some (statistical) invariance. Therefore they remarkably resemble a box of gas molecules in thermal equilibrium, and statistics becomes a natural tool.

On the other hand, compared with general images with multiscale or multi-frequency features, reaction-diffusion patterns resulting from nonlinear parabolic equations carry more geometric regularity. Therefore, nonlinear geometric filters such as the curvature could play a more important role than conventional linear ones.

3.1. Entropy and Second Law of Thermal Dynamics. Ever since Shannon's famous paper on the mathematical theory of communication and information [55], the notion of *entropy* from statistical mechanics has been playing a universally crucial role in information theory, signal and image analysis [13], and pattern learning and analysis [65, 66]. Information entropy (or the negative Shannon information measure), is a natural metric for measuring the degree of "thermal" disorder in information.

The Second Law of Thermal Dynamics, on which the entire theory of statistical mechanics stands up, says that any adiabatic (i.e., Neumann or no flux exchange across the boundaries) dynamic process must result in an increment in entropy, which leads to the *maximum entropy principle* [22]. In this section, empirical evidences are presented to illustrate the remarkable Second Law type behavior during the formation of reaction-diffusion patterns.

From now on, we shall only focus on the "activator" $v = v(x, y, t)$ (see the kinetic matrix in (2.16) of the reaction-diffusion system:

$$\begin{cases} u_t = D_1 \Delta u - uv^2 + F(1-u) \\ v_t = D_2 \Delta v + uv^2 - (F+k)v, \end{cases} \tag{3.1}$$

with generic initial conditions $u(x, y, 0) = u_0(x, y)$ and $v(x, y, 0) = v_0(x, y)$, and Neumann adiabatic conditions $\partial u/\partial \boldsymbol{n} = 0$ and $\partial v/\partial \boldsymbol{n} = 0$ along the boundary $\partial\Omega$. The initial conditions of (3.1) are usually specified by random fields near the nontrivial uniform equilibrium:

$$u_0(x, y) = u_+ + n(x, y) \quad \text{and} \quad v_0(x, y) = v_+ + m(x, y),$$

where $n(x, y)$ and $m(x, y)$ denote white noises whose variances are well dominated by the signals (u_+^2, v_+^2) (i.e., with a large signal-to-noise ratio). As practiced in [47], one could also replace u_+ and v_+ in the initial data by some compactly supported data.

The value of the activator $v(x, y, t)$ is understood as the activation level at site (x, y). Although the system is formulated deterministically, with random fields as the initial inputs, as well as in the cellular level where diffusion becomes Brownian motion, the solution $v(x, y, t)$ is better treated as a random field. For each constant $\theta \in (0, \infty)$, define the θ-level set

$$\Gamma(\theta|t) = \{ \ (x, y) \in \Omega \mid v(x, y, t) \equiv \theta \ \}. \tag{3.2}$$

[Notice that the first quadrant of the uv-plane is a kinetic invariant region, which makes it suffice to only consider $v \in (0, \infty)$.] The histogram of $v(x, y, t)$, or the empirical probability density, is defined as

$$p(\theta|t) = \frac{1}{|\Omega|} \lim_{\Delta\theta \to 0} \frac{|\cup_{\alpha \in (\theta, \theta+\Delta\theta)} \Gamma(\alpha|t)|}{\Delta\theta}, \tag{3.3}$$

where $|\cdot|$ denotes the Lebesgue measure in R^2. We then define the entropy of v at time t to be

$$S(t) = -\int_0^\infty p(\theta|t) \ln p(\theta|t) d\theta, \tag{3.4}$$

where the Boltzmann constant k_B in Gibbs' original definition has been set to 1 [22].

Our first remarkable empirical discovery is that the reaction-diffusion system appears to (approximately) follow the Second Law of Thermodynamics (see Figure 3.1).

Empirical Discovery. The entropy $S(t)$ consists of two components $S_0(t) + S_a(t)$, where (see Figure 3.1),

(1) the mean-field or smooth component $S_0(t)$ is non-decreasing, and converges to some constant A as $t \to \infty$; and

(2) the dynamically active component $S_a(t)$ is a rapidly fluctuating time series so that the signal-to-noise ratio $\lim_{t\to\infty} S_0(t)/\sigma(S_a(t)) \gg 1$, where $\sigma(\cdot)$ denotes the standard deviation.

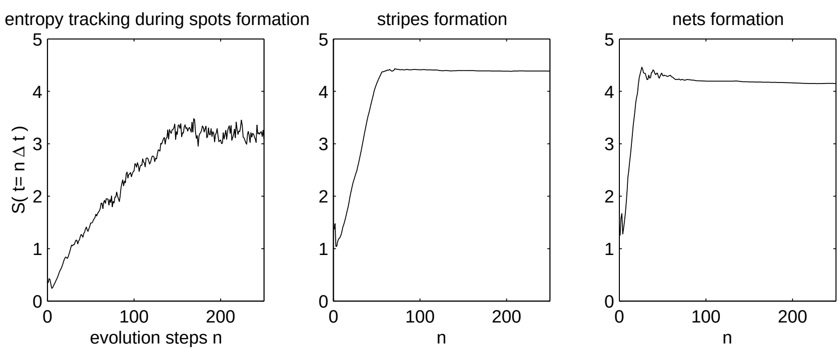

FIG. 3.1. *The entropy sequences during the formation of typical patterns.*

In Figure 3.1, the entropy curve for the spots pattern appears to be bumpier than the other two. This reflects the well known fact that the spots pattern often evolves more actively in computer simulations, with constant annihilation and creation.

While Figure 3.1 looks clear to human vision that the above two statements make perfect sense, it is yet not so trivial to actually extract the two components $S_0(t)$ and $S_a(t)$ from $S(t)$. One has to apply tools from signal estimation theory such as Bayesian reference.

Employed here is the deterministic version of Bayesian estimation [39, 10] - Tikhonov regularization [15]. Due to the monotonicity consideration on S_0, we apply Rudin-Osher-Fatemi's total variation (TV) restoration model [53, 4, 6] to extract the two components. Define an error functional for any estimation $g(t)$ by

$$e[g(t)\big|S(t)] = \int_0^T |g'(t)|_\varepsilon dt + \frac{\lambda}{2}\int_0^T (S(t) - g(t))^2 dt, \tag{3.5}$$

where T is the total running time (e.g., $T = 250\Delta t$ for Figure 3.1), and $|a|_\varepsilon = \sqrt{a^2 + \varepsilon^2}$ the regularized absolute value for some fixed small constant. The least square fitting term implies that Gaussian type fluctuations have been assumed. As in Bayesian estimation, the parameter λ is inversely proportional to the fluctuation variance, and is either tunable or estimated. We then define the mean-field (or trend) entropy S_0 and the dynamically active (or fluctuating) entropy S_a by

$$S_0(t) = \mathrm{argmin}_{g(t)} e[g(t) \big| S(t)] \quad \text{and} \quad S_a(t) = S(t) - S_0(t).$$

It is well known [4, 15, 35] that the decomposition does exist and is unique.

Plotted in Figure 3.2 are the extraction results for the examples in Figure 3.1. Notice that the dynamically active component have been plotted in the dB unit: $10 \log_{10} |S_a(t)|$, as customary in signal analysis and processing [44]. The dynamic difference is manifest in the S_a plotting between the active spots pattern and relatively quieter stripes and nets patterns. (Assuming $\Delta t = 1$ in Figure 3.1, we have taken the TV regularizer $\varepsilon = 0.01$ and $\lambda = 1$.)

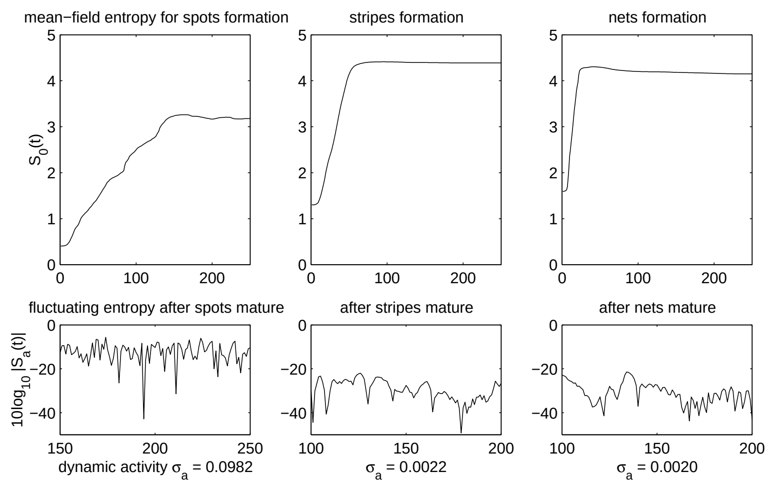

FIG. 3.2. *Extraction of the mean-field entropy $S_0(t)$ and the dynamically active entropy $S_a(t)$ by the model in Eqn. (3.5). Notice that S_a has been plotted in the dB unit: $10 \log_{10} |S_a(t)|$, as customary in signal processing. It clearly shows that the spots pattern is more dynamically active.*

3.2. Pattern Maturity $r(t)$ and Dynamic Activity σ_a. The main goal of this section is to define two important quantities based on entropies: maturity $r(t)$ and dynamic activity σ_a.

Assume the fluctuation component $S_a(t)$ has a zero mean $E[S_a(t)] \equiv 0$. Then the trend $S_0(t) = E[S(t)]$ is precisely the mean field of the entropy time series. In particular, by approximating the expectation operator by empirical averaging, ones obtains

$$A = \lim_{t \to \infty} S_0(t) = \lim_{t \to \infty} \frac{1}{h} \int_{t-h}^{t} S(t) dt, \tag{3.6}$$

where h could be any prefixed window length. We shall call A the *saturated* entropy of the pattern formation. Let β be a positive weight constant. The *ripeness index* or *maturity* is then defined by

$$r(t) = r(v(t)) = e^{-\beta (A - S_0(t))^2}. \tag{3.7}$$

Notice that $r(t) \in (0,1]$, and $\lim_{t\to\infty} r(t) = 1$. Thus the closer $r(t)$ gets to 1, the more matured the pattern appears to be. [Note: One could also try Gibbs' form $e^{-\beta(A-S_0(t))}$ since the mean entropy $S_0(t)$ has been assumed non-decreasing. The quadratic Gaussian form is adopted here since numerical extraction may never get the exact monotonicity. Also notice that the weighting constant β corresponds to the reciprocal of the temperature in Gibbs' form and of the variance in the Gaussian form.]

The ripeness index is a valuable measure for deciding a good stopping time T_s in computer simulation after the pattern has become "ripe" enough. For example, one could preset a maturity level $r_0 \in (0,1)$ close to 1, and define the stopping time by

$$T_s = T_s(r_0) = \sup\{t > 0 \mid \text{for any } s < t,\ r(s) < r_0\ \}. \tag{3.8}$$

Then arises a natural issue regarding the actual computation of the ripeness index $r(t)$. On one hand, its definition (3.7) crucially depends on the saturated entropy A, which is obtained from $S_0(t)$ by letting $t \to \infty$. On the other hand, the major goal of introducing $r(t)$ is to offer an effective strategy for automatic termination and avoid running the simulation all the way to the infinity.

The issue could be resolved by two approaches. First, if one is experimenting on a fixed system with fixed reaction-diffusion parameters, but with different initial conditions for instance, the saturated entropy A could be learned and well estimated from a single full run of a particular but generic set of initial conditions. Alternatively, one could modify the definitions in (3.7) and (3.8) so that they become progressive or *causal*, independent of any "future" development. This can be done as follows. Suppose the targeted maturity level is $r_0 = 1 - \varepsilon^2$ with $\varepsilon \ll 1$. Then (3.8) implies that for any $t_1, t_2 : T_s < t_1 < t_2$ (under the assumption that the mean entropy is non-decreasing),

$$r(t_i) \geq r_0 = 1 - \varepsilon^2, \qquad i = 1, 2.$$

On the other hand, by (3.7), to the first order,

$$r(t_i) = e^{-\beta(A - S_0(t_i))^2} \simeq 1 - \beta(A - S_0(t_i))^2.$$

Thus, again to the first order,

$$\beta(A - S_0(t_i))^2 \leq \varepsilon^2, \qquad i = 1, 2,$$

which leads to

$$|S_0(t_2) - S_0(t_1)| \leq b\varepsilon, \qquad b = \frac{2}{\sqrt{\beta}}. \tag{3.9}$$

This inspires a causal definition of the stopping time: with a being any fixed time duration,

$$T_s = T_s(\varepsilon) = \sup\{t > a\ :\ |S_0(t) - S_0(t-a)| > \varepsilon\ \}. \tag{3.10}$$

Compared with the original definition, it is *causal* and independent of future computation. In real simulation, one may add a further delay of some fixed length h so that numerical error could be diminished and the pattern could get securely matured up to the precision ε. That is, one takes $T_s(\varepsilon) + h$ as the actual stopping time.

The fluctuation entropy $S_a(t)$, though behaving like noise, does reflect important dynamic properties of matured patterns. Weak fluctuations often correspond to static *standing waves*, while stronger ones are often associated with *travelling waves* like spontaneous spots. We thus define the *dynamic activity* for any *matured pattern* to be

$$\sigma_a = \lim_{t\to\infty} E[S_a(t)^2]^{\frac{1}{2}}. \tag{3.11}$$

As before, S_a has been assumed to have zero mean. In application, the expectation operator is replaced by empirical averaging, and the limit is approximated at any moment $t > T_s + h$,

$$\sigma_a^2 \simeq \frac{1}{h}\int_{t-h}^{t} S_a^2(\tau)\, d\tau, \qquad \text{for some time interval } h.$$

For the typical patterns in Figure 3.2 for instance, $\sigma_a = 0.0982,\ 0.0022,\ 0.0020$ separately for the matured spots, stripes, and nets patterns. The difference is as big as nearly 50 times.

3.3. Patterns of Histograms via Invariant Moments. So far, we have only defined key quantities for monitoring the maturity and activity of developing patterns. In the coming two sections, we discuss how to differentiate and classify the matured patterns.

Let us first briefly review the general framework of pattern classification [17]. Let X denote a set of target objects to be classified. A feature map of dimension n associates to each object $x \in X$ a feature vector in R^n:

$$x \to \boldsymbol{F}(x) = (f_1(x), \cdots, f_n(x)).$$

For any subset $E \subset R^n$, the collection of objects $\boldsymbol{F}^{-1}(E) \subset X$ is called an E-pattern class. A finite and exclusive patten classification based on the feature map $\boldsymbol{F}$ is a finite partition of R^n:

$$R^n = E_1 \cup E_2 \cup \cdots \cup E_N, \qquad \text{for some } N,$$

and its induced partition of X into different pattern classes:

$$X = X_1 \cup \cdots \cup X_N, \qquad \text{with } X_k = \boldsymbol{F}^{-1}(E_k),\ \ k = 1 : N.$$

The main challenges of pattern classification reside in [65, 66, 49]

(a) how to establish a suitable feature map $\boldsymbol{F}$ which is neither too coarse (with substantial information loss) nor over-redundant; and

(b) how to properly partition the feature space based on either supervised or unsupervised learning.

In the current application of reaction-diffusion patterns, X is the collection of all matured patterns, and the main challenge is to properly define features and patterns. On the other hand, pattern classification could be conveniently benefited from *supervised* learning, as supervised by human vision and popular classification name tags such as *spots* and *stripes*.

Our first scheme for characterizing reaction-diffusion patterns is directly based on the histogram defined in Section 4.1:

$$p_v(\theta) = \lim_{\Delta\theta \to 0} \frac{|\{(x,y) \ : \ \theta \le v(x,y,t) \le \theta + \Delta\theta\}|}{\Delta\theta |\Omega|}, \tag{3.12}$$

where it has been assumed that $t \ge T_s$ (after the pattern matures) so that $p_v(\theta)$ is effectively independent of t (i.e., the stationary assumption).

Let M denote the collection of all continuous probability measures supported on $(0, \infty)$. Then the histogram could be considered as a feature map:

$$p : X \to M : \quad v \to p_v,$$

which maps any matured pattern v to its histogram. From 2-D to 1-D, it already achieves a significant amount of data compression by being blind to the spatial organization of each individual activation level $\Gamma(\theta)$ (see (3.2), with t dropped as well for matured patterns).

Plotted in Figure 3.3 are the histograms of three typical matured patterns corresponding to spots, stripes, and hexagon nets. It appears quite clear to human vision that the histograms have easily differentiated the three. The histogram of the stripes pattern is bimodal and almost symmetric, while those of the spots and nets are both monomodal but tilted toward the opposite ends.

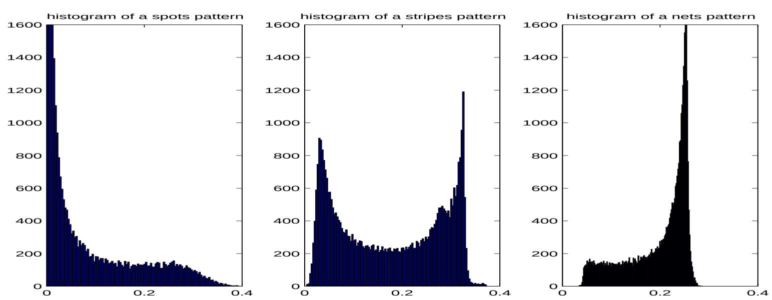

FIG. 3.3. *The histograms of three typical matured patterns.*

This inspires further data compression and feature representation. To characterize the features of a probability measure (such as symmetry, tilting, and resemblance to Gaussians, etc), one could project them onto different templates such as canonical monomials or orthogonal polynomials. Here we work with two high order *invariant moments - skewness* and *kurtosis*, which have been frequently applied in language modelling and image analysis [37].

The skewness of a given continuous probability distribution $p(\theta)$ (supported on $(0, \infty)$) is the 3rd order normalized moment:

$$s[p] = \int_0^\infty \left(\frac{\theta - m}{\sigma} \right)^3 p(\theta) \, d\theta,$$

where m and σ are the mean and standard deviation. Similarly, the kurtosis is the 4th order normalized moment:

$$k[p] = \int_0^\infty \left(\frac{\theta - m}{\sigma} \right)^4 p(\theta) \, d\theta.$$

For the Gaussian class $N(m, \sigma)$ for instance, the skewness always vanishes while the kurtosis is 3. Both the skewness and the kurtosis are invariant under linear transforms of the associated random variables, which makes them ideal and robust for pattern analysis. (See for example Mumford's ICM 2002 paper [37] on the role of kurtosis in language and natural pattern analysis. For more discussion on invariant moments such as cumulants, see for example the recent work by Rota and Shen [51].)

For a given matured pattern v, we are able to define two invariant moments $s_v = s[p_v]$ and $k_v = k[p_v]$. Thus a 2-D feature map is eventually established for the collection of matured patterns:

$$\text{matured R.-D. pattern } v \Rightarrow \quad \text{histogram } p_v \Rightarrow \quad \text{invariant moments } \boldsymbol{F}_v = (s_v, k_v). \tag{3.13}$$

This is certainly a highly *lossy* information compression scheme going from 2-D functions to 2-component vectors. It is, however, highly efficient as shown in Figure 3.4 – the skewness and kurtosis do well differentiate the common patterns.

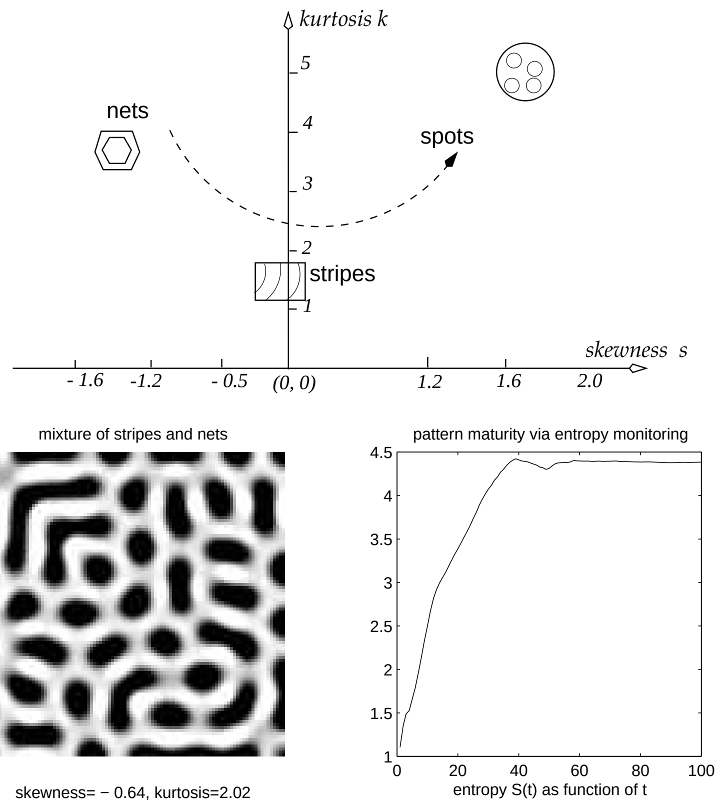

FIG. 3.4. *Typical matured patterns in the skewness-kurtosis feature plane. Some representative values of* (s_v, k_v) *from simulations: spots (1.80, 5.19) and (1.69, 4.70); stripes (-0,03, 1.47) and (-0.24, 1.51); and nets (-1.33, 3.74) and (-1.19, 3.26). For a matured mixture of stripes and nets as shown below,* $(s_v, k_v) = (-0.64, 2.02)$*, which lies between the regions of stripes and nets in the skewness-kurtosis plane. The plane also hints a natural direction from nets to stripes, and eventually to spots, which could also be observed in the comprehensive simulation results by Williams [63].*

As an example of supervised learning (by human vision), one could now easily partition the skewness-kurtosis plane to match the supervised classification of the most common patterns of spots, stripes, and nets.

For instance, for the current model, one could even try simple linear decision rules solely based on the skewness, such as:

$$
\begin{aligned}
E_{\mathrm{sp}} &= \{(s,k) \mid s > 0.8\} \\
E_{\mathrm{st}} &= \{(s,k) \mid -0.8 < s \leq 0.8\} \\
E_{\mathrm{net}} &= \{(s,k) \mid s \leq -0.8\}.
\end{aligned}
$$

Here the decision boundaries 0.8 and -0.8 have been obtained by inspection for the illustrative purpose. Generally, by properly assuming the distribution type (in the skewness-kurtosis half-plane) for each pattern, one could obtain the decision boundaries from sufficiently many simulation data, and pattern analysis methods such as cross-validation and maximum likelihood estimation (MLE) [17]. For example, in the case of identical isotropic Gaussians and MLE, pattern classification is tightly connected to Voronov partitioning of the skewness-kurtosis half-plane.

On the other hand, one could also question whether it makes perfect sense to partitions all matured patterns of a given reaction-diffusion system into three or several specific patterns. This is because that some matured patterns could be the mixtures of these discretized or quantized patterns, like the one shown at the bottom of Figure (3.4). For some patterns, even human vision cannot make a binary decision. Thus the decision boundaries must be understood as gray transitional areas instead of clear cut binary ones. In this sense, the (s,k) half-plane itself is a good representation.

3.4. Statistical Pattern Analysis via Geometric Features. Histograms of activation levels leave out the information of their geometric organization. In this section, we intend to incorporate more geometric information into the statistical schemes.

In modern image and vision analysis, linear filters have become a common tool for representing and analyzing various spatial features such as orientations, multiple scales or resolutions, and local spatial wave numbers. Some well known examples include the Gabor filters [32] and many family of linear filters in wavelets design [14, 58, 65, 66].

Compared with general image patterns such as textures and clusters of natural scenes, reaction-diffusion patterns are relatively better organized due to the nature of the system (i.e., controlled by a system of partial differential equations). Thus geometric regularity seems to be more important and pertinent than other common image features. This observation motivates the application of nonlinear geometric filters such as *curvatures*. The curvature feature has been actively employed in modern image and vision analysis, such as in *mean curvature motions* [19], and image denoising, deblurring, inpainting, and vision modeling [52, 8, 7, 9, 57].

For a smooth image $v = v(x,y)$, the *naturally signed* curvature is defined to be

$$\kappa_v = \kappa_v(x,y) = \nabla \cdot \left[\frac{\nabla v}{|\nabla v|}\right] = \mathrm{div}(\boldsymbol{n}), \qquad \boldsymbol{n} = \frac{\nabla v}{|\nabla v|}, \tag{3.14}$$

which is well defined at all regular points with nonzero gradients. Notice that κ_v is *morphologically invariant*: $\kappa_{g(v)} = \kappa_v$ for any monotonic function g, which should be the case since curvature is geometric and only belongs to level sets. To avoid the singularity on homogeneous regions where the gradients almost vanish, as well practiced in image processing [52, 9], κ_v is often regularized to

$$\kappa_v = \kappa_v(x,y) = \nabla \cdot \left[\frac{\nabla v}{|\nabla v|_\varepsilon} \right], \text{ with } |a|_\varepsilon = \sqrt{a^2 + \varepsilon^2}, \tag{3.15}$$

for some small regularization constant ε. This regularized curvature, though not strictly morphologically invariant, is well defined everywhere.

On the other hand, the curvature information on homogeneous regions is less significiant than where activation levels change rapidly, e.g., along the "edges" of spots or stripes. Therefore, instead of the *pure* curvature, we consider its weighted version:

$$\psi_v(x,y) = |\nabla v| \kappa_v = |\nabla v| \; \nabla \cdot \left[\frac{\nabla v}{|\nabla v|} \right] = \Delta v - D^2 v(\boldsymbol{n}, \boldsymbol{n}), \tag{3.16}$$

where Δv denotes the Laplacian, and $D^2 v$ the Hessian bilinear form. The ε-regularization technique (3.15) applies to ψ_v as well.

Note that the weighted curvature ψ_v is no longer morphologically invariant due to the contrast weighting $|\nabla v|$. To establish robust comparison schemes among different patterns, we shall work with the statistically normalized matured patterns, as in the definitions of skewness and kurtosis. Define the empirical mean and variance of v to be

$$\begin{aligned} m &= \frac{1}{|\Omega|} \int_\Omega v(x,y) dx dy = \int_0^\infty \theta p_v(\theta) d\theta, \\ \sigma^2 &= \frac{1}{|\Omega|} \int_\Omega (v-m)^2 dx dy = \int_0^\infty (\theta - m)^2 p_v(\theta) d\theta. \end{aligned}$$

Then we investigate the statistics of the normalized weighted curvature:

$$\phi_v = \psi_{\frac{v-m}{\sigma}} = \frac{1}{\sigma} \psi_v = \frac{1}{\sigma} (\Delta v - D^2 v(\boldsymbol{n}, \boldsymbol{n})). \tag{3.17}$$

The histogram of ϕ_v shall be denoted by $q_v(\phi)$, and Figure 3.5 displays the histograms for some typical matured patterns. It is quite clear to human vision that these geometry based histograms successfully differentiate different patterns.

Unlike the ones in the proceeding section, geometry based histograms cannot be translated along its horizontal axis, since zero curvature has its intrinsic meaning. The peaks at zero in the weighted curvature histograms in Figure 3.5 for example, reflect the fact that most sites in a matured reaction-diffusion pattern are almost homogeneous. In particular, the mean itself is a valuable statistical feature containing crucial pattern information.

Thus we define a feature map based on the first four moments:

$$v \Rightarrow \boldsymbol{F}_g = \boldsymbol{F}_g(q_v(\phi)) = (m_g, \sigma_g, s_g, k_g),$$

i.e., the mean, standard deviation, skewness, and kurtosis. The symbol g stands for geometry. From the empirical data shown in Figure 3.5, it is clear that the mean and kurtosis are outstandingly effective in differentiating the patterns. Thus one could further simplify the above four-component feature map to a two-component one:

$$v \Rightarrow \boldsymbol{F}_g = (m_g, k_g).$$

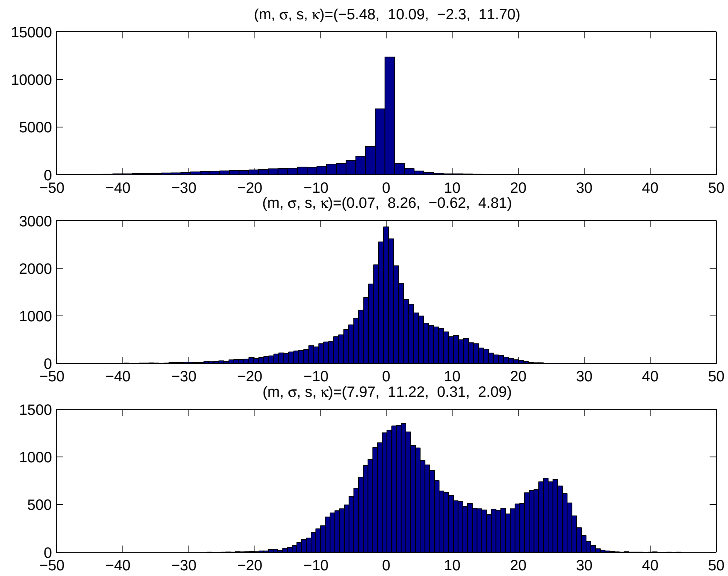

FIG. 3.5. *Histograms of the normalized weighted curvatures ϕ_v's (see Eqn. (3.17)) for typical patterns of spots, stripes, and nets. [For a good axis scale, the histograms are actually for $h \times \phi_v$ where $h = 0.01$ is the pixel size.] The title lines give the corresponding information on the mean, standard deviation, skewness, and kurtosis.*

Figure 3.6 shows the typical matured patterns on the (m_g, k_g) half-plane. It feature plane clearly resolves different patterns. The discussion on pattern quantization based on supervised learning is similar to the proceeding section.

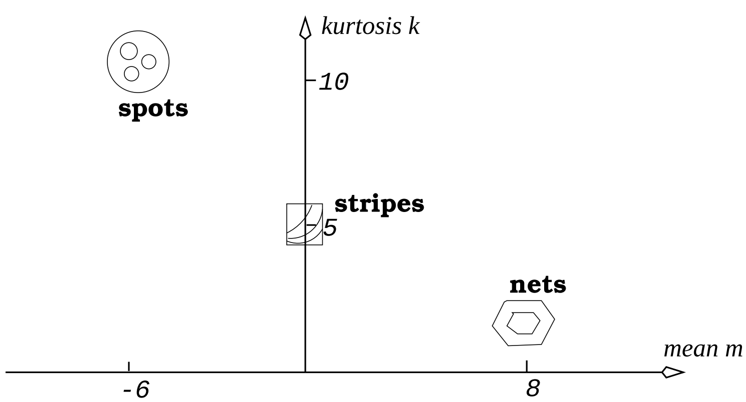

FIG. 3.6. *Typical matured patterns in the geometric mean-kurtosis (m_g, k_g) half-plane.*

3.5. Universality of the Statistical Features via Qualitative Models. In this section, we explain why the statistical features observed for some typical patterns from the Gray-Scott model are universal for general reaction-diffusion patterns.

3.5.1. Histogram features of spots, stripes and nets. A matured spots pattern typically consists of isolated and almost circular islands in a darker sea background. A nets pattern, on the other hand, could be qualitatively treated as the reversal of the regular spots pattern, in the sense that it consists of dark holes in a brighter background. This explains why the histograms of the two are approximately the reversals of each other (see Figure 3.3).

Thus it suffices to model the histogram features of the spots and stripes patterns only.

Consider a normalized circular model for a single spot in R^2:

$$v(x,y) = g(r^2) = g(x^2+y^2),$$

where the rotational generator $g(x)$ satisfies

$$g(x) \in C^1[0,\infty), \quad g(0)=1, \quad g(+\infty)=0, \quad \text{and } g \text{ decreases monotonically from 0 to } +\infty. \tag{3.18}$$

(Notice that due to the squared argument r^2, $\nabla v(0.0) = g'(0)(0,0) = (0,0)$, implying a smooth spot center.) For example, $g(r^2) = e^{-\frac{r^2}{2\sigma^2}}$ is a Gaussian spot, and

$$g(r^2) = \begin{cases} (1-r^2)^2, & r \le 1 \\ 0, & r > 1 \end{cases} \tag{3.19}$$

is a compactly supported spot. For the convenience of superimposing many isolated spots, we shall now work with a typical spot model (3.18) whose compact support is the unit disk, which implies that $g'(1) = 0$. In addition, assume the domain Ω is exactly the unit disk. Let $x = f(\theta)$ denote the inverse function of $\theta = g(x)$ with $x \in [0,1]$ and $\theta \in [0,1]$, which is well defined since g is monotonically decreasing. Then by the definition of the empirical histogram,

$$\begin{aligned} p_v(\theta) &= \lim_{\Delta\theta\to 0} \frac{|\{(x,y) : \theta \le g(r^2) < \theta + \Delta\theta\}|}{|\Omega|\ \Delta\theta} = \lim_{\Delta\theta\to 0} \frac{|\{(x,y) : f(\theta+\Delta\theta) \le r^2 < f(\theta)\}|}{|\Omega|\ \Delta\theta} \\ &= \frac{1}{|\Omega|} \lim_{\Delta\theta\to 0} \frac{\pi(f(\theta) - f(\theta+\Delta\theta))}{\Delta\theta} = -\frac{\pi}{|\Omega|} f'(\theta) = -\frac{\pi}{|\Omega|}\frac{dr^2}{d\theta}. \end{aligned}$$

As a result,

$$p_v(0^+) \propto -f'(0^+) = -1/g'(1^-) = +\infty,$$

which well explains the high peak at the left end of the histogram in the spots pattern (see Figure 3.3). Figure 3.7 displays the predicted histogram for the spot model (3.19).

In a realistic pattern when N spots almost independently float on a uniform darker background, the histogram is simply

$$p_v(\theta) = \frac{p_1(\theta)|\Omega_1| + \cdots + p_N(\theta)|\Omega_N|}{|\Omega|}, \tag{3.20}$$

where $\Omega_1 \cup \cdots \cup \Omega_N$ partitions Ω into effective influence regions associated to the N spots. When the profiles of the spots are very similar, or identical most ideally, then we have $p_v(\theta) \equiv p_1(\theta)$. For general situations

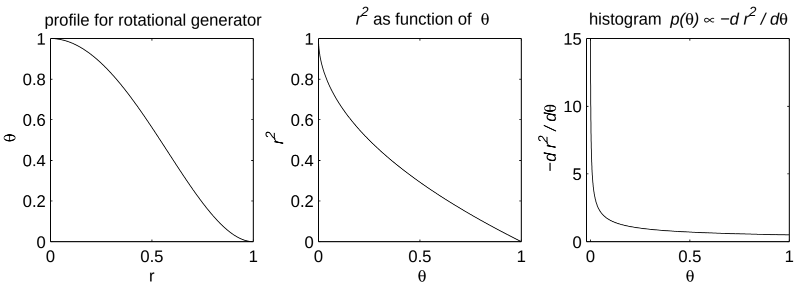

FIG. 3.7. *The histogram of a model spot:* $v(x, y) = g(r^2) = g(x^2 + y^2)$, *with* $g(s) = (1 - s)^2$ *for* $s \in [0, 1]$, *and* 0 *otherwise. This model histogram qualitatively matches that of the spots pattern in Figure 3.3.*

where the sizes and profiles of the spots do vary slightly, formula (3.20) could be thought of a smoothing (or filtering) process of the histograms of individual spots.

We now explain the histogram features for the stripes pattern. Consider a *canonical* ideal stripe along the y-direction on the square $Q = [0, 1] \times [0, 1]$:

$$v(x, y) = g(x) \quad \text{with} \quad \begin{cases} g(0) = 0, \quad g(1) = 1; \quad \text{and} \\ g'(0) = 0, \quad g'(1) = 0, \quad \text{and} \quad g' > 0. \end{cases}$$

Furthermore, we assume that g is symmetric: $g(x) = g(1-x)$ about the stripe "edge" $x \equiv 1/2$. For example, one could take

$$g(x) = \frac{1 - \cos(\pi x)}{2} = \sin^2 \frac{\pi}{2} x.$$

Let $x = f(\theta)$ denote the inverse function of $\theta = g(x)$. Then by the definition of empirical histogram, it is easy to obtain

$$p_v(\theta) = f'(\theta) = \frac{1}{g'(x)},$$

which well explains the blowups at the two ends in the histogram of the stripes pattern (see Figure 3.3), since

$$p_v(0^+) = \frac{1}{g'(0^+)} = +\infty \quad \text{and} \quad p_v(1^-) = \frac{1}{g'(1^-)} = +\infty.$$

As an example, consider $\theta = g(x) = \frac{1-\cos(\pi x)}{2}$. Since $g'(x) = \frac{\pi}{2} \sin(\pi x)$, one has

$$p_v(\theta) = \frac{2}{\pi} \frac{1}{\sqrt{1 - (2\theta - 1)^2}} = \frac{1}{\pi} \frac{1}{\sqrt{\theta(1 - \theta)}} \ .$$

Thus for this ideal model, the empirical histogram is precisely the beta distribution $B(\frac{1}{2}, \frac{1}{2})$ on (0, 1), which also well explains the bimodal pattern of the actual stripes histogram in Figure 3.3.

3.5.2. Feature interpretation for the geometric histograms. We now qualitatively explain why the geometric histograms of different patterns are noticeably concentrated on different regions (see Figure 3.5).

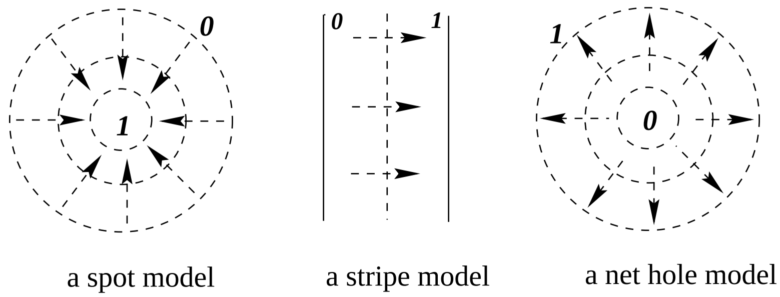

FIG. 3.8. *The normal vector fields near the "edges" of different patterns. The normal field is contractive for a spots pattern, expansive for a nets pattern, and almost constant for an ideal stripe pattern, leading to negative, positive, and almost zero divergences (i.e., curvatures).*

Consider the normal fields $\boldsymbol{n} = \nabla v / |\nabla v|$ of the patterns of spots, stripes, and nets. Near the "edges" where $|\nabla v|$ is relatively large, the normal fields are typically depicted by Figure 3.8. For a spots pattern, the normal field $\boldsymbol{n}$ is contractive and points toward the peak at the center. Therefore, its divergence $\kappa_v = \mathrm{div}(\boldsymbol{n})$ (i.e., curvature) must be negative by the divergence theorem. This explains why in Figure 3.5, the geometric histogram for $\phi_v = |\nabla v|\kappa_v$ associated to the spots pattern is mostly concentrated on the negative axis. The situation reverses for a nets pattern. For an ideal *local* stripe, the normal field is close to be constant which leads to almost zero curvatures.

On the other hand, for all the three patterns, the peaks at $\phi_v (= |\nabla v|\kappa_v) = 0$ are due to the fact that away from the transitionary edges, v is almost a constant and $|\nabla v|$ close to zero.

4. Geometric Measures of Reaction-Diffusion Patterns. In this section, we employ tools from geometric measure theory to answer questions like: what is the average spot size in a spots pattern? how many are there on a square inch on average? and what is the aspect ratio of the stripes in a stripes pattern. Many of these tools have been recently revived in mathematical image and vision analysis.

4.1. Total Variation and Curvatures. The broad applications of the total variation (TV) measure in image processing started from Rudin, Osher, and Fatemi's work on TV based image denoising and deblurring [52, 53]. From the applied harmonic analysis point of view, total variation and its sibling Besov norms are also naturally connected to wavelets based image denoising and regularization [12, 16, 35].

The total variation $\mathrm{TV}[u]$ or $\int_\Omega |Du|$ of an integrable function $v \in L^1(\Omega)$ generalizes the regular Sobolev norm $\int_\Omega |\nabla u| dxdy$, and is defined in the distributional sense [24]:

$$\mathrm{TV}[u] = \int_\Omega |Du| = \sup_{\boldsymbol{g} \in C_0^1(\Omega, B^2)} \int_\Omega u(\nabla \cdot \boldsymbol{g}) dxdy, \tag{4.1}$$

where $C_0^1(\Omega, B^2)$ denotes all compactly supported C^1 vectorial functions that take values inside the unit disk in R^2:

$$B^2 = \{\boldsymbol{g} = (g_1, g_2) \ : \ |\boldsymbol{g}| = \sqrt{g_1^2 + g_2^2} \leq 1\}.$$

Applying (4.1) to any open subset of Ω, one could define rigorously the set function $\int_\bullet |Du|$ as a *finite* Radon measure on Ω, provided that $\mathrm{TV}[u] < \infty$.

Unlike the Sobolev norm in $W^{1,1}(\Omega)$, the total variation of the indicator function χ_E of a Borel set $E \subset \Omega$ is well defined (though can be ∞), and is finite for any Lipschitz domain with finite 1-D Hausdorff measure of its boundary. In the latter case they are in fact identical. Thus total variation leads to a more general definition of the perimeter of a Borel set $E \subset \Omega$ by:

$$\mathrm{Per}(E|\Omega) = \int_\Omega |D\chi_E|.$$

Any function $u \in L^1(\Omega)$ with finite total variation is said to have bounded variation (BV). The space of all such functions is denoted by $\mathrm{BV}(\Omega)$, which is a Banach space under the BV-norm:

$$\|u\|_{\mathrm{BV}} = \|u\|_{L^1} + \int_\Omega |Du|.$$

It is the space of $\mathrm{BV}(\Omega)$ that has been recently widely applied as an approximate but efficient image model for general images [35, 9, 10].

The geometric nature is reflected in the celebrated co-area formula of Fleming and Rishel [20] and De Giorgi [23]:

$$\mathrm{TV}[u] = \int_\Omega |Du| = \int_R \mathrm{length}(u \equiv \lambda) d\lambda, \tag{4.2}$$

where $u \equiv \lambda$ denotes the λ-level set, assuming that u is regular. For general BV functions whose individual level sets are less meaningful, the co-area formula is given by

$$\mathrm{TV}[u] = \int_\Omega |Du| = \int_R \mathrm{Per}(u < \lambda|\Omega) d\lambda,$$

where $u < \lambda$ represents the set $E_\lambda = \{(x,y) \in \Omega : \ u(x,y) < \lambda\}$. Thus total variation provides a natural way to collectively sum up the lengthes of *all* the level sets, which has made it so powerful in image processing and vision analysis.

Although general smoothing operators such as heat diffusion are not continuous in $\mathrm{BV}(\Omega)$, they are often continuous in terms of the value of the TV measure. That is, suppose $u_\varepsilon \in W^{1,1}(\Omega)$ or $C^\infty(\Omega)$ is a family of mollifications of $u \in \mathrm{BV}(\Omega)$ (often spatially adaptive when Ω is bounded), so that $u_\varepsilon \to u$ in L^1 as $\varepsilon \to 0$. Then,

$$\mathrm{TV}[u_\varepsilon - u] \to 0 \ \text{ is generally untrue, while } \ \mathrm{TV}[u_\varepsilon] \to \mathrm{TV}[u] \ \text{ could be easily achieved.} \tag{4.3}$$

If the sequence is not from a mollification design, then generally only the *lower semi-continuity* could be guaranteed [24]. This observation is very useful for numerically computing the total variation norm using ordinary schemes such as finite differences (also see later sections).

For BV functions with smooth level sets, such as the indicators of smooth domains, one could further investigate high order geometric measures. For instance, the m-th order total curvature is defined by

$$K_m[u] = \int_\Omega \kappa^m |Du| = \int_R \left(\int_{u\equiv\lambda} \kappa^m ds \right) d\lambda, \tag{4.4}$$

where ds denotes the arc length element along the level lines, and $\kappa = \nabla \cdot (\nabla u / |\nabla u|)$. For instance, Euler's elastica measure $E[u] = a\mathrm{TV}[u] + bK_2[u]$ has numerous applications in image analysis ever since Mumford first introduced it into computer vision [38, 43, 34, 56, 5, 18, 54]. Theoretical difficulties on analyzing high order geometric measures still remain open (see for example, the remarkable work by the De Giorgi school Bellettini, Dal Maso, and Paolini [2]).

4.2. Binary Segmentation of Reaction-Diffusion Patterns. When one asks questions like: what is the average radius of the spots, or the average width of the stripes, it has already been assumed that the spots and stripes are isolated "objects" with clear-cut boundaries. The human vision system could carry out this so called *segmentation* task unconsciously but effortlessly even for complex images. Its mathematical modelling and algorithm design, however, have turned out to be highly nontrivial, challenging, but the most fundamental in a number of fields. For example, the celebrated Mumford-Shah segmentation model and its variations have been proven very powerful in segmenting complicated image scenes with low textures [41, 11]. Another well known method if David Marr's zero-crossing theory [33].

Reaction-diffusion patterns are more regular compared with general natural images. Thus in the current work, we adopt a simpler segmentation scheme which is directly based on binary thresholding. Let $v_2(x, y)$ be the binary image perceived by human vision from the matured pattern $v(x, y)$. Then we model v_2 by

$$v_2(x, y) = \begin{cases} 1, & v \geq \theta_c \\ 0, & v < \theta_c, \end{cases}$$

where θ_c is the critical activation level for binary thresholding. An effective threshold θ_c has to depend on the image $v(x, y)$ itself: $\theta_c = \theta_c[v]$.

In what follows, we take θ_c to be the mean value:

$$\theta_c = \theta_c[v] = \frac{1}{|\Omega|} \int_\Omega v dx dy = \int_0^\infty \theta p_v(\theta) d\theta.$$

From the statistics in the proceeding section, the histograms for typical matured stripes patterns are almost symmetric, and the two peaks at the two ends correspond to the dark valleys and bright peaks in $v(x, y)$ (Figure 3.3). Thus the mean provides a natural cut-off for the boundaries. For a typical spots pattern v, the mean is tilted toward the dark background, and the associated binary segmentation often produces "fatter" bright spots compared with human perception. The situation reverses for a typical nets pattern. However, due to the well known reaction-diffusion mechanism of short-range activation and long-range inhibition, the transition is often rapidly realized. As a result, the mean based binary segmentation cannot be too fat nor too slim. Figure 4.1 shows two examples of mean based binary segmentation of matured reaction-diffusion patterns.

From now on, we shall only work with the binary images v_2.

4.3. Geometric Analysis of Spots and Nets. As clear from Figure 4.1, all spots are almost circular. Thus consider the following approximate model of spots distribution. Suppose there are totally N non-overlapping circular white spots in the binary image v_2, and their radii are

$$r_1, r_2, \cdots, r_N, \qquad \text{subject to certain unknown random distribution.} \tag{4.5}$$

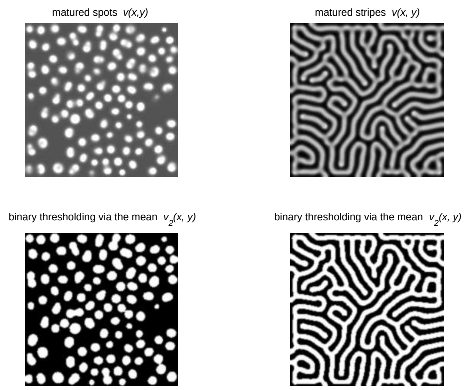

FIG. 4.1. *Two examples of mean based binary segmentation of matured reaction-diffusion patterns.*

Definition [Average Radius]. The average radius $\bar{r}$ of a binary spots pattern v_2 is defined to be:

$$\bar{r} = \langle r \rangle = \frac{r_1 + r_2 + \cdots + r_N}{N}. \tag{4.6}$$

By the definition, it appears natural to first compute all the individual radii and then have them averaged. Such computation is generally very costly due to the individual segmentation, identification, and information extraction. Part of the beauty of the geometric measure approach is that it carries out all the tasks at once.

THEOREM 4.1. *Suppose the circular spots do not overlap in a binary image v_2. Then the average radius is given by $\bar{r} = \dfrac{1}{2\pi N} \displaystyle\int_\Omega |Dv_2|$.*

Proof. Let $D_k \subset \Omega$ denote the disk region corresponding to the k-th circular spot with radius r_k, $k = 1 : N$. Then we must have

$$v_2(x, y) = \sum_{k=1}^{N} \chi_{D_k}(x, y).$$

Since D_k's have been assumed non-overlapping,

$$\int_\Omega |Dv_2| = \sum_{k=1}^{N} \int_\Omega |D\chi_{D_k}| = \sum_{k=1}^{N} \mathrm{Per}(D_k) = 2\pi \sum_{k=1}^{N} r_k.$$

This leads to the conclusion. ∎

For simplicity it has been assumed that there are no incomplete spots along the boundary of the reaction-diffusion domain Ω, which is at least a very good approximation since the population of incomplete spots is a negligible fraction. In Figure 4.1 for example, the binary image v_2 contains only two incomplete spots along the boundary, out of a total about 100 ones.

The main issue with Theorem 4.1 is that in reality N is unknown either. How to count the number without getting down to the individual level? This is again made possible by suitable geometric measures.

To start, suppose that $r_k = \bar{r} + \delta r_k$, $k = 1 : N$, with

$$\langle \delta r \rangle = 0 \quad \text{and} \quad \sigma^2 = \langle (\delta r)^2 \rangle \ll 1. \tag{4.7}$$

Then,

$$\frac{1}{\pi N}\int_\Omega v_2 dxdy = \frac{1}{\pi N}\sum_{k=1}^{N}\int_\Omega \chi_{D_k} dxdy = \frac{1}{N}\sum_{k=1}^{N} r_k^2,$$

which is precisely the second empirical moment $\langle r^2\rangle$ for the random radii $(r_k)_{k=1}^N$. Since $\sigma^2 = \langle r^2\rangle - \bar{r}^2$, we have

$$\frac{1}{\pi N}\int_\Omega v_2 dxdy = \bar{r}^2 + \sigma^2.$$

Combined with Theorem 4.1, it gives

$$\frac{[\pi N]^{-1}\int_\Omega v_2 dxdy}{[2\pi N]^{-1}\int_\Omega |Dv_2|} = \frac{\bar{r}^2 + \sigma^2}{\bar{r}} = \bar{r}(1+\varepsilon^2), \tag{4.8}$$

where $\varepsilon = \sigma/\bar{r}$. Therefore, we have proved the following theorem.

THEOREM 4.2. *Suppose the signal-to-noise ratio (SNR) $\varepsilon^{-1} = \bar{r}/\sigma \gg 1$, then the average radius is*

$$\bar{r} = \frac{2\int_\Omega v_2 dxdy}{\int_\Omega |Dv_2|}\ (1 + O(\varepsilon^2)). \tag{4.9}$$

Compared with Theorem 4.1, Theorem 4.2 does not require the *a priori* information of N, as long as the SNR is high, or equivalently, the variation of the radii is small.

The combination of Theorem 4.1 and Eqn. (4.8) also gives a formula for N.

THEOREM 4.3. *Following the assumptions of Theorem 4.2, the total number N of spots is given by*

$$N = \frac{\left[\int_\Omega |Dv_2|\right]^2}{4\pi\int_\Omega v_2 dxdy}\ (1 + O(\varepsilon^2)). \tag{4.10}$$

Note that the computation of the TV measure can be benefited from the discussion in the equation line (4.3). In all the numerical examples in the current section, we always employ the heat diffusion as the mollifier $u(x, y; t)$:

$$\begin{cases} u_t = \Delta u, & (x,y) \in \Omega; \quad \text{with} \\ \left.\dfrac{\partial u}{\partial \boldsymbol{n}}\right|_{\partial\Omega} = 0, & u(x,y,0) = v_2(x,y), \end{cases}$$

and approximate TV measure $\int_\Omega |Dv_2|$ by the Sobolev norm $\int_\Omega |\nabla u(x,y,\tau)| dxdy$ for some small time step τ (often several numerical steps). The computation of the latter is easily achieved by uniform finite differences.

As an example, we have applied the formulae in Theorem 4.2 and 4.3 to the binary spots pattern v_2 in Figure 4.1. Our numerical results show that

$$\bar{r} = \frac{A}{33} \quad \text{and} \quad N = 103,$$

where the average radius has been compared with the side length A of the square domain. Thus the average diameter of the spots is about 1/16 of A. On the other hand, out manual counting of all the spots gives $N = 107$, which has even taken into account automatic splitting (by using human vision intelligence) of seemingly merged spots pairs. Thus Theorem 4.2 and 4.3 do perform remarkably well.

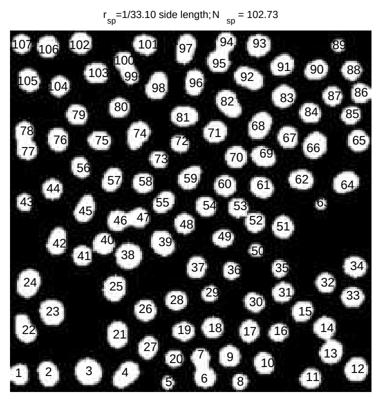

FIG. 4.2. *Estimating the average radius $\bar{r}$ and total number N of spots for a segmented binary spots pattern by Theorem 4.2 and 4.3. Our results are $\bar{r} = A/33$ and $N = 103$ where A denotes the side length of the domain. The time consuming manual counting as marked on top of the image gives $N = 107$, which has even taken into account automatic splitting of some "merged" pairs relying on human visual intelligence (e.g., pairs 6 and 7, 46 and 47, 77 and 78, 94 and 95, and 99 and 100).*

4.4. Geometric Analysis of Stripes Patterns. In this section, theory and a number of interesting tools are developed to study important geometric features of stripes patterns.

4.4.1. Zebra curves: Definitions and properties. Recall that a $C^{(m)}$ diffeomorphism $\phi : U \to V$ between two open sets in R^2 (or more generally, between two differential manifolds) is a $C^{(m)}$ map $\boldsymbol{q} = \phi(\boldsymbol{p})$ whose inverse $\boldsymbol{p} = \psi(\boldsymbol{q}) : V \to U$ exists and is also $C^{(m)}$.

Definition [Zebra Curves]. A compact set $\Gamma \subset R^2$ is said to be a $C^{(m)}$ *zebra curve* if for any $\boldsymbol{p} \in \Gamma$, there is an open neighborhood $U \subset R^2$, and a $C^{(m)}$ diffeomorphism $\phi : U \to R^2$, under which $\phi(\boldsymbol{p}) = (0, 0)$ and $\Gamma \cap U$ is mapped onto either the x-axis $R^1 \times 0$, the half axis $R^+ \times 0$, or the standard *convex* Y-junction:

$$Y_\theta = R^- \times 0 \bigcup 0 \times R^- \bigcup R^+(\cos\theta, \sin\theta), \quad \text{with} \;\; 0 \leq \theta < \pi/2. \tag{4.11}$$

Here $R^+ = [0, \infty)$ and $R^- = (-\infty, 0]$. (See Figure 4.3.)

Explanation is needed to specifi cally address the notion of *convex* Y-junctions. First, the word convexity refers to that the convex hull of Y_θ for all $0 < \theta < \pi/2$ is the whole plane R^2. It is in this sense that any Y_θ with $\theta \in (\pi/2, 2\pi)$ is not convex. Remarkably, visual inspection of any generic matured stripes pattern shows

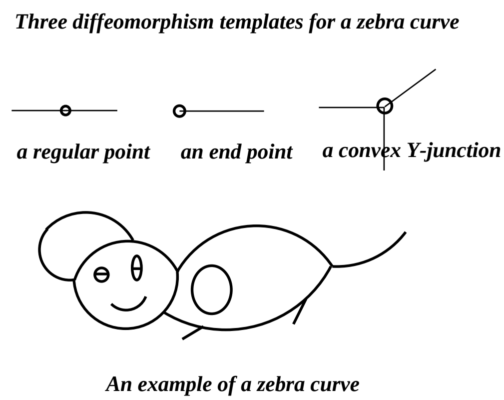

FIG. 4.3. *Zebra curves: The three diffeomorphism templates and an example.*

that all the Y-junctions seem to be convex (see Figure 4.1 for example). The essential difference between convex and non-convex Y-junctions is addressed in [27], and will also be mentioned later.

Secondly, the value $\theta = 0$ is permitted in the definition to include T-junctions, i.e., structures that are diffemorphic to $R \times R^-$. T-junctions are fundamentally important in human and machine vision [43, 28].

Finally, the notion of convexity is invariant under diffeomorphism since locally a diffeomorphism coincides with its non-degenerate Jacobian, and convexity as a linear concept is invariant under invertible linear transforms.

For convenience, we shall denote by Γ° all the regular points on a zebra curve Γ which are locally diffeomorphic to $R \times 0$, and Γ^* all the singular points on Γ which are either diffeomorphic to $R^+ \times 0$ or a convex Y-junction Y_θ.

We shall also assume that the diffeomorphisms involved are all at least $C^{(3)}$, so that second order geometric features such as curvatures could be properly defined and easily controlled.

THEOREM 4.4. *Let Γ be a $C^{(m)}$ compact zebra curve. Then, (a) the singular set Γ^* is finite; and (b) the regular set Γ° has finitely many connected components, each of which is $C^{(m)}$ diffeomorphic to R^1 or S^1.*

Proof. By definition, a singular point, whether an end point or a Y-junction, must be isolated. If Γ^* is infinite, it must have at least one limit point $\boldsymbol{q}$ in Γ since Γ is compact. Now that $\boldsymbol{q}$ can no longer be singular, it must be regular. But by definition, a regular point has at least one neighborhood that contains only regular points, which leads to contradiction. This proves (a).

By definition, the regular set Γ° is a 1-D $C^{(m)}$ submanifold in R^2. According to the well known classification theorem for 1-D connected manifolds [36], each connected component of Γ° must be diffeomorphic to either S^1 (in the compact case) or R^1 (non-compact case). Since the zebra curve is compact, the number of components diffeomorphic to S^1 must be finite. To prove (b), therefore, it suffices to show that the number of connected components diffeomorphic to R^1 is finite as well. For convenience, such components are said to be *open*.

Let C denote any open component. It can be naturally oriented and parameterized by R^1: $\boldsymbol{p} = \boldsymbol{p}(t), -\infty < t < \infty$. Let $\boldsymbol{q}_+ \in R^2$ be one of the limit points of $\boldsymbol{p}(t)$ as $t \to +\infty$, and $\boldsymbol{q}_-$ as $t \to -\infty$.

They must exist and belong to Γ since Γ is compact. Notice that $\boldsymbol{q}_{\pm}$ cannot be regular since otherwise locally near $\boldsymbol{q}_{\pm}$, Γ is diffeomorphic to $R^1 \times 0$ in R^2. Thus $\boldsymbol{q}_{\pm} \in \Gamma^*$. Let $|\Gamma^*|$ denote the cardinality of the singular set, which is finite by (a). Then by tracking the unique unordered pair of $\boldsymbol{q}_{\pm}$ for each open component, one could easily show that the total number of open components is bounded by $3 \times |\Gamma^*|/2$. This completes the proof. □

As a result of the endpoint-tracking technique in the proof, we also conclude that

COROLLARY 4.5. *Let C denote any open connected component of the regular set Γ°. Then its closure in R^2 is diffeomorphic to either a compact interval $[a, b] \times 0$ in R^2, or the unit circle S^1 with at most one corner.*

In combination, the theorem and corollary imply that the total length $L = |\Gamma|$ of a zebra curve Γ is well defined and must be finite.

Definition [Zebra Stripes Pattern]. Let Γ be any compact zebra curve and $d > 0$ a positive parameter. Then the open neighborhood

$$(\Gamma)_d = \{\boldsymbol{p} \,:\, \operatorname{dist}(\boldsymbol{p}, \Gamma) < \frac{d}{2}\} \tag{4.12}$$

is call a *zebra stripes pattern*, and d its width. For convenience, Γ is called the *skeleton* of the pattern.

The main challenge is the inverse problem: for a given open domain which is the zebra stripes pattern built upon some unknown zebra skeleton curve, how to efficiently compute its width d, and the length L?

4.4.2. Geometric measures for width, length, and singular sets. As for spots patterns, we start with simple stripe models for developing and analyzing proper geometric measures.

Let C denote a $C^{(m)}$ simple curve in R^2, parametrized by its arc-length: $\boldsymbol{r} = \boldsymbol{r}(s)$ with $0 \leq s \leq L$, where L is the total length. As a special case of zebra curves, let $(C)_d$ denote its open neighborhood defined by (4.12). Assume that the aspect ratio $L/d \gg 1$, which is tantamount to letting $d \to 0$ since L will be fixed in the analysis.

Let T and N denote the unit tangent and normal of C (see Figure 4.4). Define a map from $D = [0, L] \times R$ to R^2 by:

$$(s, t) \to \boldsymbol{p}(s, t) = \boldsymbol{r}(s) + tN(s). \tag{4.13}$$

By the implicit function theorem, there exists some $b > 0$, so that $\boldsymbol{p}(s, t)$ is a diffeomorphism from $D_a = [0, L] \times (-b/2, b/2)$ to its image in R^2. For any $d < b$, $(C|d) = \boldsymbol{p}(D_d)$ is called a *tubular neighborhood* of C in differential topology [36]. Due to its explicit formula, the tubular neighborhood $(C|d)$ is much easier to work with than $(C)_d$ defined in (4.12). On the other hand, as $d \to 0$, they coincide with each other (except at the two ends). The definition of $(C)_d$, however, is more general since it belongs to general metric spaces.

Let $C(t)$ denote the parametric curve $\boldsymbol{p}(s, t), 0 \leq s \leq L$ for any fixed t. Then

$$\boldsymbol{p}_s(s, t) = T(s) - t\kappa T(s) = (1 - t\kappa)T(s),$$

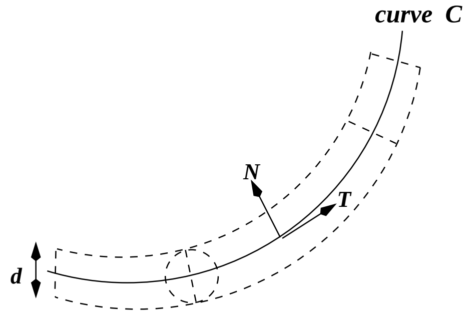

FIG. 4.4. *The tubular neighborhood $(C|d)$ of a simple curve segment C.*

where κ is the curvature. Thus the length $|C(t)|$ is give by

$$|C(t)| = \int_0^L |1 - t\kappa| ds \quad \text{leading to} \;\; \big|\, |C(t)| - L\big| \leq |t| \int_0^L |\kappa(s)| ds = |t| K_a, \tag{4.14}$$

where K_a denotes the total *absolute* curvature of C.

Now consider the total area of the tubular neighborhood $(C|d)$. Notice that

$$\boldsymbol{p}_t(s,t) = N(s) \quad \text{and} \quad \boldsymbol{p}_s(s,t) = (1 - t\kappa)T(s).$$

Therefore,

$$\text{area}(C|d) = \int_0^L \int_{-d/2}^{d/2} |\boldsymbol{p}_t \times \boldsymbol{p}_s| \; dsdt = \int_0^L \int_{-d/2}^{d/2} |1 - t\kappa| \; dsdt = \int_{-d/2}^{d/2} |C(t)| \; dt. \tag{4.15}$$

Combined with Eqn. (4.14), it leads to

$$|\text{area}(C|d) - Ld| \leq \frac{K_a}{4} d^2. \tag{4.16}$$

Also notice that both Eqn. (4.14) and (4.15) can be made more precise if we assume that

$$\frac{d}{2} \leq \min_{0 \leq s \leq L} \frac{1}{|\kappa(s)|}, \qquad \text{the lower bound of all curvature radii along } C.$$

Then $|t\kappa(s)| \leq 1$ for any $t \in [-d/2, d/2]$, and

$$|C(t)| = L - tK \quad \text{and} \quad \text{area}(C|d) = Ld, \tag{4.17}$$

where $K = \int_C \kappa(s) ds$ is the total *signed* curvature of C.

Finally, let $v_2(\boldsymbol{p}) = v_2(x,y)$ denote the (binary) indicator function of the tubular domain $(C|d)$, whose closure is assumed to be inside a reaction-diffusion domain $\Omega \subset R^2$. Then by Eqn. (4.14)

$$\int_\Omega |Dv_2| = \left| C\left(\frac{d}{2}\right) \right| + \left| C\left(-\frac{d}{2}\right) \right| + 2d = 2L + O(d(K_a + 2)). \tag{4.18}$$

Here the line segments $\boldsymbol{p}(0, [-d/2, d/2])$ and $\boldsymbol{p}(L, [-d/2, d/2])$ at the two ends have been taken into account. On the other hand, by Eqn. (4.16),

$$\int_\Omega v_2 dxdy = \text{area}(C|d) = Ld + O(K_a d^2). \tag{4.19}$$

In combination, we have proven the following theorem.

THEOREM 4.6. *Suppose the aspect ratio $\varepsilon = d/L \ll 1$, and K_a denote the total absolute curvature of C. Then*

$$d = \frac{2\int_\Omega v_2 dxdy}{\int_\Omega |Dv_2|}\,(1 + O(\varepsilon(K_a + 2))) \quad \textit{and} \quad L = \frac{1}{2}\int_\Omega |Dv_2|\,(1 + O(\varepsilon(K_a + 2))), \tag{4.20}$$

where the order symbol $O(\bullet)$ is independent of the given curve C and d.

The above analysis has been established only for a single piece of simple curve C. Note that the number 2 inside the order symbol $O(\bullet)$ is caused by the *two* endpoints of C. For a a general zebra curve Γ and its associated zebra stripes pattern $(\Gamma)_d$ defined in the proceeding section, let v_2 denote the (binary) indicator function of $(\Gamma)_d$ in the interior of a reaction-diffusion domain Ω. By applying Theorem 4.6 to each one of the connected components predicted by Theorem 4.4, and by the special counting effect (as explained later) of the *functional* total absolute curvature K_a

$$K_a = K_a(d) = \int_\Omega |\kappa||Dv_2| = \int_\Omega \left|\nabla \cdot \left[\frac{\nabla v_2}{|\nabla v_2|}\right]\right| |Dv_2|, \tag{4.21}$$

we could prove the following theorem for general zebra stripes patterns.

THEOREM 4.7. *Let Γ be a compact zebra curve with total length L, and $(\Gamma)_d$ its associated zebra stripes pattern with $\varepsilon = d/L \ll 1$. Then*

$$d = \frac{2\int_\Omega v_2 dxdy}{\int_\Omega |Dv_2|}\,(1 + O(\varepsilon K_a)) \quad \textit{and} \quad L = \frac{1}{2}\int_\Omega |Dv_2|\,(1 + O(\varepsilon K_a)), \tag{4.22}$$

This theorem has been applied to a real stripes pattern generated from the reaction-diffusion equation (see Figure 4.5). Let A denote the side length of the reaction-diffusion domain. Then our numerical results show that $d = A/30$ and $L = 15A$, both implying that travelling horizontally (or vertically), one shall encounter 15 black-white cycles on average. Notice that the task of estimating the total length L is extremely difficult for human vision.

Furthermore, if the second order curvature measures are to be employed, it is even possible to extract more detailed geometric information about a stripes pattern. This is being thoroughly investigated in our paper [27]. Below we state one of the most interesting results.

First, divide the singular set Γ^* of a zebra curve Γ into two natural categories:

$$\Gamma_+ = \{\boldsymbol{p} \in \Gamma^* : \ \boldsymbol{p} \text{ is a } Y\text{-junction}\}$$
$$\Gamma_- = \{\boldsymbol{p} \in \Gamma^* : \ \boldsymbol{p} \text{ is an end point}\}.$$

Let $z_\pm$ denote their cardinalities separately. Our goal is to develop suitable geometric measures to compute or estimate $z_\pm$.

Similar to the definition of the total absolute curvature in (4.21), define the functional total *signed* curvature of a zebra stripes patter $(\Gamma)_d$ to be

$$K = K(d) = \int_\Omega \nabla \cdot \left[\frac{\nabla v_2}{|\nabla v_2|}\right] |Dv_2|, \tag{4.23}$$

FIG. 4.5. *Estimating the average width d and the total length L of a matured stripes pattern by Theorem 4.7. Our numerical results show that $d = A/30$ and $L = 15A$ where A is the side length of the reaction-diffusion domain. Both results imply that travelling horizontally (or vertically), one should encounter 15 black-white cycles on average. Notice that estimating the total length L is an extremely challenging task even for human vision.*

where $v_2(x, y)$ denote the indicator function of $(\Gamma)_d$ in Ω.

Definition [Linear Zebra Curves]. A zebra curve Γ is said to be *linear* if any connected component of its regular set Γ° is a line segment.

Then one of the main results in [27] states that

THEOREM 4.8. *Let Γ be a $C^{(m)}$ $(m > 2)$ compact zebra curve inside a reaction-diffusion domain Ω. Then*

$$z_+ - z_- = \lim_{d \to 0} K(d) \tag{4.24}$$

$$z_+ + z_- \le \lim_{d \to 0} K_a(d), \tag{4.25}$$

and the second equality holds if and only if Γ is a linear zebra curve.

In particular, as in spline theory, defi ne

$$x^{\pm} = \frac{|x| \pm x}{2}, \qquad \text{for any real scalar } x \in R,$$

and the two half-sided total curvatures,

$$K_{\pm} = K_{\pm}(d) = \int_\Omega \kappa^{\pm} |Dv_2| \;\; \text{with} \;\; \kappa = \nabla \cdot \left[\frac{\nabla v_2}{|\nabla v_2|} \right].$$

Then Theorem 4.8 leads to the simple formulae:

$$z_{\pm} \le \lim_{d \to 0} K_{\pm}(d), \tag{4.26}$$

and the two equalities hold when Γ is a *linear* zebra curve.

Theorem 4.8 and formulae in (4.26) allow to effectively and robustly compute or estimate the singular points. They depend on the crucial assumption of convexity in the definition of Y-junctions (see the previous section). We refer to [27] for more details.

5. Conclusion. In the current paper, statistical as well as geometric theories and frameworks have been developed to identify, classify, and characterize generic patterns arising from reaction-diffusion systems.

In the vast reaction-diffusion literature that permeates biology, chemistry, and computer simulations, the current paper presents the *first* most systematic work in the efforts of combining pattern-theoretic analysis with pattern formation research. Many novel ideas and methodologies have thus been inevitably developed at the risk of being completely wrong.

The present work once again proves, that the numerous contributions by the community of modern image, vision, and pattern analysis, are *not only* applicable in the important fields of computer vision, artificial intelligence, medical imaging, astronomic data analysis, or natural language processing, *but also* universally powerful for various feature and pattern mining tasks in physical, chemical, and biological simulations.

Acknowledgments. We are profoundly grateful to Professor Hans Othmer for his inspirational teaching on biochemical pattern formation and mathematical biology, without whose generous support and encouragement this project would have been absolutely impossible.

Jackie Shen would like to thank as well Professor Bob Gulliver for his teaching on geometric measure theory and minimal surfaces, and to Doug Arnold for his gift book on vision and cognitive sciences, to his former advisors and dear friends Professors Gilbert Strang, Tony Chan, and Stan Osher for their constant support and trust in the research. Jackie Shen also wishes to thank Professor David Mumford for a visiting position in Brown's Pattern Theory Group in the spring of 2000, which in many ways has helped shape some of the key ideas of the current work.

REFERENCES

[1] L. Alvarez, F. Guichard, P.-L. Lions, and J.-M. Morel. Axioms and fundamental equations of image processing. *Arch. Rational Mech. Anal.*, 123:199–257, 1993.

[2] G. Bellettini, G. Dal Maso, and M. Paolini. Semicontinuity and relaxation properties of a curvature depending functional in 2D. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)*, 20:247–297, 1993.

[3] V. Castets, E. Dulos, J. Boissonde, and P. De Kepper. Experimental evidence of a sustained standing Turing-type non-equilibrium chemical pattern. *Phys. Rev. Lett.*, 64:2953–2956, 1990.

[4] A. Chambolle and P. L. Lions. Image recovery via Total Variational minimization and related problems. *Numer. Math.*, 76:167–188, 1997.

[5] T. F. Chan, S.-H. Kang, and J. Shen. Euler's elastica and curvature based inpaintings. *SIAM J. Appl. Math.*, 63(2):564–592, 2002.

[6] T. F. Chan, S. Osher, and J. Shen. The digital TV filter and nonlinear denoising. *IEEE Trans. Image Process.*, 10(2):231–241, 2001.

[7] T. F. Chan and J. Shen. Variational restoration of non-flat image features: models and algorithms. *SIAM J. Appl. Math.*, 61(4):1338–1361, 2000.

[8] T. F. Chan and J. Shen. Mathematical models for local nontexture inpaintings. *SIAM J. Appl. Math.*, 62(3):1019–1043, 2001.

[9] T. F. Chan and J. Shen. On the role of the BV image model in image restoration. *Amer. Math. Soc. Contemporary Mathematics*, volume on Computational PDEs and Image Processing, to appear, 2003.

[10] T. F. Chan, J. Shen, and L. Vese. Variational PDE models in image processing. *Amer. Math. Soc. Notice*, 50:14–26, 2003.

[11] T. F. Chan and L. Vese. A level set algorithm for minimizing the Mumford-Shah functional in image processing. *IEEE/Computer Society Proceedings of the 1st IEEE Workshop on "Variational and Level Set Methods in Computer Vision"*, pages 161–168, 2001.

[12] A. Cohen, W. Dahmen, I. Daubechies, and R. DeVore. Harmonic analysis of the space BV. *Revista Matematica Iberoamericana*, submitted, 2002.

[13] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., New York, 1991.

[14] I. Daubechies. *Ten lectures on wavelets*. SIAM, Philadelphia, 1992.

[15] D. C. Dobson and C. R. Vogel. Convergence of an iterative method for total variation denoising. *SIAM J. Numer. Anal.*, 34(5):1779–1791, 1997.

[16] D. L. Donoho and I. M. Johnstone. Ideal spacial adaption by wavelet shrinkage. *Biometrika*, 81:425–455, 1994.

[17] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley & Sons, Inc., New York, 2001.

[18] S. Esedoglu and J. Shen. Digital inpainting based on the Mumford-Shah-Euler image model. *European J. Appl. Math.*, 13:353–370, 2002.

[19] L. C. Evans and J. Spruck. Motion of level sets by mean curvature. *J. Diff. Geom.*, 33(3):635–681, 1991.

[20] W. H. Fleming and R. Rishel. An integral formula for total gradient variation. *Arch. Math.*, 11:218–222, 1960.

[21] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intell.*, 6:721–741, 1984.

[22] W. Gibbs. *Elementary Principles of Statistical Mechanics*. Yale University Press, 1902.

[23] E. De Giorgi. Frontiere orientate di misura minima. *Sem. Mat. Scuola Norm. Sup. Pisa*, 1960-61.

[24] E. Giusti. *Minimal Surfaces and Functions of Bounded Variation*. Birkhäuser, Boston, 1984.

[25] U. Grenander. *Lectures in Pattern Theory. I. II. and III*. Springer, 1976-1981.

[26] W. Heisenberg. Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Zeitschrift für Physik*, 43:172–198, 1927.

[27] Y.-M. Jung and J. Shen. Geometric modelling and analysis of zebra stripe patterns from reaction-diffusion systems. *In preparation*, 2003.

[28] G. Kanizsa. *Organization in Vision*. Praeger, New York, 1979.

[29] L. Lengyel and I. R. Epstein. A chemical approach to designing Turing patterns in reaction-diffusion systems. *Proc. Nat. Acad. Sci. USA*, 89:3977–3979, 1992.

[30] L. Lengyel, G. Rabai, and I. R. Epstein. Experimental and modeling study of oscillations in chlorine dioxide-iodinacid reaction. *J. Amer. Chem. Soc.*, 112:9104–9110, 1990.

[31] K. Lin, S. G. Schirmer, and E. T. Camacho Wirkus. Chemical pattern formation in reaction-diffusion systems. Technical report, MSRI Summer program on Dynamics of Low Dimensional Continua, U.C. Berkeley, 1999.

[32] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 1998.

[33] D. Marr and E. Hildreth. Theory of edge detection. *Proc. Royal Soc. London*, B:207: 187–217, 1980.

[34] S. Masnou and J.-M. Morel. Level-lines based disocclusion. *Proceedings of 5th IEEE Int'l Conf. on Image Process., Chicago*, 3:259–263, 1998.

[35] Y. Meyer. *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations*, volume 22 of *University Lecture Series*. AMS, Providence, 2001.

[36] J. W. Milnor. *Topology from the Differentiable Viewpoint*. Princeton Univ. Press, revised edition, 1997.

[37] D. Mumford. Pattern theory: The mathematics of perception. *Int'l Congress Mathematicians (ICM) 2002*, vol. III, to appear.

[38] D. Mumford. Elastica and computer vision. In C. L. Bajaj, editor, *Algebraic Geometry and its Applications*, pages 491–506. Springer-Verlag, New York, 1994.

[39] D. Mumford. *Geometry Driven Diffusion in Computer Vision*, chapter "The Bayesian rationale for energy functionals", pages 141–153. Kluwer Academic, 1994.

[40] D. Mumford and B. Gidas. Stochastic models for generic images. *Quaterly of Applied Mathematics*, 59:85–111, 2001.

[41] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Applied. Math.*, 42:577–685, 1989.

[42] J. D. Murray. *Mathematical Biology*. Springer-Verlag, New York, 1993.

[43] M. Nitzberg, D. Mumford, and T. Shiota. *Filtering, Segmentation, and Depth*. Lecture Notes in Comp. Sci., Vol. 662. Springer-

Verlag, Berlin, 1993.

[44] A. V. Oppenheim and R. W. Schafer. *Discrete-Time Signal Processing*. Prentice Hall Inc., New Jersey, 1989.

[45] H. G. Othmer. *Mathematical Models of Biological Pattern Formation*. Lecture notes and book to appear, 2003.

[46] H. G. Othmer and L. E. Scriven. On the eigenvalues of the matrix pencil $A + \mu B$. *Z. Angew. Math. Phys.*, 24:135–139, 1973.

[47] J. E. Pearson. Complex patterns in a simple system. *Science*, 261:189–192, 1993.

[48] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Machine Intell.*, 12:629–639, 1990.

[49] T. Poggio and S. Smale. The mathematics of learning: Dealing with data. *Amer. Math. Soc. Notice*, 50(5):537–544, 2003.

[50] I. Prigogine and R. Lefever. Symmetry breaking instabilities in dissipative systems. *J. Chem. Phys.*, 48:1695–1700, 1968.

[51] G.-C. Rota and J. Shen. On the combinatorics of cumulants. *J. Comb. Theory (A)*, 91(1):283–304, 2000.

[52] L. Rudin and S. Osher. Total variation based image restoration with free local constraints. *Proc. 1st IEEE ICIP*, 1:31–35, 1994.

[53] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[54] J. Shah. Elastica with hinges. *J. Visual Comm. Image Rep.*, 13:36–43, 2002.

[55] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 623–656, 1948.

[56] J. Shen. Geometric inpainting and applications. *Proc. SPIE*, vol. 4792:102–113, 2002.

[57] J. Shen. On the foundations of vision modeling I. Weber's law and Weberized TV restoration. *Physica D*, 175:241–251, 2003.

[58] G. Strang and T. Nguyen. *Wavelets and Filter Banks*. Wellesley-Cambridge Press, Wellesley, MA, 1996.

[59] D. S. Thompson, J. S. Nair, S. S. D. Venkata, R. K. Machiraju, M. Jiang, and G. Craciun. Physics-based feature mining for large data exploration. *Computing Sci. Eng.*, 4(4):22–30, 2002.

[60] A. M. Turing. The chemical basis of morphogenesis. *Philos. Trans. R. Acad. Sci.*, B 237:37–72, 1952.

[61] G. Turk. Generating textures on arbitrary surfaces using reaction-diffusion. *Computer Graphics, SIGGRAPH '91*, 25:289–298, 1991.

[62] J. D. Watson and F. H. C. Crick. A structure for deoxyribose nucleic acid. *Nature*, 171:737–738, 1953.

[63] R. Williams. Xmorphia. Web Publishing at Center for Advanced Computing Research, Caltech, Pasadena, California. http://www.cacr.caltech.edu/˜roy/.

[64] A. Witkin and M. Kass. Reaction-diffusion textures. *Computer Graphics, SIGGRAPH '91*, 25(3), 1991.

[65] S. C. Zhu and D. Mumford. Prior learning and Gibbs reaction-diffusion. *IEEE Trans. Pattern Anal. Machine Intell.*, 19(11):1236–1250, 1997.

[66] S. C. Zhu, Y. N. Wu, and D. Mumford. Minimax entropy principle and its applications to texture modeling. *Neural Computation*, 9:1627–1660, 1997.